

# Toronto Shelter Flow Exploratory Analysis\*

Dingding Wang

02/06/2022

## Abstract

According to the 2021 report on international housing affordability by Urban Reform Institute, Toronto ranked 5th the most ‘severely unaffordable’ city globally. Toronto has increasing shortage of affordable housing, resulting in many individuals experiencing homeless and increasingly rely on temporary shelters. This report will discuss the growing issue of homeless trends in Toronto.

## 1 Introduction

In this report, I aim to assess the breadth of the homeless shelter problem and to develop a methodology for shelter measurement. The spike in homelessness has been a major social issue in Canada for decades. The City of Toronto has budgeted to spend 663.2 million dollars on homeless people and supports in 2021, which is about double the amount the government spent in 2019. The government is committed to decreasing the risk of outbreaks of Covid-19 in the shelter system and continues to provide enhanced supports to homelessness. However, the number of homeless people in Canada is not trending downward.

Under the affect of COVID-19, there is a growing recognition that homelessness is a more complex social and public health phenomenon than not having a place to live. While there is evidence of links between homelessness and social factors such as government benefits, drug addiction and unemployment, few studies have examined the broader social context in which homelessness occurs from the perspective of the homeless themselves. The purpose of this report is to examine the stories of homelessness in order to understand the social conditions under which homelessness occurs and to propose theoretical explanations, as well as shelter system flow of Toronto.

## 2 Data

### 2.1 Data Source

Toronto Shelter System Flow data is collected and published by Shelter Support & Housing Administration division of the government, and is updated on a monthly basis. Data collector last updated the dataset on Jan 7th, and data is refreshed on the 15th of each month or the next business day. Data was obtained from Shelter Management Information System (SMIS) database, which is an internal tool that improves efficiency for shelter staff to check for available resources for homeless people and record.

### 2.2 Methodology and Data Collection

Data for this project was collected from Open Data Toronto portal, which is the official source for Toronto open data from City divisions and agencies. Toronto Shelter System Flow dataset is published by Shelter, Support & Housing Administration, which contains information about homeless people in Toronto who are entering and leaving the shelter system each month from January 2020 to December 2021. The dataset is refereshed monthly and last updated on Jan 7, 2022. Implementing the shelter system involves changing in many policies, funding plans and data collection procedure, which all intersect with the application that

---

\*Code and data are available at: <https://github.com/estherxwang/STA304Paper1>

power the operation of divisions, the Shelter Management Information System (Dominguez,2020). The data of overnight services was collected through SMIS in order to record number of people who are accessing these services into shelter programs. SMIS eliminates time-consuming application processes to simplify and improve shelter services, thus shelter staff are able to quickly find available beds throughout the system and speed up the flow of patients in and out. This dataset aims to help reduce homelessness in Toronto and develop a better Shelter Management Information System.

While the dataset is refreshed on a monthly basis, there exist some limitations and biases. Regarding the SMIS system, although it significantly improves the efficiency of shelter services, the shelter sites without the system and funded by other levels of government are not contained in the dataset. Since the population the dataset is supposed to represent all homeless people in Toronto, using SMIS leads to unaccuracy and incomplete data. In addition, the dataset reflect only people who use night services and exclude people who sleep outdoors specifically or use other homelessness services. Another concern is that the release date of the dataset is on the 15th of each month, which indicates that the dataset is replaced monthly when new dataset is generated and the previous analysis research may not be applicable. The conclusions drawn from the report may not be generalized to a wider population and may relate only to the participants in this study(Mabhala et al,2017).

## 2.3 Important variables

The Toronto Shelter System dataset includes 180 observations with 18 variables. By cleaning the data using mutate and select in tidyverse (Wickham, 2019), 11 variables were selected, which includes: date, population group (either refugee, family, or others), newly identified homeless people, count of actively homeless people, age under 16, age 16-24, age 25-44, age 45-64, age 65 and over, gender male, gender female, and other gender.

There are 8 primary variables from the dataset that we will be analyzing.

**Date:** The time period (month/year) for which the data is being published (format: mmm-yy)

**population\_group:** Each row represents different group: all population, chronic, refugees, families, youth, single adult and non-refugees

**Newly Identified:** People who entered the shelter system for the first time

**Actively Homeless:** People who have used the shelter system at least one time in the past three months and did not move to permanent housing.

**moved\_to\_housing:** Moved to Permanent Housing: People who were using the shelter system and have moved to permanent housing.

