Trends in Toronto Public Libraries: A Study of 2012-2022 Toronto Public Library Data*

My subtitle if needed

Liam Wall

January 25, 2024

This paper looks into four annual statistics released by the Toronto Public Library (TPL) system from 2012-2022 - visits, new registrations, circulation of materials, and computer sessions usage. It shows an almost predictable down-trend in library from 2012 to 2019 and a major hit to the TPL system during the COVID-19 pandemic. Follwowing the 2020 hit, the TPL began to bounceback but all of the statistics (bar perhaps one) do not recover to pre-pandemic levels. The only positive correlation in the entire paper is between new registrations and computer usage from 2012-2015.

Table of contents

---

*Code and data are available at: https://github.com/Lwall02/What-Are-We-Using-Libraries-For.

1 Introduction

With the ever-increasing accessibility of publications, like books, magazines, and newspapers, to name a few, one can wonder what libraries are really used for nowadays. It is well-known that a library is a place to come to read and learn. There is an immensity of resources available at most public libraries across all of the world, including public records, films, local newspapers, computers, socializing areas, and of course publications. Despite the common thought that not many people go to libraries, it is still a strong belief that the absence of libraries would be a detriment to the surrounding community (PEW RESEARCH CITATION).

To a lot of people, the ability to access the internet is nothing more than sitting at their personal home computer or using their phone connected to their wifi or cellular service. However, one does not always remember there is a significant portion of the community that is not always fortunate enough to have this level of access. In today's world, this can be one the libraries single most important amenities; access to the internet. The Toronto Public Library (TPL) system offers free wifi at every branch and almost all of their 103 branches offer free computer stations (REFERNCE TO TPL WEBSITE). In this paper, I look at TPL information from 2012 to 2022 covering the number of annual visits, new registrations, circulation of materials (called circulations), and computer station usage (called a computer sessions). I wanted to find out why people are using the TPL, when it was most popular, what is increasing or decreasing, and how popular are libraries today compared to a decade ago.

My findings show that there is a definitive decrease in every statistic from 2012 to 2022 except for new registrations. They show a decrease in yearly visitors, circulations, and computer sessions, however the number of new registrations per year has countered these trends, in fact the number of new registrations in 2022 is more than that of 2012. We can also see the large impact the COVID-19 pandemic had on the TPL with decreases in every statistic during 2020. Post-pandemic, the numbers in each of the four categories began to increase however, only new registrations were able to reach pre-pandemic levels by 2022.

The Data section of this paper provides an explanation on the TPL data used in this paper and the steps that went into making this data usable and readable for a report. The Results sections discusses in greater detail the findings from the data by showing the trends I explained. The Discussion sections provides some information on the possible shortcomings of the analysis and other events that may have an effect on the findings. The Conclusion sections summarizes the findings from this paper.

2 Data

Data used in this paper was downloaded from the Open Data Toronto website (INSERT CITI-ATTION FOR ODT). I downloaded a total of five data sets published by the Toronto Public Library: Library Branch General Information (INSERT CITATION TO WEB PAGE), Library Visitors (INSERT CITATION TO WEB PAGE), Library Workstation Usage (INSERT CITATION TO WEB PAGE), Library Card Registrations (INSERT CITATION TO WEB PAGE), and Library Circulation (INSERT CITATION TO WEB PAGE). The data was worked with entirely on the open source programming language R (R Core Team 2022). I used the following packages to clean and analyze the data: tidyverse (INSERT CITATION TO PACKAGE), ggplot2 (INSERT CITATION TO PACKAGE), patchwork (INSERT CITATION TO PACKAGE), knitr (INSERT CITATION TO PACKAGE), and kable (INSERT CITATION TO PACKAGE), dplyr (INSERT CITATION TO PACKAGE).

2.1 The Data Sets

I downloaded the five above mentioned data sets and in order to conduct some analysis I had to merge four of them into one tibble. The cleaning and merging of the data sets is explained below.

It is important to note that there were 100 to 104 different TPL branches across the Greater Toronto Area each year from 2012 to 2022. Each branch is attached with a branch code that is two capital letters. For example, the Lillian H. Smith Library located just below the University of Toronto is labeled LS. You will see these branch codes in tables and graphs later in the paper.

