Navigating Toronto's Homelessness Landscape: A Comprehensive Data Analysis of Trends, Compositional Shifts, and Projections for the City's Homeless Population Dynamics (2018-2023)

Yiyi Feng, Yingxvan Sun

March 15, 2024

Against the backdrop of economic challenges and the ramifications of the pandemic, this analysis offers an intricate exploration of Toronto's shelter system from 2018 to 2023. Unveiling nuanced shifts in the homeless population's composition across single adults, families, genders, and age groups, the study scrutinizes population dynamics. It elucidates fluctuations in the actively homeless count and discerns patterns of return to homelessness. Notably, predictive modeling forecasts a concerning surge in homelessness, underscoring the urgency for proactive strategies. These findings uncover the need for expanded shelter infrastructure to meet the escalating challenges ahead.

Table of contents

1 Introduction [1]

Undoubtedly, homeless individuals in Toronto have been facing serious challenges in accessing nighttime accommodation in recent years. As early as 2020, the general manager of the city's shelter, support, and housing administration openly stated that an average of 72 people were being turned away from shelters due to overcrowding every day, a number that only continued to rise (Gibson 2023). By the winter of 2022, this figure had surged to a staggering 168 people (National 2022). For residents living in Toronto, this reality is apparent - despite shelter occupancy rates exceeding 98% (Group 2021), a significant number of individuals still resort to sleeping on the streets, a reality witnessed firsthand by residents and students alike.

Our understanding of the changing trend in the composition of the homeless community in Toronto is little. While the overall number of homeless individuals is documented, the specific demographic shifts within this population, such as variations in age groups and gender distribution, have not been comprehensively explored. Understanding these trends is essential for developing targeted interventions and resource allocation strategies to address the diverse needs of the homeless population in Toronto more effectively.

To do this, we analyzed Toronto's homeless population, to understand the trends in this community, including changes in overall numbers and demographic composition. Leveraging data obtained from the Toronto Open Database, we aim to find the complexity and evolving trends of Toronto's homeless community. We analyze the status and demographic composition of Toronto's homeless population from 2018 to 2023, focusing on changes in status, such as the number of homeless individuals transitioning between housing and shelters, and long-term use of the shelter system, as well as demographic characteristics including gender, age, and identity.

[1] Please check https://github.com/kqlqkqlqF/Toronto-Shelter-System-Flow-between-2018-and-2023-with-Further-Predictions.git for more information.

A series of structured analyses were conducted. We begin by examining the annual counts of transitions of homeless individuals between housing, shelters, and other statuses within the Toronto shelter system from 2018 to 2023. Following this, we analyzed the composition of homeless population groups and the trend of age distribution within the homeless population. Similarly, the trend of gender composition across different population groups in the homeless community was also analyzed. Finally, we present the predictions from our linear model for future trends in the number of homeless individuals in Toronto.

Our analysis reveals several key findings. Firstly, the linear model we built suggests a steady increase in the number of homeless individuals in Toronto, indicating the need for expansion of the shelter system to accommodate more people. Additionally, our examination of the data through various figures provides insights into the nuanced dynamics of homelessness in the city. These include trends in the annual counts of homeless individuals transitioning within the Toronto shelter system, trends in the composition of homeless population groups and age distribution, and trends in gender composition across different population groups within the shelter system.

Understanding the trends and dynamics of homelessness in Toronto is of paramount importance, given its implications for public policy and social welfare. By finding out the challenges faced by the Toronto shelter system, we aim to inform policymakers and stakeholders in their efforts to address homelessness in the city.

2 Data

2.1 Data Source

The data for this analysis was collected from the shared Toronto Open Database (Shelter 2018-2023). This study utilizes and analyzes the dataset titled "About Toronto Shelter System Flow". The dataset contains lots of information on the transition of homeless individuals in Toronto shelters from January 2018 to January 2024, including gender, age, quantity, and population groups individuals belong to (such as refugees). Additionally, the dataset provides information on the number of people leaving and entering shelter systems. It is released by the Toronto Shelter, Support, and Housing Administration, updated monthly, and has a high level of credibility. Moreover, the dataset has received full marks for freshness, metadata, accessibility, completeness, and usability on the Opendatatoronto website. Therefore, we consider the content of this dataset to be highly credible and utilize it as the primary data source for this paper. However, due to the limited data available for January 2024, we did not include this portion in the data analysis, as the smaller dataset size for 2024 may lead to unexpected analysis results.

