

# Analyzing Homeless Deaths\*

## Understanding the Patterns and Implications of Homeless Deaths Over Time

Yingqi Pang

23 January 2024

### Abstract

Homelessness is a critical issue affecting communities worldwide. This paper delves into the analysis of homeless deaths, aiming to identify patterns and draw insights into the factors contributing to this tragic phenomenon. By examining available data and employing statistical tools, we explore the temporal trends, demographics, and potential implications of homeless deaths. The study aims to shed light on the gravity of the situation and proposes considerations for addressing this pressing issue.

## Contents

<b>Introduction</b>	<b>1</b>
<b>Data</b>	<b>3</b>
Data Collection . . . . .	3
Variables of interest . . . . .	3
Population, Frame or Sample . . . . .	3
<b>Visualizing the Data and The Implications</b>	<b>4</b>
Visual 1 . . . . .	4
<b>Reference</b>	<b>7</b>

## 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. 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. 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 @ref(data). In my analysis, I will explore the number of homeless deaths

---

\*Code and data are available at: <https://github.com/pangyin2/Term-Paper-1.git>

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 2021) using the dplyr (Wickham et al. 2021), knitr (Xie 2021b), tidyverse (Wickham et al. 2019), lubridate (Grolemund and Wickham 2011), janitor (Firke 2023), readr (Wickham, Hester, and Bryan 2022), tidyr (Wickham, Vaughan, and Girlich 2023), and opendatatoronto (Gelfand 2022) packages. All figures in the report are generated using ggplot2 (Wickham 2016), and tables are created using kableExtra (Zhu 2021).

## Data

In this data section @ref(data), I'll describe the data collection and processing methodology and dive into the content of the data. First, (Table @ref(fig:)) gives us a glimpse of the data.

### 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` [citeODT], 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.

### 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.

### Population, Frame or Sample

```
## [1] 0

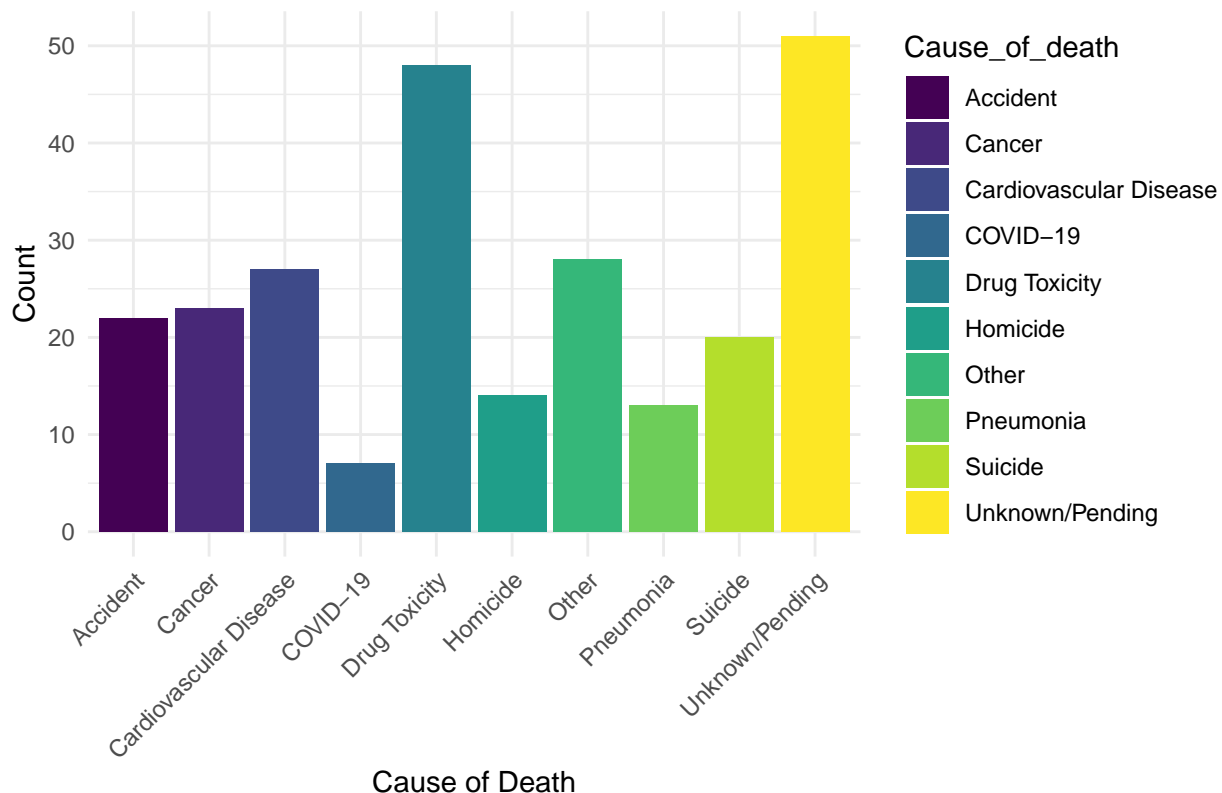
## # A tibble: 6 x 6
##   `_id` `Year of death` Cause_of_death Age_group Gender Count
##   <dbl> <fct>           <fct>         <fct>    <fct>   <dbl>
## 1     1   2017           Accident      40-59    Male     2
## 2     2   2017           Accident      60+     Male     3
## 3     3   2017           Cancer        60+     Female   1
## 4     4   2017           Cancer      40-59    Female   2
## 5     5   2017           Cancer      40-59    Male     2
## 6     6   2017           Cancer      60+     Male     4
```

