

# **Urban Divisions: Gentrification and Income Polarization in Ottawa, Canada**

Lazar Ilic

A thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the Doctorate in Philosophy Degree in Geography written under the direction of Michael Sawada and approved by the thesis committee consisting of:

Kenza Benali (University of Ottawa)

Jackie Dawson (University of Ottawa)

Anders Knudby (University of Ottawa)

And External Examiner

David Wilson (University of Illinois Urbana-Champaign)



Department of Geography, Environment and Geomatics  
Faculty of Arts  
University of Ottawa

© Lazar Ilic, Ottawa, Canada, 2023

## **Abstract**

This thesis examines urban inequalities in the Canadian context and focuses on Ottawa, the capital city.

Firstly, income inequality in the eight largest Census Metropolitan Areas in Canada, between 1971 and 2016, is examined through an analysis of income polarization and its spatial patterns. The middle-income group has declined across CMAs, while the low-income group has usually expanded. Concurrently, increasing spatial fragmentation is identified in every CMA examined. Local spatial autocorrelation identified clustering of high-income Census Tracts (CTs) suggesting that these areas are more resilient to fragmentation. Therefore, patterns of urban inequality are ones of an unequivocally disappearing middle-income population with an increasingly spatially fragmented urban income-scape.

At local levels, inequality is manifest in processes such as gentrification. To increase the spatial and temporal ability to monitor and map gentrification in a large city, artificial intelligence and Google Street View imagery were used to identify visual improvements to properties that are indicative of gentrification between 2007 and 2016. A deep Siamese convolutional neural network (SCNN) and VGG19 backbone was trained to recognize visual gentrification-like changes of properties over time. This deep mapping model achieved a 95.6% level of accuracy in identifying the visual signs of property improvement using 86110 georeferenced photographs of individual properties in Ottawa, Canada. Given that the residential/commercial property itself is the atomic object of gentrification, properties identified as having undergone a gentrification like visual change were mapped as points to produce kernel density maps that reduce noise and identify regions of high visual property change intensity (hot spots). The intensity of visual property improvements exhibited strong concordance with the

spatial pattern of building permits between 2011 and 2016. The results confirmed areas known to be undergoing gentrification and also presented areas where the process was not previously suspected of occurring.

Thirdly, a select set of census-based quantitative methods of modelling gentrification were compared between 2006 and 2016 at both the CT and the finer dissemination area unit of analysis. All models were tested against their ability to predict the density of GSV-points per unit residential area that were predicted by the Siamese deep learning model. For the CT level, two new regression models were created using all the census variables identified within the learned literature. An OLS multivariable regression model was created using backward stepwise regression, after which only age (youth), dwelling-value, income & occupation were retained. Residual spatial dependence in the OLS required a spatial linear model specification. A spatially lagged simultaneous autoregressive model (SAR Lag\_y) explained 57% of the variance in GSV-point density in Ottawa. Out of all the models tested, the SAR Lag\_y possessed the strongest spatial correlation with the original pattern of GSV point density as measured using Lee's L bivariate spatial autocorrelation statistic. A second model was produced using quasibinomial regression in order to predict the probability of a given CT being gentrified. That model achieved 91.7% accuracy. Out of the five reproduced models from the literature, one performed close to as well as our new models in predicting GSV-point density per unit residential area. While there is some agreement between models that purport to measure gentrification, there are considerable differences between models, which suggests that census-based gentrification measure should be locally focused.

## Acknowledgements

This thesis is the culmination of my PhD studies in Geography, and I remain both thankful and grateful to a number of individuals for various support over the years.

Firstly, I thank the Pajkic family who has hosted me numerous times in Toronto, both when I came to the city to play chess and to simply get away from Ottawa for a weekend every once in a while. This acknowledgement has been written on my last such trip. I especially thank my parents, Stanimir and Svetlana. Their support and parenting have been fundamental to my degrees thus far.

I thank a number of colleagues and friends at the university with whom I have played various intramural sports. They are too many to list given the amount of extracurricular activities (that helped me stay in shape), but some of them include Taite Dibdin, Sabrina Charland, Irina Georgijev, Louis-Philippe Morin, Caitlin Lapalme, Kaitlyn Anne Giffin, Caleb Chan, Ana Stefanovic, Stefan Lanceman and Polina Tarasenko.

I also thank various lab-mates and colleagues, who I have not yet mentioned, such as Amaury Zarzelli, Karine Saboui, Daniel Dumas, Fiona Davidson, Sarah Speight, Guillaume Guébin, Franny Delaney, and Mike Hackett. I thank my former housemates, especially Nguyen Hoang and Jennifer Spalding. I also thank Anna Klose for her friendship and for introducing me to Ritter Sport! Special thank you as well to Anoosh Noorizadeh Kollou and Alexandra Zemskova, who I have known since I came to this city. And, thank you to Gyöngyvér Kiss for providing me with some software, as well as friends Anastasia Grebenyuk, Andrew Re, Daniel Kostiuk and Odile Leclerc.

I thank both Susan Mowers from the Geographic, Statistical and Government Information Centre (GSG) at the University of Ottawa library, and Sherri Sunstrum from Maps, Data and Government Information Centre (MADGIC) at Carleton University's library.

I thank my department's administrative staff (Chantale Arcand and Philippe Villeneuve) for being helpful and the department in general, for funding in the form of TA-ships which helped finance part of my studies. I also thank Justice Giovanna Toscano Roccamo and the Social Planning Council of Ottawa, for having me work for them briefly. The funds from this work were handy and the results of this work, particularly an analysis of jury representation, were extremely interesting.

I also thank a number of professors in my department and beyond. Firstly, thanks to Keith Yearman, I embarked on my path in Geography. I thank Erick Howenstine and especially appreciate his work with cartograms. Of special thanks is Ian MacLachlan who I met at the 2014 Canadian Association of Geographers conference in St. Catharines. Also, Hugh Millward provided some important feedback for which I am grateful for. Of special note is Jean-Pascal Beaudoin and Thushara Hewage, under whose supervision I completed a pedagogical certificate in University Teaching from the University's Teaching and Learning Support Services. I also thank Luisa Veronis and Tamara Kotar for a number of useful consultations. I thank Elvin Wyly from the University of British Columbia, for much needed inspiration at times.

I greatly thank members of my thesis committee: Kenza Benali, Jackie Dawson, and Anders Knudby. I also thank the external examiner, David Wilson. Their input is necessary for finalizing this thesis.

Finally, and most importantly, this thesis would not have been possible without the assistance, guidance, and supervision of my advisor, Michael Sawada. Thanks to Michael, my

academic and research skills have significantly developed. The PhD endeavor of this magnitude would not have been realized without him, and my further success is in effect his success too.

Toronto – August 7, 2022.

## Preface

This PhD thesis is paper-based and consists of three papers. All were written in collaboration with my advisor, Michael Sawada, while one also involved Amaury Zarzelli. The three papers correspond to chapters two through four. The thesis starts with an introduction and ends with a synthesis/conclusions chapter.

The first paper corresponds to the second chapter. It was coauthored and is available at:

*Ilic Lazar, M. Sawada. 2021. The temporal evolution of income polarization in Canada's largest CMAs. PLoS ONE. V. 16(6): e0251430. <https://doi.org/10.1371/journal.pone.0251430>*

The second chapter began with my own ideas for what I initially thought would be two separate papers. The chapter has been greatly modified over time thanks to essential insight by Michael Sawada. Under Sawada's supervision this paper became worthy of publication, and the unexpected twists and turns in the process producing this paper were immense. During this process significant growth in my skills took place and were applied to this paper, thanks in large part to Sawada. At an early-stage Sawada identified the need to justify the usage of income as a focal variable, which is an important step that is unfortunately taken for granted by academics who have done work on income polarization. It is thanks to Sawada that I learned how to conduct bootstrapping in statistical analysis and improve my skills in the creation of quality figures. I conducted the fragmentation indices and under Sawada's direction improved one of these counts through a spatial intersection. Sawada conducted the spatial autocorrelation part of the analysis, and I processed those and other results to create select figures in the manuscript and its appendix.

There were two reviewers for this manuscript. The first identified themselves as Mario Polèse, from INRS (Institut national de la recherche scientifique) in Montreal. The second remained anonymous.

A mistake was identified in the paper shortly after publication. It was reported to the journal, but the journal is taking some time to publish the correction. The mistake was in the second table, which was added after an anonymous reviewer made a speculation. To avoid speculations, the second table was added. The table clarifies this ambiguity, but due to a transcription error discrepancies were created. Chapter two in this thesis depicts the correct figures.

The second paper was also published and is the third chapter of this thesis. It is available at:

*Ilic Lazar, M. Sawada, Zarzelli Amaury. 2019. Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. PLoS ONE. V. 14(3): e0212814. <https://doi.org/10.1371/journal.pone.0212814>*

This chapter was an immense collective effort which made a big splash in the media, as numerous international outlets reported on it in multiple languages. The work was conceptualized by Sawada. The project was so unique and new, to the extent which would not have been feasible in many other Geography departments due to a lack of necessary computing power which we had in the laboratory for Applied Geomatics and GIS Science (LAGGISS) at the University of Ottawa.

Investigation and background research was done by me, but the technical skills required were beyond my abilities and as such Sawada and Zarzelli were involved in some crucial work, particularly with coding, data curation and analysis. All three of us were involved in training the model by undertaking thousands of manual comparisons of properties side by side via an internet-based image comparison game. At one point I went out in the field and took hundreds of photographs of a neighborhood, which proved to be handy in our figures due to the journal PLoS ONE requiring all photographs to be CC-BY copyright. As such, we could not use the actual

Google Street View (GSV) photos in the published article and created example figures together with a links to the real figures on GitHub.

When the work on this chapter seemed finished, I presented the research at the 2017 Atlantic Canadian Association of Geographers Annual Conference, in Halifax. Astute advice from Hugh Millward, from Saint Mary's University, resulted in an additional component being added which strengthened the work: an analysis of building permits. I analyzed a total of 56,269 building permits, from 2011 and 2016, and pruned them down to 3986. These were later used to create kernel density maps.

The third chapter was reviewed by two anonymous reviewers. In the wake of publication, numerous media outlets reported on this research (Wu 2019; Umberto 2019a; Umberto 2019b; Umberto 2019c; Gagnon 2019; Josset 2019; uOttawa 2019; PLOS 2019a; PLOS 2019b; PLOS One 2019; Public Library of Science 2019; Vandette 2019; ResearchCareer 2019; Science Friday 2019; WBEZ Chicago 2019; Pritchard 2019; CBC 2019; Education News Canada 2019; The Map Room 2019; Mario 2019; Zarzalejos 2019; Rosales 2020; Glover 2019; Blewtt 2019; Brackley and Walker 2019).

The fourth chapter of this thesis contains the third paper. While it is in final form, it will be submitted at a later date to a peer reviewed journal that has a respectable impact factor. The citation for this paper could be:

*Ilic Lazar, M. Sawada. Forthcoming. Modeling and mapping AI-derived gentrification data with census data in a large Canadian city.*

I presented an early version of this chapter at the 2019 Canadian Association of Geographers Annual Conference in Winnipeg and received the ESRI Award for Best Graduate Student Paper Presentation. Since then, the chapter had changed substantially for the better under the auspices of Sawada. The paper represents a moment of personal growth for me, as Sawada's

guidance has taught me how to properly go about dasymetric mapping and how to conduct various regression analysis.

The paper begins with my thorough examination of previous quantitative gentrification literature which utilized census data. After identifying five studies that take interval/ratio approach to assessing gentrification. Then, I reproduced the studies using the published methods within the geographic extents of the inner-city and inner-suburbs of Ottawa.

Subsequently I conducted various regression analysis to test the predictive power of individual census variables and produced methodologies. I ran multiple regression, stepwise regression and logistic regression. Michael identified certain weaknesses in my procedures in order to produce stronger more meaningful spatial auto-regression analyses.

## Table of Contents

<b>Abstract</b> .....	ii
<b>Acknowledgements</b> .....	iv
<b>Table of Contents</b> .....	xi
<b>List of Tables</b> .....	xiv
<b>List of Figures</b> .....	xv
<b>Chapter 1. Introduction</b> .....	1
<b>Philosophical Underpinnings</b> .....	1
<b>Situating the Research</b> .....	3
<b>Layout of Study</b> .....	4
<b>Gentrification</b> .....	10
<b>Contributions to Knowledge</b> .....	14
<b>Chapter 2. The temporal evolution of income polarization in Canada’s largest CMAs</b> .....	17
<b>Introduction</b> .....	17
<b>Justifying income as a measurement of inequality</b> .....	18
<b>Inequality vs polarization</b> .....	21
<b>The Canadian context</b> .....	21
<b>Geographic approaches to measuring income polarization at the census tract level in the Canadian urban context</b> .....	22
<b>Methodology</b> .....	24
<b>Study areas</b> .....	24
<b>Data</b> .....	25
<b>Population based weights</b> .....	28
<b>Defining and measuring the middle</b> .....	28
<b>Mapping</b> .....	30
<b>Fragmentation analysis</b> .....	31
<b>Spatial autocorrelation</b> .....	34
<b>Results</b> .....	37
<b>Discussion</b> .....	45
<b>Conclusion</b> .....	55
<b>Supporting information</b> .....	56
<b>Acknowledgments</b> .....	56

<b>Chapter 3. Deep mapping gentrification in a large Canadian city using deep learning and Google Street View</b> .....	57
<b>Introduction</b> .....	57
<b>Deep mapping and the built environment</b> .....	61
<i>Google Street View (GSV) in auditing the urban environment</i> .....	61
<i>Machine mapping and deep mapping</i> .....	62
<b>Methodology</b> .....	65
<i>Study area</i> .....	66
<i>Training data</i> .....	67
<i>Siamese model</i> .....	69
<i>Detection and mapping</i> .....	72
<i>Independent validation</i> .....	74
<b>Results</b> .....	75
<b>Discussion</b> .....	76
<b>Conclusion</b> .....	85
<b>Supporting information</b> .....	87
<b>Acknowledgments</b> .....	87
<b>Chapter 4. Modeling and mapping AI derived gentrification data with census data in a large Canadian city</b> .....	88
<b>Aspects of scale</b> .....	90
<i>Independent measures of gentrification</i> .....	91
<b>Background</b> .....	92
<i>Neglect of quantitative methods</i> .....	92
<i>Geographic units and geographic contexts</i> .....	93
<i>Census variables</i> .....	95
<i>Quantitative Methods</i> .....	100
<b>Methodology</b> .....	102
<i>Study Area and Temporal Dimension</i> .....	102
<i>Reproducing Methodologies</i> .....	104
<i>Comparing CT and DA results of reproduced methodologies</i> .....	107
<i>Model building</i> .....	109
<b>Results</b> .....	117
<i>Reproducing Methodologies</i> .....	117

<i>Examining Data Variability</i> .....	118
<i>Spatial Autocorrelation</i> .....	120
<i>Correlations with dependent variable</i> .....	121
<i>OLS and linear spatial regression models</i> .....	123
<i>Quasibinomial model</i> .....	127
<i>Quantitative Inter-model comparison</i> .....	129
<b>Discussion</b> .....	133
<b>Conclusion</b> .....	140
<b>Chapter 5. Conclusion</b> .....	143
<b>Summary</b> .....	143
<b>Contributions to Geography</b> .....	146
<b>Chapter 6. References</b> .....	156
<b>Appendix I. Examining income trajectories and income polarization in Ottawa, 1971-2016.</b> .....	185
<b>Appendix II. Supplementary information for ‘The Temporal Evolution of Income Polarization in Canada’s Largest CMAs’</b> .....	208
<b>Appendix III. Additional CMA Data</b> .....	234
<b>Appendix IV. Supplementary information for ‘Deep mapping gentrification in a large Canadian city using deep learning and Google Street View’</b> .....	235
<b>Appendix V. Supplementary information for ‘Modeling and mapping gentrification with census data in a large Canadian city’</b> .....	247

## List of Tables

Table 1.1. Gentrification typologies .....	13
Table 2.1. Slopes of the three income groups (low, middle & high) and their associated 95% confidence intervals (CI) obtained via non-parametric bootstrapping of household income data. ....	41
Table 2.2. Slopes of the three income groups (low, middle & high) and their associated 95% confidence intervals (CI) obtained via non-parametric bootstrapping of individual income data. ....	41
Table 3.1. Results of model after fine-tuning. ....	59
Table 4.1. Gentrification studies which used census data and quantitative approaches.....	94
Table 4.2. Census variables utilized to measure gentrification. ....	96
Table 4.3. Approaches to measuring gentrification with census data. Dates .....	101
Table 4.4. Spatial autocorrelation for reproduced methods and new regression model predictions based on 9,999 permutations of Moran's I. ....	121
Table 4.5. Moran's I for independent variables is based on 999 permutations at the CT level..	123
Table 4.6. OLS multivariate model and SAR Lag_y (spatially lagged simultaneous autoregressive model estimated by maximum likelihood). ....	125
Table 4.7. SAR Lag_y model effects.....	127
Table 4.8. Quasibinomial model output.....	128
Table 4.9. Quasibinomial regression model. ....	129

## List of Figures

Figure 1.1. Structure of main contributing chapters. ....	10
Figure 2.1. Locations of Canada’s eight largest CMAs.....	25
Figure 2.2. The percentage of households in low, middle and high-income classes, in Canada’s largest eight CMAs. ....	40
Figure 2.3. Conventional maps (a) and Cartograms (b) of the temporal evolution of the income structure of Canada’s eight largest CMAs. ....	43
Figure 2.4. Results of spatial autocorrelation / join count analysis. ....	44
Figure 2.5. Diagrams showing fragments and fragmentation indexes for Canada’s eight largest CMAs. ....	52
Figure 3.1. The study area contains the oldest building stock in Ottawa and is within the Greenbelt.....	66
Figure 3.2. Training data collection web interface. ....	69
Figure 3.3. Siamese CNN architecture using two VGG19 branches.....	71
Figure 3.4. Example of detecting gentrification-like change within a sequence of GSV images at the same geographic location using SCNN-FC-8.....	73
Figure 3.5. Comparison between model detections and building permits.....	75
Figure 3.6. SCNN-FC-8 detections over time. ....	77
Figure 3.7. Examples of SCNN-FC-8 detections.....	81
Figure 4.1. Structure of Ottawa’s Dissemination Areas within Census Tracts (2016) for areas which are bounded by the Greenbelt and Airport.....	104
Figure 4.2. Reproduced methodologies and frequencies.....	118
Figure 4.3. Maps of DA variability within CTs.....	120
Figure 4.4. Correlation matrix between thirteen variables used in the literature and GSV-point density.....	122
Figure 4.5: A) Observed density of GSV-points per unit residential area; B) Predicted density of GSV-points per residential area from SAR Lag_y model; C) A in deciles; D) Probability of gentrification predictions from quasibinomial model.....	130

Figure 4.6. A) Pearson's R between mapped outcomes at the CT level from all models; B) Lee's L bi-variate spatial correlations. .... 132

**List of Abbreviations/Acronyms**

AIC	Akaike Information Criterion
API	Application Programming Interface
BGR	Blue-Green-Red
CMA	Census Metropolitan Area
CT	Census Tract
CI	Confidence Interval
CNN	Convolutional Neural Network
DA	Dissemination Area
ED	Edge Density
FOV	Field Of View
FI	Fragmentation Index
GSV	Google Street View
JC	Joins-Count
KDE	Kernel Density Estimation
MAUP	Modifiable Areal Unit Problem
NHS	National Household Survey
OLS	Ordinary Least Squares
SCNN	Siamese Convolutional Neural Network
SES	Socio-Economic Status
SA	Spatial Autocorrelation
SDM	Spatial Durbin Model
SSOI	Systematic Social Observation Instruments
TCM	Three City Model
TCP	Three City Project
VIF	Variance Inflation Factors

## **Chapter 1. Introduction**

### **Philosophical Underpinnings**

This research examines urban inequalities in Canada and is thus motivated by social justice – more specifically, urban social justice. The discipline of Geography is well placed for such research and has been approached from a social justice angle since the 1960s (Smith 1994, 4). Justice itself is based on addressing injustices and hinges on perceived fairness (Eckstein and Wickham-Crowley 2003, xi). Matters of justice are very much urban issues: most people live in urban areas and urban society is mired with much tension, strain, conflict and difference (Swyngedouw 2006, 80). Soja (2010, 1) opines that the spatial aspects, or spatiality, of justice are a quintessential element of justice. Therefore, the field of Geography is well situated to address matters of social justice from a spatial vantage.

As, matters of social justice are related to fairness and equity (Hay 1995) and as social justice is confronted with significant challenges relating to urban inequality under capitalism (Smith 1994, 150), the present research can be situated under social justice elements such as a) human rights, b) equity and c) diversity. Through the examination of urban differences, spatial inequalities are better understood, which can result in a more equitable distribution of services to vulnerable populations (Ilic and Sawada 2021, 8).

An important means by which social justice could be gauged is the equality of opportunity, which is spatial in nature (Israel and Frenkel 2018). This thesis is not focused with aspects of social justice such as access to healthcare, transportation, sustenance, public services, and so forth. This work centers on defining spatiality of inequality in major urban centers and then focuses on one aspect of spatial inequality: gentrification.

In the results of the second chapter, on income inequality, areas are identified in which other elements of justice can be explored in future studies. For example, do inequalities in health care access coincide with income inequality within a city? Additionally, questions on housing are ever more pressing. For example, there is much discussion about the “right to housing” (Hulchanski and Leckie 2000; Leijten and de Bel 2020; Segal 2020; Kreide 2022). In terms of equity, or impartiality, the notion of fairness and the distribution of previously mentioned aspects of urban development arise. Equity therefore intersects with diversity, a social justice tenant, as cities are comprised of diverse neighbourhoods and populations of a wide range of socioeconomic backgrounds. However, in urban studies diversity is an ambiguous word (Hodge 1980), and as such gentrification can further polarize and divide society by coopting notions of diversity (Bridge, Butler, and Lees 2012).

A significant part of this thesis (third and fourth chapters) deals with gentrification, which is especially related to human rights, because a key feature of gentrification is displacement of less affluent groups. Gentrification can be seen as a spatial injustice (Zhu and Guo 2022), because it deprives people of their previous neighborhood (LeGates and Hartman 1986) and therefore is a quintessential element of urban social justice. Displacement, a requisite component of gentrification, results in the displaced being deprived of their previous space or home, which Marris (2015, 57) notes as being extremely disruptive as it fractures crucial spatial and social relationships that cannot be easily reestablished in new settings. The loss of one’s home and neighbourhood can be seen as worse than the loss of close family (Smith 1994, 276), which exemplifies the aforementioned three social justice tenants.

## **Situating the Research**

In recent times there is a heightened concern about socioeconomic inequality around the world. Since the 1980s, inequality has deepened more than even critics have foreseen (Smith 2008, 242) and has only been exacerbated since the SARS-COVID-19 pandemic, as the divide between haves and have-nots became unequivocally apparent. Inequality has always been at the forefront of urban social justice, which is where this research is situated.

This thesis contributes to research on socioeconomic polarization, which is of interest to a variety of scholars from fields such as Anthropology, Economics, Geography, Sociology and Political Science. The term socioeconomic polarization refers to an expanding separation between those at the higher and lower end of the socioeconomic spectrum. Socioeconomic polarization can be seen as segregation through factors including, but not limited to, individual income, housing value, and poverty.

The initial motivation behind this thesis was from work by urban geographers in Canada who were increasingly focusing on urban inequality in the 2010s. They comprised the Neighborhood Change Research Project, which was in effect spawned by Hulchanski (2007; 2010) and based around his “Three City Model”. While researchers in his group applied his method in their cities, none did so in Ottawa. My early examination of inequality in Ottawa was the first step in this thesis, and an unpublished version of this work is available in Appendix I.

This thesis is divided into three main sections, each corresponding to a paper that has either been published or will be submitted for publication after taking into account comments from examiners post-defense. Each of the three sections constitute a significant contribution to research on urban inequality and thus, in a more general sense, social justice. Supplemental material is also provided for each of the chapters after the references at the end of the thesis

document.

## **Research Objectives**

For decades, the development of tools and techniques used to measure urban divides have been somewhat lacking. In particular, measuring gentrification has been somewhat elusive, as most studies have focused on avenues of qualitative investigation in the form of field surveys, discourse analysis, literature reviews, and interviews.

The research in this thesis is of a quantitative nature and aims to broaden the scope of inquiry by producing novel techniques to assess, quantify, track, and monitor the process of spatial change in urban environments.

This thesis is based on the following questions:

- (1) Is income-polarization taking place in the largest Canadian CMAs? What are the spatial patterns of this phenomena in different cities?
- (2) Can the spatial and temporal limitation of census data for measuring gentrification be surmounted through the usage of street level imagery and the use of deep learning methods?
- (3) To what extent do quantitative models of gentrification based on census data agree with each other and how well do they measure gentrification?

## **Layout of Study**

The second chapter is a multi-case study that examines income inequality in Canada's eight largest Census Metropolitan Areas (CMAs). This chapter was published in PLoS ONE (Ilic and Sawada 2021) and builds from Appendix I through an examination of how Canadian urban space has evolved between 1971 and 2016. A foundational base for the methodology of this chapter comes from MacLachlan and Sawada (1997), rather than the research group around

Hulchanski which was the basis for the endeavor in Appendix I. The methods presented in the second chapter are novel and are a conceptual contribution for measuring income polarization.

The second chapter starts by clarifying an ambiguous assumption in social inequality research, namely, income as the single/sole/focal variable to study inequality is justified. Previous studies made no attempt to do so: they assume a priori that this census variable is an appropriate measure, not just of income but by extension income as a suitable measure of inequality in general. That assumption lays the foundation for an empirical examination of changing patterns of income structures in Canada's largest cities. This endeavor begins with establishing low-, medium- and high-income groups, and compares the trends of these groups between 1971 and 2016. Such trends uncover if income polarization is indeed occurring in the select CMAs. There is a focus on household income, but also a limited comparison to individual income figures. This chapter also examines spatial fragmentation of income groups and growing urban divides, which empiricizes claims of some urban geographers, that the postmodern city is increasingly fragmented. Additionally, spatial autocorrelation in the form of a joins count is conducted to determine whether there are differences through time in the spatial clustering of low-, medium-, and high-income groups.

Are the same patterns in the largest CMAs present in smaller CMAs in other places in Canada? This question is answered in Appendix III, where trends of aforementioned income groups, between 1971 and 2016, are compared side by side for both individual and household measures of incomes in a total of 22 CMAs.

The second chapter is in essence a broader examination of uneven development in cities. An element of uneven development in mature cityscapes is gentrification (Smith 2008, xvi). Some feel that gentrification is linked to income-inequality more so than uneven development

(Chapple 2017), though there is not consensus in the academic community for this claim.

Gentrification has received an increasingly large amount of attention in recent years, as is evidenced by the sheer volume of research being published in both academic journals and books (S1 in Appendix V). However, a serious gap in terms of measuring gentrification persists, despite the fact that gentrification and policies which support it are at the forefront of forces driving urban restructuring (the other force being continued suburbanization).

The third chapter addresses that gap with conceptually new methods that use visual changes over time to properties to assess the geographic extent of gentrification at fine spatial detail across a major metropolitan area. The third chapter was published in PLoS ONE (Ilic, Sawada, and Zarzelli 2019) and it employs advanced cutting-edge methods of AI and computer vision to measure gentrification. Gentrification is difficult to quantify and measure precisely in space. Most researchers have relied on census data to assess gentrification, and some lament how the lack of data between years does not allow for a precise identification of the extent or timing of gentrification processes (Moskowitz 2017). Advances in technology and the emergence of Big Data allows for new methods of assessing gentrification.

Since the mid-2000s, significant amounts of street level imagery (Big Data) have been released on interactive virtual platforms such as Google Street View (GSV). GSV allows for the harvesting of data in many inter-census years. In the third chapter, individual properties, the atomic unit of gentrification, as seen from the street, are analyzed using artificial intelligence (AI) modelling using new deep learning techniques in computer vision. AI modeling allows for a large quantity of visual data to be analyzed for each property in a city. The results of the AI model are point locations indicating where properties exhibited the visual signs of gentrification, and these were then compared to data obtained from manually classifying individual building

permits for the same time period obtained from the City of Ottawa's databases. The strong concordance between the results obtained from machine learning and building permit analysis shows the promise of using new AI technology and Big Data for identifying gentrification processes that range from mature to initial.

This research added to scientific knowledge as it was the first to propose the use of computer vision AI in gentrification research and provided a new way of measuring and mapping gentrification in urban areas, allowing aggregation to any arbitrary geographic scale (blocks, dissemination areas (DAs), census tracts (CTs), custom geographies, etc.). Finally, the research in this chapter shows how new regions of gentrification are identifiable that may not be on the 'general radar' of researchers, the municipality and/or citizens/developers. From a social justice perspective, the rapidity by which gentrification in different levels of process can be identified allows elected officials to mitigate potential issues that negatively impact equality in urban regions through tools like planning and zoning, for example.

The work on gentrification in the third chapter was chronologically completed first, before that of the second chapter for the simple reason that the field of deep learning was still new and the academic literature was bereft of such AI applications outside of computer science - unlike today. Thus, it was important to be the first to establish this method in a field of study, namely gentrification, that is largely bereft of quantitative approaches in general. In that regard, the third chapter represents the most significant contribution to advancing the state of knowledge in gentrification studies and provided a kick-start to the field by renewing hope of quantifying the process. The paper has garnered more than 57 citations since publication in 2019.

It is important to note that in the third chapter only the visible signs of gentrification are examined. This differs from what would constitute a full measure of gentrification, and as such

remains an imperfect proxy of gentrification. The intention was to use this approach on additional urban areas that were examined in the second chapter in order to assess the role that gentrification plays within the spatial mosaic of income polarization and inequality. However, this was not possible because in July of 2016, Google's updated terms of use made it prohibitively expensive (increased prices 5000% overnight) to use their street level data. Therefore, from that point on it was no longer feasible to reproduce a comparative multi-city approach as was done in the second chapter and the overall work was drawn back to the City of Ottawa where we had sufficient data to continue to look at inequality in new and novel ways. Moreover, other avenues of street-level imagery, such as Mapillary.com (free crowd-sourced GSV), do not have the historical timeline that google provides. Thus, such widespread application to other cities in Canada will not be easily possible for some time.

The fourth chapter builds on the third by examining how well census data can measure gentrification as defined in the third chapter; strictly speaking, the visual signs of gentrification. The independent gentrification dataset produced by the AI deep learning predictions in the third chapter provided a set of point locations where gentrification-like visual changes were found in the Street-View record for properties in Ottawa. Those results provided a gentrification dataset that is completely independent of census-based models that have almost exclusively been used in the scant literature on quantitative measures of the phenomena.

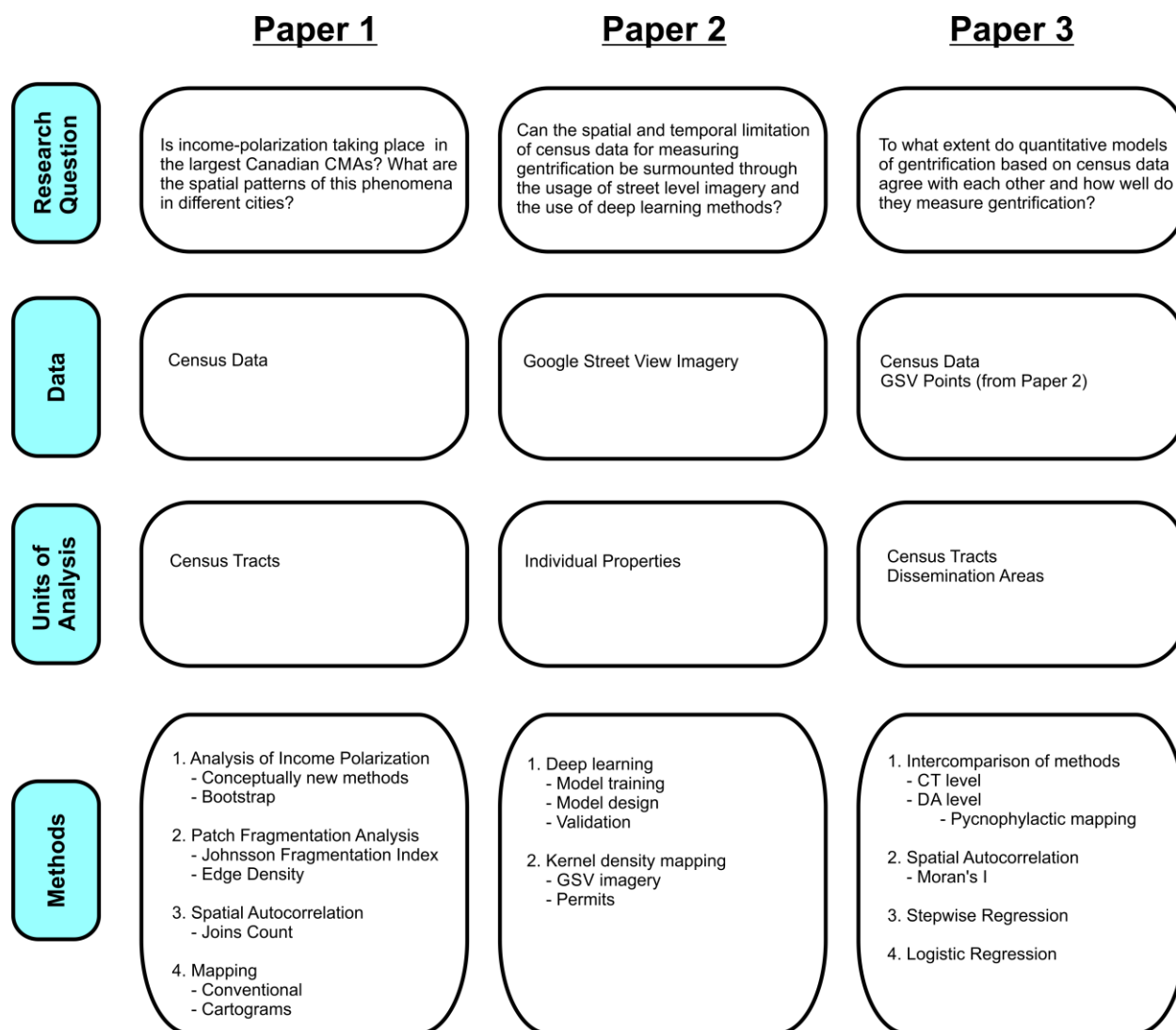
The fourth chapter starts with a review of the methods used by other researchers who employ census data to measure gentrification and identifies the census variables used across the different quantitative approaches. Five quantitative methodologies were reproduced at the CT geographic unit of analysis. Subsequently, these methodologies were applied to DAs, which are a much finer nested level of geography than the CT. The examination of gentrification at a sub-CT

unit of analysis is scant, and this may be the first academic study which examines gentrification at the DA level.

Next, the set of quantitative census-based variables were tested against the results of the third chapter. Using the census variables, informed by researchers using quantitative measures in the literature, two new statistical models to predict gentrification were developed using a combination of OLS regression, spatial simultaneous autoregression and a quasibinomial model. Results from the third chapter were employed to assess how the five reproduced gentrification methods and the two new models performed in relation to an independently derived picture of gentrification from the deep learning model. Therefore, chapter four contributes twofold to a better understanding of how census data can be used to examine gentrification through elucidating the consistency and success of different models of gentrification and also providing an inter-model comparison.

These three components produce a strong quantitative examination of urban inequality through the a) development of sound means by which to assess income polarization, b) development of means by which to assess gentrification at a fine spatial scale beyond what is possible with census data, and c) through a better grasp of how to use census data to study gentrification. The three chapters should be of much interest to urban geographers who study urban inequality and gentrification.

The layout of the three main components of the thesis is depicted in the form of a diagram (Figure 1.1.) which highlights the main differences in terms of research questions, types of data, units of analysis and methods.



**Figure 1.1. Structure of main contributing chapters.**

## Gentrification

Much of this thesis deals with gentrification and this term deserves elaboration. While gentrification is an increasingly popular word, ambiguity continues to enshroud it. The term was coined by Ruth Lazarus Glass (1964, xviii-xix) to describe urban changes taking place in the inner London neighborhood of Islington:

*“One by one, many of the working class quarters of London have been invaded by the middle classes-upper and lower. Shabby, modest mews and cottages-two rooms up and two down-have been taken over, when their leases have expired, and have become elegant, expensive residences. Larger Victorian houses, downgraded in an earlier or recent period-which were used as lodging*

*houses or were otherwise in multiple occupation-have been upgraded once again. Nowadays, many of these houses are being sub-divided into costly flats or 'houselets' (in terms of the new real estate snob jargon). The current social status and value of such dwellings are frequently in inverse relation to their size, and in any case enormously inflated by comparison with previous levels in their neighbourhoods. Once this process of 'gentrification' starts in a district, it goes on rapidly until all or most of the original working class occupiers are displaced, and the whole social character of the district is changed. There is very little left of the poorer enclaves of Hampstead and Chelsea: in those boroughs, the upper-middle class take-over was consolidated some time ago. The invasion has since spread to Islington, Paddington, North Kensington-even to the 'shady' parts of Notting Hill-to Battersea, and to several other districts, north and south of the river. (The East End has so far been exempt.) And this is an inevitable development, in view of the demographic, economic and political pressures to which London, and especially Central London, has been subjected."*

Wyly has described gentrification as “*the class transformation of urban space*”, while Hackworth (2002) has described gentrification as “*the production of urban space for progressively more affluent users*”. I would define gentrification as “class-based displacement through which individuals are displaced by means of higher purchasing power”. At its core gentrification is a replacement process (Schulman 2013) and resembles colonialization (Atkinson and Bridge 2005).

Gentrification did not begin with Glass first noticing it, and Glass herself did not focus her research on Gentrification, and even rarely came back to this word despite the proliferation of this term in subsequent decades. Research has identified gentrification prior to her work (Smith 1996; Reick 2018). Gale (2021) has described gentrification before World War II (and until the 1970s) as “embryonic gentrification.” Perhaps the earliest description of gentrification was a “reverse” process (of filtering), described by Hoyt (1939, 118-119) in the then seminal text on urban models. Other terms, all of which gentrification has replaced, include: Whitepainting (in Canada), Trendification (in Australia), Chelseafication (in London and/or Capetown) and Brownstoning (in New York City) (Williams 1976; Weston. 1982; Ley 1983). These toponyms contrast to gentrification in that they are localized and do not focus on the “gentry”, from which the term Gentrification stems from. The “gentry” are persons of gentle birth and are constituted

of a landowning elite (London and Palen 1984, 7). In contrast, it is noted that gentrifiers often come from places within cities or urban areas (Zukin 1987, 131) and as such are not predominantly from suburban or rural areas.

Gentrification has often been used to describe a process taking place in the inner city. Definitions of gentrification continued to explicitly describe it as such into the 1990s, and those measuring gentrification often saw it as a process taking place in only disadvantaged neighborhoods within the inner city (Zukin 1991, 186–187; Hammel and Wyly 1996) .

In the late 1970s, Gentrification became more nuanced, as it began to be conceptualized in a boarder manner to include more nuanced examination of various restructuring (Sassen 1991, 255). New varieties of gentrification were identified, the first being “rural gentrification”. Over time, in the 1990s and 2000s more types were identified and were eventually termed “mutations of gentrification” (Lees, Slater and Wyly 2008). More recently, gentrification expanded to non-urban realms. Gentrification of the mind (Schulman 2013), food (Alkon, Kato and Sbicca 2020), media (Cooper, 2014), and even the internet (Lingel 2021) are examples of how broad this term has become. Gentrification as such can be divided into urban (Table 1.1) and non-urban gentrification.

Table 1.1. Gentrification typologies.

<b>Urban Gentrification</b>	<b>Non-urban Gentrification</b>
Climate gentrification	Coastal Gentrification
Commercial/retail gentrification	Educational gentrification
Family gentrification	Rural gentrification
Green/eco/environmental gentrification	Wilderness gentrification
Hyper gentrification	Gentrification of cuisine / food
Mega gentrification	Gentrification of the media
Migrant-led gentrification	Gentrification of the mind
Minority-led gentrification	Gentrification of the internet
New-build gentrification	
Religious gentrification	
Slum gentrification	
Studentification	
Suburban gentrification	
Super gentrification	
Tourism gentrification	
Transit gentrification	

The research herein examines urban gentrification. The visual aspect of property improvement is almost synonymous with gentrification (Hammel and Wyly 1996, 251) and as such this is a common feature of gentrification irrespective of what type of urban gentrification one examines. While many types of gentrification are recognized, their different forms are not focused on. Instead, the focus lies on aspects of measuring gentrification, especially through visual aspects.

The research herein provides movement towards identifying gentrification at finer spatial scales than in the past, and offers potential inspiration for those who are alarmed by this process. For example, research in the last several years shows an emergence of interest in how aspects of health are intertwined with gentrification. Health applications to gentrification consist of three interrelated points: i) displacement and the stress associated with it, ii) crime, and iii) uneven development.

Those displaced by gentrification generally have worse mental health (Lim et al 2017), and this is often a consequence of elevated stress levels. Even the fear of displacement can induce stress (Guzman, Bhatia and Durazo 2005). While some research suggests that gentrification can reduce certain types of crime (McDonalds 1986), other research shows that gentrification can also induce crime (Anguelovski et al. 2018, 45). Gentrification efforts have also actively partaken in the marginalization of previous residents, as these vulnerable populations have been deemed undesirable through criminalization and labeling as disorder that should be contained (Helms, Atkinson and MacLeod 2007, 271–2). Stigma in particular can be conferred onto previous vulnerable residents, which can lead to pathologization (Smith 2010). Overall, urban regions do benefit from gentrification, even though at times certain neighborhoods might seem to be benefiting (Gibbons, Barton and Brault 2018).

Given the negatives attributes associated with gentrification, enhanced quantitative tools give insight into better understanding gentrification and this can result in means by which to attempt to manage, contain or mitigate the process. Such endeavors could be effective if places with populations in need of assistance can be easily identified. Likewise, movement towards eventual prediction of gentrification can result in pre-emptive action to mitigate the negative effects of this process.

### **Contributions to Knowledge**

While certain studies examine urban inequalities (such as gentrification) in Ottawa, the nation's capital remains understudied in terms of socioeconomic inequalities when compared to larger cities in Canada such as Montreal, Toronto, and Vancouver (MTV). The research herein addresses several gaps in the literature and in doing so contributes to urban social justice by helping delineate the spatial patterns of urban inequality, namely income polarization and

gentrification. Chapters two through four each aid in establishing a means by which to go about the identification of processes for which there is little consensus in terms of measurement. This thesis also goes a long way in terms of bridging the divide between urban geography and geomatics. For example, to date there is only one urban geographer that wrote textbooks which take a quantitative approach, the last of which has been out of print since the mid-1990s (Cadwallader 1985; 1996). Chapters two through four, herein, contribute in unique ways to geographic knowledge.

First, this research establishes a definitive means of measuring income polarization and its spatial structure in Canadian CMAs. This is achieved in the second chapter, wherein robust means by which to examine income polarization are established. These methods aid in advancing research on urban socioeconomic divisions.

Secondly, in the third chapter, a new and novel means by which to measure gentrification, without being dependent on sporadic data from the census, is demonstrated. Hence, a novel way in which to gauge the extent of gentrification in time and space is demonstrated.

Thirdly, the fourth chapter examines mapping of gentrification with census data through new methods and examines various approaches to quantify this process. The first quantitative comparison between interval/ratio census-based gentrification measures is completed together with an examination of gentrification at a finer level of geography than is typical.

These three elements constitute a significant expansion to the understanding of urban change and as such advance both Geography and Urban Studies, while examining Ottawa in particular, though spatially delineating inequality and through mapping and identifying gentrification in both time and space.

In the Canadian context, gentrification has not been seriously analyzed in a quantitative manner in decades. For gentrification studies, an analysis of census data should aim to achieve an identification of areas which have undergone gentrification. However, such analysis needs to be cognizant of the more nuanced understanding of how the definition of gentrification has changed in the last few decades. Therefore, the research herein adds to increased prominence of urban inequality research that is present post 2010 amongst social scientists from various fields (Nijman and Wei 2020).

Many scientists aim to model processes which take place, due to the predictive value that such modeling gives. This study examines such modeling and develops new techniques to measure urban divides and gentrification. The recent expansion of machine learning paves the way for addressing the research questions posed herein while providing novel contributions to the field of urban social justice in general.

## **Chapter 2. The temporal evolution of income polarization in Canada's largest CMAs**

**Abstract:** Income polarization is a pressing issue which is increasingly discussed by academics and policymakers. The present research examines income polarization in Canada's eight largest Census Metropolitan Areas (CMAs) using data at the census-tract (CT) level between 1971 and 2016. Generally, there are significant decreasing trends in the middle-income population with simultaneously increasing trends in low-income groups. The high-income groups have been relatively stable with fewer significant increasing population trends. Using conventional mapping and cartograms, patterns of the spatial evolution of income inequality are illustrated. Every CMA examined contains an increasing trend of spatial fragmentation at the patch level within each CMA's landscape mosaic. The results of a spatial autocorrelation analysis at the sub-patch, CT level, exhibit significant spatial clustering of high-income CTs as one process that dominates the increasingly fragmented landscape mosaic.

### **Introduction**

Inequality is an increasingly pressing issue. As a popular element of social justice (Sen 2000), inequality is exacerbated through uneven development at a multitude of spatial scales, exposing the contrasts between haves and have-nots. Whereas some studies concerning inequality focus on an international or national scale (Singh, Dhumale, and Cornia 2004), recent research increasingly includes local income polarization at regional and urban scales (Breau 2015; Buitelaar, Wterings, and Ponds 2018; Hulchanski 2007; Madariaga, Martori, and Oller 2019; Mondor et al. 2018; Qiu and Zhao 2019). Income polarization is a process where the middle-income group decreases in size while the two opposing poles of low-income and high-income groups expands in size.

In the Canadian context, a network of researchers who are part of the Neighbourhood Change Research Project conducted a number of studies on income inequality in select Canadian cities, the results of which are summarized in a recent publication (Grant, Walks, and Ramos

2020). Herein, we refer to, their methodological approach as the Three City Project (TCP), due to their utilization of what has come to be known as the Three City Model (TCM).

We posit that research on income polarization requires an examination of time-slices in order to determine how the spatial pattern of income groups change over time. A time-slice map portrays the spatial pattern of the income groups (low, middle, high) for a particular census year. Time-slice maps of income structure build on work by MacLachlan and Sawada (1997), who identified an income range as a methodological necessity for examining income polarization.

This present research addresses income polarization, while focusing on the following questions: Given that previous research suggests that income polarization has increased in certain Census Metropolitan Areas (CMAs) over the past few decades (MacLachlan and Sawada 1997; Hulchanski 2010; David Ley and Lynch 2012; Grant, Walks, and Ramos 2020), how has income polarization evolved in Canada's eight largest CMAs between 1971 and 2016? To what extent are low-income and high-income groups changing at the expense of the middle-income group? How is income inequality manifested spatially within CMAs? For example, are income groups more fragmented within the CMAs and, if so, is there spatial clustering of income groups?

### **Justifying income as a measurement of inequality**

Inequality is manifested in numerous ways, including but not limited to happiness, wealth, opportunities, achievements, needs, freedoms, rights, quality of life, and so forth (Sen 1992). For the purposes of this study, the type of inequality being examined is income inequality.

There are numerous multivariate indices of inequality and several inequality indices include, income in addition to various socioeconomic variables (Desilver 2015; Development Finance International and OXFAM 2018; Guan 2017; McKay 2002; Sergeant and Firth 2006)..

Furthermore, income is a dominant variable in measures of socio-economic status (SES) (Davies 1984; Townshend and Walker 2002).

Justifying income as a sole measure of inequality is usually glossed over in most income-only studies, due to an implicit assumption that income is a valid *a priori* means of measuring inequality. Axioms postulated include: that income is the “most common and encompassing measure” for examining inequality (Green and Kesselman 2006, 4), that “income is the key contributor to the well-being of Canadian families” (Green, Ridell, and St-Hilaire 2016, 4), and that income is the “most used indicator” for examining economic inequality and segregation (Buitelaar, Wterings, and Ponds 2018, 33). However, such statements regarding income are equivocal justifications at best. Thus, *a priori* claims regarding income as the most common or best measure of inequality are debatable.

Critics of income-only inequality suggest that income does not capture the extent of opportunities that people face (Sen 1992, 28). The use of only income may be problematic because it neglects deprivations which relate to other socioeconomic indicators such as unemployment, health, education and social exclusion (Sen 1999, 108). These variables however, particularly health and social exclusion, can be hard to quantify and obtain in the Canadian context.

Income is a powerful variable because it is scalar, readily available, and amenable to statistical analyses. The basic categorization of needs for which income is a prerequisite is depicted in Maslow’s hierarchy (Maslow 1943), which compartmentalizes human needs into five categories. In the context of Western industrialized countries, particularly in urban areas, income is a prerequisite for the fulfilment of the most basic needs such as food and shelter. If an individual has no income, then they would fail at realizing any of the needs in Maslow’s

hierarchy. Furthermore, individuals with larger incomes have more freedom to choose how to utilize their income to best satisfy higher-level needs. Thus, while income does not capture other aspects of deprivation, generally speaking, income is a prerequisite for a number of other dimensions such as those noted by Sen (1992, 28), including health, education and the satisfaction of basic needs.

Consequently, several scholars have chosen to focus on income as a stand-alone measure of inequality. Family well-being implicitly implies financial security which is realized through income. This is not to say that income is the ideal variable in every case. Whereas Esteban and Ray (1994, 823) claim that income is a good proxy for socioeconomic differences or similarities, others such as Walks (2016, 759–61) prefer to focus on wealth because they feel that it better captures class difference. However, obtaining wealth data is more arduous than obtaining data on income, which is the simplest and most general indicator of welfare and inequality (Logan, Taylor-Gooby, and Reuter 1992, 130).

On a final note, in recent decades, the implementation of neoliberal policies by municipal governments, through the movement from managerial to entrepreneurial means of governance, has produced a scramble to recoup revenue via imposed or increased costs using pay-for-service charges (Lightbody and Walker 1987; Harvey 1989; Heuton and Girard 2010; Donald et al. 2014; Hamel and Autin 2017). In such context, income has become increasingly important in one's ability to maintain the same life choices and activities, be they education, access to recreational spaces, or the handling of added costs for utilities and infrastructure usage. The importance of income to deprivation in general is underlined by basic income projects that are increasingly discussed and attempted in affluent countries (Spermann 2017; The Economist

2018; Hamilton and Mulvale 2019; Magnani and Piccoli 2020). Such projects would in theory ensure a minimal income through which individuals can meet their basic needs.

### **Inequality vs polarization**

This paper makes use of two terms which are similar but slightly different: income inequality and income polarization. Income inequality refers to an unequal distribution of income (Dinca-Panaitescu and Walks 2015), and, as such, can be manifested in different ways (Sittthiyot and Holasut 2020). Income inequality can exist without income polarization. Conversely, income polarization is a term that is focused on the disappearing middle-income segment of society (Schettino and Khan 2020). Under income polarization, middle-income groups decrease while low- and high-income groups increase (Marcuse 1989, 699; Townsend 1993, 99; MacLachlan and Sawada 1997, 384).

Income polarization is a type of vertical inequality. Vertical inequalities are inequalities in terms of income between either households or individuals (Saiz and Donald 2017, 1031). Other inequalities exist too. For example, horizontal inequalities are between socially constructed groups, and as such are with respect with categories including but not limited to gender, ethnicity, sexuality, and so forth (Balakrishnan, Heintz, and Elson 2016, 32). Horizontal inequalities are also those that are spatial in nature. We are interested in exploring how vertical income inequality plays out spatially, for example, to what extent is the income group mosaic changing with changing income inequality in Canadian CMAs.

### **The Canadian context**

The concept of “polarization” was popularized by Harvard’s economist Lawrence Katz in the mid-2000s (Stiglitz 2012, 300). It is under that backdrop, and subsequent interest in inequality after the economic crisis of 2007–08, that recent studies of income polarization in the

Canadian urban context commenced (Hulchanski 2007; 2010). Income inequality in Canada is quintessentially an urban phenomenon (Fong 2017, 2). Considering that Canada is one of the most urbanized countries of the world (with about four fifths of its population residing in urban areas (Statistics Canada 2018; United Nations 2018)), urban issues are national issues. Research since the 1990s shows that income inequality has increased in Canada (on national, provincial and urban scales), to the extent that the country experienced the highest rise in inequality amongst OECD countries since the mid-1990s (Banting and Myles 2013; Beach 2016; Breau 2015; Fortin et al. 2012; Osberg 2008; John Myles, Picot, and Pyper 2000; Fong 2017). Such trajectories in income composition and distribution are of concern, because governments will at some point have to seriously address this development in order to evade adverse ramifications such as economic and social instability (Osberg 2016, 333).

### **Geographic approaches to measuring income polarization at the census tract level in the Canadian urban context**

In the early 1990s, urban geography research examining divisions within cities flourished (Davis 1990; Mollenkopf and Castells 1991; Sassen 1991; Fainstein, Gordon, and Harloe 1992; Bird et al. 1993; Watson and Gibson 1995). Concomitantly, studies on income inequality and polarization in the Canadian context also emerged (Bourne 1993; MacLachlan and Sawada 1997; John Myles, Picot, and Pyper 2000). MacLachlan and Sawada (1997) detected income polarization by examining how the middle-income group changed in twenty-two Canadian CMAs. Their key contribution was the delineation of income ranges (e.g., low, middle, high) for examining income polarization. This is important, because a city's income distribution may both polarize and depolarize depending on how the middle is defined, and that therefore it is necessary to examine different ranges in order to detect if polarization is actually occurring (MacLachlan and Sawada 1997, 385, 392).

About a decade later, research on income polarization was picked up by the Three City Project (TCP). The TCP originated as a report that compared trends in income change of census tracts (CTs) in Toronto solely between two endpoints: the 1971 and 2001 censuses (Hulchanski 2007). When the second report (Hulchanski 2010) was published, the production value increased in terms of improved visuals. The second report resulted in TCP research gaining more traction and notoriety (S1 in Appendix II). In this updated review, maps for additional census years were produced which demonstrated a more detailed evolving income landscape in Toronto between 1971 and 2006. Thus, emerging concentrations of high-income and low-income areas were illustrated.

After the updated report, numerous studies ensued which examined income polarization in other Canadian urban areas (Ley and Lynch 2012; Rose and Twigge-Molecey 2013; Prouse et al. 2014; Harris, Dunn, and Wakefield 2015; Townshend, Miller, and Evans 2018). Unfortunately, the TCP research across select urban areas did not use the same analytical parameters. These differences limit ability to make broad generalizations between cities.

In effect, the TCP produced two products. The first is the “Three City Model”, which the researchers present as an examination of income polarization. However, that model actually examines income trajectories (Distasio and Kaufman 2015, 11).

The second TCP product examined income-polarization through “time-slices”. These time-slices cartographically convey, in a cartographic manner, what previous studies (MacLachlan and Sawada 1997) discussed. However, the TCP researchers did not make maps for every census year, nor did they examine the spatial patterns of income inequality across urban areas. Spatial information can support more robust conclusions regarding potential knock-on effects of spatial inequalities to, for example, urban form and function.

## **Methodology**

This research examines the temporal evolution of income polarization. We begin by investigating how various income groups have changed over time in each Census Metropolitan Area (CMA). Analyzing the trend of each income group over time allows for direct comparisons of the slopes of the trend lines that model the rates of income decline or rise among the CMAs in our study. Subsequently, an examination of the spatial distribution of income groups within each CMA is achieved using time-slice maps. Finally, time-slice maps are subjected to spatial analysis to measure fragmentation and spatial autocorrelation within CMAs.

## **Study areas**

This study examines the eight largest Census Metropolitan Areas (CMAs) in Canada: Toronto, Montreal, Vancouver, Calgary, Ottawa-Gatineau, Edmonton, Quebec City and Winnipeg (Figure 2.1).

In this study, the spatial boundaries used are the CMAs of each census year. The CMA is a popular geographic level of analysis and can be rationalized by the notion that the CMA reflects the labor market (Hulchanski 2010, 25).

For this study, within each CMA, analysis is at the census tract (CT) scale, which is the most common spatial unit at which cities and urban processes are examined (S3 in Appendix II). Data at this level of geography is available throughout all years of study.



**Figure 2.1. Locations of Canada’s eight largest CMAAs.**

### **Data**

The data for this study is from Statistics Canada and covers eight census years between 1971 and 2016. The 1976 census is considered to be a mini census, and it was only starting in 1986 that mid-decade censuses in Canada collect the same items as the start of decade censuses (Statistics Canada 2015). We do not use 2011 census data because politically motivated issues led to a non-random population sample on much data. The federal regime at the time made the long form census voluntary, resulting in a disconcerting situation whereby data scientists were effectively “muzzled” by not having access to quality data (Vinodrai and Moos 2015).

The *focal variable* for this research is average before tax household income at the CT level. In the present paper we usually refer to this census variable as simply income. The full census variable names used are provided in the supplemental material (S4 in Appendix II). Previous research used income at the household level (MacLachlan and Sawada 1997, 389), but without presenting a very strong rationale for the choice. Households may represent atomic units of the urban landscape more so than individuals, as individuals pool their assets towards major expenditures. The largest such cost remains the same throughout decades: housing / shelter costs (MacLachlan and Sawada 1997, 389; Hulchanski 2010, 21). Using average household income in lieu of average individual income should yield similar results (Hulchanski 2010, 24). Finally, out of convenience it is easier to use average household income because the census provides the former in all census years that we examine. However, to assess the robustness of our results under different assumptions, we repeated our income trend analysis using the aforementioned methodology on individual-based income.

The missing data for 1976 and 2011 was imputed using linear interpolation. Moritz et al (2015) tested the accuracy of numerous imputation techniques on time-series which had various levels of periodicity, trend, random and non-random missing data. They found that linear interpolation was the most accurate means of imputation for non-periodic datasets with a trend and missing data that is assumed random in nature. While our time-series of high-, middle- and low-income is quite short (ten census years), it exhibits no obvious seasonality but does show trends. As such we choose to use linear interpolation to impute the values for 1976 and 2011 prior to statistical analyses.

We use income figures before taxation, because after-tax income data was not available until the 2006 census (Statistics Canada 2017b), and median income was not collected every

year. Hence, to maintain consistency for all years, we opt for the common pre-tax average for all analyses. In the literature there is no consensus as to the ideal measure of income that should be used when measuring income polarization (MacLachlan and Sawada 1997, 389; Hulchanski 2007, 6). Thus, we feel confident in using before-tax income through time to assess the temporal evolution of income polarization.

The total number of households was not available in the 1971 census. However, numbers of households within income ranges were provided. Hence, a sum of the households in all the ranges gives a number sufficiently close to the total number of households, varying only due to rounding issues (because Statistics Canada often rounds figures to the nearest number ending in 5 or 0).

The average income of households in the CMA was not available for the 1971 census. This is the only year in our study which has such an anomaly and hence the average income figure had to be derived using the total number of households with income, and the average household income in each CT. The formula for the CMA's average household income is:

$$AI_c = \sum_{k=0}^n \left( \frac{HH \times AI_{hht}}{T_{HH}} \right)_k$$

Where  $AI_c$  = Average Income of CMA, where  $n$  = total number of CTs,  $AI_{hht}$  = Average Household Income of CT,  $HH$  = Households with income in CT,  $T_{HH}$  = Total households in study area.

This method was validated on data sets from years where average household income was provided for the entire CMA. For census years in which Statistics Canada provides average income figures, our calculated averages were within a fraction of a percent of the provided figures.

All data that was used in this study has been made available on the following GitHub page: [https://github.com/lazarification/IncomePolarization\\_CA\\_CMAs](https://github.com/lazarification/IncomePolarization_CA_CMAs).

### **Population based weights**

Previous research has used population weighted income for census units when examining inequality in North American cities (Jargowsky 1997, 148; John Myles, Picot, and Pyper 2000, 3; Chen, Myles, and Picot 2012, 882). Such methodological decision is in lieu of treating all CTs in an area as having equal weights of one.

Assigning household weights is rationalized through the notion that CTs with higher population (i.e., households) have more impact on their corresponding income group (low, medium and high) than CTs with lower population. Such an approach is similar to the work of MacLachlan and Sawada (1997, 391), who calculated the sizes of middle-income groups, in various CMAs at the CT level, based on the total number of households in CTs whose average household incomes corresponded to such income group.

In our analysis we employ household weights when assessing trends of low-, middle- and high-income groups, as well as in the production of cartograms. We do not employ these weights to produce conventional maps, perform fragmentation indices, or to conduct spatial autocorrelation since these latter analyses do not consider a continuous variable.

### **Defining and measuring the middle**

The middle-income group is most often defined by using an income range, and these *vary across studies*. For example, in the Three City Project (Hulchanski 2007; 2010; Ley and Lynch 2012; Rose and Twigge-Molecey 2013; Prouse et al. 2014; Harris, Dunn, and Wakefield 2015; Townshend, Miller, and Evans 2018), the range used was  $\pm 20$  percent around the average income. This 40-percentage point range, hereafter referred to as the bandwidth, is between 80

and 120 percent of the average income. However, numerous other definitions of the middle exist. Incomes which range from -33 percent and +100 percent of the median income are considered middle-income by Kochhar et al (2015, 6). Piketty (2014, 251) defined the middle as the 40 percent of the population which is above the median and below the elite (who he classified as top 10 percent).

Interpretations are less robust in the absence of any means to invalidate observed trends that emerge from a single arbitrarily middle-income bandwidth definition. Previous research has shown that, depending on the chosen bandwidth, results could be engineered to exhibit either polarization and depolarization (MacLachlan and Sawada 1997, 392). It is, therefore, important to assess the sensitivity of results, to broadly different bandwidth definitions.

Sensitivity analysis examines multiple bandwidth definitions for the middle-income group to determine the consistency of income trends in each bandwidth between 1971 and 2016. TCP researchers did not examine multiple definitions of middle-income ranges when they analyzed time-slices in Canadian cities. Without a common definition of ‘middle’, the comparison of trends between different CMAs is fraught with potentially spurious inferences.

This study uses  $\pm 15$ ,  $\pm 20$  and  $\pm 25$  percent bandwidths to assess the sensitivity of the low-, middle- and high- income trends and also presents the average of the three as a robust trend. The mean of the three bandwidths for each time-slice provides results that are less sensitive to any single arbitrary change in magnitude induced by an income group’s bandwidth for a particular year. The average ensures that any outliers, in the form of a high-frequency variation within one of our three bandwidths is reduced in each year. We use the non-parametric mean for each time-slice and recognize that the assumptions of normality for a given year are not valid and so use Monte-Carlo methods to assess the significance of trends in our data.

Once CTs are assigned a corresponding income range, it is possible to sum their household counts to produce a total household weight in the form of a percentage of the total CMA households. This is also done for the low- and high-income categories, because one should identify if these other two groups both increase in order to be able to affirmatively deduce that income polarization is occurring (Hamnett 2003, 70).

To determine how low-, middle- and high-income groups have changed over time, graphs are produced which plot the percentage of each income group over the temporal duration of the censuses examined. From each graph, we modeled the linear trend of the mean of the three bandwidths and derived a slope value to assess if the particular CMA exhibited rise or decline in each income group over time.

To assess the statistical significance of the linear trend lines from 1971–2016, we obtained confidence intervals for the mean slopes for each income group for each CMA using nonparametric bootstrapping, replicating 10,000 bootstrapped slopes for each income group's slope by randomly sampling, with replacement.

### **Mapping**

We utilized a  $\pm 20$  percent bandwidth in creating our maps. We map every census year in which data is available and produce cartograms, weighted by the number of households in order to provide a more balanced representation of the reality of income polarization in urban areas.

Examining the spatial distribution of income polarization within the boundary of CMAs, at the CT level, provides excessive visual weight for rural CTs that are large in area but low in number of households. Large-area CTs obscure the significance of small-area CTs within the more populated city centers, hence the use of cartograms. Cartograms weight the CTs by the number of households, thereby decreasing the size of CTs with smaller numbers of households

and increasing the size of CTs with larger numbers of households that fall into a specific income category. As such, small-area CTs with high numbers of households are visually expanded in area on the map. The goal of the cartogram is to produce a topologically enforced density equalized map of a CMA. Density is equalized by changing the size of census tracts.

Cartogram production used the Gastner-Newman algorithm (Gastner and Newman 2004). Cartograms by definition yield twisted and distorted images (Dent, Torguson, and Hodler 2009, 168, 182). Contiguous cartograms require multiple iterations in order for the distorted CT shapes to become progressively closer to matching the chosen weight (Dent, Torguson, and Hodler 2009, 183), which in our case is the total number of households per CT. Therefore, we applied four or five iterations of the Gastner-Newman technique for producing each cartogram. This amount of iterations was chosen because the application of additional iterations produced images which appeared almost identical, meaning that there was a convergence of the Gastner-Newman algorithm and only minor changes in density, whereby the size of modified CTs corresponds approximately to the total household weights. Certain CT boundaries had to be corrected in certain CMAs for select years due to mistakes in source digitization (S5 in Appendix II).

### **Fragmentation analysis**

As the inequality of an urban area changes, we expect concomitant changes in the spatial distribution of the three income groups as a consequent of numerous urban processes related to income polarization. For example, gentrification, among others, leads to changes in the income status of adjacent census tracts, due to, for example, spillover effects (Davidson and Lees 2005, 1169). As such, fragmentation does not show income inequality itself, rather it reveals the spatial structure that can stem from income inequality. Hence, while fragmentation is not per se income-polarization, we are using it as an additional tool to examine the consequences of income-

polarization spatially which can provide insight on whether the income trends lead to a more or less divided urban landscape.

