Toronto's Most Studious Wards: An Analysis of Annual Toronto Public Library's Circulation and Visits by Wards Pre and Post-COVID*

Francesca Ye

January 25, 2024

Toronto Public Library data sets on branch information, annual circulation, and annual visits were used to examine library usage patterns by ward between 2018 and 2022. Based on the analysis of annual circulation and visits by ward over this five year period, there is evidence that library usage significantly dropped in 2020 when comparing values to 2018 and 2019. Looking beyond the initial outbreak of COVID-19, it also appears that library usage is beginning to recover, although not to pre-pandemic levels, which highlight differences in pre and post-pandemic library usage. It would be beneficial to expand the scope of the analysis to more years before and after 2020 to get a larger sample of library usage patterns in addition to analysis of monthly circulation and visitiation data against a timeline of government mandated COVID-19 measures.

Table of contents

*The full code and data used in this paper is available at: https://github.com/Francesca-Ye/tpl_activity_analysis

1 Introduction

Public libraries have an intrinsic value in communities regardless of if they are used for research, educational, entertainment, or recreational purposes. The availability of public libraries in addition to their material and programming are thought to have an impact on the quality of life in communities. Public libraries can influence a community's number of jobs, median household income, and level of educational attainment (Chow and Tian 2021). As a result of the outbreak of COVID-19 which was formally declared a pandemic on March 11, 2020 by the World Health Organization (Organization 2023), libraries had their operations affected. Ranging from full closures due to lock downs or social distancing measures which restricted capacity, the landscape upon which public libraries operated rapidly changed. Hamid R. Jamali and Philip Hider's "Changes in the Usage Pattern of Public Library Collections during the COVID Pandemic" in *Collection Management* looked into the usage patterns of two library networks in Sydney, Australia. They found that physical loans experienced an approximately half million decrease (Jamali and Hider 2023).

The purpose of this report is to identify and analyze the patterns of total circulation and visits to Toronto Public Library branches by ward between 2018 and 2022 to see if the COVID-19 pandemic and ensuing measures against it had an effect on the usage patterns of circulation and visitation metrics. This report utilized Toronto Public Library's annual circulation and visitation in addition to general information about each branch including the ward they reside in as data sources. To examine whether or not there are differences in pre and post-pandemic usage patterns, graphical analysis was conducted in order to see changes in usage over the five year period in addition to mean and standard deviation statistics. Report findings reveal that there is a shift in usage patterns starting in 2020 in comparison to data from 2018 and 2019. These variances in circulation and visitation data could be attributed to pandemic measures.

The Data section of this report aims to discuss the characteristics of the data sets used in addition to the methods of cleaning applied in order to get workable data. Succeeding, the Results section highlights trends and patterns discovered in statistical and graphical analysis. The Discussion section notes possible analytical limitations and measurement errors in the

data sets in addition to overall insights of the report's analysis. To summarize the report's content, the Conclusion section briefly touches on the overall insights that can be distilled.

2 Data

Data used in this report were taken from the Open Data Toronto Online Portal (Gelfand 2022). The specific data sets sourced from Open Data Toronto included *Library Branch General Information* (Toronto Public Library 2023a), *Library Circulation* (Toronto Public Library 2023b), and *Library Visits* (Toronto Public Library 2023c). The data sets were imported into RStudio where the open-source statistical programming language, R (R Core Team 2023), was used for cleaning and analysis. The raw data sets were extracted and saved with functionalities of the `readr` (Wickham, Hester, and Bryan 2023) package. Data cleaning was aided by the use of the `janitor` (Firke 2023), `tidyverse` (Wickham et al. 2019), and `dplyr` (Wickham et al. 2023) packages. To create tables and graphs for data visualization of the datasets, the `ggplot2` (Wickham et al. 2019) and `knitr` (Xie 2014) packages were used.

2.1 Library Branch General Information

This data set was published by the Toronto Public Library and records the current characteristics of its library branches. The data set is refreshed on an annual basis and includes data regarding branch code, branch name, materiality, address, postal code, website, telephone, square footage, parking spots, workstations, ward number, ward name, present site date, neighbourhood number, neighbourhood name, latitude, longitude, and services offered (Toronto Public Library 2023a). The data used in this report was captured and accurate as of January 19, 2024.

Basic data cleaning was first applied to the column value names in order to ensure readability. As this is a report focused on investigating annual circulation and visits data by ward and year, it was decided that rows of branches that had a value of `0` in the 'physical_branch' which indicated that they did not have a physical location that patrons could enter were removed from the data set. The reason for this removal was to focus on public library branches that were physically accessible to patrons. Additionally, `branch_code`, `ward_name`, and `ward_no` were the only column values that were kept in the cleaned data set. This is because the purpose of this data set was to provide a way for the branch codes present in the *Library Circulation* and *Library Visits* data sets to be assigned their respective ward number for graphical analysis. The final cleaned version of the *Library Branch General Information* the data set has these changes reflected (Table 1).