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.

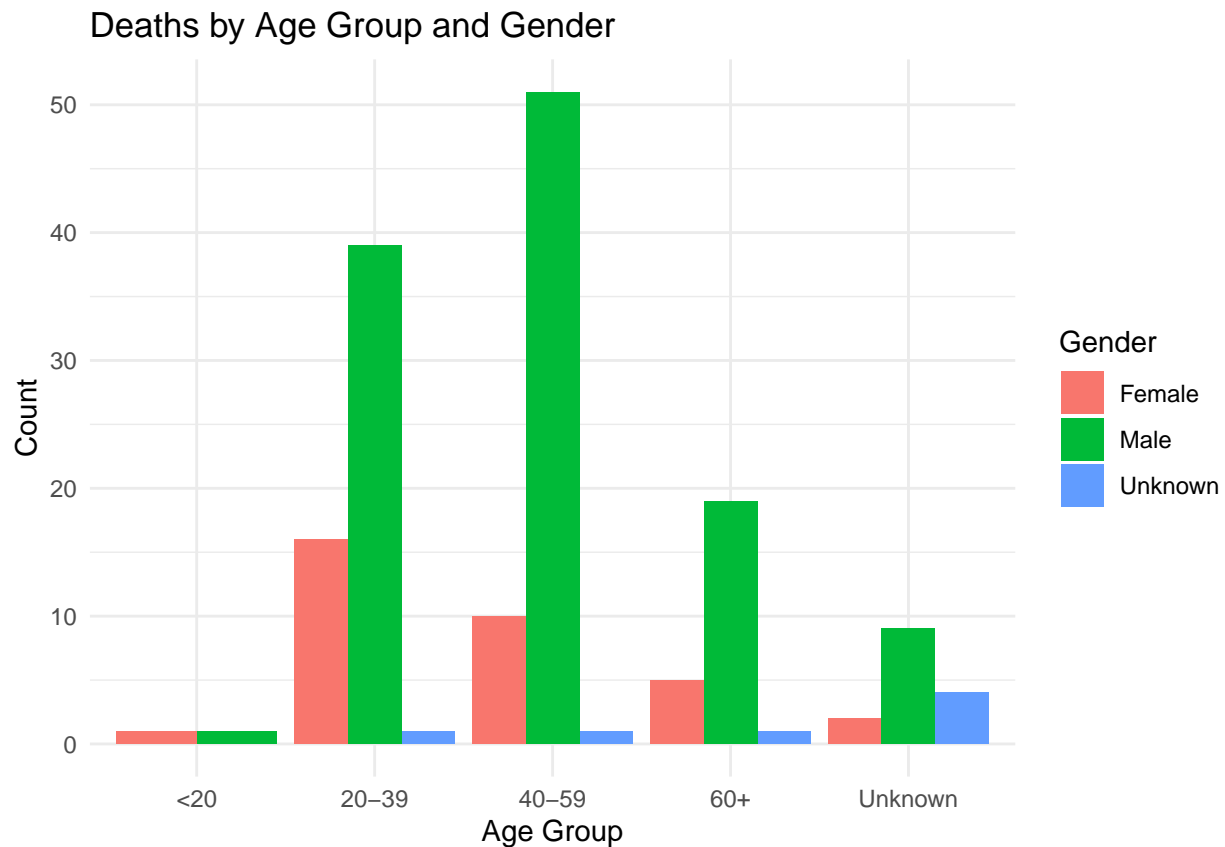
## Visualizing the Data and The Implications