2.1.1 The Library Branch General Information

The Library Branch General Information data set was only used in two ways for this paper. It provided a key to attach a libraries branch code with its full name and address, and it separated the physical branches from the 'non-physical' branches. The physical branches are the libraries you can visit, the 'non-physical' libraries are services like a bookmobile or a virtual library. The 'non-physical' libraries are shown in Table 1. (SHOW AND REFERNCE TABLE OF THE NONPHYSICAL LIBRARIES).

Table 1: This shows the 12 non-physical branches as of 2023 along with their TPL given name

Branch Code	Branch Name	Physical Branch (1 - Yes/ 0 - No)
AL	Answerline	0
BKONE	Bookmobile One	0
BKTWO	Bookmobile Two	0

Branch Code	Branch Name	Physical Branch (1 - Yes/ 0 - No)
DS	Departmental Staff	0
HLS	Home Library Service	0
IL	Interloan	0
LD	Literacy Deposits	0
ME	Merril Collection	0
OS	Osborne Collection	0
PR	Automated Phone System	0
SB	Sunnybrook Hospital	0
VIR	Virtual Library	0

It is important to separate and remove the 'non-physical' libraries from all the data sets because they introduce very different values in each of the categories I am looking at. The 'non-physical' libraries recorded their visits, registrations, circulations, and computer sessions in different ways. For example, the visits would have been clicks to a website or number of visitors into a mobile library. In this paper, I am focused on the popularity of the physical library and so this data does not serve to tell me how a physical library is performing.

2.1.2 The Library Card Registration, Circulation, Visits, and Workstation Usage

These four data sets, The Library Card Registration, Circulation, Visits, and Workstation Usage, serve as the essential data for this paper. As they are also released by the TPL system, they follow the exact same branch code organization. Each of the four data sets gives data for each branch code for every year from 2012 to 2022. They are very similar data sets to each other in that there is the year, branch code, and number of the value that is to be shown (visits, registrations, circulations, or sessions).

2.2 Cleaning the Data

The first step of cleaning the data so that I could analyse it in an efficient way was to combine the visits, registrations, circulations, and sessions, into one table according the year and branch code. In that way all the information stays according to the correct library it was taken from and the correct year. The second step was to use the clean_name() function. The third step was to remove the 'non-physical' branches so that the data I analyse is from real, physical libraries only. A sample of final table with each of the four statistics aligned with branch code and year is found in Table 2. (SHOW A TABLE OF THE COMPLETE INFO I USED). An explanation of the four variables is found in Results.

Table 2: This shows 10 rows of the final table used for analysis. It contains every branches information for each year from 2012-2022 organized such that the most visited library per year is at the top

Year	Branch Code	Visits	Registrations	Circulations	Sessions
2012	CL	1493796	10291	1703346	484789
2012	TRL	1137016	8682	355367	727822
2012	AB	522309	4939	515235	132024
2012	WS	488396	2013	542341	105207
2012	FV	442214	4064	890759	164649
2012	AG	434320	3166	1061437	140407
2012	RI	421962	3082	791674	131921
2012	CED	415753	4750	731057	240834
2012	ND	407488	2755	489994	127600
2012	BRW	392397	2015	518725	41350

3 Results

First, before I get into the results, I will explain what each variable is that you will see in this section. The 'Visitors' describes the number of visitors per year at a certain library branch. The 'Registrations' describes the number of library cards registered at a certain library branch per year. The 'Circulation' describes the number of materials that were checked out and returned during the year at a certain library branch. The 'Sessions' describes the number of computer sessions at a library's public workstation per year. A computer session lasts a duration of at least 15 seconds.