From a municipal perspective and under the purview of social justice, understanding socioeconomic fragmentation of income groups within CMA boundaries can aid in developing urban strategies aimed to decrease spatial inequalities in access to services, health and so forth. Additionally, the knowledge of the spatial configuration of vulnerable populations can aid in developing strategies to mitigate undesirable socioeconomic conditions by better targeting places and populations in need.

Patch fragmentation studies usually involved examination of landscapes in an ecological context (Simmons 1964; Januszewski 1968; Igbozurike 1974; Barrington and Ilbery 1987; Johnsson 1995; Demetriou, Stillwell, and See 2013; Wang, Blanchet, and Koper 2014). Likewise, the urban spatial structure is characterized as fragmented (Angel, Parent, and Civco 2012, 249). Consequently, fragmentation studies have been applied to demographic data (Crews and Peralvo 2008), and more recently in urban contexts (Reis, Silva, and Pinho 2016; Bagheri et al. 2019; Delmelle 2019).

Herein, the socio-economic fragmentation of each time-slice map is calculated. A fragment or patch is defined as any contiguous set of CTs that belong to the same income category (low, middle or high). These groups of CTs create a mosaiced landscape composed of high-, middle- and low-income patches. CTs with data suppression are not considered in the analysis. Water bodies are present in many CMAs, and in effect form lacunas of empty space that divide CTs. Major water bodies (large rivers, bay inlets, oceans, etc) were present in some CMAs and were treated as dividing lines that separate fragments/patches. Thus, groups of

contiguous CTs were split. However, we did not split individual CTs which were multipart polygons.

As water bodies are not present in all CMAs, it was necessary to remove them from certain digital boundary files.

Two patch fragmentation indexes were used: the Johnsson Fragmentation Index and Edge Density. The lower limits of both of these indexes approach zero in situations when the total area examined is large and there are either few fragments or when there is relatively small fragment edge length. Conversely, the upper limit is boundless.

Johnsson's Fragmentation Index is given as (Johnsson 1995, 212):

$$Fragmentation\ Index = \frac{(M - 1)}{(N - 1)}$$

where M is the total number of fragments (or map regions), and N is the total number of areal units (or pixels in the raster). For this index it is therefore necessary to obtain the total number of fragments and to convert the CTs to raster format in order to obtain the total number of pixels. The cell size for conversion was set to 50 m<sup>2</sup> (only the absolute value of the index is sensitive to differing cell sizes of 10 m<sup>2</sup>, 100 m<sup>2</sup> and 500 m<sup>2</sup>, but the trend invariant).

Edge Density is a more common fragmentation index and can be expressed as (Crews and Peralvo 2008, 73):

$$Edge\ Density = \frac{\sum e}{a}$$

where e is the perimeter of each class of polygons, e.g., middle income, and a is the total area in question.

Both fragmentation methods contain area related terms in their denominators. The methods are effective when comparing similar spatial extents, but their results are respectively

incomparable when the study area's spatial extent changes drastically even though both are normalized by area. An example is provided in the supplemental material (S6 in Appendix II), demonstrating the necessity of calculating fragmentation in this research using a constant sized study area for each CMA.

For each city, in order to construct a constant study area for the fragmentation indices, boundary files for each CMA for all census years were spatially intersected. The resultant common area was used for fragmentation analysis. The areas outside this common boundary mainly were comprised of water bodies and sparsely populated rural areas in most cases.

Constructing a constant spatial boundary through time using geometric intersection resulted in numerous slivers or spurious polygons. Such polygons result from discrepancies in the digitization of CT boundaries in different census years—often the boundaries for the same CT do not perfectly overlap from one year to the next. These spurious polygons are a larger concern for the Johnsson Index, and hence were not used towards the total fragment count.

Using a constant study area essentially results in the Johnsson Index becoming effectively a count of Fragments divided by a constant area, while the Edge Density measure amounts to a sum of edge lengths divided by a constant area.

### **Spatial autocorrelation**

The fragmentation indices analyze the high-, middle- and low-income mapped CMA mosaic at the patch level for each year. A patch is composed of one or more CTs with the same income status. As such, the patch-level pattern is an emergent pattern resulting from the non-spatial process of income polarization as well as spatial organizational processes operating at the level of the CTs. As such, to complement the patch fragmentation indices, we assessed the degree of spatial autocorrelation within each of the income groups for each CMA and each time-

slice. The examination of changes in the spatial autocorrelation of each income group can help us understand what spatial processes (clustering/dispersion) are giving rise to the patches and whether or not patches tend to grow or shrink over time with changes in the income trends.

A nominal level measure of spatial autocorrelation (SA) called the Joins-Count (JC) is used herein to compute spatial autocorrelation for each time-slice map (Cliff and Ord 1981, 63–65). Calculation of the joins-count measure utilized the ‘joinscount.mc’ function in the package *spdep* (Bivand and Wong 2018) in R 4.0 (R Core Team 2020).

The JC enumerates the number of ‘joins’ between polygons with categorical labels. A join occurs whenever two CTs share a boundary. On a time-slice map, each CT is a polygon that is contiguous with its neighbouring polygons. For example, if two CTs are contiguous via a shared border then that border represents a single ‘join’ between the two CTs. The JC measure counts how many times a join is found between polygons that belong to the same nominal category (e.g., between two middle-income CTs) or between different nominal categories (e.g., middle and high or middle and low or high and low).

Joins counts are performed using a binary spatial weights matrix. The analyses for each time-slice was completed using a Queen’s case neighbourhood definition, wherein a given CT included all surrounding CTs as neighbours if they share a boundary or point along their boundary. Because spatial autocorrelation results can be sensitive to the definition of neighbourhoods used, we tested the sensitivity of the results by repeating the analysis using both a Rooks (sharing only a linear boundary) and K-nearest-neighbour (with  $k = 5$  so that each CT has 5 neighbours) neighbourhood definition.

In each map, there are three income categories (low, middle and high), and the JC was used to determine the number of low-low, middle-middle and high-high joins. If, for example,

CTs for the middle-income group tend to be joined more often to other middle-income group CTs, then that income group on a time-slice map is an example significant positive spatial autocorrelation of middle-income CTs. Positive SA indicates that the income category tends to be spatially clustered rather than dispersed within the CMA.

For a given time-slice map, to estimate the expected value and the variance for each income category, 999 simulated joins-count measures were produced by shuffling all three income group labels on the map and each time counting the number of joins for each income group and adding the observed number of joins as one possible outcome. Next, the mean and variance for that simulation was calculated and stored. Then, for each map that same procedure was repeated 1000 times to create two normally distributed distributions, one containing the means and the other the variances. The average of the means and variances of the reference distribution then approximates the true expected value and variance for an infinite number of joins-counts measures for the given time-slice map. Finally, for a given map, the observed number of joins for each income category was subtracted from the average of the reference distribution and divided by the standard deviation to arrive at a standardized z-value for each income group in each time-slice in each city. Higher z-values indicate stronger positive spatial autocorrelation for a given income group whereas lower z-values near zero indicate weaker positive spatial autocorrelation and negative z-values indicate significant negative spatial autocorrelation. When assessing a given plot for significance, we only interpret the trendline if at least one of the time slices exhibits a z-value that corresponds to the Bonferroni corrected rate of  $\alpha = 0.005$  which corresponds to  $z = 2.576$ . The reasoning is such: significant spatial autocorrelation values across the time slices are no different from a random shuffling of the three groups among the CTs and thus a trend would be spurious.

For a given time-slice and income group, a joins-count measure that increases over time suggests a clustering process, wherein similar income groups are more often found together (by inference of sharing numerous joins), perhaps because of spatial segregation of income groups for example. Conversely, if the joins-count measure is decreasing over time then this suggests that there is a tendency for CTs with different income group memberships to be adjacent. Under a decreasing joins-count scenario, an income group would be more likely to find a different income group adjacent to itself and thus the group in question would be more dispersed than chance would predict. Such a scenario can be explained by increased fragmentation of CTs of that income group within urban landscape mosaic.

The joins-count z-values were plotted over time and the linear trend was modelled and bootstrapped to obtain confidence intervals for the 24 panels (one per three income groups in every CMA) using the same procedure as was used for the average of the three bands for income groups detailed above.

## **Results**

The statistically significant negative slopes for the middle-income group trend line in all cities (Figure 2.2) based on household (Table 2.1) or individual (Table 2.2) income data implies a steady erosion of the middle-income group since 1971 for all CMAs except for Quebec City and Vancouver. Additionally, all CMAs have significant positive trends for the low-income group under either household income (Table 2.1) or individual income (Table 2.2), with the exception of Quebec City in the latter. Given that these are proportions weighted by the number of households, the low-income group is expanding within most CMAs at the cost of the middle-income group. For household income, high-income groups exhibit increasing trends but only half are statistically significant (Table 2.1). Three-quarters of the high-income group trends are

significant if we consider individual income (Table 2.2). In general, middle and low-income groups show similar trends for both types of income, household or individual. The greatest variability is around the low and middle-income groups. For all subsequent analysis and mapping we utilized the average household income data only.

Figure 2.2 illustrates that, for a given bandwidth definition (e.g.,  $\pm 15\%$ ,  $\pm 20\%$ ,  $\pm 25\%$ ), there is variability in the percentage of households over time within any of the three income-groups. In many cities, the census to census variability between the different definitions of the middle differ only in magnitude between the three bandwidths. For example, in several cities, one could choose any bandwidth, say  $\pm 15\%$ , and the others ( $\pm 20\%$  or  $\pm 25\%$ ) would be a simple translation of the curve along the percentage axis. However, in other cases, the variability for a given bandwidth is less predictable for any randomly chosen census year.

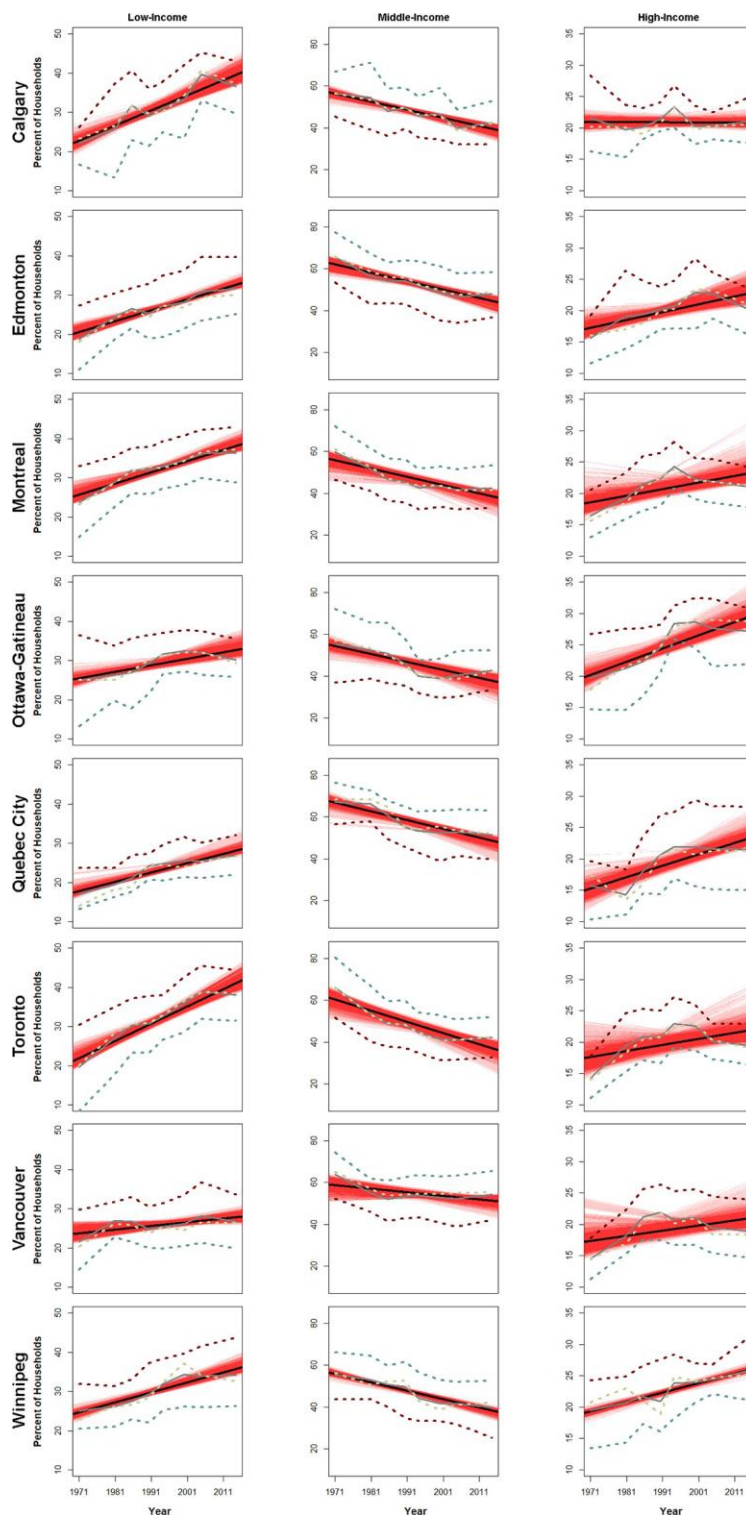
In many of the middle-income graphs (Figure 2.2), from 2006 onwards, the bandwidth curves stop declining and at the same time many low-income group trends plateau. There are fewer significant positive trends in the high-income group across the CMAs, thus the middle-income group has not shifted by equal amounts to the low- and high-income groups. Growth of the lower-income groups over the higher group has been previously observed by (Pahl 1988, 260).

Maps (Figure 2.3) provide insight into how the income groups are spatially structured through time. For considerations of space in the main body of this paper, maps are provided for only the time-period endpoints of 1971 and 2016 for each CMA (Figure 2.3). However, sets of maps are provided for each CMA for every year in the supplementary material (S7 in Appendix II).

Conventional maps show how the CMAs have expanded (and in some cases temporarily contracted) in size over time (S6 in Appendix II). Between 1971 and 2016, both high- and low-income groups have expanded from the center towards outer regions of many CMAs.

CMAs include suburban and rural census tracts (CTs) which have large areas. The density equalization effect of the cartograms is effective in reducing the strong visual weight given to rural CTs with smaller numbers of households. Likewise, cartograms increase the size of CTs that are closer to the central areas of CMAs, which are territorially small but contain equal or greater numbers of households.

In every CMA, there has been a significant decline in the middle-income trend since 1971. Spatially, the gap left by the middle-income group decline is illustrated by cartograms that show expansions of low- and high-income areas (Figure 2.3 and S6 in Appendix II). Concurrently, the low-income areas (in red) have expanded considerably.



**Figure 2.2. The percentage of households in low, middle and high-income classes, in Canada's largest eight CMAs.** Each column has a different y axis, but each panel within the same column has the same axis range. Different definitions of the middle are shown (blue:  $\pm 25$  percent of average, yellow:  $\pm 20$  percent of average, dark red:  $\pm 15$  percent of average), with grey being the composite average. Confidence intervals of the linear trend of the average of the three bandwidths are shown in red. A bootstrapped confidence interval that does not cross zero indicates a statistically significant result.

**Table 2.1. Slopes of the three income groups (low, middle & high) and their associated 95% confidence intervals (CI) obtained via non-parametric bootstrapping of household income data.**

<b>CMA</b>	<b>Low (95% CI)</b>	<b>CI<sub>mid</sub></b>	<b>CI<sub>high</sub></b>
Calgary	0.37 [0.27, 0.48] *	-0.37 [-0.47, -0.27] *	-0.00 [-0.03, 0.04]
Edmonton	0.27 [0.21, 0.32] *	-0.39 [-0.52, -0.26] *	0.12 [0.04, 0.21] *
Montreal	0.28 [0.19, 0.36] *	-0.38 [-0.57, -0.17] *	0.10 [-0.02, 0.21]
Ottawa-Gatineau	0.16 [0.06, 0.25] *	-0.37 [-0.55, -0.17] *	0.20 [0.12, 0.29] *
Quebec City	0.23 [0.16, 0.29] *	-0.41 [-0.54, -0.25] *	0.18 [0.10, 0.27] *
Toronto	0.42 [0.31, 0.52] *	-0.52 [-0.77, -0.29] *	0.09 [-0.03, 0.26]
Vancouver	0.09 [0.02, 0.16] *	-0.17 [-0.35, 0.01]	0.08 [-0.05, 0.21]
Winnipeg	0.25 [0.20, 0.31] *	-0.39 [-0.46, -0.34] *	0.15 [0.14, 0.16] *

A slope is significant statistically when its confidence interval does not include zero and these are indicated by an \* .

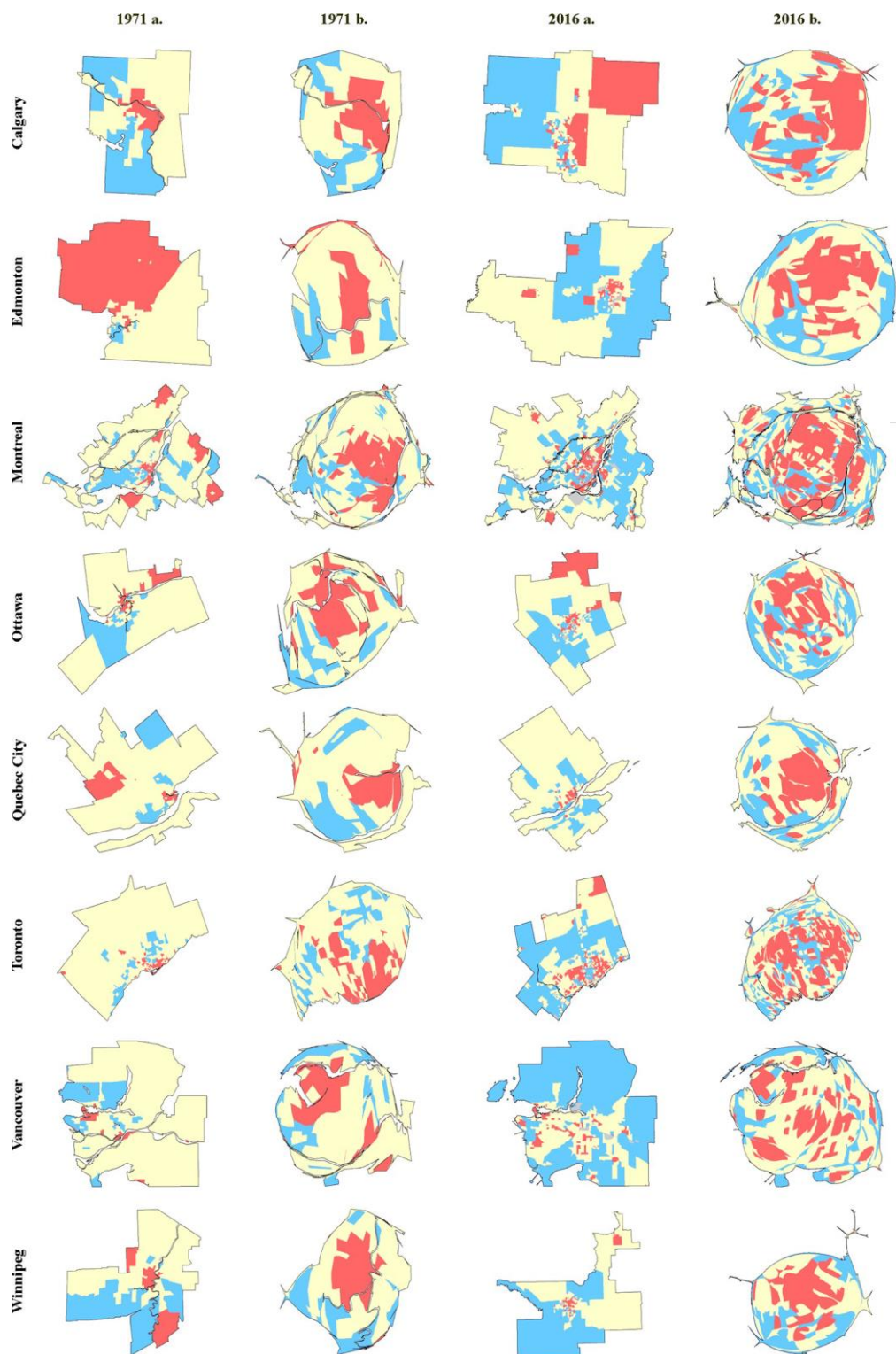
**Table 2.2. Slopes of the three income groups (low, middle & high) and their associated 95% confidence intervals (CI) obtained via non-parametric bootstrapping of individual income data.**

<b>CMA</b>	<b>Low (95% CI)</b>	<b>CI<sub>mid</sub></b>	<b>CI<sub>high</sub></b>
Calgary	0.57 [0.45, 0.74] *	-0.65 [-0.86, -0.52] *	0.09 [0.05, 0.13] *
Edmonton	0.35 [0.27, 0.43] *	-0.48 [-0.60, -0.32] *	0.13 [0.05, 0.18] *
Montreal	0.14 [0.10, 0.18] *	-0.20 [-0.26, -0.14] *	0.06 [0.04, 0.09] *
Ottawa-Gatineau	0.08 [0.05, 0.15] *	-0.12 [-0.23, -0.04] *	0.05 [-0.02, 0.12]
Quebec City	0.02 [-0.02, 0.06]	-0.05 [-0.14, 0.02]	0.03 [-0.01, 0.8]
Toronto	0.53 [0.41, 0.66] *	-0.62 [-0.76, -0.48] *	0.09 [0.07, 0.13] *
Vancouver	0.40 [0.35, 0.48] *	-0.51 [-0.61, -0.43] *	0.11 [0.08, 0.16] *
Winnipeg	0.19 [0.05, 0.29] *	-0.32 [-0.50, -0.09] *	0.13 [0.05, 0.24] *

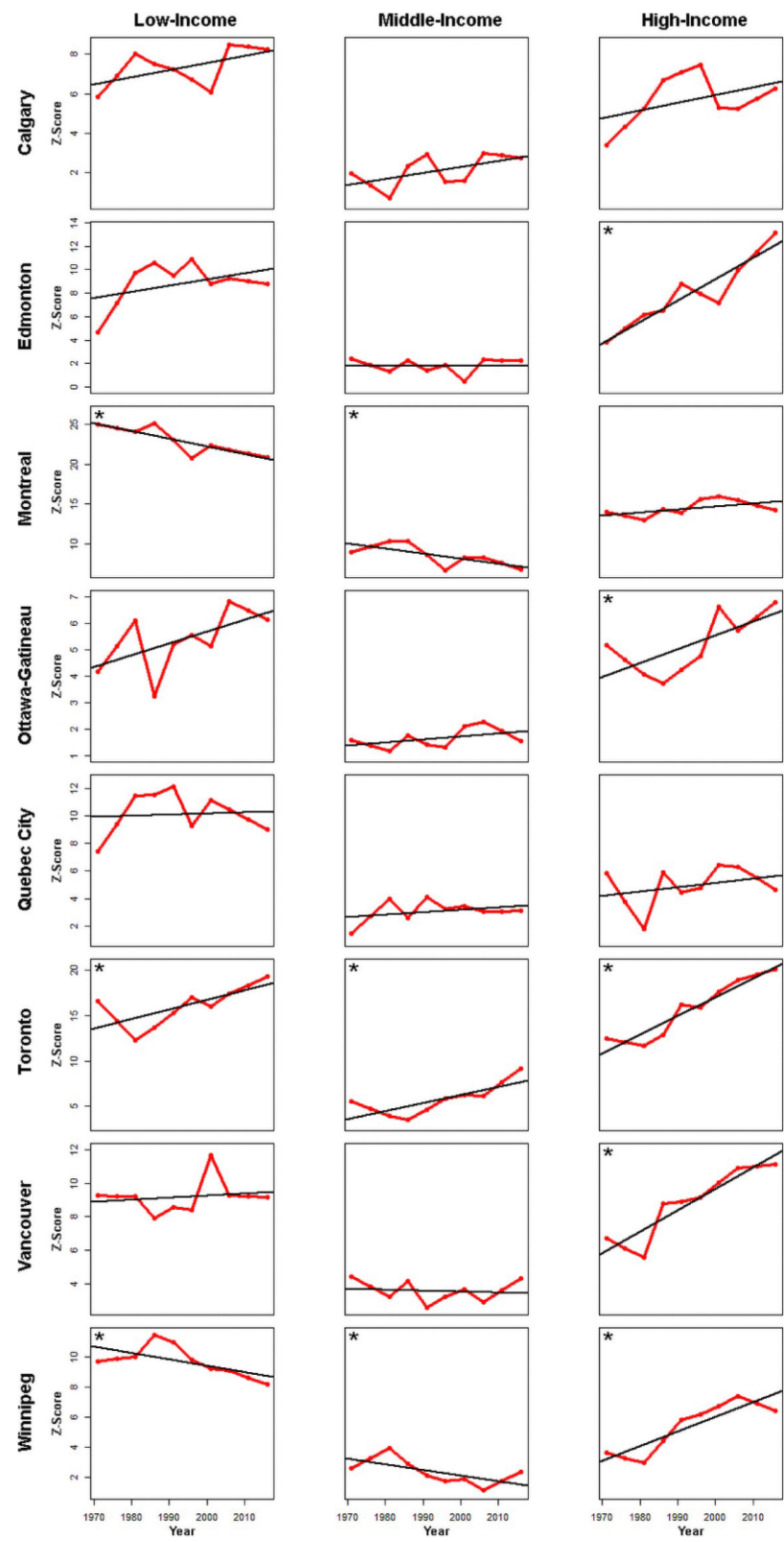
A slope is significant statistically when its confidence interval does not include zero and these are indicated by an \* .

Visually, all CMAs appear more fragmented over time (Figure 2.3 and S6 in Appendix II). Both conventional maps and cartograms support this observation, as over the decades an increasing number of separate fragments have emerged in every CMA. This is perhaps most visible in Vancouver's CMA, where one can observe a mosaic of many fragments forming rather than large clusters of CTs.

There has been an almost continual increase in fragmentation between income groups in every CMA. Some regions, such as Ottawa-Gatineau exhibit continual increases in fragmentation. Others have moments of stability before fragmentation resumes on an upward trend. Examination of the indices calculated on the constant area, Montreal, Toronto and Calgary show the steepest increases, while Winnipeg, Quebec City and Ottawa-Gatineau exhibit moderate increases and, finally, Edmonton and Vancouver show the smallest increases in fragmentation since 1971. The indices calculated on the full complement of CTs in each year tend to show very low increases in fragmentation and some decreases (Ottawa-Gatineau and Winnipeg).



**Figure 2.3. Conventional maps (a) and Cartograms (b) of the temporal evolution of the income structure of Canada's eight largest CMAAs.** Low-income CTs (less than 80 percent of average) are in red, middle-income CTs ( $\pm 20$  percent of the average) are in yellow, high-income CTs (above 120 percent of the average) are in blue, and other (no data, or no households) CTs are in grey. Additional maps for each census year for all eight CMAAs are provided in the Appendix (S7 in Appendix II).



**Figure 2.4. Results of spatial autocorrelation / join count analysis.** Confidence intervals from bootstrapping are available in the supplementary material (S9 in Appendix II). Significant trends at an  $\alpha = 0.05$  are shown by an \* in the upper left corner of a panel.

Montreal and Winnipeg exhibit statistically significant declining trends of spatial autocorrelation over the period of study for the middle-income group and, as such, middle-income CTs tend to be adjacent less often over time (Figure 2.4). In Toronto, the low-income trend is significant and positive, increasing spatial autocorrelation over time. Thus, low-income Toronto CTs tend to be found next to each other more often than chance would predict. That is also true for the medium and high income CTs in Toronto. Conversely, in Montreal and Winnipeg, the negative trend in low-income spatial autocorrelation means that low-income CTs have successively had fewer low-income neighbours over time. The trends in the remaining CMAs are not significant but tend to indicate a positive slope and increasing spatial autocorrelation over time. Except for Calgary, Montreal and Quebec City, there are significant positive trends of spatial autocorrelation for the high-income groups. Neither Ottawa nor Edmonton exhibited any significant spatial autocorrelation in the middle-income class over time. Confidence intervals for trend line slopes of each CMA's Join-Count data are provided in the supplemental material (S9 in Appendix II). The results of Join-Count analyses using the other two neighbourhood configurations yields similar results (S10 and S11 in Appendix II).

## **Discussion**

Prior to the 2010s, there were few studies mapping income polarization in Canada. This research examined the trends in low-, middle- and high-income groups in Canada's largest CMAs. First and foremost, we would like to note that each CMA has its own unique spatial and temporal specificities, whereby differences can be manifested in varying patterns of spatial segregation: unevenness, clustering, exposure, etc (Massey and Denton 1988; Townshend and Walker 2002).

We presented income trends using both the household (Table 2.1) and individual (Table 2.2) level incomes. Some researchers prefer to use data on the level of individuals rather than households because they feel that using household data introduces complexities involving changing household size over time (Bourne and Hulchanski 2020, 26). Household size has decreased over time (Townshend and Walker 2020), so it is possible that due to the rise in single individual households there might be a decrease in such incomes. Equally possible is that household incomes might increase in some areas in situations where multiple low-income individuals reside together (Bourne and Hulchanski 2020, 26). However, there are all sorts of reasons for decreasing household size, including the fact that households have fewer children now compared to the past. It is therefore important to not generalize suppositions across all time and space.

In a few chapters of the recent Three City Model (TCM) publication (Grant, Walks, and Ramos 2020), it is noted that, per their methodology, Gini coefficients are higher when examining household-income than when examining individual-income. However, earlier research underlines the shortfalls of examining Gini coefficients when examining income inequality and polarization (MacLachlan and Sawada 1997). To avoid overgeneralizing and speculating with the aforementioned structural arguments, we presented trends of low-, middle- and high-income groups in Figure 2.2 and Tables 2.1 and 2.2.

The results between Tables 2.1 and 2.2 are in some ways similar. There are statistically significant decreasing trends in the middle-income groups for all but two CMAs (Tables 2.1 and 2.2; Figure 2.2). Almost every CMAs' income structure is consistent with various conceptualizations of polarization, whose defining feature is the disappearance of the middle whereby the two poles (low and high) increase (Pahl 1988; Marcuse 1989).

The results for household and individual income data are somewhat similar. Income based data is sometimes more extreme, which suggests that the household-based income data is in some cases a more conservative means of assessing income polarization. Results for Vancouver and Quebec City tend to differ from the other CMAs. For example, no income-group trends are significant in Quebec City for individual-based income data. The converse is the situation with Vancouver, where middle and high-income trends are not significant when household-based income data is examined. As such, there are some differences induced by different income measures.

The low-income groups for all CMAs, except for Quebec City, exhibited significant increasing trends. Of the pairs of datasets where both types of data were statistically significant, in CMAs other than Montreal, Ottawa-Gatineau, and Winnipeg, the individual-based income data showed larger increases than the household-based income data.

Middle-income group trends were decreasing in every CMA, though Quebec City's results for individual-based income and Vancouver's results for household-based income were not statistically significant. Once again, apart from Montreal, Ottawa-Gatineau, and Winnipeg individual based data gave higher extremes than household-based data when comparing how the middle-income groups decreased.

High-income groups give somewhat different results when comparing household- and individual-based income approaches. The household income results which show only half the CMA's having statistically significant increasing slopes, whereas the individual income approach shows all CMAs except for Ottawa-Gatineau and Quebec City having statistically significant increasing slopes.

The middle-income group has not been equally divided between the high- and low-income groups and there is a clear asymmetry. With few exceptions, the low-income groups have increased significantly more than high-income groups. Such phenomenon is distinctly different from classic polarization, which assumes an equal division of the middle to the high and low ends, and our observations could be termed “lumpenization” (opposite of embourgeoisement).

Historically, it is far easier for one to move from middle-income status downwards, in part because the upper-class is exclusionary. Aside from the trend lines, within the bandwidth curves themselves (Figure 2.2), the general plateauing of the low-income groups from 2006 onwards and slight increases in the middle-income groups at the same time, may be an indication of stabilization. Patterns of slow or no erosion of the middle-class would be expected in the 1970s, and to an extent the 1980s, as these were a period in which Keynesian economic policies fostered a strong welfare state (Banting 1997; Myles 1998; Brenner and Theodore 2002, 350). However, our results are different, in that they show a steady erosion of the middle-income group during this period. This observation could be due to natural short-term variability of the available census data.

With a shorter record, our trends and inferences would have been different. Our time series was constrained by data availability from 1971 onwards and this is a relatively short period. As such, we are unable to assess any periodicity that might be present in the bandwidth curves themselves and make the assumption that the variation from census to census is random around the trends we produced. As with any time series, and in particular short ones, taking any random interval or start/end point will lead to trends that are in opposition to the full series trend.

We chose the start point of this study as 1971 because the 1970s ushered in a new epoch of re-emerging neighbourhood inequality in Canadian cities (Harris 2020, 39; Grant, Walks, and

Ramos 2020, 255). Therefore, we would expect lower levels of change were the start-point of our analysis moved up to a later date.

The high-income bandwidth curves (Figure 2.2) exhibit a decline around 2001, in almost every CMA. However, the high-income group also exhibits the most variability over the period of record and, similarly to the observations on the middle- and low-income groups, such a decline might simply be a spurious observation resulting from natural periodicity that exists at a greater temporal scale than our observed record. Periodicity is conjecture as we have no way of knowing without a longer time series. Thus, the reliance on the global trend is a more robust way to interpret income polarization over time—as opposed to examining changes in the bandwidth curve direction between individual census years and attempting to explain each ‘kink’. The ‘kinks’ are due to some political/social/local process and we cannot conclusively say that the variability through the period of record is just random variation, it could very well be labor market responses to national and/or provincial policies, changes of national, provincial and municipal governments or any combination thereof plus random variability. Unfortunately, that depth of analysis is beyond the scope of the current research but could prove fruitful in more specific research that aims to understand the observed variability from census year to census year.

We used three bandwidths and analyzed the trend of the average. For clarity, only the middle-income bandwidth (20%) was mapped for all the cities, but similar trends of an expanding low-income area, as depicted in red on our maps (Figure 2.3 and S6 in Appendix II), can be observed irrespective of the range used to define middle-income. These maps were not produced, but the household density equalizing cartograms show similar patterns under other bandwidths of the middle-income group.

The spatial distribution of the low-income area (Figure 2.3) leads to the inference that suburban areas are increasingly becoming low-income areas. Such spatial patterning is consistent with research which has shown that between 1986 and 2006, poverty and impoverished places have increasingly shifted from the inner-city towards suburbs of large Canadian CMAs (Ades, Apparicio, and Séguin 2012; Pavlic and Qian 2014).

CMAs are increasingly divided, creating a landscape mosaic of concentrated disadvantaged and advantaged groups that we infer from the increasing trends in patch fragmentation over time in all CMAs (Figures 2.3 and 2.5). Every CMA has become more fragmented (Figure 2.5, last two columns). Increasing fragmentation over time and the division of space that ensues is a defining feature of postmodern urban landscapes (Dear and Flusty 1998, 66).

The increasing fragmentation of income groups has implications because socioeconomic class remains an important factor in one's development and opportunities in life. Opportunities in employment, education, health and access to amenities, including health, food, recreational areas and services, are structured by inequalities in a divided urban landscape. For example, affluent parts of cities have more resources and hence opportunities than disadvantages areas (Holifield 2001; Quastel 2009; McKenzie 2014; Stehlin and Tarr 2017; Banzhaf, Ma, and Timmins 2019). The juxtaposition of concentrated advantaged groups and an expanded area of disadvantaged groups produces negative prospects for social cohesion and may contribute to an increased potential of detrimental social backlash.

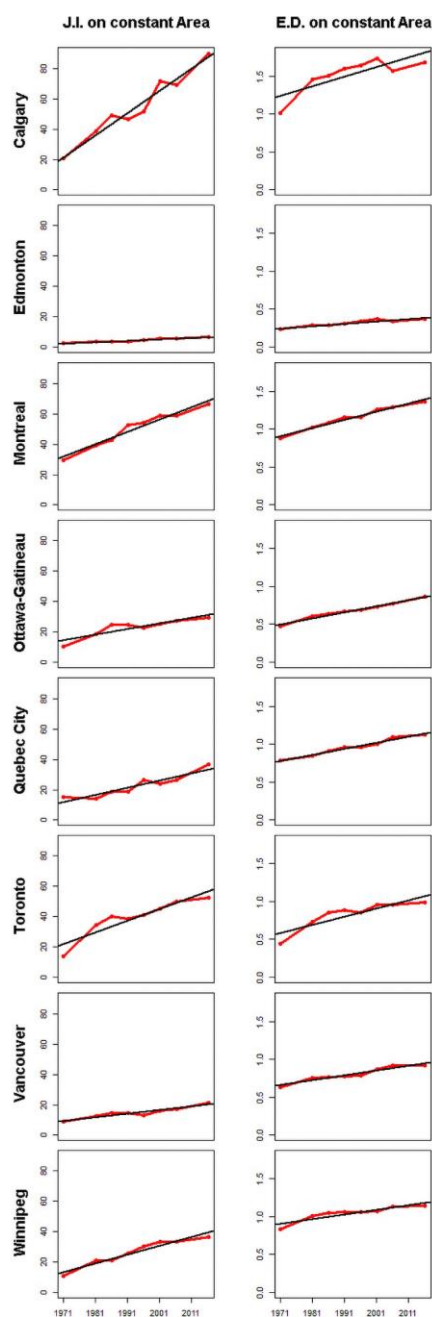
While the fragmentation indices inform as to what is happening at the spatial scale of the patches that are created by the distribution of high-, medium- and low-income groups in the CMA mosaics, the spatial autocorrelation analysis deals with the behavior of the constituent

census tracts (CTs) that create the patches within the urban landscape. The changing nature of spatial autocorrelation among the constituent CTs over time can help to reveal potential spatial processes/interactions that have led to the patch formations through the changes in income status of individual CTs over time. Over time, changes in positive spatial autocorrelation can inform as to the nature of spatial interaction between income-group CTs. For example, do certain income-group CTs attract (cluster), repulse (disperse) or show no change in interaction over time?

With respect to the spatial behavior of individual income groups, the statistically significantly increasing positive spatial autocorrelation trends in all but three of the CMAs for the high-income group equates to a form of spatial interaction whereby, in most CMAs, the high-income CTs tends to attract more high income CTs to form their patches over time. The remaining three CMAs also exhibit positive trends in the spatial autocorrelation that, while non-significant, does reinforce the observation. The growth of high-income patches is due to a spatial process of clustering of high-income CTs. Processes underlying this can be many, for example, gentrification in general or spillover gentrification effects in particular, would be local processes that can lead to this type of clustering among high-income CTs. Other processes include professionalization (Hamnett 1994), a process in which individuals become upwardly socially mobile through the attainment of professional skills.

Gentrification is a socio-spatial process that is operating in every CMA among the inner-city CTs around downtown cores (Figure 2.5), and this is well documented in past research (Caulfield 1994; Ley 1996a; Anderson et al. 2005; Logan and Vachon 2009; Peterson 2013; Zwicker, Supernant, and Luckert 2017; Bunce 2018; Ilic, Sawada, and Zarzelli 2019; Evans, Collins, and Chai 2019). Maps of inner-city areas and their creation are provided in the appendix (S12 in Appendix II). Coincident with gentrification of the downtown core, many CTs will show

increases in average household incomes. Such change reflects a more affluent population creating emerging clusters of wealth which replace previous populations that had lower or middle-incomes.



**Figure 2.5. Diagrams showing fragments and fragmentation indexes for Canada’s eight largest CMAAs.** The first and second columns show results from two fragmentation indices on a constant area over time—the Johnson fragmentation index (F.I. on constant area) and Edge Density (E.D. on a constant area). Johnson Fragmentation Index values were multiplied by a million. Confidence intervals from bootstrapping are available in the supplementary material (S8 in Appendix II). Every panel in this figure has significant trends at an  $\alpha = 0.05$ .

The low-income CTs exhibited, for the most part, trends of positive spatial autocorrelation that increased over time, but these were not statistically significant. While the total area within the CMAs that is occupied by the low-income group is growing, that growth may be more opportunistic: middle-income CTs are scavenged wherever they occur. However, the positive trends over time are suggestive of weaker process whereby some low-income CTs are more likely to end up near low-income CTs to form patches. An additional potential factor that contributes to this result is a compositional effect, as there is an increase over time of one- and two-person income households (Massey and Denton 1988; Townshend and Walker 2002). However, the rise of one and two person households is not spatially equal or homogeneous and hence one should be wary of generalizing this observation across all temporal and spatial domains.

The middle-income group CTs in Winnipeg and Montreal show a significant decreasing trend in spatial autocorrelation consistent with CTs where the space around them is dissimilar with respect to income-groups over time. That is consistent with dispersion of middle-income CTs within the landscape mosaic over time as well as fragmentation at the CT level and patch level. The decreasing magnitude of positive spatial autocorrelation could be due to the disappearing middle-income groups in these CMAs. However, six of the eight CMAs have no significant trends in the spatial autocorrelation of the middle-income groups over time. While fragmentation is increasing at the patch level, the behavior of middle-income CTs clustering with other middle-income CTs has not changed over time and this is likely due to the fact that this group has been disappearing from all CMA mosaics. Perhaps, from a spatial perspective, that is a weakness: middle-income CTs are not increasingly attracting other middle-income CTs to their patches and so are more vulnerable to 'predation'.

Perhaps the lack of statistically significant positive or negative trends in the spatial autocorrelation among middle- and low-income CTs, makes these more vulnerable to being appropriated by the CTs of the high-income group that shows increasing clustering over time. In short, the CMAs are becoming more fragmented and only the high-income group CTs exhibit increasing spatial organization in the form of clustering over time. The high-income group CTs are the most spatially organized and in some sense this group is in ‘control’ of the mosaic. Similar research has found that urban areas all have their own unique spatial and temporal specificities, whereby differences can be manifested in varying patterns of spatial segregation: unevenness, clustering, exposure, etc (Massey and Denton 1988; Townshend and Walker 2002).

The underlying explanation that has produced the income polarization and emergent fragmented spatial mosaics is multifaceted and includes neoliberal economic policies (which work towards abating the welfare state), globalization, and immigration. Delving deeply into these components is beyond the scope of our study and evidence, but in the following paragraphs it is worth conjecturing how these mechanisms could potentially influence the patterns we have found.

The dismantling of the Canadian welfare state can be seen through the dismantling of various benefits, and these vary both among provinces and within provinces (Banting and Myles 2013; Battle, Mendelson, and Torjman 2006). Additionally, the 1980s and 1990s saw the polarization of the labor market (Green and Sand 2015), and consequently a reduction of the population in the middle-income category is to be expected.

Immigration plays a role in the evolving income structure of the city and is the product of two concurrent forces. The first is that suburban regions are increasingly becoming reception zones of immigrants as the inner-city becomes less affordable (Walks and Maaranen 2008a),

whereas the second is an unfortunate pattern whereby incomes of recent immigrants, in relation to that of native born Canadians, have been significantly decreasing since the 1980s (Li 2000; Picot and Hou 2003; Walks 2015, 154). In Anglo-America, immigrant reception areas are increasingly in suburban areas more so than the inner-city (Murdie 2008; Ray and Preston 2009; Singer 2013), and coincidentally many suburban areas are comprised of low-income CTs.

In the US, the urban structure is extremely divided (often along racial lines) (Fainstein, Gordon, and Harloe 1992; Jargowsky 1997; Hackworth 2005; Wilson 2006; Surgue 2014; Rothstein 2017; Mallach 2018). A dated but popular argument made by Goldberg and Mercer (1986) suggests that Canadian cities are ostensibly distinct from their American counterparts. However, as this paper demonstrates empirical evidence of the rising income polarization in Canadian cities as well as how that trend is leading to a more spatially divided urban mosaic that resembles American counterparts.

## **Conclusion**

Canada's fortuity in terms of egalitarianism is uncertain. As noted, amongst OECD countries, Canada has seen the highest increase of income inequality in recent years (Banting and Myles 2013; Beach 2016; Breau 2015; Fortin et al. 2012; Osberg 2008; Myles, Picot, and Pyper 2000; Fong 2017). Inequality is exemplified at the urban scale, where we illustrate a disappearing middle-income populace both spatially and temporally that is expressed by an increasingly spatially fragmented urban income mosaic. This study has demonstrated the unequivocal decreases of middle-income groups over time and coincident increases in the low-income groups in the eight largest CMAs in Canada.

This research finds significant growth in low-income areas within every CMA examined, many of which are located in declining areas that are often referred to as inner-suburbs.

Concurrently, levels of fragmentation have increased universally with only the high-income groups exhibiting any increase in spatial organization through the process of spatial clustering of similar CTs. This study provides a base for a more thorough examination of the lived experiences of individuals in these increasingly divided Canadian urban landscape mosaics.

Finally, our research has identified places in CMAs where policy makers can focus on, the actions of which may lead to the fostering of middle-income communities and the confrontation of what some have termed the “menace of suburban decline” (Smith, Caris, and Wyly 2001). Such endeavor may mitigate or reverse the decline of middle-income groups in metropolitan areas. Whereas this research does not postulate remedies to inequality and polarization, it should be noted that in order to successfully combat the aforementioned problems, a multi-pronged approach is required (Walks 2015, 171).

### **Supporting information**

Appendix II. Supplementary information for ‘The Temporal Evolution of Income Polarization in Canada’s Largest CMAs’

### **Acknowledgments**

We would like to thank the time and feedback that the editor, Celine Rozenblat, and the reviewers, Mario Polèse and one anonymous reviewer for their detailed comments and suggestions that clearly improved this manuscript.

This work is a contribution to the Ottawa Neighbourhood Study ([www.neighbourhoodstudy.ca](http://www.neighbourhoodstudy.ca)).

### Chapter 3. Deep mapping gentrification in a large Canadian city using deep learning and Google Street View

**Abstract:** Gentrification is multidimensional and complex, but there is general agreement that visible changes to neighbourhoods are a clear manifestation of the process. Recent advances in computer vision and deep learning provide a unique opportunity to support automated mapping or ‘deep mapping’ of perceptual environmental attributes. We present a Siamese convolutional neural network (SCNN) that automatically detects gentrification-like visual changes in temporal sequences of Google Street View (GSV) images. Our SCNN achieves 95.6% test accuracy and is subsequently applied to GSV sequences at 86110 individual properties over a 9-year period in Ottawa, Canada. We use Kernel Density Estimation (KDE) to produce maps that illustrate where the spatial concentration of visual property improvements was highest within the study area at different times from 2007–2016. We find strong concordance between the mapped SCNN results and the spatial distribution of building permits in the City of Ottawa from 2011 to 2016. Our mapped results confirm those urban areas that are known to be undergoing gentrification as well as revealing areas undergoing gentrification that were previously unknown. Our approach differs from previous works because we examine the atomic unit of gentrification, namely, the individual property, for visual property improvements over time and we rely on KDE to describe regions of high spatial intensity that are indicative of gentrification processes.

#### Introduction

Working in London, UK, Glass (1964), coined the term ‘gentrification’ to describe the process whereby the working class is displaced by upper classes (the gentry) in urban space. Class displacement resulting in gentrification has been observed in many western cities (Laska and Spain 1980; Palen and London 1984; Smith and Williams 1986). However, the conceptualization of gentrification has expanded over time to focus on the driving factors of the process such as culture and consumption (Caulfield 1994; Ley 1996b; Butler 1997), political economy and production (Smith 1996; Wilson 1990; 1991; Wyly and Hammel 1999) and, indeed, the definition has extended to include different groups of gentrifiers ranging from marginal (Rose 1984) and middle class (Butler 1997) to the working class (Paton 2014). Current

thought on gentrification is intersectional, examining how various discourses are implicated, such as bodies (Wilson, Wouters, and Grammenos 2004) or industrial spaces (Curran 2004) and, as such, gentrification has been identified as taking place where it was once not seen as possible (Newman and Ashton 2004). Recent years have witnessed an expansion of the topic into tourism (Gravari-Barbas and Guinand 2017), planning and policy impacts (Bunce 2018), environment (Gould and Lewis 2017; Curran and Hamilton 2018), and many other specialized topics (Krase and DeSena 2016; Lees, Shin, and López Morales 2016; Moskowitz 2017; Schlichtman, Hill, and Patch 2017; Albet and Benach 2017; Curran 2018). The Economist tweeted a quote by Dyckhoff claiming that gentrification is “the most significant force in Western cities in the second half of the 20th century.” (Economist 2017a; 2017b)

When a higher socioeconomic class subjugates an existing urban locale, there are visible changes to the building stock. Visually/aesthetically, housing stock re-investment is the most acute indicator of gentrification (Hammel and Wyly 1996, 251). As indicators of gentrification, visible changes to building stock can manifest in several ways. Given that most studies on gentrification are concerned with the discourse, politics and socio-cultural aspects that shape the gentrification process, the characteristic aesthetic changes taking place are often ignored even though they are the primary external indicator of the process. As a result, there is a literature gap in measuring and quantifying the visual expression of the gentrification processes, probably because as Hammel and Wyly (1996, 248) rightly note, the visual expressions of gentrification processes are difficult to observe and measure.

In short, property improvement over time is an important sign of gentrification and is perhaps one of the most important yet overlooked indicators of the process. Jager (1986) has demonstrated that various types of renovation and restoration initiatives serve to assert social

rank, because housing confers the attributes of social status and prestige. Redfern (2003, 2353) notes that gentrification cannot emerge if property improvement is not present. As Gentrification is “the production of urban space for progressively more affluent users” (Hackworth 2002, 815), it is fair to assume that a number of visual indicators reveal this process in a given location.

Some scholars have analyzed gentrification based on the direct auditing of the built environment. Direct auditing methods are those that require people to manually examine the GSV images (or the real world) for perceptual attributes (Wu et al. 2014; Lafontaine, Sawada, and Kristjansson 2017). Using direct auditing methods, Hammel and Wyly (1996, 250 Table 1.) sought visible signs of reinvestment in both single family homes and multi-unit buildings. Features observed in single unit homes included property repainting, changes in the degree of ornamentation, reconstruction of porches, steps, window replacement, fences, and so forth. Meanwhile, multi-unit housing would have some other visible features such as entryway and signage changes, porch furniture, new brickwork/siding, and the like (Hammel and Wyly 1996, 250–251).

**Table 3.1. Results of model after fine-tuning.**

	# of Training batches	Number of epochs per batch	Iterations with batch size of 24	Validation accuracy	Test accuracy
Fully connected layers (SCNN-FC)	8	50	70000	92.2	-
Fully connected layers and top 4 convolutional (SCNN-FC-4)	8	50	70000	94.6	-
FC + top 8 convolutional layers (SCNN-FC-8)	4	50	35000	95.2	95.6%

Heidkamp and Lucas (2006) examined additional visual changes in gentrification processes, including non-residential upgrades, the renovation of commercial space, improvements to the street infrastructure, and even the noticeable presence of vehicles that represent upgraded transportation. Streets were sometimes redesigned to imitate cobblestones of the past. The transformation of commercial space is noteworthy because it too gentrifies and caters to consumer preferences of the higher class (Hubbard 2017).

The aforementioned studies identify numerous visual impacts that indicate gentrification to properties. However, technological availability limited those studies to small spatial domains (a few city blocks) because of the onerous time demands that primary observational work requires (Heidkamp and Lucas 2006, 113). It is useful to note that Google Street View (GSV), by itself, reduces the time required for direct auditing research by eliminating field logistics in the real-world and this leads to as much as 50% time savings when collecting data (Lafontaine, Sawada, and Kristjansson 2017; Ploeger et al. 2016).

Google Street View (GSV) data provides opportunities to analyze urban space for visual gentrification-like changes at the atomic level of the gentrification process: the property level—the smallest spatial unit upon which a gentrification process can select/act upon—and at frequencies higher than the bi-decadal or decadal census. Moreover, with GSV, potentially every urban structure can be examined for visual changes through time (back to 2007 in most cases in the United States and Canada). While GSV yields large quantities of data, manual examination of each individual structure's GSV images over time for the hallmark visual indicators of gentrification is untenable in a large urban center. However, recent advances in machine learning and computer vision can be leveraged to examine each structure over time in a large urban center

to produce detailed maps of visual changes. Machine learning applied to GSV imagery has the potential to revolutionize the study of the spatial expression of the gentrification process.

The intention of this paper is not to debate the potential benefits or drawbacks of gentrification but rather to assess the potential of machine learning to detect the visual process of gentrification. Herein we propose a machine learning approach using deep convolutional neural networks to examine hundreds of thousands of GSV images to detect the associated visual impacts that are characteristic of urban renewal and change.

### **Deep mapping and the built environment**

#### *Google Street View (GSV) in auditing the urban environment*

GSV has been used to uncover intricate details in neighborhoods using direct auditing methods. For example, research in the UK suggests that GSV is an accurate tool for ascertaining if a neighborhood is struggling (Odgers et al. 2012, 1015). GSV has also been used to virtually evaluate physical environmental characteristics along cycling routes (Vanwolleghem et al. 2014). In general, the use of GSV has a longer history within the development of systematic social observation instruments (SSOIs) applied to the urban environments (Clarke et al. 2010; Badland et al. 2010; Griew et al. 2013; Kelly et al. 2013; Rundle et al. 2011; Wilson et al. 2012).

Few gentrification studies have been conducted using Google Street View (GSV) as a data source. A study by Hwang and Sampson (2014) was one of the first to use GSV to measure gentrification in some manner. Street-level panoramic imagery offers a unique perspective because the update interval is relatively short compared to other data sources used in gentrification studies such as census data. GSVs of a property are separated by time intervals of between one to three years. Such imagery can be used to understand short-term gentrification in addition to detecting the process in the initial stages.

### *Machine mapping and deep mapping*

The production of spatially detailed maps illustrating gentrification processes across entire cities is beyond the ability of most research that uses GSV in direct auditing methodologies. There have been limited studies that have used machine learning techniques with GSV for the sole purpose of mapping perceptual attributes of urban spaces. We refer to the general approach of using machine learning techniques such as computer vision algorithms (GIST, Texton histograms, etc..) applied to street-level imagery for mapping the perceptual environment as ‘machine mapping’. Within machine mapping, we propose “deep mapping” to denote the use of deep learning approaches that employ convolutional neural networks (CNN) trained with street-level imagery to produce spatial data that is used for mapping or within other geocentric analyses.

*Machine mapping.* Salesses, Schechtner, & Hidalgo (2013) produced the first large-scale machine mapping study using GSV. They mapped the perception of safety in select US cities. They used crowd-sourced data from over 7000 individuals from 91 countries who responded to a website with a game that showed pairwise comparisons of GSV images from Boston, New York, Linz and Salzburg and asked the question “Which looks safer?”, the left or right image. Given that each image was compared to numerous others in the online game, the authors were able to compute a ‘Q-score’ for each image that provided a relative ranking of a each of the images along the perceptual dimension considered. The crowd-sourced dataset they produced was made available and called Place Pulse 1.0 (Salesses, Schechtner, and Hidalgo 2013).

Naik et al. (2014; 2015) subsequently used the Place Pulse 1.0 dataset to compute a ‘Streetscore’ for 21 US cities. They used the Trueskill ranking algorithm to rank the images in the Place Pulse 1.0 dataset. Using those rankings together with corresponding state-of-the-art

feature vectors (GIST, HOG2x2, SIFT etc.), they employed support vector regression to predict the perceived Streetscore for over one million GSV images. Their study produced the first high resolution street-level maps derived from computer vision and machine learning techniques.

Naik et al. (2017) used the Streetscore to predict the ‘Streetchange’ of over one and a half million GSV images in several US cities. The change in Streetscore values for a given pair of images at successive time intervals at the same location was used to compute the ‘Streetchange’. They only included the years of 2007 and 2014. A positive Streetchange score indicated that the perception of visual safety of a street increased between two time periods. This study was innovative in that it applied machine learning and computer vision techniques to GSV images to map out physical improvements at fine spatial scales in order to provide a basis to test three theories of urban change. Porzi et al. (2015), however, found that predicting the perceived safety level between two GSV images, the process upon which Streetchange is based, is significantly improved using CNN when compared to traditional techniques that are based on feature extraction (GIST, HOG etc..) and ranking support vector regression algorithms.

*Deep mapping.* Within machine mapping studies, there have been several applications of deep mapping. For example, to better understand which features (e.g., canal, valley, trees) lead to scenic beauty, Seresinhe et al. (2017) used a CNN to extract features from images that had been rated for their scenicness through crowd-sourcing. They then used a CNN to predict the scenicness of unrated images to produce a map of scenicness across London, UK.

In 2017, Gebru et al. (2017) used a CNN and 50 million GSV images to classify and map 22 million vehicle types in the US and compare the results to census data. Liu et al (2017) used a CNN and street-level images taken by Baidu (China’s equivalent to GSV) to map out the quality of buildings across Beijing, China. Similarly, a team has combined GSV images and Google

Maps 3D as input to a CNN in order to map building frontage quality on a structure by structure basis across London, UK (Law, Shen, and Seresinhe 2017). More recently, CNNs have been utilized to classify buildings into semantic categories in several cities using Street View images (Kang et al. 2018).

Dubey et al. (2016) developed the Place Pulse 2.0 dataset containing over one million pairwise comparisons of GSV images, from several countries, along eight perceptual dimensions. In contrast to Naik et al. (2014; 2017), Dubey et al. (2016) used a deep mapping approach and trained a Siamese fully-convolutional neural network to both predict pairwise comparisons for perceptual attributes as well as a second CNN with fully connected top layers to rank images according to a given perceptual attribute. The prediction accuracy of Dubey et al. (2016) was ~73%, largely because Place Pulse 2.0 data is heterogenous, derived from pairwise image comparisons among 28 different countries. People from different cultural milieus will introduce ontological and semantic inconsistencies in the dataset and this ultimately affects the generalization potential of a CNN model. However, the biases in the dataset may be relatively minor. Salesses, Schechtner, and Hidalgo (2013), found little bias due to age, gender or location (US vs. non-US) in the Place Pulse 1.0 dataset but that was relatively small compared to Place Pulse 2.0. To improve predictability of deep mapping models, Blečić et al. (2018) collected over 17,000 GSV images in Italy and, with each being rated on a perceived ‘walkability’ scale of 1–5, they trained a CNN to rate GSV images and map their walkability scores for several Italian cities. The authors achieved accuracy of 78% and 99% (for ‘1-class-off’). The results of Blečić et al. (2018), suggest that deep mapping applications may achieve improved prediction performance when trained with regionally consistent GSV datasets.

The aforementioned studies demonstrate that machine learning techniques applied to GSV imagery provide a monumental leap in our ability to classify and map urban space for several kinds of perceptual/semantic attributes. For example, a machine should be able to learn how to compare two sequential images and identify if a gentrification-like visual change has taken place. By mapping the locations of GSV sequences that contain a visual change, gentrification-like visual changes can be mapped at highly detailed spatial scales over large spatial domains.

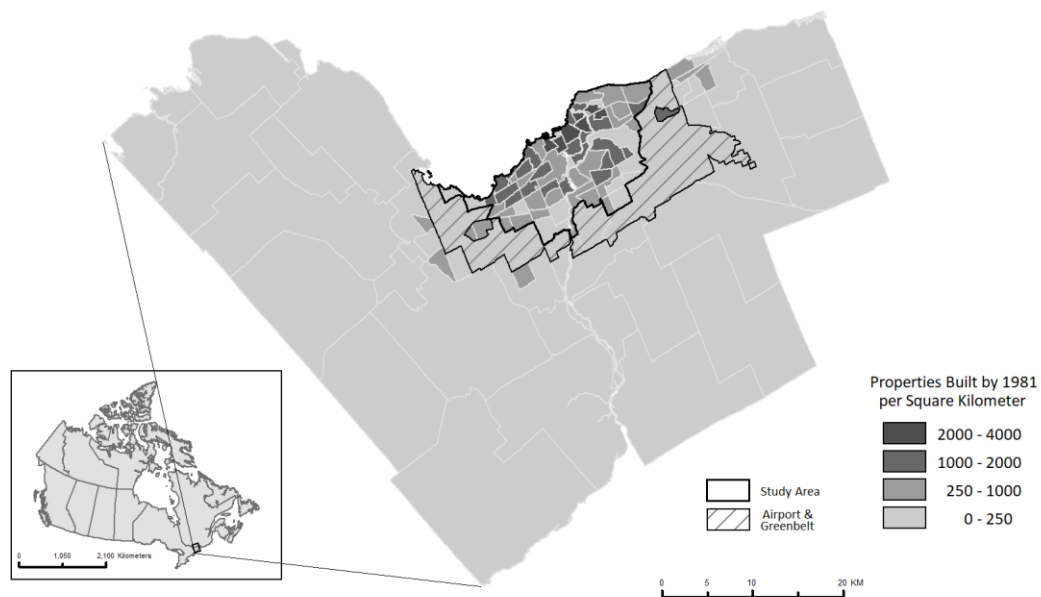
### **Methodology**

Herein, we adopt a deep mapping approach and employ a Siamese-CNN (SCNN) to detect improvements in the frontage quality (building structure plus the front of the property) of individual properties using imagery through time from GSV in Ottawa, Canada. Our approach differs from previous works because we examine the atomic unit of gentrification, namely, the individual property, for visual improvements over time that are indicative of gentrification processes. Because individual properties can undergo a visual gentrification-like change anywhere in geographic space, the SCNN becomes a detector and we rely on mapping the SCNN's detections to identify clusters of visual changes in space. Our reasoning is as follows: If there is a spatial concentration (a high intensity) of increasingly positive changes in the physical appearance of several properties in close proximity, then this spatial concentration is indicative of a gentrification-like process. However, any given region of high spatial intensity cannot definitively confirm gentrification as the causal factor: to do so would require information on the socioeconomic and cultural changes within and around that space. Nevertheless, as discussed in the introduction, housing stock re-investment and the visual changes that ensue are an acute indicator of gentrification (Hammel and Wyly 1996). Therefore, while not confirmative, regions

exhibiting a high intensity of visual property upgrades provide a spatial hypothesis that can be tested by examination of other factors implicated in the gentrification process.

### *Study area*

The City of Ottawa is Canada’s capital and has a population of 934,000 (metropolitan, 1.32 million) in 2016 (Statistics Canada 2017a). Within the city boundaries, the region containing the oldest building stock (Figure 3.1) is most likely susceptible to gentrification pressures. The region of the city that falls within the ‘Greenbelt’—an agricultural land use zone intended to stop urban sprawl—contains the highest densities of pre-1981 structures and thus defines our study area. For every property within the study area, the SCNN will search for gentrification-like visual changes in the GSV time sequence of images. Moreover, the study area is the only contiguous region that contains GSV images beginning in 2007 (Section S3 in Appendix IV).



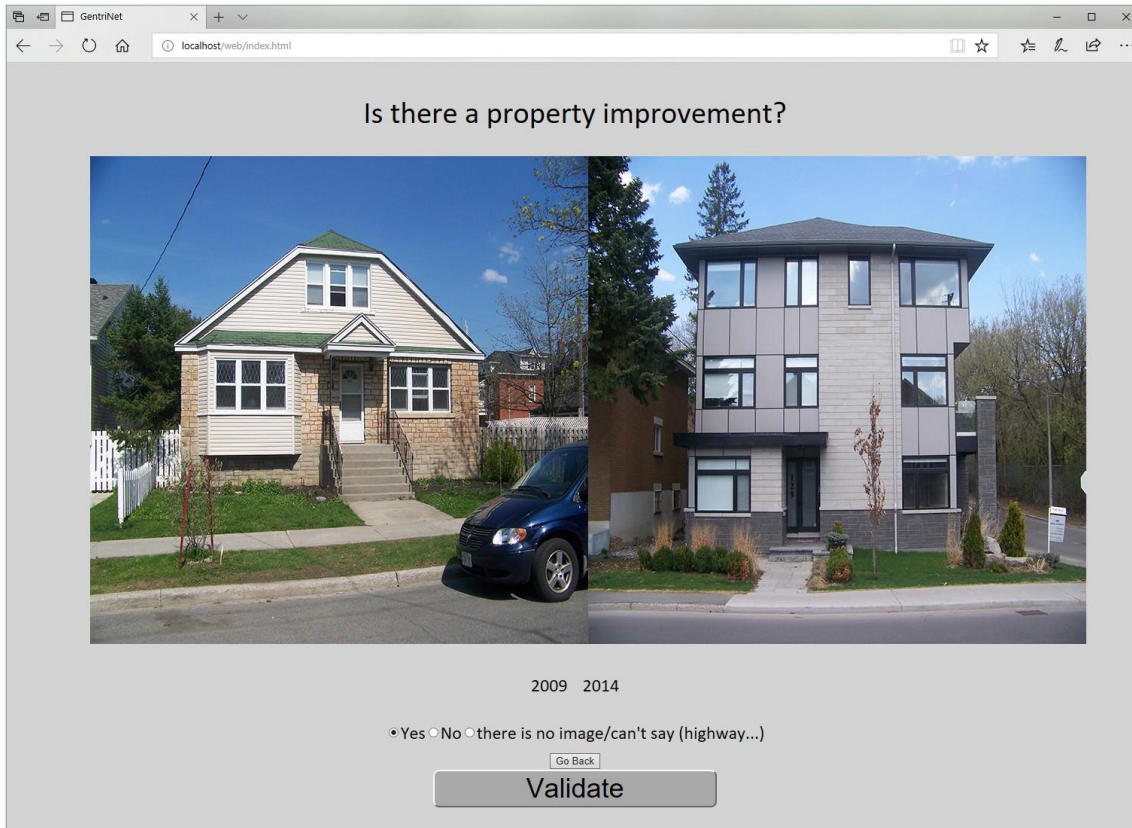
**Figure 3.1. The study area contains the oldest building stock in Ottawa and is within the Greenbelt.**

### *Training data*

Google Maps Application Programming Interface (API) was used to access panoramas (full 360° hemispherical views) of every property within the study area as well as within suburban regions that were adjacent to the greenbelt. When this work was completed, prior to July 16th, 2018, our use of GSV complied with the Google Maps APIs Terms of Service. The City of Ottawa building footprints and street network (City of Ottawa, n.d.) were used to calculate a vector perpendicular to the street and looking at each building structure. The origin (longitude / latitude), look direction and field of view were fed to the Street View API to clip a GSV image from all the time-stamped panoramas at each location. There were 157,303 properties at which panoramas were available, leading to a final count of 593,723 clipped GSV images (Figure A(A) in Appendix IV). We then selected a subset of 9462 properties, containing 16,224 individual images, from several neighbourhoods for training the SCNN (Figure A(B) in Appendix IV).

