Exploring the Dynamics of Homeless Deaths in Toronto*

Yingqi Pang

24 January 2024

Abstract

In the urban landscape of Toronto, homelessness remains a poignant and complex issue, especially in the context of mortality. This study delves into the intricate patterns of homeless deaths, utilizing comprehensive data analysis to uncover underlying trends and contributing factors. By integrating statistical methods and data from Toronto's public records, our research highlights significant temporal and demographic patterns in these deaths. The findings not only illuminate the severity of the situation but also provide critical insights for policymakers and social services. This paper not only aims to enhance understanding of the homeless mortality crisis but also to inform strategies for mitigation and support for this population.

Contents

*Code and data are available at: https://github.com/pangyin2/Homeless-Deaths-in-Toronto.git

1 Introduction

Homelessness is a serious challenge in Canada, drawing attention to the harsh realities faced by the most vulnerable groups among us. On the bustling streets, amongst towering skyscrapers and thriving businesses, the plight of the homeless is often overlooked, their struggles overshadowed by the vibrant rhythm of the city (Hwang 2000). In Toronto, every night there are thousands of people without a home. This includes those living in shelters and temporary housing, as well as those living outdoors or in other locations. The impact is complex, enduring, and for many, devastating. Homelessness affects some of the most vulnerable residents in our community and leads to persistent health inequalities. Homeless individuals are at high risk of illness and have a higher mortality rate than the general population. In Toronto, the mortality rate among homeless shelter-using males is 8.3 times higher for those aged 18-24, 3.7 times higher for those aged 25-44, and 2.3 times higher for those aged 45-64 compared to the general male population in those age groups. The likelihood of death for homeless women aged 18-44 is 10 times higher than for the average Toronto woman (Hwang 2001). Early studies have documented the heavy burden of disease borne by the homeless due to mental illness and addiction, medical conditions, tuberculosis, HIV infection, and trauma. These illnesses, coupled with severe poverty and often insufficient access to healthcare, result in a high mortality rate among the homeless.

In this project, I will examine the data on homeless deaths in Toronto from 2017 to 2023, which is extensively described in the Data Section. In my analysis, I will explore the number of homeless deaths categorized by cause, the distribution of homeless deaths across different age groups, and the annual trends in homeless deaths. The analysis will be conducted in R (R Core Team 2020) using the tidyverse (Wickham et al. 2019),readr (Wickham, Hester, and Bryan 2024), dplyr (Wickham et al. 2021) and opendatatoronto (Gelfand 2020) packages. All figures in the report are generated using ggplot2 (Wickham 2016), and tables are created using kableExtra (Zhu 2021).

2 Data

In this data section, I'll describe the data collection and processing methodology and dive into the content of the data.

2.1 Data Collection

All the data used are sourced from the City of Toronto Open Data Portal, titled "Homeless Deaths by Cause." Using the R package opendatatoronto (Gelfand 2020), we loaded the data into an R script titled "Data Collection and Processing." This data is uploaded and funded by Toronto Public Health and is updated biannually, with the last update being on September 29, 2023, up to February 6, 2022. Starting from January 2017, Toronto Public Health (TPH) began tracking the deaths of homeless individuals to more accurately estimate the number of deaths and their causes. TPH is responsible for the data collection, analysis, and reporting. The Shelter, Support, and Housing Administration (SSHA) and health and social service agencies that support the homeless share information about deaths with TPH, and some of the data is verified by the Office of the Chief Coroner of Ontario (OCCO). For this data collection initiative, homelessness is defined as "a situation where an individual or family does not have stable, permanent, appropriate housing, or the immediate prospect, means, and ability to acquire it." The data includes 4 variables: Year of death, Cause of death, Age group, Gender, Count.

