Tracking Homelessness: A Statistical Analysis of Toronto's Shelter System Flow*

Jiaxuan Song

December 9, 2025

This report is based on the Toronto Shelter System Flow dataset from Toronto Open Data, providing an in-depth analysis of how the city's homelessness services are functioning. The data offers valuable insights into the movement of people within the shelter system, helping to measure Toronto's progress in reducing homelessness. The goal of this analysis is to better understand the persistence and patterns of homelessness, ultimately providing evidence to support efforts to reduce homelessness to rare, brief, and non-recurring instances.

Table of contents

*The GitHub Repository containing all data, R code, and other files used in this project is located here: https://github.com/Jiaxuan-Song/Toronto-Shelter-System-Flow-.git

1 Introduction

As society continues to evolve, homelessness has become a major challenge faced by cities around the world, including Toronto. The causes of homelessness are complex and multifaceted, typically involving personal, economic, social, and structural factors. Economic difficulties, such as low income, unemployment, and high housing costs, are some of the primary reasons behind homelessness. Additionally, personal factors like mental health issues, substance abuse or addiction, serious illnesses, and family breakdowns also increase the risk of individuals becoming homeless (Nishio et al. 2017). Long-term homelessness not only severely impacts an individual's physical and mental health but also places a heavy burden on healthcare systems, social welfare services, and public resources. Homeless individuals are often at higher risk of illness and have a lower quality of life, which makes it even harder for them to escape poverty and secure stable housing (Goodman, Saxe, and Harvey 1991). Given these challenges, addressing homelessness has become increasingly urgent.

In response to this issue, the City of Toronto has implemented a range of measures designed to support people experiencing homelessness. These measures include emergency shelters, respite services, and additional programs like hotel/motel initiatives and warming centers, all aimed at providing essential shelter and resources to those in need (Hwang 2000). However, providing shelter alone is not enough. To better address homelessness and develop effective policies, the city has established a statistical system to track and manage individuals accessing these services. Through this system, Toronto can not only monitor homelessness trends in real-time but also allocate resources more accurately, helping vulnerable populations move towards long-term stability. This study aims to analyze the Toronto Shelter System Flow data (Gaetz 2020) to explore patterns of homelessness, assess the impact of socio-economic and demographic factors on shelter usage, and identify potential disparities within the system, providing insights for improving services. The data will be processed and analyzed using R, focusing on demographic characteristics such as age, gender, and ethnicity in relation to shelter use, with particular attention to the movement and patterns of various groups. Through this analysis, we hope to offer more insights into the mobility of homeless populations in Toronto, helping to shape more targeted policies and promote a fairer and more efficient shelter system in the future.

2 Data

This dataset is sourced from the City of Toronto's Open Data Portal(Gelfand 2022) and records the flow of individuals through the city's shelter system. The data is collected via the Shelter Management Information System (SMIS) and includes personal information of those using shelter services, covering various groups such as families, youth, and single adults (*Shelter Management Information System (SMIS)* 2024). The dataset is updated monthly, providing the latest data from the previous month to ensure timely information. It contains multiple variables, such as shelter service types, entry and exit times, offering crucial insights into the shelter system's operations. To ensure accuracy, certain fields are updated within two weeks after a person exits the shelter (Gaetz 2020).

In this study, the data was processed using R programming language (R Core Team 2022), primarily utilizing the tidyverse package (Wickham et al. 2019) for cleaning and analysis. For simplicity, some demographic characteristics (such as age groups) were consolidated into broader categories. Finally, data visualization was conducted using ggplot2 (Wickham 2016), illustrating the usage patterns and trends across different groups in the shelter system, helping to reveal the dynamics and characteristics of the homeless population.

2.1 Measurement

The data used in this analysis comes from the Shelter Management Information System (SMIS) (*Shelter Management Information System (SMIS)* 2024), which intakes individuals accessing emergency shelters, respites, and allied services such as hotel/motel programs and warming centres in the City of Toronto. The dataset, referred to as the "Shelter System Flow data (Gaetz 2020)," records the number of people experiencing homelessness who access these services, with a specific focus on individuals who have used the system within the past three months and are considered actively homeless. Actively homeless individuals are defined as those who have not been discharged to permanent housing during the reporting period.The SMIS data is updated monthly on the 15th (or the next business day). As new data is generated, an updated dataset replaces the previous month's data. This process ensures that any discrepancies in client discharge status, which may remain open for up to two weeks post-intake, are rectified before the final data is published. The system only captures information on individuals who have accessed overnight shelter services, and individuals sleeping outdoors or using non-city-funded services are not reflected in the dataset.

