Vizualizing Toronto Homess Shelter data Across a Global Pandemic*

A Look Into How The Pandemic disrupted the Flow Of Homeless Shelters And How The Situation Has Improved Since

Adam Labas

06 February 2022

Abstract

In addition to a global pandemic, homlessness and housing is a self apparent issue in the city of Toronto. Taking a stroll through certain areas in the downtown core, it's almost impossible to miss homeless tent cities and homeless people lined outside of shelters awaiting food or entry in the afternoon. Upon further examination of data collected from the Open Data Portal of the City of Toronto, it was easy to find interesting trends in homeless shelter data in the years 2020 and 2021. We found trends in the diminishing use of shelters by various demographic of homeless people and age groups. This paper attempts to visualize these trends as well as touch on some problems that one might experience in trying to extract real meaning from the data.

Contents

*Code and data are available at: https://github.com/adam-labas/adam_labas_paper_1_opendatatoronto

1 Introduction

The year 2020 was no doubt a year of unspoken discomfort and uncharted territories for everyone. For myself, the year 2020 started of like all the others. Winter commutes to school and winter festivities like ice skating with friends. As the semester started to develop and I started to get into a routine with my new classes, we started hearing of the corona virus which had supposedly arrived from China and was rapidly making its way around the globe. Although this seemed as nothing but a farce and a scene from a movie, the reality and gravity of the situation rapidly started to be apparent. As school was brought to an abrupt stop after the World Health Organization had declared COVID-19 as a pandemic on March 11th 2020 (https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020), the magnitude on the situation started to manifest itself in everyone daily lives. However, many lives were affected much more than others.

While I was mostly confined to the four corners of my home for the vast part of 2020. Being indoor during the cold Canadian winter has never been more of a challenge. However, it does require a brief explanation to claim that some people under no confinement at all, would be in a much more grave situation. For long, homelessness has been a problem in the city of Toronto. Going for a stroll through certain areas in the downtown core, it's almost impossible to miss homeless tent cities and homeless people lined outside of shelters awaiting food or entry in the afternoon. The privilege that people who live in homes have is immense and the difficulties that homeless people face daily is a reality that no one ever wants to live.

In this analysis, I had a look at Toronto Homeless shelter data, which is described with great extent in the Data Section 2. In my analysis, I decided to focus on how the global pandemic has impacted the homeless community of Toronto by creating visuals on the progression of the shelter usage thought the 24 months from January 2020 to December 2021. To further break down the analysis, I decided to look into the distribution of the aforementioned data but segregating the data not only by the various age group bins provided in the data sent but also the different demographic population groups provided. In general, it was found that shelter usage dropped drastically, immediately after the start of the pandemic and has recently been slowly trending back to its original usage flow. In this paper, I will attempt to take you on a journey to explore the data that was used to produce the findings, I will visually present to you the findings and I we will speak about the implications of the findings and later on in the Limitations Section 4, we will discuss the limitations of our data and problems that this may have when wanting to extract true, ethical, unbiased information from the implications of the findings.

2 Data

In this Data Section 2, I will provide a look into the data acquisition and processing methodology as well as a deep dive into the contents of the data. First of, (Table **??**) give us a glimpse of the data.

2.1 Data Collection

All the data that is used in this paper was retrieved from the City of Toronto Open Data Portal and is titled "Toronto Shelter System Flow." Using the R package opendatatoronto (Gelfand 2020), we loaded the data in the R-script titled *Data Acquisition and Processing*. The data is uploaded and funded by the City of Toronto and is refreshed monthly, last being updated on January 7th 2022 as of February 6th 2022. As the webpage in the Open Data Toronto Portal explains, "the City of Torontofunds and operates services dedicated to people experiencing homelessness in Toronto." They go on to explain that the services which they provide like emergency shelters, hotels/motels, warming centers etc. all use the Shelter Management Information System (SMIS) to track and log people entering the municipally funded homeless shelter system. When a new person enters the SMIS for the first time, many data points are collected. These variables are presented in our data. They include 14 variables like the current month and year for which the count is being recorded, the number of people who are returning/moving to/from housing, returning to shelter and more importantly for our analysis, the population group and the age group which the homeless person fits in.