Table 1: Homeless Deaths in Toronto

_id	Year of death	Cause_of_death	Age_group	Gender	Count
1	2017	Accident	40-59	Male	2
2	2017	Accident	60+	Male	3
3	2017	Cancer	60+	Female	1
4	2017	Cancer	40-59	Female	2
5	2017	Cancer	40-59	Male	2
6	2017	Cancer	60+	Male	4

2.2 Variables of interest

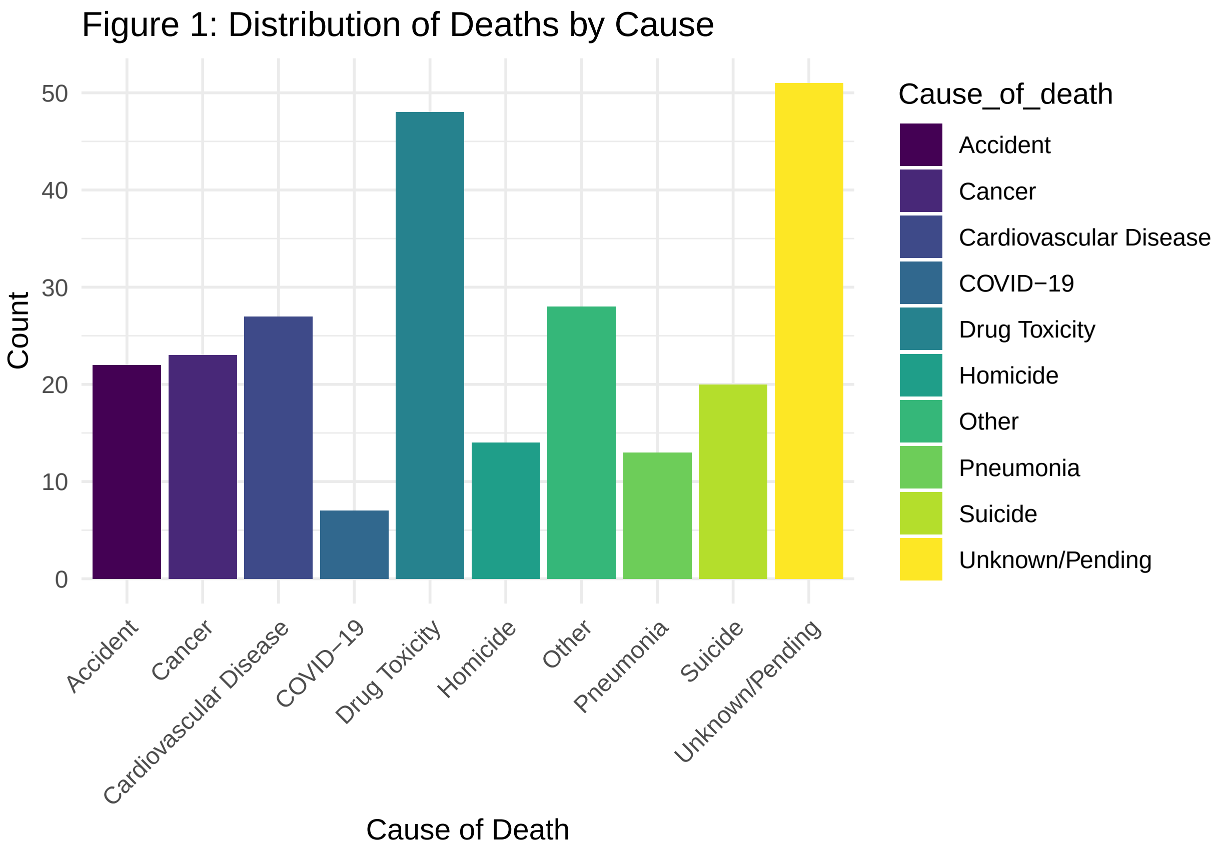
The Cause of Death is divided into 10 categories: Accident, Cancer, Cardiovascular Disease, COVID-19, Drug Toxicity, Homicide, Pneumonia, Suicide, Other, and Unknown/Pending. The data only reflects the deaths reported to TPH by SSHA, community partners, and the coroner's office. About 25% of the reported death cases have an unknown or pending cause of death. To protect privacy, less than 2% of the causes of death are categorized as 'Other'. The age groups in the dataset are divided as follows: under 20, 20 to 39, 40 to 59, over 60, and unknown. Gender is divided into Female, Male, and Unknown. The data does not determine the status of indigenous people, as in 70% of the reported cases, indigenous status is reported as unknown or missing.