2.2 Results

After loading the dataset using the R programming language (R Core Team 2022), the `tidyverse` (Wickham et al. 2019) package was used to generate graphs. In doing so, R code was adapted from Alexander (2023).

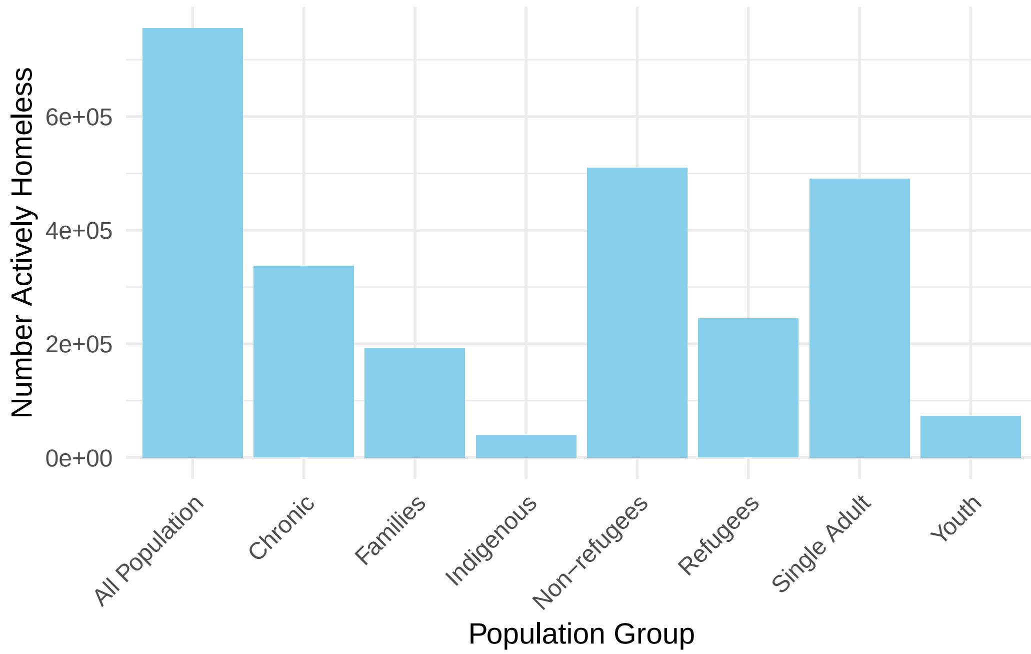

Figure 1: Actively homeless individuals across different population groups in Toronto 2018-2024

Figure 1 shows significant differences in the number of homeless individuals across various population groups. Certain groups, such as families and those experiencing chronic homelessness, have relatively higher numbers of homeless individuals, indicating that these groups may have a greater need for shelter services. In contrast, groups like youth or single adults have comparatively fewer homeless individuals, suggesting that these groups may rely less on shelter services. Additionally, the graph highlights a notable number of homeless individuals in the refugee group, which may reflect the challenges refugees face during the resettlement process.