2.2 Variables of interest

Population group is divided into seven categories, All Population, Chronic, Families, Indigenous, Non-refugees, Refugees, Single Adult and Youth, with Indigenous only being a population group which the SMIS began tracking in January 2021. Certain group definitions are important. First off, the population group Youth is defined by the Open Data Toronto Portal as "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." Next, the age groups in the data set are divided as such: peoples aged under 16, aged 16 to 24, aged 25 to 44, aged 45-64 and aged 65 and above. Importantly, we also use the date variable to

2.3 Data Processing

To process our data, we began by loading all the necessary libraries like opendatatoronto (Gelfand 2020) which was used to load the data into R (R Core Team 2020), a package called lubridate (Grolemund and Wickham 2011) which will be used to process date data which later proves to be quite tricky. We also used load in libraries knitr (Xie 2021), tidyverse (Wickham et al. 2019), tidyr (Wickham 2021), janitor (Firke 2021), and dplyr (Wickham 2021) at some points in the process for data processing and pdf document generation.

After loading the data into our local environment and saving its as a cvs called raw.csv, we began to create subsets of the data with the desired variables. we created a new column called total which creates a count of the total number of users in the SMIS data base for each month, by each population group. This was done by simply summing up all the values for the number of people in each age group. We append this to the data set and then create a new data set called clean where we removed the variables which we did not desire like returned from housing and only kept the date, the age groups, the total and the population groups.

In addition to the totals column, I also needed a way to track the progression of data use over time. The raw data sent does provide a date column but this date deemed to be very problematic. The date in the data set is of class character and is in the form mmm-yy. For example, "Jan-20." Because the class is a charterer, I attempted to use the as.Date() function in the Lubridate package Grolemund and Wickham (2011) in order to convert the the date to class date, so that the visuals that we produce treat is as such and we can get the data in chronological order rather than in alphabetical order. R (R Core Team 2020) is very picky about how it likes its dates to be formatted and because there is no day data provided and we only have month and year, I needed to use two different work around to get chronological time data. The first was to add a Month column which is class numerical with values 1-24. These values represent the amount of months elapsed since the start of 2020. For example, March 2021 will have the month number 15. This way, we can organize the data chronologically.

Table 1: First ten rows of a dataset of shelture usage without All Population

Population Group	Aged less than 16	Aged 16-24	Aged 25-44	Aged 45-64	Aged 65	Total	Month
Chronic	429	432	1096	1205	307	3469	1
Refugees	969	341	1177	427	27	2941	1
Families	1438	238	790	231	9	2706	1
Youth	0	987	0	0	0	987	1
Single Adult	0	0	3103	2621	495	6219	1
Non-refugees	471	886	2716	2425	477	6975	1
Chronic	405	435	1029	1251	309	3429	2
Refugees	961	334	1174	407	26	2902	2
Families	1434	226	798	219	9	2686	2
Youth	0	957	0	0	0	957	2

Alternatively, we also added a day column where the day is the last day of every month (note 2020 was a leap-year and thus the last day in February 2020 is the 29th). We then created a column where the three values month, day and year where combined into a string with the proper formatting and then that date was converted using the as.Date( ) to a value of class date.

We also created 2 new data sets, no_all_population and all_population which removed the all population data as well as removed all non All Population data respectively. Each data set (raw.csv, clean.csv, all_pop.csv and no_all_pop.csv) are all available on GitHub through the link on the title page.

2.4 Population, Frame or Sample

It is important to discuss the reach of our data. As previously mentioned, the City of Toronto funds services for people experience homelessness and the data we have acquired for this paper reflects all the data collected in the SMIS system in the years 2020 and 2021. This is important because it brings to light one of the first limitations of the data which we will discuss more in Section 4. Although the SMIS data is quite expansive with 323 data points, it fails to capture the entire population of homeless people in the city of Toronto as it does not take into consideration homeless people who sleep on the street or those only using services which are not government funded. Thus, in this paper, we are only deal with a sample of the population, but a big one nonetheless.

