Mapping Toronto's Libraries*

Are Libraries Distributed Equally Throughout the City?

Hadi Ahmad

24 January 2024

This report analyzes the Library Branch General Information dataset, which is made available to the public from Open Data Toronto. It uses the library branches, square footage, and ward information to determine how well libraries are spread across the city. In summary, [TO DO].

Table of contents

---

*Code and data are available at: https://github.com/hadi-q/toronto-public-libraries.

1 Introduction

Libraries are a critical form of social infrastructure in modern society. They serve many functions, ranging from providing access to books and WiFi, to programs on self-improvement, to providing places of shelter to study or work. More importantly, however, they offer most or all of this functionality for free. Thus, libraries are not just places for any specific demographic: they act as third spaces, allowing users of different ages, ethnicities, and backgrounds to all exist in the same space. In a society where most services are monetized or paid, libraries remain a vestige of a more egalitarian era.

The Toronto Public Library in particular excels in these functions due to its size, cross-branch integration, and volume of services available. With 100 physical branches in its network, users can borrow books from any branch and have it delivered to their home branch. This increases the access to information that Toronto Public Library users benefit from. As well, users can borrow movies, internet hotspot devices, video game consoles, and even get free passes for museums or other attractions. A Toronto Public Library card also gives users access to third party platforms like Libby, OverDrive, Hoopla, and Consumer Reports, providing additional online resources that users can benefit from.

In October 2023, the Toronto Public Library was victim to a major cybersecurity attack that took down their website and online systems. This incident also led to some user data being compromised to hackers. As of January 2024, their website services remain offline. This prevents users from being able to borrow or renew books online, create or renew a library card, or access many of Toronto Public Library's services. As a result, physical proximity to a library branch has become more important than ever before, as it remains the only way to borrow books or access other library services until full services are restored.

This paper aims to analyze the distribution of Toronto Public Library branches across the 25 wards of Toronto. The distribution of branches will be viewed in two ways: number of branches and square feet of branch space within a given ward. The resulting analyses will determine if certain wards are over- or underrepresented in terms of the quantity and area of

library branches. Future analyses may explore demographic factors to determine if libraries are disproportionately located in neighbourhoods of higher or lower income, to get a sense of equality of access to libraries.

The remainder of this paper is structured as follows. Section 2

2 Data

All relevant data was sourced from the Open Data Toronto portal, and extracted using the opendatatoronto library for R (Gelfand 2022). In particular, one dataset was used for the graphs and analyses in this paper [TO DO]

(Table commented out due to unresolved errors)

```
# #| label: library-branches
# #| fig-cap: Sample of dataset
# #| echo: false

# analysis_data |>
#   kable(col.names = c("Branch", "Square Footage", "Ward Number", "Ward Name", "Year Buil
#     digits = 2)

# make column names syntactically valid; copied from https://stackoverflow.com/questions/1
# analysis_data <- make.names(analysis_data)
# copied from https://github.com/EthanSansom/torontocitytrees/blob/main/outputs/paper/pape
# error: Error in dimnames(x) <- dn :
#  length of 'dimnames' [2] not equal to array extent
# see here: https://stackoverflow.com/questions/12985653/what-does-length-of-dimnames-1-no

# head(analysis_data, 12) |>
#   kable(
#     col.names = c("Branch", "Square Footage", "Ward Number", "Ward Name", "Year Built"),
#     booktabs = TRUE
#   )
```

Talk more about it.

And also planes (fig-planes). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

```
# label: fig-planes
# fig-cap: Relationship between wing length and width
# echo: false
# warning: false
# message: false

# analysis_data <- read_csv(here::here("outputs/data/analysis_data.csv"))

# analysis_data |>
#   ggplot(aes(x = width, y = length)) +
#   geom_point(alpha = 0.8) +
#   theme_minimal() +
#   labs(x = "Wing width (mm)",
#        y = "Wing length (mm)")
```

Talk way more about it.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in [Appendix -sec-model-details].

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in tbl-modelresults.

```
# echo: false
# eval: true
# warning: false
# message: false

# library(rstanarm)

# first_model <-
#   readRDS(file = here::here("outputs/models/first_model.rds"))
```

```
# echo: false
# eval: true
# label: tbl-modelresults
# tbl-cap: "Explanatory models of flight time based on wing width and wing length"
# warning: false

# modelsummary::modelsummary(
#   list(
#     "First model" = first_model
#   ),
#   statistic = "mad",
#   fmt = 2
# )
```

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In fig-ppcheckandposteriorvsprior-1 we implement a posterior predictive check. This shows...

In fig-ppcheckandposteriorvsprior-2 we compare the posterior with the prior. This shows...

```
# eval: true
#| echo: false
#| message: false
#| warning: false
# label: fig-ppcheckandposteriorvsprior
# layout-ncol: 2
# fig-cap: "Examining how the model fits, and is affected by, the data"
# fig-subcap: ["Posterior prediction check", "Comparing the posterior with the prior"]

# pp_check(first_model) +
#   theme_classic() +
#   theme(legend.position = "bottom")

# posterior_vs_prior(first_model) +
#   theme_minimal() +
#   scale_color_brewer(palette = "Set1") +
#   theme(legend.position = "bottom") +
#   coord_flip()
```

B.2 Diagnostics

C fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

D fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

E LLMs

Statement on LLM usage: no LLMs were used in the making of this paper.

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.