

Using web data for evidence-informed event management with R

Yoo Young Lee, Web and Digital Initiatives Librarian, uOttawa Library, yooyoung.lee@uOttawa.ca (@yooylee)

Cecilia Tellis, Head, Design and Outreach, uOttawa Library, cecilia.tellis@uOttawa.ca (@ceejaytee_ott)

DOI 10.17605/OSF.IO/P4NFB

Objectives

- To use previous event data to better plan library outreach activities, optimize library space, and assist in offering alternative spaces to requestors
- To uncover trends and patterns in event data at uOttawa
- To get a bird's eye view of events offered over time at uOttawa

Background and context

The exploratory data analysis undertaken was rooted in our desire to practice more evidence-informed library outreach (Somerville & Kloda, 2016). Engagement with students, staff and faculty is increasingly done through events and activities, alongside more typical workshop and classroom instruction, and liaison librarian work. Academic libraries have recognized the need to create innovative spaces that encourage, promote and support experiential learning, cross-disciplinary collaboration and scholarship, and opportunities to partner with other services on campus that support the student experience. But with these spaces come requests from campus partners, and while libraries tend to want to say “yes” at every opportunity for strategic partnerships, Owens and Bishop (2018) quite rightly note that while “this doesn’t mean agreeing to every request that comes in; it does mean helping facilitate and problem solve by suggesting alternatives” (p. 81).

