

Active Homeless Population: A Growing Imbalance in Proportions*

Yuxin Sun

September 22, 2024

In this report, the dataset on the actively homeless population was analyzed. The data shows that the non-refugee population is much larger than the refugee group, while single adults significantly outnumber those with families. In terms of gender distribution, males continue to dominate the overall population. Over time, from 2018 to 2024, there has been a marked increase in chronic homeless individuals, refugees, and single adults, indicating significant growth within these categories. These trends underscore the importance of tailored interventions to address the growing challenges faced by these groups. Future research should investigate the underlying causes of these increases and evaluate the impact of current policies aimed at supporting these vulnerable populations.

Table of contents

1	Introduction	2
2	Data	3
2.1	Overview	3
2.2	Raw data	3
2.3	Cleaning data	4
2.4	Results	6
3	Discussion	8
4	Limitations and future	8

*A GitHub Repository containing all data, R code, and other files used in this investigation is located here:
https://github.com/Yuxin-Sun-Caroline/Active_Homeless_Imbalance_Toronto

A Appendix	10
A.1 Dataset and Graph Sketches	10
A.2 Data Cleaning	10
A.3 Attribution Statement	10
References	11

1 Introduction

The rising homeless population in Toronto is leading to significant increases in violence, both within the shelter system and on the streets (News 2024). From 2014 to 2021, a intense increase in critical incidents in Toronto shelters has rised to a shocking number of 9,982 per year, and 4,102 of them are alleged violence, including physical assaults, verbal abuse, threats of death and harm and harassment of staff and residents (Gibson 2024). This issue not only affects the safety and well-being of individuals experiencing homelessness but also plays a critical role in shaping the city’s overall image and quality of life. With growing concerns about public safety and social stability, understanding the distribution and demographics of the active homeless population becomes increasingly essential. By analyzing the population patterns, policymakers and community leaders can better address the root causes of homelessness and implement more effective solutions.

The Shelter Management Information System (SMIS) is used by the City of Toronto to track and manage individuals accessing the city’s shelter services. The system measures the population by recording those who use emergency shelters, respites, and other temporary accommodations, defining the “actively homeless” population as individuals who have accessed shelter services at least once in the past three months and have not been discharged to permanent housing(Gelfand 2024). A recent 2024 study shows that Toronto shelters maintain a nightly occupancy rate of 98%, with 77% of the homeless population experiencing homelessness for over six months(Victor 2024).

This paper utilizes data from the Toronto Shelter System Flow, encompassing 604 samples of the homeless population in Toronto between 2018 and 2024. It aims to analyze the distribution of different target populations and identify the major group within the actively homeless population. Additionally, this paper seeks to observe and assess the growth trends of this primary group, providing insights that may guide future policies to address homelessness more effectively.

This paper finds clear evidence that a significant portion of the actively homeless population falls within the categories of ‘single adult,’ ‘non-refugee,’ and ‘male,’ identifying these groups as the primary demographic for active homelessness in Toronto. Additionally, a noticeable growth trend in both the ‘single adult’ and ‘refugee’ populations highlights the urgency of

implementing targeted interventions to support these groups. The findings suggest that immediate actions are necessary to address the needs of these vulnerable populations and to prevent further increases in homelessness.

The structure of this report is as follows: Section 2 provides an overview of the dataset and methodology. Section 2.4 presents key findings from the analysis, while Section 3 discusses the implications of these results. Section 4 concludes limitations of the analysis of this data together with recommendations for policy and future research directions.

2 Data

2.1 Overview

For this analysis, the R programming language (R Core Team 2023) was used along with packages such as *dplyr* (L. H. Hadley Wickham Romain François 2023) and *knitr* (Xie 2023) to clean and organize the data. The *here* (Kirill Müller 2020) was used to find reference files. The *opendatatoronto* package (Gelfand 2022) was utilized to download the raw data directly from Toronto’s Open Data Portal, followed by data cleaning and analysis using the *tidyverse* (Wickham et al. 2019) suite of packages. The *ggplot2* (Wickham 2016), *RColorBrewer* (Neuwirth 2022), *scales* (O. iD. Hadley Wickham Thomas Lin Pedersen 2023) and *dplyr* (L. H. Hadley Wickham Romain François 2023) to visualize the graphs. This process ensured the accuracy and reliability of the analysis, allowing for insights into the trends within Toronto’s shelter system over time.