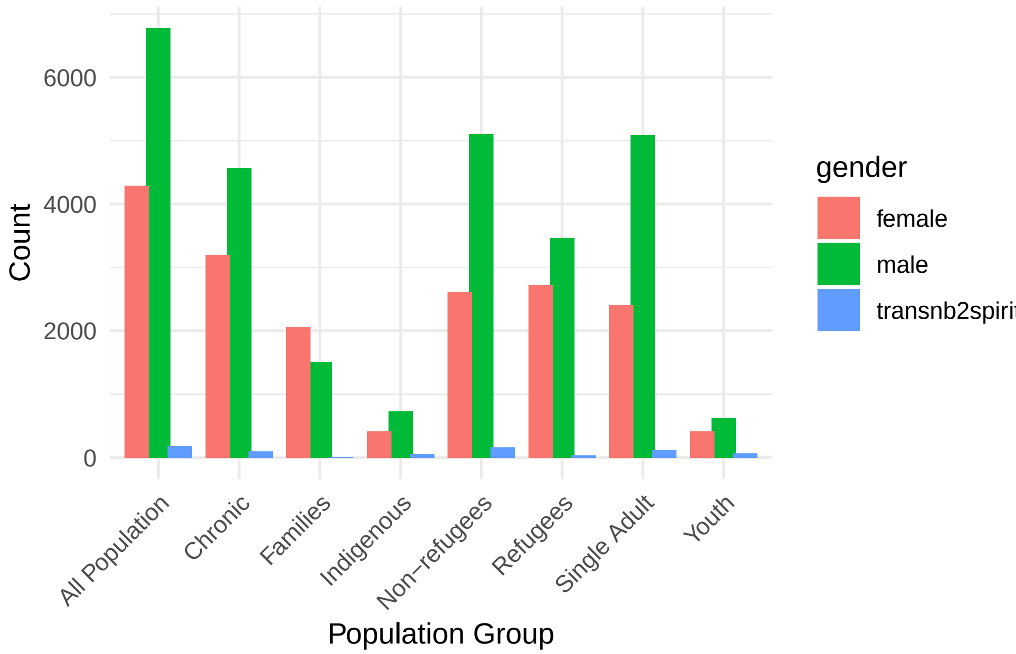

Figure 2: gender breakdown of actively homeless individuals across different population groups in Toronto 2018-2024

Figure 2 presents the gender breakdown of actively homeless individuals across various population groups in Toronto.Overall, males (blue) dominate the homeless population in several groups, particularly in the "All Population" and "Single Adult" categories, where the number of homeless males is much higher compared to females (red) and non-binary individuals (green).Females (red) have a relatively higher representation in the "Families" and "Refugees" groups, indicating that women in these groups may be at greater risk of homelessness. In contrast, non-binary and transgender individuals (green) have a lower presence across all groups, suggesting underrepresentation or fewer recorded homeless individuals in this category.

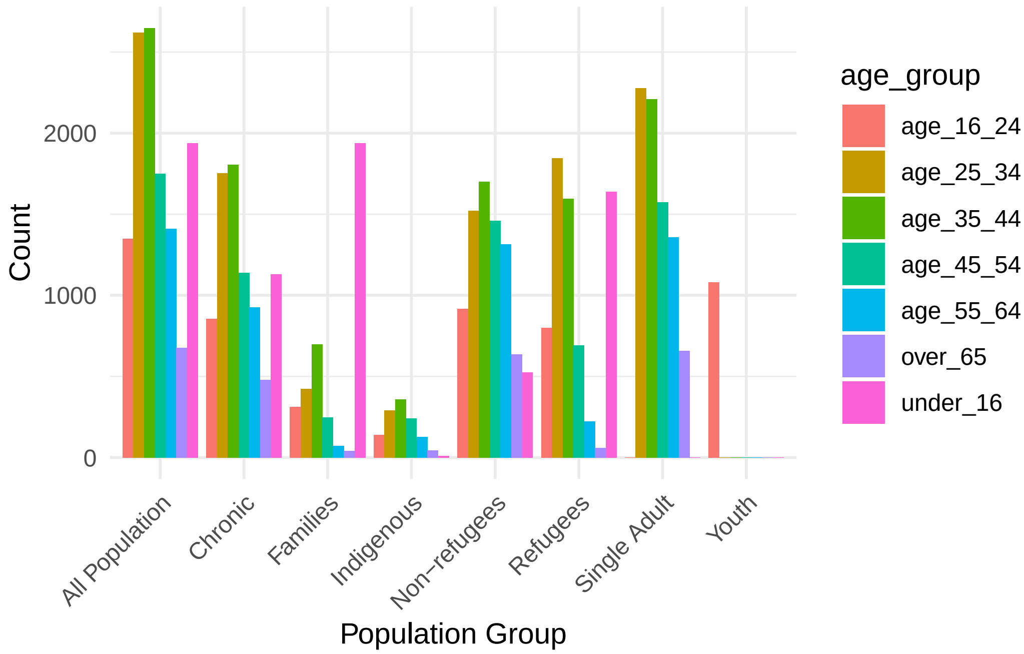

Figure 3: Age distribution of actively homeless individuals across different population groups in Toronto (2018-2024)

Figure 3 illustrates the age distribution of homeless individuals across various population groups, highlighting differences in the representation of age groups within each category. In the All Population group, those aged 16-24, 25-34, and 35-44 are the most represented, making up the majority of the homeless population. Among the Chronic Homeless group, middle-aged individuals aged 35-44 and 45-54 are dominant. The Families group shows a more balanced distribution, with a notable presence of individuals under 16, indicating that many homeless families include minors. The Indigenous and Refugee groups have a higher concentration of individuals aged 25-34 and 45-54, while the Single Adults group is primarily composed of those aged 35-44 and 45-54. As expected, the Youth group is concentrated in the 16-24 age range. Overall, the graph shows that individuals aged 25-44 make up a significant portion of the homeless population across multiple groups, while specific groups like Families and Youth include more younger individuals. This insight helps identify age-related needs within each group for targeted interventions.