2.2 Features

Upon entering shelters, individuals utilizing shelter services are required to provide their name, age, gender, and group affiliation, which are recorded in the database. This dataset only records homeless individuals using overnight shelter services and does not include those utilizing other welfare policies, such as receiving free food or vaccinations. In the data, homeless individuals are divided into five age groups: under 16, 16 to 24, 25 to 44, 45 to 64, and over 65. Gender is categorized as male, female, and transgender/non-binary/two-spirit. The population group includes chronic, refugees, families, youth, single adults, non-refugee, and indigenous. however, the indigenous group was only included in the statistics starting from January 2022, as the authors of the dataset stated their intention to collect more detailed data, hence adding the subdivision. Here, chronic refers to homeless individuals who have continuously used shelter services for more than 180 days. Regarding the documentation of homeless transitions, the dataset provides six subdivisions: newly identified, return from housing, return to shelter, moved to housing, became inactive, and actively homeless. Newly identified refers to people who entered the shelter system for the first time; returned from permanent housing refers to people who previously used the shelter system, then moved to permanent housing, and have now returned; returned to shelter refers to people who were previously using the shelter system, then did not use the system for 3 months or longer, and have now returned; moved to permanent housing refers to people who were using the shelter system and have moved to permanent housing; became Inactive refers to people who last accessed shelter services three months ago; actively homeless refers to people who have used shelter services at least one time in the past three months and have not moved to permanent housing.

2.3 Methodology

The cleaned data were analyzed and performed using R (R Core Team 2022) with `tidyverse` (Wickham et al. 2019), `here` (Müller 2020), `rstanarm` (Goodrich et al. 2020), `ggplot2` (Kassambara 2023), `knitr` (Xie 2023), `tibble` (Müller and Wickham 2022), `stringr` (Wickham 2020), `lubridate` (Grolemund and Wickham 2020), `janitor` (Firke 2023), `testthat` (Wickham and RStudio 2020), and `kableExtra` (Zhu 2021).

Due to the clarity of the data itself, our data cleaning process primarily focused on converting the raw data dates into the yyyy-mm-dd format and selecting the data needed for producing each figure respectively to ensure code was organized and minimize the amount of code in the final QMD file. For the first chart, intended to reveal Trends in the Annual Counts of Homeless Individuals Transitioning within the Toronto Shelter System between 2018 and 2023, we retained the data for the six transition status categories from monthly data and aggregated data within the same year for ease of subsequent chart generation. For the second chart, revealing Trends in the Composition of Homeless Population Groups and Age Distribution within the Toronto Shelter System between 2018 and 2023, we removed all data except for age and population groups. Similarly, for the third chart, Trends in Gender Composition of

Homeless Population Across Different Population Groups within the Toronto Shelter System between 2018 and 2023, we retained only gender and group affiliation data. The final linear model utilized the cleaned dataset for the second figure, and additional cleaning was performed. It is important to note that we excluded data from January 2024 in all chart data to avoid errors.

3 Result

3.1 Trends in Homelessness Transitioning within the Toronto Shelter System

Year	Returned from Housing	Returned to Shelter	Newly Identified	Moved to Housing	Became Inactive	Actively Homeless
2018	776	6055	14442	8135	10420	110052
2019	873	6175	13124	8293	11037	117078
2020	891	5308	7617	6094	9302	99777
2021	989	5024	8297	3409	8690	99580
2022	940	4116	9795	4385	8738	117222

Figure 1: Number of Homeless Individuals Transitioning between Housing, Shelters, and Other Status Within the Toronto Shelter System between 2018 and 2023

The data provided in Figure 1 and Figure 2 offers insight into the dynamic changes in the homeless population within the Toronto shelter system from 2018 to 2023.

