Understanding The Toronto Shelter System Flow Pattern*

A Preliminary Analysis

Yifan Tong

January 23, 2024

The shelter system, often funded by the city, provides a temporary stay to people who are homeless. The system if of vital importance to our society as it acts as a socail safety net to protect the vulnerable population and its usage pattern sheds light on policy formulation. Using the real world data on shelter system flow in the city of Toronto, the paper identifies several patterns. The paper finds that males are more likely to be homelss than females. In addition, the trend of total homeless people does not coincide with the trend of new homeless people and people moving out of homeless status, indicating factors like macroeconomic conditions and policies play a vital role in the process.

1 Introduction

The shelter system, often funded by the city, provides a temporary stay to people who are homeless. The shelter system is of vital importance to the society for several reasons. Firstly, by allowing everyone to have a roof over his head, the system protects the dignity and safety for one of the most vulnerable populations (Morse 1999). Secondly, the shelter system can detect potential health issues for the homeless and offer additional assistance, acting as another layer of social safety net (Morse 1999). In addition, the shelter system can yield economic benefits by reducing the cost on potential crimes and unemployment, improving the economic efficiency in the long run (Paat et al. 2021). Lastly, the shelter system, often as the last resort for the low-income population, provides data and insights on the vulnerable population. Analyzing the data, such as the composition and trend, would shed light on the evolving economic conditions for the vulnerable population and provides policy insights (Brown et al. 2017).

*Code and data are available at: https://github.com/YifanTong02/Term-Paper1/tree/main

Despite the importance of the shelter system, the scarcity of data often limits researchers from investigating the social issue. Motivated by this gap, the research uses real world data from the City of Toronto, whose observations range from 2018 Jan. to 2023 Dec. Since the data covers pre-pandemic, pandemic, and post-pandemic era, it can also help us understand the effect of pandemic on the housing pressure and offers housing policy suggestions for the post-pandemic era.

The research renders the following main findings. Firstly, males are much more likely to be homeless than females. Secondly, the number of actively homeless people can vary drastically, featuring high volatility but no clear seasonal pattern. The number of actively homeless people first declined during pandemic and then increased sharply when the pandemic approached an end. Thirdly, the number of newly identified homeless people and the number of people finding a new house have different trends with the actively homeless people, indicating factors like macroeconomic conditions could play a vital role in the diverging trends.

The paper is organized as follows. The data section presents a brief introduction on the context and characteristics of the data, which also includes preliminary data processing. The discussion section presents statistical and visual findings on the data. The conclusion section summarizes the findings and extend them to relevant policy aspects.

2 Data

2.1 Data Collection and Context

The data set to be used in this research is sourced from the City of Toronto Open Data Portal using R package `opendatatoronto` (Gelfand 2022). The original file is named toronto-shelter-system-flow in the website, although the file is renamed in the Github repository. The data is monthly that records the entering and leaving in the shelter system, and includes more details about demographics and previous housing status. Analyzing the data can reveal who is using the system and the trend of people experiencing homeless. Since observations range from 2018 Jan. to 2023 Dec, both pre-pandemic and past-pandemic periods are covered, allowing us to understand the effect of pandemic on housing pressure. The data was last refreshed on Jan. 14 2024, and the data is of high quality in terms of freshness and completeness.

2.2 Measurement and Variables of Interest

The original data is mostly count data, under which each cell represents the count for a specific population under the category determined by the row and column. Each row represents a certain group for a certain month. Several variables of interest would help us investigate the matter. The first variable is date, which records the date of the observation and is especially helpful in analyzing the trend. The second variable of interest is actively_homeless, which

identifies the number of people who have used the shelter service in the past three months. The variable helps capture the need for the shelter service. The third variable of interest is the newly_identified, which counts the number of people who entered the system for the first time. Combining the second and the third variable would help us understand whether the need for sheltering is mostly persistent or temporary. Besides, the gender_male and gender_female variables count the number of males and females during intake in that month. The last variable of interest is the moved_to_housing, which can be considered as the number of people who are leaving the shelter system because they have a house to live in. This variable helps capture the outflow of the system. For example, one can reasonably conjecture that more people would leave the housing system when unemployment rate becomes lower. The count data, however, suffers from the limitation that it does not scale and can be misleading, thus we sometimes calculate the percentage for more robust analysis.