2.2 Raw data

The dataset used in this analysis is part of the “Toronto Shelter System Flow Data (Gelfand 2024),” which is regularly updated and published by the City of Toronto (Gelfand 2022). This dataset provides comprehensive details on the shelter system’s capacity, usage, and flow of individuals across different shelters in the city. The dataset captures various variables, including the number of actively homeless individuals, categorized by demographic groups such as gender, age, and refugee status. Key variables analyzed in this report include “Shelter Capacity” and “Occupancy Rate,” which track the shelter system’s capacity and how much of it is being used, respectively, as well as “Actively Homeless Population,” a variable reflecting the number of individuals experiencing homelessness. This data is crucial for understanding the demographic shifts in Toronto’s homeless population and the system’s ability to accommodate these changes.

2.3 Cleaning data

In the process of cleaning the homeless dataset, several adjustments were made to ensure accuracy and facilitate analysis. First, all instances of percentage values were stored as strings with a “%” symbol, making them non-numeric. These entries were cleaned by removing the “%” symbol and converting the values to numeric form, allowing for proper quantitative analysis. Additionally, the original dataset contained a number of columns with complex or inconsistent names, such as `age16.24` and `gender_transgender.non.binary_or_two_spirit`, which were standardized and renamed to more intuitive labels like `age_16_24` and `gender_transgender_non_binary_or_two_spirit`, respectively. This improved clarity and streamlined the data manipulation process. Furthermore, all unnecessary columns that were not directly related to the analysis were retained but kept consistent for future exploration.

Table 1: Sample of cleaned homeless data

ID	Date	Population Grouping	Returned From Housing	Returned to Shelter	Newly Identified	Moved to Housing	Became Inactive	Actively Homeless	Age <16	Age 16-24	Age 25-34	Age 35-44	Age 45-54	Age 55-64	Age 65+	Male	Female	Trans/binary	Non-binary	Two-spirit	Pop Group
			46	494	1106	508	862	7958	1233	1111	1341	1560	1318	973	422	496	3912	83	100.0		
1	Jan 18	All Pop-ulation	46	494	1106	508	862	7958	1233	1111	1341	1560	1318	973	422	496	3912	83	100.0		
2	Jan 18	Chronically Homeless	1	29	317	111	111	2532	223	346	313	403	505	495	247	163	270	30	31.8		
3	Jan 18	Refugees	4	32	651	328	75	2408	914	241	414	547	204	66	22	121	9177	12	30.3		
4	Jan 18	Families	0	14	561	321	57	2277	1232	187	276	411	130	23	18	968	1308	1	28.6		
5	Jan 18	Youth	12	64	116	44	116	924	0	924	0	0	0	0	0	571	1312	41	11.6		
6	Jan 18	Single Adult	34	416	429	143	689	4757	0	0	1065	1149	1188	950	404	342	1292	41	59.8		

A sample of cleaned data is represented above with Table 1. After loading the dataset using the R programming language (R Core Team 2023) and the `here` package (Kirill Müller 2020), the `tidyverse` (Wickham et al. 2019) package was used to generate graphs. In doing so, R code was adapted from Alexander (2023). For each of the categories in this report, we have specific definitions to represent the different groups involved. **All populations** refers to the total number of individuals without demographic breakdown, offering an overall count.

Chronic homelessness is represented by people who meet federal criteria, either by having stayed in shelters for 180 nights in the past year or for a total of at least 546 nights over the past three years. **Families** represent individuals staying in family-designated services, while **youth** refers to unaccompanied individuals aged 16 to 24 who are not part of a family. **Single adults** are those who do not fall under youth or family classifications. **Refugees** are individuals identified as such during shelter intake or through specific programs, while **non-refugees** are those who do not meet the refugee criteria. Finally, **Indigenous** refers to individuals who have self-identified as First Nations, Métis, or Inuit during at least one intake recorded in the Shelter Management Information System (SMIS). These categories help clarify distinctions and target interventions more effectively.