**returned\_from\_housing:** Moved to Permanent Housing: People who were using the shelter system and have moved to permanent housing.

**returned\_to\_shelter:** Moved to Permanent Housing: People who were using the shelter system and have moved to permanent housing.

**no\_recent\_shelter\_use:** People who were using the shelter system previously, but have not used the shelter system in the past three months. Some other communities may call this indicator "Moved to Inactive".

Since our research focus is population group, understanding the official definition of population variables is essential to answering the research question:

**All populations:** The total number of individuals represented in the report without any demographic breakdown.

**Chronic:** People who meet one of the criteria as per the federal definition of chronic homelessness. Either the person has recorded a minimum of 180 overnight stay in the past year (365 days) or has recurrent overnight stays over the past three years with a cumulative duration of at least 546 nights.

**Families:** Individuals who are recorded as staying in a family designated overnight service.

**Youth:** Unaccompanied youth and includes people who are between 16 and 24 years old by the last date of the reporting month and are not members of a family as defined above.

**Single Adult:** Individuals who are neither classified as youth nor are members of a family.

**Refugees:** People who either identify as refugees upon intake to a shelter system or whose intake is completed into a program designated for refugees. Applies to all household members as answered by the household head.

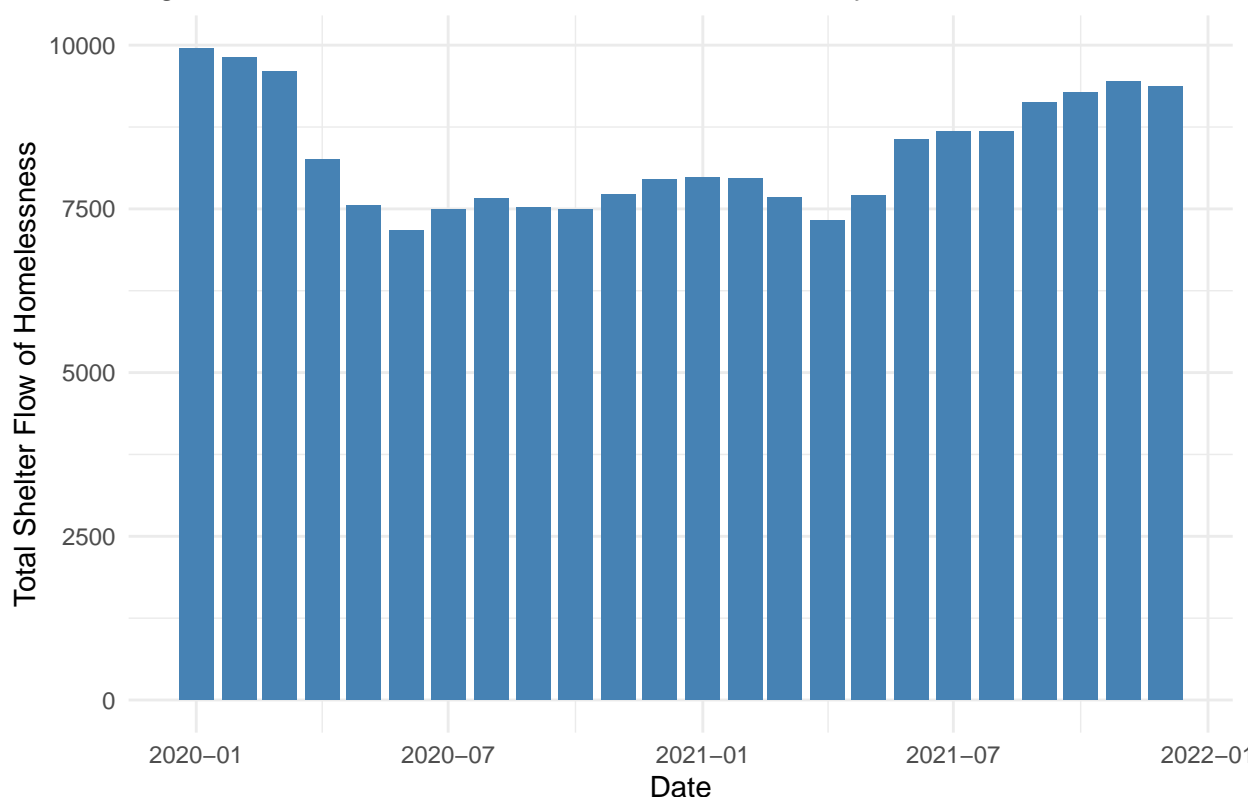
**Non-Refugees:** People who are not categorized as refugee per the definition above.