3 Discussion

The study had conducted an in-depth analysis of the composition of Toronto's homeless population, focusing on homelessness across different demographic groups. Through the analysis of various population categories—such as chronically homeless individuals, refugees, families, youth, and single adults—we found that chronically homeless individuals and single adults make up a significant portion of the homeless population, highlighting the systemic challenges these groups face (Aubry, Klodawsky, and Coulombe 2012). Additionally, the gender and age distribution revealed disparities within the homeless population. For example, males dominate across all groups, while females and youth are more concentrated in the refugee and family categories. These findings underscore the complexity of homelessness and not only enhance our understanding of the homeless population but also provide critical data to support future policy interventions (Spetter 1996).

Our findings align with previous research (Nishio et al. 2017), confirming that chronically homeless individuals, due to a lack of stable social support and housing opportunities, often remain in long-term homelessness. This trend is particularly evident in the single adult group, who frequently lack family support and face more severe economic and mental health challenges. Furthermore, the higher proportion of females in the refugee and family groups may be due to the fact that these groups are more likely to consist of family units, which face unique challenges in adapting to new environments and accessing resources.

At the same time, youth homelessness highlights the vulnerability of this group during the transition into adulthood. Without adequate employment and educational opportunities, they are at higher risk of falling into a cycle of homelessness. The underreporting of gender minorities, such as transgender and non-binary individuals, may be related to systemic recording deficiencies, reflecting the "invisibility" (Norris and Quilty 2021) of these groups within the homeless population and suggesting the need for greater attention and data collection in the future.

While this study provides valuable insights into Toronto's homeless population, it also has several limitations. First, the study only includes individuals who used overnight services, excluding those who sleep outdoors or use other non-municipally funded services, potentially underestimating the total number of homeless individuals. Second, the proportion of gender minorities in the data is low, indicating that these groups may be systematically overlooked or underreported within the homeless population. Additionally, the time range of the study limits our ability to analyze long-term trends, preventing us from fully understanding how homelessness evolves over time and across different spatial contexts. Finally, the study relied solely on municipal data, possibly missing key subgroups such as migrant or informal populations.

A Appendix

A.1 Dataset and Graph Sketches

Sketches depicting both the desired dataset and the graphs generated in this analysis are available in the GitHub Repository.

A.2 Data Cleaning

The data cleaning process involved filtering out some of the columns from the raw dataset and renaming some of the data entries for clarity and simplicity.

A.3 Attribution Statement

"Contains information licensed under the Open Government Licence – Toronto" (**tphlicense?**).

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. https://tellingstorieswithdata.com/.

Aubry, Tim, Fran Klodawsky, and Daniel Coulombe. 2012. "Comparing the Housing Trajectories of Different Classes Within a Diverse Homeless Population." *American Journal of Community Psychology* 49: 142–55.

Gaetz, Stephen. 2020. "Making the Prevention of Homelessness a Priority: The Role of Social Innovation." *American Journal of Economics and Sociology* 79 (2): 353–81.

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

Goodman, Lisa A, Leonard Saxe, and Mary Harvey. 1991. "Homelessness as Psychological Trauma: Broadening Perspectives." *American Psychologist* 46 (11): 1219.

Hwang, Stephen W. 2000. "Mortality Among Men Using Homeless Shelters in Toronto, Ontario." *Jama* 283 (16): 2152–57.

Nishio, Akihiro, Ryo Horita, Tadahiro Sado, Seiko Mizutani, Takahiro Watanabe, Ryosuke Uehara, and Mayumi Yamamoto. 2017. "Causes of Homelessness Prevalence: R Elationship Between Homelessness and Disability." *Psychiatry and Clinical Neurosciences* 71 (3): 180–88.

Norris, Michelle, and Aideen Quilty. 2021. "Unreal, Unsheltered, Unseen, Unrecorded: The Multiple Invisibilities of LGBTQI+ Homeless Youth." *Critical Social Policy* 41 (3): 468–90.

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

Shelter Management Information System (SMIS). 2024. City of Toronto. https://www.toronto.ca/community-people/community-partners/emergency-shelter-operators/shelter-management-information-system/.

Spetter, Victoria C. 1996. "As Government Assistance Decreases, Homelessness Increases: A Closer Look at Welfare, Housing and Homelessness." In *Hybrid*, 3:111. HeinOnline.

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.