Visual 1

Distribution of Deaths by Cause

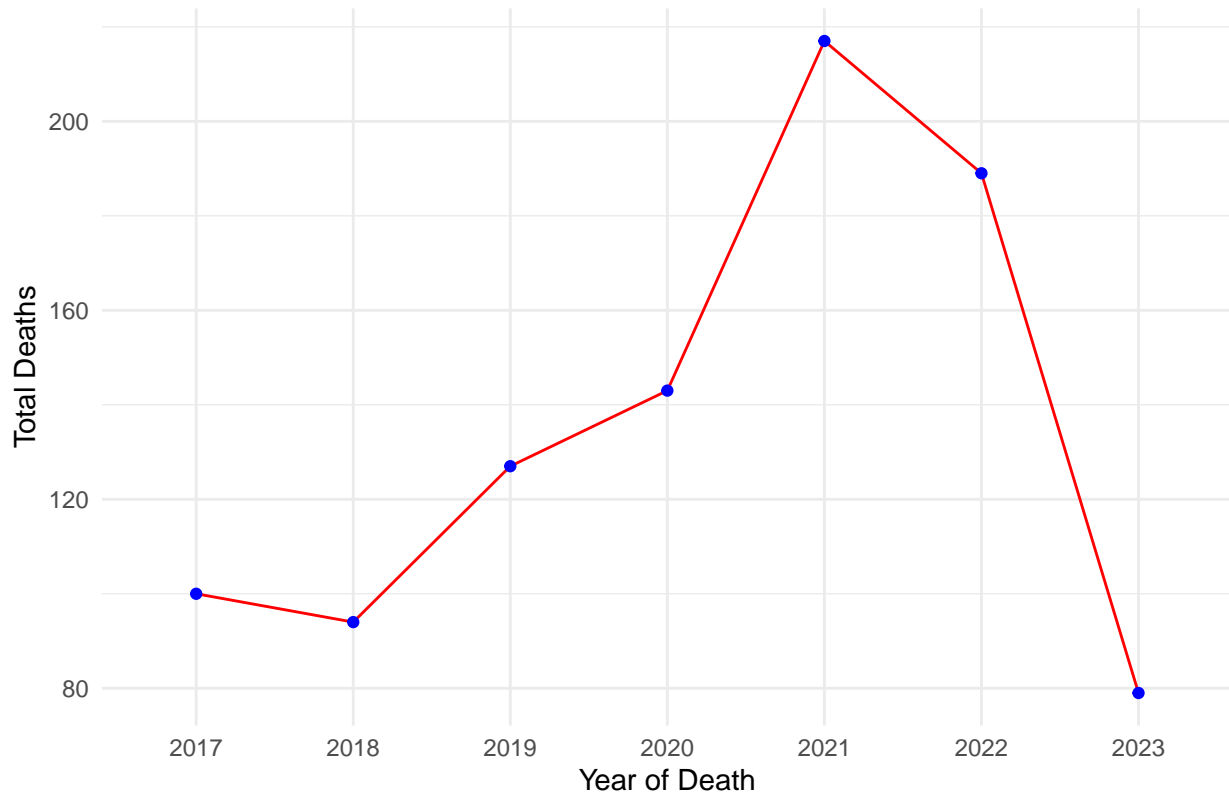


This bar plot provides us with a stark visualization of the tragic endpoints faced by Toronto's homeless population. Each bar, colored distinctly to represent a different cause of death, stands as a reminder of the varied threats to life that disproportionately affect those without a home. The height 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.



This plot illustrates the distribution of homeless deaths across different age groups, further categorized by gender. The x-axis represents the age groups, the y-axis shows the count of deaths, and the color coding differentiates between genders. Observations: This plot is useful for identifying any trends or disparities in death counts among different age groups and between genders. For instance, if one age group has a significantly higher count, it indicates a higher vulnerability within that group. Similarly, gender-based disparities can also be identified.

Yearly Trends in Homeless Deaths



This line plot shows the total number of homeless deaths per year. The x-axis represents the years, 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. Observations: This plot is particularly useful for identifying any year-over-year changes or trends in the data. An upward trend would indicate an increasing problem, while a downward trend could suggest improvements in the situation. It also helps in understanding the impact of any specific events or policy changes over time on the homeless population's mortality rate.

## Reference