The GSV dataset used to train the model was comprised of 16,224 paired image comparisons, of which, 1307 (at 1275 unique locations) were positive and 14,917 pairs (at 8594 unique locations) were negative for a gentrification-like visual change. The sum of unique locations for positive and negative physical changes ( $n = 9869$ ) is greater than the sum of unique locations for the entire training set ( $n = 9462$ ) because, at a given location, there can be multiple pairs that have a negative for change or positive for change. Following the data collection protocol of Place Pulse 1.0 (Salesses, Schechtner, and Hidalgo 2013) and Place Pulse 2.0 (Naik et al. 2015; Naik et al. 2017), we created the training dataset using a webpage in which a property (long,lat) was selected and the user was shown two sequential GSV images within the sequence of time-stamped images at that location and, for each pair, asked to respond to the

question: ‘Is there a property improvement?’ (Figure 3.2). The authors of this paper undertook the comparisons, with it being understood that we were concerned with property improvements that are identifiable as gentrification-like visual changes. A gentrification-like visual change to a property is anything that provides evidence of a significant reinvestment in the property being viewed, including such things as the removal of an older house and construction of a new house, significant updates to the facade of an existing structure, or a clear improvement in landscaping and other relevant visual changes that are suggestive of the visual cues of gentrification including those identified by (Heidkamp and Lucas 2006; Hammel and Wyly 1996). Unlike the Place Pulse dataset, our line of investigation was centered on the atomic unit of gentrification: the individual property, which includes the building structure and frontage together rather than a default GSV camera orientation. Thus, using the webpage (Figure 3.2), we were able to record when a GSV image pair exhibited a gentrification-like visual change between each successive image for the same property. The result of each comparison was then saved in a file, recording the unique GSV assigned image identifier (panoid) for the left image, right image and yes or no for property improvement (S2 in Appendix IV).



**Figure 3.2. Training data collection web interface.** To comply with CC-BY copyright the photos are the author's and presented for illustrative purposes. The photos are similar to those from GSV.

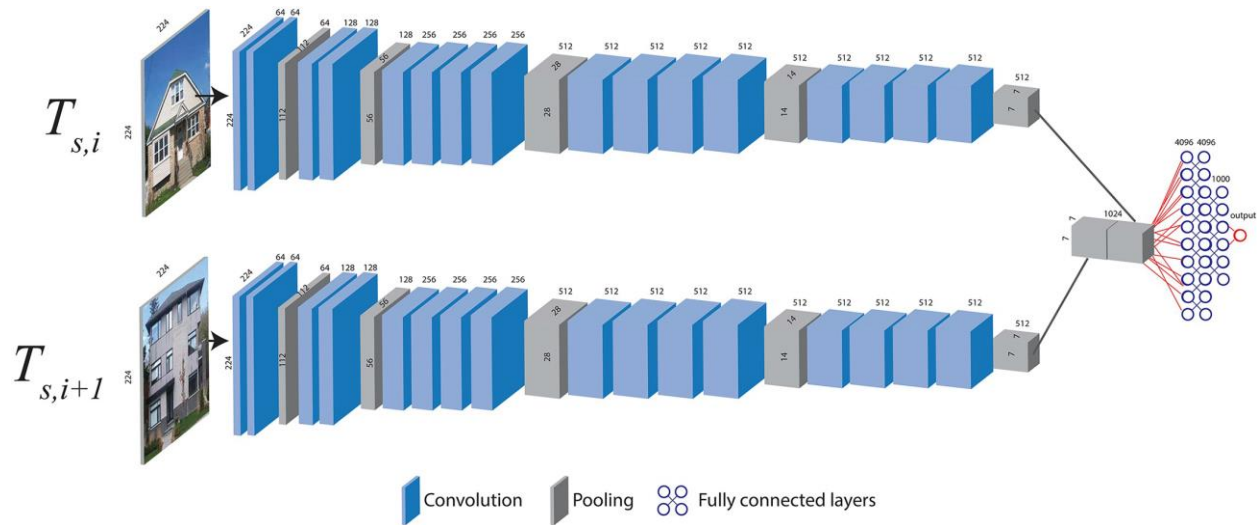
The training dataset was subsequently randomly shuffled and 40% was held back for model validation ( $n = 3245$  pairs) and testing ( $n = 3245$  pairs). The remaining 9734 pairs were used for model training.

### *Siamese model*

We trained a Siamese CNN (SCNN) to detect the visual improvements to a given property between two successive GSV images (Figure 3.3). We programmed the SCNN in python using the Keras library (Chollet and others 2015) and a tensorflow backend (S2 in Appendix IV). An excellent introduction to convolutional neural networks can be found in the book 'Deep Learning With Python' by Francois Chollet (2018).

To train the Siamese model (Figure 3.3), a transfer learning approach was employed (Razavian et al. 2014; Yosinski et al. 2014), whereby the two VGG19 (Simonyan and Zisserman 2014) convolutional branches (Figure 3.3) were initialized using weights trained on the ImageNet (Deng et al. 2009) challenge database (Berg, Deng, and Li 2010). These VGG19 model weights are the result of training on 1.3 million images with 1000 categories (Simonyan and Zisserman 2014). Because the ImageNet weights were produced on a band order of Blue-Green-Red (BGR) images of size 224x224 that were mean-centered across the entire ImageNet database (Simonyan and Zisserman 2014), our dataset of RGB images were likewise resized and channel-wise color normalized by converting to BGR order and mean-centered by subtracting the mean of each image band using the same means computed on the ImageNet dataset (B-103.939,G-116.779,R-223.68).

First, the fully connected fusion layers of the Siamese model (Figure 3.3) and their weights were trained. In the first stage of training, the VGG19 convolutional weights, in both branches, and their resultant feature maps were held constant while the fully connected fusion layers were trained (Glorot normal initialization). To avoid overfitting and for computational tractability, mini-batch training was employed, and three dropout layers were added to the fully connected layers.



**Figure 3.3. Siamese CNN architecture using two VGG19 branches.** The top branch convolves the first image,  $T_{s,i}$ , in the sequence of images at a unique geographic location  $s$ , ( $T_s = \{T_{s,i}: i \in 2 T_s\}$ ,  $s = (x, y)$ ), and the bottom branch convolves the second image  $T_{s,i+1}$  in the sequence. The last pooling block in each branch are concatenated and flattened before being fed to a set of three fully connected layers to produce the output probability using a sigmoid activation. The numbers along the sides of each convolutional layer represent the tensor dimensions (columns and rows), and the numbers across the top of the layers represent the tensor filter block's depth. Photos are the author's and are presented for illustrative purposes.

All training was completed using mini-batches of 4200 image pairs from the training set with image augmentation of each mini-batch. For each pair, within each mini-batch, both images had the same random augmentation applied. An image augmentation consisted of a random image warp (Wong et al. 2016) in some combination of rotation, height and width shift, zoom, shear and horizontal flipping within pre-specified thresholds. The augmentation was the same for the left and right images. Augmentation was used to avoid overfitting and to ensure that only the most relevant features were reinforced within the CNN during training (Wong et al. 2016).

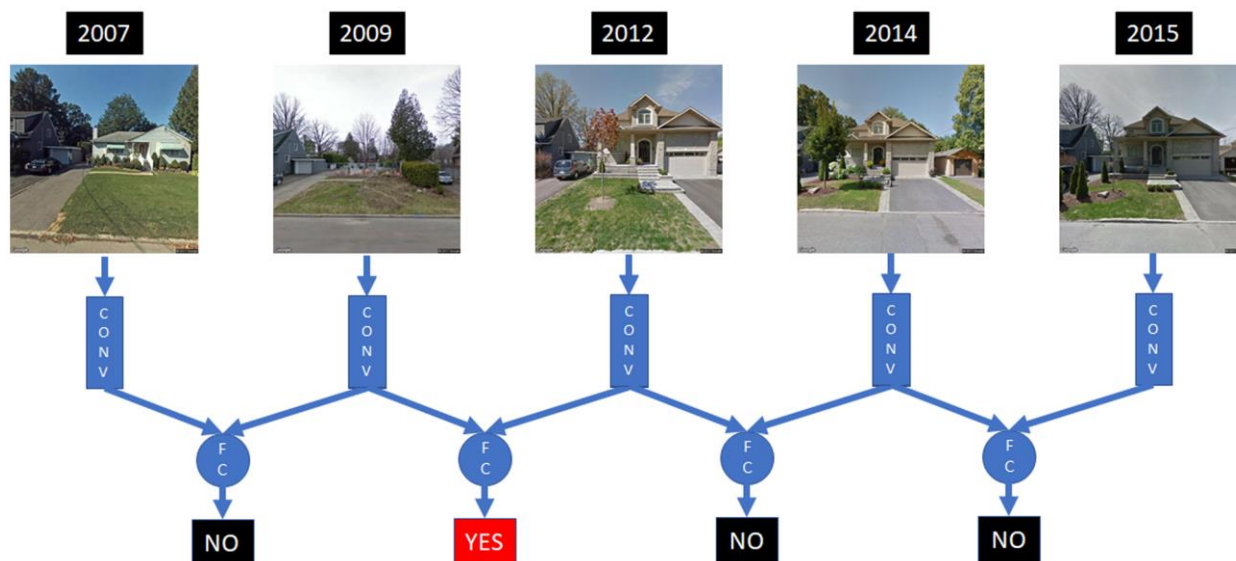
To complete the training of the fully connected layers, eight augmented mini-batches were produced. For each of the eight mini-batches, the model was trained using stochastic gradient descent and 50 epochs with a batch size of 24 for each iteration. After training the fully connected layers (SCNN-FC in Table 3.1), the top four convolution layers were fine-tuned together for a further eight mini-batches (SCNN-FC-4 in Table 3.1). Finally, four mini-batches

were used to fine-tune the fully connected layers and top eight convolution layers in the branches of the VGG19 Siamese model (SCNN-FC-8 in Table 3.1). Because the training dataset was imbalanced in favor of negative cases, class weights were employed to balance model parameter weight penalties during training. The class weights were representative of the overall dataset with 10% importance assigned to negative cases and 90% importance assigned to positive cases. Training time was approximately 33 hours on a single NVIDIA 1080ti GPU.

The final test accuracy of the fully trained model SCNN-FC-8 (~235 M trainable parameters) was 95.6% (Table 3.1), an AUC of 0.84 and an F1 Score of 0.72. All subsequent analyses presented herein use that model. To crudely assess the sensitivity of our model to the partitioning of the GSV training dataset, we repeated the entire training process as described above using a different random permutation of the dataset, that was subsequently partitioned into training, validation and test arrays. This second model achieved a final validation accuracy of 95.2% and test accuracy of 94.4%, an AUC of 0.82 and an F1 Score of 0.66 (S5 in Appendix IV).

### *Detection and mapping*

Using SCNN-FC-8, we detected on channel-wise normalized GSV image sequences at 86110 unique locations (these locations contained 403216 images in total) within the urban extent of Ottawa that contains the oldest building stock (area within the Greenbelt (Figure 3.1)). A single example of the detection process is illustrated in Figure 3.4. From the detected set, we retained those locations (points) that were identified as positive for a gentrification-like visual change as well as the year at which the change was completed. The positive detections and their location were converted to point features and used in Kernel Density Estimation (KDE) to produce mapped results.



**Figure 3.4. Example of detecting gentrification-like change within a sequence of GSV images at the same geographic location using SCNN-FC-8.** In this example, the GSV image from 2007 and 2009 are input into the SCNN-FC-8 and the model detects whether there is a gentrification-like visual change. Only the images between 2009 and 2012 are detected as exhibiting a gentrification-like visual change in the sequence of GSV images. Each detection is recorded along with the geographic coordinates ( $s_{x,y}$ ) of the sequence for subsequent mapping. In order to comply with CC-BY copyright, these photos are the author's and are not GSV imagery and so are provided for illustrative purposes only. Figure with the GSV sequence for this example can be found here: [https://github.com/laggiss/DeepMapping/blob/master/GSVFIGS/fig4\\_online.png](https://github.com/laggiss/DeepMapping/blob/master/GSVFIGS/fig4_online.png).

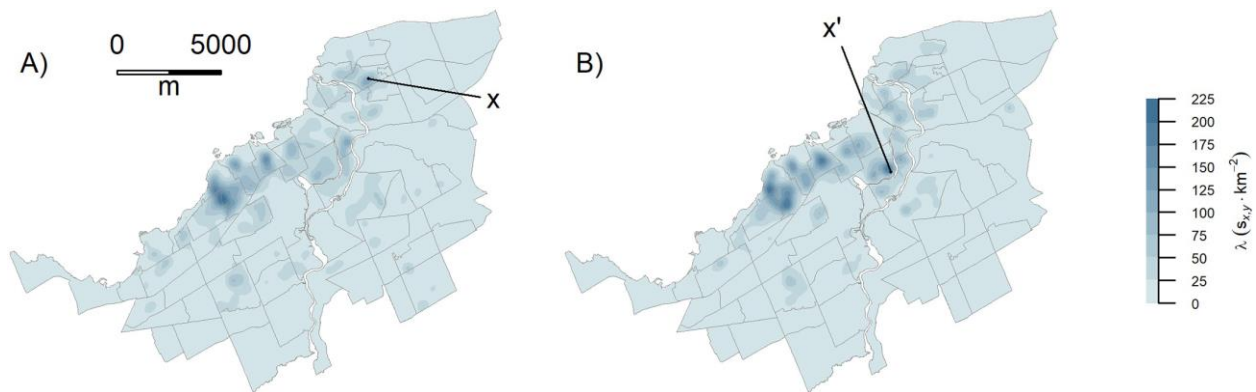
SCNN-FC-8 will detect a gentrification-like visual change between successive GSV images whenever and wherever they occur and cannot judge whether such changes are a consequence of a *bona fide* gentrification processes or regular property maintenance/home improvement. There are many types of gentrification-like visual changes that can occur to individual structures that, when in relative spatial isolation, are not associated with a potential gentrification process, despite mimicking the range of visual changes in the training dataset that we associate with gentrification for individual properties. These are visual analogues of gentrification-like visual changes. So, the model will detect gentrification-like visual changes across the spatial domain of study wherever they arise. We stipulate, however, that such changes must be numerous and occur in relatively close spatial proximity in order to define a potential gentrification process. Therefore, we are concerned with identifying the pattern of the local

spatially varying intensity,  $\lambda$ , where  $\lambda = n/A$ , the number of locations,  $n$ , divided by the area,  $A$ . We used Kernel Density Estimation (KDE) with a Gaussian kernel and bandwidth of 200 m to produce maps that illustrate the spatially varying intensity of gentrification related property improvements within the study area. We used the ‘density()’ function within the R package Spatstat 1.56–0 (Baddeley, Rubak, and Turner 2016) to produce the KDE estimates.

The use of KDE does not imply anything about the continuity/discontinuity of the underlying gentrification process: KDE is used to identify the likely ‘hot-spots’ of gentrification-like visual changes that are a product of the underlying spatial process of gentrification.

#### *Independent validation*

To assess the reliability of the deep mapping approach to detect the property improvements that are visually suggestive of gentrification, we compared the mapped results from SCNN-FC-8 to the spatial distribution of building permits in the City of Ottawa from 2011 to 2016. We used the same Kernel Density estimation procedure described above to map the building permits. These permits do not account for unlawful and/or non-permitted work that takes place but represents the best independent direct measure of changes to individual structures, some of which are expressed visually—depending on the nature of the permitted work. Moreover, not all aesthetic upgrades to a property including landscaping, painting, changes to the façade (siding etc.) require building permits. Within the study area, there were a total of 56,269 permits containing 22,631 non-standardized unique descriptions of the permitted works. Because SCNN-FC-8 can only assess a property as seen from the street, the descriptions of each permitted work were used to prune the dataset to a total of 3986 permits that would most likely produce a visual change that would be viewable from the street (Section S4.1.1 in Appendix IV).



**Figure 3.5. Comparison between model detections and building permits.** A) Kernel density surface of SCNN-FC-8 detections; B) KDE of building permits. Both panels represent data from 2011 onwards. See text for the explanation of the labelled locations ( $x$  is within Vanier North neighbourhood and  $x'$  is within a neighbourhood named Glebe-Dow's Lake). Also see Section S4.1.2 in Appendix IV for further comparisons.

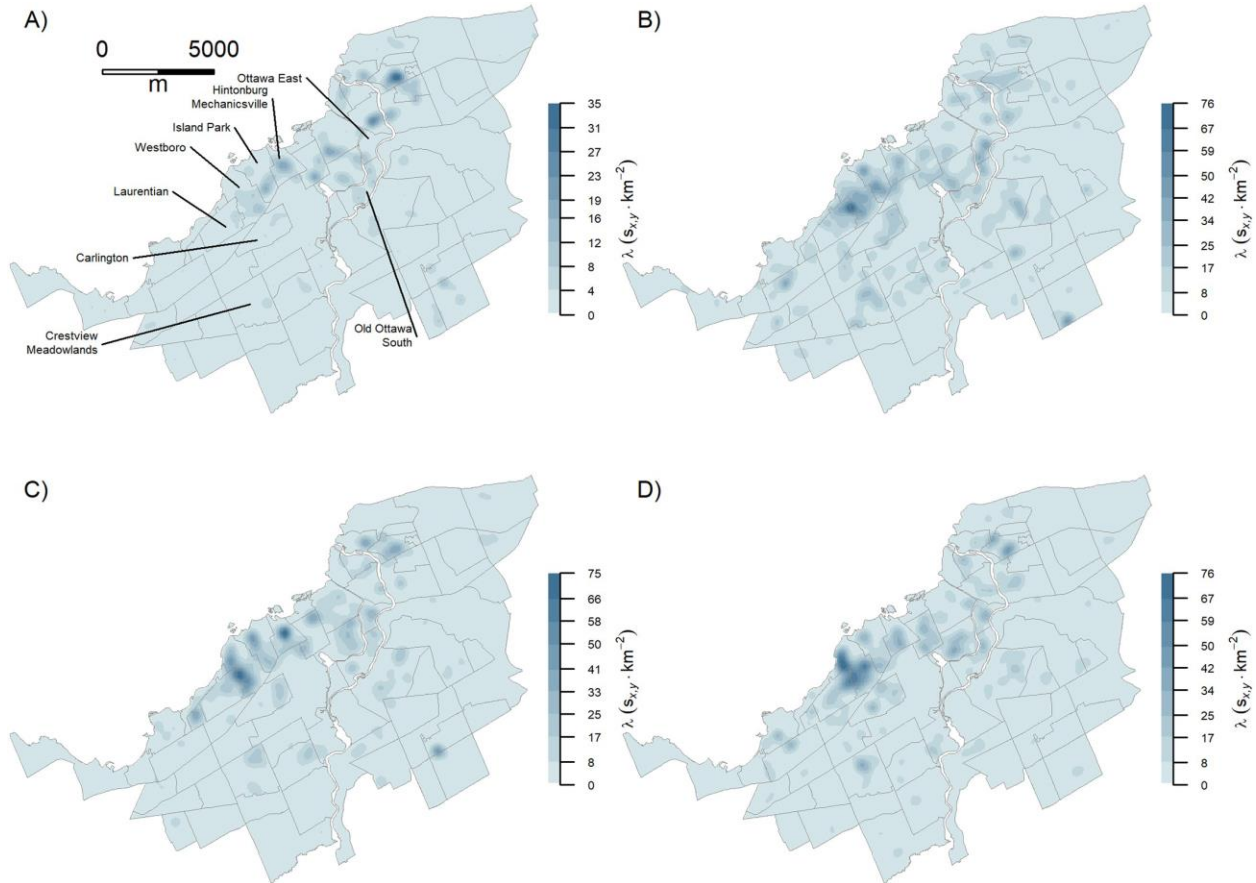
## Results

SCNN-FC-8 detected 3483 instances of gentrification-like visual changes at a total of 2922 unique locations. When multiple visual changes were detected for a property, we retained the most recent date when producing KDE maps. The KDE maps of the detected results and the building permits exhibit very similar patterns (Figure 3.5). Two notable differences in the patterns are labelled in Figure 3.5. The first difference (labelled  $x$  Figure 3.5A) is due to false positive detections that were identified because of a change in the GSV camera, whereby the images along one street were offset for the same geographic locations between 2007 and 2009. The second large difference (labelled  $x'$  on Figure 3.5B) represents permits within a multi-hectare redevelopment of the city's ageing cattle dome/football stadium, the majority of which could not be seen in GSV. A difference map of the standardized intensities highlights both  $x$  and  $x'$  as spatially distinct regions of differing intensities between the KDE of visual changes detected by SCNN-FC-8 and the KDE of building permits (Figure F(A) in Appendix IV).

KDE maps can be produced within the study area at intervals spanning 2–3 years as defined by the update interval of GSV images (Figure 3.6). Because of sparse GSV coverage in 2016 (Figure D in Appendix IV), we combined all the positive detections for 2015–2016 (Figure 3.6D). Changes which are taking place can be seen in areas which are typical for gentrification in classic literature—inner city areas which are close to amenities and which have pleasant architecture. For example, the region of high intensity that is occurring between the Westboro and Laurentian neighbourhoods exhibits a shift northward, locally, in intensity between 2012–2014 and then expands appreciably between 2014–16 (Figure 3.6). However, some areas within the more established regions such as Crestview-Meadowlands, Island Park and Ottawa East are also identified as having regions of high intensity of gentrification-like visual changes to properties from 2012 onwards (Figure 3.6).

## **Discussion**

Despite the pervasiveness of gentrification in modern cities, the focus on social, economic and cultural discourse around the phenomenon has led to a neglect in the development of methods that quantify the temporal and spatial evolution of the phenomenon itself. Typically, to identify gentrification, census data is analyzed over time to identify socioeconomic changes in census tract structure (Ley 1992; Wyly and Hammel 1999; Walks and Maaranen 2008a; Naik et al. 2017). However, as our maps illustrate, gentrification-like visual processes are often localized, particularly in the initial stages and have no natural respect for artificial census boundaries. In general, census data places restrictions on the ability to quantify the process of gentrification at arbitrary temporal and spatial domains.



**Figure 3.6. SCNN-FC-8 detections over time.** A) 2007–2009—with select neighborhoods labelled; B) 2009–2012; (c) 2012–2014; (d) 2014–2015/16.

Spatially, census data implicitly assumes that variables are homogenous within the census unit. Thus, the lack of within-census unit variation limits the spatial resolution of analysis and interpretation to the scale of the census unit. Moreover, because census units are artificial spatial units (e.g., census tracts), interpretations are confounded by the modifiable areal unit problem’s (MAUP) analytical effects of zoning and scale (Openshaw 1983). By way of illustration, if a bona fide gentrification process spans several census unit boundaries, there may not be enough compositional changes in key census indicators within any one of the spatial units to indicate that a gentrification process itself has taken or is taking place, increasing the chances of false negatives when mapping gentrification. Consequently, the delineation of gentrifying or gentrified

areas becomes ambiguous (Heidkamp and Lucas 2006). Temporally, the frequency at which gentrification can be gauged from census data is at five- to ten-year intervals (in Canada and the U.S.). As such, gentrification may often be detected only after it has progressed to a large extent within census tracts. To study the gentrification process spatially, optimally, requires being unconstrained from the spatial and temporal limitations of census data by precisely locating when and where the process is within geographic space.

Our results show that a deep mapping approach centered on the sequence of GSV images of an individual property is successful in spatially resolving likely gentrification processes within urban areas and has potential, upon detailed analysis of the GSV data, to provide a high spatial and temporal resolution that could be of benefit in testing theories of urban renewal. For example, to test the invasion hypothesis, Naik et al. (2017) aggregated their Streetchange and Streetscores to census tracts and showed that a neighborhood spillover effect was present by establishing that the Streetscores of adjacent neighborhoods increased the Streetchange of a given neighbourhood positively through time. Our KDE results potentially allow for the decomposition of spillover effects over time. Recall, for example, that our results detected that the center of high intensity between the Laurentian and Westboro neighbourhoods in 2007–9 moved both north and south along the neighbourhood boundary by 2014–2016 (Figure 3.6). Thus, by focusing on the property as the unit of comparison, we can see a potential spillover effect within the KDE results. However, that being said, we cannot rule out that the observed adjacent and sequential clusters are not due to some spontaneous influence.

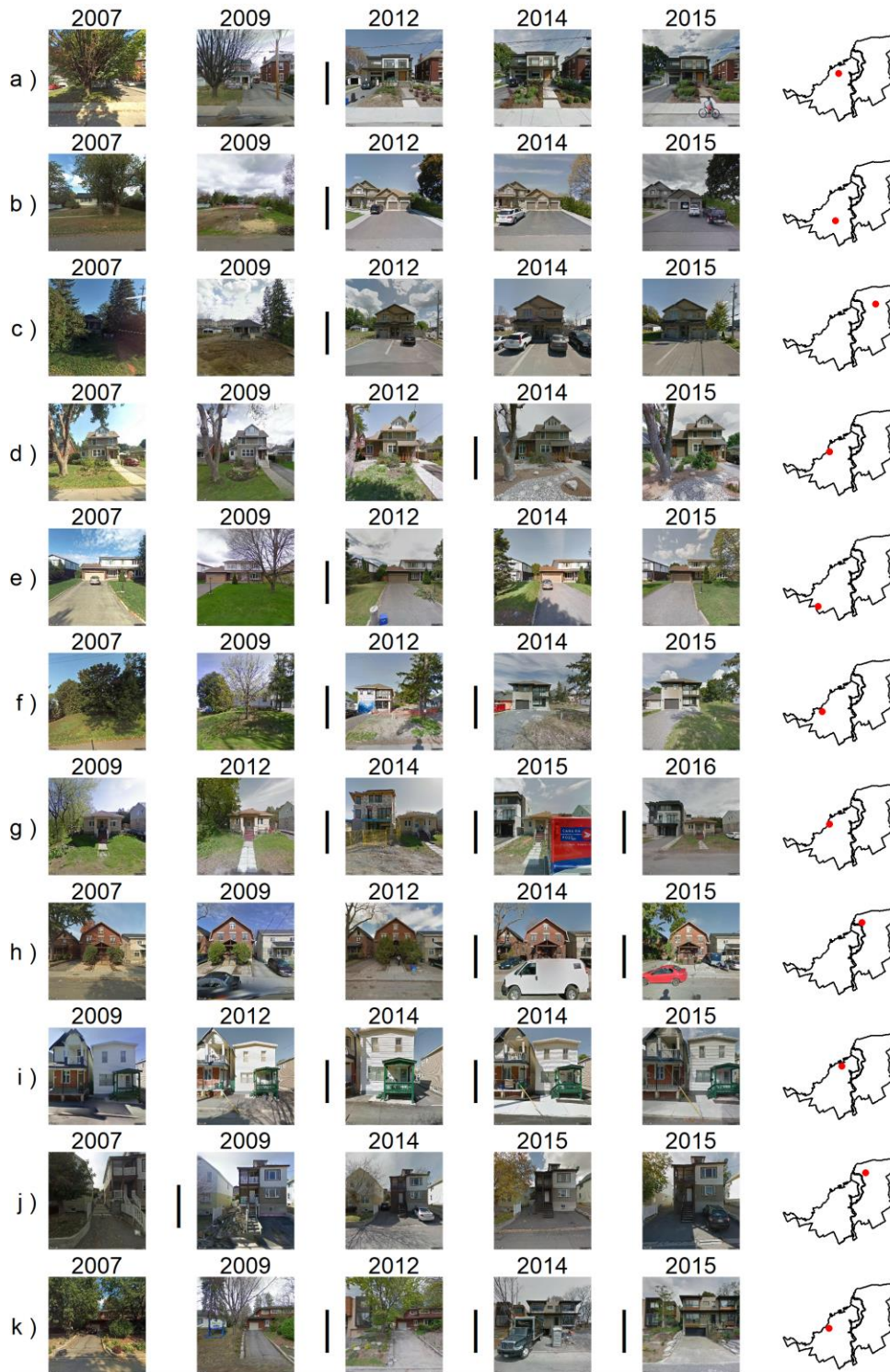
In general, our results indicate those areas that contain re-investment in the housing stock as exhibited by a progressively upward improvement of aesthetic appearance of properties that is typical of gentrification (Figure 3.7).

When SCNN-FC-8 examined a sequence of GSV images for a property, most of the time only one instance of gentrification-like visual change was found. Typical visual features found when a single instance of change was detected, include an older structure being torn-down and a new more expensive structure being created (Figure 3.7A to 3.7C) or a significant change in landscaping and/or façade (Figure 3.7D). Figure 3.7A to 3.7D also illustrate that the model is quite robust to the removal of a structure, vegetation phenology, varying sky/cloud conditions and exposure. The robustness of CNN models with respect to variations in the photographic quality and weather variations equate to less image pre-processing when compared to some other machine mapping studies like that of Naik et al. (2015, 2014, 2017) who, for example, had to control for trees and sky when using feature extraction and support vector regression in order to detect semantic qualities of GSV images.

SCNN-FC-8 detected approximately 16% of the properties as containing more than one instance of a gentrification-like visual change: 399 properties showed two instances of change, 57 showed three instances of change and 16 showed four instances of change. When capital is invested in the built environment, it is also immobilized for varying lengths of time (Weber 2002). Developers and investors aim to minimize housing production and holding costs, and from that perspective, multiple major upgrades or redevelopment to property over several years should generally be infrequent to maximize returns on investment.

In cases where multiple detections of visual change were found at the same location, there were several reasons. Foremost, most false positive detections occurred when the camera location was either offset or had a considerably different field of view (FOV) in a subsequent GSV at the same location—the coordinates that were sent to the Google API. This coordinate offset/FOV meant that a GSV image in at least one year had the camera pointing either between

two structures, whereas previously it was pointed at one structure or vice versa. The sudden appearance of a second structure entering the GSV image led to SCNN-FC-8 incorrectly identifying a gentrification-like visual change (Figure 3.7E, 3.7I and 3.7J). When such effects occurred, examination of the activation patterns of the last convolutional block tended to show strong activation on the neighboring structure (S6 in Appendix IV). These offset/FOV-induced false positive detections were pronounced in at least one neighbourhood (Vanier North) (Figure 3.5A).



**Figure 3.7. Examples of SCNN-FC-8 detections.** Each row represents the time series of GSV images for the unique location specified in the map plot at the end of each row. The vertical bar (|) symbol between any two successive years indicates that the SCNN-FC-8 detected a visual gentrification-like change between the two GSV images. (a to d) Normal cases of a single change being detected across the series of images at a given unique location. (e to k) represents cases where the model detected more than one change or detected a false positive change within the sequence for some other reason as detailed in the text.

SCNN-FC-8 was robust to smaller changes in offset/FOV (Figure 3.7A to 3.7D). A common issue with using GSV images in deep mapping is obfuscation of the subject by a large vehicle (Naik et al. 2017). If a large vehicle was obscuring a property in one image and gone in the next, the model would detect a false positive gentrification-like visual change (Figure 3.7G, 3.7H and 3.7K). Such objects tended to be strongly activated upon in the last convolutional block causing the differences in the visual appearance of a property to force a positive decision by SCNN-FC-8, even though no gentrification-like visual change was evident. In other cases, multiple-detections at the same location were not false positive detections but were due to multi-year construction periods on the property (Figure 3.7F, 3.7G and 3.7K). In some of those cases, more than one structure was within the FOV of the GSV image, the first instance of change was due to an adjacent structure undergoing modification that was then followed by a transformation to the second structure (Figure 3.7G and 3.7K). In these cases, our model would detect multiple changes to a property beginning with the major structural change and followed by subsequent updates to the new façade if construction spanned more than one year or one property. Because two instances of change occurred in 399/472 cases of multiple change that were detected by SCNN-FC-8, we retained only the most recent date at which a gentrification-like visual change occurred when producing our maps. Improvement of properties was not limited to residential properties. Our model also detected commercial gentrification, which can be observed through the improvement of store fronts and significant upgrading to streetscapes.

To understand the emergent pattern produced by the model detections we used Kernel Density Estimation (KDE). When there are spatial clusters of SCNN-FC-8 detected changes that occur within historically stagnant urban areas, these changes are more likely to indicate a gentrification process occurring locally. Because our interpretation of gentrification relies

heavily on the spatial intensity of gentrification-like visual changes, local or regional disasters could lead to renewed property reconstruction and regions of high intensities of gentrification-like visual change that are not driven by gentrification. However, in Ottawa during the period of study, there were no large-scale disasters that would have induced local clusters of property reconstruction.

Our reliance on KDE to identify hot-spots of gentrification-like visual change assumes that there is a background intensity of physical property improvements that are non-gentrification related and follow an inhomogeneous Poisson process. It would be reasonable to assume that this inhomogeneous process is random through space and time contingent on the distribution of structures in a city and secondly their age. Our current mapping process includes the unknown background intensity estimate within the KDE maps. However, because the full GSV dataset within the study area is already contingent on the spatial distribution of building stock, a spatially random process contingent on that building stock would be unlikely to produce the high-intensities we find on our KDE maps of the model detections. In other words, by focusing on intensity as the basis for inference about the gentrification process, we can ignore the background noise because spatially isolated properties exhibiting gentrification-like visual changes contribute little to the intensity surface at any location. Further research is required to understand and estimate the background process of gentrification-like visual changes within a city which, when controlled for, would lead to more accurate and focused spatial estimates of intensity.

The validation of SCNN-FC-8 showed that the spatial patterns of our model and that of permits were strikingly similar (Figure 3.5)—with a few minor variations due to false positives in the model detections. When a permit is issued, GSV images at the permit location might show

an improvement that same year. Sometimes the improvement might appear only a year or two later, due to the month in which the permit was issued and the length of time necessary for construction. This lag time would partly explain minor differences observed between the KDE maps of permits and the model detections. Other differences may be in part due to the criteria for retaining permits. It was important to differentiate suburbanization from gentrification, as these are clearly different processes. When examining permits, “new build gentrification” was retained. This type of new construction might not be picked up by our model and can be a contentious matter in terms of whether or not such developments represent gentrification at all (Davidson and Lees 2005; 2010). Finally, in some cases, a region of high intensity in SCNN-FC-8 detections could be due to false positive detections—as was the case in Vanier North (Figure 3.5) caused by offset/FOV differences between sequential GSV images along one street. In such cases, KDE shows regions of high intensity that would be greater than the background noise. Without careful examination of the KDE results, such false-positive regions could be attributed to a gentrification process.

Our KDE maps of gentrification-like visual changes agree with the recent accounts of gentrification in the City of Ottawa. The greatest intensity of gentrification-like visual changes were detected in the region of Westboro/Laurentian followed by Hintonburg/Mechanicsville, both of which are local gentrification hot-spots in Ottawa (MacMillan 2011; Turcotte 2016; Butler 2017; City of Ottawa 2010). In addition, we found high intensities in regions we did not know were undergoing gentrification like the neighbourhoods of Crestview-Meadowlands, Island Park and Ottawa East among others. While we have noted the issues in Vanier North, some of the detections were valid and agree with observations of urban densification and new build gentrification over the past decade in that neighborhood (Benali 2013). Similarly, we find

moderate intensity in Old Ottawa South, a neighbourhood that contained gentrification largely prior to our study (Leffers and Ballamingie 2013), but just across the northern boundary we found Ottawa East to be currently undergoing intense gentrification.

A distinct advantage of using a deep mapping approach based on CNNs, at least in our case, is the robustness of the model with respect to GSV images containing different weather, exposure and vegetation phenology. However, we worked in a restricted spatial domain and expect that false-positive model detections will be, for the most part, contained in the random background. Still, using a larger GSV dataset for detection could lead to areas of high intensity of gentrification-like visual changes being identified that are due to weather or exposure issues. With more expansive GSV datasets it becomes increasingly important to have examples of all possible GSV image conditions represented within the subset of data used for training.

## **Conclusion**

By taking a deep mapping approach to detecting gentrification, we have shown that it is possible to indicate precisely where and when gentrification processes are happening in an urban area. By focusing on the detection of gentrification-like visual changes to individual properties, SCNN-FC-8 results could be aggregated to any arbitrary geography and thereby specify the proportion of any arbitrary spatial unit that has gentrified. One advantage of doing so would be to test scale and zoning effects of the MAUP on results that are tied to a set of artificial boundaries (such as census tracts or neighbourhoods). For example, with SCNN-FC-8, mapped results are able show if two blocks gentrify around a boundary. This could aid in validating or decomposing the results of census-based inferences about gentrification in urban areas. While the SCNN-FC-8 model was developed and tested in the context of Ottawa, we believe that similar results can be reproduced in other urban contexts within peer countries. It remains to be

seen what or if similar results can be replicated in different contexts and the degree to which we can detect different kinds of gentrification (such as slum gentrification). Definitions and visual indicators of gentrification can differ based on cultural or architectural norms across different countries. With a relatively small but regionally consistent training dataset, other North-American locales would be able to train the SCNN-FC-8 architecture using a transfer learning approach.

Given the relative ease with which deep learning models can now be applied, thanks to high level APIs like Keras (Chollet and others 2015), a relatively modest time investment can produce highly spatially and temporally resolved maps of the gentrification process. Defining the rate and location of gentrification-like visual changes can directly benefit municipalities at a number of levels: the tax-assessor can use the information to model future property tax income scenarios; city budget and planning can use detailed maps of gentrification in order to prioritize infrastructure upgrades; the transportation office can determine where traffic-calming measures or bicycle lanes might be required in rapidly gentrifying areas; the building permit office could examine results to aid in locating non-permitted work; and, city zoning can use the information in rezoning application reviews; in the current neoliberal era, public housing is ever scarcer and the mapping of gentrification could indicate areas in need of assistance in response to displacement. An investor could speculate on where to invest within a city by examining where gentrification is happening and where the likely spill-over will occur. Finally, from a competition perspective, retail location analysis would benefit significantly by establishing new locations where gentrification is happening or ensuing.

## **Supporting information**

Appendix II. Supplementary information for ‘Deep mapping gentrification in a large Canadian city using deep learning and Google Street View’. Figure A. GSV panorama distributions used in this research. Figure B. Frequency distributions. Figure C. Spatial distribution of properties that have a panorama in 2007 and also in 2015–2016. Figure D. Spatial distribution of all accessed GSV images through time. Figure E. Training and validation of SCNN-FC-8 (Epoch 1 to min(loss)). Figure F. Comparison of KDE maps. Figure G. Comparison of A) SCNN-FC-8 results with B) the replicate model results. (PDF)

## **Acknowledgments**

This work was supported by and is a contribution to the Ottawa Neighbourhood Study ([www.neighbourhoodstudy.ca](http://www.neighbourhoodstudy.ca)).

## Chapter 4. Modeling and mapping AI derived gentrification data with census data in a large Canadian city

**Abstract:** The complexity of assessing gentrification with census data is demonstrated through the reproduction of select quantitative methodologies in the context of Ottawa, Canada, between 2006 and 2016 at the Census Tract (CA) and finer Dissemination Area (DA) units of analysis. The performances of reproduced methodologies were compared to data obtained from previous research that used computer vision via deep learning to create a point dataset of properties in Ottawa that exhibited visual signs of gentrification-like improvements between 2007 and 2016. This unique dataset provided a census-independent measure of gentrification. Herein, two new multivariable models are produced: A spatial simultaneous regression model (SAR Lag<sub>y</sub>) to predict the density of the visual signs of gentrification and second quasibinomial model to predict the presence/absence of gentrification in Ottawa's CTs. Qualitative mapping and quantitative comparisons were undertaken between all models using Lee's L measure of bivariate spatial correlation. Our new models are superior to those methods in the literature with our SAR Lag<sub>y</sub> model explaining 57% of the variance in gentrification-like visual changes in the city and the quasibinomial model achieving a 91.7% accuracy for predicting the presence of gentrification. An intermodel comparison shows that those models that examine fewer census variables tend to perform the worst when predicting gentrification in Ottawa and performance is partially a function of the spatial structure of the variables used for a particular urban area.

### Introduction

Since Gentrification was coined by Ruth Lazarus Glass in the 1960's, Slater (2011, 571) suggests that the literature has expanded significantly in response to interest in the subject itself and more poignantly because of social justice needs in identifying and analyzing urban inequalities. Gentrification is a replacement process (Schulman 2012, 14), whereby a progressively more affluent group (Hackworth 2002, 815) replaces the previous less affluent one, typically within a subspace of the urban milieu. However, gentrification is difficult to observe and measure which produces uncertainty in delineating it in geographic space (Hammel and Wyly 1996, 248).

Gentrification research has expanded drastically since the 1970s (S1 in Appendix V; Loretta Lees, Slater, and Wyly 2008a, 244), but the overwhelming majority of gentrification research is conducted in a qualitative manner with quantitative approaches being shunned (Reades, De Souza, and Hubbard 2019). The prevalence of the qualitative methodology can be observed in its overrepresentation in gentrification-oriented readers and collected volumes (Atkinson and Bridge 2005a; Loretta Lees, Slater, and Wyly 2008b; Brown-Saracino 2010; Loretta Lees, Shin, and López-Morales 2015; 2016). Of course, the goals of quantitative and qualitative research on gentrification are somewhat different, with the former being concerned with identifying the emergence and spatial delineation of the process and socioeconomic factors that can explain the process within some space, whereas the latter concentrates largely on the reasons for the process itself, often within one or two subregions of a larger metropolitan area.

Since the end of the 1970s, a slow but limited stream of research has emerged which quantifies gentrification using data obtained largely from the census (Table 4.1). However, the literature does not reveal any consensus as to how to measure gentrification using census data (Barton 2016), nor are there agreed upon census variables that should be used for such endeavors. As such, research assessing the variability between the different methods is critically needed to ascertain whether the spatial delineation of gentrification between different methods is robust.

Herein, we conduct a thorough examination of previous methods that use census data to measure gentrification quantitatively, and then proceed to investigate gentrification occurring within Census Tracts (CTs) and Dissemination Areas (DAs). We reproduce five methodologies which produce continuous interval/ratio measures of gentrification and compare these to the new results obtained from two novel regression models built to predict gentrification.

Earlier research has shown that visual signs gentrification can be observed at the sub-CT level through the use of computer vision (Ilic, Sawada, and Zarzelli 2019). Hence, in the present research, census data is used as independent variables to model gentrification in Ottawa, Canada and compare these results to the visual signs of gentrification previously identified using computer vision and deep learning (Ilic, Sawada, and Zarzelli 2019) in addition to the five quantitative methods derived from the literature.

### **Aspects of scale**

Spatially delineating gentrification as precisely and finely as possible is important for an accurate depiction of its characteristics and extent. Therefore, from a mapping perspective, the judicious choice of census geography is crucial. Almost all studies that use census data to map gentrification do so at the census tract (CT) level (Table 4.1). However, the scale of gentrification has nothing to do with the scale of geography and their areal units. Using CTs to measure gentrification results in the implicitly defining of gentrification as the change in magnitude/variance of census data.

The paucity of gentrification studies at scales finer than the CT was noted as a concern as far back as the early 1980s (Spain 1981, 16; Gale 1985, 28). While CTs are small areas, they typically contain thousands of people within their borders (zones). Scale matters: An area can appear to be heterogenous at one scale and homogenous at another (Hodge 1980, 199). Such scale and zoning effects are by definition part of the modifiable areal unit problem affecting variance due to spatial aggregation (Parenteau and Sawada 2011; Fotheringham and Wong 1991; Openshaw 1983; Tuson et al. 2019). For example, a single CT may contain several sub-neighborhoods, and if gentrification is localized, it is entirely probable that some sub-regions of a CT might be gentrifying, but not to a magnitude that is sufficient to change the gentrification

status of the containing CT (Spain 1981, 16). Such an inability to detect gentrification is particularly prevalent when using census-based metrics of gentrification or thresholds that determine gentrification/non-gentrification. As such, gentrification can be missed in geographic spaces and only noticed in subsequent censuses, if at all.

Simply put, the borders of CTs cannot support a spatially accurate and precise representation of gentrification (Lee and Mergenhausen 1984, 513) because the scale of neighborhood change may not be coincident with the spatial units of measurement. Therefore, some scholars argued that data at the CT level does not support a spatially accurate and precise representation of gentrification (Schuler, Kent, and Monroe 1992). Measuring gentrification at any area-based level, including dissemination areas (DAs), can suffer from the same drawbacks as those at the CT level, because DAs also ignore the extent of existing gentrification. In Canada, DAs are the smallest level of geography within which National Household Survey (NHS) data is reported and are fully nested within the larger CTs. However, using smaller geographic units than CTs provides the opportunity to identify gentrification with an increased spatial precision.

### ***Independent measures of gentrification***

Providing an independent dataset that delineates the spatial extent of gentrification over time, at the urban scale of a large metropolitan area, does not generally exist to test census-based models against. In previous research (Table 4.1), quantitative studies of gentrification provide no independent map of gentrification since the ‘true’ extent is approximated by changes in census variables over time. The lack of a census-independent measure of gentrification is thus challenging when one aims to assess a census model’s accuracy. We undertake a two-pronged approach to assessing the different census-based models of gentrification: First, model intercomparisons are used to determine the agreement between different models of the

gentrification process as they play out spatially, and second, we assess the different models against the independent measure of gentrification provided by Ilic, Sawada, and Zarzelli (2019). Whereas none of the existing measures of gentrification within the learned literature (Table 4.1) aim to model the visual signs of gentrification, the work of Ilic, Sawada, and Zarzelli (2019) is based on deep learning of street-level images over time in Ottawa and has no relation to census data and thus provides an independent variable that is sufficient to assess the performance of different census-based models against. The spatial configuration of the visual signs of gentrification may or may not be the ‘best’ measure of the process, but it is one that is defined and produced completely independently of census data.

## **Background**

Our literature review starts by examining the lack of quantitative methods in gentrification research. Subsequently we investigate the national and spatial contexts of where gentrification has been studied in a quantitative manner using census data. Next, the most common variables used to measure gentrification are examined, and finally we take stock of the quantitative methods used for gentrification research.

### ***Neglect of quantitative methods***

Most gentrification research is done through qualitative methods which often produce an ethnographic account of a neighborhood based on sociological interviews or an examination of media, discourse, or rhetoric from involved actors or stake holders. In contrast, there are not many quantitative gentrification mapping studies in the learned literature. One of the first<sup>1</sup> books which brought together research on gentrification was an anthology by Palen and London (1984). This book did not include anything on quantitative techniques to measure gentrification. The next gentrification book came out in 1986 (Smith and Williams 1986). Some chapters included

quantitative methods that were used to back up the existing narrative rather than being the focus of the research.

Subsequent books dealing with case studies of gentrification also failed to tackle the challenge of dealing with quantitative analysis, as they mostly focused on qualitative methods (Caulfield 1994; Smith 1996; Butler 1997; Muniz 1998; Brown-Saracino 2009; Huse 2014; Paton 2014; Tissot 2015; Krase and DeSena 2016). Books on international examples of gentrification likewise contained few quantitative approaches (Atkinson & Bridge, 2005 and Lees et al, 2015). The most comprehensive book on gentrification (Loretta Lees, Slater, and Wyly 2008a) also focuses on qualitative examples of gentrification, but a book of collected works by the same authors (Loretta Lees, Slater, and Wyly 2008b) includes a small-scale quantitative study (by Hammel and Wyly) from the mid-1990s (one chapter out of the forty). Finally, another book of collected works on gentrification (Brown-Saracino 2010) contains thirty chapters but completely lacks quantitative approaches.

The quantitative void in the approaches to studying gentrification exists largely because of the decline of positivist methodologies and quantitative approaches by many scholars of human geography (Dorling 2005, 254–55; Wyly 2011).

### ***Geographic units and geographic contexts***

A thorough literature review was conducted to compile a list of thirty-two studies that use census data to analyze gentrification (Table 4.1). Research comes from books that deal with gentrification, as well as from journal articles.

**Table 4.1. Gentrification studies which used census data and quantitative approaches.** Dates with an asterisk indicate multiple studies by the author(s) with same methodology.

	Year	USA	Canada	UK	China	Sub-CT	CT	District/Ward	Large areas	Zip-Code
Lipton	1977	x					x			
Spain	1981	x							x	
Ley	1985*		x				x			
Marcuse	1986	x					x			
Schaffer, Smith	1986	x					x	x		
Nelson	1988	x					x			
Fillion	1991		x				x			
Schuler, Kent, Monroe	1992	x				x				
Wyly, Hammel	1999*	x					x			
Atkinson	2000			x				x		
Bostic, Martin	2003	x					x			
Freeman	2005	x					x			
Meligran, Skaburskis	2005		x				x			
Heidkamp, Lucas	2006	x				x				
Walks, Maaranen	2008		x				x			
He	2010				x			x		
McKinnish, Walsh, White	2010	x					x			
Ellen, O'Reagan	2011	x					x			
Grube-Cavers, Patterson	2014		x				x			
Ding, Hwang, Divringi	2016	x					x			
Flanagan, Lachapelle, El-Geneidy	2016	x					x			
Gibbons, Barton	2016	x					x			
Lim et al	2017	x						x		
Steinmetz-Wood et al	2017		x				x			
Glaeser et al	2018	x								x
Smith, Phillips, Kinton	2018			x		x				
Loukaitou-Sideris, Gonzalez, Ong	2019	x					x			
Reades, De Sousa, Hubbard	2019			x			x			
Rigolon, Stewart, Gobster	2020	x					x			
Yonto, Thill	2020*	x					x			
Walks, Hawes, Simone	2021		x				x			
Rucks-Ahidiana	2021	x					x		x	
		21	7	3	1	3	23	4	2	1
		65.63%	21.88%	9.38%	3.13%	9.38%	71.88%	12.50%	6.06%	3.13%

The overwhelming majority of studies, in Table 4.1, take place in the US and Canada, perhaps because this is where quantitative studies on gentrification emerged and where census data is easily accessible. Almost all these studies occur at the urban level. Gentrification is not strictly an urban phenomenon, as some have examined this process in the rural context (Phillips 1993; Smith, Phillips, and Kinton 2018).

The spatial unit of analysis in gentrification research is overwhelmingly the census tract (CT) (Table 4.1), which is not surprising, considering that the CT is the most common unit of

analysis in urban geography (Ilic and Sawada 2021, S3). Some studies examine gentrification within smaller areas such as a single neighborhood, the inner city, or other locally constrained geographic areas. Others examine gentrification at multiple geographic units (Schaffer and Smith 1986). Few studies examine gentrification at large scales (Spain 1981; Walks, Hawes, and Simone 2021), and only one study chose zip codes as the geographic unit of analysis (Glaeser, Kim, and Luca 2018) (Table 4.1).

Some researchers (both quantitative and qualitative) examine gentrification only after the fact (Prouse et al. 2015; Tierney and Petty 2015), or in other words after gentrification was hypothesized or known to have occurred in an area (Holm and Schulz 2018, 253). Such neighborhood-based approaches have been called for (Rose 1996, 161), but have the weakness of being less robust, as their findings are limited to a narrow spatial scale and have little to no predictive power at best.

### *Census variables*

There is no consensus on a common set of census variables to use in quantifying gentrification (Table 4.2). Past debates even highlight a strong disagreement as to which type of data is appropriate. For example, in the late 1980s, Ley and Smith had exchanges in which they disagreed about the proper way to assess gentrification: through change in household social status or the revalorization of the housing stock (Smith 1987; Ley 1987).

Such debates are scarce, and researchers instead sidestep justifying the selection of census variables for measuring gentrification (McKinnish, Walsh, and Kirk White 2010; Ellen and O'Regan 2011; Grube-Cavers and Patterson 2015; Steinmetz-Wood et al. 2017; Rucks-Ahidiana 2021). One variable alone may not be sufficient to measure gentrification (Bostic and Martin 2003, 2431), but many times the variables chosen are simply implicitly assumed to be

adequate, as the researcher in question would surmise that past studies were without fault. In other words, there is a lack of criticality in justifying variable selection using a priori principles.

**Table 4.2. Census variables utilized to measure gentrification.** Dates with an asterisk indicate multiple studies by the author(s) with same methodology. The variables represent the final selection of variables used by authors to measure gentrification in their models and results. Some of these researchers initially started with more census variables. The methodologies of studies in grey were reproduced in this study.

Author(s)	Year	Education	Income	Occupation	Rent	Race	Home value	Home ownership	Age	Pop. change	Dwellings rented	Poverty	Housing Age	Dwelling density	Tot.Var
Lipton	1977	x	x												2
Spain	1981					x									1
Ley	1985*	x		x											2
Marcuse	1986	x													1
Schaffer, Smith	1986					x									1
Nelson	1988		x												1
Fillion	1991		x												1
Schuler, Kent, Monroe	1992	x				x			x						3
Wyly, Hammel	1999*	x	x	x						x					4
Atkinson	2000			x											1
Bostic, Martin	2003	x	x	x		x		x	x						6
Freeman	2005	x					x						x		3
Meligran, Skaburskis	2005		x		x									x	3
Heidkamp, Lucas	2006		x	x				x				x			4
Walks, Maaranen	2008	x	x	x	x						x				5
He	2010			x											1
McKinnish, Walsh, White	2010		x												1
Ellen, O'Reagan	2011		x												1
Grube-Cavers, Patterson	2015	x	x	x	x			x							5
Ding, Hwang, Divringi	2016	x			x		x								3
Flanagan, Lachapelle, El-Geneidy	2016	x	x			x	x	x							5
Gibbons, Barton	2016	x	x		x										3
Lim et al	2017	x	x		x										3
Steinmetz-Wood et al	2017	x	x		x				x						4
Glaeser et al	2018	x				x			x						3
Smith, Phillips, Kinton	2018	x													1
Loukaitou-Sideris, Gonzalez, Ong	2019	x	x		x	x									4
Reades, De Sousa, Hubbard	2019	x	x	x			x								4
Rigolon, Stewart, Gobster	2020	x		x											2
Yonto, et al	2020*	x	x	x											3
Walks, Hawes, Simone	2021		x		x		x								3
Rucks-Ahidiana	2021	x	x		x		x								4
		21	20	11	10	7	6	4	4	1	1	1	1	1	

The most common variables used in gentrification studies are at the ecologic level and represent variables that are either derived from individuals within the census unit, such as education, income, occupation, race, age, and home ownership or variables related to market forces such as rent and home value.

Out of twenty-eight studies that used either income and/or education, eight used only education, seven used only income and thirteen used both. Only four studies did not use either income or education. Eleven of the thirty-two studies identified used occupation and ten used rent. These individual and market related variables are easily justified with relation to class distinction and class change within the process of gentrification, for example:

### *Education*

A population which is more affluent tends to have higher levels of education. Post-secondary degrees/diplomas would in theory lead to employment with higher remuneration and therefore a rising level of education in a geographic unit is considered as a means of gauging the extent of change when an area is experiencing an upward movement of social status.

Furthermore, groups with more education (and consequent income), in general, are more likely to enhance the aesthetics of their properties and contribute to visually changing their neighborhoods. This outer expression of affluence has formed the basis for examining gentrification at the property level using manual means and machine learning (Hammel and Wyly 1996; Hwang and Sampson 2014; Ilic, Sawada, and Zarzelli 2019). Of note is that groups such as artists (often considered to be early gentrifiers) did not necessarily have high incomes.

Why then do some researchers prefer education over income, for measuring gentrification? The answer stems from work in the 1980s by Ley (1985), who was adamantly opposed to income as an indicator of gentrification. This was rationalized by research which suggested that early gentrifiers were often relatively low-income groups such as but not limited to students and artists. Some (Freeman 2005, 471) use vague language such as “perhaps education is a better marker of class than income”, with suggestions of income fluctuating through time whereas educational attainment does not.

### *Income*

As gentrification is an upward class transformation of space, it makes sense that income is an important variable. It has been noted that gentrifiers have a higher income than the native or in-situ population they are displacing (Spain 1980, 28), and this in part stems from how gentrification is often defined, whereby a higher class replaces a lower class. Baldassare (1984,

92) suggests that rising incomes are the most dependable indicator of urban revitalization (ie gentrification).

A strong rationale for using income is that gentrification represents an upward class turnover of population, and that income has been shown to be the dominant variable in measuring socio-economic status (Davies 1984; Townshend and Walker 2002).

A higher class has higher income than the class below. Even some, who focus on other variables, note that income is also reasonably reliable and plausible as an indication of gentrification (Marcuse 1986, 176), and hence it is not surprising that it is one of the most popular variables for measuring gentrification (Table 4.2).

### *Occupation*

Occupational outcomes are often contingent on levels of education and previous work experience. Occupations have varying incomes, and it is assumed that gentrifiers are made up more so of a demographic which has white-collar jobs. In contrast, the blue-collar manual labor work force is more likely to be displaced by gentrification. Hence, measures of occupation are often seen in gentrification studies. In the US and Canadian census, many categories for types of employment exist, and these change over time. For example, Schaffer and Smith (1986, 350) note that between 1970 and 1980 occupation definitions changed to such an extent that they were no longer comparable.

### *Race*

While race and/or ethnicity remains a pertinent variable in the gentrification literature, it is more applicable to the US context, given the historical and temporal specificities of high levels of segregation in the US. Major US cities tend to have high levels of racial segregation. Even

though the civil rights movement occurred decades ago, race remains inseparable from many urban processes in the US.

Marcuse (1986, 176) has noted that, as with income, race and/or ethnicity is a reasonable and theoretically plausible measure for measuring gentrification. At various moments race has been less popular for measuring gentrification, perhaps in part because class and gender have predominated in both qualitative and quantitative approaches used in gentrification studies (Lees 2000, 400). In general, race has emerged as an important factor of gentrification, especially as it pertains to the displacement of Blacks and Latinos in neighborhoods cities of the US (Betancur 2002).

#### *Age*

Early thinking on gentrification focused on explaining the process through demographic change. This line of thought is sometimes referred to as the “demographic hypothesis”, and it examines variables including changing age structures, family size, female labor force participation (Ley 1986, 527). Examining such variables, Ley found that the concentration of the age group between 20 and 35 is significantly associated with gentrification. While this literature may be dated, more recent research suggests that a decline in the proportion of children and elderly is an indicator of gentrification (Brown-Saracino 2009, 22).

However, many different types of gentrification exist. For example, under family gentrification (Marcuse 1986, 164, 176; Karsten 2003; Goodsell 2013; Karsten 2014; Cain 2020) it is families who come and gentrify, and therefore the numbers of youth also increases. Families could also take the form of dink (dual/double income, no kids) status. Therefore, it remains unclear if changes in age categories alone are a good indicator for gentrification in all contexts.

### *Rent and house value*

Characteristics of the built environment, such as house value or average monthly rent, are captured by the census and are relevant to gentrification. Impacts of gentrification include an increase of property values and rent (Atkinson and Bridge 2005b, 5). If an area is appreciating, it is logical to assume that rent and home values would increase. Likewise, if these two variables are decreasing, it may be a fair assumption that an area is declining.

When gentrification takes place, an upgrade to the housing is an almost universal consequence. Such modifications have the effect of raising both property values and rents.

### ***Quantitative Methods***

Whereas studies focused primarily on describing gentrification until the 1970s; by the 1980s, thanks in part to newly released census data (Nelson 1988, 17), limited gentrification research moved towards the use of quantitative methodology. Table 4.3 shows the different approaches to examining gentrification. About two thirds of the studies mapped gentrification, whereas the others simply examined gentrification in a quantitative manner. Usually it is geographers, more so than researchers from other fields, who are interested in the spatial aspects (ie mapping) of gentrification.

**Table 4.3. Approaches to measuring gentrification with census data.** Dates with an asterisk indicate multiple studies by the author(s) with same or similar methodology.

Author(s)	Year	Approach			Methods							Mapping	
		Continuous	Threshold	Other	Logical	% Change	PCA	Index	Net Change	Stepwise	Other	Yes	No
Lipton	1977		x								x	x	
Spain	1981	x				x							x
Ley	1985*	x					x					x	
Marcuse	1986		x			x						x	
Schaffer, Smith	1986		x			x						x	
Nelson	1988		x			x						x	
Filion	1991	x										x	
Schuler, Kent, Monroe	1992		x								x	x	
Wyly, Hammel	1999*		x						x			x	
Atkinson	2000			x				x					x
Bostic, Martin	2003		x		x								x
Freeman	2005		x		x								x
Meligran, Skaburskis	2005		x				x					x	
Heidkamp, Lucas	2006		x						x			x	
Walks, Maaranen	2008			x			x					x	
He	2010	x									x	x	
McKinnish, Walsh, White	2010		x					x					x
Ellen, O'Reagan	2011		x			x							x
Grube-Cavers, Patterson	2015		x		x							x	
Ding, Hwang, Divringi	2016		x		x							x	
Flanagan, Lachapelle, El-Geneidy	2016	x									x	x	
Gibbons, Barton	2016		x		x								x
Lim et al	2017		x				x					x	
Steinmetz-Wood et al	2017		x		x								x
Glaeser et al	2018			x							x		x
Smith, Phillips, Kinton	2018		x								x	x	
Loukaitou-Sideris, Gonzalez, Ong	2019		x		x						x	x	
Reades, De Sousa, Hubbard	2019	x									x	x	
Rigolon, Stewart, Gobster	2020		x		x							x	
Yonto, et al	2020*	x					x					x	
Walks, Hawes, Simone	2021			x							x		x
Rucks-Ahidiana	2021		x		x								x
		7	21	4	9	5	3	2	2	2	9	21	11
		21.88%	65.63%	12.50%	28.13%	15.63%	9.38%	6.25%	6.25%	6.25%	28.13%	65.63%	34.38%

Most quantitative gentrification research utilizes thresholds to indicate when a geographic unit of analysis is gentrified (Table 4.3). Hence, studies would have results that show CTs grouped into categories such as gentrified, partially gentrified, not gentrified, and so forth (Nelson 1988; Wyly and Hammel 1999; Bostic and Martin 2003; Freeman 2005; Heidkamp and Lucas 2006; Meligrana and Skaburskis 2005; Walks and Maaranen 2008; McKinnish, Walsh, and White 2010; Grube-Cavers and Patterson 2015; Ding, Hwang, and Divringi 2016; Lim et al. 2017; Steinmetz-Wood et al. 2017; Loukaitou-Sideris, Gonzalez, and Ong 2019).

However, thresholds are arbitrary and unreliable because any figures right below the threshold are discarded, whereas anything right above the threshold is included. In reality, boundaries are fuzzy, and membership is a matter of degrees (likelihood) and not kind.

Table 4.3 also shows some of the methods used to measure gentrification. The most popular method is a logical model. These models usually require that a number of census attributes exhibit a particular change. When these align, then gentrification is labeled as having occurred, as the threshold has been reached. But such threshold approaches are limited by the fact that gentrification is a process which is not necessarily ever complete, as neighbourhoods are rarely homogenous (Curran and Hamilton 2012, 1034). Additionally, as space is “always under construction” (Massey 2005, 9), it is entirely possible for a neighbourhood to undergo multiple rounds of successive gentrification such that groups which were once displacers become the displaced. Therefore, there is reason to assume that an approach using continuous data is desirable in order to assess the extent of gentrification rather than only if it has occurred.

## **Methodology**

This study uses census data from 2006 and 2016 to assess gentrification, in Ottawa Canada. First, five methodologies that examine gentrification at census tracts (CTs) in an interval/ratio manner are reproduced to assess gentrification at the CT level. This is followed by using four of these methodologies at a finer level of analysis: the dissemination area (DA). Subsequently, the difference between using CTs and DAs is examined by assessing how each geographic unit’s results differ among the models and by comparing the spatial autocorrelation of the results that the respective methodologies create. Finally, regression analysis is conducted, and the results of new models are compared to those of the reproduced methodologies to compare their predictive power.