Table 1: Sample of Cleaned Library Branch Data

Branch Code	Ward Number	Ward Name
AB	1	Etobicoke North
ACD	20	Scarborough Southwest
AD	3	Etobicoke-Lakeshore
AG	22	Scarborough-Agincourt
AH	8	Eglinton-Lawrence
AN	4	Parkdale-High Park
AP	5	York South-Weston
BB	16	Don Valley East

2.2 Library Circulation

This data set was published by the Toronto Public Library and captures the annual circulation of materials at its various branches. The data set is refreshed on an annual basis and includes data regarding the year, branch code, and total circulation. The data used in this report was captured and accurate as of January 19, 2024. According to Open Data Toronto, circulation is counted as any first-time checkouts and renewals of material. This data set also covers physical and electronic materials in its count. Open Data Toronto also notes that circulation totals are a measure of how many of these aforementioned transactions were processed by each respective branch. They also note that online renewals of physical materials are associated with the home branch of the renewing customer (Toronto Public Library 2023b).

Basic cleaning was first applied to the column value names to ensure readability as a CSV file. Similar to what was done for the cleaning of *Library Branch General Information*, non-physical branches were removed from the data set in addition to the branch code of `VR` which represented electronic circulation totals. This is because the scope of the report focuses only on physical branches that are accessible to patrons. Rows that did not contain `year` values between 2012 and 2017 were also removed since the scope of this report focuses on a five-year span between 2018 and 2022. The `subset()` function was used to remove rows containing multiple argument values. (Bobbitt 2021) The aforementioned changes were reflected in the cleaned data set (Table 2).

Table 2: Sample of Cleaned Circulation Data

Year	Branch Code	Total Circulation
2018	AB	328536
2018	ACD	303855
2018	AD	155864
2018	AG	623148

Table 2: Sample of Cleaned Circulation Data

Year	Branch Code	Total Circulation
2018	AH	214013
2018	AN	201476
2018	AP	85101
2018	BB	156836

2.3 Library Visits

This data set was published by the Toronto Public Library and records the annual number of visits at each branch. The data set is refreshed on an annual basis and includes data regarding the year, branch code, and total visits. The data used in this report was captured and accurate as of January 19, 2024. According to Open Data Toronto, visits were measured using sensors at the entrances of library branches and counts were recorded during scheduled operating hours. They note that visit totals were sometimes estimated when sensor issues were occurring. They also state that since visits were based on entries to the branch, the same individuals might be counted multiple times due to re-entrance in one day (Toronto Public Library 2023c).

Basic cleaning was applied to this data set's column value names to ensure readability. Following what was done on the two previous data sets, non-physical branches were removed from the data set since the scope of the report is focused on physical branches accessible to patrons. Additionally, the `BKONE` branch code was removed since it represented the annual visits of the two bookmobiles. The reason for this removal was due to its lack of fixed physical presence which could be found within one specific ward in Toronto. Rows that did not contain `year` values between 2012 and 2017 were also removed since the scope of this report focuses on five years between 2018 and 2022. The `subset()` function was used to remove rows containing multiple argument values (Bobbitt 2021). These changes to the raw data set are viewable in the cleaned table (Table 3).

It is also important to note that this data set is incomplete. For 100 physical branches over five years of collections, there should be 500 rows; however, only 492 are present. For the purposes of analysis, no null or 0 value was created as a placeholder for the missing data to not skew the results. However, the absence of these data points in calculating ward totals is a possible limitation of the analysis results which will be further discussed.

Table 3: Sample of Cleaned Visits Data

Branch Code	Year	Total Visits
2018	AB	328536
2018	ACD	303855
2018	AD	155864

Table 3: Sample of Cleaned Visits Data

Branch Code	Year	Total Visits
2018	AG	623148
2018	AH	214013
2018	AN	201476
2018	AP	85101
2018	BB	156836

2.4 Annual Library Circulation and Visits by Ward

To get data sets that feature the annual circulation and visits total for each ward, the cleaned library branch data set was appended to each respective cleaned data set using the `merge()` function. Data sets were merged by `branch_code` which allowed for the corresponding ward numbers found in the library branch data set to be appended to the rows of annual total data that are sorted by branch code. (Bobbitt 2022) After the data sets were merged, they were grouped by ward number and then a year to be summarized by total visits or circulation. This process resulted in a data set producing the total annual circulation number for each ward and year between 2018 and 2022 as seen in its table (Table 4). It also produced a data set producing the total annual visit number for each ward and year between 2018 and 2022 as seen in its resulting table (Table 5). The reason for creating these data sets was to be able to produce graphs which visualize annual circulation and visit totals by ward instead of branches.