2.3 Data Cleaning and Processing

The data processing and future data analysis is conducted in the R (R Core Team 2022) environment. Additional packages used include `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `knitr` (Xie 2023), and `lubridate` (Grolemund and Wickham 2011). `tidyverse` (Wickham et al. 2019) is used for changing column content, adding new columns, and deleting columns in the data. `ggplot2` (Wickham 2016) is used for visualizing the data. `lubridate` (Grolemund and Wickham 2011) is used for converting some character to date format in R. `knitr` (Xie 2023) is used to generate tables and the whole report. Table 1 (Table 1) presents the summary statistics after data cleaning and processing.

Table 1: Summary Statistics

newly_identified	actively_homeless	gender_male	gender_female	moved_to_housing
Min. : 368.0	Min. : 7506	Min. :4690	Min. :2570	Min. :235.0
1st Qu.: 718.8	1st Qu.: 8575	1st Qu.:5384	1st Qu.:3075	1st Qu.:343.2
Median : 845.5	Median : 9558	Median :5854	Median :3584	Median :500.0
Mean : 881.4	Mean : 9291	Mean :5728	Mean :3438	Mean :503.4
3rd Qu.:1106.0	3rd Qu.: 9984	3rd Qu.:6088	3rd Qu.:3754	3rd Qu.:637.0
Max. :1469.0	Max. :10849	Max. :6638	Max. :4096	Max. :952.0

3 Discussion

3.1 Gender Composition

Figure 1 (Figure 1) analyzes the average male and female intake by year. In the whole City of Toronto population, one should expect the number of males to be roughly equal to the number of females. If males and females are equally likely to be homeless, one should expect rather similar average male and female shelter intakes by year. However, Figure 1 (Figure 1) shows that in every year, the number of males are much more than the number of females. In fact, the average number of homeless males almost double than the number of homeless females. The pattern remains persistent over years. There could be several likely causes explaining why males are consistently more likely to be homeless. Firstly, it could be because males are more likely to have a criminal record than females, and the criminal record reduces one's probability to find a job and thus find a house. Secondly, it could be because males and females have on average different mental health conditions, and the mental health condition further prevents someone from finding a place to rent or live. Lastly, this could be attributed to a matter of taste. Females may be more likely to reserve a substantial portion of their income for housing. Although the causes are debatable, the phenomenon calls for the attention of policy makers to correct the gender imbalance in shelter seekers.