Best practices for library programming and related topics like library marketing continue to be well researched and discussed in lively in-person and online forums. There is also a new and growing body of library literature on assessing and measuring the value of outreach initiatives which its emphasis on analysis *post* the outreach event (Garcia, 2013; Farrell & Mastel, 2016; German & LeMire, 2018). However, we found a lack of literature at the intersection of “library programming” and “data analysis” at the *pre*-outreach stage.

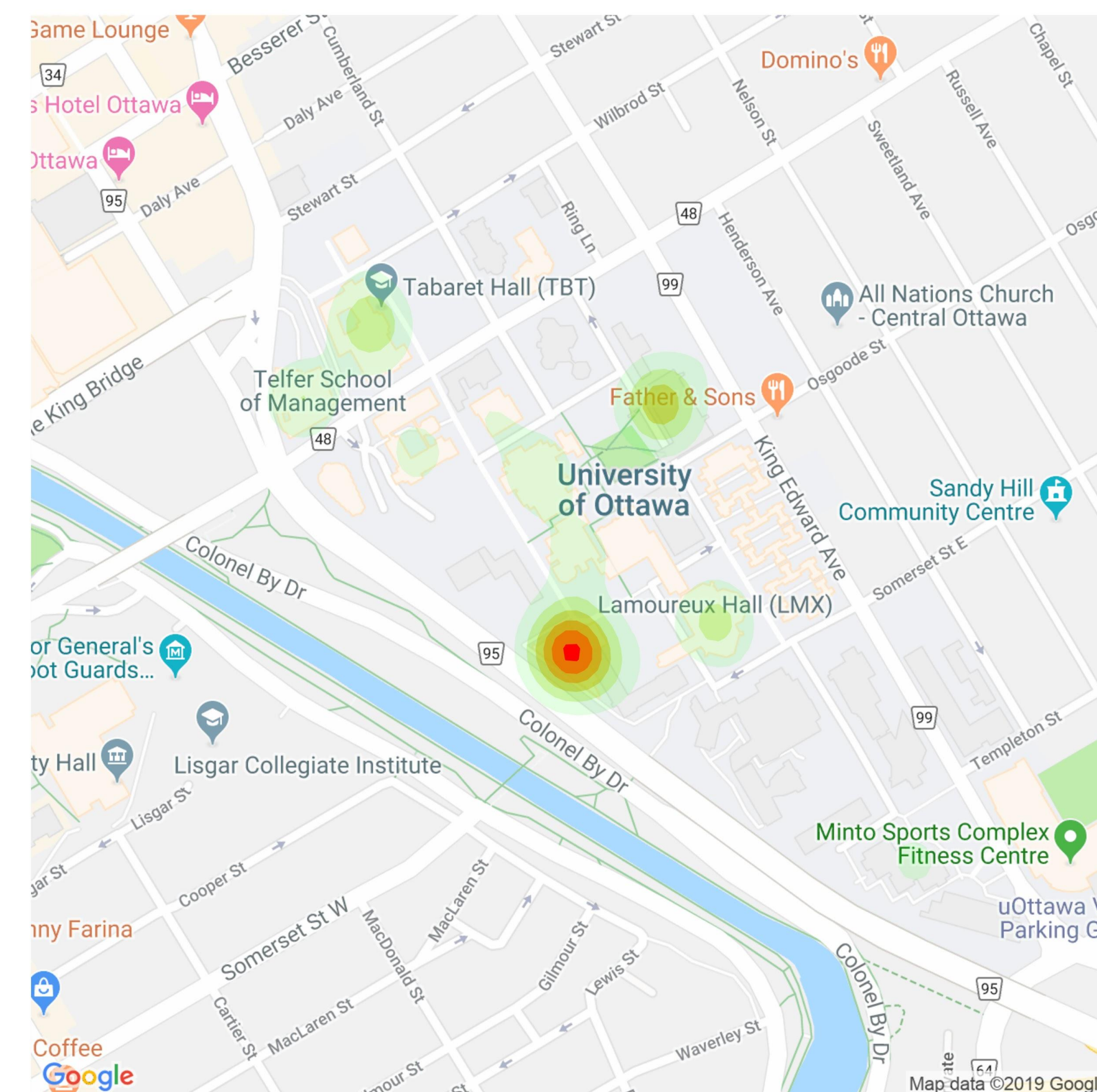
This led to exploring an approach to using data mining techniques. There is no doubt that there is value that can be derived from mining data from openly accessible websites and databases. It can even be said that the insights are “amazing” (Carpenter, 2016). In our view, the use of data scraped from the “official” event calendar of our institution, and analyzed using a variety of R packages, revealed patterns that we can use to adjust our programming. As Nicholson (2006) explains, “[t]hese patterns are then seeds for more thorough explorations, which may result in new visualizations, new reports, and even new aggregate measures” (p.794).