**Indigenous:** The person who has self-identified as First Nations (status, non-status, treaty, non-treaty), Metis or Inuit in at least one of the intake(s) completed in SMIS (Shelter Management Information System).

## 2.4 Data Cleaning

Using the `tidyverse` (Wickham, 2019) and `dplyr` (Wickham, 2018) package, I cleaned the data and extracted the important variables to start my exploratory analysis procedure. Since my goal of this data analysis report is to examine the relationship between the flow of shelter in Toronto and potential factors, I first filter the population group with `filter()` function to find cases of different population groups and combine them. Then, I used the `rename()` function to efficiently correct the variables that are not well formatted for better analysis. Our dataset focuses the data collected from the past two years, so the date variable is also important. I converted the year-month date variable to formal date format for better visualizations. Lastly, important variables were selected for further analysis.

Figure 1: Total Shelter Flow of Homelessness by Month



From the bar plot generated for shelter flow of homelessness by month, we can observe the trend of flow of shelter varies by season. There is a noticeable decrease in the number of total shelter flow from March to April in 2020. Despite the potential reason of lockdown policy, weather may play an important role in it. In winter, the shelter is occupied the most, particularly from January to March; on the contrary, the flow is much lower in summer. As Canada's largest city, the climate of Toronto is continental with extremely cold

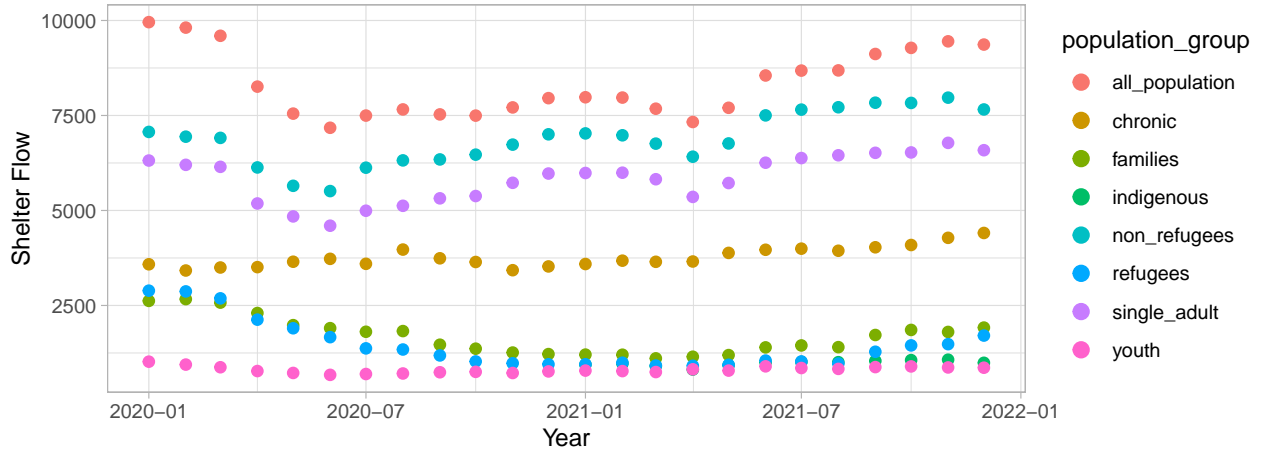
winters and warm summers. The winter runs from late November with the average temperature about  $-5.5^{\circ}\text{C}$  ( $22^{\circ}\text{F}$ ), while summers are warm with highs around  $26/27^{\circ}\text{C}$  ( $79/81^{\circ}\text{F}$ ). The temperature difference is a convincing explanation of the change in shelter flows corresponding to months. We can also notice that the trend does not seem to be affected by Covid-19 throughout the months.

sum_mean	sum_sd
3901.017	2820.985

According to the table, the mean and standard deviation of sum of shelter flow are relatively approximately 3901 and 2821.