Table 4: Sample of Merged Annual Circulation Data by Ward

Ward Number	Year	Total Circulation
1	2018	517444
1	2019	483554
1	2020	202214
1	2021	243070
1	2022	321187
2	2018	1130647
2	2019	1118318
2	2020	546293

Table 5: Sample of Merged Annual Vists Data by Ward

Ward Number	Year	Total Visits
1	2018	517444
1	2019	483554
1	2020	202214
1	2021	243070
1	2022	321187
2	2018	1130647
2	2019	1118318
2	2020	546293

2.5 City-Wide Annual Circulation and Visits

Based on the two previous data sets which aim to display total annual circulation and total by ward, they were also used to create city-wide totals for each year. The data sets were grouped by year and ward totals of circulation and visits were summed. This resulted in one data set which provides the total annual circulation of the Toronto Public Library's physical and fixed location branches (Table 6) in addition to another data set which provides the total annual visits of the Toronto Public Library's physical and fixed location branches (Table 7).

Table 6: Total Annual Circulation of Toronto Public Library Materials (Physical and Fixed Location Branches Only) Data Set

Year	Total Circulation
2018	22753255
2019	21900991
2020	10227385
2021	12938365
2022	15520469

Table 7: Total Annual Visits of Toronto Public Library Branches (Physical and Fixed Location Branches Only) Data Set

Year	Total Visits
2018	17514253
2019	17193858
2020	5540037
2021	4017202

Table 7: Total Annual Visits of Toronto Public Library Branches (Physical and Fixed Location Branches Only) Data Set

Year	Total Visits
2022	9527633

3 Results

Bar graphs were created with `ggplot2`(Wickham et al. 2019) to visualize the cleaned data sets. A bar graph for each year between 2018 and 2022 showing the total circulation by each ward was created. The y-axis scale for all of these graphs started at 0 and ended at 1,600,000 with increments increasing by 200,000 to capture all data values. Additionally, a bar graph for each year between 2018 and 2022 showing the total visits to libraries by each ward was created. The y-axis scale for all of these graphs started at 0 and ended at 2,400,000 with increments increasing by 200,000 to capture all data values.

3.1 Comparing Annual Circulation by Ward Between 2018 to 2022

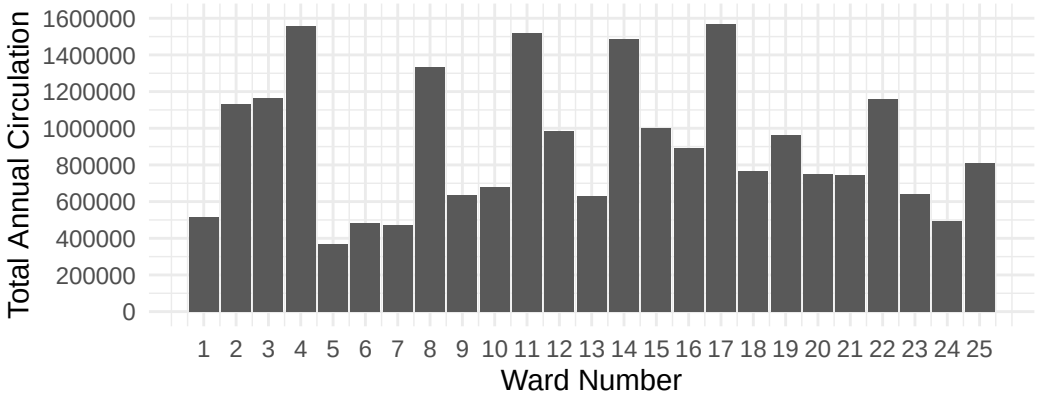

Figure 1: Total 2018 Annual Circulation of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