Selected Literature Cited

- Carpenter, T. (2016, February 11). Text and Data Mining Are Growing and Publishers Need to Support Their Use – An AAP-PSP Panel Report. Retrieved February 10, 2019, from <https://scholarlykitchen.sspnet.org/2016/02/11/text-and-data-mining-are-growing-and-publishers-need-to-support-their-use-an-aap-psp-panel-report/>
- Nicholson, S. (2006). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing & Management*, 42(3), 785–804. <https://doi.org/10.1016/j.ipm.2005.05.008>
- Somerville, M.M. and Kloda, L. A. (2016) Academic libraries. In: D. Koufogiannakis & A. Brettle (Eds.), *Being evidence based in library and information practice*. Facet, London, pp. 93-104

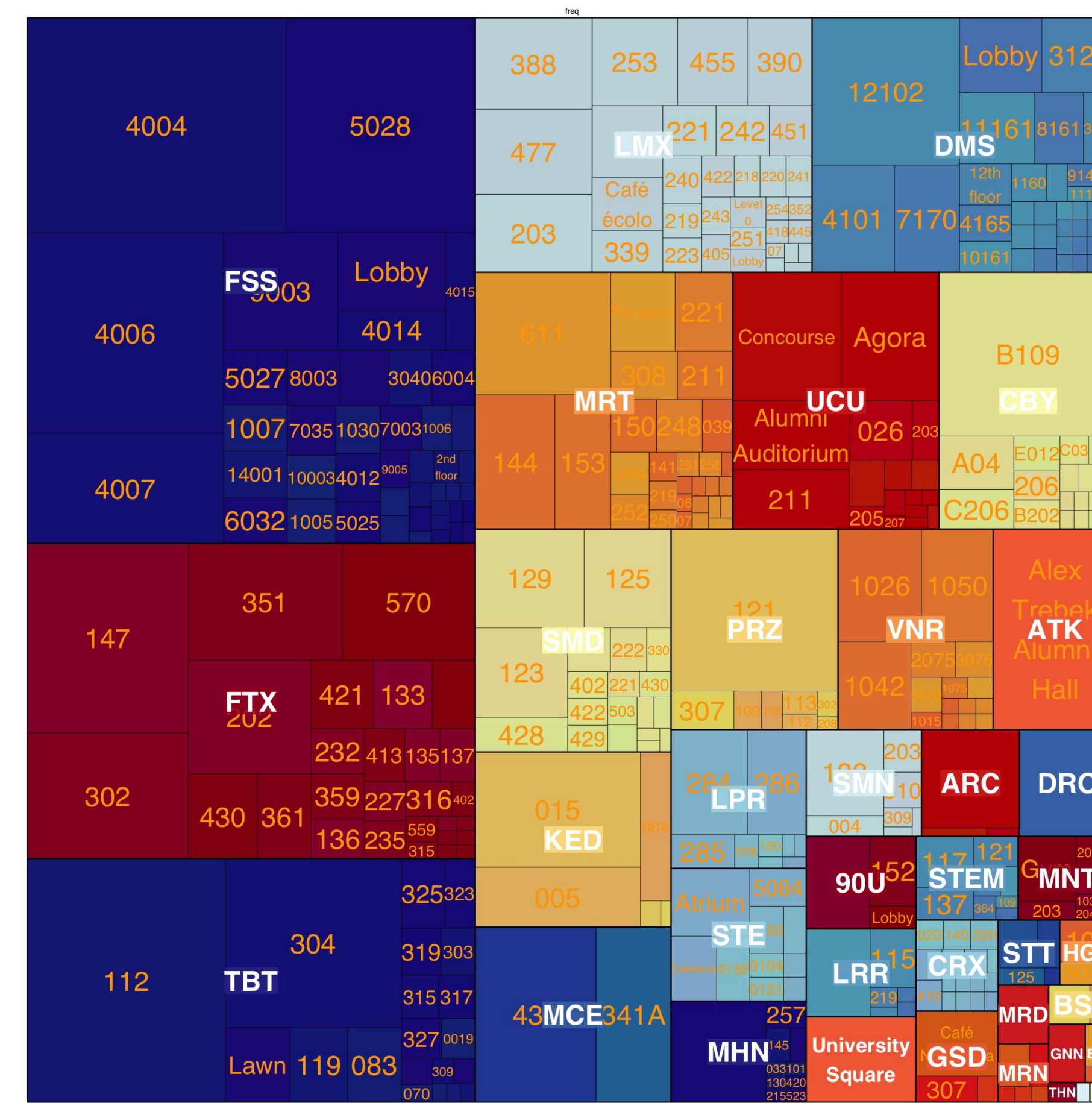
Data Visualizations

Location