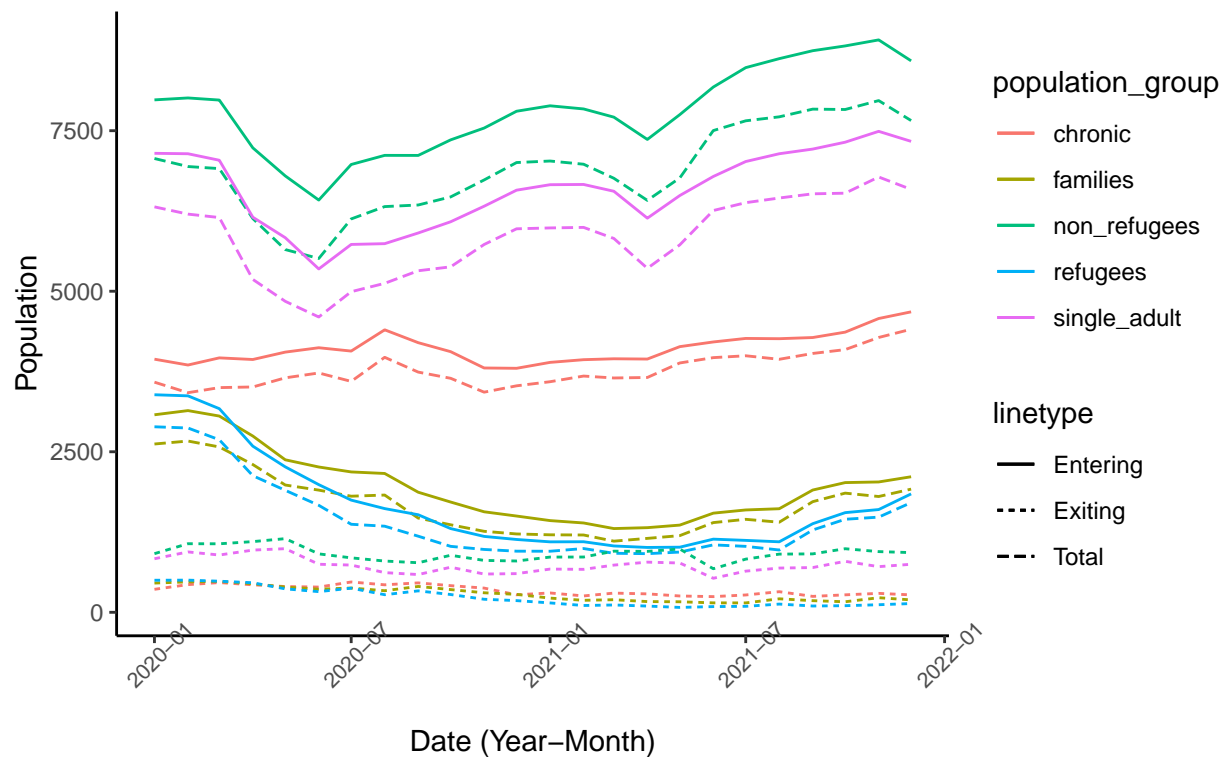
date	population_group	sum_positive	sum_negative	sum_tot
2020-01-01	all_population	11368	-1412	9956
2020-01-01	chronic	3942	-358	3584
2020-01-01	refugees	3388	-499	2889
2020-01-01	families	3076	-455	2621
2020-01-01	youth	1144	-123	1021
2020-01-01	single_adult	7148	-834	6314