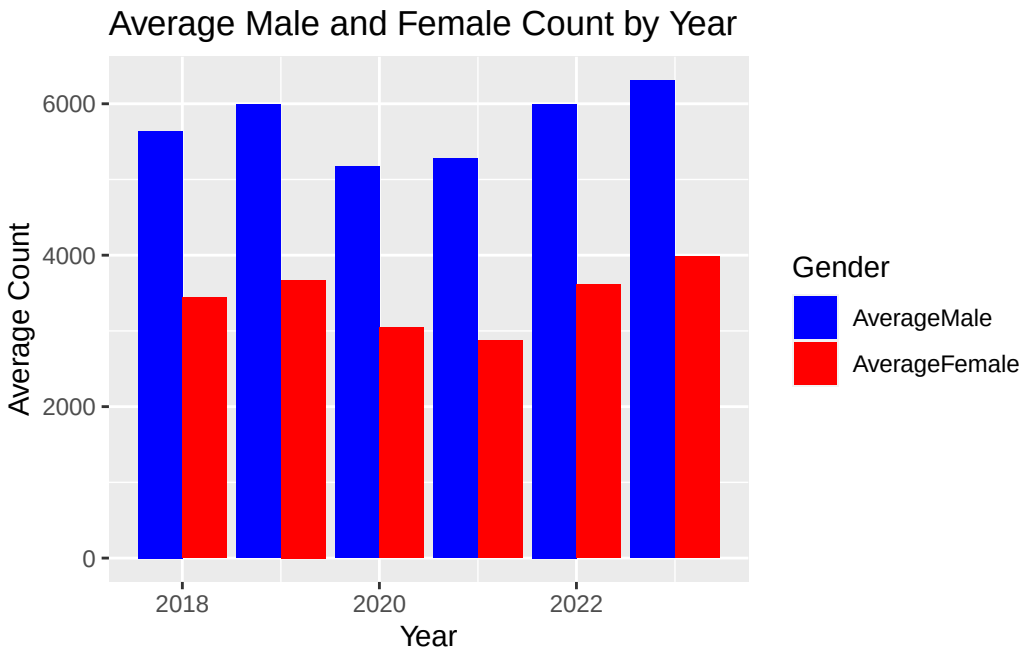

Figure 1: Gender Composition by Year

3.2 Time Trend

Figure 2 (Figure 2) analyzes the trend of actively homeless individuals. Overall, this number seems rather volatile. In 2018, there is first a rise in the number of actively homeless people and the number becomes less volatile during 2019. In 2020, when the pandemic first hits, there is a sharp decline in the number of actively homeless people. This could be due to several reasons. For example, the government may have issued subsidy for housing or prohibited landlords from evicting tenants during pandemic. When the pandemic eases and relevant policies ends, there is a sharp rise in the number of actively homeless people, and the number peaks in the beginning of 2023. Overall, there is a very slight upward trend, but it is hard to tell if the trend is statistically significant from the plot. In addition, there is no clear seasonality on the data, indicating being homeless has no or rather weak relationship with the month or season.

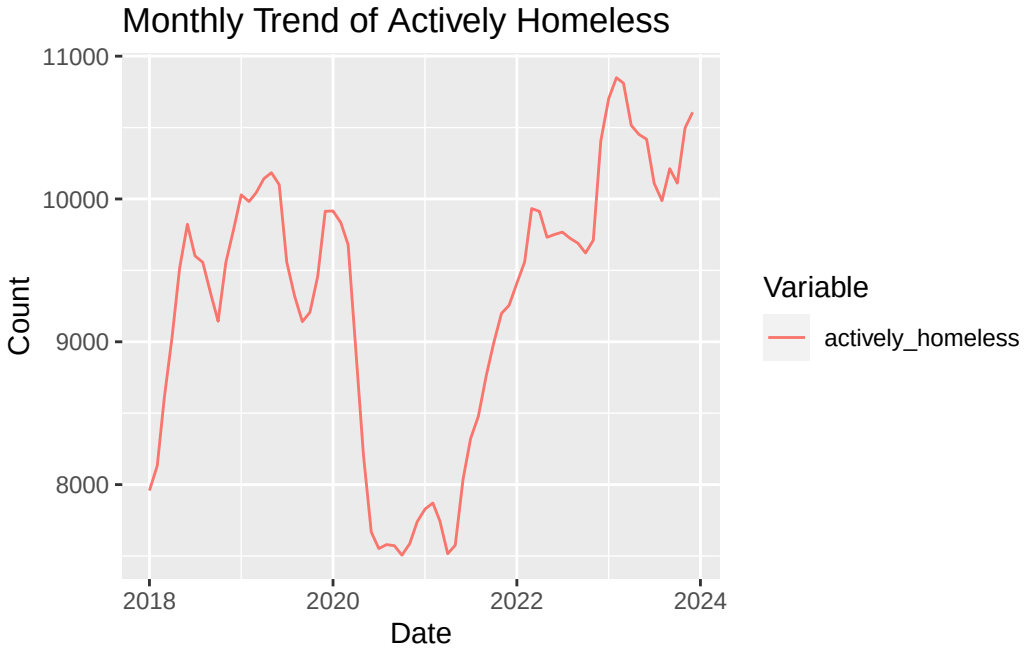

Figure 2: Trend of Actively Homeless

Figure 3 (Figure 3) investigates the time trend on the number of people who are newly identified into the shelter system and people who have moved out of the system to a house. The trend of newly_identified is similar to the trend of actively_homeless in the sense that it experiences a sharp drop and rise during the pandemic. The drop could be explained by relevant policies that offer subsidies and prohibit eviction. Nevertheless, there is no upward trend for newly_identified, potentially because the economy is getting better which limits the new inflow to the shelter system. Notice that newly_identified does not count the intake of people who have used shelter in recent periods, so the different overall trend compared to that in figure 2 (Figure 2) suggests the significant rise of homeless people after the pandemic is mainly

attributed to people who are homeless before the pandemic. For the moved_to_housing variable, one could see that it first decreases until hitting a bottom in the beginning of 2022. Afterwards, the number of people moving to a new house increases overall. Observe the fact that the increase of moved_to_housing lagged the increase of actively_homeless, which implies other factors like macroeconomic conditions must play a role in the diverging trend.