Now, lets have a look at some of the final data we will be working with. They are illustrated below in Table 1

In Table 1, we notice that there are alot of zeros but nowhere in the data do we see *NA* values. The reason for the zeros is the definition of various demographics. As mentioned, a Youth is defined as only being able to be someone aged 16 to 24 so we expect not to be able to see any values for Youth under any columns other than that age range.

3 Visualizing the Data and The Implications

3.1 Visual 1

In this analysis, I am interested at looking at how the the CVODI-19 pandemic has affected the flow of homeless shelter usage over the years 2020 and 2021 for the various population and age groups. In this section, we will display and explain the various visuals we were able to meaningfully extract from the data and the implications we are able to extract from them.

Lets look at a graph which shows the usage of homeless shelters by people aged under the age of 16 over 2020 and 2021.

In this graph (Figure 1), we can see that shortly after the start of 2020, the use of government funded shelters

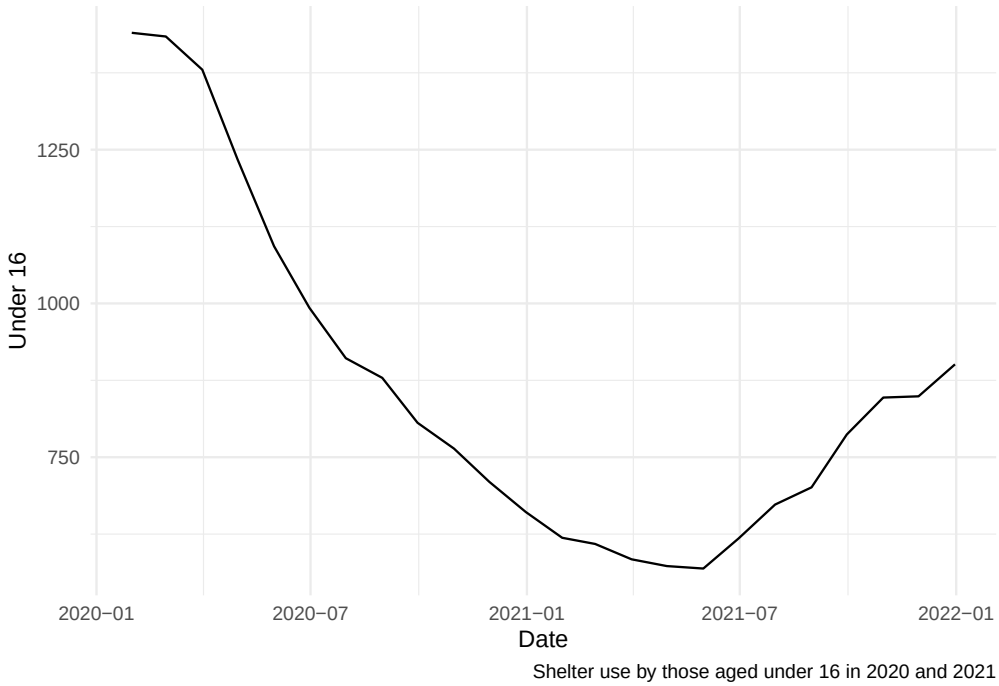

Figure 1: Shelter use by those aged under 16

in the SMIS database have been rapidly declining until June 2021 where it started to increase just as fast as it declined. Although at a surface level, it might seem like something very positive that shelter usage is decreasing, the fact of the matter is that homeless people simply do not disappear at such a rapid rate as illustrated in Figure 1. Perhaps this finding is unique to people under the age of 16. Lets take a look at similar graphs for all the other age groups and we will then talk about the implications of the findings. We will use the Patchwork package (Pedersen 2020) to visualize all the data together.

