Mapping Toronto's Libraries, Part 2*

Understanding the Relationship Between Income and Library Locations

Hadi Ahmad

25 April 2024

The report analyzes the Library Branch General Information dataset from Open Data Toronto, and adds additional insights from StatCan data about incomes and other variables at a neighbourhood or census tract level. The goal of this paper is to build on the original paper's analysis to determine whether there is a correlation between a neighbourhood's income and its accessibility to Toronto Public Library branches.

Table of contents

---

*Code and data are available at: https://github.com/hadi-q/toronto-public-libraries-income. This paper builds on the work of a previous paper by the author, at: https://github.com/hadi-q/toronto-public-libraries.

1 Introduction

It is difficult to underestimate the role that libraries have played in shaping knowledge in the modern world. From the stores of the Library of Alexandria in centuries past, to the freely accessible Victorian England institutions that brought class and sophistication to ordinary citizens and not just members of nobility, libraries have repeatedly been staples of advanced societies, serving as bellwethers of progress. With the functions of providing information and education, access to resources, systems to categorize and archive materials, and creating diverse social environments, they transcend class and cultural barriers to provide public access to resources.

Beyond their typical functions, libraries serve as community spaces, where residents of a neighbourhood can come together, meet, socialize, and learn together, free from the shackles of a broader late-capitalistic society that aims to monetize all free real estate. This highlights the importance of libraries within local communities, in terms of proximity and space made available to patrons. It stands to reason, then, that more accessible libraries would allow more people to access their services.

But are all libraries and their resources made equal? Is it better to have more branches with less space, or fewer branches with greater capacity? Who stands to benefit the most from libraries, and do they really get those opportunities? These are the questions that led to the creation of both the previous paper and this one.

The Toronto Public Library excels in its function compared to most public library systems. Creating a network of interconnected branches allows patrons to freely access resources across branches, with a particular focus on its online system and third party subscriptions provided for free to cardholders. This reduces physical barriers to accessing information. However, at the time of publication of the previous paper, the Toronto Public Library system had fallen victim to a major cybersecurity attack that took down most services from October 2023 to January 2024. At the end of January, some core online services were reinstated, bringing respite to residents who were previously forced to access the then-limited amenities solely in branch. By the publication of this paper in April 2024, all Toronto Public Library systems have been fully restored, marking an end to roughly 6 months of access limitations to services like finding books from the catalog or renewing library cards.

The previous paper analyzed the distribution of the Toronto Public Library system's 100 branches to determine how well distributed its branches are at a ward level. Its key findings included the fact that, despite large variances in the physical size of branches, there was less variance in the square footage of library space available per ward. However, it did not consider other factors that could determine accessibility of branches, like proximity to public transportation or the income of the neighbourhoods that do have branches compared to those that do not. This paper attempts to address that shortfall, by using household income and population data from Statistics Canada and introducing variables like median and average household income and population density.

To build on the work of the previous paper, this paper will use Statistics Canada data on Toronto districts and compare it to `opendatatoronto` data on the distribution and details of Toronto's libraries. In particular, by focusing on income and broader demographic factors, this paper aims to determine the role that income and population density plays in the distribution of library branches across the city. Further analyses may look at transit accessibility to see if libraries really are accessible to all.

The Data Section 2 of this paper explores the `opendatatoronto` and Statistics Canada data holistically, outlining tools and methods used to analyze the data and some preliminary observations from the data. The Results Section 3 introduces more observations found from the data analysis. The Discussion Section 4 ties back the findings to its real world relevance. Lastly, the Conclusion Section 5 summarizes key findings and outlines future areas of study.

1.1 Approach

To determine the equality of access of libraries in Toronto, wealth is a key consideration. After all, libraries are more likely to be used by those from less wealthy backgrounds, those who cannot afford to buy copies of books and keep them at home. But how does one measure wealth? By definition, wealth is a measure of total assets owned by an individual. This sort of data would have to consider assets like real estate and stocks, while also deducting liabilities like mortgages or lines of credit. Information of this nature is not readily available, and even if it was, it could pose privacy issues to individuals. Aggregate data could be used instead, such as by categorizing neighbourhoods or wards by wealth. That begs the question: how else can wealth be measured at sub-municipal levels?