Second, the number of library branches did not stay the same every year. There were anywhere from 100 to 104 library branches, as mentioned in the introduction, however there was only ever 96 to 100 active physical branches. This can be seen in Figure 1.
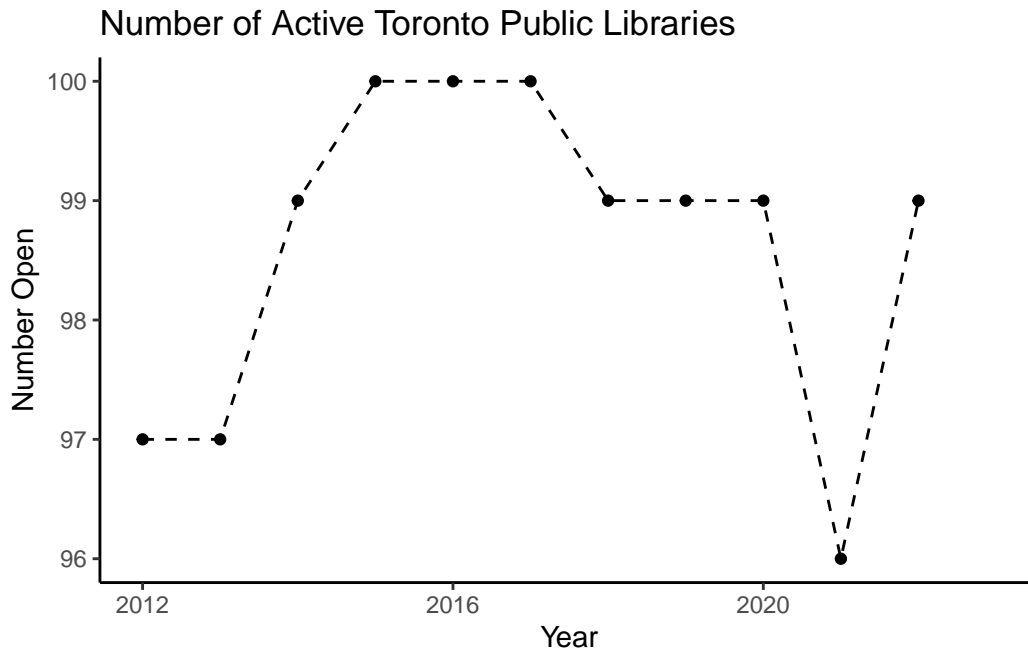
Number of Active Toronto Public Libraries

Figure 1: This shows the variation in the number of library branches each year

The results of this analysis followed from looking at the trends of the data over the years. The first step was to try and see how the libraries performed each year based on the four variables listed above. The four scatter plots in Figure 2 show the count of each variable from 2012 - 2022. In this format it is hard to see whether visits are increasing or decreasing, especially from 2012 - 2019, and the same goes for the other three variables. The biggest takeaway from this view is that the COVID-19 pandemic did effect the libraries in 2020. It is hard to say exactly how much the libraries were effected but we can see that the visits, registrations, circulations, and sessions, have no very large values in 2020.
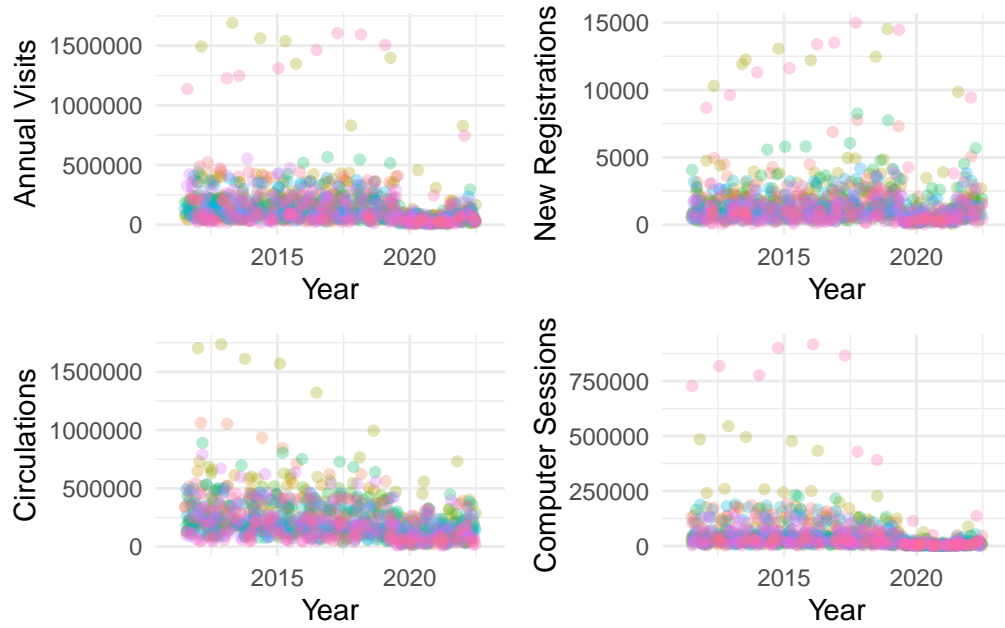
Figure 2: A visual of all the data used for this analysis, organized by year. Each color represents a value for a library branch