## ***Study Area and Temporal Dimension***

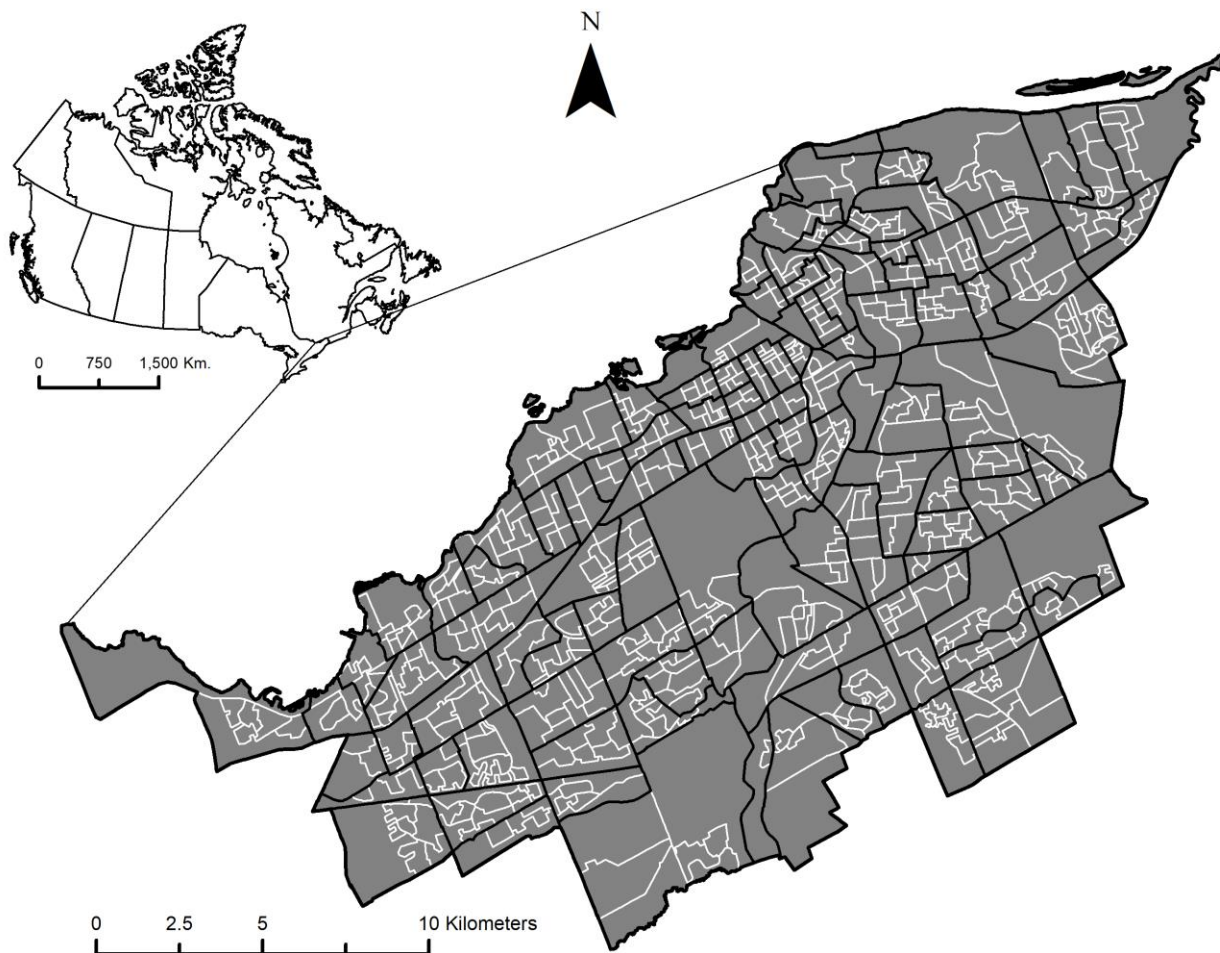
The Ottawa-Gatineau Census Metropolitan Area has approximately 1.4 million people in both Ontario and Quebec. Within this CMA, the largest jurisdiction in both territory and

population is the City of Ottawa, which has surpassed a million denizens in 2019 (City of Ottawa 2019). Ottawa got amalgamated with many of its suburbs at the turn of the millennium (Rosenfeld and Reese 2003). At 2790 km<sup>2</sup>, Ottawa has by far the largest area of any large municipality in the US and Canada (S2 in Appendix V). As much of this territory is comprised of rural communities and agricultural land (such as the greenbelt), our study area is constrained.

In earlier research we chose to examine gentrification in the part of Ottawa which is bounded by greenbelt, because this area contained the older building stock in the city and hence is most likely to experience gentrification (Ilic, Sawada, and Zarzelli 2019). The temporal dimension in earlier research covered the time period between 2007 and 2016. This time frame is close to the two census years which serve as our census data end points: 2006 and 2016.

The geographic units of analysis are both CTs and DAs (Figure 4.1). In Canada, DAs are the smallest level of geography within which National Household Survey (NHS) data is reported and are nested within the larger CTs. This study area contained 769 DAs in 2016, which are located within 112 CTs. The average population of DAs and CTs in our study area rounds to 624 and 4283 respectively.

In the literature, there is an absence of DAs being used as the spatial unit for measuring gentrification. The reason for this is multifaceted. Firstly, DAs have existed for a short time: since 2001. Secondly, data is often repressed at the DA level due to low populations. Thirdly, the boundary changes of DAs over successive censuses induces an added level of complexity when trying to compare data across censuses. In other words, the zonal configuration of DA geography is more different from census to census when compared to that of CTs, which results in the direct comparisons of gentrification results problematic from a quantitative perspective.



**Figure 4.1. Structure of Ottawa's Dissemination Areas within Census Tracts (2016) for areas which are bounded by the Greenbelt and Airport.**

### *Reproducing Methodologies*

Five methods in the literature are identified which use census data to map gentrification in an interval/ratio manner in lieu of using a threshold method.

#### *a. Ley*

Ley (1985) produced an index to examine gentrification between 1971 and 1981 in major Canadian CMAs. In subsequent publications, additional census dates were also examined (Ley 1988; 1992; 1996a; 1996b). The variables used were measures of education and occupation: the percentage changes in levels of individuals with university degrees, and the percentage of the work force in highly skilled employment categories (Ley 1992, 232). Ley's index was the average

of these two variables, and a derived variable called social status change, which he called gentrification, was the difference of the index between two census years. Ley produced an index to examine gentrification between 1971 and 1981 in major Canadian CMAs. In subsequent publications, additional census dates were also examined (Ley 1988; 1992; 1996b; 1996a). The variables used were measures of education and occupation: the percentage changes in levels of individuals with university degrees, and the percentage of the work force in highly skilled employment categories (Ley 1992, 232). Ley's index was the average of these two variables, and a derived variable called social status change, which he called gentrification, was the difference of the index between two census years.

*b. Fillion*

Fillion (1991) examined gentrification between 1971 and 1986, in Toronto, by using the relative change quotient of median pre-tax household income.

*c. He*

He (2010) examined gentrification in Shanghai. He's variable for examining gentrification was employment categories. Unlike Ley, He did not include groups such as artists into the analysis, as researchers had started to reject this group as being potential gentrifier populations (Walks and Maaranen 2008b, 10). He examined the change in the Location Quotient of the potential gentrifier population to map gentrification.

*d. Flanagan et al.*

In an examination of cycling infrastructure and gentrification in Portland and Chicago, Flanagan, Lachapelle and El-Geneidy (2016) used the sums of Z scores of several variables to create their gentrification index. The variables examined were the percentage changes: in population density, percentage of white population, percentage of homeownership, percent of people with higher education, median home value, and median household income.

*e. Yonto et al.*

In their study of gentrification in Charlotte, Yonto and his colleagues used a similar methodology to that of Ley (Yonto and Schuch 2020; Yonto and Thill 2020). Their index differs in that it uses income in addition to education and occupation data. The type of income data used was the logarithm of median household income. The output of this modified version of Ley's model was standardized using z-scores prior to being combined to form the index.

The last two studies use median household income. Median figures are typical for data from the US. In our study, average household income is used, because certain DA level data required unsuppressing, which is not possible on median figure data. We address the data suppression issue by unsuppressing DAs using CT data. In our study area, almost all cases of DA suppression are such that one DA within a CT requires unsuppressing. Hence, by using averages and counts one can get a close estimate of the suppressed value. Datasets for owners and renters were not possible to unsuppress due to levels of suppression being too high within some CTs. Therefore, when several DAs in a CT lack data, it is not possible to accurately estimate the suppressed values. The lack of this owner/renter variable did not allow for the reproduction of Flanagan et al's (2016) methodology at the DA level. Among other variables, the income variables had higher rates of suppression, with about twenty DAs requiring unsuppression.

The change in DA boundaries between successive censuses is addressed through Dasymetric Mapping, which is a method that until now has not been used in gentrification studies. Dasymetric Mapping is a thematic mapping technique which refines attributes in a map using ancillary geographic data on which parts of a geographic area people do not live in such as parks, commercial buildings and so forth (Dent, Torguson, and Hodler 2009, 117; Mennis 2009, 728). Dasymetric mapping is often used with population data. Basic area interpolation with a binary dasymetric mask focuses the spatial variables before re-allocation to new DA units. In

effect, all census data for all years is transferred to a single DA zonal configuration prior to analysis and mapping.

A crucial step in binary Dasymetric Mapping is the creation of a binary mask which delineates areas that are relevant to the target variable that is going to be transferred from one geographic zonation to another. As gentrification is a process affecting people, the first component of our dasymetric mask was residential land-use zoning, which was obtained from the City of Ottawa. By pruning out only zoning types that contain possible residential areas, a representation of spatial population distribution as the ancillary data for Dasymetric Mapping was achieved. After filtering zoning to only include residential zoning types (S3 in Appendix V), property-parcel data was used to remove streets from the land-use zoning file, as people do not reside there. The removal of streets also helped in the later reaggregation step, as DA borders are often digitizing along streets. Hence, the delineation of residential areas allowed for one zonal configuration of DAs across all years to map gentrification. This was achieved by using a basic area interpolation of the binary mask intersected with the census data in order to reallocate all necessary census variables to a common DA geography.

Examining gentrification at a level of geography that is finer than a CT is useful because gentrification is a localized phenomenon and/or begins as a localized phenomenon. Dasymetric Mapping allows for a finer identification of spatial variables at locations of interest. Hence, the use of Dasymetric Mapping fills a void in the quantitative gentrification literature by offering a technique for gentrification researchers to use at smaller levels of geography than the CT.

### ***Comparing CT and DA results of reproduced methodologies***

After reproducing the aforementioned methodologies, results on gentrification were analyzed: via mapping, variability of DAs, and spatial autocorrelation.

### *Mapping*

Through mapping, two sets of choropleth maps were produced. The first set used data from the reproduced methodologies at both CT and DA geographic units. Because the output of the different methods naturally differ in absolute values, each map's data was sorted by quintile classification to allow for the distribution of data to be standardized throughout the maps so that the maps can be compared.

The second set of maps are called frequency maps. They offer insight into how frequently CTs and DAs appear in each quintile across the reproduced methodologies.

### *DA Variability*

The first part to examining DA variability is to assess each CT's data range, which was calculated by subtracting the minimum DA value from the maximum DA value. The results were mapped in deciles in order to show more patterns than quintiles.

Next, a set of maps which show only the top quintile of each methodology's results at the CT level were mapped. It is assumed that if gentrification is present that it would be in the highest quintile. However, to see how results at the DA level compare to those of the CT level, one more set of maps was created that show the portion of DAs in the top quintile within each CT. A final set of maps was made which shows the difference between the aforementioned two maps. The final set can be termed discrepancy maps. In effect, this last set of maps is an examination of how often CTs appear in the top quintile in relation to how often DAs appear in the top quintile.

Data for discrepancy maps can in theory range from -1 to 1. CTs in the top quintile were assigned a value of 1 and the remainder with a 0. DAs within CTs were given a value indicative to the portion of DAs within the respective CT. Hence, these values can also range from 0 to 1.

The resulting dataset contains CTs with values in three categories : from -1 to 0, 0, and 0 to 1. Positive values would indicate that the CT unit of analysis identifies top quintiles more than DAs, whereas negative values indicate that DAs identify the top DAs more. The maps produced had data breaks at one-third intervals above and below zero.

Quantitative comparisons between methods were also completed as described below (Independent variables).

### *Spatial Autocorrelation*

Spatial autocorrelation was used to complement the analysis of DA variability. Spatial autocorrelation can help in examining processes that lead to self-similarity of neighboring values, as it examines how similar the values of the reproduced methods at the CT/DA level are to surrounding or adjacent CTs/DAs. The spatial autocorrelation of each of the five reproduced study outputs and the independent variables were measured using Moran's I (Cliff and Ord 1981; Bivand and Wong 2018) with pseudo p-values estimated using 9,999 permutations of the mapped outputs. Spatial autocorrelation was measured for both the DA and CT level results as well as new model outputs and all independent variables.

### *Model building*

Regression analysis, a traditional form of supervised machine learning, was used to develop two novel models using GSV-point density as the dependent variable. First, ordinary least squares (OLS) regression was used to specify the optimal set of variables using stepwise regression. Next, because model residuals contained spatial autocorrelation, a spatially autoregressive model was used as the final model to predict GSV-points per unit residential area. Subsequently, another approach using a quasibinomial regression method was used to predict the probability of gentrification after dichotomizing the dependent variable. Finally, all models were

compared, including the five reproduced methodologies using bivariate spatial correlation techniques.

### *Dependent variable*

In previous research, we spatially delineated the visual signs of gentrification across all of the Ottawa metropolitan area by using an artificial intelligence (AI) based deep learning model applied to Google Street View (GSV) images between 2007 and 2016 (Ilic, Sawada, and Zarzelli 2019). Our computer vision model created a dataset of point observations where each point was associated with an address (geocode) and was labeled as either gentrified or non-gentrified, so called GSV-points hereafter. We then mapped the GSV-points and used a kernel density estimation (KDE) to remove noise and show where the visible signs of gentrification at the property level were high/low through the time period. While no single GSV-point in that research means that gentrification is taking place in a neighborhood per se, a concentration of GSV-points indicates gentrification. We assume that a higher concentration of such points means more gentrification as highlighted by higher densities on the KDE surface.

There are minor CT boundary changes between 2016 and 2006 as provided by Statistics Canada. The net effect of these slight boundary changes on the allocation of GSV-points is such that 43.1% of the 109 CTs have different counts of GSV-points contained within them in one year compared with the other. While that percentage seems large, the total number of points that could be allocated to different CTs due to boundary changes was only 211, representing less than 3.1% of the 6819 GSV-points from the combined primary and replicate experiment of Ilic, Sawada, and Zarzelli (2019). Thus, to account for this potential misallocation, the average number of GSV-points per CT was used based on counting the number of GSV-points within the 2006 and then the 2016 CTs. The reason for the different allocations during this time-period are

largely due to GSV-points that are nearly collinear with CT boundaries, which as coordinates derived from street-level imagery, in general are along major streets that coincide with census boundaries.

The area of a census unit has no a priori relation to the area of properties (the gentrification unit) found within a census unit, such as a CT. There are many places within census units that are “not gentrifiable”, areas such as parks, water bodies, hospital grounds, school grounds, cemeteries, and so forth. These areas vary among census units. We thus tally residential zoning area as the denominator to create the dependent variable of GSV-points per unit area of residential property zoning (see also *Reproducing Methodologies*). More specifically, because of CT boundary changes between 2006 and 2016, the residential zoning area was calculated for both years’ boundaries and then averaged as the denominator for the GSV-point density measure which comprises the dependent variable in this research. Hereafter, we refer to this either as the ‘dependent variable’ or the ‘GSV-point density’ to distinguish that from GSV-points which are the raw model output from Ilic, Sawada, and Zarzelli (2019) models.

Prior to averaging the GSV-point counts, three CTs contained no GSV-points from the first model run of Ilic, Sawada, and Zarzelli (2019), namely census tracts 2.04, 14.00, & 137.02. Ilic, Sawada, and Zarzelli (2019) provided a replicate experiment that produced GSV-points from a second independently trained deep mapping model. The GSV-points from that replicate experiment were used to fill in the missing data within the three aforementioned CTs that had no points within them. In the replicate GSV-points, only one CT (11.03) contained a count of zero that was likewise filled with the count of two GSV-points from the main experiment of Ilic, Sawada, and Zarzelli (2019).

The CTs for Parliament Hill (47.00), was removed from analysis due to low populations and 137.03 and 50.00 were removed due to unreported data in 2016 for some of the independent variables.

### *Independent variables*

The literature identifies many variables (Table 4.2) that can be used to measure gentrification using census data. A selection of these variables is one of the starting points to model gentrification. As there is a lack of consensus as to which variables should be selected, a holistic approach was taken by which a broad range of variables was examined. The variables from other researchers (Table 4.2) and additional ones were used.

For example, three age categories are identified: young adults (ages 20 to 35), elderly (ages 65+) and youth (younger than 20). These categories reflect what researchers state is associated with gentrification (Ley 1986; Brown-Saracino 2009). The unemployment variable is not typically used to assess gentrification, but it is mentioned by Flanagan et al. (2016) as associated with the process. They did not use this variable directly in modeling because the US census did not collect such data in 1990.

A total of thirteen variables were chosen for examining gentrification (see Table 4.5). A detailed breakdown of census variable names is provided in the supplementary material (S4 in Appendix V). Their percentage change between 2006 and 2016 was examined. Some variables had to be derived. For example, dwelling density required the dwelling count to be divided by the area. Others used US definitions. For example, racial categories in gentrification studies in the US typically include Blacks and/or Hispanics. In Canada, these groups were extracted from the list of visible minorities.

Among the independent variables there are numerous measures of income and education. After testing the dependent variable against twelve measures of income and four measures of education (S5 in Appendix V) the strongest, according to Pearson's R, for each was selected, namely, average pre-tax household income and post-secondary education (trades, college, and university degrees/diplomas).

### *Model Building*

Ordinary least squares (OLS) linear regression assumes the dependent variable is normally distributed, as are the residuals in addition to the latter being homoscedastic and lacking spatial autocorrelation. The GSV-points per unit zoning area were not normally distributed. Given that gentrification over a decade time period at the CT level is not going to be frequent across the city, the GSV-point density was low-value biased and right-skewed or heavy-tailed. Because of this, a Lambert W transformation was chosen to bring the dependent variable to normality using the R package Lambert W (Goerg 2011; 2015; 2022) because it performed best (largest p-value) using the Shapiro-Wilk normality (SWN) test (Shapiro and Wilk 1965) among the other common transformations for right-skewed data (logarithmic, square-root and arcsin) (S6 in Appendix V).

Two statistical models were built, the first being a multivariable regression model to predict the continuous density of GSV-points and the second being a quasibinomial model to predict the probability of a CT having been gentrified over the study period. To model the probability required that we transform the dependent variable into a dichotomous outcome of gentrified vs non-gentrified, despite the theoretical weakness of this approach to identifying gentrification. We used a quasibinomial general linear model after transforming the dependent variable to a binary variable that was defined by those CTs within the top 10% of the Lambert W

transformed GSV-point density values. For the multivariable regression model, linear multivariable OLS regression was utilized but spatial autocorrelation was present in the model residuals and subsequently, we used a multivariable spatially lagged simultaneous autoregressive model estimated by maximum likelihood (SAR Lag\_y). The independent variables remained the same for both models, and the independent variables were all percentage changes between census variables (2006-2016). The independent variables were standardized for the quasibinomial model to achieve better interpretability of the coefficients and odds ratios.

For the multivariable model, we first assessed the variables from Table 4.2 that were significantly linearly related to the dependent variable using Pearson's R. Next, using only those variables that had a significant linear relation to the dependent variable, we used a backwards stepwise regression for model specification and because of the nested structure we selected the combination of variables that provided the lowest value for the Akaike information criterion (AIC). Subsequently, we assessed the variance inflation factors (VIF), normality, heteroskedasticity and spatial autocorrelation of the multivariable OLS model residuals. To test the VIF of models we used the `vif()` function of the R package `car` 3.1-0 (Fox and Monette 1992; Fox 2016; Fox and Weisberg 2019). To test for heteroskedasticity we used Koenker's studentized Breusch-Pagan test (Breusch and Pagan 1979; Koenker 1981; Krämer and Sonnberger 1986). To test for normality, we used the Shapiro-Wilks test (Shapiro and Wilk 1965) and for spatial autocorrelation we used the Moran's I test for residual autocorrelation (Cliff and Ord 1981). To undertake a spatial model specification search, we used Lagrange multiplier diagnostics for spatial dependence in linear models (Anselin 1988; Anselin et al. 1996) followed by the common factor hypothesis using likelihood ratio tests based on a full spatial

Durbin model (SDM) against the more parsimonious nested model alternatives (spatial lag vs. spatial error model) (LeSage and Pace 2009).

Since the specification search for a spatial linear model with spatial dependency is affected by the nature of the spatial weights matrix that is used, the sensitivity of the spatial model specification search was assessed using two row-standardized weights matrices: A) a queen's case contiguity, whereby for a given CT, any surrounding CT that shares a boundary or single point along the boundary is considered adjacent, and; B) a k-nearest neighbours definition whereby for a given CT's centroid the five ( $k = 5$ ) closest CT centroids, as the crow flies, are considered neighbours. We settled on queen's case contiguity for all spatial analyses given that the SAR model specification did not change under  $k=5$  nearest neighbours when compared to Queen's case contiguity and the linkages between CTs is more realistic of CT interactions regarding gentrification, that is, the level of visual gentrification as measured by the GSV-point density in one CT should be most similar to adjacent CTs densities, more so than those CTs that have no direct contiguity to the CT in question.

For the quasibinomial model, we divided the dependent variable into deciles. By dividing the dependent variable into deciles, we choose the topmost decile as the 'gentrified' one, thereby approximating a threshold-based method for determining which CTs underwent gentrification between 2006 and 2016. For example, the top decile represents the highest 10% of GSV-point density values, the second highest represents the highest 20%, and so forth. We then coded the dependent variable as a binary variable with the value one (1) if the CT was within the top 10% of GSV-point density values and zero (0) otherwise. Because logistic type models do not assume linearity between independent and dependent variables, we began the modelling process by creating a null model with the outcome as a constant function together with a second model

containing nine independent variables (excluding age(young adults), dwellings-owned, dwellings-rented, & income as discussed later in the results). Subsequently, a forward stepwise regression was used to determine the final model. A quasibinomial model was used for the final model because of its resilience to inflated standard errors as the model can account for variance above and beyond that of the binomial logit model, which can help in this situation given the spatial autocorrelation present in the independent variables.

A further issue with treating the dependent variable as dichotomous, based on the chosen 10% threshold, is that we avoid creating a zero biased variable when using this threshold-based method because it assumes gentrification is a rare event. In other words, there is a strong class imbalance between 1s (gentrified - small counts) and 0s (not-gentrified - large counts). As such, we used class weights and maximum likelihood to estimate the quasibinomial regression parameters, whereby the weights of the two classes (1s and 0s) were approximately inversely proportional (1 and 1.2 as numerators for the 0 and 1 category respectively) to the number of occurrences of each for a given top 10% categorization method. Weights were further constrained to unity, which has the effect of equalizing the class imbalance in the model since not doing so would strongly bias any binomial type model towards predictions of the majority class.

Finally, for intermodal comparison, Lee's L (Lee 2001; 2004) was used to assess the bivariate spatial dependence between the different mapped model outputs. Lee's L integrates both point-to-point correlations, namely Pearson's R but also includes the topology of mapped observations as defined by a spatial weights matrix in Moran's I. Lee's L borrows a spatial smoothing scalar term from global Moran's I to account for spatial autocorrelation within each pattern and then integrates this with the point-to-point Pearson's R correlation to assess the

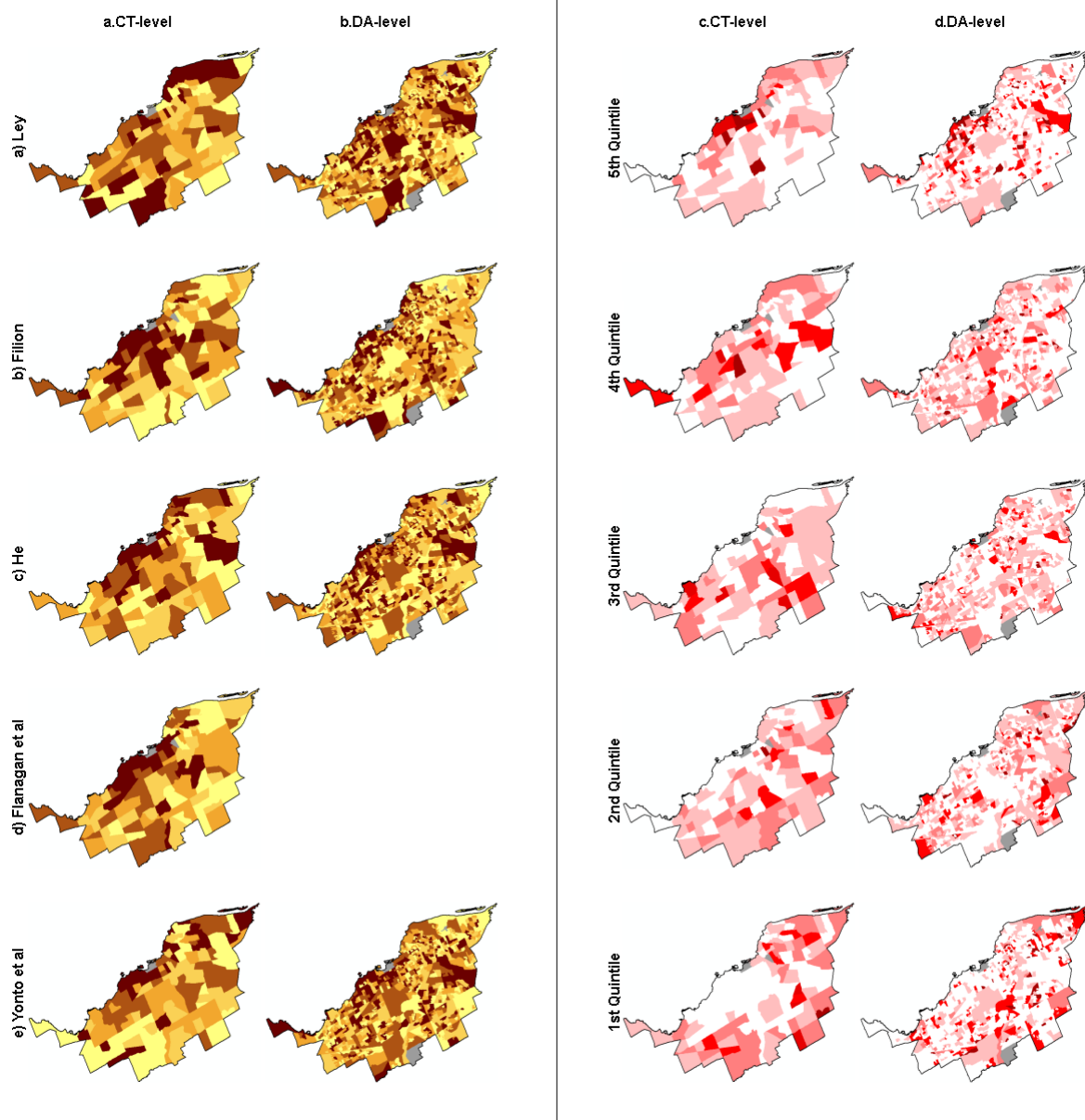
spatial correlation between two mapped patterns (Lee 2001; 2004). In general, then, Lee's L measures the similarities/dissimilarities of the bivariate association and the common co-occurrence of similar values between variables (e.g., the similarity in spatial clustering of values between layers). Lee's L was computed using 9,999 permutations. Pearson's R was also computed for comparative purposes.

## **Results**

### ***Reproducing Methodologies***

An initial comparison of each reproduced methodology's results at CT and DA geographic units show that, as expected, there is more variability at the DA level (Figure 4.2, columns a vs. b). Data for each reproduced study was classified into quintiles, whereby darker colors indicate higher quintiles. Ley's (1985) methodology at the CT level shows a much higher dispersion of the top quintile than other methodologies, which happen to have the highest quintile more concentrated close to the central parts of the city. Dispersion of top quintile results at the DA level are however very spatially dispersed for all methodologies.

The spatial patterns of reproduced methodologies led the production of frequency maps, which are an examination of how frequently each quintile appeared in CTs and DAs (Figure 4.2, columns c and d). Here too, DA level result had more variability than that of the CT level. The darker areas show more overlap between methodologies, whereas lighter shades show less overlap (Figure 4.2, columns c and d). The top quintile, at the CT level is more likely to indicate gentrification and most often appears in central areas of the city, except for Yonto et al's (Yonto and Schuch 2020; Yonto and Thill 2020) method. The methodologies have few instances of the lowest quantile appearing in central areas at the CT level, though again, there is increased variability at the DA level.



**Figure 4.2. Reproduced methodologies and frequencies.** Column a and b show results of the reproduced methodologies at census tract (CT) and dissemination area (DA) levels of geography respectively. Data is distributed into quintiles, whereby the darkest quintile is the highest. Columns c and d display maps which show how frequently each CT or DA appear in each quintile. White indicates a frequency of zero, and as the shades get darker, the more frequently the geographic area has appeared in the given quintile.

### *Examining Data Variability*

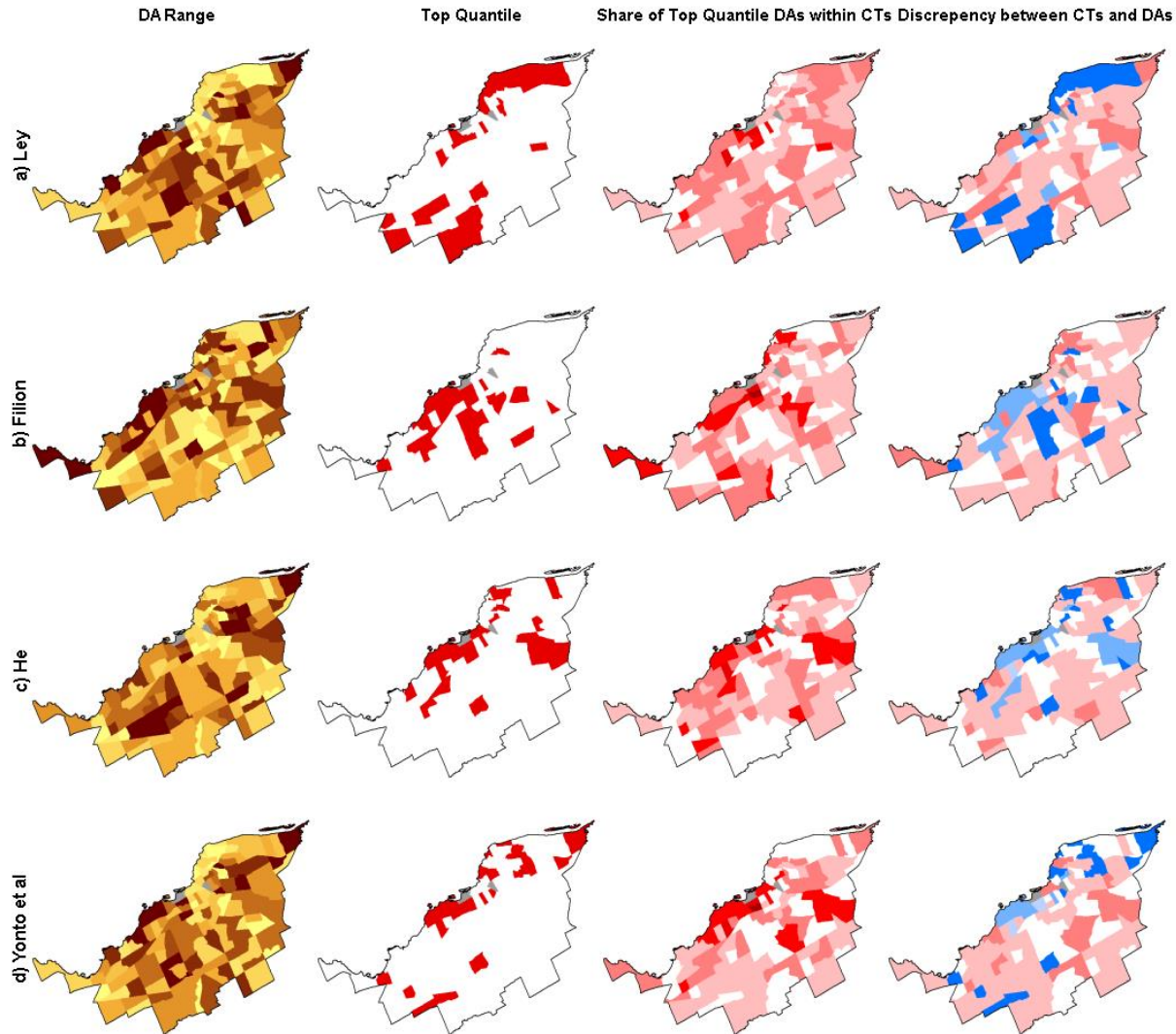
The patterns of maps in Figure 4.2 led to the examination of DA variability within CTs (Figure 4.3). The first column in Figure 4.3 shows the range of DA data within each CT. Darker

polygons have more variability. The subsequent column shows only the top quintile per each map. The map is essentially bivariate, where top quintile CTs have a value of 1 and others have a value of zero. The subsequent column shows the percentage of top quintile DAs which are within each CT. In this column values can range between zero and one. If for example, five out of ten DAs were in the top quintile, then the value of the CT in the map would be 0.5. Therefore, the darker the polygon, the higher the share of top quintile (ie gentrified) DAs.

The second and third column both represent the top quantile, or areas most likely to undergo gentrification. Whereas the second column shows a binary landscape, the third column suggests that there the urban landscape is clearly much more variable.

The last column shows the discrepancy between the second and third columns. This column is in effect a difference map. Blue areas suggest that CTs overestimate gentrification in relation to DAs whereas red areas suggest the converse.

Figure 4.3 in essence provides the same message as Figure 4.2: that spatial patterns are much more varied at the DA level than at the CT level.



**Figure 4.3. Maps of DA variability within CTs.** The first column shows the range of DA data variability within each CT. The subsequent column shows only the top quintile per each map, whereas the subsequent column shows the percentage of top quintile DAs which are within each CT. The last column shows the discrepancy between the second and third columns and is in effect a difference map.

### *Spatial Autocorrelation*

Spatial autocorrelation (SA) results of the predictions of all model outputs shown in Figure 4.2 are given in Table 4.4. Putting aside the SAR Lag<sub>y</sub> and Quasibinomial spatial autocorrelation values for now, the SA results from Yonto et al.'s (Yonto and Schuch 2020; Yonto and Thill 2020) and Ley's (1985) methodologies exhibit non-significant spatial autocorrelation at the CT level, the Moran's I values tend to indicate these methods exhibit no

spatial autocorrelation and thus no spatial self-similarity or self-dissimilarity of index values among the CTs and thus are no different from a random permutation of the values among the census units. The SA results of He's (2010) and Flanagan et al.'s (2016) methodologies have moderately weak positive SA at the CT level. In terms of spatial coherence at the CT level, Flanagan et al.'s methodology is the strongest of the five reproduced methods, followed by He's (2010) and then Fillion's (1991), showing successively weaker spatial autocorrelation. All exhibit significant but very weak spatial autocorrelation at the DA level (Flanagan et al. was excluded).

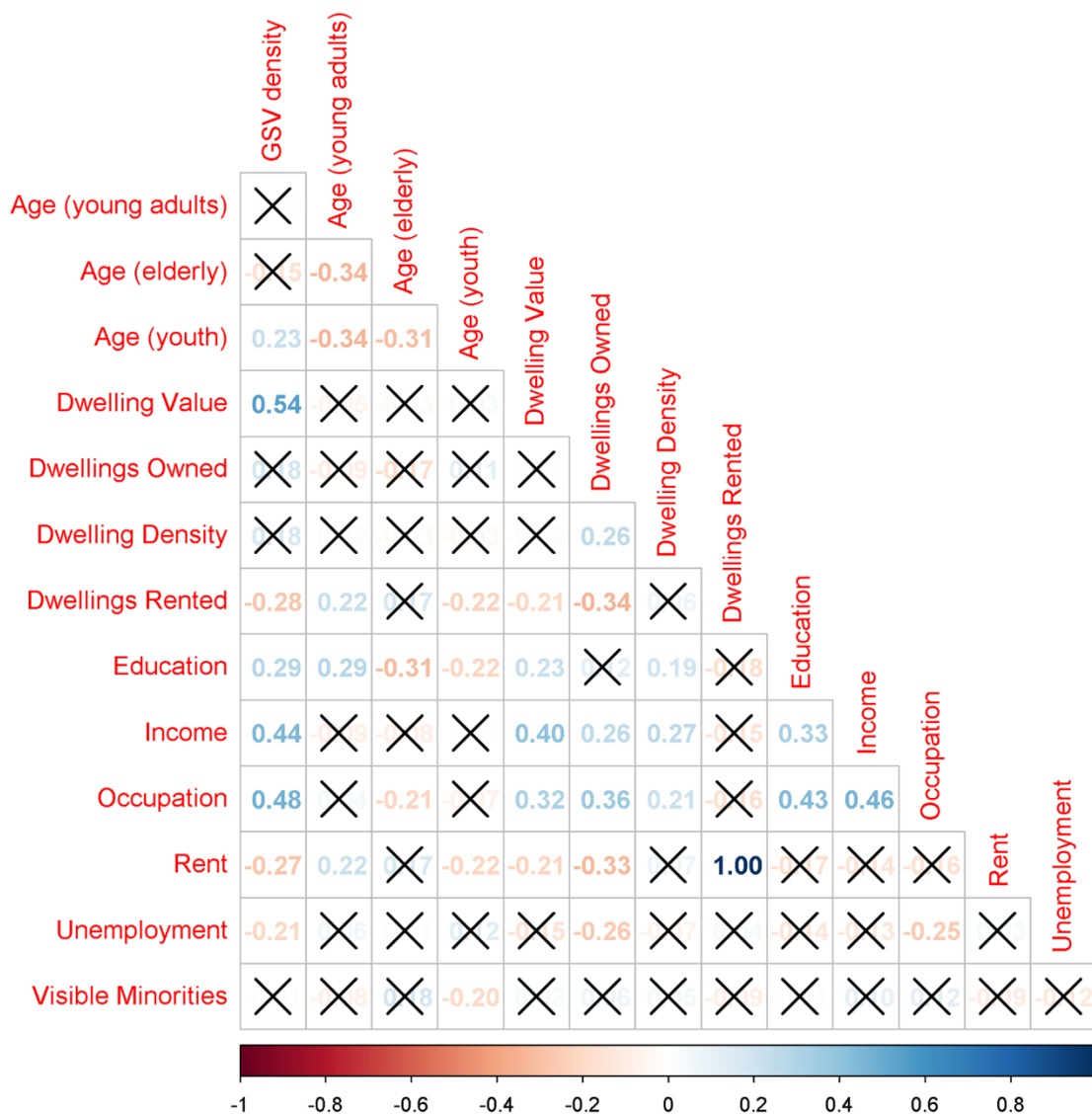
**Table 4.4. Spatial autocorrelation for reproduced methods and new regression model predictions based on 9,999 permutations of Moran's I.**

	CT level Moran's I	DA level Moran's I
Ley	-0.005	0.067**
Filion	0.150**	0.084**
He	0.204***	0.058**
Yonto Et al	-0.105	0.085**
Flanagan Et al	0.344***	
SAR Lag_y	0.719***	
Quasibinomial	0.489***	

Note: \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

### *Correlations with dependent variable*

Linear correlations between the GSV-point density and the thirteen census variables show that only eight variables have statistically significant p-values (Figure 4.4). The strongest correlations with the dependent variable are dwelling value, occupation, income and education and age (youth). Increasing values of those variables tend to be associated with increasing values of GSV-point density. Alternatively, increases in rent, dwellings rented and to a marginal extent, unemployment are associated with decreased GSV-point density.



**Figure 4.4. Correlation matrix between thirteen variables used in the literature and GSV-point density.** A represents the dependent variable after the Lambert W transformation to normality. Abbreviated census variable names are used. See S4 in Appendix V for full names and details.

Spatial Autocorrelation results (Moran's I) of thirteen census variables, at the CT level, indicates that ten variables were statistically significant (Table 4.5). The highest values were of dwelling value, income and occupation. These values are more likely to be clustered together than the others. Spatial autocorrelation on dwellings owned, unemployment, and visible minorities were not statistically significant.

**Table 4.5. Moran's I for independent variables is based on 999 permutations at the CT level.** Abbreviated census variable names are used, while the supplemental material (S4 in Appendix V) has corresponding official names.

Short form in S4	Variable abbreviation	Moran's I
B	Age (young adults)	0.180***
C	Age (elderly)	0.135**
D	Age (youth)	0.129*
E	Dwelling Value	0.296***
F	Dwellings Owned	0.027
G	Dwelling Density	0.124*
H	Dwellings Rented	0.127*
I	Education	0.088*
J	Income	0.261***
K	Occupation	0.206**
L	Rent	0.117*
M	Unemployment	-0.007
N	Visible Minorities	0.077

Note: \*p<0.05,\*\*p<0.01,\*\*\*p<0.001

### *OLS and linear spatial regression models*

Of the thirteen variables identified in the literature, only age(youth), dwelling-value, dwellings-rented, education, income, occupation, rent & unemployment had statistically significant linear relations with the dependent variable as measured by Pearson's R (Figure 4.4). However, dwellings-rented and rent are strongly collinear variables ( $R= 0.998$ ,  $p < 0.0001$ ) and thus only rent was retained for model building. After backward stepwise model selection for the multivariable OLS model, we chose the model with the lowest AIC and this was  $GSVpoint\ density \sim \beta_0 + \beta_1 Age(youth) + \beta_2 Dwelling\ value + \beta_3 Income + \beta_4 Occupation$  (AIC 793.79) (S7 in Appendix V). Variance inflation factors (VIF) were all below a value of two, suggesting multicollinearity was not an issue. Neither the Breuch-Pagan test for heteroskedasticity nor the Shapiro-Wilk test for normality of regression residuals for the OLS model could be rejected (Table 4.6). However, the OLS residuals exhibited statistically significant weak positive spatial autocorrelation (Table 4.6). As such, a specification search

commenced for a spatial regression model. The robust form of the Lagrange multiplier diagnostics for spatial dependence pointed to a spatially lagged simultaneous autoregressive model (SAR Lag<sub>y</sub>). Moreover, the common factor hypothesis compared a full spatial Durbin lag model to nested more parsimonious alternatives and the full spatial Durbin model was not significantly different from the SAR Lag<sub>y</sub> model via the likelihood ratio test (S8 in Appendix V). Therefore, the spatially lagged simultaneous autoregressive model (SAR Lag<sub>y</sub>) was chosen whereby the spatial interaction is modelled by including a spatially lagged dependent variable within the set of independent variables.

**Table 4.6. OLS multivariate model and SAR Lag\_y (spatially lagged simultaneous autoregressive model estimated by maximum likelihood).** The parentheses beneath Beta coefficient are the standard errors).

	<i>Dependent variable:</i>	
	GSV-point density (Lambert W transform)	
	OLS (1)	SAR Lag_y (2)
Age (youth)	83.878 ** (25.515)	55.764 * (22.164)
Dwelling Value	84.653 *** (17.740)	57.004 *** (15.966)
Income	38.271 (21.812)	26.136 (18.771)
Occupation	122.080 *** (32.853)	90.485 ** (28.904)
Constant	-13.609 (9.648)	-15.254 (8.285)
Rho	-	0.47***
Moran's I (residuals)	0.18**	-0.096
Shapiro-Wilk	0.99	0.99
Breusch-Pagan	1.17	4.00
Observations	109	109
R <sup>2</sup>	0.470	0.57 <sup>#</sup>
Adjusted R <sup>2</sup>	0.449	-
Log Likelihood	-511.71	-499.97
sigma <sup>2</sup>	733.8	538.8
Akaike Inf. Crit.	1,035.41	1,013.93
Residual Std. Error	27.088 (df = 104)	-
F Statistic	23.024 *** (df = 4; 104)	-
Wald Test	-	27.167 *** (df = 1)
LR Test	-	23.481 *** (df = 1)
<i>Note:</i>	*p<0.05,**p<0.01,***p<0.001	
	<sup>#</sup> Nagelkerke pseudo R <sup>2</sup>	

Koenker's studentized Breusch-Pagan tests of the residuals from both the OLS and SAR Lag\_y regression models showed no significant difference from homoskedasticity or Gaussian normality (Table 4.6). In addition, the SAR Lag\_y model had a higher likelihood, lower AIC and

no residual spatial autocorrelation when compared to the OLS model. The SAR Lag<sub>y</sub> model, like the OLS model exhibits significant coefficients only for age (youth), dwelling-value and occupation. The spatial autoregressive parameter,  $\rho$ , was statistically significant, signifying that as the GSV-point density in surrounding CTs increases, so does the density value for a given CT - above and beyond what can be explained by the independent variables alone. Moreover, the significant  $\rho$  also means that the OLS coefficients are biased and that direction is upward because OLS is unable to assign any proportion of variance within the dependent variable to a spatially lagged process (LeSage 2008). The SAR Lag<sub>y</sub> model thus contains a feedback effect whereby surrounding values of the dependent variable have an effect on the given CTs value and that given CTs value also effects surrounding density values (LeSage 2008).

The magnitude of the SAR Lag<sub>y</sub> model coefficients for the independent variables are not directly interpreted in the same way as OLS betas (as partial derivatives) due to feedback from the spatial lag term – which incorporates the spatially lagged dependent variable and its  $\rho$  coefficient within the set of independent variables. Thus, changes ascribed to any given independent variable must be broken down into the average direct, indirect and total impacts (LeSage 2008; LeSage and Pace 2009; LeSage and Fischer 2008). All independent variables in the SAR Lag<sub>y</sub> model had statistically significant impacts based on 9,999 permutations to arrive at pseudo-p-values. For age (youth), dwelling-value & occupation all impacts are statistically significant and for income no impacts were significant (Table 4.7). All of the significant impacts are positive, like their pseudo-beta counterparts in the SAR Lag<sub>y</sub>, thereby indicating that increasing positive differences in these variables are associated with increasing GSV-point density. The total impacts are not overly dominated by either direct or indirect impacts, which are approximately somewhat equal in magnitude, with the indirect impacts slightly lower, which

is common, although the indirect impacts are relatively large. Only age (youth), dwelling-value & occupation & occupation have significant indirect impacts. Those indirect effects come from the GSV-point density in surrounding CTs effecting a given CTs value of the dependent variable. The relatively large indirect effects are due to spatial spillover effects in the model (LeSage and Fischer 2008).

**Table 4.7. SAR Lag\_y model effects.**

Variable	direct	indirect	total
Age (youth)	58.7**	47.3 <sup>+</sup>	105.9*
Dwelling Value	59.9***	48.3*	108.3***
Income	27.5	22.1	49.7
Occupation	95.2**	76.7*	171.9**

Note: <sup>+</sup>p = 0.056, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

### *Quasibinomial model*

The quasibinomial forward stepwise regression models excluded age (young adults), dwellings-owned, dwellings- and income which were removed to bring the VIF for the remaining models all below a value of three. The backwards stepwise process provided a model producing the lowest AIC, which contained dwelling-value, occupation, age (youth), age (elderly), & unemployment (unemployment was marginally significant at  $p = 0.071$  and retained). Those independent variables were selected for the final quasibinomial model (S9 in Appendix V). The overall accuracy of the quasibinomial model was ~92% with a statistically significant kappa = 0.65 (Table 4.8). The prediction rate for gentrified CTs was 100% and 91% for non-gentrified CTs. There were a total of nine false positives produced by the model for the zero class and the AUC was 0.98. The approximate Nagelkerke's pseudo-R<sup>2</sup> was 0.83.

**Table 4.8. Quasibinomial model output.**

Confusion Matrix			Accuracy measures	
	Gentrified	Non-gentrified	Total Accuracy	91.74%
Predicted gentrified	10	0	95% CI	(0.849, 0.962)
Predicted non-gentrified	9	90	Positive Accuracy	100.00%
			Negative Accuracy	90.91%
			Kappa	0.647
			No Information Rate	0.826
			<i>p</i> -Value [Acc > NIR]	0.005
			Mcnemar's Test P-Value	0.008
			Sensitivity	0.526
			Specificity	1
			Pos Pred Value	1
			Neg Pred Value	0.909
			Prevalence	0.174
			Detection Rate	0.092
			Detection Prevalence	0.092
			Balanced Accuracy	0.763

All independent variables included in the quasibinomial model produced statistically significant coefficients (Table 4.9). Thus, a positive increase in any of these variables means that gentrification becomes more likely. This is exhibited via the odds ratios. For example, dwelling-value, occupation & age (youth) have the smallest *p*-values, suggesting that these three variables have the strongest association with the probability of gentrification. A unit increase in the differences represented in dwelling-value, increases the odds of gentrification by around 59, whereas a unit increase in occupation increases the odds of gentrification by a factor of 9.8.

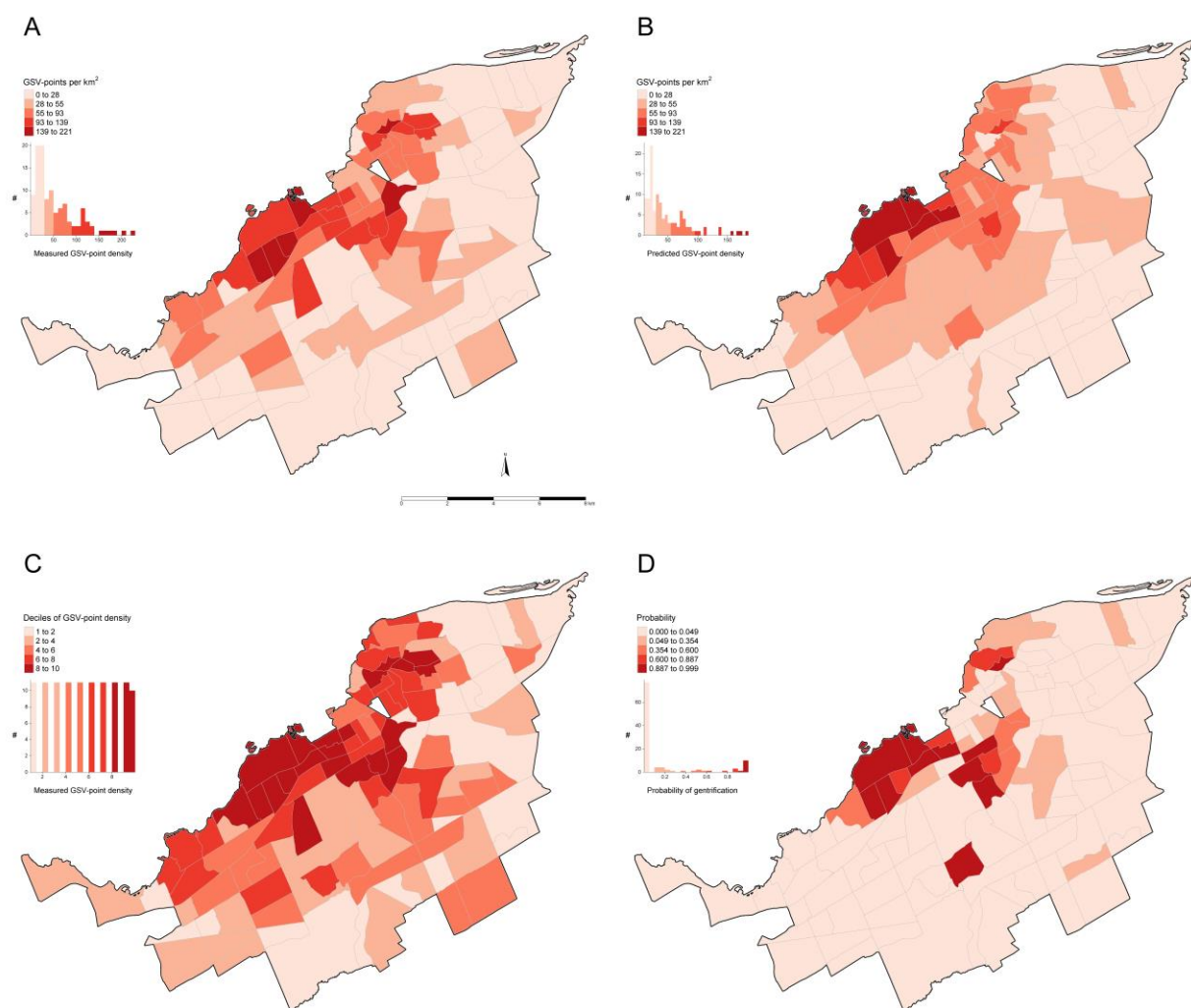
**Table 4.9. Quasibinomial regression model.** Standard errors are in parentheses below each beta coefficient. OR: Odds ratio

	<i>Dependent variable:</i>	
	Binary GSV-point density	OR
Dwelling Value	4.0871 *** (0.9584)	59.56
Occupation	2.2802 *** (0.4716)	9.77
Age(youth)	1.8216 *** (0.5048)	6.18
Age(elderly)	1.9531 ** (0.6163)	7.05
Unemployment	0.8060 ** (0.2651)	2.23
Constant	-5.7276 *** (1.3705)	-
Observations	109	
AUC	0.981	
Akaike Inf. Crit.	31.85	
Null deviance	23.3	
Residual deviance	5.3 (103 df)	
Dispersion parameter	0.0781	
Note:	**p<0.01; ***p<0.0001	

### *Quantitative Inter-model comparison*

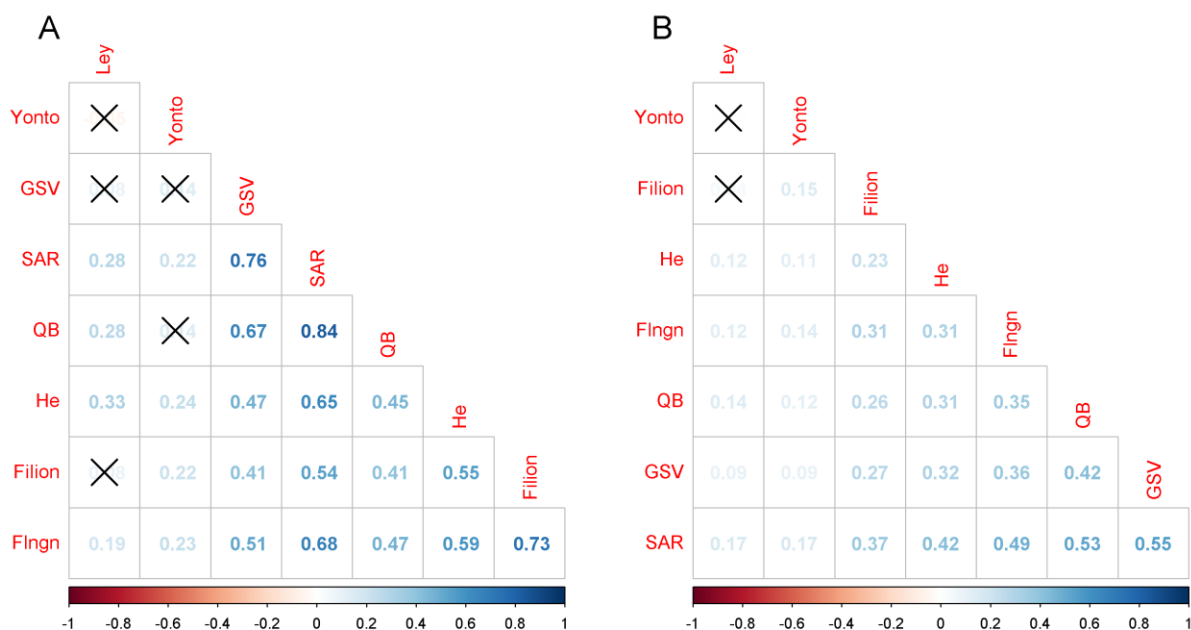
The map of the dependent variable (Figure 4.5A) defines what our model output is tested against. The observed GSV-point density map predictions (Figure 4.5B) share the general features of high density towards the central regions of the study area. They both show the highest densities in the same regions. Likewise, one can see where the top decile of GSV-point density exists in Figure 4.5C where the darkest colors (highest density) tend to be where the observed GSV-point density and the SAR Lag<sub>y</sub> also have the darkest colors. The general patterns in density are similar across the maps within the region dominated by the top 20% of GSV-point density values. The quasibinomial model output has the highest probabilities of gentrification within the dark red region of the observed GSV-point values (Figure 4.5C).

Using the mapped predictions from the five reproduced models (those in the first column of Figure 4.2 and the second column of Figure 4.5 and the dependent variable in Figure 4.5A), two correlation matrices were created: a Pearson's R point-to-point correlation matrix (Figure 4.6A) and a Lee's L spatial correlation matrix (Figure 4.6B), including the quasibinomial and SAR Lag<sub>y</sub> predictions. We do this to determine which models provided the best predictions of gentrification as measured by the observed GSV-point density.



**Figure 4.5:** A) Observed density of GSV-points per unit residential area; B) Predicted density of GSV-points per residential area from SAR Lag<sub>y</sub> model; C) A in deciles; D) Probability of gentrification predictions from quasibinomial model. Note: legend value scales are different among the panels.

For purposes of comparison, we used the probabilities from the quasibinomial regression rather than converting these to a binary outcomes. We ordered the variables within the correlation matrices using hierarchical clustering. By sorting the correlations in that manner, it becomes easier to see which model predictions have the highest magnitudes of correlation and similarity via spatial clustering. For example, in Figure 4.6A, He (2010) and the GSV-point density are moderately linearly related ( $R = 0.47$ ). Conversely, looking at Figure 4.6B, the relation is lower ( $L = 0.32$ ) when spatial correlation is included in the correlation. Most of the time, Pearson's  $R$  values suggest stronger relations than using Lee's  $L$  because of the incorporation of spatial structure of the patterns. The absolute values of the two different correlation coefficients are not directly comparable. Nevertheless, when considering the spatial/correlation coefficients of all the variables, the relative differences between Flanagan et al. (2016)'s and GSV point density in Figure 4.6A is high enough to incorrectly conclude that their map is more similar to the observed GSV point density than when considering the spatial correlation (Figure 4.6B). The quasibinomial and SAR Lag<sub>y</sub> had the strongest relation to the GSV-point density. This makes sense since they were built to predict that dependent variable specifically, whereas the other methodologies were used to predict a given construct of gentrification. Hence, this means that our best models are the SAR Lag<sub>y</sub> and the quasibinomial probability map. However, Flanagan et al. (2016) also performs best out of the models not devised in this paper.



**Figure 4.6. A) Pearson's R between mapped outcomes at the CT level from all models; B) Lee's L bi-variate spatial correlations.** GSV - GSV point density per unit residential area; SAR - predictions from SAR Lag<sub>y</sub> model; QuasiB - predicted probability of gentrification from quasibinomial model; Ley – Ley (1985) model predictions; Filion – Filion (1991) model predictions; He – He (2010) predictions; Yonto – Yonto (Yonto and Thill 2020; Yonto and Schuch 2020) predictions; Flnagn – Flanagan et al. (2016) model predictions. NOTE: The order of variables within the matrices are based on a hierarchical clustering method and this means panel A and B have different orders based on within group similarities between methods and the dependent variable, GSV. All coefficients are statistically significant at the  $\alpha = 0.05$  level except for those correlations marked by an 'x'. No corrections for multiple testing were deemed necessary since these matrices are mainly used for interpretation. The order of variables is different between A and B.

Figure 4.6 also shows that the inter-model comparisons between the reproduced studies vary in their ability to predict the same CTs as being gentrified or otherwise. Yonto's (Yonto and Schuch 2020; Yonto and Thill 2020), Ley (1985) and He (2010) have the lowest correlations between themselves and better performing models, as measured against the dependent variable, whereas Filion (1991) and Flanagan et al. (2016) have much stronger correlations with each other and with the dependent variable.

## Discussion

We did not expect all thirteen census variables to be important in modelling, but we expected education and income to be the strongest stand-alone variables, as they were so prevalent in earlier studies (Table 4.2). Figure 4.4 shows that education was not one of the top three strongest variables. Both education and income variables are not included in the final SAR Lag\_y model. Education is simply a weaker variable and perhaps the income variable has some parsimony with dwelling-value and occupation.

As defined by Lesage (2008), the SAR Lag\_y model (a model nested within the full spatial Durbin model) incorporates the effects of the spatially lagged neighbouring region's GSV point density, the neighbouring density values, the magnitude of the region's density itself, and the region's unique composition of independent variable values. We observe that the spatially lagged variable possesses a similar amount of indirect power as the direct effects of the three significant independent variables. The SAR Lag\_y model has a theoretical basis because it assumes that the neighbouring values of GSV-point density would influence a given CT's density because of spillover effects (Davidson and Lees 2005; 2010). However, at the same time because CTs are artificial boundaries, we could have equally treated the spatial dependence in the OLS model residuals as a nuisance error because the boundaries have no a priori relationship to any given gentrification process in most cases. We tested a SAR error model but it produced a greater AIC than the SAR Lag\_y. However, its AIC was less than the original OLS model (S10 in Appendix V).

By transforming the dependent variable to a dichotomous outcome, we assume that gentrification is present when a CT falls within the top 10% of GSV density values. More specifically, a dichotomous outcome assumes that gentrification is a process that is complete for

such neighborhoods coded as ones and absent or incomplete for those coded zero. This dichotomous coding provides, in our opinion, a liberal number of gentrified CTs which were ten in total. While the assumptions of a dichotomous outcome may or may not be realistic because of the uncertainty of knowing when a place is gentrified, it simplifies the modelling process because it allowed for the use of continuous variables to determine the probability of a CT having gentrified.

Although the models were all successful to some extent, the quasibinomial model possessed the highest pseudo- $R^2$  value when compared to the SAR Lag<sub>y</sub> model. However, these were predicting different aspects, in the former case, a threshold-based method was applied based on the top 10% GSV-point density values and in the latter, the GSV-point density values themselves were predicted and had a lower pseudo  $R^2$ . One advantage of the quasibinomial model is the ability to map the probability of belonging to the 'gentrified' class or category thereby providing the ability for an interested party to assess potentially gentrifying CTs according to some threshold of probability. The SAR Lag<sub>y</sub> model provides a continuous outcome in GSV-points per unit residential area and could be used similarly to predict where gentrification is happening at some density.

Considering the spatial correlations among the mapped predictions from each of the methods used in this paper, our quasibinomial probability map and SAR Lag<sub>y</sub> spatial autoregressive models had the highest bivariate spatial correlations with the dependent variable. The SAR Lag<sub>y</sub> model contained four variables (age (youth), dwelling value, income and occupation) which happened to be the most spatially autocorrelated census variables that were examined (Tables 4.4 and 4.7). The method of Flanagan et al. (2016) also produced a map that was strongly correlated to the dependent variable. If one was to compare the predicted maps

using only Pearson's R, conclusions regarding similarity between the values and their patterns would be less accurate. When two different mapped patterns are compared with Pearson's R only, the correlations are higher than the corresponding spatial correlation. This is typical of Lee's L (Lee 2001; 2004) because Pearson's R values are higher because point-to-point comparisons do not take account of the lack of spatial independence and thus are duplicative, thereby decreasing p-values and inflating Type I error rates. Lee's L can be interpreted as the correlation between two mapped patterns when accounting for spatial autocorrelation. Even for identical maps Lee's L will not take on a value of one, like Pearson's R would. In such a case Lee's L is determined by the spatial structure of the two patterns only, that is to say, the spatial smoothing scalar.

Regarding Figure 4.4, Ley's (1985) method had no relation to our independent measure of gentrification. Yonto et al.'s (Yonto and Schuch 2020; Yonto and Thill 2020) model fared somewhat better, but as it is based on a poorly performing model, it also fails to yield fruitful results. Of course, those two models exhibited no significant spatial autocorrelation and so were technically no different from a random process. A random process, when compared to the highly structured GSV-point density, will have little spatial correlation as evidenced by lower values for Lees L and even Pearson's R. The studies by Filion (1991) and He (2010) performed roughly half as well as that of Flanagan et al. (2016) when compared to the dependent variable. They both exhibited weak positive but significant spatial autocorrelation. Both of these two models only examined one census variable for measuring gentrification. Filion's (1991) methodology used income, which has significant spatial autocorrelation and their model performs the third best of the reproduced methodologies. Filion (1991, 559) noted that using income alone may

give an incomplete picture of gentrification, whereas He (2010) simply had a more limited set of data to work with.

Regarding Figure 4.6, Ley's (1985) and Yonto et al.'s (Yonto and Schuch 2020; Yonto and Thill 2020) method had no significant linear relation to our independent measure of gentrification and both only possessed a very weak positive spatial correlation. The studies by Flanagan et al. (2016), Filion (1991) and He (2010) all performed consistently between the correlation and spatial correlation measures from a relative viewpoint. He (2010) and Filion (1991) did not perform as well as Flanagan et al. (2016) and that is likely due to the fact that those two models only examined one census variable which is insufficient to explain the variance in the GSV-point density.

Except for SAR Lag<sub>y</sub> and the quasibinomial model, none of the other models tested in our research were strictly intended to predict the visual signs of gentrification and thus, the generally poor performance of some of the models may be due to the inappropriate choice of the dependent variable or a difference between our concept of gentrification and the one intended by the original author. It is entirely possible that, the comparisons were spurious since none of the studies claim to measure/predict the visual signs of gentrification or intensity of gentrification as measured by the density of GSV-points. However, given that our dependent variable is constant and has been shown to have validity in Ottawa, the comparisons are not without merit but should still be cautiously interpreted. In general, no single census-based model captures all the variation in the GSV point density and thus a general gentrification model is unlikely to apply across different urban regions using census data alone.

Measuring gentrification remains a complicated endeavor, as this research demonstrates. Gentrification is context dependent and can take place on small area, resulting in difficulties to

ascertain its extent. Furthermore, census dates are often insufficient markers of the start and end points of when an area has experienced gentrification, as has been lamented by Moskowitz (2017). Therefore, two caveats require elaboration. The first is that our dependent variable, the visible expression of gentrification as seen through Google Street View (GSV) imagery is imperfect, as the true extent of gentrification can be greater. Perhaps an alternative source of data can be building permits. However, the City of Ottawa's online public records do not have such permits digitized into a comprehensive list dating back to 2006.

The second caveat is that, as Curran (2018, 64) notes, gentrification is almost never complete. Curran's statement is not elaborated but can be explained in multiple ways. Firstly, neighborhoods generally are not homogeneous and thus if part of a neighborhood gentrifies the other part might not (Curran and Hamilton 2012, 1032). Secondly, it is not typical that everyone gets displaced in one instance of gentrification. An area's residents can be increasingly squeezed as time goes on and as gentrification reaches more mature stages. Thirdly, gentrification can displace those who previously displaced others. In this sense, even though gentrification occurred once at one point in time, it may keep going.