A notable trend observed is the fluctuation in the annual count of newly identified individuals entering the shelter system. Peaking in 2018 and gradually declining over the following years, the trend suggests a persistent need for shelter services, with a gradual increase observed from 2020 onwards. Another significant aspect of the data is the return of individuals from permanent housing to the shelter system. This phenomenon is concerning as it indicates potential housing instability for the population. Fortunately, the numbers have remained relatively stable from 2018 to 2023, and they are lower than those transitioning to permanent housing each year. However, the declining trend in individuals moving to stable housing from 2018 to 2021 is worrisome. Although there is a slight increase in 2022 and 2023, it remains below the 2018 levels, suggesting a decrease in individuals transitioning to permanent housing. The data also captures changes in individuals returning to shelters and those becoming inactive within the shelter system. While there is a gradual decline in the number of individuals becoming inactive homeless over the five years, the decrease is minimal, similar to the trend in individuals returning to shelters. Without further explanation, it's challenging to draw reliable conclusions on the decrease in the number of inactive homeless individuals or the

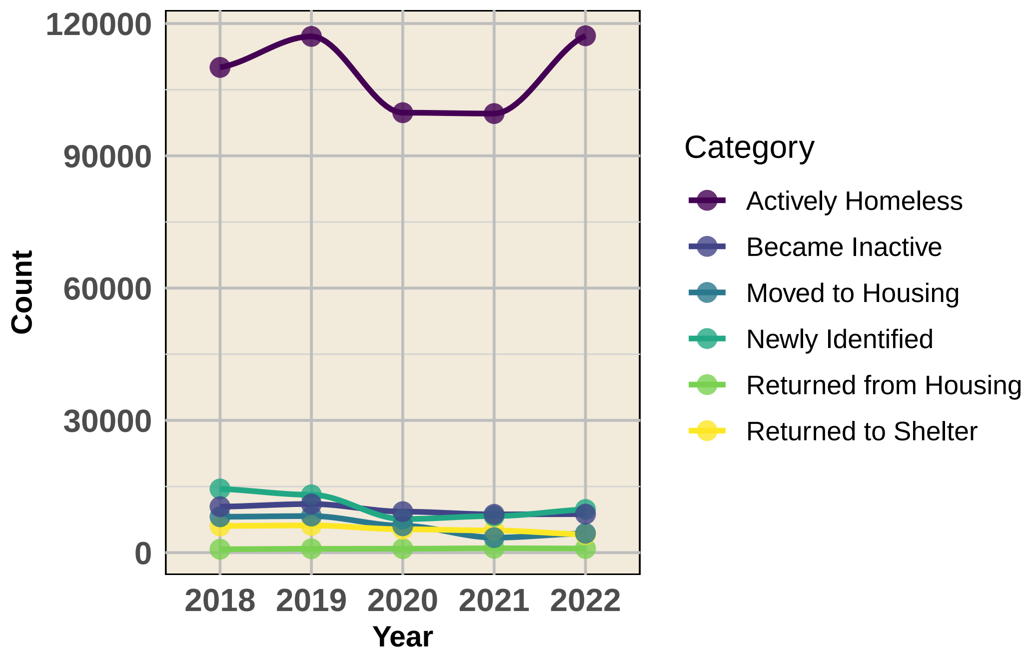

Figure 2: Trends in the Annual Counts of Homeless Individuals Transitioning within the Toronto Shelter System between 2018 and 2023

reasons behind individuals returning to shelters since their experience after leaving the shelter system is unknown. They could be moving to other cities, finding housing, or dying. Therefore, if there's no further detail for the reason they left the shelter system, it is hard to understand the implications of the data fully, and no conclusion can be drawn.

Lastly, individuals from the actively homeless category highlight the ongoing demand for shelter services and support. Although there was a decline in the number of actively homeless individuals in 2020 and 2021, it surged to new highs in 2022 and 2023. This suggests a significant increase in the number of individuals utilizing overnight shelter services in the past two years, indirectly indicating a potential increase in Toronto's homeless population from 2022 to 2023. In conclusion, these data paint a complex picture of homelessness within the Toronto shelter system, with a trend of increasing actively homeless individuals.