One consideration is real estate prices. Colloquially, neighbourhoods often get classified as wealthy, middle class, or working class by residents. In the case of Toronto, Rosedale and Forest Hill may be considered wealthy, while Agincourt may be considered middle class, and Crescent Town working class. But these neighbourhood divisions do not neatly line up with available data. For instance, Toronto Regional Real Estate Board's (TRREB) quarterly report on community pricing aggregates some neighbourhoods, like Rosedale-Moore Park and Agincourt South-Malvern West. These classifications are different from Toronto's own 25 ward divisions and Statistics Canada's census tracts. Thus, the data cannot be compared across sources. As well, the TRREB spreadsheet in question was last published in the fourth quarter of 2021 - newer monthly reports are available only as formatted PDFs, which make data extraction a challenge. Even still, the spreadsheet version of the report is access restricted to only TRREB members, not the general public.

Beyond the incongruencies in data, real estate prices may not be a great way to compare wealth. Some neighbourhoods may have higher prices, but only because the properties themselves are larger. Price per square foot would thus be a better metric, as it accounts for differences in interior space. Unfortunately, square footage is rarely listed on resale properties in Ontario. While price per square foot is a metric frequently used in new condo sales, resale properties

make up the vast proportion of real estate sales in Toronto, and thus only relying on new figures would not be reflective of the market. Even if price per square foot was considered, there can be great variance the number of rooms in a place of a certain square footage. For instance, 700 sq ft could represent a 1 bedroom apartment in one building or a 2 bedroom + den unit in another building. This means that there can be large differences in the utility of a space beyond just its square footage. So average price for units of each distinct bedroom count could be a better metric than purely average selling prices, but that information is only available at the aggregate municipal level, not by ward. Even still, the utility of a space could come in various forms: houses may have backyards, attics, or basements, while condos may have balconies attached to the unit or common elements shared by all residents. As a result, forms of utility are also not created evenly, challenging any means to compare them.

One feasible way to consider wealth beyond the aforementioned pitfalls is to use income. Statistics Canada makes income data available at a census tract, ward, municipal, provincial, or national level, with detailed breakdowns of pre-tax and post-tax individual and household income. Thus, quality data is available to compare to library distribution data obtained from the previous paper. Federal electoral district data as published by Statistics Canada also matches up with ward data from Toronto, aligning to the 25 ward model used previously. While income is not a guaranteed means to compare wealth - it is possible to have high income and low net worth, low income and high net worth, or other combinations in between - it serves as a reasonable stand-in.

2 Data

Data related to Toronto Public Library branches 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, which included information about all Toronto Public Library branches, their addresses, phone numbers, square footage, year of construction of the building, among other variables like whether there as a park adjacent to the library, if adult literacy classes are available, and GPS coordinates of the building. Statistics Canada data was obtained from their 2021 Census of Population through their website. This data includes income, demographics, languages spoken, among other variables for each ward or electoral district in Canada.

2.1 Data Tools

The data was generated, extracted, and cleaned using R (R Core Team 2022), leveraging functions from `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2023), `readr` (Wickham, Hester, and Bryan 2023), `janitor` (Firke 2023), `opendatatoronto` (Gelfand 2022), `knitr` (Xie 2014), `here` (Müller 2020), and `arrow` (Richardson et al. 2024).

2.2 Overview of Datasets

The Toronto Public Library dataset was published by the Toronto Public Library to share information on its network, and was last updated on 29 June, 2023 in accordance with its annual updating guideline. The chosen variables of analysis were the branch name, the square footage of the library, the ward number and ward name of where the library is located, and the year of construction of the site. The Statistics Canada dataset was released on 9 February, 2022 after the 2021 Census was conducted, and was last updated on 15 November, 2023. Canadian censuses are conducted every 5 years, and so this dataset also references the previous 2016 census values.

`Table 1` captures details about library branches in the city. Across the 100 branches, size ranges from just 554 square feet to over 400,000 square feet. This suggests a greater range in sizes of libraries, but at a macro level, what's more important is how those branches are allocated across wards. Aggregating data across wards is more useful for making comparisons across the city.