We cannot pinpoint a precise cutoff as to where gentrification has taken place in a particular area. Throughout this research, we focused on the highest quintile of CTs undergoing various measures, not because those areas necessarily underwent gentrification, but because the highest quintile is most indicative of gentrification, as it is in the highest quintile that the most significant increases of socioeconomic indicators occurred. It is possible that a better measurement might be the top decline of census tracts (CTs) or dissemination areas (DAs).

In each quintile map (Figures 4.2, 4.3 & 4.4) the darkest areas represent the highest quintile, and therefore the most likely places that experienced gentrification. Most of these areas

are in or near the central part of the city. However, as an anomaly, CT 20.02 is in a suburban area. Areas that strongly identify in the first quintile are likely to be areas undergoing decline and are often located in places that are commonly referred to as the inner suburbs, aligning with earlier research which finds that suburban decline is taking place in such spaces (Pavlic and Qian 2014; Ilic and Sawada 2021).

As expected, the DA level maps (Figure 4.2) show a finer level of detail and variability of data than the CT level maps. For example, some CTs may not be in the top quintile, across reproduced methodologies, but may contain DAs which are in the top quintile. Conversely, some CTs in top quintiles, across reproduced methodologies, contain DAs that are not in the top quintile. The fourth column of Figure 4.2 shows that few CTs from the 2nd and 3rd column overlap. Where they do overlap, it is usually areas that had values of zero rather than areas with values of 1. However, the patterns across reproduced studies do not show many similarities in terms of which areas are overrepresented or underrepresented in terms of DA variability.

All model intercomparison was based on the agreement on which CTs or DAs were in the same quintile across the models. Models were also quantitatively assessed for their linear relation with the GSV-points per unit residential area which was the only independent data available for gentrification over the same time period in this study in absence of a 'true' map of gentrification in the city. It is entirely possible that, the comparisons were spurious for the fact that none of the studies claim to measure/predict the visual signs of gentrification or intensity of gentrification as measured by the density of GSV-points.

We note that GSV data is imperfect for two main reasons. The first is that while GSV data is available for many years, its starting point is in 2007, which is after the census in 2006. As the GSV data and census years do not line up perfectly, some amount of uncertainty is

induced. The second reason was already mentioned and pertains to GSV data being imagery visible from the street, and as such does not capture all development that occurs on a property because there are instances when properties experience redevelopment on their interiors or in their backyards. Such redevelopments may not be picked up by our GSV model because they are not visible from the street. Not all areas have a lot of GSV data, whereas sometimes GSV data can be obfuscated.

Measuring contemporary gentrification is more arduous than in the past because gentrification as a process has mutated (Lees, Slater, and Wyly 2008a), in the sense that researchers have identified many types of gentrification. Ottawa reflects the nuanced gentrification landscape. Different processes can contribute to gentrification, and how these processes come together can vary between neighborhoods. It is therefore important that one proceeds with caution when trying to examine gentrification on a larger scale, as nuanced types of gentrification could be easy to miss.

There are numerous types of gentrification that conventional measures using census data could have a hard time recognizing. For example, Ottawa's neighbourhood Sandy Hill underwent some gentrification in our study period. However, the type of gentrification in Sandy Hill is mostly studentification. Over the years an ever-greater number of students resides in this neighborhood, as the number of families decreased. Students are not typical gentrifiers, as they tend to have lower incomes and often rely on other means of finance which includes but is not limited to: financial support from their parents, student loans (particularly OSAP), various scholarships, and their own savings. It is the absentee landlords that improve the appearance of a property in order to attract higher net rents in this area of Ottawa.

Another type of contemporary gentrification which may be more difficult for traditional gentrification measures to identify is intensification. This is the case of the neighborhood Westboro, which did not previously experience decline as typically takes place in gentrifying areas. What sets Westboro apart from many other neighborhoods is the extent of which intensification is occurring (Willing 2021). In what has been termed “over-intensification”, a single property could see several units replace what was once one home. Therefore, such contemporary changes add a dimension of complexity when examining the gentrification process on a large area.

On a final note, there may be other means to measure gentrification. For example, researchers from Germany conducted shift-share analysis with real-estate data in lieu of census data to examine gentrification in Berlin (Holm and Schulz 2018). Such data sources in the North American context are scarce (an exception is Kary(1988)) and remain underexamined and uncovering such data may be fruitful endeavors for future quantitative gentrification research.

## **Conclusion**

The endeavor of measuring gentrification requires more model intercomparisons, given that, in this small study, the discrepancies between model abilities to predict where the phenomena is/has happened models are quite large. It is clear that the definition of gentrification is fundamental to measuring model success and/or model production. In this research, we had two very specific definitions of gentrification, namely, the density of GSV points per residential area and the top 20% of density values being equivalent to gentrification between the years 2007 and 2016. In the studies we reproduced, the definition of gentrification was based on the model output and with no independent data to compare the outputs to, gentrification becomes a hypothesized process based on the model that produced the output.

Whereas many researchers do not justify their selection of census variables, or simply base their methodology on past research, herein a holistic examination of gentrification with a variety of variables was conducted. There seems to be a relation between the strength of the spatial autocorrelation in the variables that studies used and the studies performance against the dependent variable. However, more work is required to determine generalities, but the weaker the spatial structure of the model output, the worse it performed against a structured variable like GSV-point density.

A major issue facing modelling of gentrification is which census variables to choose. We reproduced models from different urban areas and variables that might have a high degree of spatial autocorrelation in one city may not in another. When comparing models using spatial correlation indices like Lee's L, if one model lacks significant spatial structure it will not perform well against a structured independent variable. Of course, we are assuming that the process of gentrification is positively autocorrelated and this may or may not be the case in reality, but like most spatial processes measured in arbitrary spatial units, it is likely to be spatially autocorrelated, particularly if spillover effects have had time to work.

Two main models for examining gentrification in Ottawa were produced. The first was a spatially lagged simultaneous autoregressive model (SAR Lag\_y), which explaining roughly 57% of the variance of GSV-point density using three census variables. The second was a quasibinomial regression model which achieved an accuracy of 91.7% accuracy. However, these models were built to predict the dependent variable of GSV-point density or our dichotomous variant and therefore they should perform better than models not produced to do that task.

The research reproduced various studies and then applied these methodologies to census tracts (CTs) and dissemination areas. Our results suggest that the sub-CT analysis can serve as a

basis for future studies, as this lower level of geography allows for a finer reflection of the pathways of gentrification and neighborhood change.

Through the comparison of our model with deep mapping results of gentrification developed at the property level from visual indicators in Google Street View data, we are able to better determine how inferences based on CT measures compared in time and space with an approximation of the ‘true’ gentrification processes in urban space.

### **Endnote**

1. The London and Palen (1984) book was the first to explicitly be about gentrification. Earlier volumes exist which do not specifically label urban renovation and revitalization as gentrification (Clay 1979; Laska and Spain 1980).

## Chapter 5. Conclusion

### Summary

This thesis contributes to urban social justice by providing a better means by which to analyze patterns and processes of urban inequality in Canada. Broadly speaking, this work deals with uneven development, whereas the two main processes examined herein are income polarization and gentrification.

The first chapter established the following research questions:

- (1) Is income-polarization taking place in the largest Canadian CMAs? What are the spatial patterns of this phenomena in different cities?
- (2) Can the spatial and temporal limitation of census data for measuring gentrification be surmounted through the usage of street level imagery and the use of deep learning methods?
- (3) To what extent do quantitative models of gentrification based on census data agree with each other and how well do they measure gentrification?

The second, third and fourth chapters address the aforementioned questions respectively and in doing so are contributions to the broader field of urban studies and more specifically urban geography and geomatics. While the methods herein contribute to urban social justice, they focus on delineating spatial patterns of inequality rather than philosophical and/or policy elements which may be more ambiguous. In effect this thesis contributes to the empiricization of processes which are mainly discussed through conjecture by those who deal with qualitative paths of inquiry.

The second chapter addressed the first research question by establishing a means by which to examine income polarization in urban areas and does so in a more robust manner than

previous research in a few ways. Firstly, it was necessary to justify income as a sole or only variable of analysis for inequality research. After all, examining income inequality would be of little use if there was no relation between this variable and inequality in general. As such, and for the first time, this variable was explicitly justified as a general measure of inequality with insight from both theory and past research before examining income polarization. Analytical results show that the middle-income group has declined in every Census Metropolitan Area (CMA) examined between 1971 and 2016. Additionally, the low-income group generally expanded. Spatial fragmentation analysis helped in further understanding the divided urban landscape and the findings show that spatial fragmentation of income groups has taken place in each examined CMA through time. Lastly, using spatial autocorrelation, high-income groups were found to be increasingly spatially similar to nearby census tracts or loosely speaking, coalescence of high-income through time was evident. Overall, the middle-income population has dramatically declined in Canada's largest cities.

All the findings from the second chapter contribute to urban social justice as they raise alarm to increasingly divided urban landscapes and offer insight into the spatial patterns of such processes. Letki and Mieriņa (2015) observed that polarization leads to the less affluent becoming more isolated. In the present research, this isolation can be seen through heightened fragmentation of low-income groups, resulting in various impairments such as having limited access to resources and potentially being trapped in circles of poverty and marginalization. Such situations are compounded by the reality of the poor being confronted with the increased stigmatization (Taylor-Gooby 2013). These realities suggest impaired opportunities of low-income groups, which is precisely a focal point of social justice.

The second chapter has significant connections and correlations to urban social justice, as it quantifies processes that are quintessential elements of the increasingly divided urban landscape and while doing so offers spatial information through which one could examine the spatiality of inequality. The subsequent two chapters are also related to social justice in similar ways, but they deal with a more focused examination of a type of uneven development taking place: gentrification.

The third chapter addressed the second research question by overcoming the limitation of assessing gentrification at sporadic census years by developing an artificial intelligence to analyze Google Street View (GSV) imagery for the visual signs of gentrification between 2007 and 2016. The model produced points on properties that underwent a gentrification-like visual improvement over the time period. Isarithmic maps (so called 'heat maps) were produced from the identified properties (points or GSV points) that underwent a gentrification like change using kernel density estimation. Areas both known to experience gentrification and those not thought to experience gentrification were identified and this was validated through concordance with an analysis of building permits issued between 2011 and 2016.

The third chapter examined the last research question by mapping and modelling gentrification between 2006 and 2016 in Ottawa. Starting with an examination of the dominance of qualitative approaches to gentrification, the chapter follows up with a comprehensive examination of previous quantitative gentrification research that used census data. Five approaches from the literature that measure gentrification at an interval/ratio level were reproduced at both the census tract (CT) and finer dissemination area (DA) geographic units in Ottawa. At the CT level, a comparison was made of the predictive power of reproduced methodologies and of results generated through multiple types of regression modelling.

Earlier research helped pave the way for a holistic examination of thirteen variables used in quantitative analysis of gentrification. Each variable was examined individually before using OLS stepwise regression. The reproduced methodologies, the SAR Lag\_y and quasibinomial models were built to predict GSV-point density (called GSV-points) produced in the proceeding chapter. An OLS model contained spatially autocorrelated residuals and so a stronger spatially lagged simultaneous autoregressive model (SAR Lag\_y) was used. That model contained four census variables: Age (youth), Dwelling Value, Income and Occupation. Furthermore, a form of logistic regression called quasibinomial modeling achieved an overall 91.7% accuracy at predicting which CTs underwent gentrification during the time period studied. The SAR Lag\_y model, whose results explain 57% of the variance of GSV-points, performed best.

Considerable differences between methods leads to the conclusion that gentrification is highly spatially dependent on local and temporal specificities. For example, certain urban areas might have more government related employment, whereas others might be more tourism based while some might have more industries of a particular type, or educational institutions. Hence, the base of economic activities is likely to have impacts in terms of what type of gentrification and urban redevelopment takes place. Therefore, efforts to measure gentrification should be locally focused. There is no general set of variables that can be used for all models of gentrification and successful ones will vary by region and potentially across one municipality.

### **Contributions to Geography**

The third and fourth chapters of this thesis noted the dominance of qualitative methods in gentrification research. Since the 1960s, much has been said against quantitative methods. This form of inquiry has been perceived as antithetic to social justice initiatives (Cokley and Awad 2013) and geographers have at time condemned approaches such as mapping (Harvey 1973,

144). This is indicative of the reality that many human geographers have simply stopped counting and measuring (Dorling 2005, 254). They have abandoned mapping in general, and qualitative methods have sidelined quantitative methods (Dorling 2005, 255). This is unfortunate considering that income polarization is significantly affecting the social structure of cities over time and because gentrification is at the forefront of urban restructuring (Paton 2014, 31). Therefore, understanding these processes better can be the first steps towards an improvement and move towards a more equitable society. The mechanisms by which such changes can happen are beyond the scope of this study, but they would involve policy decisions which are aimed at reducing class divides and a move away from neoliberalism. A quantitative approach is nested within radical positivist methodology, which has realistic potential in forwarding urban social justice (Wyly 2011, 906). As such, the research herein offers pathways to bridge the divide between quantitative methodology and social justice oriented research.

Chapters two, three and four of this study provide new means by which to quantify both income inequality and gentrification, which are important elements of contemporary spatial change in urban contexts. Such methods can be beneficial for any analysis which seeks to assess change in urban environments. For policy makers, such studies point to areas which are undergoing decline as well as those which are undergoing investment – perhaps what the late Neil Smith referred to as the seesawing of capital on the urban scale (Smith 2008). When there is much decline or a concentration of low-income groups emerging across space, then such information is potentially useful to direct resources to aid the inhabitants of such places by state officials with the power and will to do so. When there is a rapid appreciation of a neighborhood, gentrification is usually occurring through the redevelopment of housing and the replacement of previous inhabitants by new ones (condo construction along Rideau street in Ottawa is a stark

example). In general, policy makers should plan to protect the most vulnerable populations, implement mitigative measures to minimize the displacement pressures brought with development and investment or real estate speculation.

The second chapter, as an anonymous review stated: “*represents a very rigorous empirical analysis of a very complex issue...and sets a new benchmark for the spatial-temporal analysis of income polarization in Canadian CMAs.*” It is precisely the rigor in this chapter which sets it apart from earlier work on income polarization. Part of the rigor is simply due to temporal aspects, as more data was available for analysis. Past researchers, such as MacLachlan and Sawada (1997), had only three census years of data that spanned only 20 years, to work with. However, using insight from such studies, as well as from theory, lessons learned were applied to structure a statistical analysis that uses an array of techniques such as bootstrap sampling, fragmentation analysis, spatial autocorrelation and cartogram production. These components constitute some of the rigor which earlier studies lacked.

Recent studies on income polarization generally take a descriptive methodological approach, whereby spatial patterns are described without statistical analysis. This group of researchers is cognizant that there are more rigorous methods (Rose and Twigge-Molecey 2013, 22), but the lack of them in urban geography poses a void. In this sense, the benchmark established in the second chapter is such that it serves as an example for others to try to surpass or at least match.

The third chapter provides a novel method with which to measure gentrification. The study demonstrated the first use artificial intelligence (AI) in qualitative assessment of gentrification and is one of less than half-a-dozen using deep learning papers that exist for assessing the urban perceptual environment. MIT’s César Hidalgo, a pioneer in that regard,

described it as “*a well-executed, solid piece of work*” and as a “*stepping stone*” for future research (Wu 2019). The study is novel in that it processes every property image, in Ottawa, that was available on Google Street View over a nine-year time-span. Furthermore, the study shows that an AI system can be trained to identify the visual cues of gentrification over a large urban area. Identified properties/points were mapped and kernel density estimation (heat maps) was used to remove noise caused by spatially isolated random developments which were not gentrification, i.e., “one structure does not make gentrification”. Hence, this provides for a unique confluence of urban geography, AI, and GIS.

When discussing the third chapter, Winifred Curran observes that the paper exemplifies gentrification as “*you know it when you see it*” and that the research “*is a great way to demonstrate that Google and big data can be used to address local concerns*” (Wu 2019), as there are many ways to challenge gentrification and to provide help for those who are being squeezed out of their present communities. I would hope that the research in this chapter would be used for purposes of providing assistance to those that gentrification hurts, as the identification of places where gentrification is taking place is one means by which to target policy at places in need. However, there are many potential benefits that speculators and the city can take advantage of with this spatial information. The latter especially can benefit from planning perspectives which include but are not limited to service provision, property tax assessments, building permit enforcement, and open-space allocation.

The fourth chapter tackles the arduous task of modeling gentrification with census data. It is difficult to measure gentrification, because there is no consensus on how to do so. In effect, quantitative studies modeling gentrification using census variables have assumed that their model-output is the de-facto definition of gentrification. The lack of consensus as to which

census variables should be used to measure gentrification serves as a testament to there being no universal quantitative definition of gentrification. For any field to become a fruitful field of study within urban science, as in any science, an established definition is the first fundamental step in moving a field forward. In that regard, the last chapter is innovative in that it shows the lack of consensus to be a truth given the wide range of definitions in the literature. It would not be out of place to argue that some authors are mislabeling what they are measuring as ‘gentrification’, especially if they are using inappropriate variables. Their measurements may simply be the changes of their selected variables rather than gentrification. Moreover, aside from the methodological contributions in that chapter, it was demonstrated that an independent definition of gentrification could be used to build models based on socioeconomic variables as independent variables that are used to predict GSV-point density from the deep learning model developed in the third chapter.

One of the fourth chapter’s contributions is that it is the first intercomparison study of gentrification using census data that use models which result in an interval/ratio variable. Others who have completed intercomparison research have a weakness in that they use inadequate gentrification terminology and in doing so default to a threshold method for identifying gentrification. The narrow conceptualization of gentrification centers around gentrification being defined as a process that takes place in previously disadvantaged neighborhoods. Researchers often define such disadvantaged areas with income, and in doing so constrain their study area to census tracts whose average income was less than that of the urban area’s average. This thinking, while fine in the 1990s, is outdated, because gentrification has mutated and changed such that any space can undergo gentrification, not just places with lower incomes or those limited to the so called “*inner city*” (Lees, Slater, and Wyly 2008). Each rung on the social class ladder

requires lower rungs (other than the first). Even upper-middle-class regions can be gentrified (Lees 2003; Butler and Lees 2006), and this is being observed in Ottawa. Being up to date on gentrification research is necessary to pick up on such nuances, as they are fundamental. Such changes to definitions are missed when one simply accepts dated methodologies without critical scrutiny.

The fourth chapter also examines gentrification at the sub-CT level. While sub-CT mapping has been done in gentrification, it remains uncommon. What makes the research in this chapter novel is that it is first academic attempt to map gentrification at the Dissemination Area (DA) level. Explicating the variation of data at the sub-CT level is an additional contribution. Unfortunately, DAs are limited to only Canada, and as such might be less interesting to those whose researchers who are abroad. No attempt was made to build models at the DA level. Given the modifiable areal unit problem and its impacts on aggregation, the CT-level models were not overly strong at that coarser level of aggregation. Hence, DA-level models would have performed even worse at a finer level of aggregation. Lastly, the fourth chapter also poses new modeling approaches by which to examine gentrification: a spatially lagged simultaneous autoregressive model (SAR Lag\_y) and a quasibinomial regression model.

These unique contributions advance the field of urban geography as they develop quantitative methods that help bring a greater level of understanding to spatial processes which are too often limited to qualitative inquiry.

### **Potential Future Research**

There are a variety of ways, both quantitative and qualitative, in which future research can build on the work that this thesis presents.

The second chapter addresses income polarization and those conducting research on this matter are faced with the daunting task of how to design such a study. There is limited research on this matter, and without concrete consensus in the scientific community regarding how income polarization should be measured, one has many options for approaching such research from different angles.

Whereas the research herein examined income polarization with a definition of income based on “household” income, one can examine income polarization using different income measures to assess model robustness. The second chapter and Appendix III have a limited examination of “individual” income data. Results from individual income were not mapped or examined further. Other definitions of income could include “family” income. Additional measures of income that can be considered are pre-tax and post-tax income. As post-tax income started to be collected by the Canadian census starting in 2006, such analysis might not be limited at the present time. However, as time goes on the possibility to obtain additional data to make a meaningful analysis on trends since 2006 with post-tax data appears.

Data from each census that was used in the second chapter for bootstrapping has a substantial impact on results due to resampling. It is not clear if the results from a particular census are outliers or by chance occurrences (data from 2020 or 2021 would for example be an anomaly due to SARS COVID-19, were such data used). Hence, additional data can be obtained from revenue Canada for each year beyond those when the census was held in order to discern precisely when income polarization was occurring. Honing in on income polarization with precise dates can be insightful to uncover impacts of other processes on income polarization. The difficulty with such endeavor is that it would be costly. However, having such data would be

interesting because it would allow for one to discern natural breaks in income trends over time and opens up the possibility to identify outlier years, should there be any.

In the second chapter low-income areas were delineated, and they may be areas of concern. While low-income areas might not be poverty areas per se, it is fair to assume that such areas face more hardships than middle- and high-income areas. They are therefore more in need than other areas. Hence, analysis can be conducted which examines such spaces individually. There is also an opportunity to compare patterns of income polarization with patterns of gentrification in order to ascertain to what extent does gentrification impact income polarization or vice-versa. Finally, the second chapter also serves as a basis for examining the lived experiences of individuals who inhabit the divided urban landscape.

The third chapter provided state-of-the-art methods for identifying gentrification. However, as mentioned earlier, Google's current terms of use for Street View data make repeating such endeavors prohibitively expensive. One may have an opportunity to retrieve image data from other sources. Some of the largest ones are Mapillary (Global crowd sourced), Streetside (from Bing Maps) (USA mainly), Yandex Maps (in the former Soviet Union), and Tencent maps (in China). However, numerous other sources exist in select countries. A limitation of these other sources is their imagery's temporal dimension. The timeline view that GSV provides is, at the moment, superior to these other sources (in urban contexts of the US and Canada at least) due to the availability of locational imagery across many years.

Researchers also have the possibility of creating their own streetscape data using 360-degree cameras, as the cost of such equipment has declined in recent years. Such approach can provide imagery in the winter, a time of year which remains understudied in terms of urban visual spaces. Producing own data has the potential to have a fuller data set than is on GSV, as in

certain urban areas some streets are not displayed due to no imagery or due to owners who used their right to obfuscate their properties. One of the biggest advantages of GSV is the temporal record of photos for each location. However, some countries restrict what Google can collect. For example, GSV data in Germany is limited to only the largest cities and has no data in small cities, towns, and rural places. Generally, in terms of location, even in Canada, GSV has very little street imagery in rural areas. Hence, self-produced imagery in rural areas may yield insight into rural gentrification but would require a long-term commitment beyond the scope of most research studies. Finally, street imagery can be used to detect areas undergoing rapid urban decline, though such scenarios are more pronounced in certain US urban contexts (e.g., the greening of Detroit).

The fourth chapter provided insight into examining gentrification using census data. An impactful way that the research can be further expounded would be through the development of a more accurate dependent variable for regression analysis (e.g., one that is not dependent on the outward visual signs of property improvement). One means of doing so would be through an examination of building permit data. Obtaining such data could be tricky, in Ottawa, as digitized building permit data is available starting in only 2011. Such endeavors are clearly time intensive but can provide interesting means of further analysis.

The fourth chapter is nevertheless a steppingstone for further examination of gentrification using both census data and quantitative means in general. For example, data to examine gentrification at fine geographic units of analysis can be obtained from less conventional sources such as real estate institutions (Holm and Schulz 2018).

Lastly, the fourth chapter is a sort of encouragement to researchers to use data on a ratio-interval scale. This scale has a higher level of complexity and precision, than for example

categorical or nominal data, and the research herein paves the way for further such studies in the future whereby a common workflow to modelling or detecting gentrification becomes as close as one can get to a common definition of the phenomenon.

### **Closing Remarks**

This thesis examines urban divides and inequality in Canada's largest CMAs generally and in Ottawa locally. New methods and techniques have been successfully established and in doing so both urban geography and geomatics have received unique contributions.

Firstly, it was established that income polarization has taken place in the eight largest CMAs in Canada between 1971 and 2016. Secondly, street level imagery has been used to measure gentrification between 2007 and 2016 in Ottawa, using artificial intelligence. Thirdly, and lastly, quantitative models for measuring gentrification (in Ottawa) were compared at various geographic units of analysis, and new models for examining gentrification were developed.

This research provides novel quantitative methods for assessing urban inequalities over time. Each major contributing chapter both sets a benchmark and is a steppingstone for further research. In a sense this research ostensibly leaves as many questions as it provides answers, for it poses a path for examination of urban inequalities in other locations and spatial scales.

## Chapter 6. References

- Ades, Josefina, Philippe Apparicio, and Anne-Marie Séguin. 2012. "Are New Patterns of Low-Income Distribution Emerging in Canadian Metropolitan Areas?" *Canadian Geographer* 56 (3): 339–361.
- Albet, Abel, and Núria Benach, eds. 2017. *Gentrification as a Global Strategy: Neil Smith and Beyond*. Abingdon: Routledge.
- Alkon, Alison Hope, Yiki Kato, and Joshua Sbicca, eds. 2020. *A Recipe for Gentrification: Food, Power, and Resistance in the City*. New York: New York University Press.
- Anderson, Ayoka, James Platt, Ian Skelton, Geoff Butler, Paul Chorney, Eric Funk, and Brian Grant. 2005. "Social Investment in the Inner City: Community and Capital in West Broadway, Winnipeg." *Canadian Journal of Urban Research* 14 (1): 8–31.
- Angel, Shlomo, Jason Parent, and Daniel L. Civco. 2012. "The Fragmentation of Urban Landscapes: Global Evidence of a Key Attribute of the Spatial Structure of Cities, 1990–2000." *Environment and Urbanization* 24 (1): 249–283.
- Anguelovski, Isabelle, Margarita Triguero-Mas, James JT Connolly, Panagiota Kotsila, Galia Shokry, Carmen Pérez Del Pulgar, Melissa Garcia-Lamarca, Lucia Argüelles, Julia Mangione, Kaitlyn Dietz, and Helen Cole. 2020. "Gentrification and health in two global cities: a call to identify impacts for socially-vulnerable residents." *Cities & Health* 4 (1): 40–49.
- Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Springer Science+Business Media, Kluwer Academic Publishers.
- Anselin, Luc, Anil K. Bera, Raymond Florax, and Mann J. Yoon. 1996. "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics* 26 (1): 77–104.
- Atkinson, Rowland, and Gary Bridge, eds. 2005a. *Gentrification in a Global Context*. London: Routledge.
- Atkinson, Rowland, and Gary Bridge, eds. 2005b. "Introduction." In *Gentrification in a Global Context*, edited by Rowland Atkinson and Gary Bridge, 1–17. London: Routledge.
- Bacchi, Umberto. 2019a. "New driveway? Home improvements lead AI to hidden gentrification." *Thomson Reuters Foundation*, 13 March 2019. <https://news.trust.org/item/20190313174839-fb4kz/>.
- Bacchi, Umberto. 2019b. "New driveway? Home improvements lead AI to hidden gentrification." *Reuters*, March 13, 2019. <https://www.reuters.com/article/us-global-cities-gentrification-idUSKCN1QU2JI>.

- Bacchi, Umberto. 2019c. "New driveway? Home improvements lead AI to hidden gentrification." *The Peninsula*, March 13, 2019. <https://thepeninsulaqatar.com/article/14/030/2019/New-driveway-Home-improvements-lead-AI-to-hidden-gentrification>.
- Baddeley, Adrian, Ege Rubak, and Rolf Turner. 2016. *Spatial Point Patterns: Methodology and Applications with R*. Champan & Hall/CRC Interdisciplinary Statistics Series. Boca Raton ; London ; New York: CRC Press, Taylor & Francis Group.
- Badland, Hannah M., Simon Opit, Karen Witten, Robin A. Kearns, and Suzanne Mavoa. 2010. "Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits?" *Journal of Urban Health* 87 (6): 1007–1016.
- Bagheri, Nasser, Philip J. Batterham, Luis Salvador-Carulla, Y. Chen, Andrew Page, Alison L. Calear, and Peter Congdon. 2019. "Australian neighborhood social fragmentation index and its association with spatial variation in depression across communities." *Social Psychiatry and Psychiatric Epidemiology* 54 (10): 1189–1198.
- Balakrishnan, Radhika, James Heintz, and Diane Elson. 2016. *Rethinking Economic Policy for Social Justice: The Radical Potential of Human Rights*. London: Routledge.
- Baldassare, Mark. 1984. "Evidence for Neighborhood Revitalization." In *Gentrification, Displacement and Neighborhood Revitalization*, edited by John Palen and Bruce London, 90–102. Albany: State University of New York Press.
- Banting, Keith. 1997. "The Social Policy Divide: The Welfare State in Canada and the United States." In *Degrees of Freedom: Canada and the United States in a Changing World*, edited by Keith Banting, George Hoberg, and Richard Simeon, 267–309. Montreal & Kingston: McGill-Queen's University Press.
- Banting, Keith, and John Myles. 2013. "Introduction: Inequality and the Fading of Redistributive Politics." In *Inequality and the Fading of Redistributive Politics*, edited by Keith Banting and John Myles, 1–39. Vancouver: University of British Columbia Press.
- Banzhaf, Spencer, Lala Ma, and Christopher Timmins. 2019. "Environmental Justice: The Economics of Race, Place, and Pollution." *Journal of Economic Perspectives* 33 (1): 185–208.
- Barrington, Martin J., and Brian W. Ilbery. 1987. "Farm Fragmentation: A Case Study from Coventry's Urban Fringe." *Geoforum* 18 (2): 237–245.
- Barton, Michael. 2016. "An Exploration of the Importance of the Strategy Used to Identify Gentrification." *Urban Studies* 53 (1): 92–111.
- Battle, Ken, Michael Mendelson, and Sherri Torjman. 2006. "Towards a New Architecture for Canada's Adult Benefits." Ottawa: The Caledon Institute of Social Policy. <https://maytree.com/wp-content/uploads/594ENG.pdf>.

- Beach, Charles M. 2016. "Changing Income Inequality: A Distributional Paradigm for Canada." *Canadian Journal of Economics* 49 (4): 1229–1292.
- Benali, Kenza. 2013. "La densification urbaine dans le quartier Vanier : germe d'un nouveau urbain ou menace pour le dernier îlot francophone de la capitale canadienne ?" *Cahiers de géographie du Québec* 57 (160): 41.
- Berg, Alex, Jia Deng, and Fei-Fei Li. 2010. "Large Scale Visual Recognition Challenge 2010 (ILSVRC2010)." <http://www.image-net.org/challenges/LSVRC/2010/>.
- Betancur, John J. 2002. "The Politics of Gentrification: The Case of West Town in Chicago." *Urban Affairs Review* 37 (6): 780–814.
- Bird, John, Barry Curtis, Tim Putnam, George Robertson, and Lisa Tickner, eds. 1993. *Mapping the Futures: Local Cultures, Global Change*. London: Routledge.
- Bivand, Roger S., and David W.S. Wong. 2018. "Comparing Implementations of Global and Local Indicators of Spatial Association." *Test* 27 (3): 716–748.
- Blečić, Ivan, Arnaldo Cecchini, and Giuseppe A. Trunfio. 2018. "Towards Automatic Assessment of Perceived Walkability." In *Computational Science and Its Applications – ICCSA 2018*, edited by Osvaldo Gervasi, Beniamino Murgante, Sanjay Misra, Elena Stankova, Carmelo M. Torre, Ana Maria A.C. Rocha, David Taniar, Bernady O. Apduhan, Eufemia Tarantino, and Yeonseung Ryu, 10962: 351–365. Cham: Springer International Publishing.
- Blewtt, Taylor. 2019. "Wellington West is facing a moment of reckoning as gentrification looms." *Ottawa Citizen*, May 09, 2019. <https://ottawacitizen.com/business/local-business/wellington-west-is-facing-a-moment-of-reckoning-as-gentrification-looms>.
- Bostic, Raphael W., and Richard W. Martin. 2003. "Black Home-Owners as a Gentrifying Force? Neighbourhood Dynamics in the Context of Minority Home-Ownership." *Urban Studies* 40 (12): 2427–2449.
- Bourne, Larry S, and David J Hulchanski. 2020. "Inequality and Neighbourhood Change: Context, Concept, and Process." In *Changing Neighbourhoods: Social and Spatial Polarization in Canadian Cities*, edited by Jill L Grant, Alan Walks, and Howard Ramos, 5–30. Vancouver: UBC Press.
- Bourne, L.S. 1993. "Close Together and Worlds Apart: An Analysis of Changes in the Ecology of Income in Canadian Cities." *Urban Studies* 30 (8): 1293–1317.
- Brackley, Chris, and Nick Walker. 2019. "Researchers are using AI to track gentrification across Ottawa." *Canadian Geographic*, December 3, 2019. Researchers are using AI to track gentrification across Ottawa. <https://canadiangeographic.ca/articles/researchers-are->

using-ai-to-track-gentrification-across-ottawa/.

- Breau, Sébastien. 2015. "Rising Inequality in Canada: A Regional Perspective." *Applied Geography* 61: 58–69.
- Brenner, Neil, and Nik Theodore. 2002. "Cities and the Geographies of 'Actually Existing Neoliberalism.'" *Antipode* 34 (3): 349–379.
- Breusch, T.S., and A.R. Pagan. 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* 47 (5): 1287–1294.
- Bridge, Gary, Tim Butler, and Loretta Lees, eds. 2012. *Mixed Communities: Gentrification by Stealth?* Bristol: Policy Press.
- Brown-Saracino, Japonica. 2009. *A Neighborhood That Never Changes: Gentrification, Social Preservation, and the Search for Authenticity*. Chicago: The University of Chicago Press.
- Brown-Saracino, Japonica, ed. 2010. *The Gentrification Reader*. New York: Routledge.
- Buitelaar, Edwin, Anet Weterings, and Roderik Ponds. 2018. *Cities, Economic Inequality and Justice: Reflections and Alternative Perspectives*. London: Routledge.
- Bunce, Susannah. 2018. *Sustainability Policy, Planning and Gentrification in Cities*. London: Routledge.
- Butler, Don. 2017. "Ottawa Is Changing before Our Eyes: As We Hit Canada's 150th, Where Is the Capital Going?" *Ottawa Citizen*, January 5, 2017. <https://ottawacitizen.com/news/local-news/ottawa-is-transforming-before-our-eyes-as-we-hit-canadas-150th-where-is-the-capital-going>.
- Butler, Tim. 1997. *Gentrification and the Middle Classes*. Aldershot: Ashgate.
- Butler, Tim, and Loretta Lees. 2006. "Super-Gentrification in Barnsbury, London: Globalization and Gentrifying Global Elites at the Neighbourhood Level." *Transactions of the Institute of British Geographers* 31 (4): 467–487.
- Cadwallader, Martin. 1985. *Analytical Urban Geography: Spatial Patterns and Theories*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Cadwallader, Martin. 1996. *Urban Geography: An Analytical Approach*. Upper Saddle River, NJ: Prentice Hall.
- Cain, Taylor. 2020. "A Place for 'Families Like Us': Reproducing Gentrification and Gentrifiers in Two Boston Neighborhoods." Ph.D. dissertation. Boston University.
- Caulfield, Jon. 1994. *City Form and Everyday Life: Toronto's Gentrification and Critical Social*

- Practice*. Toronto: University of Toronto Press.
- CBC. 2019. "U of O project using Google Street View to spot gentrification." *CBC News*, March 18, 2019. <https://www.cbc.ca/news/canada/ottawa/streetview-gentrification-study-university-ottawa-1.5055031>.
- Chapple, Karen. 2017. "Income Inequality and Urban Displacement: The New Gentrification." *New Labor Forum* 26 (1): 84–93.
- Chen, Wen-Hao, John Myles, and Garnett Picot. 2012. "Why Have Poorer Neighbourhoods Stagnated Economically While the Richer Have Flourished? Neighbourhood Income Inequality in Canadian Cities." *Urban Studies* 49 (4): 877–896.
- Chollet, François. 2018. *Deep Learning with Python*. Shelter Island, New York: Manning Publications Co.
- Chollet, François and others. 2015. Keras. <https://keras.io>.
- City of Ottawa. 2010. "Neighbourhood Plan Hintonburg and Mechanicsville." Ottawa. <http://ottawa.ca/calendar/ottawa/citycouncil/occ/2010/02-24/pec/10%20-%20ACS2010-ICS-CSS-0002%20DOCUMENT%20%20%20EN%20Vars.pdf>.
- City of Ottawa. 2019. "The City of Ottawa Has Reached One Million Residents!" June 14, 2019. <https://ottawa.ca/en/news/city-ottawa-has-reached-one-million-residents>.
- City of Ottawa. n.d. "Open Data Ottawa." <http://data.ottawa.ca/>.
- Clarke, Philippa, Jennifer Ailshire, Robert Melendez, Michael Bader, and Jeffrey Morenoff. 2010. "Using Google Earth to Conduct a Neighborhood Audit: Reliability of a Virtual Audit Instrument." *Health and Place* 16 (6): 1224–1229.
- Clay, Phillip. 1979. *Neighborhood Renewal: Middle-Class Resettlement and Incumbent Upgrading in American Neighborhoods*. Lanham: Lexington Books.
- Cliff, A D, and J K Ord. 1981. *Spatial Processes: Models and Applications*. London: Pion.
- Cooper, Brittney. 2014. "Hollywood's post-racial mirage: How pop culture got gentrified." *Salon*, March 25, 2014. [https://www.salon.com/2014/03/25/hollywoods\\_post\\_racial\\_mirage\\_how\\_pop\\_culture\\_got\\_gentrified/](https://www.salon.com/2014/03/25/hollywoods_post_racial_mirage_how_pop_culture_got_gentrified/).
- Crews, Kelley A., and Manuel F. Peralvo. 2008. "Segregation and Fragmentation: Extending Landscape Ecology and Pattern Metrics Analysis to Spatial Demography." *Population Research and Policy Review* 27 (1): 65–88.
- Curran, Winifred. 2004. "Gentrification and the Nature of Work: Exploring the Links in Williamsburg, Brooklyn." *Environment and Planning A* 36 (7): 1243–1258.

- Curran, Winifred. 2018. *Gender and Gentrification*. Abingdon: Routledge.
- Curran, Winifred, and Trina Hamilton. 2012. "Just Green Enough: Contesting Environmental Gentrification in Greenpoint, Brooklyn." *Local Environment* 17 (9): 1027–1042.
- Curran, Winifred, and Trina Hamilton, eds. 2018. *Just Green Enough: Urban Development and Environmental Gentrification*. Abingdon: Routledge.
- Davidson, Mark, and Loretta Lees. 2005. "New-Build 'Gentrification' and London's Riverside Renaissance." *Environment and Planning A* 37 (7): 1165–1190.
- Davidson, Mark, and Loretta Lees. 2010. "New-Build Gentrification: Its Histories, Trajectories, and Critical Geographies." *Population, Space and Place* 16 (5): 395–411.
- Davies, Wayne. 1984. *Factorial Ecology*. Aldershot: Gower Publishing Company Limited.
- Davis, Mike. 1990. *City of Quartz: Excavating the Future in Los Angeles*. London: Verso.
- Dear, Michael, and Steven Flusty. 1998. "Postmodern Urbanism." *Annals of the Association of American Geographers* 88 (1): 50–72.
- Delmelle, Elizabeth C. 2019. "The Increasing Sociospatial Fragmentation of Urban America." *Urban Science* 3 (1).
- Demetriou, Demetris, John Stillwell, and Linda See. 2013. "A New Methodology for Measuring Land Fragmentation." *Computers, Environment and Urban Systems* 39: 71–80.
- Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." In 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255.
- Dent, Borden D, Jeffrey S Torguson, and Thomas W Hodler. 2009. *Cartography: Thematic Map Design*. 6th ed. Boston: McGraw Hill.
- Desilver, Drew. 2015. "The Many Ways to Measure Economic Inequality." Fact Tank. Pew Research Center. <https://www.pewresearch.org/fact-tank/2015/09/22/the-many-ways-to-measure-economic-inequality/>.
- Development Finance International, and OXFAM. 2018. "The Commitment to Reducing Inequality Index 2018: A Global Ranking of Governments Based on What They Are Doing to Tackle the Gap between Rich and Poor." <https://oxfamlibrary.openrepository.com/bitstream/handle/10546/620553/rr-commitment-reducing-inequality-index-2018-091018-en.pdf>.
- Dinca-Panaitescu, Mihaela, and Alan Walks. 2015. "Income Inequality, Income Polarization, and

- Poverty: How Are They Different? How Are They Measured?" Toronto: Neighbourhood Change Research Partnership. <http://neighbourhoodchange.ca/documents/2015/12/inequality-polarization-poverty-definitions.pdf>.
- Ding, Lei, Jackelyn Hwang, and Eileen Divringi. 2016. "Gentrification and Residential Mobility in Philadelphia." *Regional Science and Urban Economics* 61: 38–51.
- Distasio, Jino, and Andrew Kaufman. 2015. "Neighbourhoods of Change." In *The Divided Prairie City: Income Inequality Among Winnipeg's Neighbourhoods, 1970–2010*, edited by Jino Distasio and Andrew Kaufman, 2–12. Winnipeg: Institute of Urban Studies, The University of Winnipeg.
- Donald, Betsy, Amy Glasmeier, Mia Gray, and Linda Lobao. 2014. "Austerity in the City: Economic Crisis and Urban Service Decline?" *Cambridge Journal of Regions, Economy and Society* 7 (1): 3–15.
- Dorling, Danny. 2005. "Counting and Measuring." In *Questioning Geography*, edited by Noel Castree, Alisdair Rogers, and Douglas Sherman, 241–257. Malden: Blackwell Publishing.
- Dubey, Abhimanyu, Nikhil Naik, Devi Parikh, Ramesh Raskar, and César A. Hidalgo. 2016. *Deep Learning the City: Quantifying Urban Perception at a Global Scale*. Vol. 9905 LNCS.
- Eckstein, Susan Eva, and Timothy P. Wickham-Crowley. 2003. "Preface." In *What Justice? Whose Justice? Fighting for Fairness in Latin America*, edited by Susan Eva Eckstein and Timothy P. Wickham-Crowley, xi–xiv. Berkeley: University of California Press.
- Economist, The. 2017a. "How to Bring Cities Back from the Brink." May 4, 2017. <https://www.economist.com/news/books-and-arts/21721627-age-spectacle-story-revival-how-bring-cities-back-brink?fsrc=scn/tw/te/bl/ed/architectureandthe21stcenturycityhowtobringcitiesbackfromthebrink>.
- Economist, The. 2017b. "Gentrification: 'The Most Significant Force in Western Cities in the Second Half of the 20th Century'?" [Http://Econ.St/2pcY0WI](http://Econ.St/2pcY0WI)." Tweet. @theeconomist (blog). May 6, 2017. <https://twitter.com/theeconomist/status/860310125931843585>.
- Economist, The. 2018. "Not Finished: The Lapsing of Finland's Universal Basic Income Trial." *Economist* (United Kingdom) 414 (9086).
- Education News Canada. 2019. "Tracking urban gentrification, one building at a time." *Education News Canada*, March 19, 2019. <https://educationnewscanada.com/article/education/level/university/1/755761/tracking-urban-gentrification-one-building-at-a-time.html>.
- Esteban, Joan-Maria, and Debraj Ray. 1994. "On The Measurement of Polarization."

- Econometrica* 62 (4): 819–851.
- Evans, Joshua, Damian Collins, and Cher-Ann Chai. 2019. “On Thin Ice: Assembling a Resilient Service Hub.” *Area* 51 (3): 451–460.
- Fainstein, Susan, Ian Gordon, and Michael Harloe, eds. 1992. *Divided Cities: New York & London in the Contemporary World*. Oxford: Blackwell.
- Filion, Pierre. 1991. “The Gentrification-Social Structure Dialectic: A Toronto Case Study.” *International Journal of Urban and Regional Research* 15 (4): 553–574.
- Firth, Caislin L, Benoit Thierry, Daniel Fuller, Meghan Winters, and Yan Kestens. 2021. “Gentrification, Urban Interventions and Equity (GENUINE): A Map-Based Gentrification Tool for Canadian Metropolitan Areas.” *Statistics Canada, Catalogue No. 82-003-X, Health Reports*, 32 (5): 15–28.
- Flanagan, Elizabeth, Ugo Lachapelle, and Ahmed El-Geneidy. 2016. “Riding Tandem: Does Cycling Infrastructure Investment Mirror Gentrification and Privilege in Portland, OR and Chicago, IL?” *Research in Transportation Economics* 60: 14–24.
- Fong, Francis. 2017. *Income Inequality in Canada: The Urban Gap*. Toronto: Chartered Professional Accountants Canada.
- Fortin, Nicole, David A. Green, Thomas Lemieux, Kevin Milligan, and W. Craig Riddell. 2012. “Canadian Inequality: Recent Developments and Policy Options.” *Canadian Public Policy* 38 (2): 121–145.
- Fotheringham, A.S., and D.W.S. Wong. 1991. “The Modifiable Areal Unit Problem in Multivariate Statistical Analysis.” *Environment & Planning A* 23 (7): 1025–1044.
- Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models*. 3rd ed. Los Angeles: SAGE.
- Fox, John, and Georges Monette. 1992. “Generalized Collinearity Diagnostics.” *Journal of the American Statistical Association* 87 (417): 178–183.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. 3rd ed. Los Angeles: SAGE.
- Freeman, Lance. 2005. “Displacement or Succession? Residential Mobility in Gentrifying Neighborhoods.” *Urban Affairs Review* 40 (4): 463–491.
- Gagnon, Charles-Antoine. 2019. “Les quartiers urbains vus au fil des ans.” le Droit, March 13, 2019. <https://www.ledroit.com/2019/03/14/les-quartiers-urbains-vus-au-fil-des-ans-62182bb3eb93d5ee88f58d0ab8a09974/>.

- Gale, D. 1985. "Demographic Research on Gentrification and Displacement." *Journal of Planning Literature* 1 (1): 14–29.
- Gale, Dennis E. 2021. *The Misunderstood History of Gentrification: People, Planning, Preservation, and Urban Renewal, 1915-2020*. Philadelphia: Temple University Press.
- Gastner, Michael T., and M.E.J. Newman. 2004. "Diffusion-Based Method for Producing Density-Equalizing Maps." *Proceedings of the National Academy of Sciences of the United States of America* 101 (20): 7499–7504.
- Gebru, Timnit, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. 2017. "Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods across the United States." *Proceedings of the National Academy of Sciences of the United States of America* 114 (50): 13108–13113.
- Gibbons, Joseph, Michael Barton, and Elizabeth Brault. 2018. "Evaluating gentrification's relation to neighborhood and city health." *PLoS ONE*, 13(11).
- Glaeser, Edward, Hyunjin Kim, and Michael Luca. 2018. "Measuring Gentrification: Using Yelp Data to Quantify Neighborhood Change." Working Paper 24952. Cambridge: National Bureau of Economic Research. <http://www.nber.org/papers/w24952>.
- Glass, Ruth. 1964. "Introduction: Aspects of Change." In *London: Aspects of Change*, edited by Ruth Glass, E.J. Hobsbawm, Harold Pollins, W. Ashworth, J.H. Westergaard, William Holford, Margot Jefferys, John A. Jackson, and Sheila Patterson, xiii–xlii. London: MacGibbon & Kee.
- Glover, Ellen. 2019. "Gentrification hot spots." Science Node, April 24, 2019. <https://sciencenode.org/feature/Gentrification%20from%20the%20Ground.php>.
- Goerg, Georg M. 2011. "Lambert W Random Variables-a New Family of Generalized Skewed Distributions with Applications to Risk Estimation." *Annals of Applied Statistics* 5 (3): 2197–2230.
- Goerg, Georg M. 2015. "The Lambert Way to Gaussianize Heavy-Tailed Data with the Inverse of Tukey's h Transformation as a Special Case." *The Scientific World Journal*.
- Goerg, Georg M. 2022. "LambertW: An R Package for Lambert W x F Random Variables."
- Goldberg, Michael, and John Mercer. 1986. *The Myth of the North American City: Continentalism Challenged*. Vancouver: University of British Columbia Press.
- Goodsell, Todd L. 2013. "Familification: Family, Neighborhood Change, and Housing Policy." *Housing Studies* 28 (6): 845–868.
- Gould, Kenneth Alan, and Tammy L. Lewis. 2017. *Green Gentrification: Urban Sustainability*

- and the Struggle for Environmental Justice*. Abingdon: Routledge.
- Grant, Jill L, Alan Walks, and Howard Ramos, eds. 2020. *Changing Neighbourhoods: Social and Spatial Polarization in Canadian Cities*. Vancouver: University of British Columbia Press.
- Grant, Jill L, Alan Walks, and Howard Ramos. 2020. "Evaluating Neighbourhood Inequality and Change: Lessons from a National Comparison." In *Changing Neighbourhoods: Social and Spatial Polarization in Canadian Cities*, edited by Jill L Grant, Alan Walks, and Howard Ramos, 252–279. Vancouver: UBC Press.
- Gravari-Barbas, Maria, and Guinand Sandra, eds. 2017. *Tourism and Gentrification in Contemporary Metropolises: International Perspectives*. London: Routledge.
- Green, David A., and Benjamin M. Sand. 2015. "Has the Canadian Labour Market Polarized?" *Canadian Journal of Economics* 48 (2): 612–646.
- Green, David, and Jonathan Kesselman. 2006. "Dimensions of Inequality in a Just Society." In *Dimensions of Inequality in Canada*, edited by David Green and Jonathan Kesselman, 1–32. Vancouver: University of British Columbia Press.
- Green, David, W Craig Ridell, and France St-Hilaire. 2016. "Income Inequality in Canada: Driving Forces, Outcomes and Policy." In *Income Inequality: The Canadian Story*, edited by David A. Green, W. Craig Riddell, and France St-Hilaire, 1–73. Montreal: The Institute for Research on Public Policy.
- Griew, P., M. Hillsdon, C. Foster, E. Coombes, A. Jones, and P. Wilkinson. 2013. "Developing and Testing a Street Audit Tool Using Google Street View to Measure Environmental Supportiveness for Physical Activity." *International Journal of Behavioral Nutrition and Physical Activity* 10.
- Grube-Cavers, Annelise, and Zachary Patterson. 2015. "Urban Rapid Rail Transit and Gentrification in Canadian Urban Centres: A Survival Analysis Approach." *Urban Studies* 52 (1): 178–194.
- Guan, Dabo. 2017. "An Index of Inequality in China." *Nature Energy* 2 (10): 774–775.
- Guzman, Carolina, Rajiv Bhatia and Chris Durazo. 2005. "Anticipated Effects of Residential Displacement on Health: Results from Qualitative Research." Pew Charitable Trusts. <https://www.pewtrusts.org/~media/assets/2005/hiareporttrinityplazahousingredevelopment.pdf>.
- Hackworth, Jason. 2002. "Postrecession Gentrification in New York City." *Urban Affairs Review* 37 (6): 815–843.
- Hackworth, Jason. 2005. "Emergent Urban Forms, or Emergent Post-Modernisms? A

- Comparison of Large U.S. Metropolitan Areas.” *Urban Geography* 26 (6): 484–519.
- Hamel, Pierre, and Grégoire Autin. 2017. “Austerity Governance and the Welfare Crisis in Montreal.” *Alternate Routes: A Journal of Critical Social Research* 28: 165–188.
- Hamilton, Leah, and James P. Mulvale. 2019. “‘Human Again’: The (Unrealized) Promise of Basic Income in Ontario.” *Journal of Poverty* 23 (7): 576–599.
- Hammel, Daniel J., and Elvin K. Wyly. 1996. “A Model for Identifying Gentrified Areas with Census Data.” *Urban Geography* 17 (3): 248–268.
- Hamnett, Chris. 1994. “Socio-Economic Change in London: Professionalization Not Polarization.” *Built Environment* 20 (3): 192–203.
- Hamnett, Chris. 2003. *Unequal City: London in the Global Arena*. London: Routledge.
- Harris, Richard. 2020. “Plus Ça Change: Neighbourhood Inequality in Canadian Cities since 1900.” In *Changing Neighbourhoods: Social and Spatial Polarization in Canadian Cities*, edited by Jill L Grant, Alan Walks, and Howard Ramos, 31–52. Vancouver: UBC Press.
- Harris, Richard, Jim Dunn, and Sarah Wakefield. 2015. “A City on the Cusp: Neighbourhood Change in Hamilton since 1970.” Research Paper 236. Toronto: Neighbourhood Change Research Partnership. <http://neighbourhoodchange.ca/documents/2015/04/neighbourhood-change-in-hamilton-since-1970.pdf>.
- Harvey, David. 1989. “From Managerialism to Entrepreneurialism: The Transformation in Urban Governance in Late Capitalism.” *Geografiska Annaler: Series B, Human Geography* 71 (1): 3–17.
- Hay, Alan M. 1995. “Concepts of Equity, Fairness and Justice in Geographical Studies.” *Transactions of the Institute of British Geographers* 20 (4): 500–508.
- He, Shenjing. 2010. “New-Build Gentrification in Central Shanghai: Demographic Changes and Socioeconomic Implications.” *Population, Space and Place* 16 (5): 345–361.
- Heidkamp, C. Patrick, and Susan Lucas. 2006a. “Finding the Gentrification Frontier Using Census Data: The Case of Portland, Maine.” *Urban Geography* 27 (2): 101–125.
- Helms, Gesa, Rowland Atkinson, and Gordon MacLeod. 2007. “Editorial: Securing the City: Urban Renaissance, Policing and Social Regulation.” *European Urban and Regional Studies* 14 (4): 267–276.
- Heuton, Robert A, and Brienne Girard. 2010. “Fiscal Austerity and Urban Innovation: The

- Challenges Facing Canadian Cities.” Paper presented at the Canadian Political Science Association Annual Meeting, June 2, Montreal. <https://www.cpsa-acsp.ca/papers-2010/Heuton.pdf>.
- Hodge, David C. 1980. “Inner-City Revitalization as a Challenge to Diversity?: Seattle.” In *Back to the City: Issues in Neighborhood Renovation*, edited by Shirley Bradway Laska and Daphne Spain, 187–203. New York: Pergamon Press.
- Holifield, R. 2001. “Defining Environmental Justice and Environmental Racism.” *Urban Geography* 22 (1): 78–90.
- Holm, Andrej, and Guido Schulz. 2018. “GentriMap: A Model for Measuring Gentrification and Displacement.” In *Gentrification and Resistance: Researching Displacement Processes and Adaptation Strategies*, edited by Ilse Helbrecht, 251–277. Wiesbaden: Springer VS.
- Hoyt, Homer. 1939. *The Structure And Growth Of Residential Neighborhoods In American Cities*. Washington, D.C.: Federal Housing Administration.
- Hubbard, Phil. 2017. *The Battle for the High Street: Retail Gentrification, Class and Disgust*. London: Palgrave Macmillan.
- Hulchanski, David. 2007. “The Three Cities within Toronto: Income Polarization among Toronto’s Neighbourhoods, 1970 to 2000.” 41. Research Bulletin. Toronto: University of Toronto Centre for Urban & Community Studies.
- Hulchanski, David. 2010. “The Three Cities Within Toronto: Income Polarization Among Toronto’s Neighbourhoods, 1970–2005.” Toronto: University of Toronto Cities Centre.
- Hulchanski, David, and Scott Leckie. 2000. “The Human Right to Adequate Housing - 1945-1999 - Chronology of United Nations Activity.” Geneva: Centre on Housing Rights and Evictions.
- Huse, Tone. 2014. *Everyday Life in the Gentrifying City: On Displacement, Ethnic Privileging, and the Right to Stay Put*. Surrey: Ashgate.
- Hwang, Jackelyn, and Robert J. Sampson. 2014. “Divergent Pathways of Gentrification: Racial Inequality and the Social Order of Renewal in Chicago Neighborhoods.” *American Sociological Review* 79 (4): 726–751.
- Igbozurike, Matthais Uzo. 1974. “Fragmentation In Tropical Agriculture: An Overrated Phenomenon.” *Area* 6 (2): 321–325.
- Ilic, Lazar, M. Sawada, and Amaury Zarzelli. 2019. “Deep Mapping Gentrification in a Large Canadian City Using Deep Learning and Google Street View.” *PLoS ONE* 14 (3).
- Ilic, Lazar, and M. Sawada. 2021. “The Temporal Evolution of Income Polarization in Canada’s

- Largest CMAs.” *PLoS ONE* 16 (6).
- Israel, Emil, and Amnon Frenkel. 2018. “Social Justice and Spatial Inequality: Toward a Conceptual Framework.” *Progress in Human Geography* 42 (5): 647–665.
- Jager, Michael. 1986. “Class Definition and the Esthetics of Gentrification: Victoriana in Melbourne.” In *The Gentrification of the City*, edited by Neil Smith and Peter Williams, 78–91. Boston: Allen & Unwin.
- Januszewski, Józef. 1968. “Index of Land Consolidation as a Criterion of the Degree of Concentration.” *Geographia Polonica* 14: 291–296.
- Jargowsky, Paul A. 1997. *Poverty and Place: Ghettos, Barrios, and the American City*. New York: Russel Sage Foundation.
- Johnsson, Katarina. 1995. “Fragmentation Index as a Region Based GIS Operator.” *International Journal of Geographical Information Systems* 9 (2): 211–220.
- Josset, Christophe. 2019. “La gentrification des quartiers mieux comprise grâce aux algorithmes.” *L'Express*, March 13, 2019. [https://www.lexpress.fr/sciences-sante/sciences/la-gentrification-des-quartiers-mieux-comprise-grace-aux-algorithmes\\_2066711.html](https://www.lexpress.fr/sciences-sante/sciences/la-gentrification-des-quartiers-mieux-comprise-grace-aux-algorithmes_2066711.html).
- Kang, Jian, Marco Körner, Yuanyuan Wang, Hannes Taubenböck, and Xiao Xiang Zhu. 2018. “Building Instance Classification Using Street View Images.” *ISPRS Journal of Photogrammetry and Remote Sensing* 145 (A): 44–59.
- Karsten, Lia. 2003. “Family Gentrifiers: Challenging the City as a Place Simultaneously to Build a Career and to Raise Children.” *Urban Studies* 40 (12): 2573–2584.
- Karsten, Lia. 2014. “From Yuppies to Yupps: Family Gentrifiers Consuming Spaces and Re-Inventing Cities.” *Tijdschrift Voor Economische En Sociale Geografie* 105 (2): 175–188.
- Kary, Kenneth J. 1988. “The Gentrification of Toronto and the Rent Gap Theory.” In *The Changing Canadian Inner City*, edited by Trudi E. Bunting and Pierre Filion, 53–72. Waterloo: Department of Geography.
- Kelly, Cheryl M., Jeffrey S. Wilson, Elizabeth A. Baker, Douglas K. Miller, and Mario Schootman. 2013. “Using Google Street View to Audit the Built Environment: Inter-Rater Reliability Results.” *Annals of Behavioral Medicine* 45 (SUPPL.1): S108–S112.
- Kochhar, Rakesh, Richard Fry, D’Vera Cohn, Kim Parker, Claudia Deane, Renee Stepler, Anna Brown, et al. 2015. “The American Middle Class Is Losing Ground: No Longer the Majority and Falling behind Financially.” *Social & Demographic Trends*. Washington D.C.: Pew Research Center. [https://www.pewresearch.org/wp-content/uploads/sites/3/2015/12/2015-12-09\\_middle-class\\_FINAL-report.pdf](https://www.pewresearch.org/wp-content/uploads/sites/3/2015/12/2015-12-09_middle-class_FINAL-report.pdf).

- Koenker, Roger. 1981. "A Note on Studentizing a Test for Heteroscedasticity." *Journal of Econometrics* 17 (1): 107–112.
- Krämer, Walter, and Herald Sonnberger. 1986. *The Linear Regression Model Under Test*. Wien: Physica-Verlag Heidelberg.
- Krase, Jerome, and Judith N. DeSena. 2016. *Race, Class, and Gentrification in Brooklyn: A View from the Street*. Lanham: Lexington Books.
- Kreide, Regina. 2022. "Global (in)Justice and the Human Right to Housing. A Practice-Based Approach." *Critical Review of International Social and Political Philosophy* 25 (1): 107–127.
- Lafontaine, Sean J.V., M. Sawada, and Elizabeth Kristjansson. 2017. "A Direct Observation Method for Auditing Large Urban Centers Using Stratified Sampling, Mobile GIS Technology and Virtual Environments." *International Journal of Health Geographics* 16 (1).
- Laska, Shirley Bradway, and Daphne Spain, eds. 1980. *Back to the City: Issues in Neighborhood Renovation*. New York: Pergamon Press.
- Law, Stephen, Yao Shen, and Chanuki Seresinhe. 2017. "An Application of Convolutional Neural Network in Street Image Classification: The Case Study of London." In *Proceedings of the 1st Workshop on GeoAI: AI and Deep Learning for Geographic Knowledge Discovery, GeoAI 2017*, 5–9.
- Lee, Barrett A., and Paula M. Mergenhagen. 1984. "Is Revitalization Detectable?: Evidence from Five Nashville Neighborhoods." *Urban Affairs Review* 19 (4): 511–538.
- Lee, Sang-Il. 2001. "Developing a Bivariate Spatial Association Measure: An Integration of Pearson's  $r$  and Moran's  $I$ ." *Journal of Geographical Systems* 3 (4): 369–385.
- Lee, Sang-Il. 2004. "A Generalized Significance Testing Method for Global Measures of Spatial Association: An Extension of the Mantel Test." *Environment and Planning A: Economy and Space* 36 (9): 1687–1703.
- Lees, Loretta. 2000. "A Reappraisal of Gentrification: Towards a 'Geography of Gentrification.'" *Progress in Human Geography* 24 (3): 389–408.
- Lees, Loretta. 2003. "Super-Gentrification: The Case of Brooklyn Heights, New York City." *Urban Studies* 40 (12): 2487–2509.
- Lees, Loretta, Hyun Bang Shin, and Ernesto López-Morales, eds. 2015. *Global Gentrifications: Uneven Development and Displacement*. Bristol: Policy Press.

- Lees, Loretta, Hyun Bang Shin, and Ernesto López Morales. 2016. *Planetary Gentrification*. Cambridge, UK: Polity Press.
- Lees, Loretta, Tom Slater, and Elvin Wyly. 2008a. *Gentrification*. New York: Routledge.
- Lees, Loretta, Tom Slater, and Elvin Wyly, eds. 2008b. *The Gentrification Reader*. London: Routledge.
- Leffers, Donald, and Patricia Ballamingie. 2013. "Governmentality, Environmental Subjectivity, and Urban Intensification." *Local Environment* 18 (2): 134–151.
- LeGates, Richard T., and Chester Hartman. 1986. "The Anatomy of Displacement in the US." In *The Gentrification of the City*, edited by Neil Smith and Peter Williams, 178–200. Boston: Allen & Unwin.
- Leijten, Ingrid, and Kaisa de Bel. 2020. "Facing Financialization in the Housing Sector: A Human Right to Adequate Housing for All." *Netherlands Quarterly of Human Rights* 38 (2): 94–114.
- LeSage, James, and Robert Kelly Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press, Taylor & Francis Group.
- LeSage, James P. 2008. "An Introduction to Spatial Econometrics." *Revue d'Economie Industrielle* 123 (3): 19–44.
- LeSage, James P., and Manfred M. Fischer. 2008. "Spatial Growth Regressions: Model Specification, Estimation and Interpretation." *Spatial Economic Analysis* 3 (3): 275–304.
- Letki, Natalia, and Inta Mieriņa. 2015. "Getting Support in Polarized Societies: Income, Social Networks, and Socioeconomic Context." *Social Science Research* 49 (1): 217–233.
- Ley, David. 1983. *A Social Geography of the City*. New York: Harper & Row.
- Ley, David. 1985. *Gentrification in Canadian Inner Cities: Patterns, Analysis, Impacts and Policy*. Ottawa: Canada Mortgage and Housing Corporation.
- Ley, David. 1986. "Alternative Explanations for Inner-City Gentrification: A Canadian Assessment." *Annals of the Association of American Geographers* 76 (4): 521–535.
- Ley, David. 1988. "Social Upgrading in Six Canadian Inner Cities." *Canadian Geographer* 32 (1): 31–45.
- Ley, David. 1992. "Gentrification in Recession: Social Change in Six Canadian Inner Cities, 1981-1986." *Urban Geography* 13 (3): 230–256.
- Ley, David. 1996a. *The New Middle Class and the Remaking of the Central City*. Oxford:

Oxford University Press.