**Figure 2: Shelter Flow of Population Groups by Month**



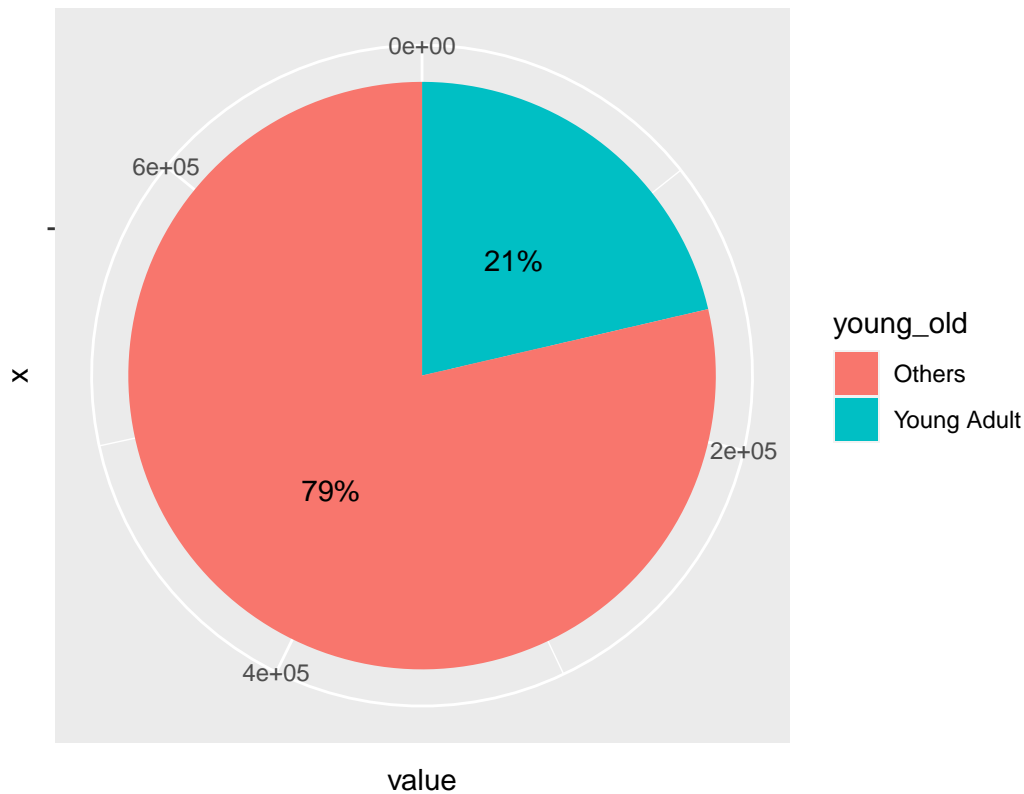
According to the plot of shelter flow of population groups by month, we can observe that the three variables among population groups that fluctuate the most are non refugees, single adults and chronic. It is also worth noticing that the shape of trend between non\_refugees and single adults are almost identical, which is a significant indicator that there might be a strong correlation between these two variables. To dive into the potential relationships among variables and the shelter flow of both entering and exiting, I generated a complete diagram below.

Figure 3: Toronto Shelter System Flow by Population Group  
Flow of People Entering and Exiting the Shelter in Each Month



According to the above plot indicating flow of people entering and exiting the shelter by month, we can still easily notice that the correlation of both entering and exiting between non\_refugees and single adult are very high. A potential reason may be that young adults who drop out of school at an early age may be more likely to be unemployed or addicted to drugs. The unemployment rates of young men and young women aged 15 to 30 and not in school are respectively 15.5% and 13.7% by the year of 2020 (Government of Canada, 2021). To support our hypothesis, I am conducting a further analysis of shelter flow according to age group.

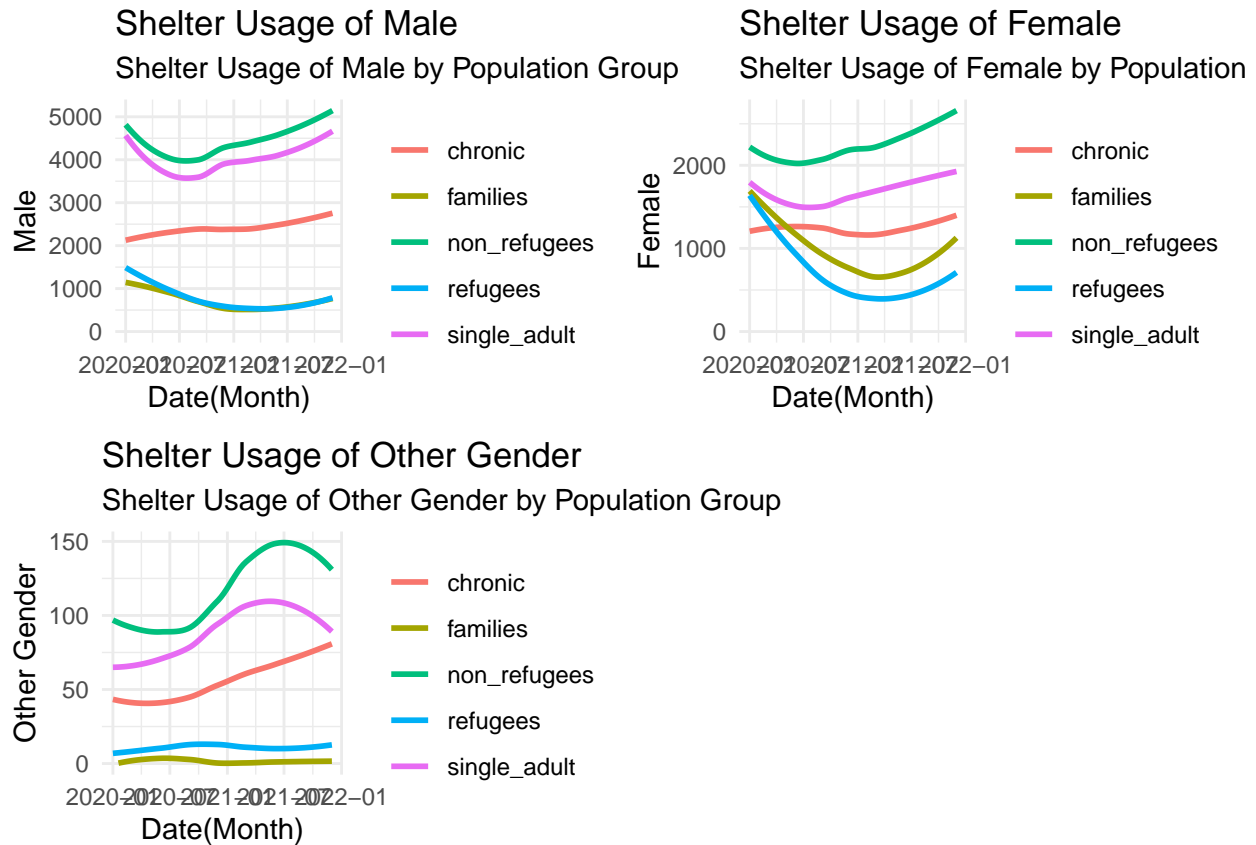
Figure 4: Comparison between Young Adults and Others