3.2 Composition of Homeless Population Groups and Correcponding Age Distribution

Before beginning the analysis of Figure 3, several points need to be explained. Firstly, within the population groups depicted in this graph, apart from refugees and non-refugees, the other options are not mutually exclusive. This means that a person can simultaneously belong to the chronic, families, and refugee groups, but cannot be a refugee and a non-refugee at the same

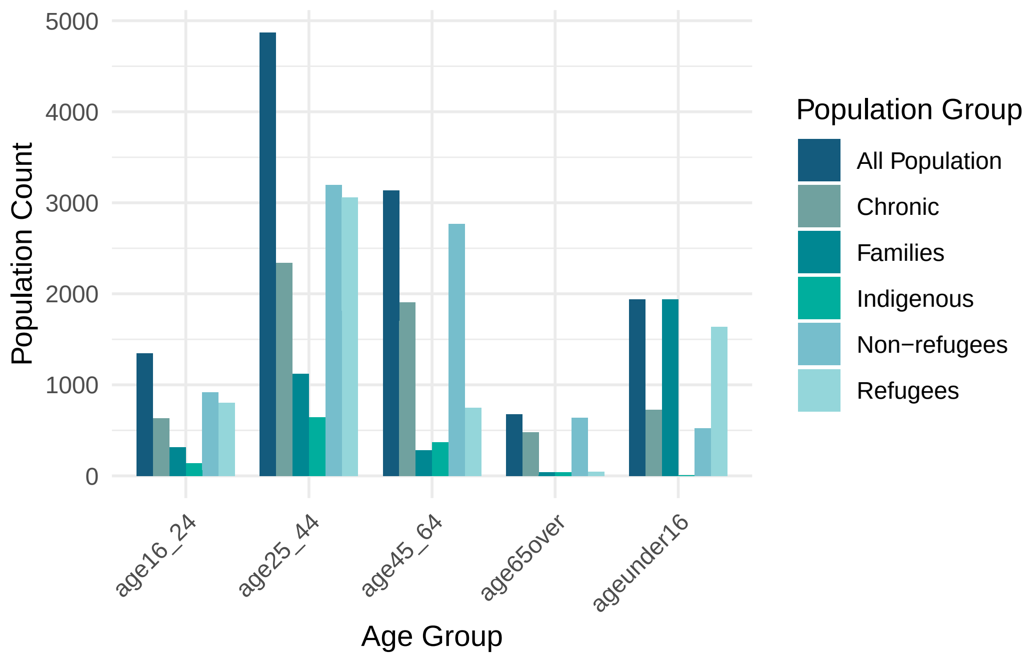

Figure 3: Composition of Homeless Population Groups and Age Distribution within the Toronto Shelter System between 2018 and 2023

time. This explains why the sum of individuals in these population groups exceeds the number in all population categories. Secondly, as mentioned earlier, the number of individuals in the indigenous group was only recorded starting in 2020, so the actual number of indigenous individuals who have used the shelter system should be higher than what is shown in the graph.

Individuals aged between 25 and 44 constitute the largest portion of the homeless population, while those aged between 45 and 64 represent the second largest group, and those aged 65 and above represent the smallest group. From the perspective of population groups, non-refugees constitute the largest group, followed by refugees and then chronic individuals, with indigenous individuals being the smallest group. From this distribution trend, we can infer the following information:

Firstly, except for the age group under 16, chronic individuals constitute a significant portion of the homeless population in other age groups, indicating that a considerable number of homeless individuals rely on overnight shelter services for an extended period and find it difficult to secure permanent housing, thus perpetuating their homelessness.

Secondly, although refugees constitute a smaller proportion of the homeless population compared to non-refugees overall, in the age group under 16, the number of refugees far exceeds the proportion of non-refugees. This suggests that there are a significant number of parenting teenagers or even younger children among the refugee population who are homeless. Additionally, we observe that the proportion of families in the age group under 16 also significantly increases, slightly surpassing the proportion of refugees and almost equaling the "all population" category. This evidence confirms our previous conclusions and supports the notion that these young refugee homeless individuals are likely to be wandering alongside their refugee family members.

Thirdly, unlike refugees, homeless individuals categorized as non-refugees are mainly distributed in the age group of 25 and above. Particularly in the age group of 65 and above, the number of non-refugees is almost equivalent to the "all population" category.

In summary, chronic individuals are prominent across age groups, highlighting challenges in securing permanent housing. Refugees tend to enter the shelter system as a whole family with their young children, indicating a concerning trend of homelessness among refugee families with children, while the non-refugees dominate the older age groups.