2.3 Data Processing

In the data preparation stage of our R script, we start by checking for missing values within the homeless_deaths dataframe. We use the sum(is.na(homeless_deaths)) function call to tally all the NA (not available) entries across the entire dataset. If any missing values are found, we remove the corresponding rows with the drop_na() function from the dplyr package to ensure that our subsequent analysis is not skewed by incomplete data. Following this, we need to ensure that the columns representing categorical data are correctly recognized as factors, which is important for certain types of analysis that require categorical variables. We use the as.factor() function to convert the 'Year of death', 'Gender', 'Age group', and 'Cause of death' columns to factors. Finally, to verify our data cleaning steps, we use the head(homeless_deaths) function call to display the first few rows of the now cleaned dataset. This provides a quick glimpse of the dataset structure and the initial entries, ensuring that our dataset is ready for analysis.

3 Visualizing the Data and The Implications

3.1 Visual 1



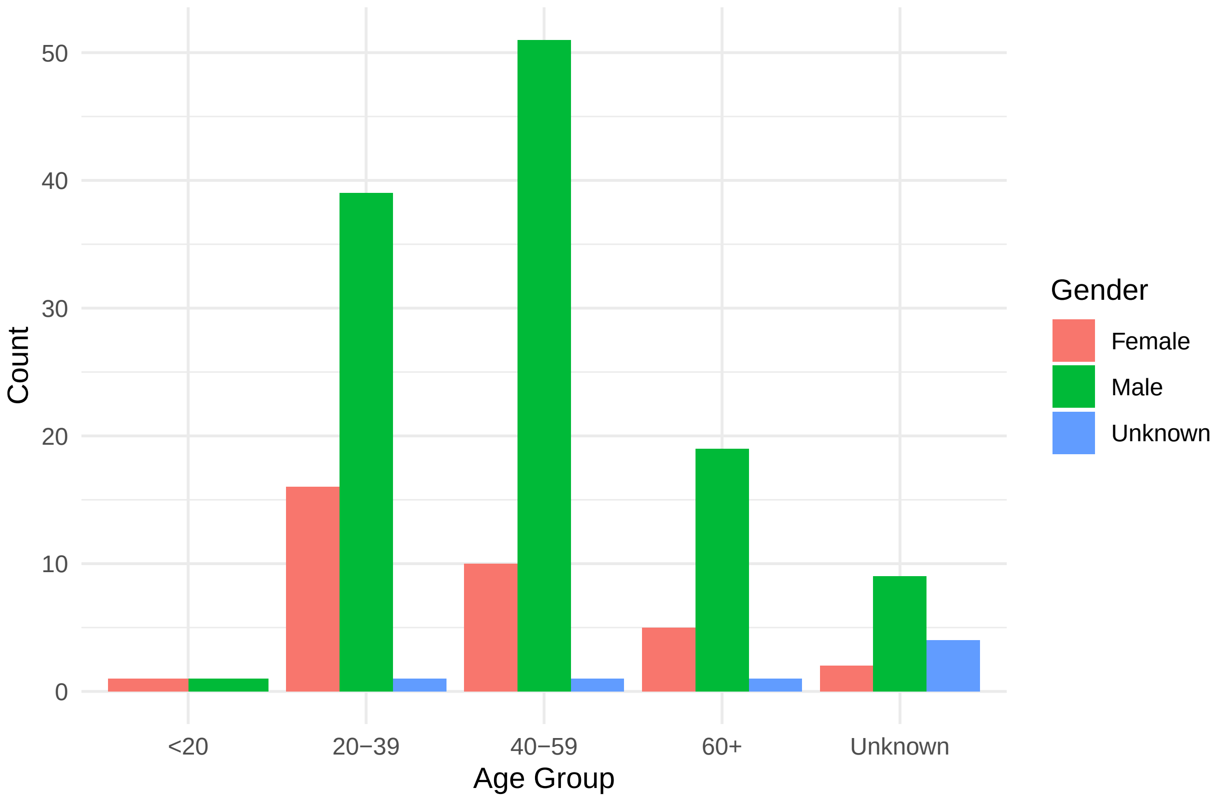
This bar plot provides us with a stark visualization of the tragic endpoints faced by Toronto's homeless population. Each bar, colored distinctly to represent a different cause of death, stands as a reminder of the varied threats to life that disproportionately affect those without a home. The height of each bar not only quantifies these threats but also symbolizes the scale of intervention needed to address them.