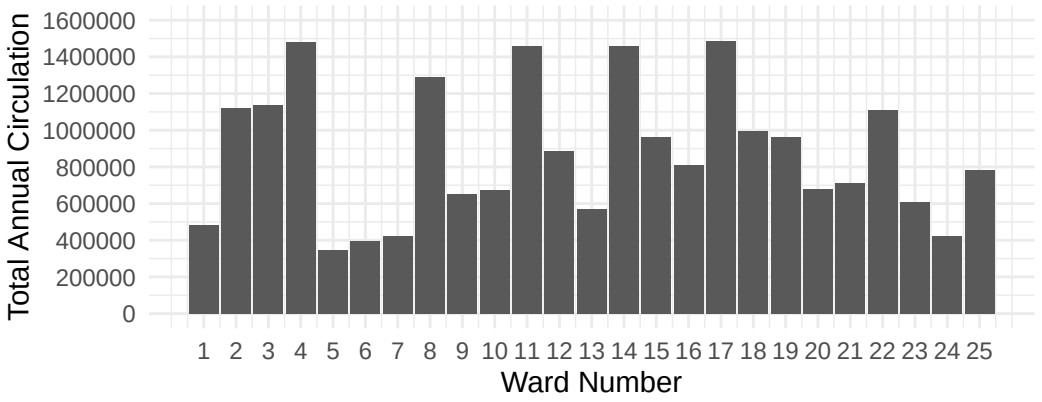

Figure 2: Total 2019 Annual Circulation of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

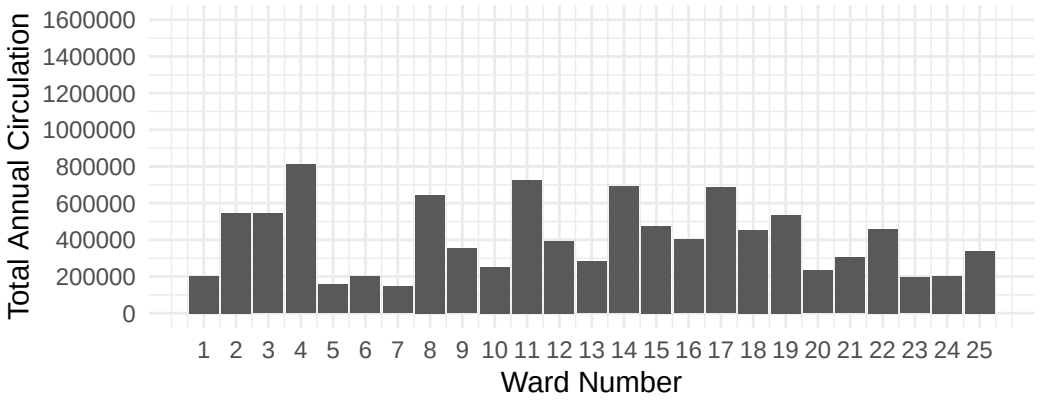

Figure 3: Total 2020 Annual Circulation of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

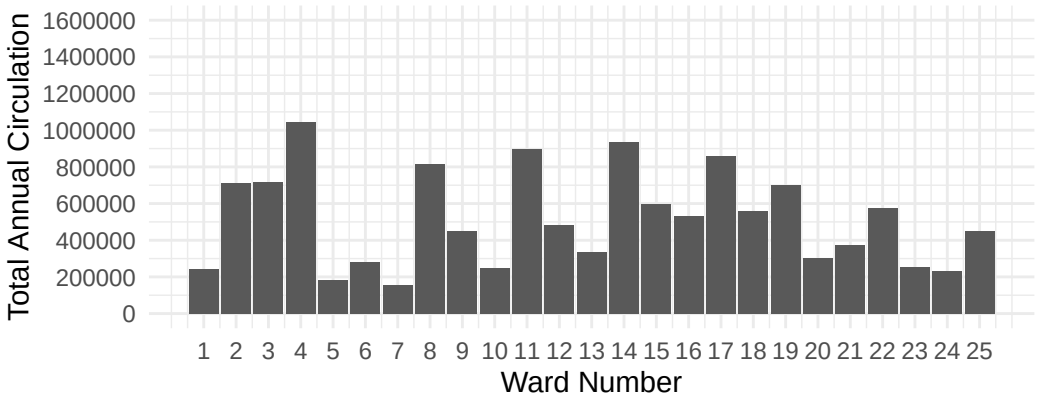

Figure 4: Total 2021 Annual Circulation of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