Above in Figure 2 we see all the data on the usage of SMIS homeless shelters across all age groups. Across all age groups, we notice a similar trend. After the start of the pandemic, there is a rapid decrease in shelter usage. Lets focus on this to begin.

3.2 Implication 1

As I've previously noted, homeless people simply do not disappear at the rate which Figure 2 suggests. Homeless shelters are known for often being crowded and not offering the best of living conditions. Although it is realistic sense that to offer the most amount of homeless people with shelter, capital restricts the quality of the service provided: there is an important tradoff between the quality and the quantity of the service provided. I believe that this element is exactly the cause of the rapid decline in shelter use at the start of the pandemic. Here is why. We have learned that the Corona virus and all of its variants are highly contagious and social distancing is one of the main ways that public health officials have advised the global population to try and reduce the spread of the virus. Because shelters can be quite crowded, especially in the winter months when the temperature outdoors drops, it would appear from the above graphs that even with the temperatures slightly increasing outdoors at the start of spring in march, many homeless people including those aged under 16 opted to sleep outdoors where although they might be defenseless against the cold, they are protecting themselves against a possible COVID-19 infection.

This is definitely tragic and one might wonder how can we remedy this situation. The answer can be found in the second half of the graph.

So far, what we have done is comment on the implications of the first half of the graph where shelter usage was at a rapid decline. Now, let's look closely at the second part of the graph.