- Ley, David. 1996b. "The New Middle Class in Canadian Central Cities." In *City Lives and City Forms: Critical Research & Canadian Urbanism*, edited by Jon Caulfield and Linda Peake, 15–32. Toronto: University of Toronto Press.
- Ley, David, and Nicholas Lynch. 2012. "Divisions and Disparities in Lotus-Land: Socio-Spatial Income Polarization in Greater Vancouver, 1970-2005." Research Paper No. 223. Toronto: Cities Centre, University of Toronto. <http://neighbourhoodchange.ca/documents/2012/10/divisions-and-disparities-in-lotus-land-socio-spatial-income-polarization-in-greater-vancouver-1970-2005-by-david-ley-nicholas-lynch.pdf>.
- Li, Peter S. 2000. "Earning Disparities between Immigrants and Native-Born Canadians." *Canadian Review of Sociology and Anthropology* 37 (3): 289–311.
- Lightbody, James, and David Walker. 1987. "Canadian Municipalities and Conditions of Fiscal Austerity in International Perspective: A Report and Some Observations." *Urban History Review* 15 (3): 274–278.
- Lim, Sungwoo, Pui Ying Chan, Sarah Walters, Gretchen Culp, M Mary Huynh, and L. Hannah Gould. 2017. "Impact of Residential Displacement on Healthcare Access and Mental Health among Original Residents of Gentrifying Neighborhoods in New York City." *PLoS ONE* 12 (12).
- Lingel, Jessa. 2021. *The Gentrification of the Internet: How to Reclaim Our Digital Freedom*. Oakland: University of California Press.
- Liu, Lun, Elisabete A. Silva, Chunyang Wu, and Hui Wang. 2017. "A Machine Learning-Based Method for the Large-Scale Evaluation of the Qualities of the Urban Environment." *Computers, Environment and Urban Systems* 65: 113–125.
- Logan, Jennifer, and Marc Vachon. 2009. "Gentrification and Rental Management Agencies: West Broadway Neighbourhood in Winnipeg." *Canadian Journal of Urban Research* 17 (2): 84–104.
- Logan, John, Peter Taylor-Gooby, and Monika Reuter. 1992. "Poverty and Income Inequality." In *Divided Cities: New York & London in the Contemporary World*, edited by Susan Fainstein, Ian Gordon, and Michael Harloe, 129–150. Oxford: Blackwell.
- London, Bruce, and John Palen. 1984. "Introduction: Some Theoretical and Practical Issues Regarding Inner City Revitalization." In, *Gentrification, Displacement and Neighborhood Revitalization*, edited by J John Palen and Bruce London, 4–26. Albany: State University of New York Press.
- Loukaitou-Sideris, Anastasia, Silvia Gonzalez, and Paul Ong. 2019. "Triangulating Neighborhood Knowledge to Understand Neighborhood Change: Methods to Study

- Gentrification.” *Journal of Planning Education and Research* 39 (2): 227–242.
- MacLachlan, Ian, and Ryo Sawada. 1997. “Measures of Income Inequality and Social Polarization in Canadian Metropolitan Areas.” *Canadian Geographer* 41 (4): 377–397.
- MacMillan, Blake. 2011. “ANT in Westboro: An Examination of the Applicability of Actor-Network Theory in Gentrification Studies.” Ottawa: Library and Archives Canada = Bibliothèque et Archives Canada.
- Madariaga, Rafa, Joan Carles Martori, and Ramon Oller. 2019. “Wage Income Inequality in Catalonian Second-Rank Cities.” *Annals of Regional Science* 62 (2): 285–304.
- Magnani, Riccardo, and Luca Piccoli. 2020. “Universal Basic Income with Flat Tax Reform in France.” *Journal of Policy Modeling* 42 (2): 235–249.
- Mallach, Alan. 2018. *The Divided City : Poverty and Prosperity in Urban America*. Washington, DC: Island Press.
- Marcuse, Peter. 1986. “Abandonment, Gentrification, and Displacement: The Linkages in New York City.” In *Gentrification of the City*, edited by Neil Smith and Peter Williams, 153–177. Boston: Allen & Unwin.
- Marcuse, Peter. 1989. “‘Dual City’: A Muddy Metaphor for a Quartered City.” *International Journal of Urban and Regional Research* 13 (4): 697–708.
- Mario. 2019. “Google Street View permet aussi de suivre la gentrification des quartiers.” *FZN / Fredzone*, March 24, 2019. <https://www.fredzone.org/google-street-view-permet-aussi-de-suivre-la-gentrification-des-quartiers-887>.
- Marris, Peter. 2015. *Loss and Change, Revised Edition*. London: Routledge.
- Maslow, A.H. 1943. “A Theory of Human Motivation.” *Psychological Review* 50 (4): 370–396.
- Massey, Doreen. 2005. *For Space*. London: SAGE Publications Ltd.
- Massey, Douglas S., and Nancy A. Denton. 1988. “The Dimensions of Residential Segregation.” *Social Forces* 67 (2): 281–315.
- McDonald, Scott C. 1986. “Does Gentrification Affect Crime Rates?” *Crime and Justice* 8: 163–201.
- McKay, Andrew. 2002. “Defining and Measuring Inequality.” Briefing Paper No.1 (1 of 3). Inequality Briefing. Overseas Development Institute and University of Nottingham. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3804.pdf>.
- McKenzie, Brian S. 2014. “Access to Supermarkets among Poorer Neighborhoods: A

- Comparison of Time and Distance Measures.” *Urban Geography* 35 (1): 133–151.
- McKinnish, Terra, Randal Walsh, and T. Kirk White. 2010. “Who Gentrifies Low-Income Neighborhoods?” *Journal of Urban Economics* 67 (2): 180–193.
- Meligrana, John, and Andrejs Skaburskis. 2005. “Extent, Location and Profiles of Continuing Gentrification in Canadian Metropolitan Areas, 1981-2001.” *Urban Studies* 42 (9): 1569–1592.
- Mennis, Jeremy. 2009. “Dasymetric Mapping for Estimating Population in Small Areas.” *Geography Compass* 3 (2): 727–745.
- Mollenkopf, John Hull, and Manuel Castells, eds. 1991. *Dual City: Restructuring New York*. New York: Russel Sage Foundation.
- Mondor, Luke, Deborah Cohen, Anum Irfan Khan, and Walter P. Wodchis. 2018. “Income Inequalities in Multimorbidity Prevalence in Ontario, Canada: A Decomposition Analysis of Linked Survey and Health Administrative Data.” *International Journal for Equity in Health* 17 (1).
- Moritz, Steffen, Alexis Sardá, Thomas Bartz-Beielstein, Martin Zaefferer, and Jörg Stork. 2015. “Comparison of Different Methods for Univariate Time Series Imputation in R.” Research Paper. Cologne University of Applied Sciences. <https://arxiv.org/ftp/arxiv/papers/1510/1510.03924.pdf>.
- Moskowitz, Peter. 2017. *How to Kill a City: Gentrification, Inequality, and the Fight for the Neighborhood*. New York: Nation Books.
- Muniz, Vicky. 1998. *Resisting Gentrification & Displacement: Voices of Puerto Rican Women of the Barrio*. New York: Garland Publishing.
- Murdie, Robert. 2008. “Diversity and Concentration in Canadian Immigration: Trends in Toronto, Montréal and Vancouver, 1971–2006.” 42. Research Bulletin. Toronto: University of Toronto Centre for Urban & Community Studies.
- Myles, John. 1998. “How to Design a ‘liberal’ Welfare State: A Comparison of Canada and the United States.” *Social Policy and Administration* 32 (4): 341–364.
- Myles, John, Garnett Picot, and Wendy Pyper. 2000. “Neighbourhood Inequality in Canadian Cities.” 160. Analytical Studies Branch Research Paper Series. Ottawa: Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2000160-eng.pdf>.
- Naik, Nikhil, Scott Duke Kominers, Ramesh Raskar, Edward L Glaeser, and César Hidalgo. 2015. “Do People Shape Cities, or Do Cities Shape People? The Co-Evolution of Physical, Social, and Economic Change in Five Major U.S. Cities.” HKS Working Paper No. 061. <https://ssrn.com/abstract=2698292>.

- Naik, Nikhil, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser, and César A. Hidalgo. 2017. "Computer Vision Uncovers Predictors of Physical Urban Change." *Proceedings of the National Academy of Sciences of the United States of America* 114 (29): 7571–7576.
- Naik, Nikhil, Jade Philipoom, Jade Philipoom, Ramesh Raskar and César Hidalgo. 2014. "Streetscore-Predicting the Perceived Safety of One Million Streetscapes." In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 793–799.
- Nelson, Kathryn P. 1988. *Gentrification and Distressed Cities: An Assessment of Trends in Intrametropolitan Migration*. Madison: The University of Wisconsin Press.
- Newman, Kathe, and Philip Ashton. 2004. "Neoliberal Urban Policy and New Paths of Neighborhood Change in the American Inner City." *Environment and Planning A* 36 (7): 1151–1172.
- Nijman, Jan, and Yehua Dennis Wei. 2020. "Urban Inequalities in the 21st Century Economy." *Applied Geography* 117.
- Odgers, Candice L., Avshalom Caspi, Christopher J. Bates, Robert J. Sampson, and Terrie E. Moffitt. 2012. "Systematic Social Observation of Children's Neighborhoods Using Google Street View: A Reliable and Cost-Effective Method." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 53 (10): 1009–1017.
- Openshaw, Stan. 1983. *The Modifiable Areal Unit Problem*. Concepts and Techniques in Modern Geography 38. Norwich: Geo Books.
- Osberg, Lars. 2008. *Quarter Century of Economic Inequality in Canada: 1981-2006*. Toronto: Canadian Centre for Policy Alternatives.
- Osberg, Lars. 2016. "What's So Bad about Increasing Inequality in Canada?" In *Income Inequality: The Canadian Story*, edited by David A. Green, W. Craig Riddell, and France St-Hilaire, 299–343. Montreal: The Institute for Research on Public Policy.
- Pahl, R.E. 1988. "Some Remarks on Informal Work, Social Polarization and the Social Structure." *International Journal of Urban and Regional Research* 12 (2): 247–267.
- Palen, John, and Bruce London, eds. 1984. *Gentrification, Displacement and Neighborhood Revitalization*. Albany: State University of New York Press.
- Parenteau, Marie-Pierre, and Michael C. Sawada. 2011. "The Modifiable Areal Unit Problem (MAUP) in the Relationship between Exposure to NO<sub>2</sub> and Respiratory Health." *International Journal of Health Geographics* 10.
- Paton, Kirsteen. 2014. *Gentrification: A Working-Class Perspective*. Surrey: Ashgate Publishing.

- Pavlic, Dejan, and Zhu Qian. 2014. "Declining Inner Suburbs? A Longitudinal-Spatial Analysis of Large Metropolitan Regions in Canada." *Urban Geography* 35 (3): 378–401.
- Peterson, Kyle. 2013. "Gentrification through Public Participation? Acceptance and Resistance in Calgary's Inner Suburbs." Ph.D. dissertation. University of Calgary.
- Phillips, Martin. 1993. "Rural Gentrification and the Processes of Class Colonisation." *Journal of Rural Studies* 9 (2): 123–140.
- Picot, Garnett, and Feng Hou. 2003. "The Rise in Low-Income Rates Among Immigrants in Canada." 11F0019MIE-No. 198. Analytical Studies Branch Research Paper Series. Ottawa: Statistics Canada. <https://www150.statcan.gc.ca/n1/en/catalogue/11F0019M2003198>.
- Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge: Harvard University Press.
- Ploeger, S.K., M. Sawada, A. Elsabbagh, M. Saatcioglu, M. Nastev, and E. Rosetti. 2016. "Urban RAT: New Tool for Virtual and Site-Specific Mobile Rapid Data Collection for Seismic Risk Assessment." *Journal of Computing in Civil Engineering* 30 (2).
- PLOS. 2019a. "Tracking urban gentrification, one building at a time." *EurekaAlert!*, March 13, 2019. <https://www.eurekaalert.org/news-releases/634680>.
- PLOS. 2019b. "Tracking urban gentrification, one building at a time: New deep-mapping computer model uses Google Street View panoramas to track gentrification at the individual property level." *Science Daily*, March 13, 2019. [www.sciencedaily.com/releases/2019/03/190313140555.htm](http://www.sciencedaily.com/releases/2019/03/190313140555.htm).
- PLOS One. 2019. "Tool tracks gentrification across cities." Science Media Centre of Canada, March 13 2019. <http://sciencemediacentre.ca/site/tool-tracks-gentrification-across-cities/>.
- Porzi, Lorenzo, Samuel Rota Bulò, Bruno Lepri, and Elisa Ricci. 2015. "Predicting and Understanding Urban Perception with Convolutional Neural Networks." In *Proceedings of the 23rd ACM International Conference on Multimedia - MM '15*, 139–148. Brisbane, Australia: ACM Press.
- Pritchard, Jenn. 2019. "Locally developed AI program scans neighbourhoods for gentrification." *City News*, March 18, 2019. <https://ottawa.citynews.ca/local-news/locally-developed-ai-program-scans-neighbourhoods-for-gentrification-1328431/>.
- Prouse, Victoria, Jill L Grant, Martha Radice, Howard Ramos, and Paul Shakotko. 2014. "Neighbourhood Change in Halifax Regional Municipality, 1970 to 2010: Applying the 'Three Cities' Model." Halifax: Dalhousie University. <http://neighbourhoodchange.ca/documents/2014/03/neighbourhood-change-in-halifax-regional-municipality-1970-to->

2010-applying-the-three-cities-model.pdf.

- Prouse, Victoria, Jill L Grant, Howard Ramos, and Martha Radice. 2015. "Assessing Neighbourhood Change: Gentrification and Suburban Decline in a Mid-Sized City." Working Paper. Halifax: Dalhousie University. [http://theoryandpractice.planning.dal.ca/\\_pdf/neighbourhood\\_change/vprouse\\_et\\_al\\_2015.pdf](http://theoryandpractice.planning.dal.ca/_pdf/neighbourhood_change/vprouse_et_al_2015.pdf).
- Public Library of Science. 2019. "Tracking urban gentrification, one building at a time." *Tech Xplore*, March 13, 2019. <https://techxplore.com/news/2019-03-tracking-urban-gentrification.html>.
- Qiu, Leiju, and Daxuan Zhao. 2019. "Urban Inclusiveness and Income Inequality in China." *Regional Science and Urban Economics* 74: 57–64.
- Quastel, Noah. 2009. "Political Ecologies of Gentrification." *Urban Geography* 30 (7): 694–725.
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Ray, Brian, and Valerie Preston. 2009. "Are Immigrants Socially Isolated? An Assessment of Neighbors and Neighboring in Canadian Cities." *Journal of International Migration and Integration* 10 (3): 217–244.
- Razavian, Ali Sharif, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition." In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 512–519.
- Reades, Jonathan, Jordan De Souza, and Phil Hubbard. 2019. "Understanding Urban Gentrification through Machine Learning." *Urban Studies* 56 (5): 922–942.
- Redfern, P.A. 2003. "What Makes Gentrification 'Gentrification'?" *Urban Studies* 40 (12): 2351–2366.
- Reick, Philipp. 2018. "Gentrification 1.0: Urban transformations in late-19th-century Berlin." *Urban Studies* 55 (11): 2542–2558.
- Reis, José P., Elisabete A. Silva, and Paulo Pinho. 2016. "Spatial Metrics to Study Urban Patterns in Growing and Shrinking Cities." *Urban Geography* 37 (2): 246–271.
- ResearchCareer. 2019. "Google maps trendy spread." *ResearchCareer*, March 14, 2019. <https://www.researchcareer.com.au/archived-news/google-maps-trendy-spread>.
- Rose, D. 1984. "Rethinking Gentrification: Beyond the Uneven Development of Marxist Urban Theory." *Environment & Planning D: Society & Space* 2 (1): 47–74.
- Rose, Damaris. 1996. "Economic Restructuring and the Diversification of Gentrification in the

- 1980s: A View from a Marginal Metropolis.” In *City Lives and City Forms: Critical Research and Canadian Urbanism*, edited by Jon Caulfield and Linda Peake, 131–172. Toronto: University of Toronto Press.
- Rose, Damaris, and Amy Twigge-Molecey. 2013. “A City-Region Growing Apart? Taking Stock of Income Disparity in Greater Montréal, 1970-2005.” Research Paper 222. Toronto: Cities Centre, University of Toronto. <https://tspace.library.utoronto.ca/bitstream/1807/94389/1/222%20Rose%20CityRegion%202013.pdf>.
- Rosenfeld, Raymond, and Laura Reese. 2003. “The Anatomy of an Amalgamation: The Case of Ottawa.” *State and Local Government Review* 35 (1): 57–69.
- Rothstein, Richard. 2017. *The Color of Law : A Forgotten History of How Our Government Segregated America*. New York: Liveright Publishing Corporation.
- Rosales, Maria. 2020. “Gentrificación es detectada y mapeada por Inteligencia Artificial.” *Cambio16*, April 21, 2019. <https://www.cambio16.com/gentrificacion-es-detectada-y-mapeada-por-inteligencia-artificial/>.
- Rundle, Andrew G., Michael D.M. Bader, Catherine A. Richards, Kathryn M. Neckerman, and Julien O. Teitler. 2011. “Using Google Street View to Audit Neighborhood Environments.” *American Journal of Preventive Medicine* 40 (1): 94–100.
- Saiz, Ignacio, and Kate Donald. 2017. “Tackling Inequality through the Sustainable Development Goals: Human Rights in Practice.” *International Journal of Human Rights* 21 (8): 1029–1049.
- Salesses, Philip, Katja Schechtner, and César A. Hidalgo. 2013. “The Collaborative Image of The City: Mapping the Inequality of Urban Perception.” *PLoS ONE* 8 (7).
- Sassen, Saskia. 1991. *The Global City: New York, London, Tokyo*. Princeton: Princeton University Press.
- Schaffer, Richard, and Neil Smith. 1986. “The Gentrification of Harlem?” *Annals of the Association of American Geographers* 76 (3): 347–365.
- Schettino, Francesco, and Haider A. Khan. 2020. “Income Polarization in the USA: What Happened to the Middle Class in the Last Few Decades?” *Structural Change and Economic Dynamics* 53: 149–161.
- Schlichtman, John Joe, Marc Lamont Hill, and Jason Patch. 2017. *Gentrifier*. UTP. Toronto: University of Toronto Press.
- Schuler, Jill R., Robert B. Kent, and Charles B. Monroe. 1992. “Neighborhood Gentrification: A Discriminant Analysis of a Historic District in Cleveland, Ohio.” *Urban Geography* 13 (1): 49–67.

- Schulman, Sarah. 2012. *The Gentrification of the Mind: Witness to a Lost Imagination*. Berkeley: University of California Press.
- Science Friday. 2019. "A Street-Level View Of Neighborhood Change." Science Friday, March 15, 2019. <https://www.sciencefriday.com/segments/a-street-level-view-of-neighborhood-change/>.
- Segal, Sahar. 2020. "The International Human Right to Adequate Housing: An Economic Approach." *Chicago Journal of International Law* 20 (2): 486–529.
- Sen, Amartya. 1992. *Inequality Reexamined*. Cambridge: Harvard University Press.
- Sen, Amartya. 1999. *Development as Freedom*. Oxford: Oxford University Press.
- Sen, Amartya. 2000. Chapter 1 Social Justice and the Distribution of Income. Vol. 1. *Handbook of Income Distribution*. Elsevier: 59–85.
- Seresinhe, Chanuki Illushka, Tobias Preis, and Helen Susannah Moat. 2017. "Using Deep Learning to Quantify the Beauty of Outdoor Places." *Royal Society Open Science* 4 (7): 170170.
- Sergeant, Jamie C., and David Firth. 2006. "Relative Index of Inequality: Definition, Estimation, and Inference." *Biostatistics* 7 (2): 213–224.
- Shapiro, S.S., and M.B. Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52 (3–4): 591–611.
- Simmons, A. J. 1964. "An Index of Farm Structure, with a Nottinghamshire Example." *East Midlands Geographer* 5 (3): 255–261.
- Simonyan, Karen, and Andrew Zisserman. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition." In ArXiv:1409.1556.
- Singer, Audrey. 2013. "Contemporary Immigrant Gateways in Historical Perspective." *Daedalus* 142 (3): 76–91.
- Singh, Ajit, and Rahul Dhumale. 2004. "Globalization, Technology, and Income Inequality: A Critical Analysis." In *Inequality Growth and Poverty in an Era of Liberalization and Globalization*. Oxford: University of Oxford.
- Sitthiyot, Thititthep, and Kanyarat Holasut. 2020. "A Simple Method for Measuring Inequality." *Palgrave Communications* 6 (1).
- Smith, Christopher B.R. 2010. "Socio-spatial stigmatization and the contested space of addiction treatment: Remapping strategies of opposition to the disorder of drugs." *Social Science &*

- Medicine* 70 (6): 859–866.
- Smith, Darren, Martin Phillips, and Chloe Kinton. 2018. “Wilderness Gentrification: Moving ‘off-the-Beaten Rural Tracks.’” In *Handbook of Gentrification Studies*, edited by Loretta Lees and Martin Phillips, 363–89. Cheltenham: Edward Elgar Publishing.
- Smith, David. 1994. *Geography and Social Justice*. Oxford: Blackwell Publishers.
- Smith, Neil. 1987. “Commentary: Gentrification and the Rent Gap.” *Annals of the Association of American Geographers* 77 (3): 462–465.
- Smith, Neil. 1996. *The New Urban Frontier: Gentrification and the Revanchist City*. London: Routledge.
- Smith, Neil. 2008. *Uneven Development: Nature, Capital, and the Production of Space*. 3rd ed. Athens: University of Georgia Press.
- Smith, Neil, Paul Caris, and Elvin Wylie. 2001. “The ‘Camden Syndrome’ and the Menace of Suburban Decline: Residential Disinvestment and Its Discontents in Camden County, New Jersey.” *Urban Affairs Review* 36 (4): 497–531.
- Smith, Neil, and Peter Williams, eds. 1986. *Gentrification of the City*. Boston: Allen & Unwin.
- Soja, Edward W. 2010. *Seeking Spatial Justice*. Minneapolis: University of Minnesota Press.
- Spain, Daphne. 1980. “Indicators of Urban Revitalization: Racial and Socioeconomic Changes in Central-City Housing.” In *Back to the City: Issues in Neighborhood Renovation*, edited by Shirley Bradway Laska and Daphne Spain, 27–41. New York: Pergamon Press.
- Spain, Daphne. 1981. “A Gentrification Scorecard.” *American Demographics* 3 (10): 14–19.
- Spermann, Alexander 2017. “Basic Income in Germany: Proposals for Randomised Controlled Trials Using Nudges.” *Basic Income Studies* 12 (2).
- Statistics Canada. 2015. “History of Census of Canada.” December 30, 2015. <https://www12.statcan.gc.ca/census-recensement/2011/ref/about-afpropos/history-histoire-eng.cfm>.
- Statistics Canada. 2017a. “Focus on Geography Series, 2016 Census.” Statistics Canada. <https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Index-eng.cfm>.
- Statistics Canada. 2017b. “After-Tax Income.” Dictionary, Census of Population, 2016. September 13, 2017. <https://www12.statcan.gc.ca/census-recensement/2016/ref/dict/pop004-eng.cfm>.
- Statistics Canada. 2018. “Canada Goes Urban.” May 17, 2018. <https://www150.statcan.gc.ca/n1/>

pub/11-630-x/11-630-x2015004-eng.htm

Stehlin, John G., and Alexander R. Tarr. 2017. "Think Regionally, Act Locally?: Gardening, Cycling, and the Horizon of Urban Spatial Politics." *Urban Geography* 38 (9): 1329–1351.

Steinmetz-Wood, Madeleine, Rania Wasfi, G. Parker, Lisa Bornstein, Jean Caron, and Yan Kestens. 2017. "Is Gentrification All Bad? Positive Association between Gentrification and Individual's Perceived Neighborhood Collective Efficacy in Montreal, Canada." *International Journal of Health Geographics* 16 (1).

Stiglitz, Joseph. 2012. *The Price of Inequality*. New York: WW Norton & Company.

Surgue, Thomas. 2014. *The Origins of the Urban Crisis: Race and Inequality in Post-War Detroit*. 3rd ed. Princeton: Princeton University Press.

Swyngedouw, Erik. 2006. "Introduction to David Harvey." In *Divided Cities: The Oxford Amnesty Lectures 2003*, edited by Richard Scholar, 79–82. Oxford: Oxford University Press.

Taylor-Gooby, Peter. 2013. "Why Do People Stigmatise the Poor at a Time of Rapidly Increasing Inequality, and What Can Be Done about It?" *Political Quarterly* 84 (1): 31–42.

Tierney, Sean, and Clint Petty. 2015. "Gentrification in the American Heartland? Evidence from Oklahoma City." *Urban Geography* 36 (3): 439–456.

Tissot, Sylvie. 2015. *Good Neighbors: Gentrifying Diversity in Boston's South End*. London: Verso.

The Map Room. 2019. "Using Street View to Spot Gentrification." The Map Room, March 21 2019. <https://www.maproomblog.com/2019/03/using-street-view-to-spot-gentrification/>.

Townsend, Peter. 1993. "Underclass and Overclass: The Widening Gulf between Social Classes in the 1980s." In *Sociology in Action: Applications and Opportunities for the 1990s*, edited by Geoff Payne, and Malcolm Cross, 91–118. London: Macmillan.

Townshend, Ivan, Byron Miller, and Leslie Evans. 2018. "Socio-Spatial Polarization in an Age of Income Inequality: An Exploration of Neighbourhood Change in Calgary's 'Three Cities.'" Toronto: Neighbourhood Change Research Partnership. <http://neighbourhoodchange.ca/documents/2018/04/socio-spatial-polarization-in-calgary.pdf>.

Townshend, Ivan, and Ryan Walker. 2002. "The Structure of Income Residential Segregation in Canadian Metropolitan Areas." *Canadian Journal of Regional Science* 25 (1): 25–52.

- Townshend, Ivan, and Ryan Walker. 2020. "Urban Change through the Lens of Demography, Life Course, and Lifestyles." In *Canadian Cities in Transition: Understanding Contemporary Urbanism*, edited by Marcus Moos, Tara Vinodrai, and Ryan Walker, 6th ed., 110–133. Don Mills: Oxford University Press.
- Turcotte, Bradley. 2016. "Houses of Hintonburg Project Documents a Changing Neighbourhood." *Kitchissippi Times*, February 5, 2016. <https://kitchissippi.com/2016/02/05/houses-of-hintonburg-instagram/>.
- Tuson, M., M. Yap, M.R. Kok, K. Murray, B. Turlach, and D. Whyatt. 2019. "Incorporating Geography into a New Generalized Theoretical and Statistical Framework Addressing the Modifiable Areal Unit Problem." *International Journal of Health Geographics* 18 (1).
- United Nations. 2018. "World Urbanization Prospects 2018 Country Profiles." <https://population.un.org/wup/Country-Profiles/>.
- uOttawa, 2019. "Tracking urban gentrification, one building at a time." uOttawa Newsroom, March 13, 2019. <https://www.uottawa.ca/about-us/media/news/tracking-urban-gentrification-one-building-time>.
- Vandette, Kay. 2019. "New tool uses Google Street View to track neighborhood gentrification." Earth.com, March 14, 2019. <https://www.earth.com/news/track-neighborhood-gentrification/>.
- Vanwolleghem, Griet, Delfien Van Dyck, Fabian Ducheyne, Ilse De Bourdeaudhuij, and Greet Cardon. 2014. "Assessing the Environmental Characteristics of Cycling Routes to School: A Study on the Reliability and Validity of a Google Street View-Based Audit." *International Journal of Health Geographics* 13.
- Vinodrai, Tara, and Markus Moos. 2015. "Appendix A: Do We Still Have Quality Data to Study Canadian Cities?" In *Canadian Cities in Transition*, edited by Pierre Filion, Markus Moos, Tara Vinodrai, and Ryan Walker, 5th ed., 436–439. Don Mills: Oxford University Press.
- Walks, Alan. 2015. "Growing Divisions: Inequality, Neighbourhood Poverty, and Homeless in the Canadian City." In *Canadian Cities in Transition*, edited by Pierre Filion, Markus Moos, Tara Vinodrai, and Ryan Walker, 5: 151–174. Don Mills: Oxford University Press.
- Walks, Alan. 2016. "Homeownership, Asset-Based Welfare and the Neighbourhood Segregation of Wealth." *Housing Studies* 31 (7): 755–784.
- Walks, Alan, Emily Hawes, and Dylan Simone. 2021. "Gentrification in Large Canadian Cities: Tenure, Age, and Exclusionary Displacement 1991-2011." *Urban Geography* 42 (5): 603–633.

- Walks, Alan, and Richard Maaranen. 2008a. "The Timing, Patterning, & Forms of Gentrification & Neighbourhood Upgrading in Montreal, Toronto, & Vancouver, 1961 to 2001." Research Paper 211. Cities Centre, University of Toronto: Centre for Urban and Community Studies. <http://www.urbancentre.utoronto.ca/pdfs/publications/RP211Walk-Maaranen-Gentrification1960-2001.pdf>.
- Walks, R. Alan, and Richard Maaranen. 2008b. "Gentrification, Social Mix, and Social Polarization: Testing the Linkages in Large Canadian Cities." *Urban Geography* 29 (4): 293–326.
- Wang, X., F.G. Blanchet, and N. Koper. 2014. "Measuring Habitat Fragmentation: An Evaluation of Landscape Pattern Metrics." *Methods in Ecology and Evolution* 5 (7): 634–646.
- Watson, Sophie, and Katherine Gibson, eds. 1995. *Postmodern Cities and Spaces*. Oxford: Blackwell.
- WBEZ Chicago. 2019. "Frans de Waal, Inactive Ingredients, Street View, and Gentrification." March 15, 2019, Part 2. *WBEZ Chicago*, March 15, 2019. <https://www.wbez.org/stories/frans-de-waal-inactive-ingredients-street-view-and-gentrification-march-15-2019-part-2/540fcb2e-4e66-491b-a066-ae60cd217ab5>.
- Weber, Rachel. 2002. "Extracting Value from the City: Neoliberalism and Urban Redevelopment." *Antipode* 34 (3): 519–540.
- Weston, Julia. 1982. Gentrification and Displacement: An Inner-City Dilemma. *Habitat* 25 (1): 10–19.
- Williams, P. R. 1976. "The role of institutions in the inner London housing market: the case of Islington." *Transactions of the Institute of British Geographers* 1 (1): 72–82.
- Willing, Jon. 2021. "Intensification Worries Some Westboro Residents." *Ottawa Citizen*, February 12, 2021. <https://www.pressreader.com/canada/ottawa-citizen/20210212/281539408655621>.
- Wilson, David. 1990. "Institutions, Agency, and Microscale Gentrification: A Tale of Two Neighborhoods." *Journal of Urban Affairs* 12 (3): 267–283.
- Wilson, David. 1991. "Urban Change, Circuits of Capital, and Uneven Development." *Professional Geographer* 43 (4): 403–415.
- Wilson, David. 2006. *Cities and Race: America's New Black Ghetto*. London: Routledge.
- Wilson, David, Jared Wouters, and Dennis Grammenos. 2004. "Successful Protect-Community Discourse: Spatiality and Politics in Chicago's Pilsen Neighborhood." *Environment and*

- Planning A* 36 (7): 1173–1190.
- Wilson, Jeffrey S., Cheryl M. Kelly, Mario Schootman, Elizabeth A. Baker, Aniruddha Banerjee, Morgan Clennin, and Douglas K. Miller. 2012. “Assessing the Built Environment Using Omnidirectional Imagery.” *American Journal of Preventive Medicine* 42 (2): 193–199.
- Wong, Sebastien C., Adam Gatt, Victor Stamatescu, and Mark D. McDonnell. 2016. “Understanding Data Augmentation for Classification: When to Warp?” In *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 1–6.
- Wu, Katherine J. 2019. “Could Google Street View Help Track Gentrification?” *PBS - NOVA*, sec. NOVA Next, March 13, 2019. <https://www.pbs.org/wgbh/nova/article/could-google-street-view-help-track-gentrification/>.
- Wu, Yu-Tzu, Paul Nash, Linda E. Barnes, Thais Minett, Fiona E. Matthews, Andy Jones, and Carol Brayne. 2014. “Assessing Environmental Features Related to Mental Health: A Reliability Study of Visual Streetscape Images.” *BMC Public Health* 14 (1).
- Wyly, Elvin. 2011. “Positively Radical.” *International Journal of Urban and Regional Research* 35 (5): 889–912.
- Wyly, Elvin. 2012. “The Future of Social Housing - Elvin Wyly.” Renters Union, Dec 3, 2012. <https://www.youtube.com/watch?v=nCITM7h-jzE>.
- Wyly, Elvin K., and Daniel J. Hammel. 1999. “Islands of Decay in Seas of Renewal: Housing Policy and the Resurgence of Gentrification.” *Housing Policy Debate* 10 (4): 711–771.
- Yonto, Daniel, and Claire Schuch. 2020. “Developing and Ground-Truthing Multi-Scalar Approaches to Mapping Gentrification.” *Papers in Applied Geography*, 352–368.
- Yonto, Daniel, and Jean-Claude Thill. 2020. “Gentrification in the U.S. New South: Evidence from Two Types of African American Communities in Charlotte.” *Cities* 97.
- Yosinski, Jason, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. “How Transferable Are Features in Deep Neural Networks?” In *Advances in Neural Information Processing Systems*, 4: 3320–3328.
- Zarzalejos, Ana. “La inteligencia artificial ya es capaz de detectar la gentrificación con imágenes de Google Street.” *Business Insider*, March 25, 2019. <https://www.businessinsider.es/google-street-view-puede-detectar-gentrificacion-antes-pase-392477>.
- Zhu, Jieming, and Yan Guo. 2022. “Social Justice in Spatial Change: Transition from Autonomous Rural Development to Integrated Urbanization in China.” *Cities* 122.
- Zukin, Sharon. 1987. “Gentrification: Culture and Capital in the Urban Core.” *Annual Review of*

*Sociology* 13: 129–147.

Zukin, Sharon. 1991. *Landscapes of Power: From Detroit to Disney World*. Berkeley: University of California Press.

Zwicker, Heather, Kisha Supernant, and Erika Luckert. 2017. “Social Mobility: Charting the Economic Topography of Urban Space.” *Television and New Media* 18 (4): 375–388.

## **Appendix I. Examining income trajectories and income polarization in Ottawa, 1971-2016.**

**Abstract:** The growth of income and wealth inequality has become a contemporary research topic and is especially relevant in Canadian contexts as the country is experiencing elevated income inequality. Herein, income inequality in Ottawa between 1971 and 2016 is examined. This study applies Hulchanski's three city model, and delineates how Hulchanski's methodology produces two products: an analysis of income trajectories and an analysis of income polarization. Additionally, an alternative means of examining income trajectories to examine the changing income landscape of the capital city is presented. Large divides in the city have taken place between the years that are examined and as such Ottawa is shown to be an increasingly polarized urban area.

### **Introduction**

Some researchers feel that the growth of income and wealth inequality has been recognized as the greatest social threat of our time (Dorling 2015, 1). Income inequality is especially relevant in the Canadian context, as in recent decades the country is experiencing one of the highest rises in income inequality in the developed world (Banting and Myles 2013; Beach 2016; Breau 2015; Fortin et al. 2012; Osberg 2008; Myles, Picot, and Pyper 2000; Fong 2017). In Canada, inequality is observed through the increasing polarization and divide between haves and have-nots. The general understanding of polarization entails a declining middle with concomitant expansion of two opposite poles.

This study starts with replicating the popular Three City Model (TCM) as it was initially applied to Toronto (Hulchanski 2010). Hulchanski's TCM became influential for examining income inequality in Canadian urban contexts and has in some form been applied by scholars in cities such as Calgary (Townshend, Miller, and Evans 2018), Halifax (Prouse et al. 2014), Hamilton (Harris, Dunn, and Wakefield 2015), Montreal (Rose and Twigge-Molecey 2013) and Vancouver (Ley and Lynch 2012).

The TCM essentially examines income trajectories, or the percentage change of individual income past a certain threshold, between two census years. Each TCM study also investigated income polarization, which is a form of income inequality whereby the disappearance of the middle-income group occurs while both low- and high-income groups expand.

This present study addresses income inequality in Ottawa through an analysis of income polarization and income trajectories, while aiming to achieve the following three objectives:

- (1) Reproduce the Three City Model (as initially done in Toronto) in Ottawa's context in order to examine income trajectories.
- (2) Examine income polarization in the same manner as Hulchanski.
- (3) Assess class change in Ottawa.

This paper has several sections, beginning with an explanation of the products of the TCM. With this necessary base established, the methodology explains how data is procured, how Hulchanski's methodology is reproduced, and poses a different method with which to assess class change through an examination of income trajectories.

### **The Three City Model**

Studying income polarization and income trajectories within cities stems from studies that examine urban inequality. Such research gained prominence in the 1990s, as scholars examined increasing socioeconomic divisions and their consequences (Davis 1990; Mollenkopf and Castells 1991; Sassen 1991; Fainstein, Gordon, and Harloe 1992; Bird et al. 1993; Watson and Gibson 1995). For example, in the American context, Molencof and Castells (1991) described New York City as divided into three parts.

Researchers examined such processes in Canada too (Bourne 1993a; MacLachlan and Sawada 1997; Myles, Picot, and Pyper 2000). MacLachlan and Sawada's (1997) contribution is

of significance to later research, because they outlined important principles for analyzing income polarization.

The Three City Model (TCM) was presented as a tool to analyze income polarization between two time periods (referred to as 'endpoints' henceforward) separated by several decades (Hulchanski 2007). Contrary to its appearance, the TCM measures income trajectories rather than income polarization (Distasio and Kaufman 2015, 11). Hence, per the TCM, Hulchanski envisioned the urban landscaped as divided into three cities: one with a trajectory of income growth of more than 20 percent (1st city), one with income decline of the equal amount (3rd city), and a somewhat stagnant middle-ground where income change ranged between a -20 and +20 percent trajectory (2nd city).

An updated report (Hulchanski 2010) was subsequently issued, which went in more depth, both in terms of production value and methodological explanations. Subsequent studies using the TCM were conducted on Census Metropolitan Areas (CMAs) such as Vancouver (Ley and Lynch 2012), Montreal (Rose and Twigge-Molecey 2013), Halifax (Prouse et al. 2014), Hamilton (Harris, Dunn, and Wakefield 2015), and Calgary (Townshend, Miller, and Evans 2018). A recent publication ties together these studies in a more uniform and coherent manner (Grant, Walks, and Ramos 2020), perhaps serving as inspiration or a basis for future researchers.

However, most researchers who use the TCM also examine income polarization separately of income trajectories. For any given study area, analyzing income polarization is done through the examination of yearly time-slices of the study area's income structure. Sometimes data from time slices is portrayed on histograms, while at other times the data is also mapped. This point will be returned to shortly.

The reports using the TCM differ in methodology, as there is little consistency in the choice of threshold for income trajectories which categorize an area as belonging to one of the three cities. Hulchanski (2007; 2010) uses a 20 percent change in income as a threshold, but subsequent studies use smaller thresholds. Some used a 15 percent threshold (Ley and Lynch 2012; Rose and Twigge-Molecey 2013), whereas others used 10 percent thresholds (Prouse et al. 2014; Harris, Dunn, and Wakefield 2015; Townshend, Miller, and Evans 2018). These subsequent studies chose smaller thresholds because Toronto's threshold was too large to produce a similar story of divided cities (Grant, Walks, and Ramos 2020, 263–64). Hence, the selection of smaller percentages was chosen in order to engineer results whereby categorizations for the respective three cities would be more significant.

An additional point of difference is present among the TCM studies: the endpoints of the studies. The earlier studies used census data between the 1971 and 2006 (Hulchanski 2010; Ley and Lynch 2012; Rose and Twigge-Molecey 2013), whereas the later ones have 1981 census data and custom data for 2010 from Revenue Canada as endpoints (Prouse et al. 2014; Harris, Dunn, and Wakefield 2015; Townshend, Miller, and Evans 2018). In the recent publication by this group of researchers, it was suggested that using 1971 may include fringe areas which were not developed (J. Grant, Walks, and Ramos 2020, 262). However, this is not the case, as CMAs were all much smaller in geographic size in 1971, and hence did not include such areas within their borders.

The unit of analysis in the TCM was the census tract (CT). In urban contexts, CTs are often used because they provide stable borders for comparisons across time. In suburban areas, CTs tend to split due to population growth. To compare income trajectories, TCM studies had to standardize CTs of two endpoints. Standardization means that CT configurations from two years

had to conform. Per the TCM, this is achieved by assigning values from one census year onto the geographical configuration of another year.

What separates Hulchanski's (2010) study is that it is the only one which mapped out each examined time-slice for each census year (excluding 1986) and nicely conveyed the information in a figure on the back cover of its publication. The other TCM practitioners also made time slice maps, but not for every census year. They chose to demonstrate their yearly data by using histograms, though they all excluded data from the 1986 census for unexplained reasons (Ley and Lynch 2012, 20; Rose and Twigge-Molecey 2013, 13; Prouse et al. 2014, 22; Harris, Dunn, and Wakefield 2015, 10; Townshend, Miller, and Evans 2018, 5). The Vancouver and Montreal study also excluded data from the 1996 census for likewise unexplained reasons (Ley and Lynch 2012, 20; Rose and Twigge-Molecey 2013, 13).

Mapping produces an additional layer of insight into changing spatial patterns of the study area's income structure, but here too the TCM practitioners are not consistent in their approach. The difference surrounds units being mapped. Hulchanski's studies used standardized CTs from the 2001 census year when mapping yearly time-slices. The CT configuration from this year was chosen because it was the last year available when his first report (Hulchanski 2007) was conducted. The updated study (Hulchanski 2010) used data from 2006 adjusted to 2001 CT boundaries. Vancouver's study used CTs from 1971 as the unit of analysis when histograms were produced (Ley and Lynch 2012, 20). Vancouver's study and all other TCM studies mapped CTs using their respective years.

The yearly time slices have three categories which define income groups based on their percentage of income in relation to the average CMA income. High-, medium- and low-income groups. These reflect income over 120%, between 80% & 120%, and under 80% of the CMA

average income respectively. In some studies, additional categories of very-high (above 140%) and very-low (below 60%) appears in maps.

Finally, the geographic extent of Hulchanski's two studies were set to a constrained spatial area which covers the extent of the post-amalgamated city of Toronto. This area roughly reflects the inner-city and inner-suburbs (Walks 2007, 169). Other TCM practitioners examine the whole CMA.

### **Methodology**

This study has a few components. The first focuses on reproducing the methodology of Hulchanski's (2010) study in order to gauge income polarization and income trajectories. Approaches from other TCM practitioners are also applied, in order to assess how results would differ under different income trajectories.

Next, income polarization is mapped, per Hulchanski's methods. Additional sets of maps are produced, using different definitions of middle-income, to determine if the same outcomes result from different assumptions. However, reproducing the original study is not a straightforward task. The two main challenges include procuring data and defining the spatial extent of the study area.

Finally, this study wraps up by measuring class change of Census Tracts (CTs) class through the use of difference maps. Such approach overcomes the weakness of income trajectories, for they disregard actual class change. For example, if CTs, whose average incomes are high, become even more affluent, they should not be labeled as on an upward swing as they are already at the top.

### *Study Area*

Per Hulchanski's methodology, the study area should correspond to a limited spatial extent which is roughly the extent of the inner city and inner suburbs. In Ottawa this would reflect the built-up part of the city which is bounded by the greenbelt. This region represents older housing stock in the city, as mapped in a recent study (Ilic, Sawada, and Zarzelli 2019, 6). The unit of analysis in the present study is the CT, the most common unit of analysis in urban geography (Ilic and Sawada 2021, S3).

### *Data Procurement*

Data for this study comes from every Canadian census held between 1971 and 2016, with the exception of 1976 and 2011. Data from the long form of the 2011 census is inadequate as it is not from a random sample (Vinodrai and Moos 2015). Data from the 1976 census is not used, as that was the last mid-decade census which was not a full census.

The most common variable used to measure inequality is income (Buitelaar, Wterings, and Ponds 2018, 33), and the TCM uses average individual pre-tax income. Average income data for individuals is available only starting in 1996. It may be plausible to assume that TCM practitioners had data prior to 1996 assembled through custom tabulations. In this study, data which is not readily available is simply derived from available data.

Data from every examined census has average income figures for males and females which allows for the possibility to derive average individual income using the following formula:

$$AI_i = \frac{AI_m \times SM_{wi} + AI_f \times SF_{wi}}{2}$$

Where  $AI_i$  = average income of individuals,  $AI_m$  = average income of males,  $AI_f$  = average income of females.  $SM_{wi}$  = share of males with income, and  $SW_{wi}$  = share of females with

income. The formula's accuracy was tested with data from select census years in which data was provided for individuals and for males and females.

This study examines individual income in relation to CMA income. As the average income of the CMA is available for every year except for 1971, the following formula was derived based on a similar formula that was applied household incomes (Lazar Ilic and Sawada 2021, 6).

$$AI_c = \sum_{k=0}^n \left( \frac{I_{wi} \times AI_i}{TI_{wi}} \right)_k$$

Where  $AI_c$  = Average Income of study area,  $I_{wi}$  = total individuals with income,  $AI_i$  = Average Income of Individuals, and  $TI_{wi}$  = total individuals with income. As with the previous formula, this one was verified using data from subsequent census'.

### *Three City Model*

In this study, the original methodology that Hulchanski employed is used to examine Ottawa's income trajectories between 1971 and 2016. The usage of all three thresholds that TCM practitioners used is employed, meaning that trajectories of 10, 15 and 20 percent would be examined and mapped. Additionally, maps are produced which show only rise and decline in average individual income.

### *Income Polarization*

Analyzing income polarization will be done through an examination of time slices as done by Hulchanski (2010). However, insight from other researchers (MacLachlan and Sawada 1997; Hamnett 2003) is incorporated in order to test if income-polarization is indeed occurring.

MacLachlan and Sawada (1997) noted that it is necessary to examine multiple definitions of middle-income when analyzing income-polarization, as their results demonstrated that

concurrent polarization and de-polarization can be observed on the same dataset depending on how middle-income is defined. TCM practitioners define the middle-income in the range from 80% to 120% percent of average income, in their time-slice maps. In the present study, two additional bands will be examined: a shorter band of 85% to 115% and a longer one of 75% to 125%. Income polarization would indeed be taking place if it is observed under multiple conditions.

Hamnett (2003, 70) noted that, when examining income polarization, it is necessary to measure changes in sizes of all three groups' (top, middle and bottom) earning or income distributions in order to gauge how they changed over time. This allows for the determination if income-polarization is indeed taking place. For example, if the middle-income group declines, while the low-income group increases, without an increasing of the high-income group, then lumpenization (expansion of lower classes) rather than polarization may be taking place (Ilic and Sawada 2021, 17).

### *Difference Maps*

To assess social change, it is important to be able to account for class change. The TCM's method of income trajectory delineation is not adequate in such endeavor, as income trajectories can occur while the social structure of a CT remains constant. Through difference maps, class change can be accounted for when examining income change.

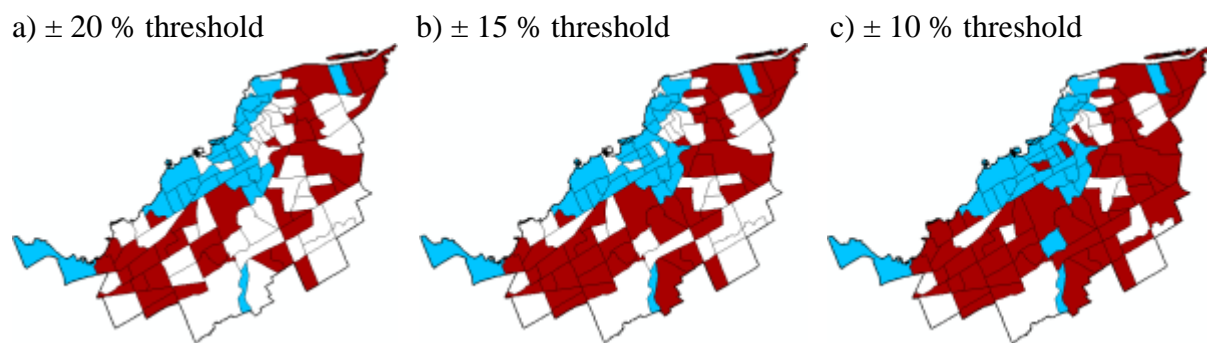
There are many categories for classifying the urban landscape. For example, a recently study envisaged Toronto as being divided into ten categories (McGuire 2012). This study uses three types of difference maps, in order to investigate how patterns may differ under different assumptions.

The first difference map uses a three-class categorization of the income landscape, whereas the second difference map uses a five-class categorization. The three-class division separates classes into the following ranges in relation to the average income: 1) under 80%, 2) 80% to 120%, and 3) over 120%. The five-class division adds an under 60% and over 140% category. Both of these categorizations correspond to how the TCM products time slices. Finally, a third difference map incorporates ten categories, starting at under 60% and increasing at 10% intervals until reaching the above 140% category.

To produce difference maps, maps are converted from vector to raster format. Each CT's class structure is assigned a number prior to conversion, with the lowest class being assigned a '1' and the highest one getting a '3', '5' or a '10' respective of how many total classes are used. Difference maps are produced by applying a minus function to two maps, which in our case are our study's time-slice endpoints (1971 and 2016). The minus function creates a new map by subtracting the values of one year from the other, in effect showing the change in assigned values of each CT. Hence, in the final map, a CT with a higher number indicates more upward class turnover. Conversely, CTs with negative values indicate decline. In effect, difference maps show class change and may be effective in depicting where gentrification has occurred in the city.

## **Results**

The size of the threshold for measuring income trajectories results in differing results (Figure 1 & Table 1). The smaller the threshold, the greater the first and third cities, which are depicted by blue and red respectively. The twenty percent threshold shows that City 2 has a plurality of CTs. This changes as the threshold decreases. The fifteen percent threshold shows that a plurality of CTs are in City 3, while the ten percent threshold results in City 3 containing the majority of CTs.



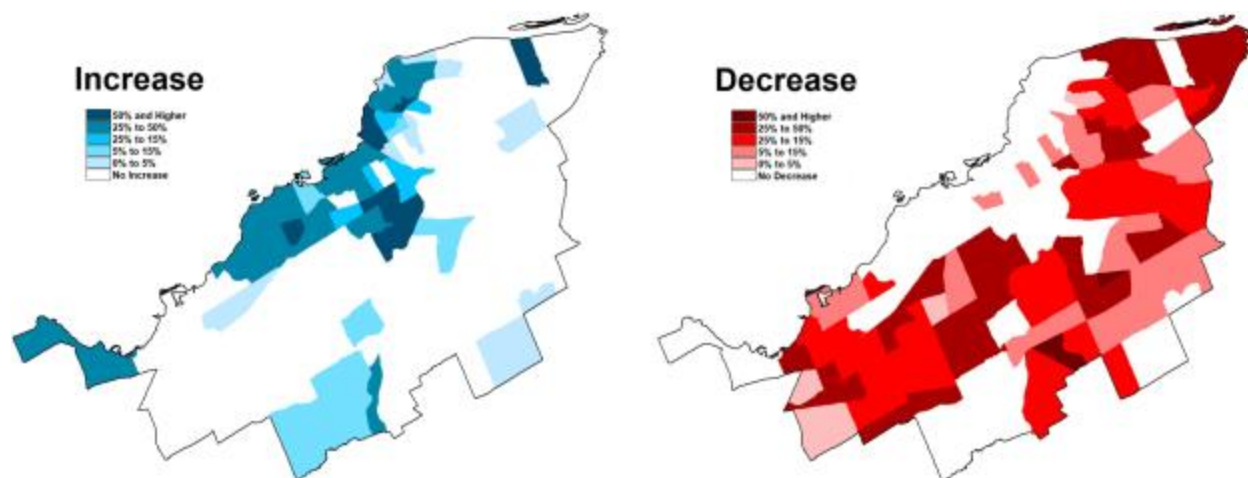
**Figure 1. Change in the Census Tract Average Individual Income as a Percentage of the Ottawa-Gatineau CMA Average, 1970-2015.** Blue – City #1 (increase above threshold), White – City # 2 (Increase or decrease is less than threshold), Red – City # 3 (Decrease more than threshold).

Table 1. Census Tracts within respective areas

Threshold	City # 1	City # 2	City # 3
20 Percent	28 (25.00%)	45 (40.18%)	39 (34.82%)
15 Percent	31 (27.68%)	34 (30.36%)	47 (41.96%)
10 Percent	33 (29.46%)	19 (16.96%)	60 (53.57%)

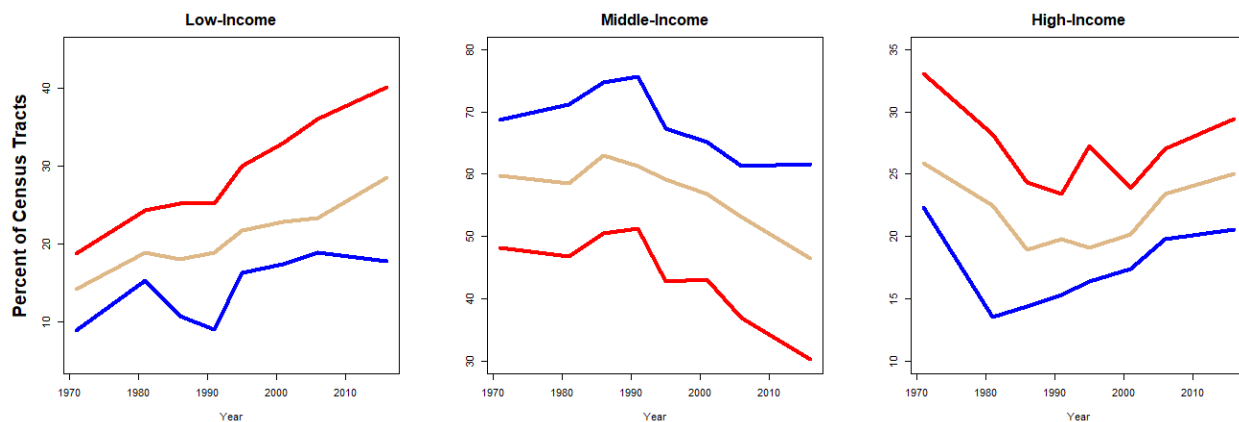
Mapping the change of income between 1971 and 2016 indicates that the 1st City is mostly confined to parts of the inner-city (Figure 2). Some parts of the city, primarily the central areas, have experienced notable income increases. These include areas such as the Glebe, Westboro, Hintonburg, New Edinburgh, Old Ottawa South, and the Market. Areas outside the central part of the city include places such as Crystal Bay and Rothwell Heights.

Many places, particularly those between the inner-city and greenbelt are undergoing significant decline in average incomes. These include areas such as Wateridge Village, Beacon Hill, Cardinal Heights, Hawthorne Meadows, Rideauview, Fisher-Meadowlands, Bayshore, Ledbury, Briar Green-Leslie Park, and Hunt-Club East.



**Figure 2. Increase and decrease of incomes in Ottawa between 1971 and 2016.**

Different definitions of middle-income yield different patterns, certain trends are consistent throughout (Figure 3). In each definition, the low-income CTs have increased and middle-income CTs have declined. High-income CTs have initially declined, and then started to grow. The decline stops at different moments depending on which definition of middle-income is used. Middle-income CTs grew until 1986 or 1991, depending on which definition one uses. The 1990s then resulted in significant decline of such CTs.

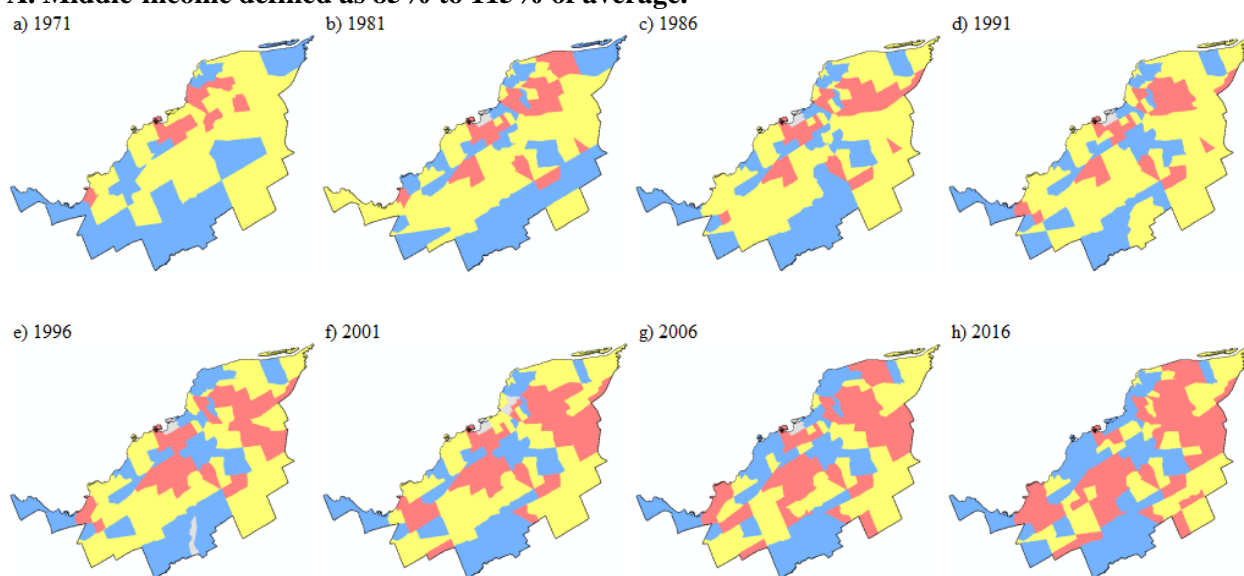


**Figure 3. Change in high-, middle- and low-income CTs between 1971 and 2016.** Blue shows the largest definition of middle income (+/- 25 percent of the average), and red shows the lowest definition (+/- 15 percent of the average).

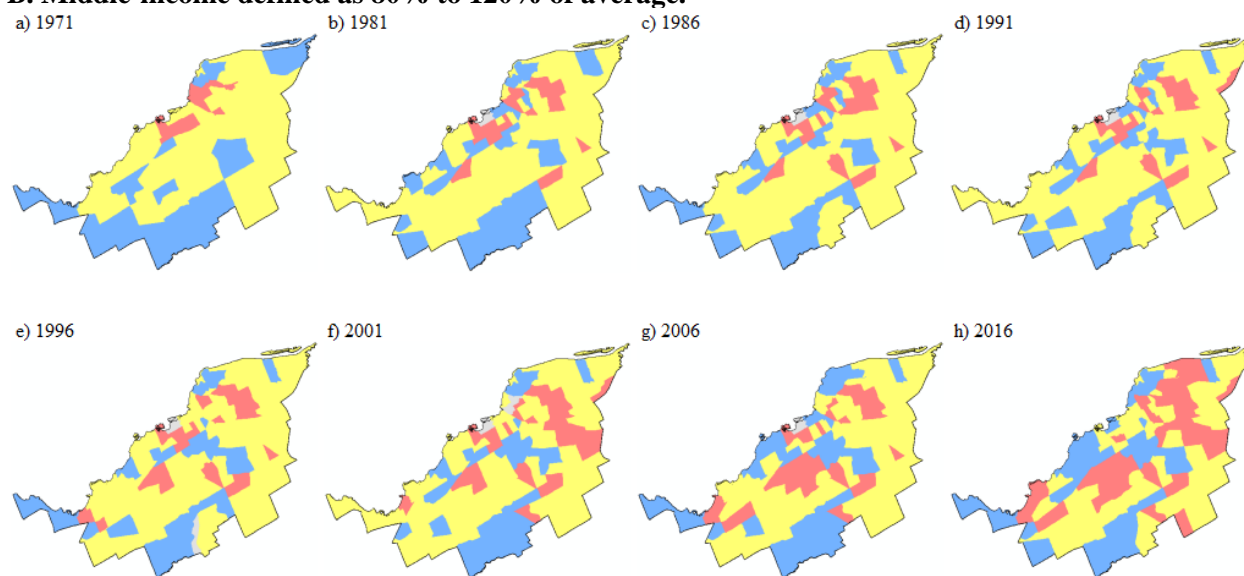
Mapping time-slices yields insight into how Ottawa's income landscape has changed (Figure 4). The three sets of maps show a similar story, whereby low-income areas have

significantly expanded over time, and shifted in location away from being primarily in the inner-city to being in the inner suburbs. The emergence of many pockets of high- and low-income areas has produced a landscape that resembles a mosaic. In 2016, the most affluent are in areas such as the Glebe, Westboro, Island Park-Wellington Village, Rockcliffe Park, and Rothwell Heights, whereas areas predominantly populated by low-income denizens include Carlington, Vanier, Ledbury, and Hawthorne Meadows.

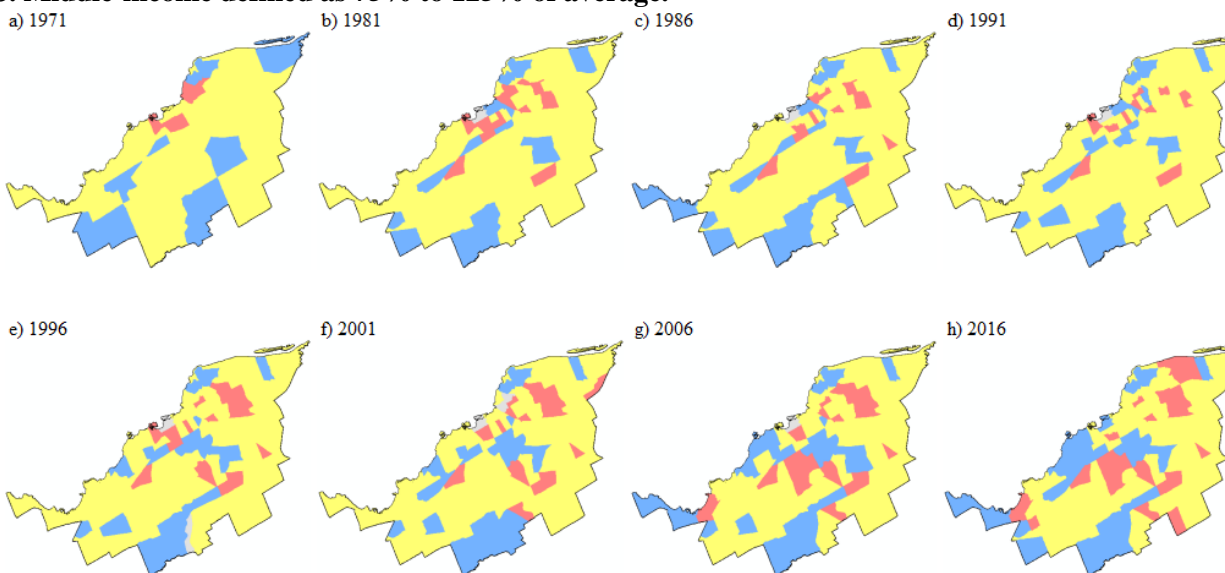
**A. Middle-income defined as 85% to 115% of average.**



**B. Middle-income defined as 80% to 120% of average.**

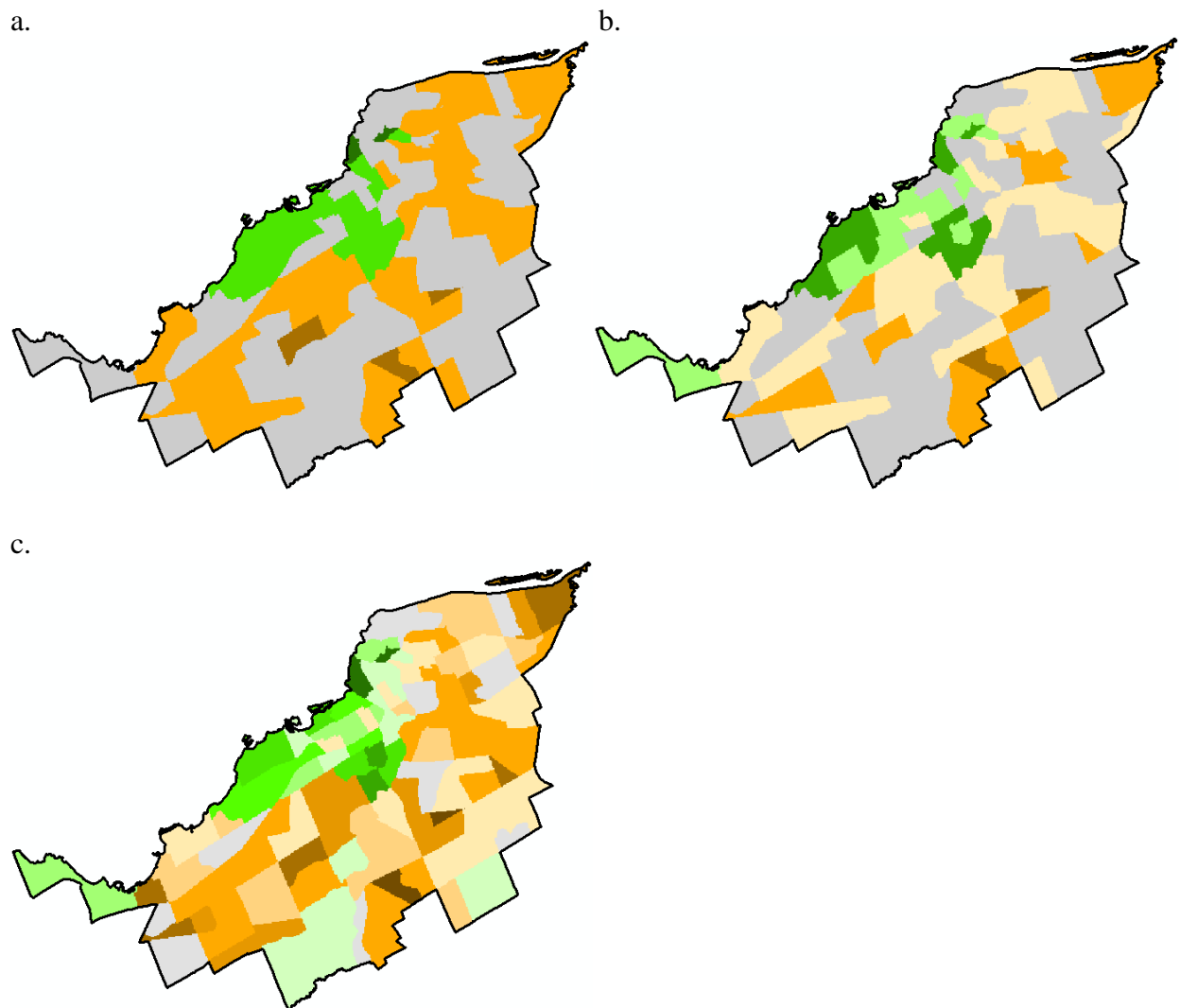


**C. Middle-income defined as 75% to 125% of average.**



**Figure 4. Time-slices of Ottawa's income-structure, 1971-2016.**

Maps yield different results depending on how many income categories (three, five and ten) were assigned to the study area (Figure 5). Using only three categories gives a somewhat generalized map. As the number of categories increases, patterns emerge which convey more detailed extents of CT income decline or increase. The Market and Lindenlea are areas which have appreciated the most. Parts of the Glebe and Old Ottawa South have also appreciated, but the ten category map shows that they did not change to the extent of the Market and Lindenlea. Likewise, decline is most vivid in parts of Hunt Club West, and Ledbury.