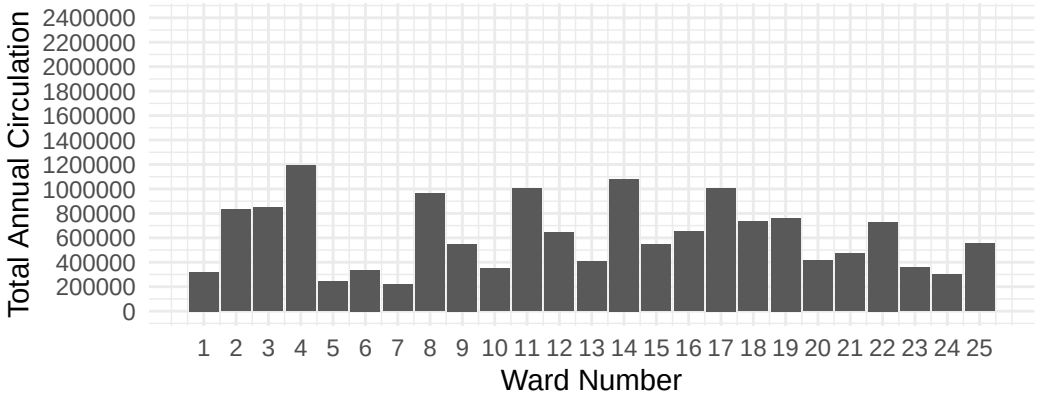

Figure 5: Total 2022 Annual Circulation of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

As seen when broadly viewing Figure 1, Figure 2, Figure 3, Figure 4, and Figure 5 as a set, there appears to be a sharp decrease in library circulation starting in 2020. In 2018 and 2019, the mean annual circulation across wards was 910,130.2 and 876,039.6 respectively. This is comparatively higher than 2020 which has a mean of 409,095.4. This mean was higher in 2021 at 517,534.6 transactions. There was also an increase in 2022 at 620,818.8.

Between 2018 and 2020, it appears that the total number of circulation transactions was trending towards an annual decrease which is reflected in Figure 1, Figure 2, and Figure 3 when viewing the annual graphs as a set. Looking at the individual ward totals of circulation in these graphs reveals an overall decrease for each ward between these years. Similarly,

between 2020 and 2022, there appears to be a positive trend in total circulation by ward when viewing the bar changes for each ward through Figure 3, Figure 4, and Figure 5.

In 2018 and 2019, Ward 17 - Don Valley North, had the highest total circulation transactions. Libraries found in Don Valley North are Ethennonnhawahstihnen', Fairview, Hillcrest, and Pleasant View. For the same years, Ward 5 - York South-Weston had the lowest total circulation transactions. Libraries found in York South-Weston are Amesbury Park, Evelyn Gregory, Mount Dennis, and Weston. These ward totals are viewable in Figure 1 and Figure 2.

Between 2020 and 2023, Ward 4 - Parkdale-High Park had the highest total circulation transactions. Libraries found in Ward 4 include Annette Street, High Park, Jane/Dundas, Parkdale, Runnymede, and Swansea Memorial. Ward 7 - Humber River-Black Creek had the lowest total circulation transactions during this period. Libraries found in Ward 7 include Black Creek, Humber Summit, Jane/Sheppard, Woodville Park, and York Woods. These ward totals are viewable in Figure 3, Figure 4, and Figure 5.

Comparatively, Ward 17 in 2018 had a circulation total of 1,568,603 to Ward 4 in 2020 and 2022's totals of 811,568 and 1,190,959. This indicates that 2020 did not see many circulation transactions across all wards by looking at the highest values. By looking at the highest values of these years, there is evidence the annual circulation total by ward hit its trough in 2020 after a year-over-year decline starting in 2018 and began to recover from then onward.

3.2 Comparing Annual Visits by Ward Between 2018 to 2022

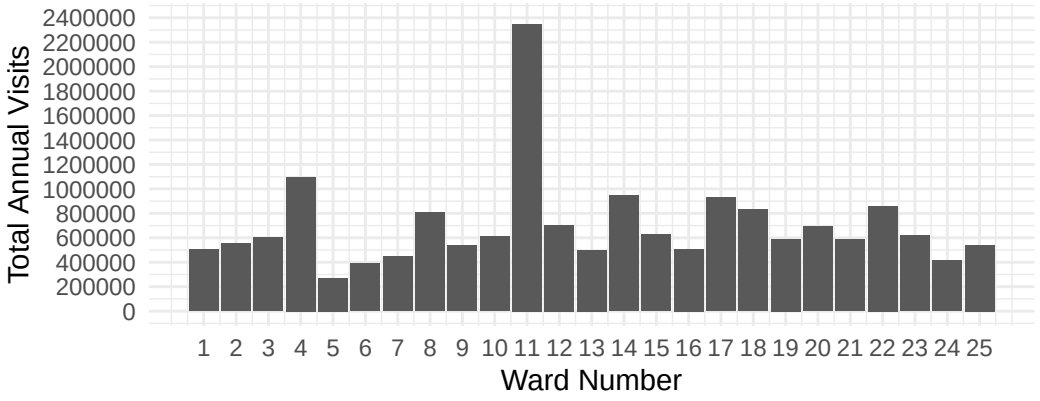

Figure 6: Total 2018 Annual Visits of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