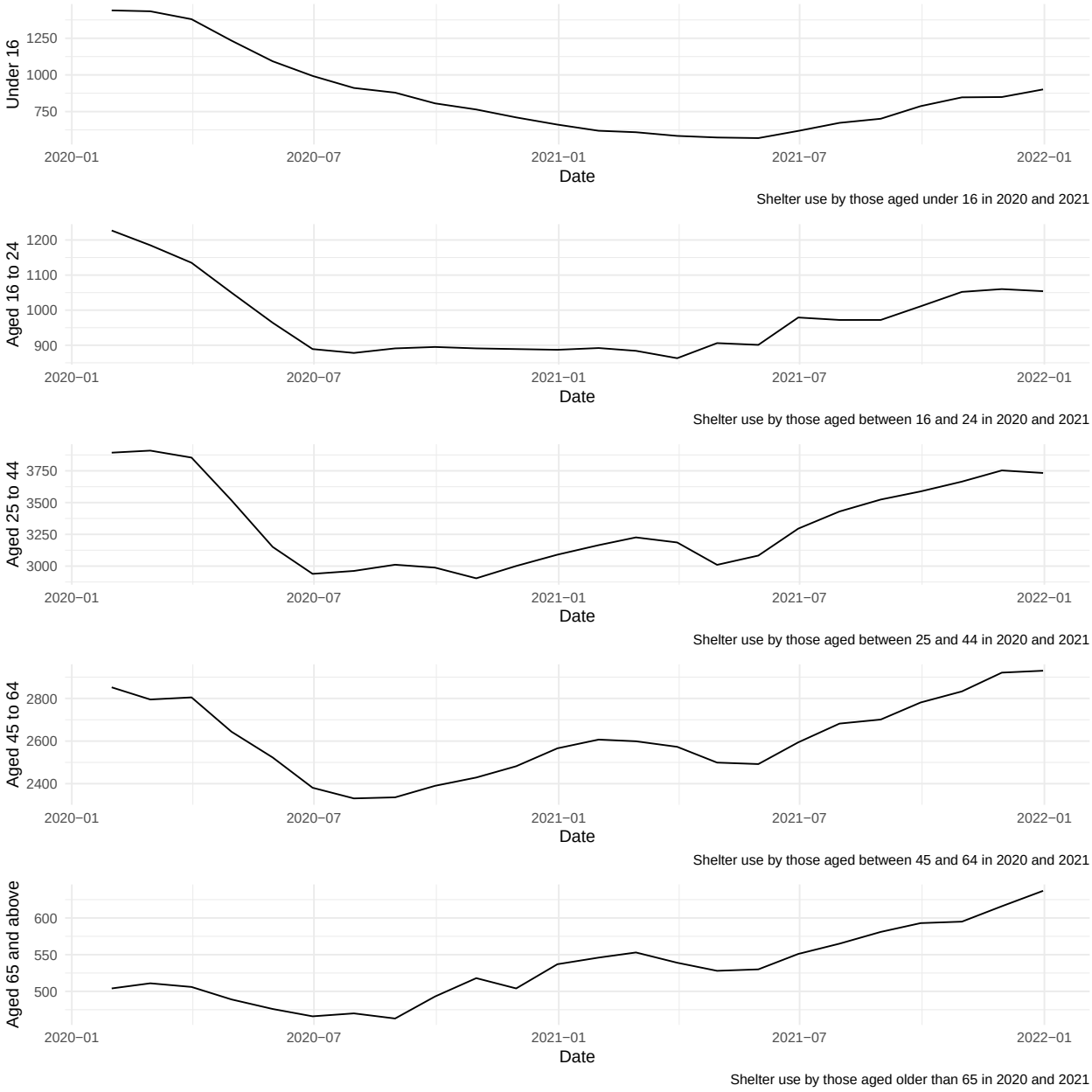

Figure 2: All age groups over time.

In the second half of the graph, we notice the opposite trend taking place. The shelter usage increases seemingly just as fast as it declines. This is great to see but will makes us wonder what could have caused this transition at the midway point of 2021. In my estimation, vaccination and education has been the solution. Because a world class race began to take place, scientists and epidemiologists from around the globe were able to develop vaccines in very short period of time in contrast to historical standards. As such, vaccination was widely available for all those wanting to get vaccinated and from the visuals above, it would appear that a combination of cooling temperatures and peoples garnered trust in the effectiveness of vaccines lead to confidence being restored to homeless folk who are looking for somewhere to stay at night.

3.3 Visuals 2

Moving on, lets see what implications we can draw when we separate the data by population group.