Upon examining the graph, it becomes evident that certain causes of death are particularly prevalent within the homeless community. For instance, a noticeably higher bar for causes like Drug Toxicity or Cardiovascular Disease suggests that these are significant contributors to mortality and thus warrant specific attention from healthcare providers and policymakers. The visualization implores us to consider the underlying factors: Is it the lack of access to timely medical intervention, the rigors of street life, or the psychological burden of homelessness that exacerbates these conditions? It also raises questions about the adequacy of harm reduction programs and the availability of emergency medical services tailored for the homeless.

Contrastingly, the data shows that the least number of homeless deaths were attributed to COVID-19. This finding could potentially reflect the success of public health interventions or the underreporting of COVID-19 deaths due to challenges in post-mortem viral testing. It may also bring attention to the effectiveness of rapid response measures taken to protect this vulnerable group during the pandemic. However, it should be interpreted with caution due to the overall uncertainty that accompanies healthcare provision for the homeless. Furthermore, this analysis must not end with mere observation. It calls for a response—a societal and systemic push towards housing stability, improved healthcare access, and targeted support for mental health and addiction. The data does not merely reflect the past but should inform our actions to shape a more equitable future.

3.2 Visual 2

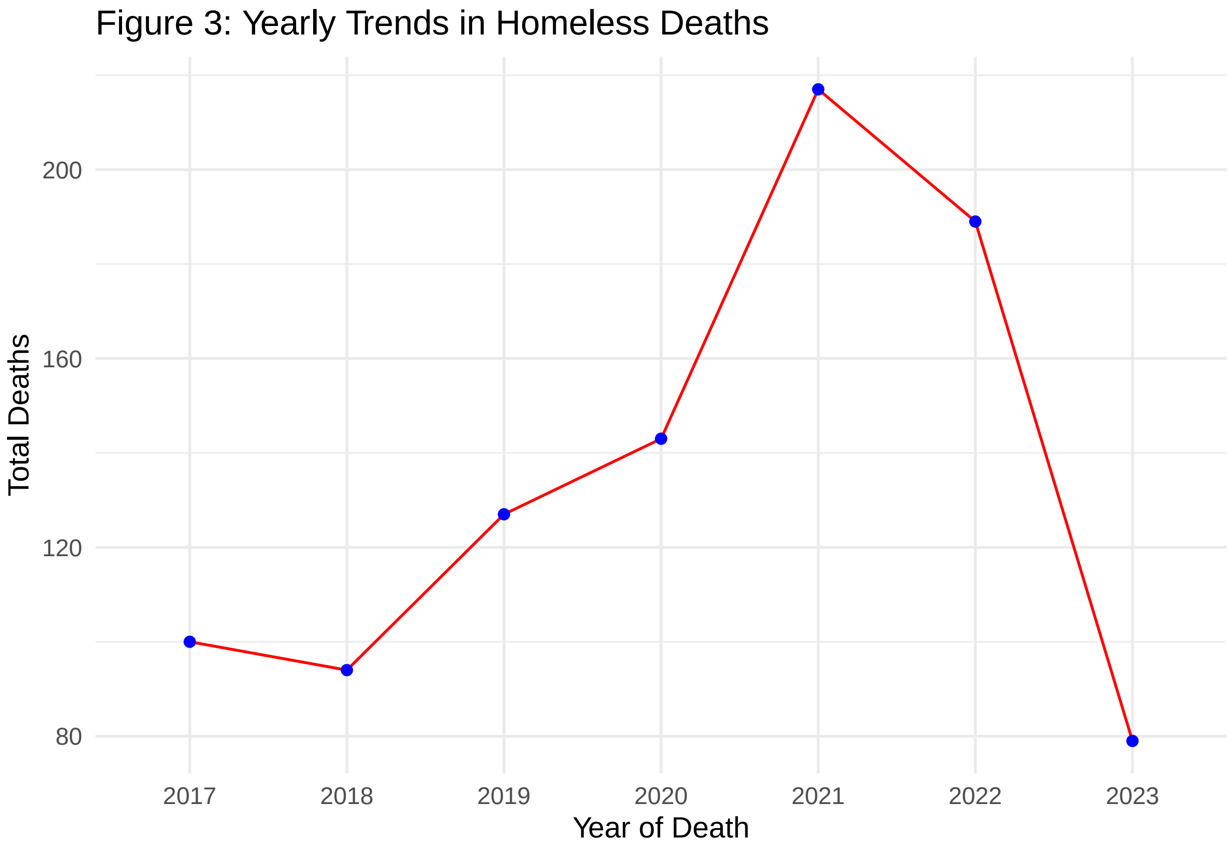
Figure 2: Deaths by Age Group and Gender