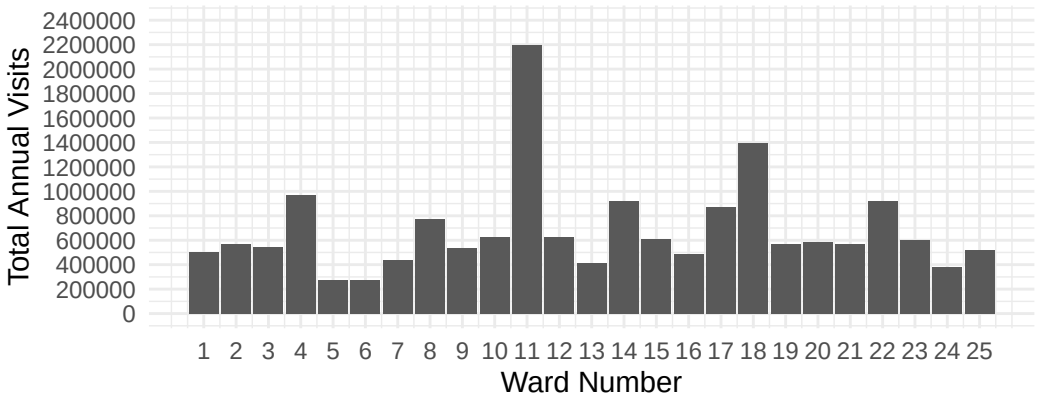

Figure 7: Total 2019 Annual Visits of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

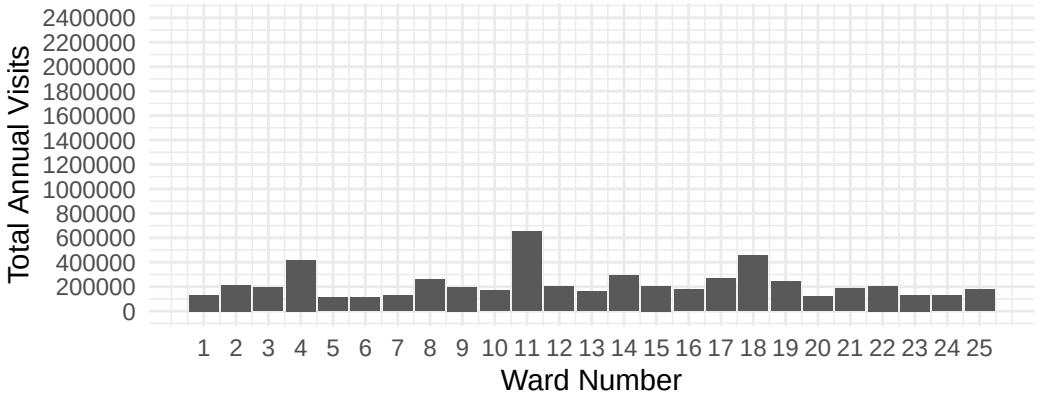

Figure 8: Total 2020 Annual Visits of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

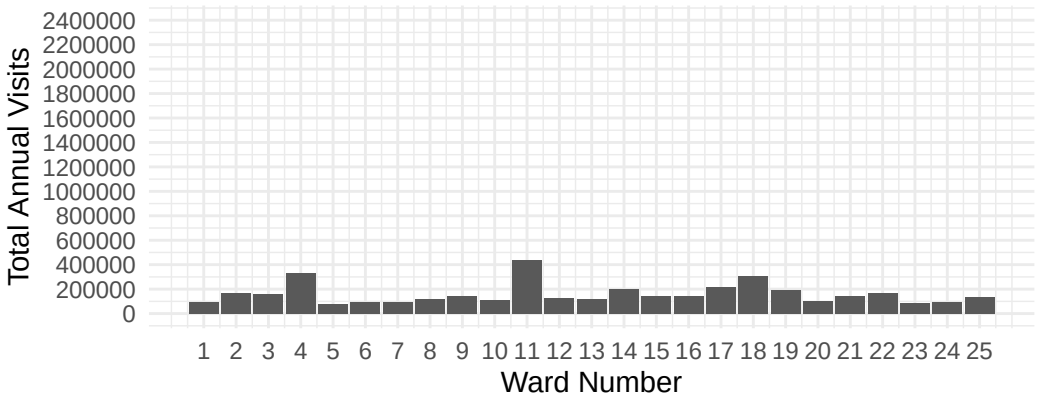

Figure 9: Total 2021 Annual Visits of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