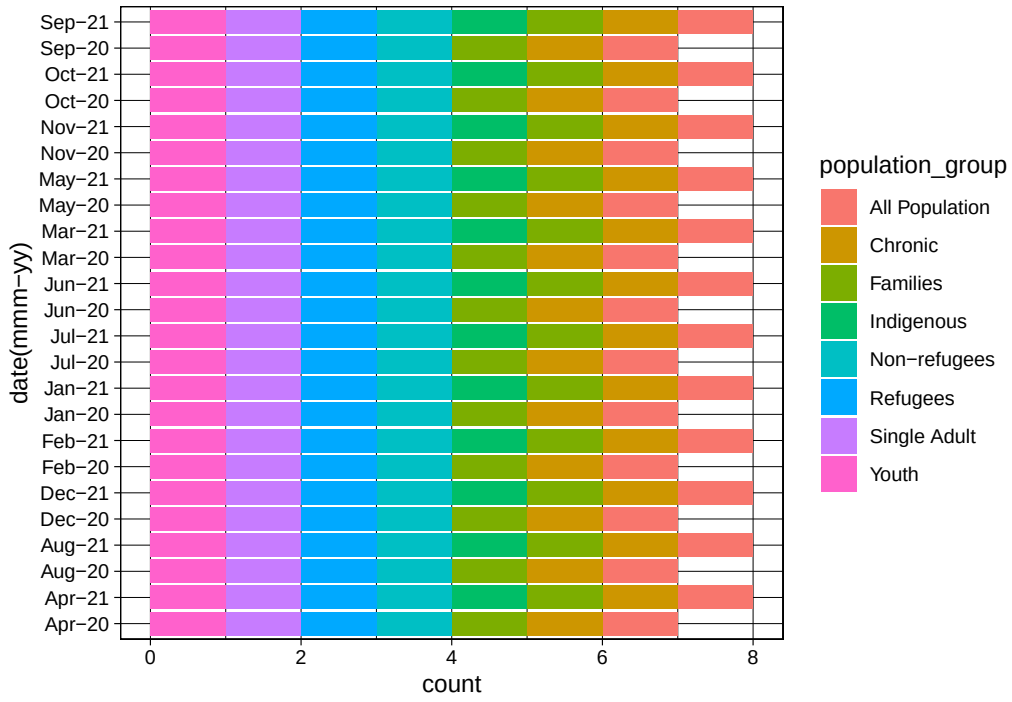

Figure 3: There are no Indeginous data for 2020

In Figure 3, we notice that for all the months which are in 2020, there is one missing bar. The bar that is missing is the Indigenous population which is not a metric that was tracked in 2020 and only started being tracked in 2021.

3.4 Implications 2

We have long know that most (if not all) of the land that we use in Canada to live and operate through our daily lives once belonged to the first nations that lived here. Although it is not as simple as this, the land belongs to them and they have often historically gotten the wrong side of the stick as it pertains to their treatment in society and the settlements that they have come to terms with, with the Federal government. As a result, many indigenous folk decide to leave their reserves, where there are high rates of murders, shootings and drug addiction, for a shot at better prospects in the cities. This is often not the case. Indigenous people of all ages find themselves in hard situation and in many cases, end up homeless. It is hard to understand how indigenous folk who have already suffered so much at the hands of the Canadian governments at all levels, have been omitted as a variable for all these years up until 2021.