According to the pie chart we generated for comparison between young adults and other age groups, there is no significant evidence that young adults dominate the homeless shelter flow. Now, let's move to gender analysis.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 2 rows containing missing values (geom_smooth).
```



### 3 Conclusion

There has always been a debate that the number of years foreign-born adults lived in Canada was significantly associated with risk for homelessness. Experts argue that ‘There’s just not enough affordable housing options’, and the recent released government figures show that a growing number of newcomers to Canada either end up living in shelters or become homeless. As the president of the Canadian Alliance to End Homelessness, Tim Richter believes that refugees are being forced to settle in homeless shelters in response to shortage of housing capacity(Wright, 2019). Based on the analysis above, we can conclude that the usage of shelter and shelter flow is significantly impacted by weather and the growing number of non\_refugees in Toronto. Extreme weathers makes it extremely difficult for homeless people to stay out overnight and rises the shelter flow, and the raising number of non-refugees also result in significant increase in shelter usage.

## Reference

- City of Toronto. (2021, November 15). Shelter System Flow Data. City of Toronto. Retrieved February 3, 2022, from <https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/shelter-system-flow-data/>
- Jesse Donaldson, Tim Richter, & Tanya Gulliver (2013): The State of Homelessness in Canada 2013. Toronto: Canadian Homelessness Research Network Press. Retrieved February 3, 2022, from <https://www.homeless-hub.ca/sites/default/files/SOHC2103.pdf>
- Mabhala, M., Yohannes, A., & Griffith, M. (2017, August 22). Social Conditions of becoming homelessness: Qualitative analysis of life stories of Homeless Peoples - International Journal for Equity in health. Retrieved February 08, 2022, from <https://doi.org/10.1186/s12939-017-0646-3>
- Dominguez, S. (2021, January 15). Lessons from the frontlines of service delivery. Retrieved February 08, 2022, from <https://codefor.ca/blog/lessons-from-the-frontlines-of-service-delivery/>
- Climate - toronto (Canada). (n.d.). Retrieved February 08, 2022, from <https://www.climatestotravel.com/climate/canada/toronto>
- Wright, T. (2019, August 09). Growing number of newcomers, refugees ending up homeless in Canada: Studies | CBC News. Retrieved February 08, 2022, from <https://www.cbc.ca/news/politics/homeless-newcomers-refugees-canada-studies-1.5242426>
- Government of Canada, S. (2021, July 26). Study: Youth employment in Canada. Retrieved February 08, 2022, from <https://www150.statcan.gc.ca/n1/daily-quotidien/210726/dq210726b-eng.htm>
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2021. Tidy: Tidy Messy Data. <https://CRAN.R-project.org/package=tidy>.
- 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, and Kirill Müller. 2021. Dplyr: A Grammar of Data Manipulation. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org> ———. 2021. *Forcats: Tools for Working with Categorical Variables (Factors)*. <https://CRAN.R-project.org/package=forcats>.