3.3 Gender Composition Across Homeless Population Groups

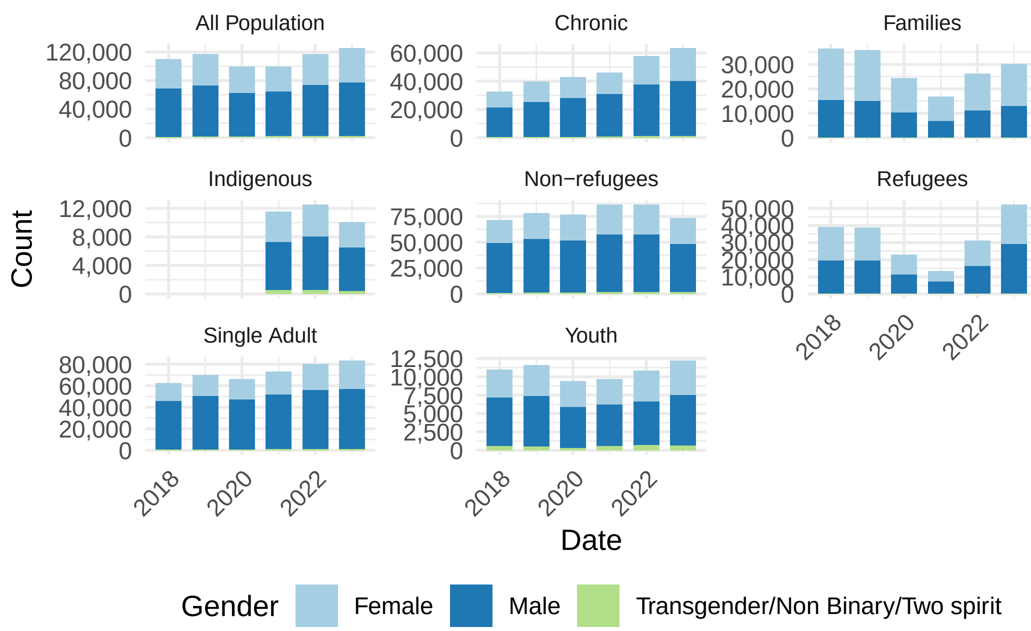

Figure 4: Trends in Gender Composition of Homeless Population Across Different Population Groups within the Toronto Shelter System between 2018 and 2023

The data presented in Figure 4 provides insights into the gender dynamics within Toronto's homeless population across different population groups from 2018 to 2023. A consistent trend observed throughout the years is the higher representation of males compared to females and individuals identifying as transgender, non-binary, or two-spirit. This imbalance suggests a prevalent gender disparity among those experiencing homelessness, with males comprising a larger proportion of the homeless population.

However, within specific population groups, variations in gender composition emerge. For instance, in the chronic population group, which likely encompasses individuals experiencing long-term homelessness, both males and females are prominently represented, indicating a diverse demographic within this subgroup. In contrast, the families population group, which likely consists of homeless families with children, demonstrates a higher count of females compared to males and individuals identifying as transgender, non-binary, or two-spirit. This gender distribution within homeless families may be influenced by factors such as caregiving responsibilities and access to support services, indicating the unique dynamics within this subgroup. The "Refugee" population group exhibits a relatively balanced distribution between males and females, with a smaller count of individuals identifying as transgender, non-binary, or two-spirit. This suggests a diverse composition within refugee populations experiencing homelessness, possibly influenced by factors such as migration patterns and resettlement challenges.

Furthermore, the data reveal temporal fluctuations in the chronic population group, which demonstrates a consistent increase in count from 2018 to 2023. This upward trend suggests a growing number of individuals experiencing long-term homelessness.