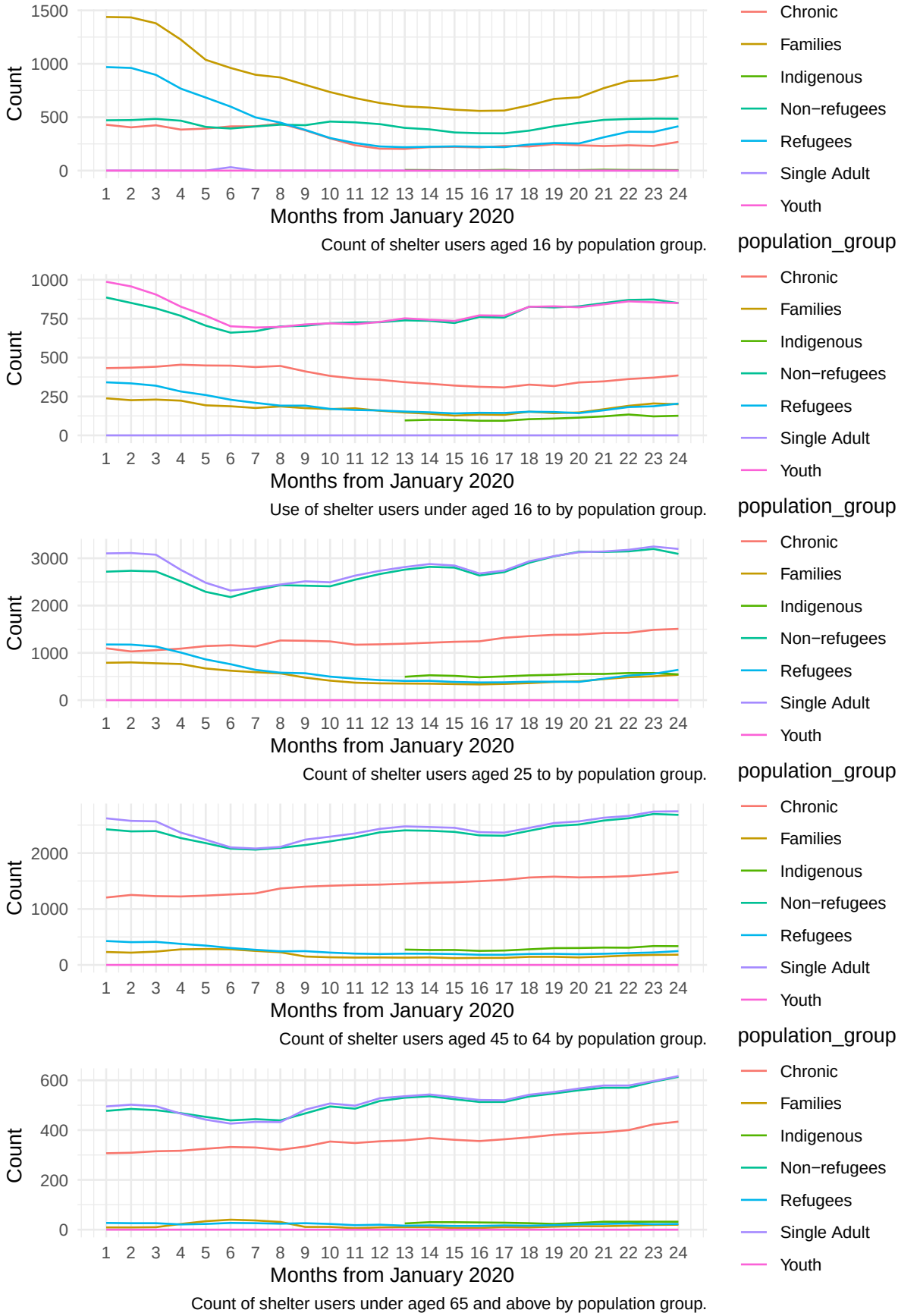

Figure 4: There are no Indeginous data for 2020

3.5 Visuals 3

3.6 Implications 3

In Figure 4, further implications can be drawn on the situation of the Indigenous homeless community. While in the other population groups, we see some change over time in the use of homeless shelters (Youth is the other exception as the value for youth will always be 0 except when the age group is 16 to 24 years old.). When we look at the Indigenous data, we find that in all age groups, there is little to no variation in the shelter usage. In my estimation, this is further reinforcing Visuals 3 and the implications we drew from it. There has been a lot of systemic racism throughout Canadian institutions towards minorities like indigenous folk and the data shows this clearly.

4 Limitations

In this limitations section, I want to focus on the data which is provided to us and have a look into why it might not be easy to take the implications mentioned above at face value

4.1 Uneven Bin width for Age Groups.

The first major flaw I've found in the data is that the bins that the age groups are recorded in are not all the same size. For example, the first bin has a size of 16 years, the second bin from ages 16 to 24 has width 9, the next two have width 19 with the last ranging in theory from 65 to infinity. Although it is possible to understand why the maintainers of the data choose to have a demarcation point of the bins at 65 and 16, it is very hard to understand why they opted to divide the interior portion into varying bin sizes.

We know in data science that when plotting bar charts, the bin width can have a big implication on the presentation of the data and the things that we can infer about its distribution. If the bin width is too small or too large, we might reduce or add characteristics to the data that aren't actually representative of it.

4.2 What about the others?

The next and last limitation that we have when using this data as our only primary source is that we don't have any data on the homeless people who remain sleeping outdoors and on the homeless people who use shelter systems for assistance but not those funded by the government or upheld by the SMIS.

5 Next Steps

Following on from the latter section, we can make some pretty good suggestions on ways to improve the data so that the findings we have will be less ambiguous and more ethical to use with confidence. First off, we can start by creating eve bins that homeless people checking into SMIS will have to answer. An even better alternative would be to collect the exact age and date of birth of homeless people (date of birth so that the age can be increased at the appropriate date every year). In this manner, the data scientist who gets a hold of the data will have the liberty to conduct his analysis and chose the bins which work best for the task he is working om, instead of being obstructed by barriers built into the data. This will not only ease the job of the data scientist but will also improve the reproducibility of his workflow and finding.

Next, the city might want to look at how it can improve and expand the SMIS system as to keep track of all the homeless shelters and services in the City of Toronto, as well as find ways to consider and count homeless folk who sleep on the street

References

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*.

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. *Tidyr: Tidy Messy Data*.

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.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.