**Figure 5. Difference maps, showing class change from 1971 to 2016.** Maps a, b and c show the products of three, five and ten categories being used to produce such maps.

## Discussion

The present study shows that Ottawa has become an increasingly divided city. The spatial pattern of Ottawa's income structure has come to resemble a mosaic by 2016. This patchwork of various income structures is indicative of what Hackworth (2005; 2007) termed the "Neoliberal Spatial Fix", where certain suburban areas are increasingly depreciating and central areas are

increasingly appreciating in value. Contrary to common notions of polarization, the decline of middle-income areas occurred with an expansion of low-income areas without an expansion of high-income ones. Hence, a different phenomenon is taking place which might not necessarily be income polarization per se.

The usefulness of the Three City Model remains unclear. The present study shows that one can engineer results to fit a desired narrative based on which threshold of income change is chosen. The bigger the income trajectory threshold, the larger the second city. Conversely, lower thresholds produce a smaller second city, and in the case of Ottawa, a higher third city. However, regardless of what threshold one chooses, the second city remains less than half of all CTs.

Mapping income polarization shows common patterns irrespective of what definition of middle-income is used. In terms of raw numbers, the middle-income group has steadily decreased since the late 1980s or 1990s. It is difficult to pinpoint an exact year when the middle started to decline, as the census is not conducted on a yearly basis, nor do all definitions of the middle produce the same maximum extent of the middle-income group. In spatial terms, the maps (Figure 4) show that the low-income area has generally expanded. Initially located only in inner-city areas, near the city core, the low-income areas are increasingly located in suburban areas. Furthermore, by 2016, very few CTs near the center of the city had average low-incomes.

Using difference maps to examine income trends appears effective. As the number of income categories increases, the patterns produced yield insight into the rates of change such that it is possible to determine degrees of appreciation or depreciation of incomes within those CTs which we already know to have changed in a particular manner. For example, patterns among the Glebe, Glebe-Annex, Old Ottawa South and Old Ottawa East are present in the ten category map, which are not under the three category map.

Difference maps are decent proxies for measuring gentrification. This is backed up through observations in the literature as to where gentrification has occurred and is occurring in Ottawa (Rubis 1982; Vachon 1992; Ley 1996; MacMillan 2010; Benali 2013; Leffers and Ballamingie 2013; Thayer 2014; Ilic, Sawada, and Zarzelli 2019). Using only income as the sole variable to measure gentrification is present in the literature (Filion 1991; McKinnish, Walsh, and Kirk White 2010), and it has been argued to be the most important variable for measuring such processes (Smith 1987, 463). This is plausible because the rise in incomes is almost always present when gentrification occurs (LeGates and Hartman 1982). Furthermore, income is the most important variable for socioeconomic status and as it is a scalar variable that is available across all census', it is both effective and convenient for measuring gentrification.

In the early 1990s, false predictions were made which postulated that gentrification was bound to decrease in relevance (Bourne 1993a; 1993b). In one such piece, Bourne (1993b, 186) suggested that a methodological failure when measuring gentrification is found in indices. While indices capture growth, they do not implicitly mean that class turnover occurred. Bourne observed this amongst some of the richest areas of Toronto. We also observe similar patterns, whereby significant income growth is visible affluent areas such as Rockcliffe Park and Rothwell Heights (Figure 2). However, these affluent areas do not exhibit change in difference maps, and as such Bourne's concern is addressed (Figure 5).

Finally, it is important to note that this study uses a constrained spatial domain reflective of Ottawa's inner-city and inner-suburbs. Hence, the study does not take into consideration edge cities in the form of new suburbs (for example, areas such as Barrhaven, Kanata and Orleans), nor places across the provincial border (the amalgamated city of Gatineau, whose center is situation in the Hull district). Therefore, this study does not take into account potential migration

whereby high-, middle- and low-income people change locations to these more distant areas. However, previous research shows somewhat results in terms of income polarization when the whole CMA is examined (Ilic and Sawada 2021).

## **Conclusion**

The present research is a contribution amongst many which indicate that Canadian cities are experiencing higher levels of inequality and socio-economic divides. Ottawa's spatial structure has significantly changed in the last several decades, and this has been exemplified using the various methods. The present research shows how different assumptions in the Three City Model yield different results. Therefore, the conclusions that this model yields are not very robust, nor can they be generalizable on all scales. However, the use of difference maps proves to be effective for measuring class turnover of census tracts.

Examining income Examining income polarization shows more robust results in terms of spatial patterns. This is not the first study, and likely will not be the last study in this nature in the Canadian context, but the methods posed can be applied elsewhere. The identified spatial patterns show worrying trends. For example, as many areas in Ottawa undergo a process of suburban decline, it is imperative to have appropriate and targeted planning in order to ameliorate such worsening trends taking place in select parts of the city. Likewise, it is important to have adequate planning to address the changes taking place in the central areas of the city. For example, adequate infrastructure (especially low-income housing) should be in place for the low-income population which is increasingly being displaced.

## S1. Appendix.

Supplementary information for Examining income trajectories and income polarization in Ottawa, 1971-2016.

### 15 Percent Band

1971		1981		1986		1991		1996		2001		2006		2016	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
37	33.04%	32	28.23%	27	24.32%	26	23.42%	30	27.27%	26	23.85%	30	27.03%	33	29.46%
54	48.21%	52	46.85%	56	50.45%	57	51.35%	47	42.73%	47	43.12%	41	36.94%	34	30.36%
21	18.75%	27	24.32%	28	25.23%	28	25.23%	33	30.00%	36	33.03%	40	36.04%	45	40.18%

### 20 Percent Band

1971		1981		1986		1991		1996		2001		2006		2016	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
29	25.89%	25	22.52%	21	18.92%	22	19.82%	21	19.09%	22	20.18%	26	23.42%	28	25.00%
67	59.82%	65	58.56%	70	63.06%	68	61.26%	65	59.09%	62	56.88%	59	53.15%	52	46.43%
16	14.29%	21	18.92%	20	18.02%	21	18.92%	24	21.82%	25	22.94%	26	23.42%	32	28.57%

### 25 percent band

1971		1981		1986		1991		1996		2001		2006		2016	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
25	22.32%	15	13.51%	16	14.41%	17	15.32%	18	16.36%	19	17.43%	22	19.82%	23	20.54%
77	68.75%	79	71.17%	83	74.78%	84	75.68%	74	67.27%	71	65.14%	68	61.26%	69	61.61%
10	8.93%	17	15.32%	12	10.81%	10	9.01%	18	16.36%	19	17.43%	21	18.92%	20	17.86%

## References

- Banting, Keith, and John Myles. 2013. "Introduction: Inequality and the Fading of Redistributive Politics." In *Inequality and the Fading of Redistributive Politics*, edited by Keith Banting and John Myles, 1–39. Vancouver: University of British Columbia Press.
- Beach, Charles M. 2016. "Changing Income Inequality: A Distributional Paradigm for Canada." *Canadian Journal of Economics* 49 (4): 1229–92.
- Benali, Kenza. 2013. "Urban Densification in Ottawa's Vanier District: A Sign of Urban Renewal or a Threat to the Last Remaining French-language Bastion in the Canadian Capital?" *Cahiers de Géographie Du Québec* 57 (160): 41–68.
- Bird, John, Barry Curtis, Tim Putnam, George Robertson, and Lisa Tickner, eds. 1993. *Mapping the Futures: Local Cultures, Global Change*. London: Routledge.
- Bourne, L.S. 1993a. "Close Together and Worlds Apart: An Analysis of Changes in the Ecology of Income in Canadian Cities." *Urban Studies* 30 (8): 1293–1317.
- Bourne, L.S. 1993b. "The Myth and Reality of Gentrification: A Commentary on Emerging Urban Forms." *Urban Studies* 30 (1): 183–89.
- Breau, S. 2015. "Rising Inequality in Canada: A Regional Perspective." *Applied Geography* 61: 58–69.
- Buitelaar, Edwin, Anet Weterings, and Roderik Ponds. 2018. *Cities, Economic Inequality and Justice: Reflections and Alternative Perspectives*. London: Routledge.
- Davis, Mike. 1990. *City of Quartz: Excavating the Future in Los Angeles*. London: Verso.
- Distasio, Jino, and Andrew Kaufman. 2015. "Neighbourhoods of Change." In *The Divided Prairie City: Income Inequality Among Winnipeg's Neighbourhoods, 1970–2010*, edited by Jino Distasio and Andrew Kaufman, 2–12. Winnipeg: Institute of Urban Studies, The University of Winnipeg.
- Dorling, Danny. 2015. *Inequality and the 1%*. London: Verso.
- Fainstein, Susan, Ian Gordon, and Michael Harloe, eds. 1992. *Divided Cities: New York & London in the Contemporary World*. Oxford: Blackwell.
- Filion, Pierre. 1991. "The Gentrification-Social Structure Dialectic: A Toronto Case Study." *International Journal of Urban and Regional Research* 15 (4): 553–74.
- Fong, Francis. 2017. *Income Inequality in Canada: The Urban Gap*. Toronto: Chartered Professional Accountants Canada.

- Fortin, Nicole, David A. Green, Thomas Lemieux, Kevin Milligan, and W. Craig Riddell. 2012. "Canadian Inequality: Recent Developments and Policy Options." *Canadian Public Policy* 38 (2): 121–45.
- Grant, Jill L, Alan Walks, and Howard Ramos, eds. 2020. *Changing Neighbourhoods: Social and Spatial Polarization in Canadian Cities*. Vancouver: University of British Columbia Press.
- Grant, Jill L, Alan Walks, and Howard Ramos. 2020. "Evaluating Neighbourhood Inequality and Change: Lessons from a National Comparison." In *Changing Neighbourhoods: Social and Spatial Polarization in Canadian Cities*, edited by Jill L Grant, Alan Walks, and Howard Ramos, 252–79. Vancouver: UBC Press.
- Hackworth, Jason. 2005. "Emergent Urban Forms, or Emergent Post-Modernisms? A Comparison of Large U.S. Metropolitan Areas." *Urban Geography* 26 (6): 484–519.
- Hackworth, Jason. 2007. *The Neoliberal City: Governance, Ideology, and Development in American Urbanism*. Ithaca: Cornell University Press.
- Hamnett, Chris. 2003. *Unequal City: London in the Global Arena*. London: Routledge.
- Harris, Richard, Jim Dunn, and Sarah Wakefield. 2015. "A City on the Cusp: Neighbourhood Change in Hamilton since 1970." Research Paper 236. Toronto: Neighbourhood Change Research Partnership. <http://neighbourhoodchange.ca/documents/2015/04/neighbourhood-change-in-hamilton-since-1970.pdf>.
- Hulchanski, David. 2007. "The Three Cities within Toronto: Income Polarization among Toronto's Neighbourhoods, 1970 to 2000." Research Bulletin. Toronto: University of Toronto Centre for Urban & Community Studies.
- Hulchanski, David. 2010. "The Three Cities Within Toronto: Income Polarization Among Toronto's Neighbourhoods, 1970–2005." Toronto: University of Toronto Cities Centre.
- Ilic, Lazar, M. Sawada, and Amaury. Zarzelli. 2019. "Deep Mapping Gentrification in a Large Canadian City Using Deep Learning and Google Street View." *PLoS ONE* 14 (3).
- Ilic, Lazar, and M. Sawada. 2021. "The Temporal Evolution of Income Polarization in Canada's Largest CMAs." *PLoS ONE* 16 (6).
- Leffers, Donald, and Patricia Ballamingie. 2013. "Governmentality, Environmental Subjectivity, and Urban Intensification." *Local Environment* 18 (2): 134–51.
- LeGates, Richard, and Chester Hartman. 1982. "Gentrification-Caused Displacement." *The Urban Lawyer* 14 (1): 31–55.

- Ley, David. 1996. *The New Middle Class and the Remaking of the Central City*. Oxford: Oxford University Press.
- Ley, David, and Nicholas Lynch. 2012. "Divisions and Disparities in Lotus-Land: Socio-Spatial Income Polarization in Greater Vancouver, 1970-2005." Research Paper No. 223. Toronto: Cities Centre, University of Toronto. <http://neighbourhoodchange.ca/documents/2012/10/divisions-and-disparities-in-lotus-land-socio-spatial-income-polarization-in-greater-vancouver-1970-2005-by-david-ley-nicholas-lynch.pdf>.
- MacLachlan, Ian, and Ryo Sawada. 1997. "Measures of Income Inequality and Social Polarization in Canadian Metropolitan Areas." *Canadian Geographer* 41 (4): 377–97.
- MacMillan, Blake. 2010. "ANT in Westboro: An Examination of the Applicability of Actor-Network Theory in Gentrification Studies." Ottawa: Carleton University. <https://curve.carleton.ca/ef48c8ac-8f21-4932-8890-3523e0f0491f>.
- McGuire, Liam. 2012. "The Ten Cities of Toronto: Patterns of Socio-Economic Inequality and Polarization Throughout The Toronto Census Metropolitan Area." Masters Thesis, Vancouver: University of British Columbia. [http://ibis.geog.ubc.ca/~ewyly/students/McGuire\(2012\).pdf](http://ibis.geog.ubc.ca/~ewyly/students/McGuire(2012).pdf).
- McKinnish, Terra, Randal Walsh, and T. Kirk White. 2010. "Who Gentrifies Low-Income Neighborhoods?" *Journal of Urban Economics* 67 (2): 180–93.
- Mollenkopf, John Hull, and Manuel Castells, eds. 1991. *Dual City: Restructuring New York*. New York: Russel Sage Foundation.
- Myles, John, Garnett Picot, and Wendy Pyper. 2000. "Neighbourhood Inequality in Canadian Cities." 160. Analytical Studies Branch Research Paper Series. Ottawa: Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2000160-eng.pdf>.
- Osberg, Lars. 2008. *Quarter Century of Economic Inequality in Canada: 1981-2006*. Toronto: Canadian Centre for Policy Alternatives.
- Prouse, Victoria, Jill L Grant, Martha Radice, Howard Ramos, and Paul Shakotko. 2014. "Neighbourhood Change in Halifax Regional Municipality, 1970 to 2010: Applying the 'Three Cities' Model." Halifax: Dalhousie University. <http://neighbourhoodchange.ca/documents/2014/03/neighbourhood-change-in-halifax-regional-municipality-1970-to-2010-applying-the-three-cities-model.pdf>.
- Rose, Damaris, and Amy Twigge-Molecey. 2013. "A City-Region Growing Apart? Taking Stock of Income Disparity in Greater Montréal, 1970-2005." Research Paper 222. Toronto: Cities Centre, University of Toronto. <https://tspace.library.utoronto.ca/bitstream/1807/94389/1/222%20Rose%20CityRegion%202013.pdf>.
- Rubis, Darryl. 1982. "The Changing Character of Ottawa's Golden Triangle." Kingston: Queen's

University.

- Sassen, Saskia. 1991. *The Global City: New York, London, Tokyo*. Princeton: Princeton University Press.
- Smith, Neil. 1987. "Commentary: Gentrification and the Rent Gap." *Annals of the Association of American Geographers* 77 (3): 462–65.
- Thayer, Jill. 2014. "What Has Happened to the Neighbourhood? Gentrification in Hintonburg and Mechanicsville: Theories and Approaches." Ottawa: Carleton University.
- Townshend, Ivan, Byron Miller, and Leslie Evans. 2018. "Socio-Spatial Polarization in an Age of Income Inequality: An Exploration of Neighbourhood Change in Calgary's 'Three Cities.'" Research Paper 241. Toronto: Neighbourhood Change Research Partnership. <http://neighbourhoodchange.ca/documents/2018/04/socio-spatial-polarization-in-calgary.pdf>.
- Vachon, Marc. 1992. "The Morphogenesis of Bank Street in Ottawa, Canada." *Geoscope* 20 (2): 1–12.
- Vinodrai, Tara, and Markus Moos. 2015. "Appendix A: Do We Still Have Quality Data to Study Canadian Cities?" In *Canadian Cities in Transition*, edited by Pierre Filion, Markus Moos, Tara Vinodrai, and Ryan Walker, 5th ed., 436–39. Don Mills: Oxford University Press.
- Walks, Alan. 2007. "The Boundaries of Suburban Discontent? Urban Definitions and Neighbourhood Political Effects." *Canadian Geographer* 51 (2): 160–85.
- Watson, Sophie, and Katherine Gibson, eds. 1995. *Postmodern Cities and Spaces*. Oxford: Blackwell.

## **Appendix II. Supplementary information for ‘The Temporal Evolution of Income Polarization in Canada’s Largest CMAs’**

### **Contents**

**S1. Three City Project Traction**

**S2. Population ranges of CTs in CMAs.**

**S3. Comparison of four geographic scales**

**S4. Full census variable names**

**S5. Shapefile errors and corrections**

**S6. Fragmentation Index tests on Spatial scales**

**S7. Maps of CMAs**

**S8. Confidence intervals of bootstrapped Fragmentation Index data**

**S9. Confidence intervals of bootstrapped Joins-Count data (Queens spatial weight matrix)**

**S10. Trends of Joins-Count data and Confidence intervals of bootstrapped Joins-Count data (Rooks spatial weight matrix)**

**S11. Trends of Joins-Count data and Confidence intervals of bootstrapped Joins-Count data (KNN-5 weight matrix)**

**S12. Inner City Extents**

## S1. Three City Project Traction

To examine traction of the Three City Project, data was tabulated per individual study per year. The data source was Google Scholar, and was obtained on August 1, 2019.

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2007-2018
<b>Toronto</b>	0	10	16	14	26	28	44	46	75	43	67	60	403
<b>Vancouver</b>							1	5	10	8	5	8	37
<b>Montreal</b>								2	2	2		3	9
<b>Halifax</b>								1	5	2	3	2	13
<b>Hamilton</b>									1	4	6	6	17
<b>Calgary</b>												1	1
<b>Total:</b>	0	10	16	14	26	28	45	54	93	59	81	80	

The search criteria for each city was simple the name of each study. Below are the search criteria and the links which they yielded. Once the link was obtained, the “custom range” option allowed for searching for total citation numbers in individual years.

Toronto: “The three cities within Toronto”

[https://scholar.google.com/scholar?cites=17836097183110024094&as\\_sdt=2005&scioldt=0,5&hl=en](https://scholar.google.com/scholar?cites=17836097183110024094&as_sdt=2005&scioldt=0,5&hl=en)

Vancouver: “Divisions and disparities in lotus-land: Socio-spatial income polarization in greater Vancouver, 1970-2005”

<https://scholar.google.com/scholar?client=firefox-b-d&um=1&ie=UTF-8&lr&cites=799435445609479635>

Montreal: “A City-Region Growing Apart? Taking Stock of Income Disparity in Greater Montréal, 1970-2005”

[https://scholar.google.com/scholar?sxsrf=ACYBGNSNpR4u0t8Wp8fOCauWac2MIlwY\\_w:1579500673657&uact=5&um=1&ie=UTF-8&lr&cites=9157432325877169501](https://scholar.google.com/scholar?sxsrf=ACYBGNSNpR4u0t8Wp8fOCauWac2MIlwY_w:1579500673657&uact=5&um=1&ie=UTF-8&lr&cites=9157432325877169501)

Hamilton: “A city on the cusp: Neighbourhood change in Hamilton since 1970”

<https://scholar.google.com/scholar?client=firefox-b-d&um=1&ie=UTF-8&lr&cites=14588784678578239141>

Halifax : “Neighbourhood Change in Halifax Regional Municipality, 1970 to 2010: Applying the “Three Cities” Model”

<https://scholar.google.com/scholar?client=firefox-b-d&um=1&ie=UTF-8&lr&cites=9530011185545756078>

Calgary: “Socio-Spatial Polarization in an Age of Income Inequality: An Exploration of Neighbourhood Change in Calgary’s “Three Cities””

[https://scholar.google.com/scholar?client=firefox-b-d&sxsrf=ACYBGNTd\\_Pm7830yZPAoXasf7egA\\_57X\\_A:1579500866522&uact=5&um=1&ie=UTF-8&lr&cites=3745678424521638426](https://scholar.google.com/scholar?client=firefox-b-d&sxsrf=ACYBGNTd_Pm7830yZPAoXasf7egA_57X_A:1579500866522&uact=5&um=1&ie=UTF-8&lr&cites=3745678424521638426)

## S2. Population ranges of CTs in CMAs.

### a. Calgary

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	77	985	11290	73	94.805
1981	115	1342	14987	104	90.435
1986	143	2117	9546	133	93.007
1991	153	2129	9160	144	94.118
1996	153	2091	17323	132	86.275
2001	193	617	11798	176	91.192
2006	202	308	20946	173	85.644
2016	253	1196	23798	218	86.166

### b. Edmonton

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	88	155	25735	68	77.273
1981	142	2	9907	121	85.211
1986	180	41	9119	156	86.667
1991	186	69	10424	166	89.247
1996	187	45	12078	158	84.492
2001	204	41	13390	177	86.765
2006	219	43	21628	171	78.082
2016	268	5	28192	204	76.119

### c. Montreal

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	570	45	20515	414	72.632
1981	661	4	14584	478	72.315
1986	709	122	13275	528	74.471
1991	738	120	15180	527	71.409
1996	757	261	20845	532	70.277
2001	850	160	11422	664	78.118
2006	864	246	11912	674	78.009
2016	959	5	20786	744	77.581

## d. Ottawa-Gatineau

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	120	95	20240	90	75.000
1981	178	25	7598	151	84.831
1986	191	45	10084	162	84.817
1991	208	56	10468	176	84.615
1996	215	45	12210	172	80.000
2001	236	50	16103	201	85.169
2006	248	574	12049	204	82.258
2016	277	5	16747	221	79.783

## e. Quebec City

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	98	60	14680	56	57.143
1981	126	154	12695	89	70.635
1986	138	300	11182	103	74.638
1991	152	284	10080	122	80.263
1996	152	240	12369	115	75.658
2001	165	230	7888	132	80.000
2006	165	225	9686	132	80.000
2016	181	61	9935	133	73.481

## f. Toronto

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	447	35	20120	290	64.877
1981	606	11	20583	494	81.518
1986	726	47	14864	629	86.639
1991	808	59	14213	700	86.634
1996	808	58	25437	675	83.540
2001	929	47	20635	820	88.267
2006	998	51	22724	886	88.778
2016	1150	10	23401	975	84.783

## g. Vancouver

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	179	75	24475	118	65.922
1981	245	560	10513	212	86.531
1986	273	137	13317	239	87.546
1991	298	162	12215	261	87.584
1996	298	168	19736	237	79.530
2001	386	155	11914	334	86.528
2006	409	145	13367	357	87.286
2016	474	10	13855	403	85.021

## h. Winnipeg.

Year	Total CTs w Pop	Low	High	Tracts wPop 2500 to 8000	Perc of Total
1971	106	90	12320	79	74.528
1981	135	9	13779	104	77.037
1986	147	457	11191	121	82.313
1991	155	410	7907	129	83.226
1996	157	387	10049	124	78.981
2001	164	373	8329	135	82.317
2006	167	231	9739	135	80.838
2016	173	228	20465	140	80.925

### S3. Comparison of four geographic scales

Appearance of search result counts for terms from select journals.

Many of these journals are in the top ten impact factor journals for urban studies related research. Additional journals in the list are ones in which much urban research is published, but which are not specifically limited to urban themes.

<b>Journal Name</b>	<b>Census Tract</b>	<b>Block Group</b>	<b>Dissemination Area</b>	<b>Enumeration Area</b>
Annals of the American Association of Geographers	156	51	3	8
Canadian Geographer	90	4	16	17
City	2	0	0	0
Cities	121	27	5	4
International Journal of Urban and Regional Research	42	4	1	2
Journal of Urban Economics	278	51	0	2
Professional Geography	147	42	1	5
Regional Science and Urban Economics	173	47	1	2
Urban Geography	262	55	3	6
Urban Studies	315	76	5	4
Urban Affairs Review	144	34	0	0

## S4. Full census variable names

### 1971 Census :

- Households: Income from all sources: under \$3,000
- Households: Income from all sources: under \$3,000-\$4,999
- Households: Income from all sources: under \$5,000-\$6,999
- Households: Income from all sources: under \$7,000-\$9,999
- Households: Income from all sources: under \$10,000-\$14,999
- Households: Income from all sources: under \$15,000 & over
- Households: Average income from all sources

### 1981 Census :

- Private household income - all households
- Private household income - all households - average income

### 1986 Census :

- Household income - all private households
- Average income (note: from category, "Household income - all private households")

### 1991 Census :

- Household income - all private households
- Average income, household income \$ (note: from category, "Household income - all private households")

### 1996 Census :

- Household income of private households (20% sample data)
- Average household income \$ (note: from category, "Household income of private households (20% sample data)")

### 2001 Census :

- Household income in 2000 of private households - 20% sample data
- Average household income \$ (note: from category, "Household income in 2000 of private households - 20% sample data")

### 2006 Census :

- Household income in 2005 of private households - 20% sample data
- Average household income \$ (note: from category, "Household income in 2005 of private households - 20% sample data")

### 2016 Census :

- Total – Income statistics in 2015 for private households by household size – 25% sample data
- Average total income of households in 2015 (\$)

### Sources:

1971, 1981, 1986, 1991, 1996, 2001, 2006: Canadian Census Analyzer database

2016 : Statistics Canada IVT table: “98-401-X2016043” Census Profile - Age, Sex, Type of Dwelling, Families, Households, Marital Status, Language, Income, Immigration and Ethnocultural Diversity, Housing, Aboriginal Peoples, Education, Labour, Journey to Work, Mobility and Migration, and Language of Work for Census Metropolitan Areas, Tracted Census Agglomerations and Census Tracts, 2016 Census

## **S5. Shapefile errors and corrections**

Shapefiles were created by Statistics Canada and are available on the Scholars Geoportal Database, located at <http://geo2.scholarsportal.info> . The following CMAs in the select years had errors which were corrected.

### **Montreal 1971.**

Tract 852 – This CT had two polygons. One part was mislabeled. Fixed by merging them.

### **Montreal 1981.**

Tract 411 – This CT was mislabeled as CT 418. Fixed by renaming.

Tract 560 – There were two CTs with this name. Fixed by merging.

Tract 584 – This CT was missing. It had to be redrawn.

Tract 756.01 – There were two polygons with this name. Fixed by merging.

Tract 825.01 – This CT was mislabeled as tract 852.01... Fixed by renaming.

Tract 886 – There were two polygons with this name. Fixed by merging.

Tract 887.01 – There were two polygons with this name. Fixed by merging.

Tract 887.02 – This CT was mislabeled as CT 887.01. Fixed by renaming.

Tract 904 – There were two polygons with this name. Fixed by merging.

### **Ottawa 1981.**

Tract 151.02 – contained two tracts by mistake. Fixed using the explode tool and subsequent renaming of new polygons into CTs 151.02 and 151.03.

### **Vancouver 1981.**

Tract 200 – This CT was missing. Fixed by using explode tool and subsequent renaming of new polygons into CTs 200 and 202.

Tract 237 – This CT was missing or merged into CT 243. Fixed by redrawing.

Tract 141 – This CT was mislabeled. There were two CTs labeled 140. Fixed by renaming. Tract 141 is further west than tract 140.

Tract 285.01 – This CT was mislabeled as CT 235.01. Fixed by renaming.

### **Toronto 1981.**

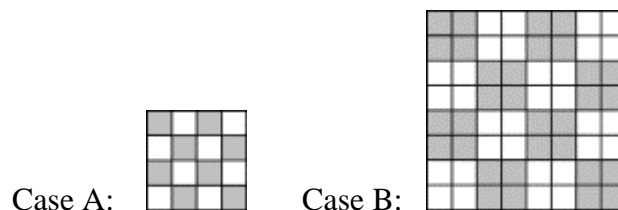
Tract 165 – This CT was mislabeled as CT 166. Fixed by renaming.

Tract 203 – This CT was mislabeled as CT 208. Fixed by renaming.

Tract 516.08 – This CT mislabeled as CT 516.07. Fixed by renaming.

### S6. Fragmentation Index tests on Spatial scales

The two following cases were examined to test the Johnson Fragmentation Index and Edge Density. Both examples prove inadequate in situations where the spatial scale is very different.



Johnson's Fragmentation Index gives a result of 1 for Case A (15 / 15) and 0.238 for Case B (15 / 63).

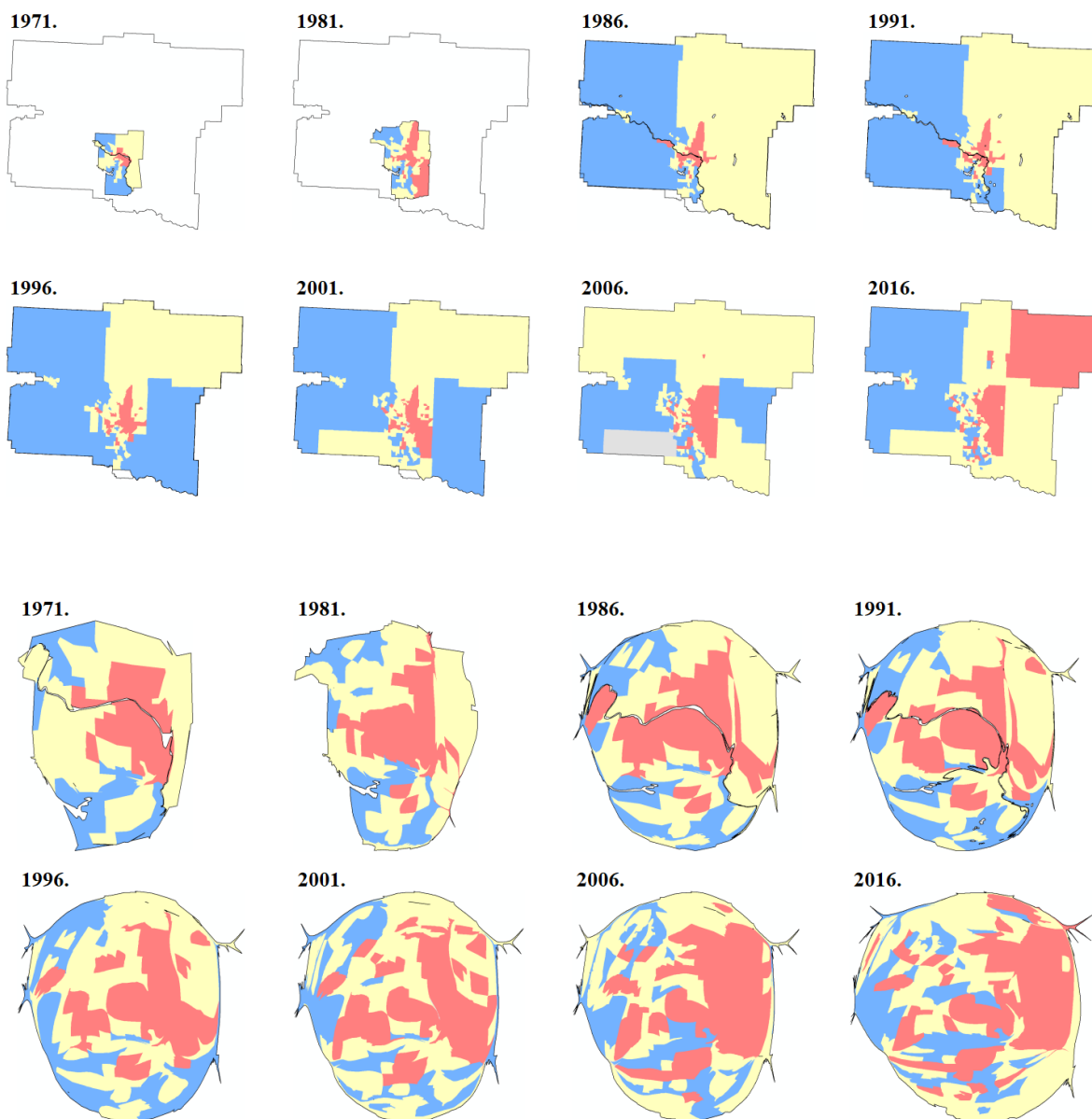
Edge Density gives a result of 4 for Case A (64 / 16) and 2 for Case B (128 / 64).

The two cases should have the same results as visually they depict the exact same pattern.

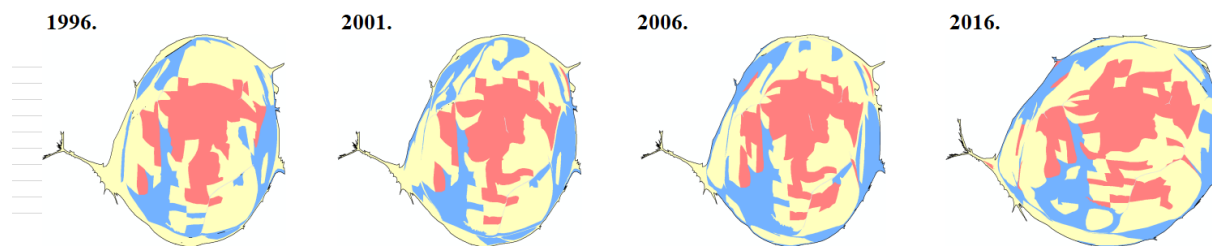
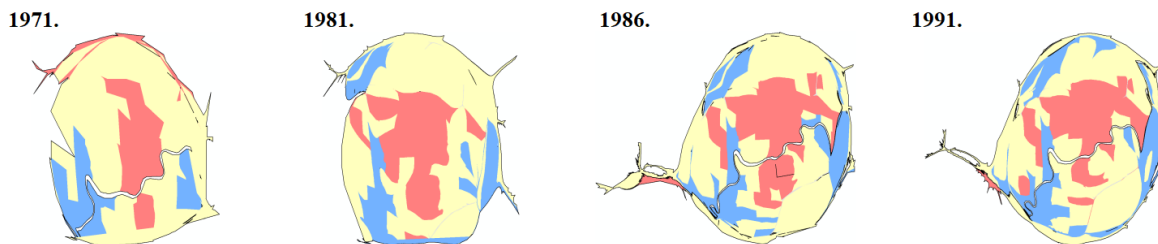
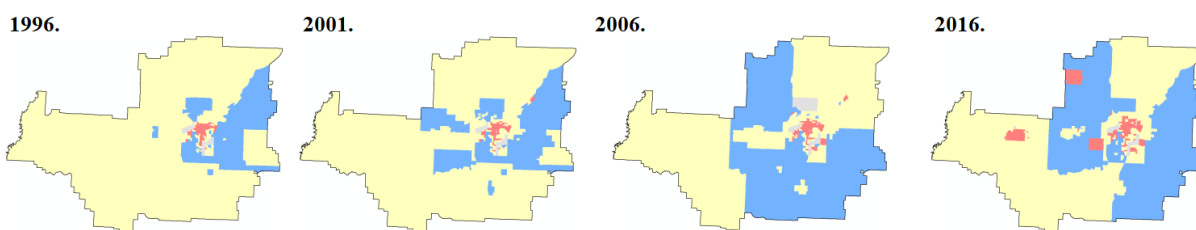
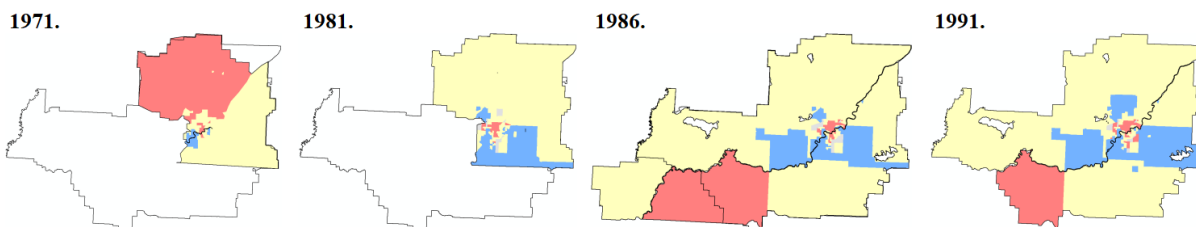
## S7. Maps

Higher resolution maps are available at the following GitHub page:  
<https://github.com/lazarification>

### a. Calgary

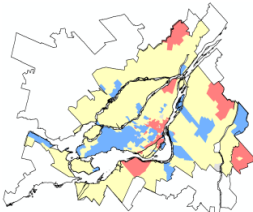


b. Edmonton

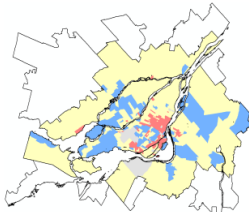


## c. Montreal

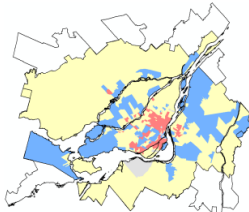
1971.



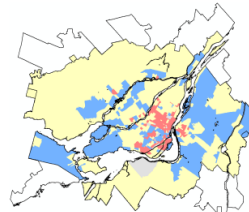
1981.



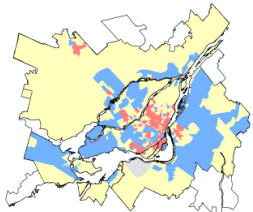
1986.



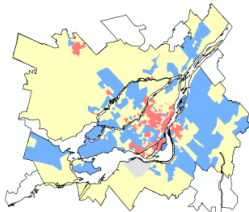
1991.



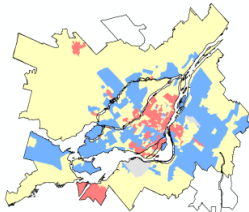
1996.



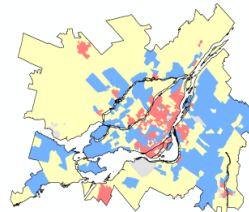
2001.



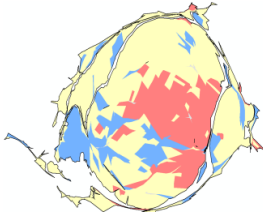
2006.



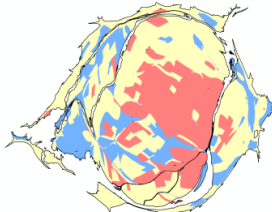
2016.



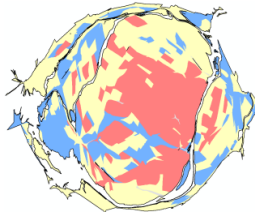
1971.



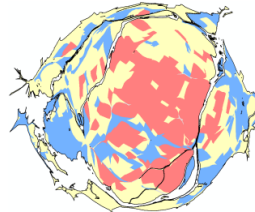
1981.



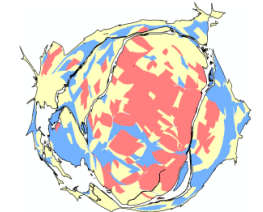
1986.



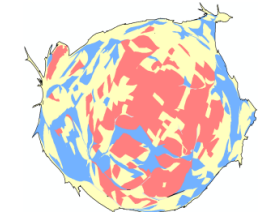
1991.



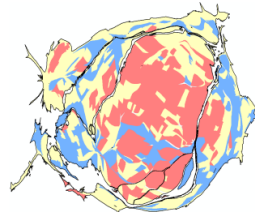
1996.



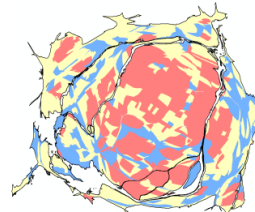
2001.



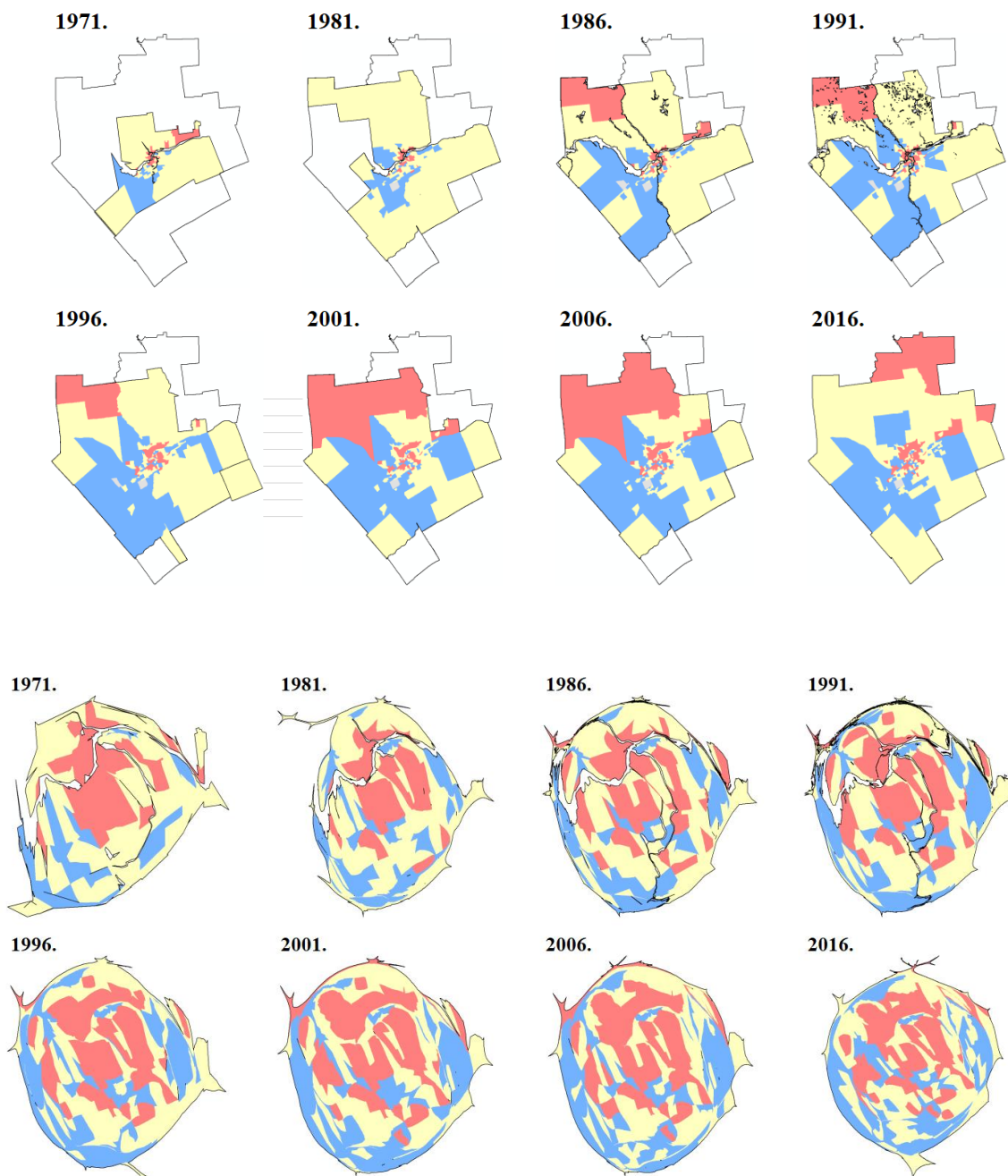
2006.



2016.



## d. Ottawa-Gatineau

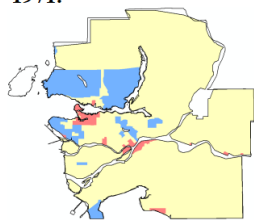




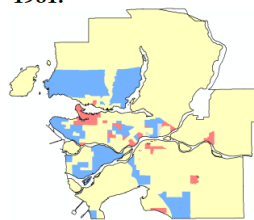


## g. Vancouver

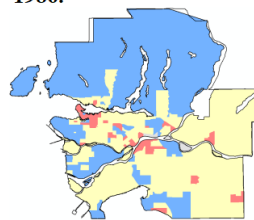
1971.



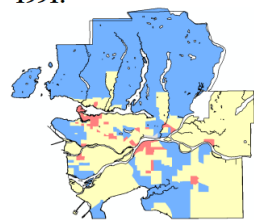
1981.



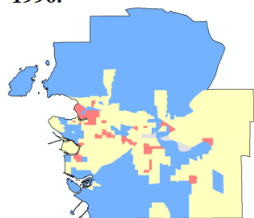
1986.



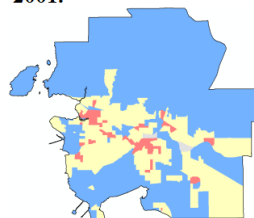
1991.



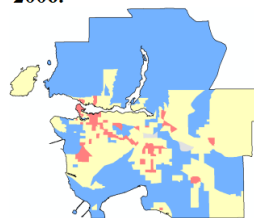
1996.



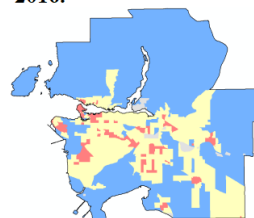
2001.



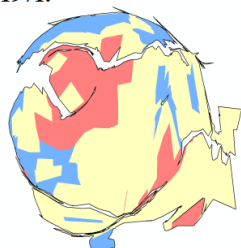
2006.



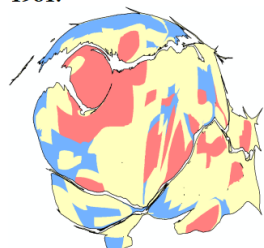
2016.



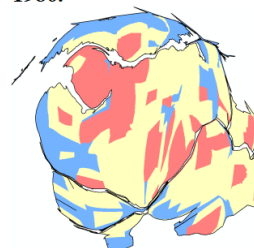
1971.



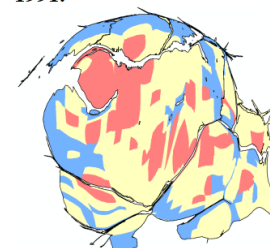
1981.



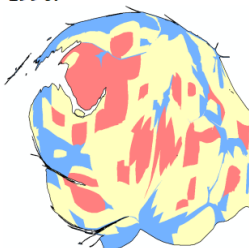
1986.



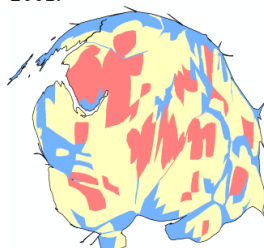
1991.



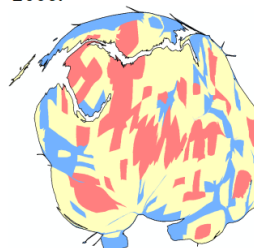
1996.



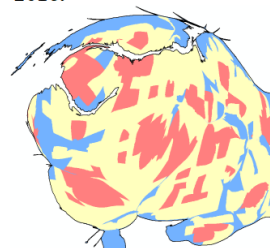
2001.



2006.

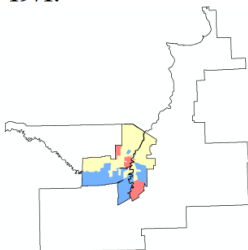


2016.

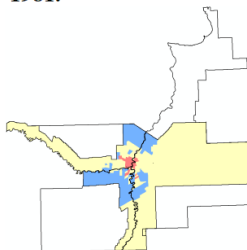


## h. Winnipeg

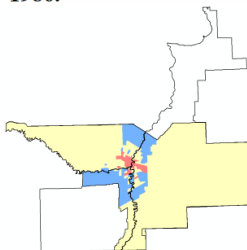
1971.



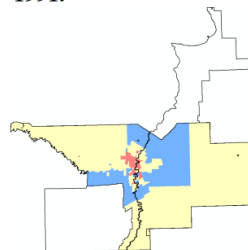
1981.



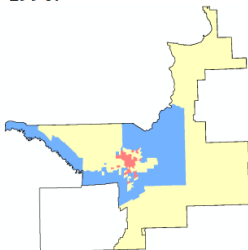
1986.



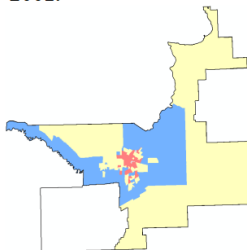
1991.



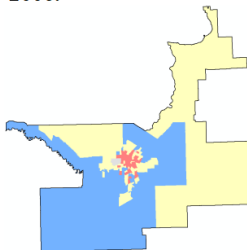
1996.



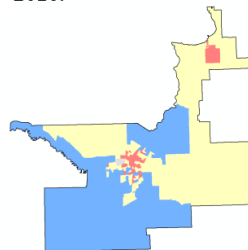
2001.



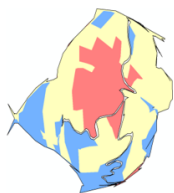
2006.



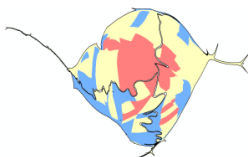
2016.



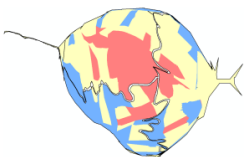
1971.



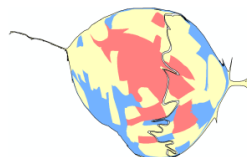
1981.



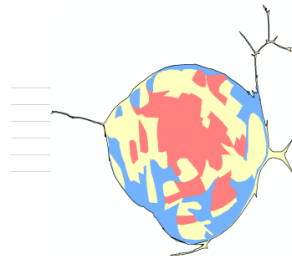
1986.



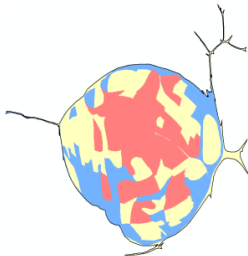
1991.



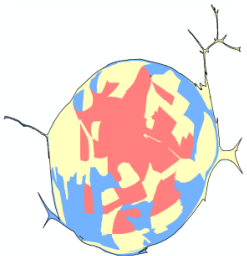
1996.



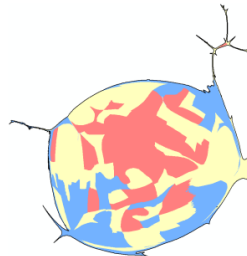
2001.



2006.



2016.



### S8. Confidence intervals of bootstrapped Fragmentation Index data

Slopes of two Fragmentation Indices (Johnsson Index and Edge Density) and their associated 95% confidence intervals (CI) obtained via non-parametric bootstrapping. All slopes are statistically significant.

<b>CMA</b>	<b>JI (95% CI)</b>	<b>CI<sub>ED</sub></b>
Calgary	1.47 [1.14, 1.80]	0.013 [0.003, 0.022]
Edmonton	0.10 [0.07, 0.12]	0.003 [0.002, 0.004]
Montreal	0.82 [0.65, 0.99]	0.011 [0.009, 0.012]
Ottawa-Gatineau	0.37 [0.19, 0.56]	0.008 [0.007, 0.009]
Quebec City	0.48 [0.29, 0.67]	0.008 [0.007, 0.009]
Toronto	0.77 [0.44, 1.10]	0.011 [0.005, 0.016]
Vancouver	0.24 [0.15, 0.33]	0.006 [0.004, 0.008]
Winnipeg	0.57 [0.42, 0.73]	0.006 [0.003, 0.009]

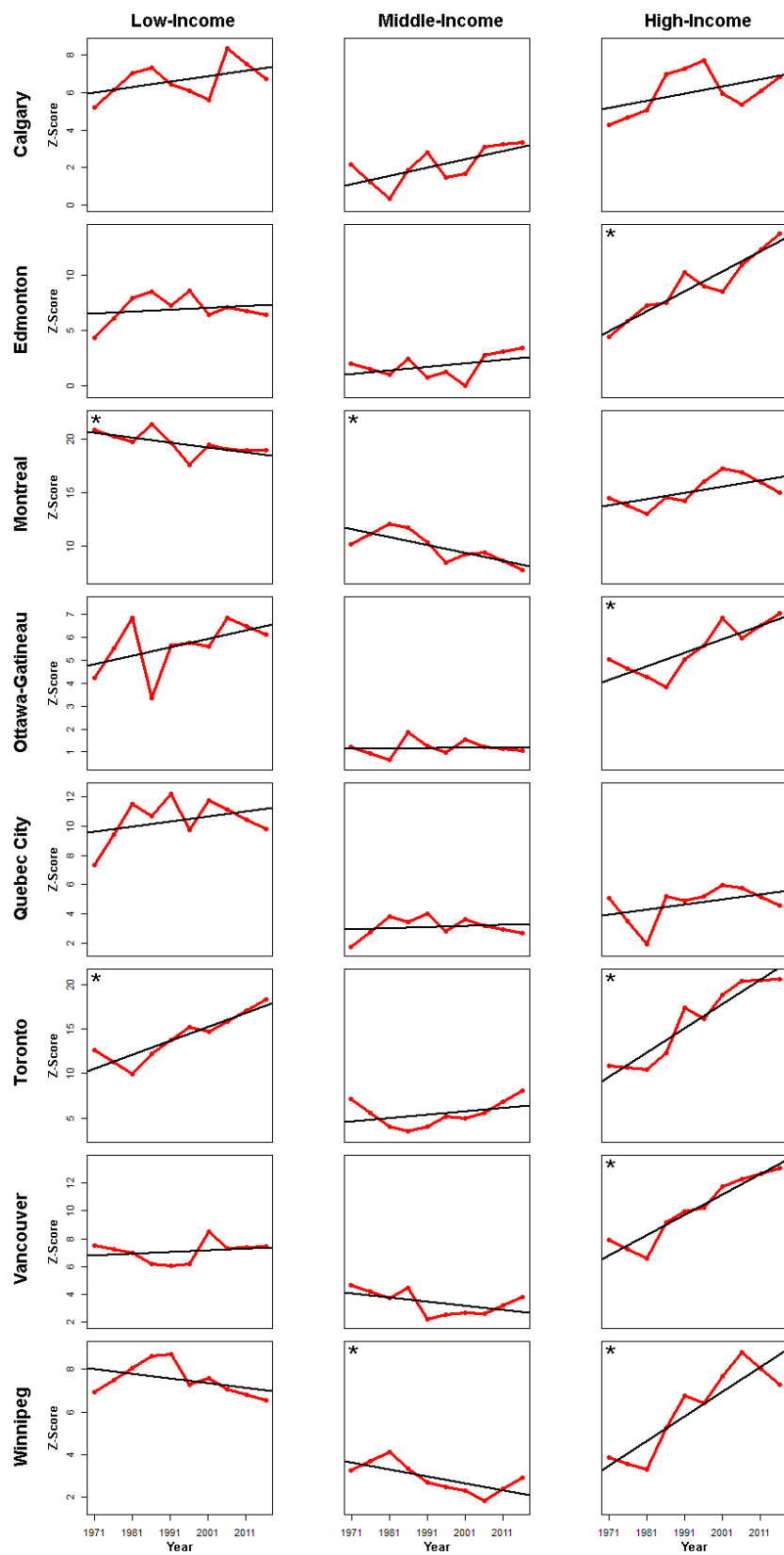
### S9. Confidence intervals of bootstrapped Joins-Count data

Slopes of the three groups (Low, Middle & High) and their associated 95% confidence intervals (CI) obtained via non-parametric bootstrapping. A slope is significant statistically when its confidence interval does not include zero and these are indicated by an \*.

<b>CMA</b>	<b>Low (95% CI)</b>	<b>CI<sub>mid</sub></b>	<b>CI<sub>high</sub></b>
Calgary	0.03 [-0.02, 0.07]	0.04 [-0.00, 0.08]	0.04 [-0.02, 0.09]
Edmonton	0.02 [-0.05, 0.08]	0.03 [-0.02, 0.09]	0.18 [0.13, 0.22]*
Montreal	-0.05 [-0.09, -0.00]*	-0.07 [-0.12, -0.03]*	0.06 [-0.00, 0.12]
Ottawa-Gatineau	0.04 [-0.01, 0.09]	0.00 [-0.02, 0.02]	0.06 [0.03, 0.09]*
Quebec City	0.03 [-0.04, 0.10]	0.01 [-0.03, 0.04]	0.04 [-0.02, 0.09]
Toronto	0.16 [0.10, 0.22]*	0.04 [-0.04, 0.11]	0.27 [0.20, 0.35]*
Vancouver	0.01 [-0.03, 0.05]	-0.03 [-0.07, 0.00]	0.15 [0.11, 0.19]*
Winnipeg	-0.02 [-0.06, 0.01]	-0.03 [-0.06, -0.01]*	0.12 [0.07, 0.16]*

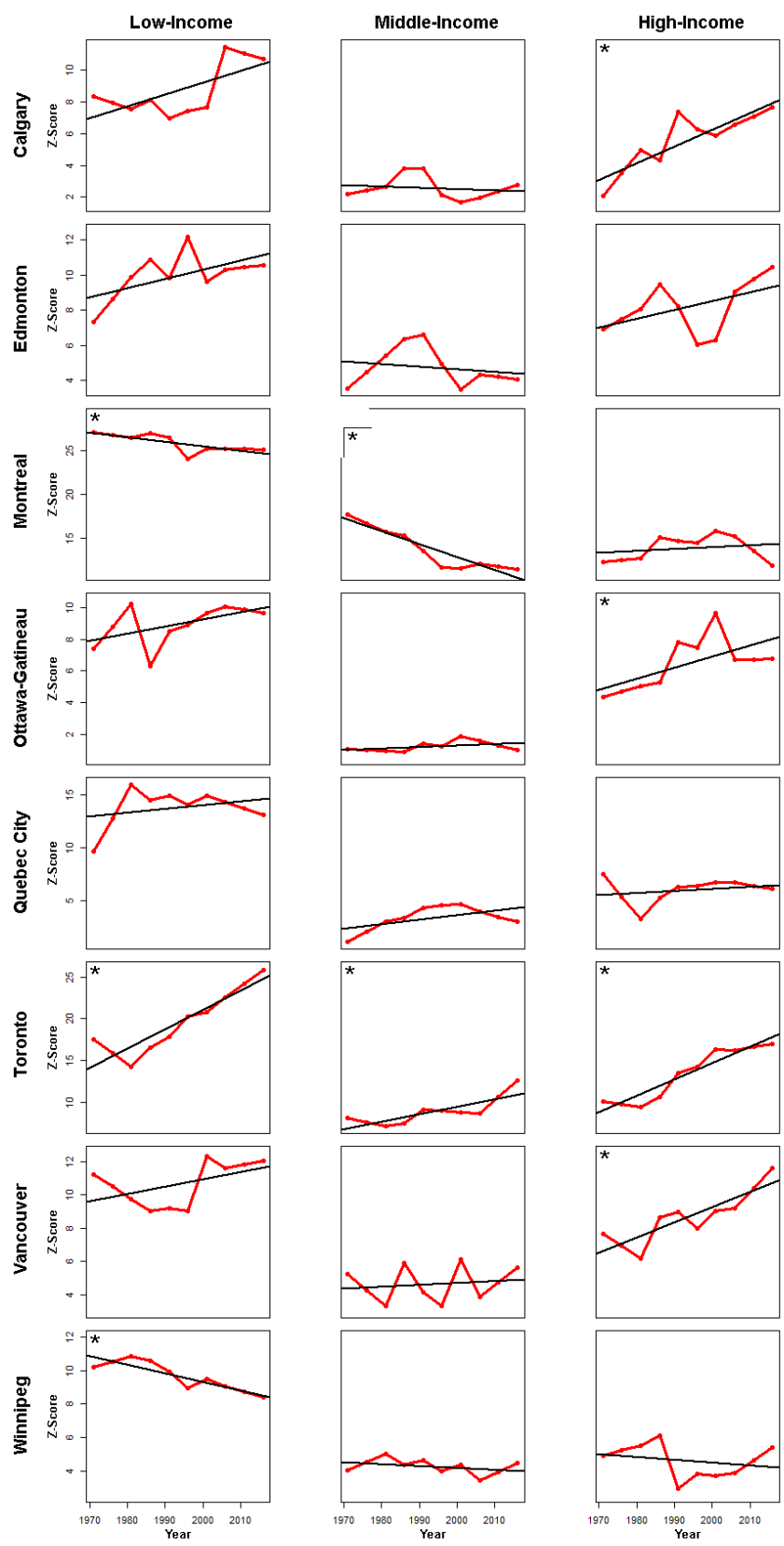
**S10. Trends of Joins-Count data and Confidence intervals of bootstrapped Joins-Count data (Rooks spatial weight matrix)**

<b>CMA</b>	<b>Low (95% CI)</b>	<b>CI<sub>mid</sub></b>	<b>CI<sub>high</sub></b>
Calgary	0.03 [-0.02, 0.07]	0.04 [-0.00, 0.08]	0.04 [-0.02, 0.09]
Edmonton	0.02 [-0.05, 0.08]	0.03 [-0.02, 0.09]	0.18 [0.13, 0.22]*
Montreal	-0.05 [-0.09, -0.00]*	-0.07 [-0.12, -0.03]*	0.06 [-0.00, 0.12]
Ottawa-Gatineau	0.04 [-0.01, 0.09]	0.00 [-0.02, 0.02]	0.06 [0.03, 0.09]*
Quebec City	0.03 [-0.04, 0.10]	0.01 [-0.03, 0.04]	0.04 [-0.02, 0.09]
Toronto	0.16 [0.10, 0.22]*	0.04 [-0.04, 0.11]	0.27 [0.20, 0.35]*
Vancouver	0.01 [-0.03, 0.05]	-0.03 [-0.07, 0.00]	0.15 [0.11, 0.19]*
Winnipeg	-0.02 [-0.06, 0.01]	-0.03 [-0.06, -0.01]*	0.12 [0.07, 0.16]*



**S11. Trends of Joins-Count data and Confidence intervals of bootstrapped Joins-Count data (KNN-5 spatial weight matrix)**

<b>CMA</b>	<b>Low (95% CI)</b>	<b>CI<sub>mid</sub></b>	<b>CI<sub>high</sub></b>
Calgary	0.07 [0.00, 0.14]	-0.01 [-0.05, 0.03]	0.11 [0.06, 0.15]*
Edmonton	0.05 [-0.00, 0.11]	-0.02 [-0.07, 0.04]	0.05 [-0.02, 0.12]
Montreal	-0.05 [-0.09, -0.02]*	-0.15 [-0.19, -0.10]*	0.02 [-0.05, 0.09]
Ottawa-Gatineau	0.05 [-0.01, 0.10]	0.01 [-0.01, 0.02]	0.07 [0.00, 0.14]*
Quebec City	0.03 [-0.05, 0.12]	0.04 [-0.00, 0.09]	0.02 [-0.04, 0.08]
Toronto	0.23 [0.15, 0.31]*	0.09 [0.04, 0.14]*	0.20 [0.14, 0.25]*
Vancouver	0.04 [-0.02, 0.10]	0.01 [-0.04, 0.07]	0.09 [0.05, 0.13]*
Winnipeg	-0.05 [-0.07, -0.03]*	-0.01 [-0.03, 0.01]	-0.02 [-0.07, 0.03]



## S12. Inner City Extents

Numerous studies have analyzed the Canadian inner-city. However, there is not a consensus as the extent of the inner-city. In the Canadian context, the inner-city is an area that is usually defined by having a large percentage of housing stock which was built before a particular date which is usually around World War II.

The census contains data on the age of the housing stock. Hence, one can see how many housing units were built during and prior particular periods of time. Using data at the census tract level, it is possible to discern which areas have an older housing stock that is reflective of the “inner-city”.

In the present study the goal is not to determine the most precise definition of what constitutes a particular CMA’s “inner-city” region. However, the maps which were produced are based on insights from the following sources:

- Bourne, L.S. 1992. Population turnaround in the Canadian inner city: Contextual factors and social consequences, *Canadian Journal of Urban Research*, 1, pp. 69-92.
- Broadway, Michael. 1992. DIFFERENCES IN INNER-CITY DEPRIVATION: AN ANALYSIS OF SEVEN CANADIAN CITIES. *Canadian Geographer / Le Géographe canadien*, 36: 189-196. doi:10.1111/j.1541-0064.1992.tb01131.x
- Broadway, Michael & Jesty, Gillian. 1998. Are Canadian Inner Cities Becoming More Dissimilar? An Analysis of Urban Deprivation Indicators. *Urban Studies*, 35(9), 1423–1438. <https://doi.org/10.1080/0042098984213>
- Carter, Tom., Polevychok, Chesya. & Sargent, Kurt. 2003. Is Winnipeg’s Aboriginal Population Ghettoized? Canada Research Chair in Urban Change and Adaptation. Research Highlight No. 2. Winnipeg: The Institute of Urban Studies, The University of Winnipeg.
- City of Calgary. Municipal Development Plan Maps. Retrieved from: [https://www.calgary.ca/\\_layouts/cocis/DirectDownload.aspx?target=http%3a%2f%2fwww.calgary.ca%2fPDA%2fpd%2fDocuments%2fmunicipal-development-plan%2fmdp-maps.pdf&noredirect=1&sf=1](https://www.calgary.ca/_layouts/cocis/DirectDownload.aspx?target=http%3a%2f%2fwww.calgary.ca%2fPDA%2fpd%2fDocuments%2fmunicipal-development-plan%2fmdp-maps.pdf&noredirect=1&sf=1)
- Ley, David. 1985. *Gentrification in Canadian inner cities: patterns, analysis, impact and policy*. Vancouver: University of British Columbia, Department of Geography.