Branch	Square Footage	Ward Number	Ward Name	Year Built
Albion	29000	1	Etobicoke North	2017
Albert Campbell	28957	20	Scarborough Southwest	1971
Alderwood	7341	3	Etobicoke-Lakeshore	1999
Agincourt	27000	22	Scarborough-Agincourt	1991
Armour Heights	2988	8	Eglinton-Lawrence	1982
Annette Street	7806	4	Parkdale-High Park	1908

Table 1: Sample of Toronto Public Library Data

In the previous paper, dataframes on library branches per ward and square footage per ward were merged into a single dataframe using the `merge` function. This was done to better understand any links between the number of branches and square footage of library space. Across the city, the number of branches per ward varied from 1 (in the case of Willowdale, home to the North York Central Library), up to 7 branches, with a mean and median of 4. By comparison, square footage of library space per ward ranged from 24,444 square feet in Scarborough North to 494,632 square feet in University-Rosedale (thanks to a significant outlier, the Toronto Reference Library). This amounted to a mean of 72,517 square feet per ward, or a median of 44,303 square feet per ward. `Table 2`, below, aggregates data into counts at a ward level.

Ward Name	Number of Branches	Total Square Footage of Branches
Beaches-East York	4	33021
Davenport	4	40312
Don Valley East	4	42129

Ward Name	Number of Branches	Total Square Footage of Branches
Don Valley North	4	97349
Don Valley West	4	40510
Eglinton-Lawrence	4	91449
Etobicoke-Lakeshore	6	61053
Etobicoke Centre	3	63055
Etobicoke North	4	43772
Humber River-Black Creek	5	68108
Parkdale-High Park	6	65855
Scarborough-Agincourt	3	40453
Scarborough-Guildwood	2	36083
Scarborough-Rouge Park	4	44834
Scarborough Centre	5	49252
Scarborough North	3	24444
Scarborough Southwest	4	48816
Spadina-Fort York	3	33784
Toronto-Danforth	7	57331
Toronto-St. Paul's	5	78705
Toronto Centre	3	27267
University-Rosedale	6	494632
Willowdale	1	168022
York Centre	2	26882
York South-Weston	4	35814

Table 2: Count and Total Square Footage of Library Branches by Ward

One issue with `Table 2` is that square footage of library space at a ward level is still a relatively meaningless count. How that space is utilized by patrons is a more important consideration, because that can be used to assess equality of access. After all, higher populations put greater pressure on resources, which can impact the ability for patrons to use the library's services. Thus, it is important to also analyze the population and population density of a ward, as `Table 3` illustrates:

Ward	Population, 2021	Population density per square kilometre
Beaches-East York	109359	6531.2
Davenport	105946	8769.1
Don Valley East	95039	4171.8
Don Valley North	113663	4669.6
Don Valley West	101959	3386.0
Eglinton-Lawrence	115832	5114.3

Table 3: Population and Population Density by Ward

While variance in population is not significant between wards - values range from 94,717 to 141,751 - there is greater variation in population density. This is due to differences in housing typologies and the size of wards. In the most populated ward, Etobicoke-Lakeshore, there has been a steady increase in the amount of condos, stacked townhomes, and other higher density housing being built in proximity to the waterfront. As a result, it has also seen a substantial 9.8% gain in population over the period from 2016 to 2021.

Ward	Branches	Total Sq Ft	Population, 2021	Population Density
Beaches-East York	4	33021	109359	6531.2
Davenport	4	40312	105946	8769.1
Don Valley East	4	42129	95039	4171.8
Don Valley North	4	97349	113663	4669.6
Don Valley West	4	40510	101959	3386.0
Eglinton-Lawrence	4	91449	115832	5114.3

Table 4: Branches and Population by Ward

Putting tables 2 and 3 together produces a more cohesive reference guide, as seen in `Table 4`. By combining the number of branches, square footage of branch space, and population, it becomes easier to see which branches are burdened by having more population in proportion to available library space.