This plot illustrates the distribution of homeless deaths across different age groups, further categorized by gender (Female, Male, Unknown). The age groups are less than 20 years old, 20 to 39 years old, 40 to 59 years old, greater than 60 years old, and unknown. The x-axis represents the age groups, the y-axis shows the count of deaths, and the color coding differentiates between genders.

The plot highlights the prevalence of homeless deaths across age and gender lines, with a stark contrast between male and female mortality rates. Notably, the mortality rate of male is obviously higher than female. Males, particularly in the 40 to 59 age group, show the highest death counts, suggesting acute vulnerabilities. Females tend to die at a younger age, potentially due to gender-specific health issues and risks such as gynecological diseases or sexual violence. This underscores the need for gender-sensitive and age-specific interventions in addressing homelessness and associated mortality.

3.3 Visual 3



This line plot shows the total number of homeless deaths per year. The x-axis represents the years (from 2017 to 2023), and the y-axis indicates the total count of deaths for each year. The red line, marked with circular points, illustrates the trend over the years. The overall trend from 2017 to 2021 is upward that indicates an increase in mortality. It suggests the worsening conditions or increased vulnerabilities within the homeless population. The peak in 2021 marks the highest mortality, but the subsequent sharp decrease in 2022 signals a significant shift. This sudden decline may reflect successful policy implementation, altered data gathering methods, or other influential variables. This plot is crucial for assessing the effects of policy and health care initiatives aimed at the homeless population over these years.

4 Limitation

The dataset used in this study, while comprehensive, has several limitations that must be acknowledged. Firstly, it only reflects the instances of homeless deaths reported to Toronto Public Health (TPH) by the Shelter, Support, and Housing Administration (SSHA), community partners, and the coroner's office, which may not represent all such deaths in the city. A significant portion, about 25%, of the reported cases have indeterminate or pending causes of death, which challenges the completeness of the dataset. Additionally, since 2019, TPH ceased reporting the locations of deaths, often unknown or unverified, which limits spatial analysis.

A considerable gap in the data is the underrepresentation of indigenous status in reported cases, with 70% of cases lacking this information. This high rate of missing information requires cautious interpretation of the data, and standard epidemiological practice requires prohibiting data publication in such cases. Furthermore, causes of death among transgender individuals are not shown due to small numbers, which could affect the understanding of the issue.

Lastly, the dataset is subject to change if TPH receives additional reports or more detailed information on existing reports. Reporting delays could alter previously published data, and the most recently released data should be considered the most complete. These limitations highlight the need for improved data collection and reporting to better understand and address the factors leading to homeless deaths in Toronto.

5 Conclusion

This paper provided a comprehensive analysis of the patterns and contributing factors to homeless deaths in Toronto through meticulous data analysis. The temporal and demographic trends illuminated by this study bring to light the severe implications of homelessness and its impact on mortality. The study's findings underscore the necessity for targeted policy interventions and enhanced support systems to address the crisis of homeless mortality. By shedding light on the acute needs of this vulnerable population, this paper serves as a foundational call to action for policymakers, healthcare providers, and social services to collaborate on creating sustainable solutions for housing stability and healthcare accessibility.

6 Reference

6.1 Generative AI Statement

I used the following generative artificial intelligence (AI) tool: ChatGPT (OpenAI 2023). I used the tool only in the graph part. I presented ChatGPT with the query "Please suggest a good title for a study on homeless deaths in Toronto." Based on the input, the AI generated several title options, reflecting a nuanced understanding of the subject matter. I selected the most fitting title from these suggestions for inclusion in my paper.

Bibliography

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

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

———. 2001. "Homelessness and Health." *Cmaj* 164 (2): 229–33.

OpenAI. 2023. "ChatGPT: Generative Pre-Trained Transformer for Conversational Agents." https://openai.com/.

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. 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.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.