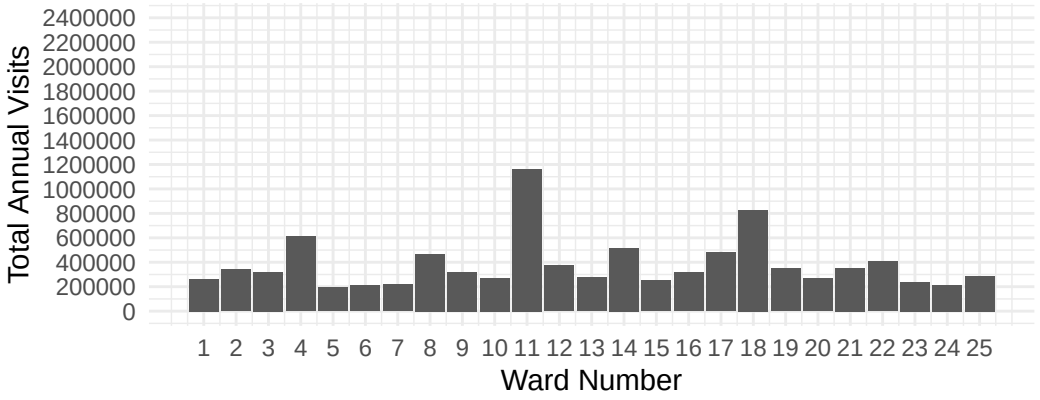

Figure 10: Total 2022 Annual Visits of Toronto Public Library Branches (Physical and Fixed Location Branches Only) by Ward

As seen in a broad view of Figure 6, Figure 7, Figure 8, Figure 9, and Figure 10 as a set, there appears to be a sharp decrease in library circulation starting in 2020. In 2018 and 2019, the mean circulation across wards was 700570.1 and 687754.3 respectively. This is comparatively higher than 2020 which has a mean of 221601.5. This mean was even lower in 2021 at 160688.1. However, the mean circulation by ward increased in 2022 at 381105.3 after two consecutive years of decrease. Although 2022's mean circulation by ward is an improvement to the two prior years, it is roughly only 50% of the means seen in 2018 and 2019.

Between 2018 and 2020, there appears to be a sharp decrease in total library visits between 2019 and 2020. This pattern is reflected in Figure 6, Figure 7, and Figure 8 when viewing the bar graphs in a set. 2018 and 2019 have visually similar graphs with six wards having

a visit total above 700,000. However, this is contrasted by Figure 8 where no ward has a visit total above 700,000. Interestingly, it appears in Figure 9 that visits to libraries across all wards were even lower than in 2020. In 2022, Figure 10 highlights total visits across all wards increasing which could indicate a future possible trend upwards after four consecutive years of decrease.

Between 2018 and 2022, Ward 11 - University-Rosedale had the highest annual total of visits. Libraries in this ward include College/Shaw, Lillian H. Smith, Palmerston, Spadina Road, Toronto Reference Library, and Yorkville. In 2020, Ward 6 - York Centre had the lowest total visits. Ward 6's libraries are Centennial and Downsview. For 2018-2019 and 2021-2022, Ward 5 - York South-Weston had the lowest total visits. Libraries found in York South-Weston are Amesbury Park, Evelyn Gregory, Mount Dennis, and Weston.

Ward 11 had visit totals of 2,349,923 in 2018, 435,484 in 2021, and 1,163,179 in 2022. By viewing the highest ward total for each year, it can be concluded that visitation decreased at a peak in 2021 across all wards. By looking at the highest values of 2018, 2020, and 2020, there is evidence the annual visit total by ward hit its trough in 2021.

3.3 City-Wide Annual Circulation and Visits

Tables that provide a summary of the city-wide total visits (Table 7) and total circulation (Table 6) for each year. Annual circulation and visitation metrics have high standard deviations of 5,502,796 and 6,349,144, respectively. The high standard deviations imply that the overall annual variance of these metrics is high, therefore indicating values are spread out from the mean. This is seen in the graphs for both usage metrics as 2020 and 2021 reveal significantly smaller values compared to 2018 and 2019. By viewing the annual data, it is clear that the graphs represent the annual data after comparing the table to the aforementioned sections.

4 Discussion

Looking over the five years of data in this report, both annual circulation and visits by ward appear to be going through a trough-shaped pattern. For both circulation and visits, 2018 and 2019 data hovered around roughly the same numbers with 2019 having slightly lower numbers. Annual circulation in this data appears to have bottomed out in 2020 where the lowest transactions were recorded across all wards. Afterwards, in 2021 and 2022, circulation totals across all wards appear to be slowly increasing; however, they have not returned to pre-2020 levels. 2021 accounted for the lowest visit total across the wards after a sharp decline in visits in 2020 whilst 2022 highlighted an increase in visit totals across all wards to levels just under pre-2020.