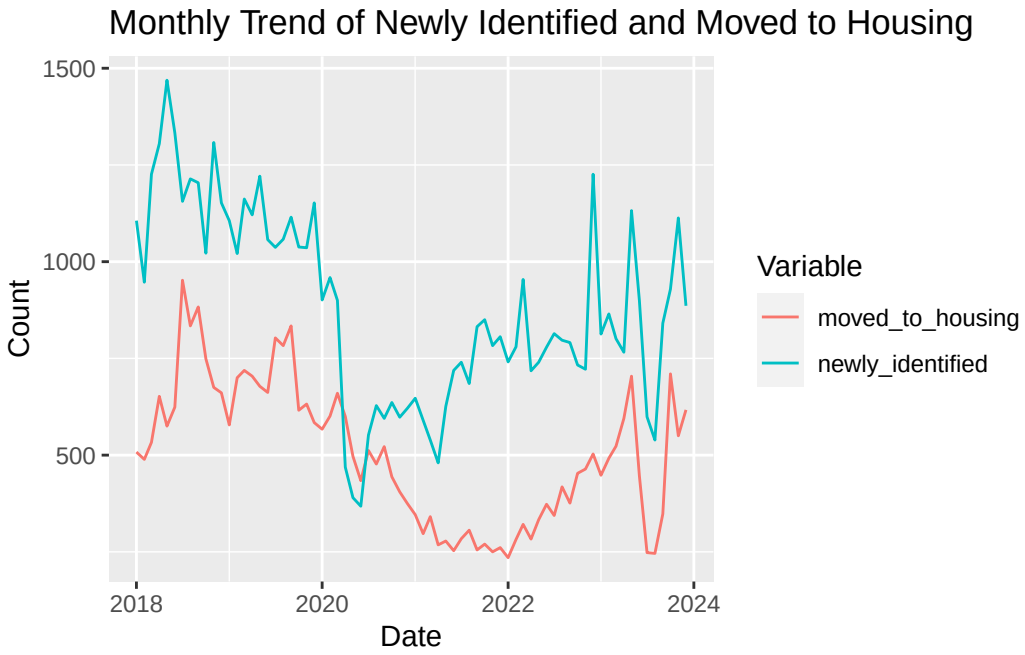

Figure 3: Trend of Newly Identified and Moved to Housing

4 Conclusion

The paper renders several main insights on the pattern of shelter system usage. Firstly, males are much more likely to be homeless than females. In fact, the chance of having a homeless male almost doubles the chance of having a homeless female. Causes can be multifaceted in explaining the gap, including different socioeconomic conditions for low-income males and females, preference for housing, and ability to find an unskilled job. In addition, the research analyzes the trend of homeless people. For the total number of actively homeless people, there is a slight upward trend over time, with substantial volatility but no seasonality. At the beginning of the pandemic, there is a sharp decline on the number of total homeless people, most likely due to different policies. The number rises again afterwards. Meanwhile, the trend of newly_identified homeless and homeless who have moved to housing do not coincide with the total homeless after the pandemic, indicating the percentage of inflow and outflow is actively determined by other factors like macroeconomic conditions.

The findings on the pattern may thus yield some policy insights. Firstly, policies that aim to reduce the number of homeless people must take gender into consideration, which could help uncover the reason behind being homeless and design more targeted measures. Secondly, more policies are needed to help people walk out of the shelter system, as data suggests a substantial portion of the shelter need is not temporary. Lastly, policies on tackling homelessness should also consider macroeconomic conditions. That means pure subsidy policies could be less effective than a combination of policies.

References

Brown, Molly, Martina Mihelicova, Jennifer Lyons, Jennifer DeFonzo, Samantha Torello, Andrés Carrión, and Allison N Ponce. 2017. "Waiting for Shelter: Perspectives on a Homeless Shelter's Procedures." *Journal of Community Psychology* 45 (7): 846–58.

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

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

Morse, Gary. 1999. "A Review of Case Management for People Who Are Homeless: Implications for Practice, Policy, and Research." *Practica [Less~ Rbs: C* 7: 1.

Paat, Yok-Fong, Jessica Morales, Aaron I Escajeda, and Ray Tullius. 2021. "Insights from the Shelter: Homeless Shelter Workers' Perceptions of Homelessness and Working with the Homeless." *Journal of Progressive Human Services* 32 (3): 263–83.

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

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.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.