2.4 Results

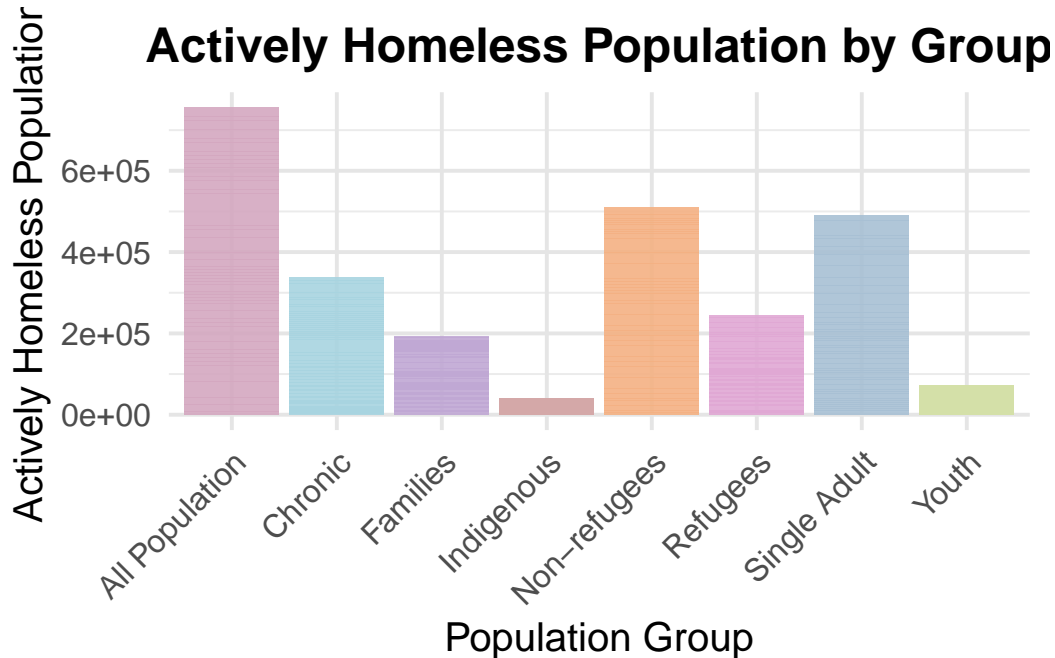


Figure 1: Actively Homeless Population By Group from 2018 to 2024 in Toronto

Figure 1 clearly illustrates the comparison and differences in the number of individuals identified as “actively homeless” across different groups in Toronto shelter data base. The “all population” category, representing the majority without specific demographic breakdown, shows a strikingly high number of homeless individuals. When comparing families with single adults, it is evident that single adults significantly outnumber families. Similarly, when comparing refugees with non-refugees, non-refugees far exceed the number of refugees.

Figure 2 shows that the vast majority of actively homeless in Toronto were males (approximately 61.6%). Females on the other hand, shows a lot less percentage by (37%), which makes approximately nearly a half less than the males. For Transgender or non-binary group, they played only (1.4%) in this distribution. However, such a low proportion could also be due to the fact that transgender individuals may not openly disclose their gender.

Figure 3 illustrates the growth history and trends of different categories within the actively homeless population from 2018 to 2024. Notably, starting from 2022, there has been a significant increase in chronic, refugee, and single adult populations, in addition to the overall “all population” category. The year 2022, marked by the initial economic impacts of the COVID-19 pandemic and emerging global geopolitical challenges, highlights the severe toll these shifts have taken on vulnerable groups, including single adults and refugees. The rise in the chronic homeless population further indicates that the long-term effects of these challenges will prolong the difficulty and pressure on efforts to restore safety and stability in communities.