3.4 Model and future prediction

3.4.1 Model Setup

The aim of our analysis is to explore the relationship between the number of months passed since January 2018 and the total number of individuals utilizing the overnight shelter system in Toronto. We seek to ascertain if there is a discernible trend in the homeless population over time and quantify the strength of this relationship. To achieve this, we employ a linear regression model utilizing data obtained from the Toronto Homeless Shelter dataset.Let y_i denote the total homeless population for each month i. The number of months passed since January 2018 is represented by x_i. The model is formulated as follows:

$$y_i \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \beta_0 + \beta_1 \times \text{AGE}_i \tag{2}$$
$$\beta_0 \sim \text{Normal}(90, 10) \tag{3}$$
$$\beta_1 \sim \text{Normal}(-0.1, 0.05) \tag{4}$$

In our linear regression model, we assign prior distributions to β_0 and β_1 to represent our initial beliefs about these parameters. For β_0, the intercept, we choose a normal distribution with a mean of 90 and a standard deviation of 10. This means we believe that the most likely value for the intercept is around 90, with values deviating from this being less likely. Similarly, for β_1, the slope, we use a normal distribution with a mean of -0.1 and a standard deviation of 0.05. This reflects our belief that the most probable value for the slope is -0.1, indicating a gradual decline in the homeless population over time. These prior distributions allow us to incorporate our assumptions about the parameters into the model, which can help improve its predictive accuracy.

The variables we are use includes: Number of Month, representing the number of months record the number of months passed starting 2018 January;Total population of homeless, representing the total number of homeless individuals used shelter overnight service in a particular month.

We hypothesize that the coefficient for the number of months passed since January 2018 (β_1) will be positive, indicating an upward trend in the homeless population over time. This hypothesis is based on the assumption that, ceteris paribus, the homeless population tends to increase gradually over time due to various socioeconomic factors.

3.4.2 Model Result

```
               Estimate Std. Error   t value     Pr(>|t|)
(Intercept)  8758.35016 221.645579 39.515113 1.335047e-49
month_number   14.60494   5.277026  2.767646 7.218617e-03
```

The key coefficients and their representation were listed below:

Intercept (β_0) : The intercept represents the estimated number of the homeless population at the beginning of the observation period, which is 2018 January.

Slope (β_1) : The slope quantifies the average change in the number of the homeless population per month.

We choose a linear regression model because we want to examine the linear relationship between month(time change) and the number of the homeless population. We assume that the change in the homeless population over time can be adequately captured by a linear trend.

3.4.3 Model Equation

$$y_i \sim \text{Normal}(\mu_i, \sigma) \tag{5}$$
$$\mu_i = \beta_0 + \beta_1 \times \text{AGE}_i \tag{6}$$
$$\beta_0 = 8758.35 \tag{7}$$
$$\beta_1 = 14.605 \tag{8}$$