3 Results

Based on the number of branches per ward, as outlined in `Figure 1`, there is a fairly normal distribution of branch counts across the wards of the city. In the context being, having an equal amount of branches per ward would appear to be ideal, but the devil is of course in the details.

The distribution of square footage of library space is the more important consideration than pure branch counts here. The ward with the greatest amount of library space, Ward 11 (University-Rosedale), is home to six libraries, yet it would still have the most library space even with only one branch: the Toronto Reference Library. Likewise, the second highest ward, Ward 18 (Willowdale), only has one library, yet it too retains a significant amount of library space. A total of 1,812,923 square feet of library space is available across the 100 branches, averaging out to 72,517 square feet per ward. After removing Toronto Reference Library, the mean falls to 54,929 square feet. The median is 44,303 square feet. See `Figure 2` for more details.
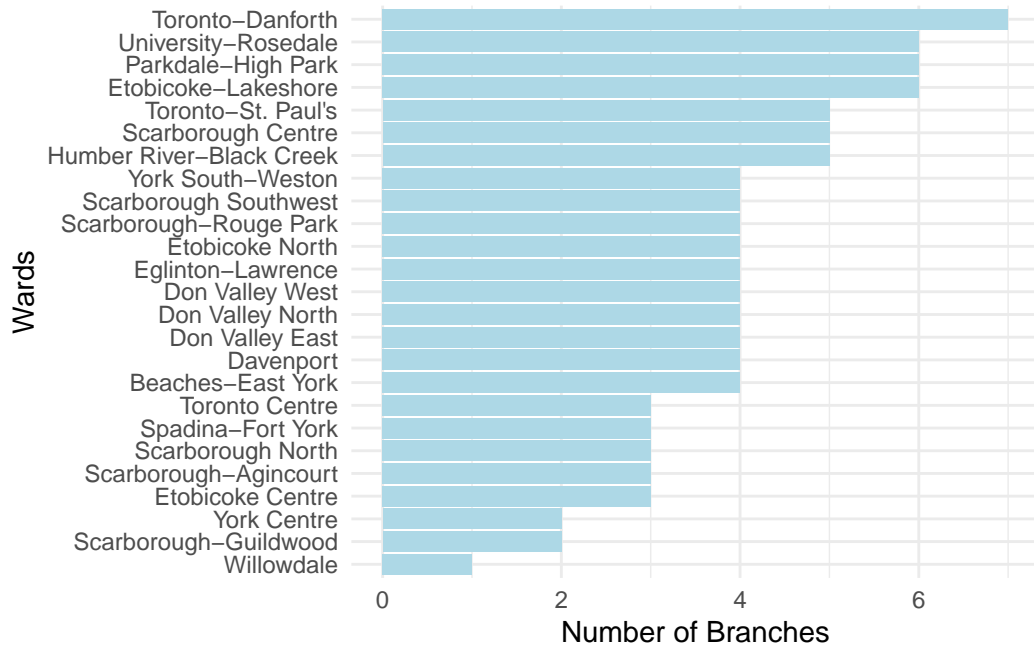
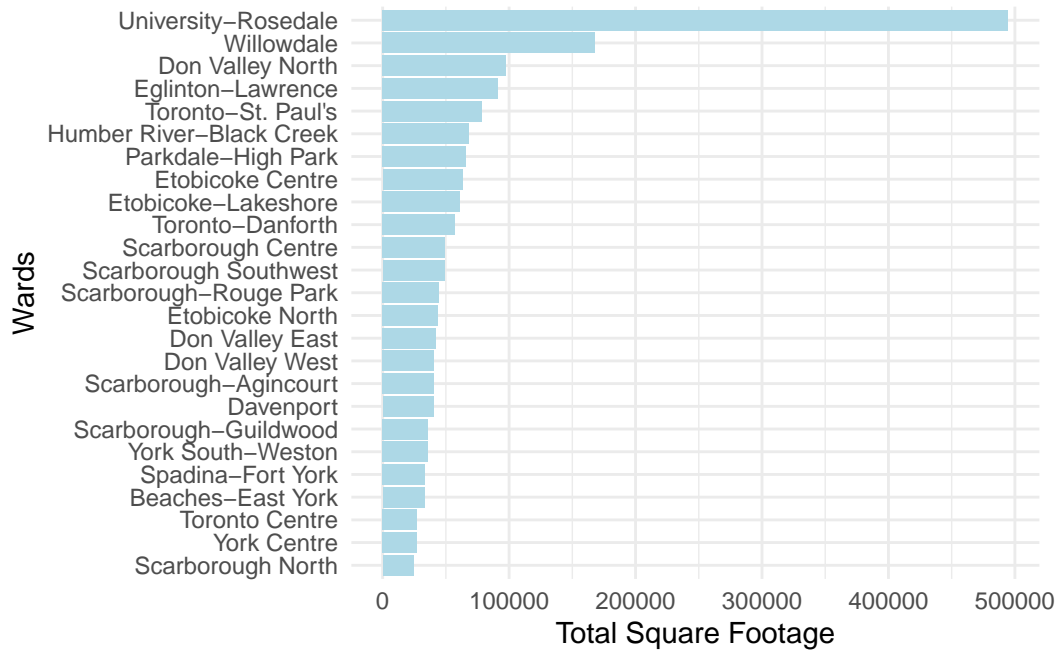Figure 1: Library Branches by Ward by Count