Gender Distribution in Population

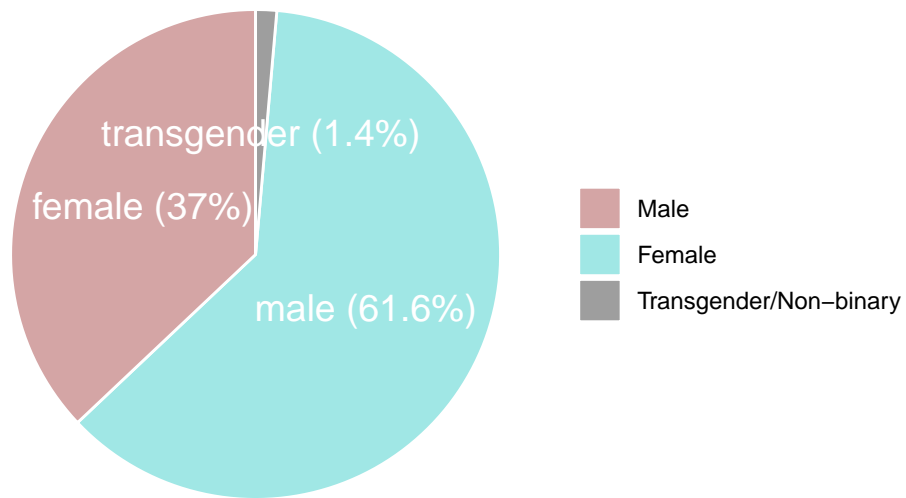


Figure 2: Porportion of different genders in Actively Homeless Population

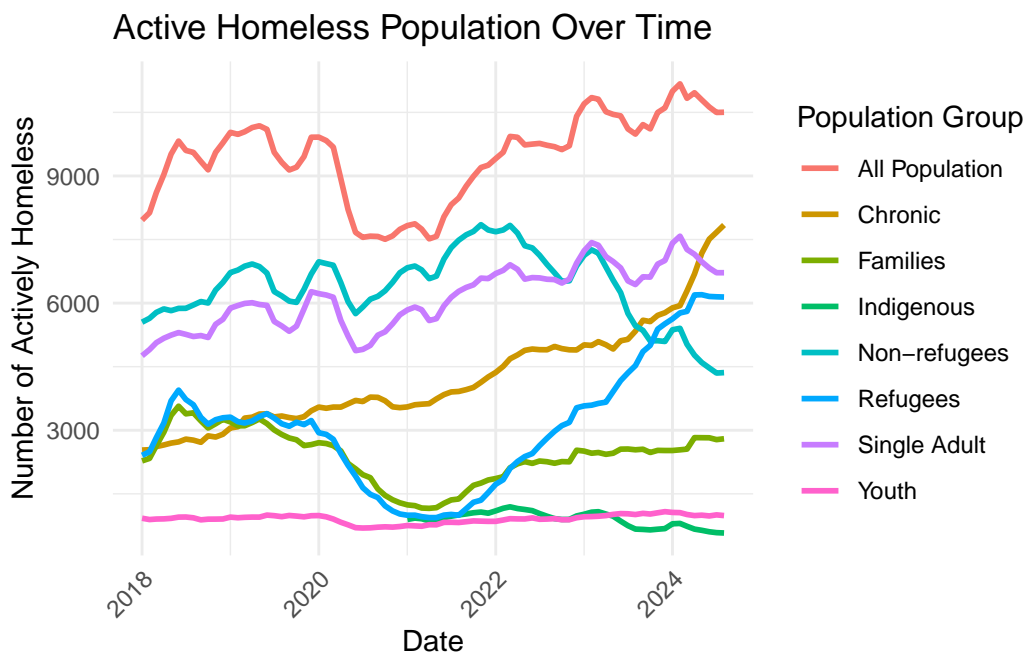


Figure 3: Active Homeless Population Change From 2018 to 2024

3 Discussion

In Figure 1, the comparison between different groups within Toronto’s shelter system clearly highlights significant disparities in the actively homeless population. The all population category, representing the majority, shows a strikingly large number of homeless individuals without demographic breakdowns. When focusing on specific groups, single adults significantly outnumber families, and non-refugees far exceed refugees. These findings align with the broader trends observed in Figure 3, which shows a marked increase in chronic, refugee, and single adult homeless populations beginning in 2022, likely driven by the economic and social effects of the COVID-19 pandemic. This sharp rise in homelessness within these vulnerable groups underscores the ongoing and prolonged impacts of these global crises, making it more difficult to restore community stability.

Gender distribution data from Figure 2 further reveals that males constitute the vast majority of Toronto’s homeless population at 61.6%, with females making up 37%. The proportion of transgender or non-binary individuals is notably low at just 1.4%, though this could be partly due to underreporting, as some individuals may not feel comfortable disclosing their gender identity. These combined findings emphasize the urgent need for targeted policies that address the distinct needs of these subgroups, especially given the increasing challenges observed in recent years.

4 Limitations and future

There are several limitations to the analysis conducted and the conclusions drawn regarding Toronto’s homeless population. Firstly, the dataset only reflects individuals who have utilized overnight shelter services, excluding those exclusively sleeping outdoors or using alternative homelessness services. As a result, this may lead to an underestimation of the total homeless population, as a significant segment of individuals experiencing homelessness may not be represented. Furthermore, shelter sites that do not utilize the Shelter Management Information System (SMIS) and are funded by other levels of government are also not included in the dataset, creating additional gaps in the analysis.

Another important limitation is that the Open Data set is updated monthly, typically on the 15th day of the month. The reason for this delay is to allow the discharge field in SMIS to remain active for two weeks, enhancing the accuracy of the final discharge records. While this system increases accuracy, it also means that the dataset for any given month is incomplete until the following update is published. Thus, the dataset is in a constant state of revision, and previous month data are replaced when a new extraction is complete, which could affect the consistency of conclusions over time.

