

Rising Trends in Shelter Resident Deaths in Toronto Over the Past 17 Years*

Haowei Fan

September 23, 2024

Shelter resident deaths are an important indicator of social stability, reflecting systemic issues such as welfare systems, housing crises, and healthcare access. This project analyzed monthly shelter resident deaths in the Toronto region from March 2007 to August 2024, focusing on trends over time and seasonal variations. The study found that deaths have been increasing, with a significant rise between 2020 and 2023. While the number of deaths remained similar in spring, summer, and fall, winter months saw a significantly higher number of fatalities. These findings provide valuable data to support targeted interventions by governments and welfare organizations to address changing needs throughout the year.

1 Introduction

In 2023, as the COVID-19 pandemic receded, the Toronto government announced plans to phase out the shelter hotels that had been providing temporary accommodations for homeless individuals. By the end of 2024, 24 shelter hotels will no longer be available due to expiring contracts, exacerbating an already overcrowded shelter system. The loss of these temporary accommodations poses significant risks not only to the safety and stability of the community but also to the well-being of those who rely on shelters for survival. The influx of new shelter residents may potentially lead to a decline in living conditions, including sanitation issues and territorial conflicts. This highlights the urgent need to understand the trends and underlying factors affecting shelter residents, especially in terms of their mortality numbers. While many sociological studies have explored the growing number of shelter residents, fewer have examined shelter resident mortality rates in detail. Given the challenges of accurately counting the transient and often untracked shelter population, studying actual death counts offers a more reliable and feasible method of understanding the impact of systemic issues. Mortality data, being both concrete and verifiable, can better inform policy decisions than estimates of shelter

*Data are available at: <https://open.toronto.ca/dataset/deaths-of-shelter-residents/>

population numbers, which are often difficult to obtain and track. This paper focuses on analyzing the number of deaths among shelter residents in Toronto from March 2007 to August 2024. The aim is to identify both long-term trends and seasonal variations in shelter resident deaths. Using simple linear regression, this study estimates the rate at which the number of deaths has increased over time. It also examines anomalies in the data by comparing yearly death rates, and assesses seasonal differences by analyzing average death rates across the four seasons. This analysis provides valuable insights into the hardships faced by shelter residents, especially during specific times of the year. This study finds three key trends: 1) the number of deaths among shelter residents has been rising over time; 2) there was a noticeable increase in deaths between 2020 and 2023; and 3) while deaths during spring, summer, and fall remained relatively consistent, winter saw significantly higher mortality rates. Understanding these trends is critical because shelter resident deaths serve as a key indicator of broader social problems, such as gaps in welfare systems, housing shortages, and insufficient healthcare services. These findings provide essential data for governments and welfare organizations to develop more targeted interventions based on the needs of shelter residents throughout the year. The remainder of this paper is structured as follows: Section 2 discusses the data sources used in this project and details the data cleaning process, followed by a visualization of the dataset’s key distribution and statistics. Section 3 outlines the statistical methods used and presents the findings in detail. Section 4 addresses the limitations of the study and offers suggestions for future research.

2 Data

2.1 Raw Data

The dataset I am using is titled “Deaths of Shelter Residents”, sourced from OpenToronto. This dataset records the number of deaths among shelter residents in the Toronto area, starting from January 2007 on a monthly basis. It also includes the gender of the deceased, categorized into three sections: male, female, and other gender, which includes Transgender, Non-binary, and Two-Spirit. The dataset is continuously updated. Table 1 contains descriptions of these variables.

Table 1: Column descriptions of the original data

Column	Description
<code>_id</code>	Unique row identifier for Open Data database.
<code>Year</code>	The calendar year being reported on.
<code>Month</code>	The month being reported on.
<code>Total decedents</code>	The total number of shelter residents who died in the reported month/year.

Column	Description
Male	The total number of male shelter residents who died in the reported month/year.
Female	The total number of female shelter residents who died in the reported month/year.
Transgender/Non-binary/Two-Spirit	The total number of transgender, non-binary, and Two-Spirit shelter residents who died in the reported month/year.

In addition to this dataset, OpenToronto also offers a similar dataset called “Deaths of People Experiencing Homelessness”, which includes four variables: id, year, month, and number of deaths. However, I chose not to use this dataset due to its limited data coverage, spanning only from May 2007 to August 2023. In comparison, the “Deaths of Shelter Residents” dataset provides more extensive data, allowing for greater accuracy and a narrower confidence interval in the results.