3.4.4 Residuals

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1826.5  -655.6   226.4     0.0   712.4  1185.1
```

1. Minimum Residual: The minimum residual is -1826.5, indicating that there is at least one observation for which the model's predicted total population was lower than the actual population by approximately 1826.5 individuals. This suggests the presence of potential outliers or factors not accounted for by the model.

2. First Quartile Residual: The first quartile (Q1) of the residuals is -655.6, implying that 25% of the observations have total populations that are less than 655.6 individuals below the model's prediction.

3. Median Residual: The median residual is 226.4, indicating that half of the observations have residuals below 226.4, suggesting a slight positive bias in the model's predictions. This means that, on average, the model tends to slightly underestimate the total population.

4. Third Quartile Residual: The third quartile (Q3) is 712.4, signifying that 75% of the observations have total populations within 712.4 individuals of the model's predictions or better. This suggests that the model provides reasonably accurate predictions for the majority of observations.

5. Maximum Residual: The maximum residual is 1185.1, suggesting that there is at least one observation for which the model's predicted total population was higher than the actual population by approximately 1185.1 individuals. Similar to the minimum residual, this could indicate the presence of outliers or unaccounted factors.

Interpretation: The distribution of residuals provides insights into the model's performance and the presence of potential outliers or discrepancies between the model's predictions and the actual observations. Further investigation into the reasons behind significant deviations can help refine the model and improve the accuracy of its predictions. Additionally, assessing the model's assumptions, such as linearity, homoscedasticity, and normality of residuals, is essential to ensure the validity of the statistical inferences drawn from the model.

3.4.5 Graphing the Model

3.4.6 Conclusion

The scatter plot below visualizes the relationship between the number of months passed since January 2018 and the total number of individuals utilizing the overnight shelter service in Toronto. The blue dashed line represents the linear regression line fitted to the data, capturing the average trend in the homeless population over time.

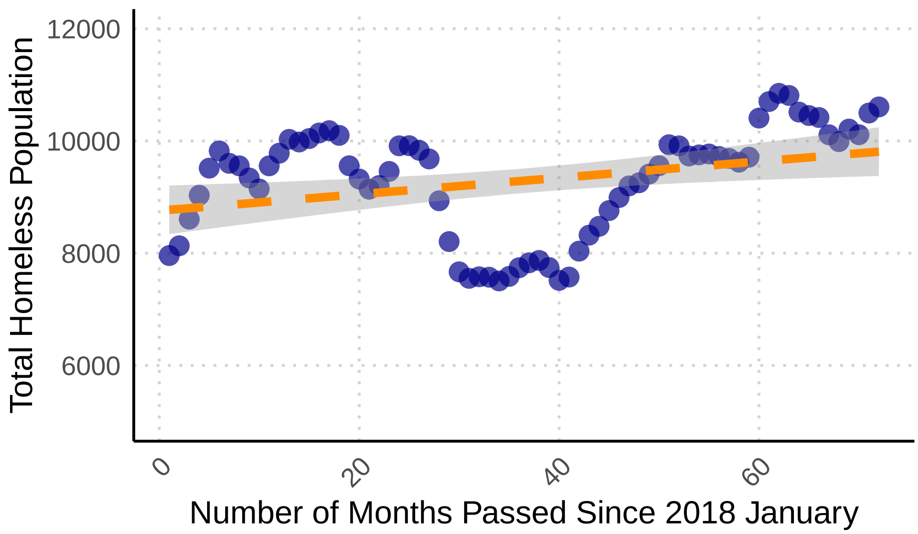

Figure 5: The Linear Relationship between Number of Month Passed and the Number of individuals use Shelter Overnight Service

The linear regression model estimates the relationship between the number of months passed since January 2018 and the total number of individuals utilizing the overnight shelter system in Toronto. The model is expressed as ($y_i \sim \text{Normal}(\mu_i, \sigma)$), where ($y_i$) represents the predicted total homeless population in month (i), (μ_i) is the mean of the normal distribution for the (i)th observation, and (σ) is the standard deviation. The mean (μ_i) is determined by the intercept (β_0) and the slope (β_1) multiplied by the number of months passed since January 2018.

The estimation results indicate that the intercept (β_0) is 8758.35, representing the estimated total homeless population in January 2018. The slope (β_1) is 14.605, indicating that for each additional month since January 2018, the total homeless population is estimated to increase by approximately 14.605 individuals. The significance of both coefficients is evidenced by their small standard errors and highly significant p-values ($p < 0.001$), indicating a strong statistical assurance in the estimates. However, the adjusted R^2 value of 0.08576 suggests that only approximately 8.6% of the variability in the total homeless population can be explained by the number of months passed since January 2018. Despite the statistically significant relationship between the independent and dependent variables, the model's overall explanatory power is relatively modest. Further analysis and consideration of additional factors may be necessary to improve the model's predictive accuracy and better understand the dynamics of homelessness in Toronto over time.

4 Discussion

4.1 Weaknesses

4.2 Further Study

bibliography

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gibson, Victoria. 2023. "Why the Homelessness Crisis Could Get Even Worse." 2023. https://www.thestar.com/news/gta/why-the-homelessness-crisis-could-get-even-worse/article_9054cdcc-f846-5a33-a9e4-f972860686a1.html.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2020. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Grolemund, Garrett, and Hadley Wickham. 2020. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

Group, Toronto Street Needs. 2021. "Toronto 2021 Street Needs Assessment Result," 11–18. https://www.toronto.ca/legdocs/mmis/2021/ec/bgrd/backgroundfile-171729.pdf.

Kassambara, Alboukadel. 2023. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2022. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

National, CBC News. 2022. "Toronto Shelters Turning Away Average of 168 People a Day." 2022. https://www.youtube.com/watch?v=vUZhZIS-XRo&t=2s.

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, Support & Housing Administration. 2018-2023. *About Toronto Shelter System Flow*. https://open.toronto.ca/dataset/toronto-shelter-system-flow/.

Wickham, Hadley. 2020. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr.

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, and RStudio. 2020. *Testthat: Unit Testing for r*. https://CRAN.R-project.org/package=testthat.

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

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.