Although this report does not specifically focus on proving causation between the COVID-19 pandemic and library usage metrics, it is clear that the annual totals of these metrics do have

differences pre and post-pandemic. The COVID-19 pandemic's lockdown and social distancing measures could be a contributing factor to the sharp decline in total circulation and visits to libraries across all wards in Toronto seen in 2020 and 2021. As indicated by the Government of Ontario, public libraries were subject to lockdown measures which could have affected patrons' ability to visit and check out material. It could also be possible that patrons' behaviour has changed post-2020 regarding public spaces and shared materials.

There are potential measurement errors in the data sets used in this report. As previously noted, the visit totals were collected by sensors at the entrances of library branches during scheduled operating hours. This means that it is possible individuals were counted multiple times in a day due to exiting and reentering. It is also important to note that the visit data set is incomplete. For 100 physical branches over five years of collections, there should be 500 rows; however, only 492 are present. 2018 data for St. Clair/Silverthorn, 2021 data for the City Hall Branch, 2021 data for Swansea Memorial, 2021-2022 data for Mount Pleasant, and 2019-2021 data for Wychwood were missing from the data set. For the purposes of analysis, no null or 0 value was created as a placeholder for the missing data to not skew the results. This is definitely a limitation to the data since ward totals that did not have this data present could be skewed lower due to the absence of data if it was omitted from the data set while the branches were still operating. Although both data sets only record annual data, the frequency of the data collection and entry methods are unknown which could be another possible limitation. As well, since the data sets only provided data up to 2022, the long-term total circulation and visit trends can not be analyzed.

Limitations also exist in the analysis of the data sets in this report. While this report does analyze and graph the annual ward totals of library circulation and visits, expanding the chronological scope of the report could provide greater insight into whether or not post-2020 data was directly affected by the pandemic and not other factors. The report organizes data by ward and therefore it might have been beneficial to compare total circulation and visits to demographic ward data to rule out other factors aside from the pandemic as causes of the severe decrease seen in 2020. Annual data is also used for the report and it is worth investigating library usage metrics against periods of lockdown or mandated social distancing on a daily or monthly basis across the same five-year period. However, monthly and daily data is not available on Open Data Toronto (Gelfand 2022). This could further support the assertion that there were changes in total circulation and visits that can be categorized as pre and post-pandemic.

5 Conclusion

This paper investigates if there was a difference in annual circulation and visit totals at the Toronto Public Library by analyzing the yearly data of these metrics by ward in order to assess what the pattern of these totals between 2018 and 2022 was in addition to whether or not there was a difference in pre and post-pandemic totals. The data and graphical analysis

show that across library circulation and visits, there was a trough pattern over the five years for both metrics. It can also be ascertained that there is a difference in pre-pandemic levels to post-pandemic as a result of the sharp decline in 2020 in addition to the positive trend of totals in 2021 and 2022 that have not yet reached the ward highs seen between 2018 and 2019. While there is a notable difference, it cannot be ascertained that these year-over-year changes were entirely due to the pandemic and related measures. Future analysis with further post-2020 data in addition to daily or monthly data to compare totals to periods of lockdown will be helpful in further investigating public library use and activity in a post-COVID-19 context.

References

Bobbitt, Zach. 2021. "R: Remove Rows from Data Frame Based on Condition." https://www.statology.org/r-remove-rows-from-data-frame-conditio /.

———. 2022. "How to Merge Data Frames by Row Names in r." https://www.statology.org/r-merge-by-rownames/.

Chow, Anthony, and Qianfei Tian. 2021. "Public Libraries Positively Impact Quality of Life: A Big Data Study." *Public Library Quarterly* 40 (1): 1–32. https://doi.org/10.1080/01616846.2019.1632105.

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.

Jamali, Hamid R., and Philip Hider. 2023. "Changes in the Usage Pattern of Public Library Collections During the COVID Pandemic." *Collection Management* 48 (4): 324–38. https://doi.org/10.1080/01462679.2023.2210070.

Organization, World Health. 2023. "Coronavirus Disease (COVID-19) Pandemic." https://www.who.int/europe/emergencies/situations/covid-19.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Public Library. 2023a. "Library Branch General Information." https://open.toronto.ca/dataset/library-branch-general-information /.

———. 2023b. "Library Circulation." https://open.toronto.ca/dataset/library-circulation/.

———. 2023c. "Library Visits." https://open.toronto.ca/dataset/library-visits/.

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.