Additionally, while extensive data on Toronto’s shelter system flow from 2018 to 2024 was analyzed, demographic information such as gender and refugee status relies on self-reporting.

This reliance on self-reported data may result in underreporting, especially among groups like transgender or non-binary individuals who may not feel comfortable disclosing their identity.

Future research should aim to include a broader spectrum of the homeless population, particularly those not using formal shelter services, to gain a more comprehensive understanding of homelessness in Toronto. Moreover, expanding the analysis to explore the long-term effects of factors like the COVID-19 pandemic on the chronic homeless population, as well as housing stability and service utilization across other Canadian cities, could provide deeper insights and inform more effective interventions. In addition to these efforts, government projects should prioritize targeted support for specific populations, including refugees, single adults, and men, who represent a significant portion of long-term homelessness. By focusing on these vulnerable groups, the government could help reduce the strain on shelters, creating more capacity and improving overall system efficiency. Addressing the unique challenges faced by these populations could not only enhance their quality of life but also alleviate pressure on the city's shelter system, allowing for more sustainable solutions.

A Appendix

A.1 Dataset and Graph Sketches

This paper analyzes the homeless population in Toronto using data from the Shelter System Flow between 2018 and 2024, with a focus on identifying the major target groups within the actively homeless population. By utilizing a sample of 604 individuals, the research examines the distribution of key demographic categories, such as ‘single adult,’ ‘non-refugee,’ and ‘male.’ The findings reveal that these categories make up a significant portion of the actively homeless population, with a marked growth trend in the ‘single adult’ and ‘refugee’ groups. This analysis underscores the urgent need for targeted policies and interventions to address the increasing demand for shelter services, particularly for single adults and refugees, in order to reduce homelessness and improve the efficiency of the city’s support systems.

A.2 Data Cleaning

To clean the data, we first load the raw CSV file using the `read.csv` function in R, ensuring that the data is read correctly into a data frame. The next step involves cleaning and standardizing the column names, which may contain dots or unconventional symbols. By using the `rename()` function from the `dplyr` package, we give the columns more descriptive and consistent names, improving readability and facilitating future analysis. After that, we handle the `population_group_percentage` column, which contains percentages in string format (e.g., “100%”). We remove the percentage sign and convert the values to numeric using `gsub()` and `as.numeric()` to make them suitable for analysis.

If the date column is in a string format, we also convert it into a proper date format using `mutate()` and `as.Date()`, which ensures that the dates can be used for plotting time series data. After these steps, we check the structure of the cleaned data using `str()` to verify that all the transformations are correctly applied, ensuring the data is ready for further analysis and visualization.

This process cleans and organizes the data, making it easier to perform insightful analysis and generate meaningful visualizations, such as line plots and bar graphs.

A.3 Attribution Statement

“Contains information licensed under the Open Government Licence – Toronto” (Toronto 1998 - 2024 2024).

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. <https://tellingstorieswithdata.com/>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- . 2024. *Opendatatoronto: Toronto Shelter System Flow*. Toronto Shelter & Support Services. <https://open.toronto.ca/dataset/toronto-shelter-system-flow/>.
- Gibson, Victoria. 2024. *Toronto's Homeless Shelters Have Become More Violent. A New Study Suggests How to Make Them Safer*. Toronto Star. https://www.thestar.com/news/gta/torontos-homeless-shelters-have-become-more-violent-a-new-study-suggests-how-to-make-them/article_34d54824-06f6-11ef-8590-872cd38d749a.html.
- Hadley Wickham, Lionel Henry, Romain François. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Hadley Wickham, ORCID iD, Thomas Lin Pedersen. 2023. *Scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>.
- Kirill Müller, Jennifer Bryan. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes*. <https://CRAN.R-project.org/package=RColorBrewer>.
- News, CDC. 2024. *City Vows to Improve Safety in Homeless Shelters as Report Says Violence on Rise*. CDC News. <https://www.cbc.ca/news/canada/toronto/shelter-safety-study-city-of-toronto-1.7188712>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto 1998 - 2024, City of. 2024. *https://Open.toronto.ca/Open-Data-License/*. <https://open.toronto.ca/open-data-license/>.
- Victor, Fred. 2024. *Facts about Homelessness in Toronto*. Imagine Canada's national Standards Program. <https://www.fredvictor.org/facts-about-homelessness-in-toronto/>.
- 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 Golemund, 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/>.