2.2 Data Cleaning

In the data cleaning process, the following steps were conducted using Rcode(R Core Team (2023)), along with the packages pandas, tidyverse, ggplot2, and dplyr:

Combining Year and Month: The variables for year and month were merged, and the month was converted to a numerical format to facilitate time series analysis. Creating a New Variable for Seasons: A new variable was created to categorize each month into one of the four seasons: March, April, and May were classified as spring; June, July, and August as summer; September, October, and November as autumn; and December, January, and February as winter. This was done to enable seasonal analysis. Removing Unnecessary Variables: All variables unrelated to this project, such as id and the three gender categories, were removed to simplify the dataset. Ensuring Equal Seasonal Data: To ensure that each season had the same length of time for comparison, I removed some samples from the beginning and end of the dataset, so that the total number of samples became a multiple of four, with each season covering the same period. Handling Missing Data: After inspecting all the samples, I confirmed that there were no missing values, so no further treatment for missing data was necessary. After the cleaning process, the dataset was reduced to three variables: Time, Total Deaths, and Season, as shown in Table 2.

Table 2: Sample of the cleaned data

year_month	season	total_decedents
2007-03	Spring	3
2007-04	Spring	1
2007-05	Spring	2

year_month	season	total_decedents
2007-06	Summer	3
2007-07	Summer	2
2007-08	Summer	2

2.3 Statistical summary of the cleaned data

Table 3 provides some basic statistics about the dataset. The project spans a total of 204 months. During this period, the highest number of deaths recorded in a single month was 19, while the lowest was 0. The average number of deaths per month is 3.848.

Table 3: Basic statistics of the cleaned data

year_month	total_decedents
Length:204	Min. : 0.000
Class :character	1st Qu.: 1.000
Mode :character	Median : 3.000
NA	Mean : 3.848
NA	3rd Qu.: 5.000
NA	Max. :19.000

Figure 1 illustrates the variation in the number of deaths per month over time. It is clear that the trend is upward, indicating an increase in the number of deaths over the observed period.

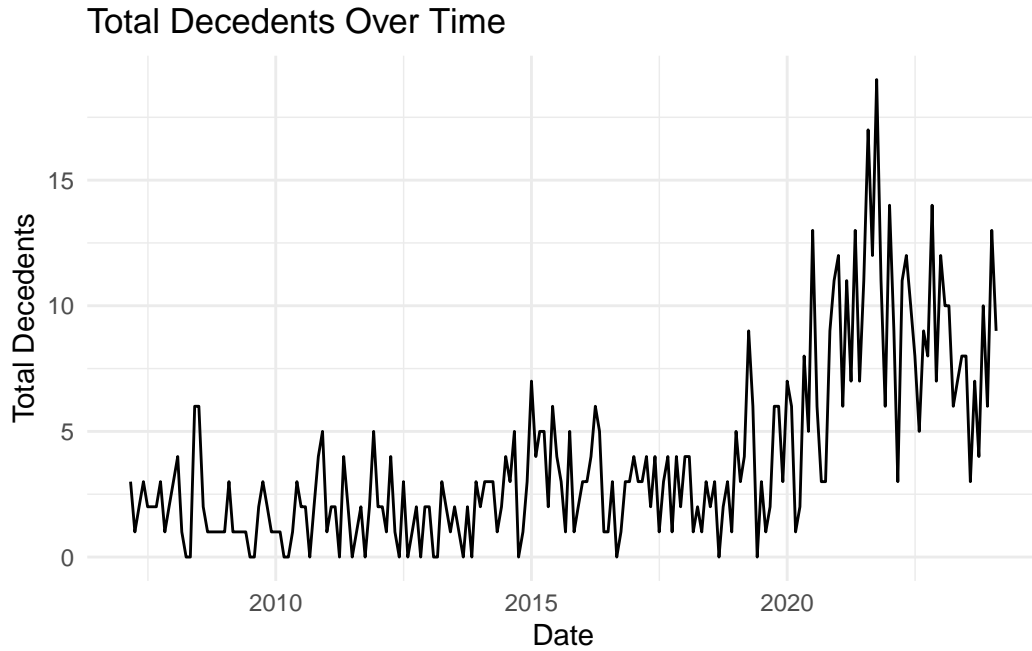


Figure 1: Total Decedents Over time

3 Result

Figure 2 presents the fitted results of a simple linear regression model. The blue line represents the predicted data, while the points indicate the actual observed data. It is evident that the number of deaths shows a clear upward trend. The slope of the line suggests that the death count is increasing at an approximate rate of 1.009105 per week.