Figure 2: Library Branches by Ward by Square Footage

Determining the square footage of library space available per person demonstrates where the inequalities lie. While York South-Weston's 35,814 square feet of space falls below the mean and median, it turns out that it is very well served by branch space. Beaches-East York, on the other hand, has similar amounts of library space (33,021 square feet), but is actually relatively deprived of branch space. This goes to show the importance of considering the people that make up a library's users when determining equality. After all, it is the patrons
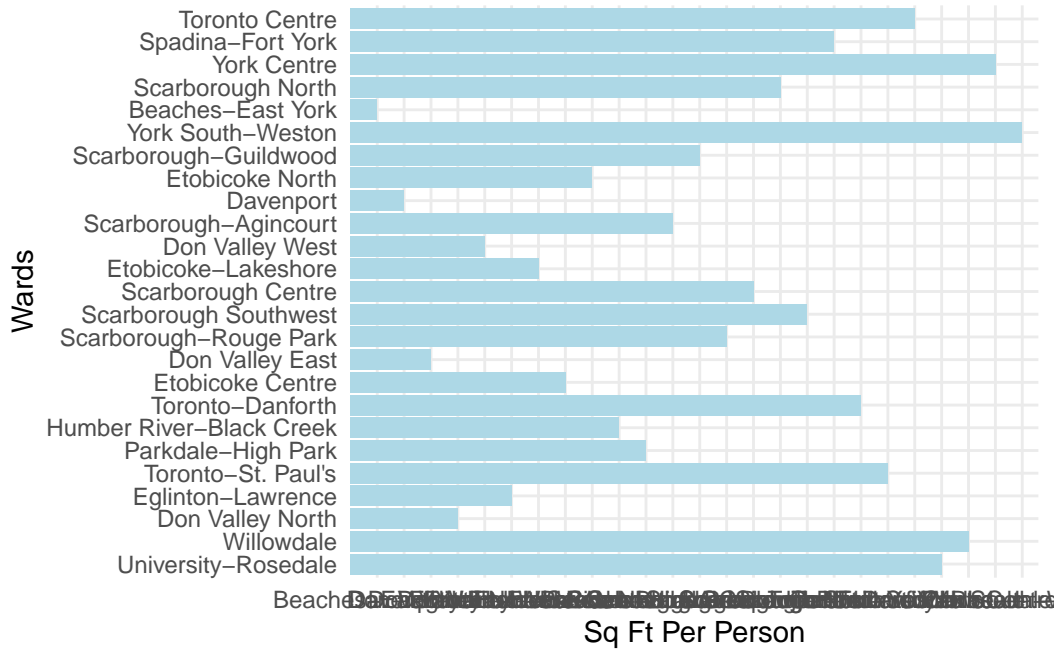

Figure 3: Library Branches by Square Footage Per Person

4 Discussion

5 Conclusion

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

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

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'.* https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.