- Ley, David. 1988. SOCIAL UPGRADING IN SIX CANADIAN INNER CITIES. *Canadian Geographer / Le Géographe canadien*, 32: 31-45. doi:10.1111/j.1541-0064.1988.tb00853.x
- Ley, David. 1992. GENTRIFICATION IN RECESSION: SOCIAL CHANGE IN SIX CANADIAN INNER CITIES, 1981-1986, *Urban Geography*, 13:3, 230-256, DOI: 10.2747/0272-3638.13.3.230
- Lezubski, Darren & Silver, Jim. 2015. High and Rising Revisited: Changes in Poverty and Related Inner City Characteristics 1996 to 2011. In, *Drawing on Our Strengths: State of the Inner City Report 2015*. Winnipeg: Canadian Centre for Policy Alternatives Manitoba Office. P. 7-34.
- Lorch, Brian. 2015. Spatial Polarization of Income in a Slow-Growth City. In, *The Divided Prairie City: Income Inequality Among Winnipeg's Neighbourhoods, 1970–2010*. Eds. Jino Distasio and Andrew Kaufman. Winnipeg: Institute of Urban Studies, The University of Winnipeg. P. 14-25.
- Savard, Charles. 2007. La lutte de la Coalition pour la sauvegarde des écoles des quartiers centraux de Québec (Vieux-Limoilou 2004-2006). Montréal: Centre de recherche sur les innovations sociales.
- Trudi Bunting, Alan R. Walks & Pierre Filion (2004) The uneven geography of housing affordability stress in Canadian metropolitan areas, *Housing Studies*, 19:3, 361-393, DOI: 10.1080/0267303042000204287
- Villeneuve, P. & Trudelle, C. (2008). Retour au centre à Québec : la renaissance de La Cité est-elle durable ? *Recherches sociographiques*, 49 (1), 25–45. <https://doi.org/10.7202/018192ar>
- Walks, R. A. (2001). The Social Ecology of the Post-Fordist/Global City? Economic Restructuring and Socio-spatial Polarisation in the Toronto Urban Region. *Urban Studies*, 38(3), 407–447. <https://doi.org/10.1080/00420980120027438>

The aforementioned sources are not in total agreement as to what constitutes the inner-city in certain cities. To an extent this is to be expected, because as decades pass, one might define the inner-city as a broader area than previously. Hence, the inner-city areas that we offer are approximate rather than definite.

Maps of the locations of inner cities within CMAs are provided:

a. Calgary, b. Edmonton, c. Montreal, d. Ottawa-Gatineau,  
e. Quebec City, f. Toronto, G. Vancouver, H. Winnipeg



## Appendix III. Additional CMA Data

Individual				Household			
CMA	Low (95% CI)	CI <sub>mid</sub>	CI <sub>high</sub>	CMA	Low (95% CI)	CI <sub>mid</sub>	CI <sub>high</sub>
Brantford	0.16 [0.08, 0.25]*	0.03 [-0.10, 0.23]	-0.20 [-0.32, -0.06]*	Brantford	0.25 [0.05, 0.48]*	-0.17 [-0.46, 0.09]	-0.08 [-0.20, 0.02]
Guelph	0.17 [0.04, 0.27]*	-0.24 [-0.39, -0.03]*	0.07 [-0.02, 0.14]	Guelph	0.36 [0.23, 0.49]*	-0.67 [-0.89, -0.46]*	0.31 [0.17, 0.50]*
Halifax	0.39 [0.29, 0.49]*	-0.57 [-0.68, -0.45]*	0.17 [0.15, 0.21]*	Halifax	0.33 [0.27, 0.40]*	-0.44 [-0.53, -0.35]*	0.10 [0.07, 0.15]*
Hamilton	0.39 [0.28, 0.49]*	-0.57 [-0.69, -0.45]*	0.17 [0.15, 0.21]*	Hamilton	0.40 [0.25, 0.56]*	-0.65 [-0.95, -0.35]*	0.25 [0.09, 0.39]*
Kingston	0.08 [-0.04, 0.20]	-0.06 [-0.24, 0.15]	-0.02 [-0.09, 0.07]	Kingston	0.31 [0.18, 0.45]*	-0.38 [-0.69, -0.10]*	0.08 [-0.10, 0.26]
Kitchener	0.36 [0.29, 0.46]*	-0.53 [-0.71, -0.37]*	0.18 [0.02, 0.28]*	Kitchener	0.63 [0.43, 0.80]*	-0.97 [-1.26, -0.62]*	0.34 [0.22, 0.48]*
London	0.36 [0.24, 0.50]*	-0.57 [-0.79, -0.37]*	0.20 [0.10, 0.30]*	London	0.33 [0.17, 0.50]*	-0.51 [-0.70, -0.31]*	0.18 [0.13, 0.21]*
Oshawa	0.16 [0.12, 0.21]*	-0.38 [-0.47, -0.32]*	0.22 [0.19, 0.26]*	Oshawa	0.20 [0.11, 0.26]*	-0.64 [-0.89, -0.44]*	0.44 [0.29, 0.62]*
Peterborough	0.12 [-0.01, 0.22]	-0.34 [-0.43, -0.23]*	0.23 [0.05, 0.45]*	Peterborough	0.33 [0.26, 0.41]*	-0.47 [-0.65, -0.19]*	0.14 [-0.10, 0.27]
Regina	0.09 [0.01, 0.15]*	0.06 [-0.06, 0.31]	-0.15 [-0.29, -0.06]*	Regina	0.12 [-0.02, 0.29]	-0.09 [-0.47, 0.38]	-0.02 [-0.34, 0.22]
Sarnia	0.41 [0.31, 0.47]*	-0.46 [-0.58, -0.43]*	0.06 [-0.01, 0.16]	Sarnia	0.53 [0.25, 0.83]*	-0.64 [-1.04, -0.28]*	0.11 [0.02, 0.27]*
Saskatoon	0.24 [0.16, 0.34]*	-0.34 [-0.54, -0.15]*	0.09 [-0.02, 0.21]	Saskatoon	0.29 [0.10, 0.47]*	-0.27 [-0.51, -0.02]*	-0.02 [-0.13, 0.15]
Sault Ste Marie	0.17 [0.02, 0.31]*	-0.30 [-0.45, -0.12]*	0.13 [0.06, 0.16]*	Sault Ste Marie	0.17 [0.01, 0.30]*	-0.49 [-0.62, -0.30]*	0.32 [0.30, 0.35]*
Sherbrooke	0.06 [-0.04, 0.14]	-0.14 [-0.37, 0.15]	0.08 [-0.10, 0.26]	Sherbrooke	0.28 [0.05, 0.53]*	-0.42 [-0.73, -0.11]*	0.14 [0.04, 0.26]*
St. Catharines	0.28 [0.24, 0.34]*	-0.44 [-0.56, -0.32]*	0.16 [0.06, 0.23]*	St. Catharines	0.44 [0.33, 0.58]*	-0.78 [-0.94, -0.63]*	0.33 [0.27, 0.39]*
St. John	0.20 [0.15, 0.25]*	-0.22 [-0.33, -0.10]*	0.02 [-0.07, 0.13]	St. John	0.30 [-0.06, 0.56]	-0.33 [-0.71, 0.17]	0.02 [-0.13, 0.15]
St. Johns	-0.17 [-0.32, -0.02]*	0.25 [-0.06, 0.61]	-0.08 [-0.32, 0.09]	St. Johns	0.02 [-0.05, 0.10]	-0.04 [-0.22, 0.17]	0.02 [-0.13, 0.15]
Sudbury	0.24 [0.17, 0.32]*	-0.44 [-0.60, -0.25]*	0.20 [0.04, 0.31]*	Sudbury	0.41 [0.15, 0.64]*	-0.69 [-0.92, -0.29]*	0.28 [0.21, 0.37]*
Thunder Bay	0.20 [0.15, 0.28]*	-0.47 [-0.75, -0.19]*	0.27 [0.04, 0.48]*	Thunder Bay	0.33 [0.19, 0.53]*	-0.68 [-0.94, -0.50]*	0.34 [0.25, 0.51]*
Trois-Riviere	0.08 [-0.02, 0.20]	-0.13 [-0.29, 0.02]	0.05 [-0.05, 0.12]	Trois-Riviere	0.29 [0.20, 0.39]*	-0.73 [-0.88, -0.62]*	0.43 [0.36, 0.52]*
Victoria	0.07 [-0.05, 0.12]	-0.12 [-0.17, -0.05]*	0.05 [0.00, 0.08]*	Victoria	0.13 [0.05, 0.20]*	-0.21 [-0.38, -0.07]*	0.07 [-0.02, 0.17]
Windsor	0.33 [0.29, 0.37]*	-0.55 [-0.75, -0.36]*	0.22 [0.08, 0.41]*	Windsor	0.44 [0.32, 0.58]*	-0.73 [-0.98, -0.53]*	0.30 [0.20, 0.40]*
Calgary	0.57 [0.45, 0.74]*	-0.65 [-0.86, -0.52]*	0.09 [0.05, 0.13]*	Calgary	0.37 [0.27, 0.47]*	-0.37 [-0.47, -0.27]*	-0.00 [-0.03, 0.04]
Edmonton	0.35 [0.26, 0.43]*	-0.48 [-0.60, -0.32]*	0.13 [0.05, 0.18]*	Edmonton	0.27 [0.21, 0.33]*	-0.39 [-0.53, -0.26]*	0.12 [0.04, 0.21]*
Montreal	0.14 [0.10, 0.18]*	-0.20 [-0.26, -0.14]*	0.06 [0.04, 0.09]*	Montreal	0.28 [0.19, 0.36]*	-0.38 [-0.57, -0.19]*	0.10 [-0.01, 0.22]
Ottawa	0.08 [0.05, 0.15]*	-0.12 [-0.23, -0.04]*	0.05 [-0.03, 0.12]	Ottawa	0.16 [0.06, 0.25]*	-0.37 [-0.55, -0.16]*	0.20 [0.12, 0.30]*
Quebec City	0.02 [-0.03, 0.06]	-0.05 [-0.14, 0.02]	0.03 [-0.01, 0.08]	Quebec City	0.23 [0.16, 0.29]*	-0.41 [-0.55, -0.25]*	0.18 [0.10, 0.27]*
Toronto	0.53 [0.41, 0.66]*	-0.62 [-0.76, -0.48]*	0.09 [0.07, 0.13]*	Toronto	0.42 [0.32, 0.53]*	-0.52 [-0.77, -0.29]*	0.09 [-0.04, 0.26]
Vancouver	0.40 [0.35, 0.48]*	-0.51 [-0.61, -0.43]*	0.11 [0.08, 0.16]*	Vancouver	0.09 [0.02, 0.16]*	-0.17 [-0.36, 0.01]	0.08 [-0.05, 0.22]
Winnipeg	0.19 [0.05, 0.29]*	-0.32 [-0.50, -0.09]*	0.13 [0.05, 0.24]*	Winnipeg	0.25 [0.20, 0.31]*	-0.39 [-0.46, -0.34]*	0.15 [0.13, 0.16]*

## **Appendix IV. Supplementary information for ‘Deep mapping gentrification in a large Canadian city using deep learning and Google Street View’**

### **Contents**

#### **S1. Introduction**

#### **S2. GitHub repository**

#### **S3. Study area and GSV data**

#### **S4. Results**

##### **S4.1 SCNN-FC-8 model training and validation**

###### **S4.1.1 Building Permits**

###### **S4.1.2 SCNN-FC-8 and building permits**

#### **S5. SCNN-FC-8 sensitivity**

#### **S6. Activation maps**

### **S1. Introduction**

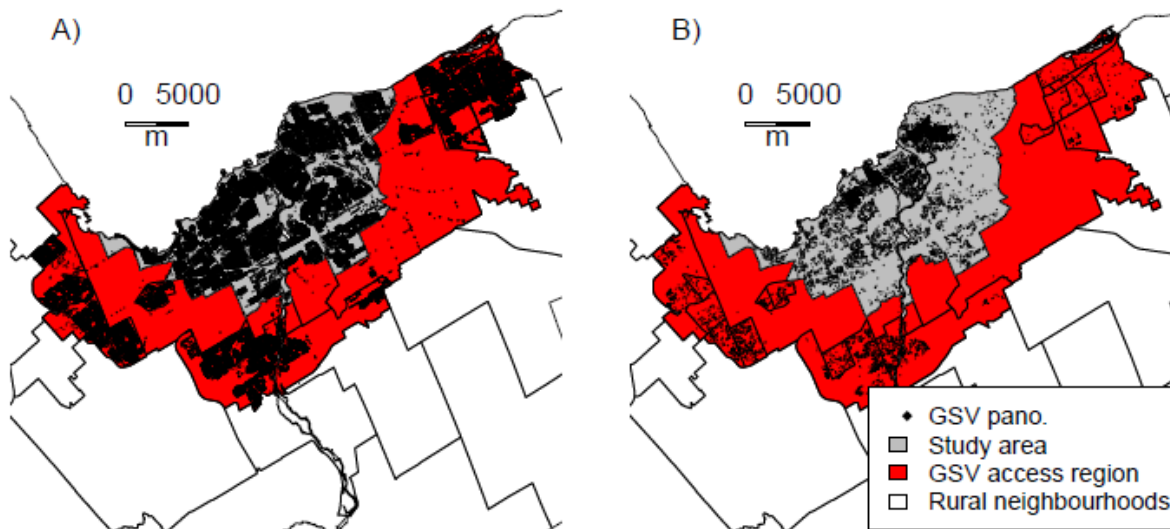
In the main manuscript we describe the use of SCNN-FC-8 to detect gentrification-like visual changes within a sequence of GSV images for a given property. This document contains supplementary explanations and figures that elaborate upon information in the main manuscript.

### **S2. GitHub repository**

The set of python scripts used to create and train SCNN-FC-8 can be found at:

\*<https://github.com/laggiss/DeepMapping>

NOTE: \*This document was created from an RMarkdown file. The RMarkdown version of this document contains all of the R-language code used to access datasets as well as create the maps (except Fig 1 in the manuscript which was created elsewhere) and graphs in this supplementary document and manuscript. However, the R code is suppressed in the version you are currently reading.



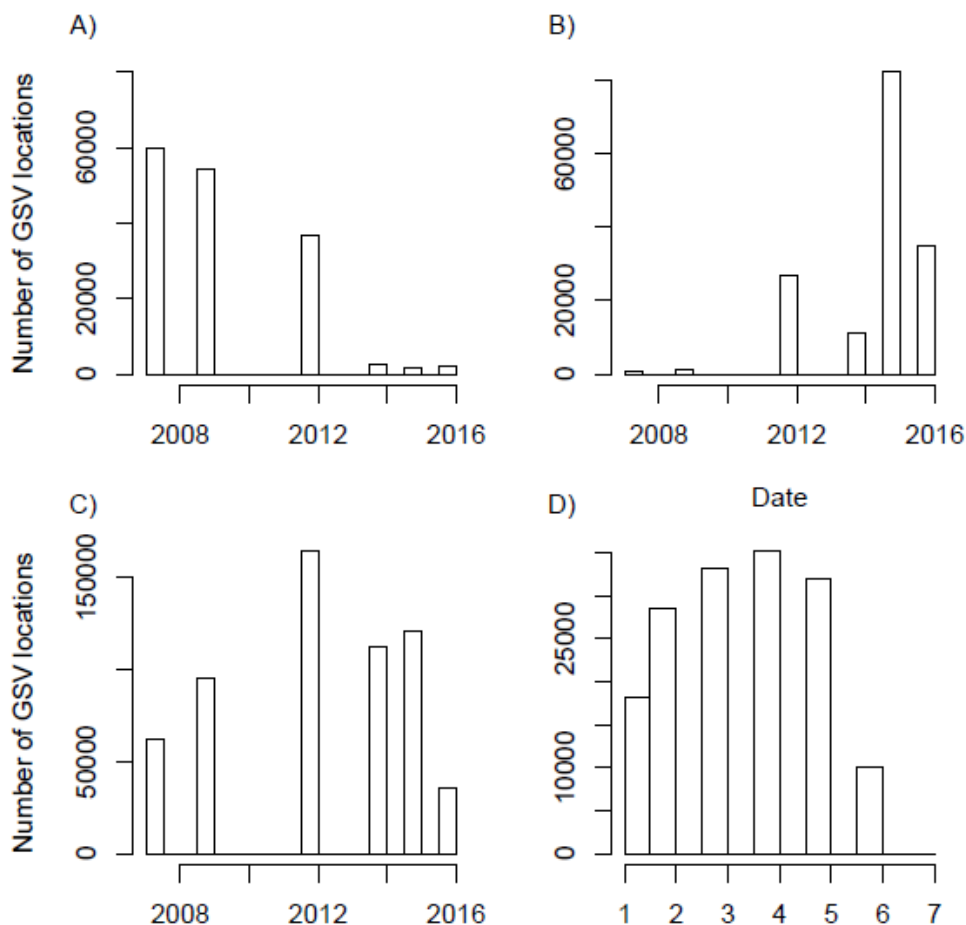
**Figure A: GSV panorama distributions used in this research. A) Distribution of all GSV panoramas accessed in this research; B) Subset of GSV panoramas within the GSV access area that were used in training, validation and testing of SCNN-FC-8.**

Should you wish to examine the raw data or code, you can find the RMarkdown document with R-language code containing all references at:  
<https://github.com/laggiss/DeepMapping/blob/master/sampling.Rmd>

### S3. Study area and GSV data

The spatio-temporal coverage of GSV imagery varies across space and time. We were primarily concerned with detecting gentrification as far back as possible within our ‘study area’. The study area corresponds to the spatial domain defined by the highest concentration of older building stock in Ottawa (see Fig 1 in manuscript). To train SCNN-FC-8, we wanted a larger region in which to find examples of gentrification-like visual change. Thus, we accessed GSV imagery within a broader region that includes our study area and a buffer region that, itself, is comprised primarily of post 1980’s suburban residential land use. We call this broader region the ‘GSV access region’ within which GSV panoramas were selected to train SCNN-FC-8 (Fig A(A)). There were 157303 properties containing panoramas within the GSV access region.

Because each property/location can have numerous panoramas over time, the total number of panoramas across the properties was 593,723. A subset of 16224 GSV images were clipped from the panoramas to extract individual buildings by sending the field of view (FOV) and camera direction to the Google Maps API. That subset was used for training, validating and testing SCNN-FC-8 (Fig A(B).

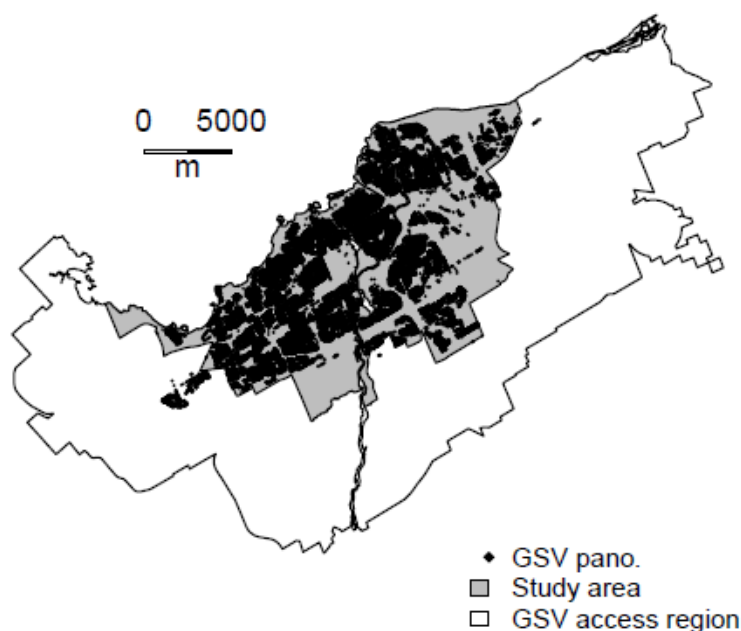


**Figure B: Frequency distributions. A) Distribution of the oldest GSV date accessed at a location; B) Distribution of the Youngest GSV date accessed at a location; C) Distribution of all GSV image dates accessed at a location; D) Distribution of the number of years of GSV imagery at each queried location.**

Next, we determine if the panoramas within the GSV access region can support analysis in the study area back to 2007. The first GSV panoramas in the city were taken in 2007 and we would like to map changes as far back as possible when describing the spatial evolution of

gentrification in the study area. Thus, the possible temporal domain for this study is from 2007-2016 (we accessed imagery for this work in the spring of 2017). However, GSV images may not be available at all time periods at all locations. In addition, we also determine which years in the interval between 2007 and 2016 contained sufficient spatial coverage.

Within the GSV access region (Fig A), the majority of properties had GSV images by 2012 (Fig B(A)) and the majority of these were fully re-imaged between 2014-2016 (Fig BB). 2012 is the year with the most panoramas (Fig B(C)). Fig B(C) provide evidence of an approximate 2-3 year update frequency of GSV imagery. The average number of years that GSV panoramas are available at a property is 3.41 Fig B(D).



**Figure C: Spatial distribution of properties that have a panorama in 2007 and also in 2015-2016.**

Examining all properties (locations) that contained an image in 2007 and one in the years 2015-2016, most clearly defines the region in which we can make inferences back to 2007 (Fig C). We cannot, for example, make inferences regarding changes in the spatial distribution of

gentrification-like visual changes between the years 2009-2010, since in 2010 there are almost no GSV panoramas.

Concerning the spatial distribution of GSV images over the study period and within the study area, we examined the spatial distribution of the GSV panorama dates through time. In Fig C, it is apparent that only the study region contains sufficient spatial coverage back to 2007 (Fig C). Google did not expand coverage into the GSV access region until 2009 in Ottawa. The study region is completely covered with panoramas in 2007,2009,2012,2014, and 2015 (Fig D). Therefore, we were confident in using SCNN-FC-8 to detect gentrification-like visual changes within the study region. Regarding the examination of gentrification through time, we can compare GSV images within the study area between the years: 2007-2009,2009-2012,2012-2014 and between 2014 and a combined 2014-2016 (Fig C).

## **S4 Results**

### **S4.1 SCNN-FC-8 model training and validation**

#### **S4.1.1 Building Permits**

The City of Ottawa building permits are organized by month and by year in excel files and each permit contains fields such as address, date of issue and the description of work as well as the cost. We concatenated permit lists from July 2011 through December 2016 which produced a list of 56,269 permits that had 22631 unique descriptions. This total was reduced, as is explained in the following.

Through permits' descriptions it was possible to trim the list by removing those which were deemed redundant, as well as to establish a list of relevant permits. After pruning the permit list, 1391 permit descriptions remained and were inspected manually, after which the total number of permits amounted to 27,509.

After removing permits belonging to structures outside the study area, 6,342 permits remained. Each remaining permit location was manually inspected on Street View, after which the list was further reduced to a final number of 2356. There were two criteria for retaining permits. The first was to only keep permits whose addresses can be seen on Street View, and the second was that the change observed to the property of the permit must have undergone a change which yielded a progressively more affluent structure. Some examples can include the addition of a car port, additions to the side of the house, or the construction of a completely new structure. However, more permits were removed (3986) than were kept. This is not to imply that the removed permit entries are of structures which were not improved, but rather that their improvements are not visible from the street. Rear additions, basement alterations, and foundation repair are examples of improvements which do not necessarily yield a visual change.

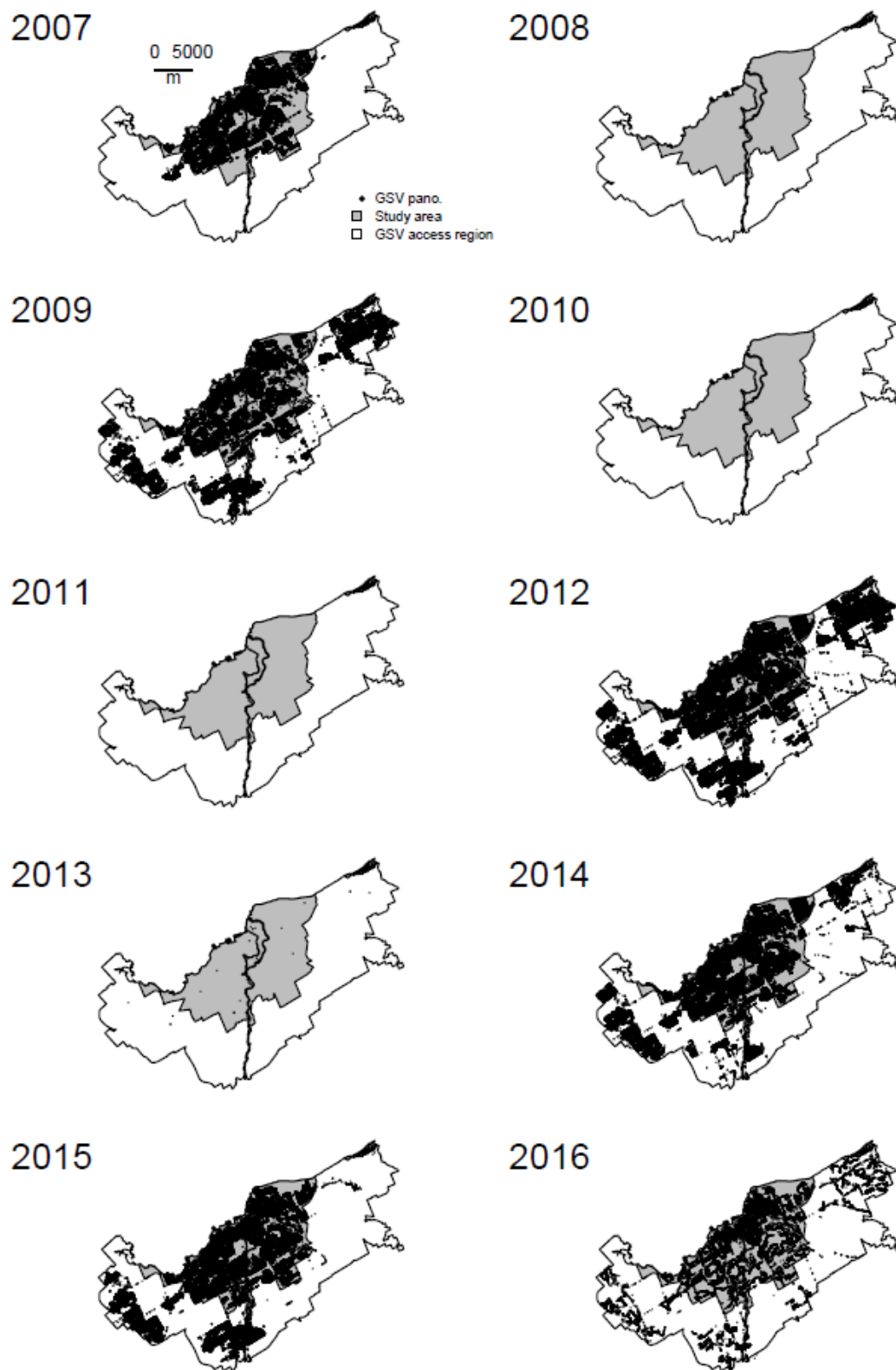
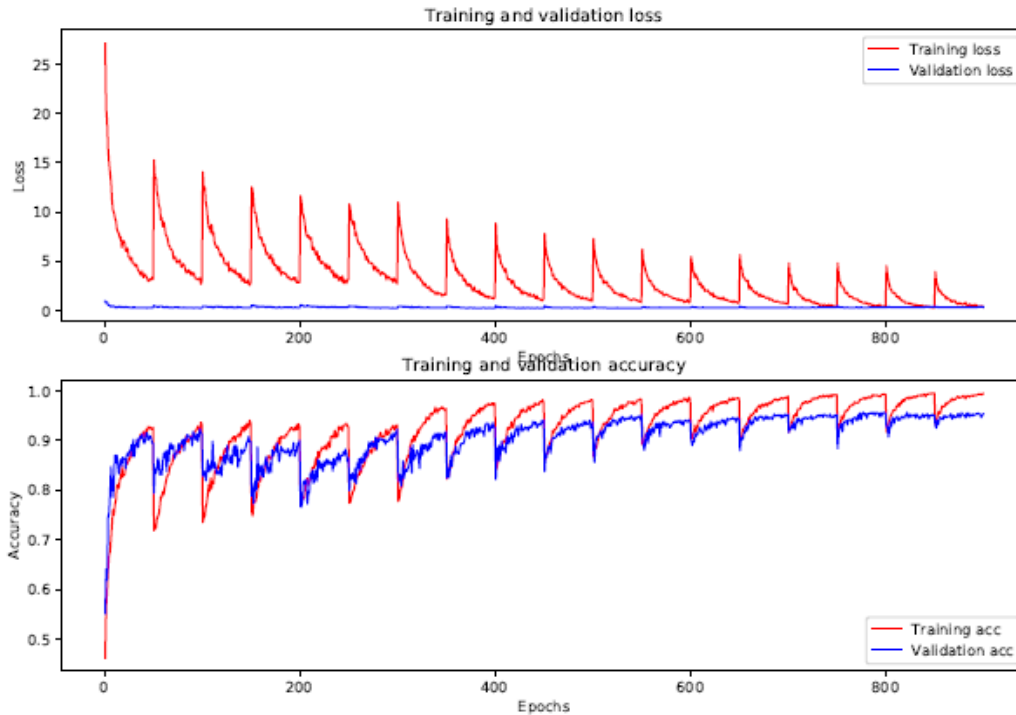


Figure D: Spatial distribution of all accessed GSV images through time.



**Figure E: Training and validation of SCNN-FC-8 (Epoch 1 to min(loss)).**

Street View lacks imagery on certain streets and segments of streets. This is most common with suburban subdivisions which are designated as privates. Some permits were deleted because their addresses were obfuscated on Street View. Other permits were deleted as the upgrades that they represented are not indicators of gentrification. This would include certain commercial infrastructure such as: industrial buildings, warehouses and storage structures. Institutional infrastructure was also removed, and included objects such as universities, schools, fire stations, churches, hospitals, community centers and transit infrastructure.

The SQL statements used to refine the permits can be found here: [https://github.com/laggiss/DeepMapping/blob/master/sql\\_statements\\_permits.txt](https://github.com/laggiss/DeepMapping/blob/master/sql_statements_permits.txt)

#### **S4.1.2 SCNN-FC-8 and building permits**

We could have used KDE for point patterns based on a network c.f.[Okabe and Sugihara, 2012]. However, strictly speaking, because the spatial support for the points relate to an area

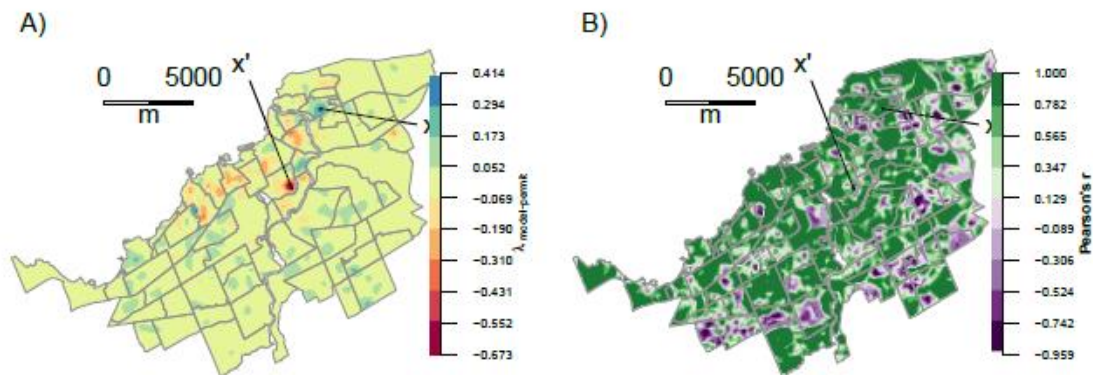
unit, the property, our point data violates the basic assumption of point pattern analyses, either in free space or on a network, both of which requires points to be representative of discrete entities. That being said, because the point locations  $(x,y)$  are a dimensional reduction of the property polygon, we use KDE on what is effectively a point process that contains dispersion at small scales - because the property areas are approximately equal. Our choice of bandwidth is much larger than the small-scale dispersion among adjacent properties and so a valid way to visualize and compare patterns is by using KDE. We are interested in the location and scale of the gentrification process as defined by the clustering of gentrification-like visual changes.

We found the pattern of building permits and SCNN-FC-8 predictions over the same period to be very similar, with only a few minor deviations (Fig 5 in Manuscript).

Fig F(A) illustrates the difference between the standardized intensity,  $\lambda$ , of the SCNN-FC-8 KDE prediction map and KDE of building permits. The two isolated regions of intensity in Fig F(A),  $x$  and  $x'$ , clearly stand out as the two largest differences between the model predictions and the building permits. As discussed in the main text,  $x'$  is due to a large development of Lansdowne Park ([https://en.wikipedia.org/wiki/Lansdowne\\_Park\\_redevelopment](https://en.wikipedia.org/wiki/Lansdowne_Park_redevelopment)), and  $x$  was due to false positive detections of gentrification-like visual change because of offset/FOV issues. In particular, in the region denoted by  $x'$ , there were a number of new homes constructed which normally would have been detected by SCNN-FC-8 within the image sequence. However, after examining the sequences of GSV images in the region of  $x'$ , a common theme became apparent in the sequence of visual changes: Initially, a park was present and that was followed by construction of a large wall. The wall was eventual removed and a new house was constructed. The change from a wall to a residential structure, or from a park to wall for that matter, was not

something that SCNN-FC-8 encountered in our training data set. Thus, such changes were not reinforced in training as being gentrification-like instances of visual change.

Fig F(B) represents the local Pearson's correlation coefficient between the Standardized intensity  $\{\lambda\}$  of the SCNN-FC-8 KDE prediction map and KDE of building permits.

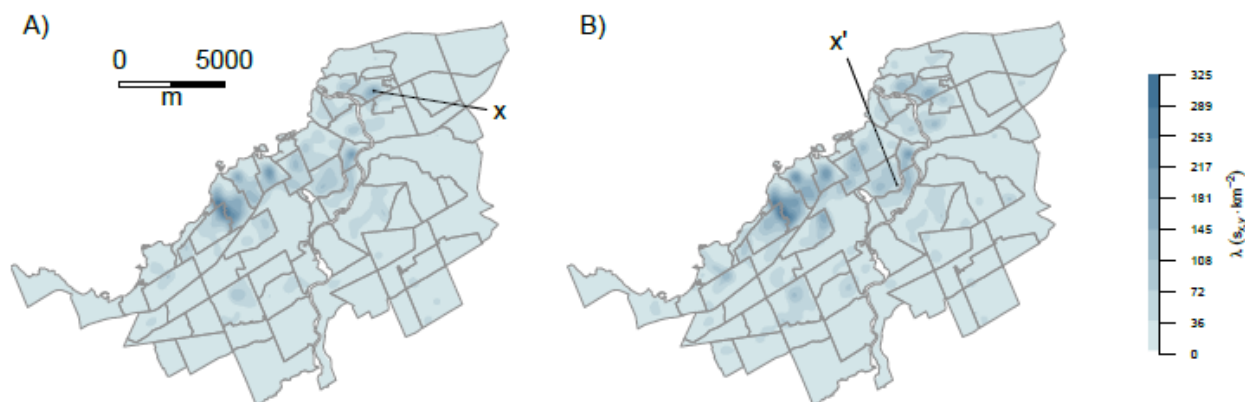


**Figure F: Comparison of KDE maps. A) Proportional difference between predicted KDE and permit KDE. B) Local Pearson's R between predicted KDE and permit KDE.**

We use the correlation coefficient  $r$  as an exploratory tool to identify local regions of similarity/dissimilarity in intensity between the model predictions and the permits. This map represents the correlation coefficient in a 5x5 moving window between the standardized values from manuscript Fig 5AB. The large amount of green area in Fig F(B), indicates that the majority of the intensity values strongly covary, which is evident in the visual inspection of Fig 5AB (Manuscript). There are strong dissimilarities or negative correlations locally, but these are generally in areas of very low spatial intensity around the southern boundary of the study region. In other words, the majority of disagreements in the model/permits comparison are in regions where there are few gentrification-like visual changes. This further reinforces the concordance seen between SCNN-FC-8 and the building permits upon visual comparison of their mapped intensities.

### S5. SCNN-FC-8 sensitivity

After training SCNN-FC-8 as described in the manuscript, we wanted a crude test of how different our results would be if we reshuffled all of the training data and re-trained the model from scratch (the model training procedure is described in the manuscript). The training data set contained all 16224 GSV images and we used 60% of these to train SCNN-FC-8, holding back 20% for model validation during training, and the last 20% were held back for testing the model once we achieved our desired model validation accuracy. All 16224 image pairs were shuffled/permutated prior to being split. To retrain the model once again, we simply reshuffled all 16224 GSV image pairs and reapplied the training procedure. After training this second model instance, all GSV image sequences were fed through the model and the results were mapped (Fig G). The spatial pattern of intensity between SCNN-FC-8 (Fig G(A)) and the second model (Fig G(B)) are nearly identical. This suggests that the results obtained using SCNN-FC-8 are not significantly biased by the train/validation/test split.



**Figure G: Comparison of A) SCNN-FC-8 results with B) the replicate model results**

## S6. Activation maps

Activation maps can be used to explore which features in a GSV image are most strongly reinforced in SCNN-FC-8. The activation maps of the last convolution block for each image in a properties sequence (Fig 8), aid in revealing the potential reasons why some false-positive detections were made by SCNN-FC-8. Generally, we expect similar activation maps on the  $t[0]$  and  $t[0]+1$  GSV images when no gentrification-like visual change is detected by SCNN-FC-8. Conversely, when a change is detected, there should be differences in which features are most strongly activated upon by SCNN-FC-8. The activation maps in Fig 8 [https://github.com/laggiss/DeepMapping/blob/master/GSVFIGS/figS8\\_online.png](https://github.com/laggiss/DeepMapping/blob/master/GSVFIGS/figS8_online.png) correspond to the images in the manuscript Fig 7-[https://github.com/laggiss/DeepMapping/blob/master/GSVFIGS/fig7\\_online.png](https://github.com/laggiss/DeepMapping/blob/master/GSVFIGS/fig7_online.png). See manuscript for further explanation.

## References

Atsuyuki Okabe and Kokichi Sugihara. *Spatial Analysis along Networks: Statistical and Computational Methods*. John Wiley & Sons, Ltd, Chichester, UK, July 2012. ISBN 978-1-119-96710-1 978-0-470-77081-8. doi: 10.1002/9781119967101. URL <http://doi.wiley.com/10.1002/9781119967101>.

**Appendix V. Supplementary information for ‘Modeling and mapping gentrification with census data in a large Canadian city’**

**Contents**

**S1. Gentrification Journal Articles**

**S2. Largest Cities in the US and Canada**

**S3. Zoning for Dasymetric Mapping Mask**

**S4. Census Variables**

**S5. Regression Results for Various Income and Education measures**

**S6. Transformations of Dependent Variable**

**S7. Backwards Stepwise Model Selection**

**S8. Lagrange Diagnostics**

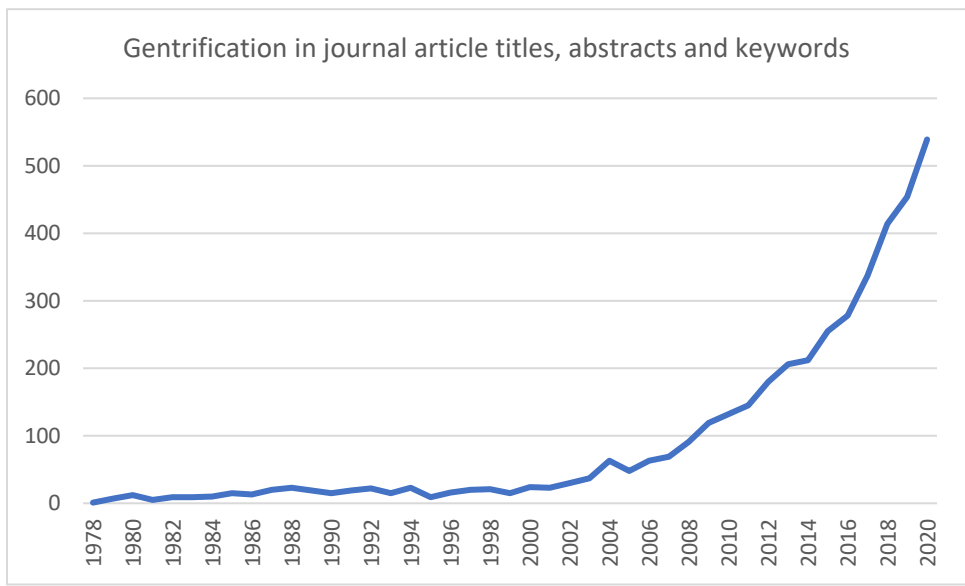
**S9. Bivariate Quasibinomial Model Output**

**S10. Spatial Error Model**

**S1. Gentrification Journal Articles**

The following graph shows the number of journal articles with gentrification in their article title, abstract or keywords. This data is obtained from the Scopus database on 15 May 2021.

1978	1	1993	15	2008	91
1979	7	1994	23	2009	119
1980	12	1995	9	2010	132
1981	5	1996	16	2011	145
1982	9	1997	20	2012	180
1983	9	1998	21	2013	206
1984	10	1999	15	2014	212
1985	15	2000	24	2015	255
1986	13	2001	23	2016	278
1987	20	2002	30	2017	337
1988	23	2003	37	2018	414
1989	19	2004	63	2019	454
1990	15	2005	48	2020	539
1991	19	2006	63		
1992	22	2007	69		



## S2. Largest Cities in the US and Canada

The following list was obtained from [https://en.wikipedia.org/wiki/List\\_of\\_North\\_American\\_cities\\_by\\_population](https://en.wikipedia.org/wiki/List_of_North_American_cities_by_population) and was subsequently modified by ranking the cities in descending order of territorial size and pruned to only include cities in the US and Canada.

Pop. Rank	Name	Country	Population	Sq. Km.
14	Ottawa	Canada	956,710	2,790 km <sup>2</sup>
17	Jacksonville	United States	868,031	2,265 km <sup>2</sup>
5	Houston	United States	2,296,224	1,733 km <sup>2</sup>
36	Oklahoma City	United States	631,346	1,608 km <sup>2</sup>
33	Nashville	United States	654,610	1,362 km <sup>2</sup>
8	Phoenix	United States	1,563,025	1,338 km <sup>2</sup>
9	San Antonio	United States	1,469,845	1,307 km <sup>2</sup>
2	Los Angeles	United States	3,971,883	1,302 km <sup>2</sup>
43	Hamilton	Canada	556,359	1,138 km <sup>2</sup>
40	Louisville	United States	615,366	1,030 km <sup>2</sup>
11	Dallas	United States	1,300,092	993.1 km <sup>2</sup>
10	San Diego	United States	1,394,928	964.5 km <sup>2</sup>
19	Indianapolis	United States	853,173	953 km <sup>2</sup>
21	Fort Worth	United States	833,319	920.9 km <sup>2</sup>
12	Calgary	Canada	1,230,915	825.3 km <sup>2</sup>
15	Austin	United States	931,830	790.1 km <sup>2</sup>
32	Memphis	United States	655,770	789 km <sup>2</sup>
1	New York City	United States	8,550,405	783.8 km <sup>2</sup>
23	Charlotte	United States	827,097	771 km <sup>2</sup>
16	Edmonton	Canada	899,447	684 km <sup>2</sup>
28	El Paso	United States	681,124	671.5 km <sup>2</sup>
3	Toronto	Canada	2,826,498	630.2 km <sup>2</sup>
4	Chicago	United States	2,720,546	606.1 km <sup>2</sup>
45	Tucson	United States	531,641	587.2 km <sup>2</sup>
20	Columbus	United States	850,106	583 km <sup>2</sup>
42	Albuquerque	United States	559,121	489.2 km <sup>2</sup>
44	Quebec City	Canada	540,994	484.1 km <sup>2</sup>
13	San Jose	United States	1,026,908	469.7 km <sup>2</sup>
25	Winnipeg	Canada	718,400	464.1 km <sup>2</sup>
6	Montreal	Canada	1,753,034	431.5 km <sup>2</sup>
27	Denver	United States	682,545	401.2 km <sup>2</sup>
35	Portland	United States	632,309	375.5 km <sup>2</sup>
29	Detroit	United States	677,116	370.1 km <sup>2</sup>
7	Philadelphia	United States	1,567,442	367 km <sup>2</sup>

37	Las Vegas	United States	623,747	352 km <sup>2</sup>
46	Surrey	Canada	526,004	316.4 km <sup>2</sup>
47	Fresno	United States	520,052	300.4 km <sup>2</sup>
24	Mississauga	Canada	761,300	292.4 km <sup>2</sup>
39	Brampton	Canada	620,700	266.7 km <sup>2</sup>
41	Milwaukee	United States	600,155	250.7 km <sup>2</sup>
38	Baltimore	United States	621,849	239 km <sup>2</sup>
31	Boston	United States	667,137	232.1 km <sup>2</sup>
26	Seattle	United States	684,451	217 km <sup>2</sup>
30	Washington, D.C.	United States	672,228	177 km <sup>2</sup>
18	San Francisco	United States	864,816	121.4 km <sup>2</sup>
34	Vancouver	Canada	648,608	115 km <sup>2</sup>

### **S3. Zoning for Dasymetric Mapping Mask**

#### Zoning used:

AM – Arterial Main Street Zone  
GM – General Mixed Use Zone  
LC – Local commercial Zone  
MC – Mixed Use Centre Zone  
MD – Mixed Use Downtown Zone  
R (R1, R2, R3, R4, R5) – Residential  
RS – Residential Single Dwelling Zone  
RI – Rural Institutional Zone  
RR – Rural Residential Zone  
TD – Transit Oriented Development Zone  
TM – Traditional Mainstreet Zone

#### Zoning not used:

AG – Agricultural Zone  
DR – Development Reserve Zone  
EP – Environmental Protection Zone  
I1 – Minor Institutional Zone  
I2 – Major Institutional Zone  
IG – General Industrial Zone  
IH – Heavy Industrial Zone  
IL – Industrial Zone  
IP – Business Park Industrial Zone  
L1 – Community Leisure Facility Zone  
L2 – Major Leisure Facility Zone  
L3 – Central Experimental Farm  
O1 – Parks and Open Space Zone  
RU – Rural Countryside Zone  
T1 – Air Transportation Facility Zone  
T2 – Ground Transportation Zone

## S4. Census Variables

Census data is available from a few places:

- IVT tables from Statistics Canada
- the database Canadian Census Analyzer
- the database Odesi

**B.** Age – Young Adults – From 20 to 35 years old (ages 20 to 34)

**C.** Age – Elderly – Above 65 (ages 65+)

**D.** Age – Youth – Under 20 years old (ages 0 to 19)

**E.** Dwelling Value

**F.** Dwellings Owned

**G.** Dwellings per square kilometers / Dwelling Density

**H.** Dwellings Rented

**I.** Education

**J.** Income

**K.** Occupation

**L.** Rent

**M.** Unemployment

**N.** Visible Minorities

**B., C., D.** : Age categories :

*Total population by sex and age groups*

**B.** Census profiles for age (young adults) :

Combine: “20 to 24 years”, “25 to 29 years”, & “30 to 34 years”

**C.** Census profiles for age (elderly) :

Combine: “65 to 69 years”, “70 to 74 years”, “75 to 79 years”, “80 to 84 years”, & “85 years and over”

**D.** Census profiles for age (youth) :

Combine: “0 to 14 years” & “15 to 19 years”

**E.** Census profiles for Dwelling Value

*“Owner-occupied private non-farm, non-reserve dwellings / Average value of dwelling \$”*

**F.** Census profiles for Dwellings Owned

*“Total number of non-farm, non-reserve private dwellings occupied by usual residents - 20% sample data / Owner-occupied private non-farm, non-reserve dwellings”*

**G.** Census profiles for Dwelling Density

*“Land area in square kilometres, 2006”*

*“Household income in 2005 of private households - 20% sample data”*

**H.** Census profiles for Dwellings Rented

*“Total number of non-farm, non-reserve private dwellings occupied by usual residents - 20% sample data / Tenant-occupied private non-farm, non-reserve dwellings”*

**I. Census profiles for Education**

*“Total population 25 to 64 years by highest certificate, diploma or degree - 20% sample data”*

+ *“Certificate, diploma or degree / Apprenticeship or trades certificate or diploma”*

+ *“Certificate, diploma or degree / College, CEGEP or other non-university certificate or diploma”*

+ *“Certificate, diploma or degree / University certificate, diploma or degree”*

**J. Census profiles for Income**

*“Household income in 2005 of private households - 20% sample data / Average household income \$”*

**K. Census profiles for Occupation**

*“All occupations”*

+ *“0 Management occupations”*

+ *“1 Business, finance and administration occupations”*

+ *“2 Natural and applied sciences and related occupations”*

+ *“3 Health occupations”*

+ *“4 Occupations in education, law and social, community and government services”*

**L. Census profiles for Rent**

*“Tenant-occupied private non-farm, non-reserve dwellings / Average gross rent \$”*

**M. Census profiles for Unemployment**

*“Unemployment rate”*

**N. Census profiles for Visible Minorities**

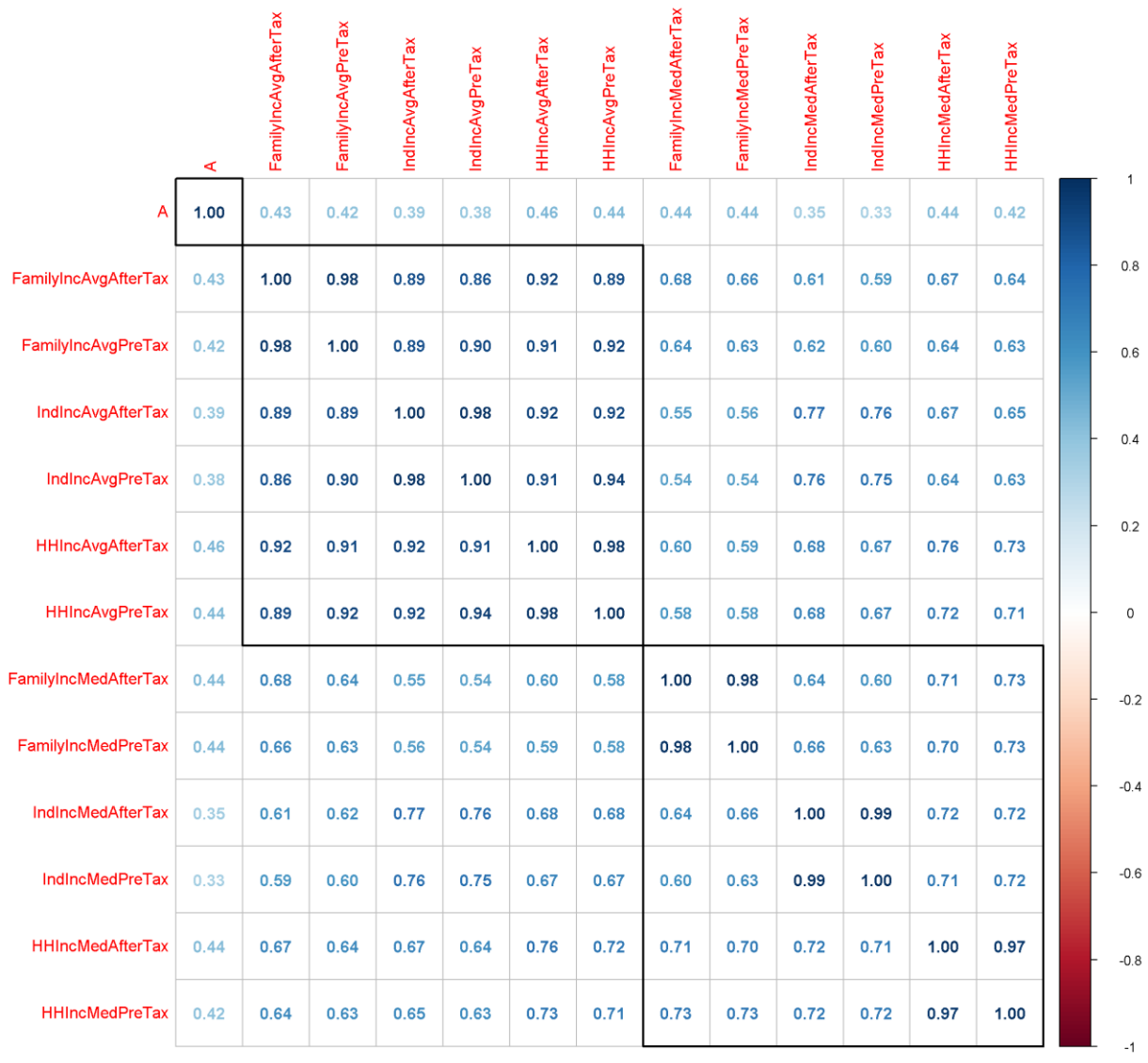
*“Total population by visible minority groups - 20% sample data”*

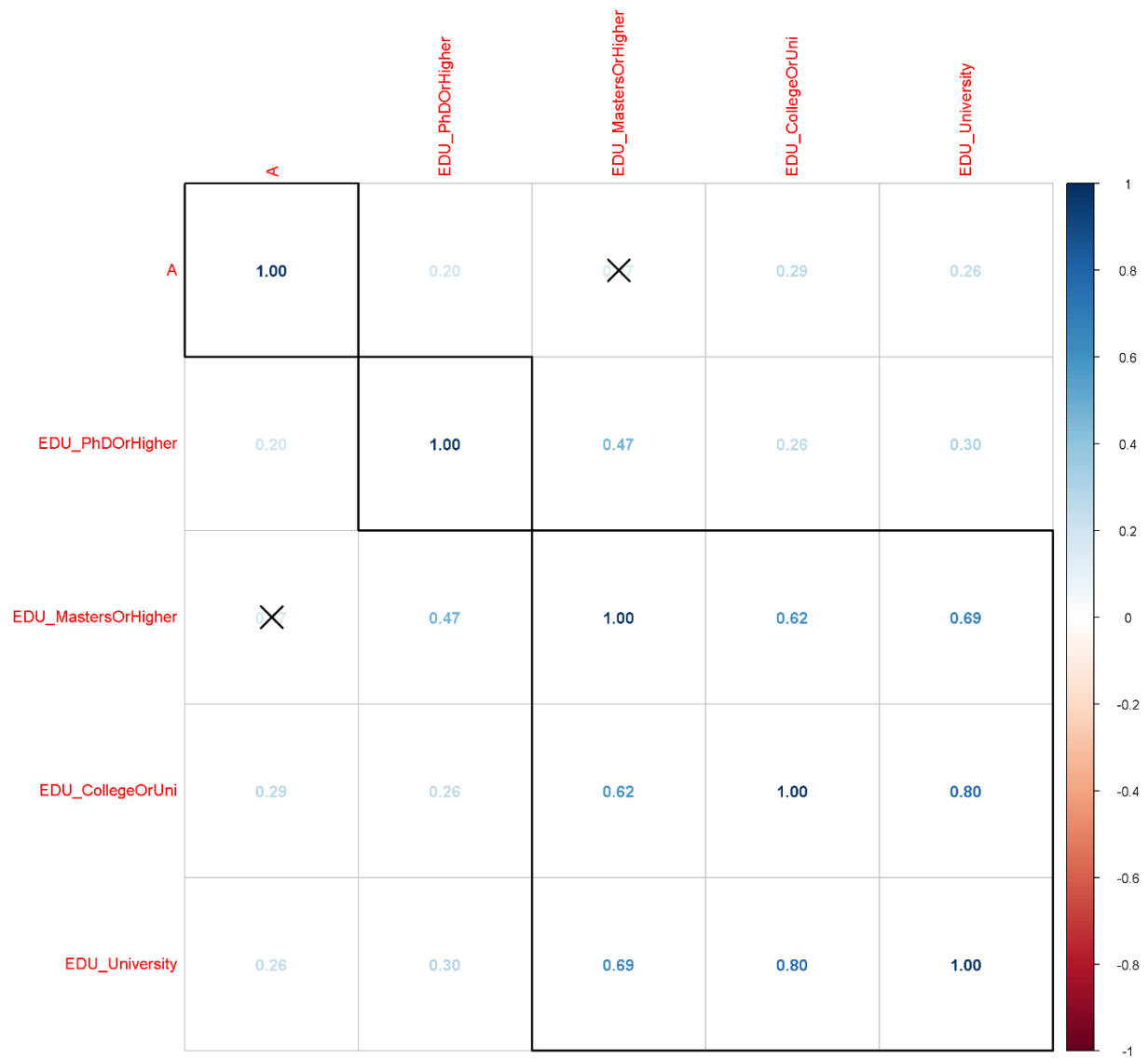
+ *“Total visible minority population / Black”*

+ *“Total visible minority population / Latin American”*

The above (B through N) are census profile names from the 2006 census. The 2016 ones are the same or similar. For example, data from the long form for the 2016 census is based on a 25% sample, whereas the 2006 census is based on a 20% sample... certain census variables note this.

### S5. Regression Results for Various Income and Education Measures





## S6. Transformations of Dependent Variable

```
shapiro.test(arc.sin)
##
## Shapiro-Wilk normality test
##
## data: arc.sin
## W = 0.98365, p-value = 0.2035
shapiro.test(square.root)
##
## Shapiro-Wilk normality test
##
## data: square.root
## W = 0.94663, p-value = 0.0002639
shapiro.test(log.e)
##
## Shapiro-Wilk normality test
##
## data: log.e
## W = 0.98373, p-value = 0.2064
shapiro.test(A$input)
##
## Shapiro-Wilk normality test
##
## data: A$input - Lambert W transformed density
## W = 0.98876, p-value = 0.5023
ctsjoined$Aprime=log.e
```

## S7. Backwards Stepwise Model Selection

```

## Start: AIC=726.1
## A ~ D + E + I + J + K + L + M
##
##      Df Sum of Sq  RSS    AIC
## - L   1    545.6 74119 724.91
## - I   1    775.0 74348 725.24
## - M   1   1178.6 74752 725.83
## <none>                73573 726.10
## - J   1   1830.8 75404 726.78
## - K   1   5838.3 79412 732.42
## - D   1   7782.3 81356 735.06
## - E   1  14155.8 87729 743.28
##
## Step: AIC=724.91
## A ~ D + E + I + J + K + M
##
##      Df Sum of Sq  RSS    AIC
## - I   1   1000.0 75119 724.37
## - M   1   1191.9 75311 724.65
## <none>                74119 724.91
## - J   1   1835.7 75955 725.57
## - K   1   6020.2 80139 731.42
## - D   1   9435.9 83555 735.97
## - E   1  15260.3 89379 743.31
##
## Step: AIC=724.37
## A ~ D + E + J + K + M
##
##      Df Sum of Sq  RSS    AIC
## - M   1   1191.0 76310 724.08
## <none>                75119 724.37
## - J   1   2283.2 77402 725.63
## - K   1   8407.8 83527 733.93
## - D   1   8542.9 83662 734.11
## - E   1  15935.1 91054 743.34
##
## Step: AIC=724.08
## A ~ D + E + J + K
##
##      Df Sum of Sq  RSS    AIC
## <none>                76310 724.08
## - J   1   2258.8 78569 725.26
## - D   1   7929.3 84239 732.86
## - K   1  10131.9 86442 735.67
## - E   1  16707.6 93018 743.66

```

## S8. Lagrange Diagnostics

```

##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = eqn, data = ctsjoined)
## weights: fcpolys.wt
##
## LMerr = 9.4413, df = 1, p-value = 0.002122
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = eqn, data = ctsjoined)
## weights: fcpolys.wt
##
## LMLag = 27.405, df = 1, p-value = 1.65e-07
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = eqn, data = ctsjoined)
## weights: fcpolys.wt
##
## RLMerr = 1.9486, df = 1, p-value = 0.1627
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = eqn, data = ctsjoined)
## weights: fcpolys.wt
##
## RLMLag = 19.912, df = 1, p-value = 8.108e-06
## Spatial Durbin specification
      Likelihood ratio for spatial linear models

data:
Likelihood ratio = 5.5453, df = 4, p-value = 0.2358
sample estimates:
Log likelihood of Durbin      Log likelihood of SAR Lag_y
      -497.1923                      -499.9650

      Likelihood ratio for spatial linear models

data:
Likelihood ratio = 17.26, df = 4, p-value = 0.00172
sample estimates:
Log likelihood of Durbin      Log likelihood of Sar Error x
      -497.1923                      -505.8225

```

## S9. Bivariate Quasibinomial Model Output

```
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
## weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10209  -0.08543  -0.08224  -0.07680   0.29472
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.03954    0.19857   0.199   0.843
## B            -1.68830    1.76285  -0.958   0.340
##
## (Dispersion parameter for quasibinomial family taken to be 0.009350793)
##
## Null deviance: 1.3863  on 108  degrees of freedom
## Residual deviance: 1.3776  on 107  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
## weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10056  -0.08450  -0.08224  -0.08083   0.26993
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.09728    0.27232  -0.357   0.722
## C            0.38313    0.75594   0.507   0.613
##
## (Dispersion parameter for quasibinomial family taken to be 0.009347809)
##
## Null deviance: 1.3863  on 108  degrees of freedom
## Residual deviance: 1.3839  on 107  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
## weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09855  -0.08484  -0.08246  -0.07970   0.29327
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.06504    0.20422   0.318   0.751
## D            1.54169    1.46603   1.052   0.295
```

```

##
## (Dispersion parameter for quasibinomial family taken to be 0.009341332)
##
##      Null deviance: 1.3863  on 108  degrees of freedom
## Residual deviance: 1.3758  on 107  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
##
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16733  -0.06644  -0.03319  -0.01808   0.44523
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.512      1.476  -5.766 7.94e-08 ***
## E              13.217      2.240   5.902 4.29e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.00775512)
##
##      Null deviance: 1.38629  on 108  degrees of freedom
## Residual deviance: 0.77033  on 107  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
##
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.14510  -0.08362  -0.07822  -0.07369   0.28749
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.04339    0.20139   0.215  0.8298
## F              4.16035    2.31495   1.797  0.0751 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.009554795)
##
##      Null deviance: 1.3863  on 108  degrees of freedom
## Residual deviance: 1.3508  on 107  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
##
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)

```

```

##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11507 -0.08262 -0.08069 -0.07867  0.29551
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1525     0.2356  -0.647   0.519
## G              2.4230     2.1289   1.138   0.258
##
## (Dispersion parameter for quasibinomial family taken to be 0.009374939)
##
## Null deviance: 1.3863  on 108  degrees of freedom
## Residual deviance: 1.3738  on 107  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16621 -0.09182 -0.07678 -0.04433  0.32592
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3914     0.2273   1.722 0.08803 .
## H             -9.1540     2.9878  -3.064 0.00277 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.009344847)
##
## Null deviance: 1.3863  on 108  degrees of freedom
## Residual deviance: 1.2289  on 107  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.12313 -0.08737 -0.07336 -0.06468  0.32023
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.6065     0.2687  -2.258 0.0260 *
## I              9.4167     2.9249   3.219 0.0017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.009109645)
##
## Null deviance: 1.3863  on 108  degrees of freedom

```

```

## Residual deviance: 1.2742 on 107 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17283  -0.07002  -0.06175  -0.04700   0.46853
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.4713     0.6306  -3.919 0.000157 ***
## J              8.0990     1.9435   4.167 6.27e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.01110521)
##
##      Null deviance: 1.3863 on 108 degrees of freedom
## Residual deviance: 1.0913 on 107 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18573  -0.07079  -0.04539  -0.02869   0.39787
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.300     0.345  -3.768 0.00027 ***
## K              19.648     3.839   5.118 1.37e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.009028973)
##
##      Null deviance: 1.38629 on 108 degrees of freedom
## Residual deviance: 0.92069 on 107 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16948  -0.09125  -0.07696  -0.04886   0.32319

```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3868      0.2282   1.695  0.0931 .
## L            -8.9918      2.9910  -3.006  0.0033 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.00942054)
##
## Null deviance: 1.3863 on 108 degrees of freedom
## Residual deviance: 1.2328 on 107 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.08679 -0.08459 -0.08367 -0.08268  0.28193
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.04565   0.22479   0.203  0.839
## M            -0.11983   0.30137  -0.398  0.692
##
## (Dispersion parameter for quasibinomial family taken to be 0.009345757)
##
## Null deviance: 1.3863 on 108 degrees of freedom
## Residual deviance: 1.3848 on 107 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = exps[[i]], family = quasibinomial(), data = ctsjoined,
##      weights = ctsjoined$T5weights)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.08841 -0.08480 -0.08388 -0.08216  0.28188
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05172   0.22766   0.227  0.821
## N            0.10372   0.24127   0.430  0.668
##
## (Dispersion parameter for quasibinomial family taken to be 0.009345861)
##
## Null deviance: 1.3863 on 108 degrees of freedom
## Residual deviance: 1.3846 on 107 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
library(caret)

```

## S10 – Spatial Error Model

```
## Call:errorsarlm(formula = eqn, data = ctsjoined, listw = fcpolys.wt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -56.15147 -17.84052  0.83107  16.44645  66.38891
##
## Type: error
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.8725     10.1757  0.1840 0.8540000
## D              55.8286     23.8950  2.3364 0.0194696
## E              60.7683     16.8418  3.6082 0.0003083
## J              21.4972     20.0775  1.0707 0.2843007
## K              100.4790     29.6192  3.3924 0.0006929
##
## Lambda: 0.51552, LR test value: 11.766, p-value: 0.00060337
## Asymptotic standard error: 0.1105
##      z-value: 4.6655, p-value: 3.0791e-06
## Wald statistic: 21.767, p-value: 3.0791e-06
##
## Log likelihood: -505.8225 for error model
## ML residual variance (sigma squared): 594.06, (sigma: 24.373)
## Number of observations: 109
## Number of parameters estimated: 7
## AIC: 1025.6, (AIC for lm: 1035.4)
```