Source: Based on all data

Room



Source: Based on all data

Event Categories

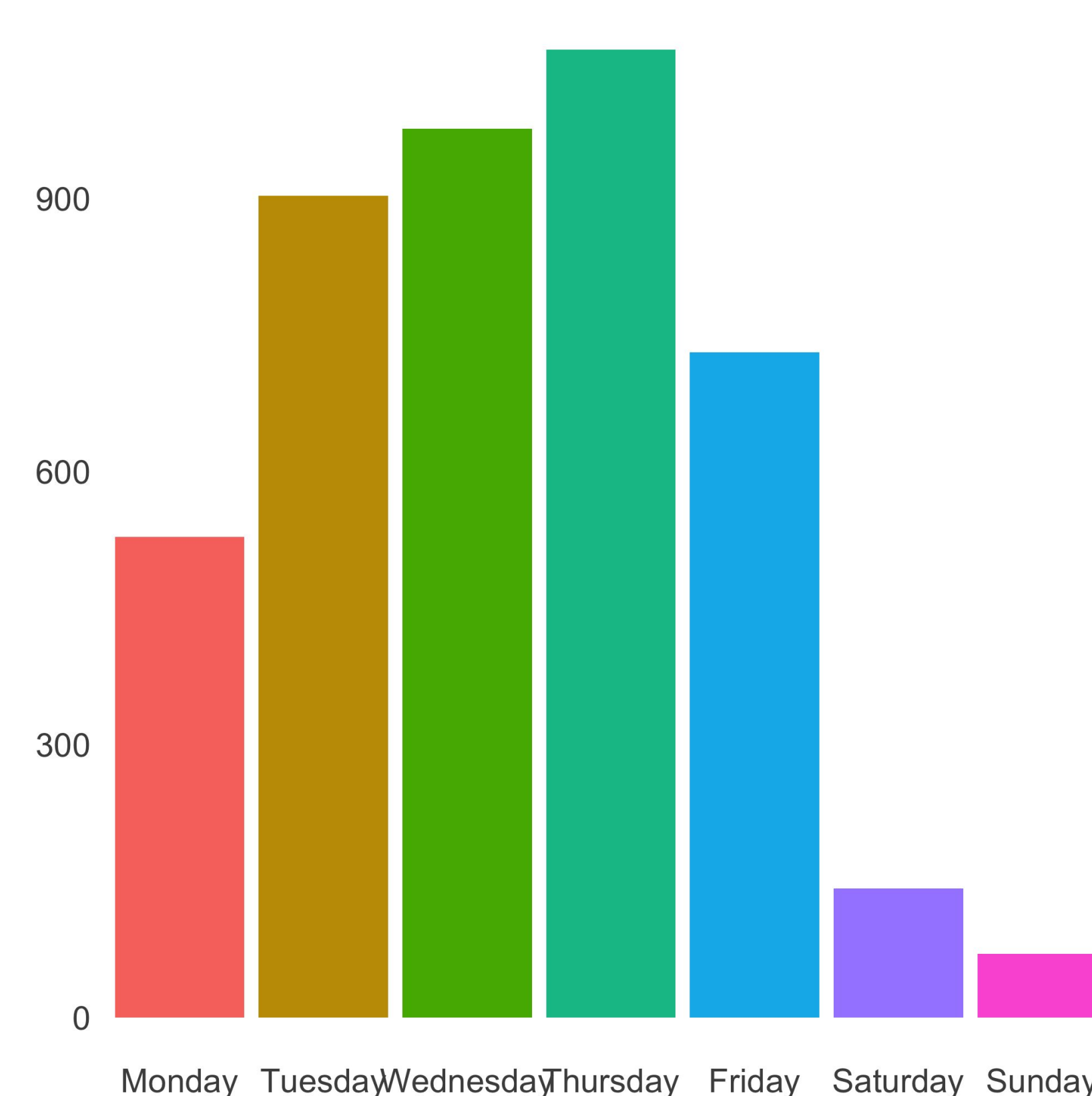


Source: Based on all data

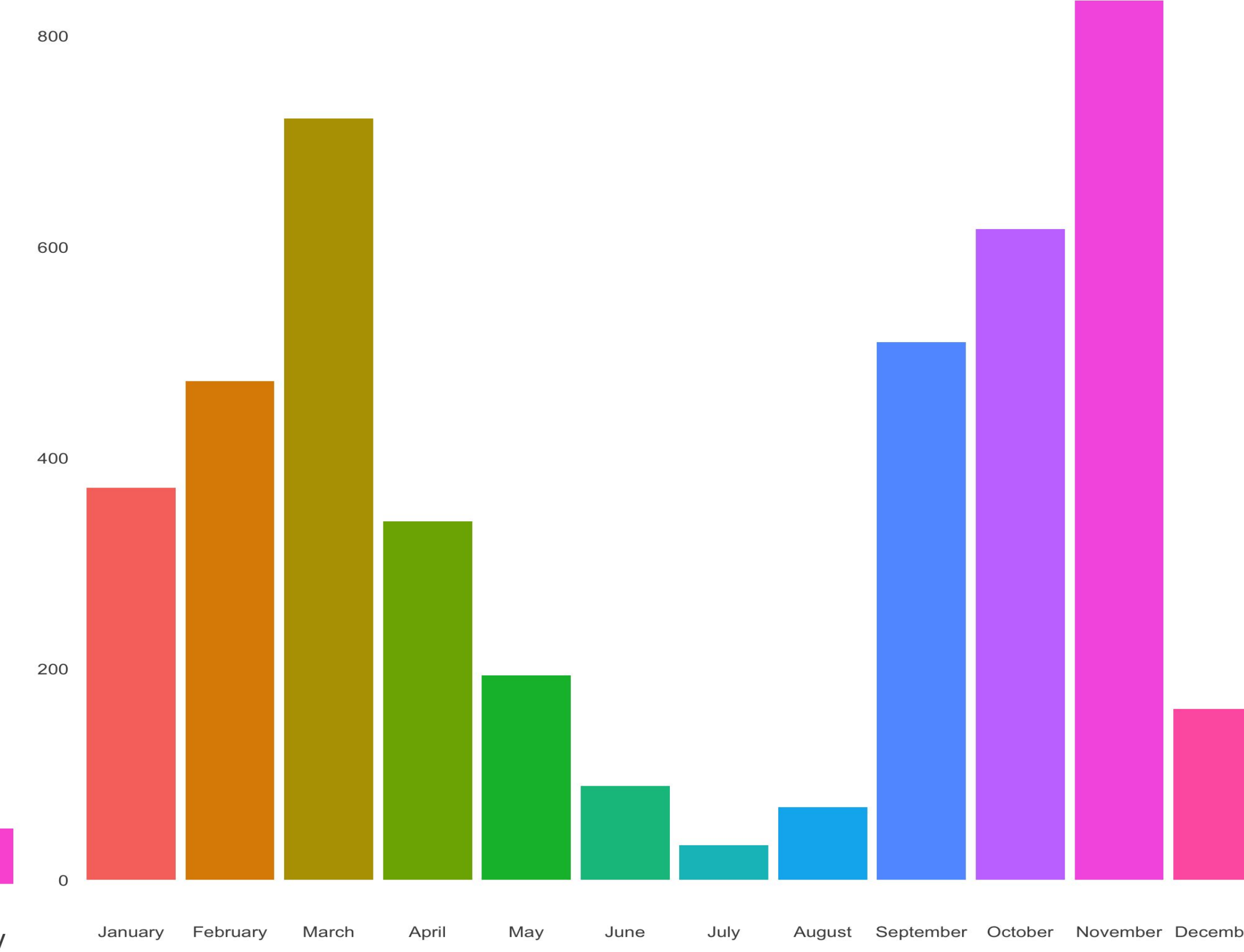


Source: Based on Library building data

Days



Months



Methodology

1. Scrape uoCal data from the website

- Downloaded 10,757 event data records on Nov. 22, 2018 from 2011 to 2019
 - Attributes: Title, Time, Date, Event categories, Organizations, Location, Room
- R packages: [Rcrawler](#) using ContentScraper
- Computing environment: Compute Canada

2. Clean uoCal data

- Removed duplicated records (EN|FR)
 - Reformatted date, time, location, room (EN|FR)
- Removed records held outside uOttawa main campus
- Removed records where date, time, location, room were empty
- Added information about latitude and longitude based on location
- Total records: 4415
- R packages: [plyr](#), [dplyr](#), [stringr](#)

3. Wrangle uoCal data

- Reshaped data for date field to day, month, and year fields using `::separate`
- Gathered event category columns into rows using `::gather` as there were event records associated with multiple events
- R packages: [tidyr](#)

4. Analyze and visualize uoCal data

- Location: Geocoding with Google Maps
 - R packages: [ggmap](#), [ggplot2](#)
 - Google Maps API key
- Room: Treemap
 - R package: [treemap](#)
- Days and Months: Barchart
 - R package: [ggplot2](#)
- Time: radar plot
 - R package: [radarchart](#)
- Event categories: Word cloud
 - R packages: [wordcloud](#), [tm](#), [RColorBrewer](#)