In Figure 2, it is hard to determine any kind of trend among the visits, registrations, circulations, and computer sessions. As stated above, it especially hard to do so from 2012 - 2019 because there is so many points crowding the same space. In order to look at the trends of this data, we look at the averages of each variable. This is visible in Figure 3.

As you can more clearly see in Figure 3, yearly visits has been steadily decreasing across all TPL. Further, yearly circulations has really been declining since 2012. When looking at the new registrations and computer usage there is more storied trend. Surprisingly, despite less visitors and less materials being checked out, computer sessions and new registrations were increasing from 2012 - 2015. From 2016 onwards, computer sessions have trended down quite significantly while new registrations has remained trending upwards.

Once COVID-19 struck in 2020, every variable took quite a hit. In Figure 3, we can see every variable has gone near the x-axis at 2020. We can also see that from 2020 - 2022, visitors, circulations, and especially new registrations have trended upwards. In fact, new registrations is averaging the same amount in 2022 as it was for nearly all of 2012 - 2019. Average computer sessions, however, has remained very low compared to its previous averages.
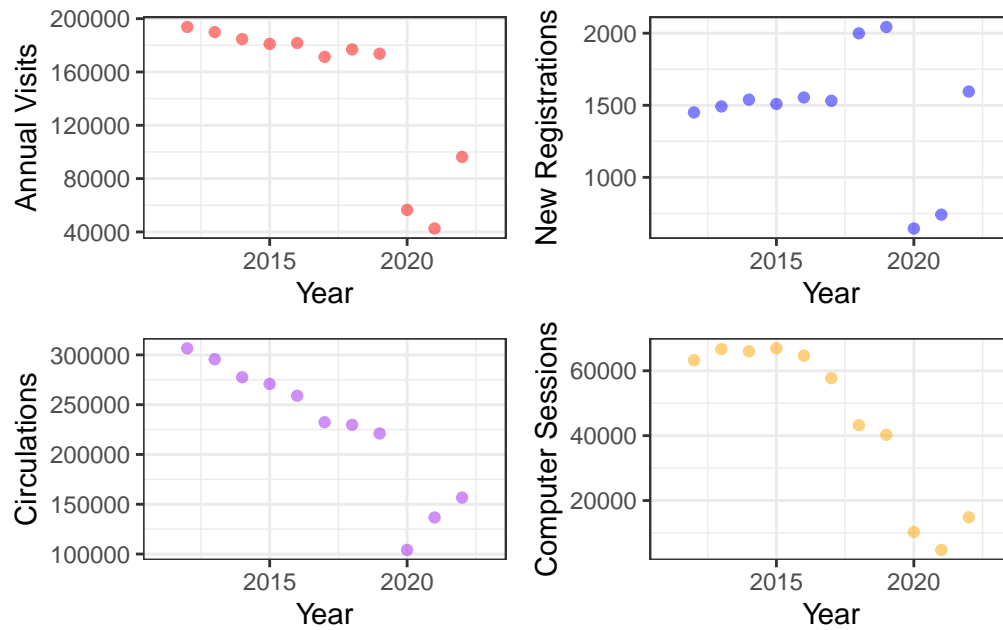
Figure 3: A visual of the average of each variable from 2012 - 2022. We can more clearly see the negative trends of each variable.

4 Discussion

5 Conclusion

6 Appendix

References

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