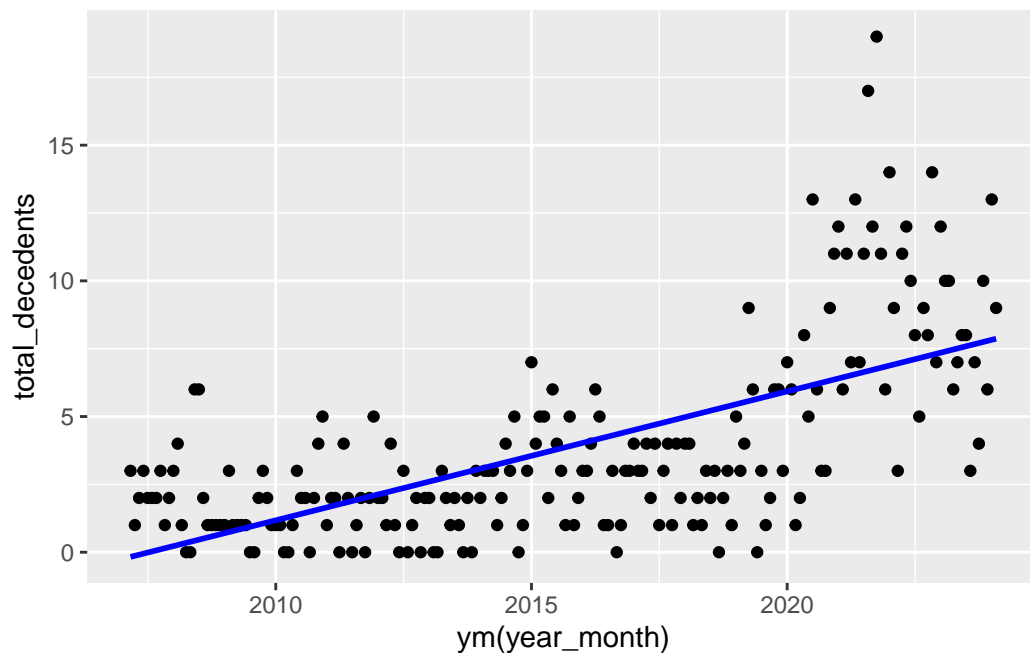


Figure 2: Simple linear regression

Figure 3 displays the total number of deaths per year, along with their respective proportions. It is important to note that I removed the data for 2007 and 2024 because they do not represent complete years. From the chart, it is evident that the number of deaths between 2020 and 2023 was significantly higher, with each of these years accounting for over 10% of the total deaths during the 15-year period. The peak occurred in 2021, reaching as high as 17.8%.

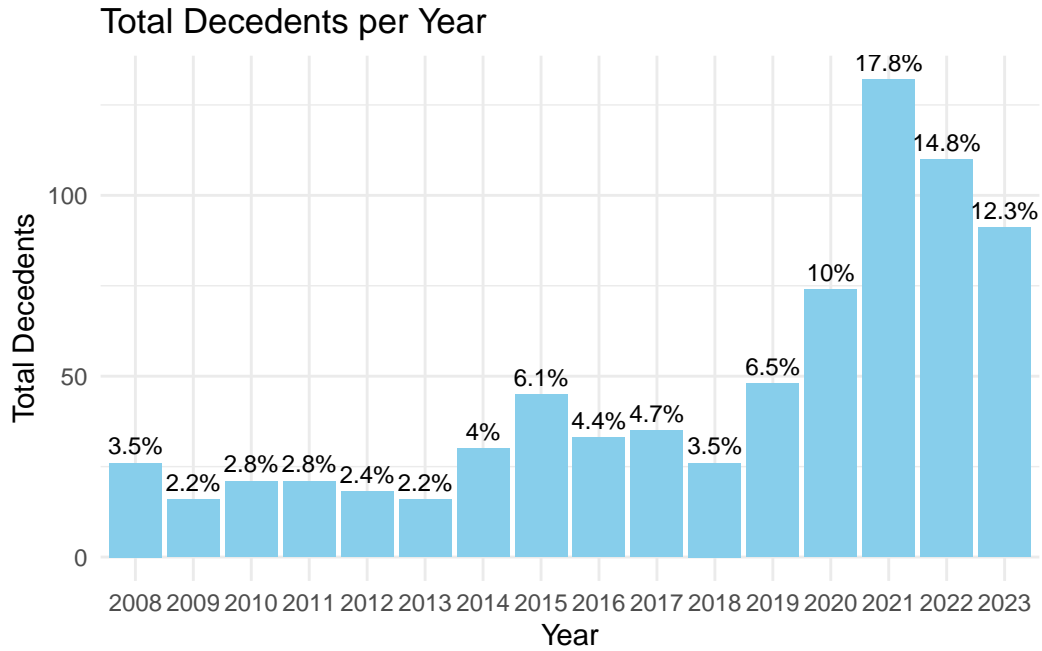


Figure 3: Total Decedents per Year

Additionally, Figure 4 shows the average number of deaths per season over the past few years. The proportions for spring, autumn, and summer are relatively similar, ranging between 23.6% and 23.8%. However, winter stands out with a significantly higher proportion, reaching 29%, notably surpassing the other three seasons.

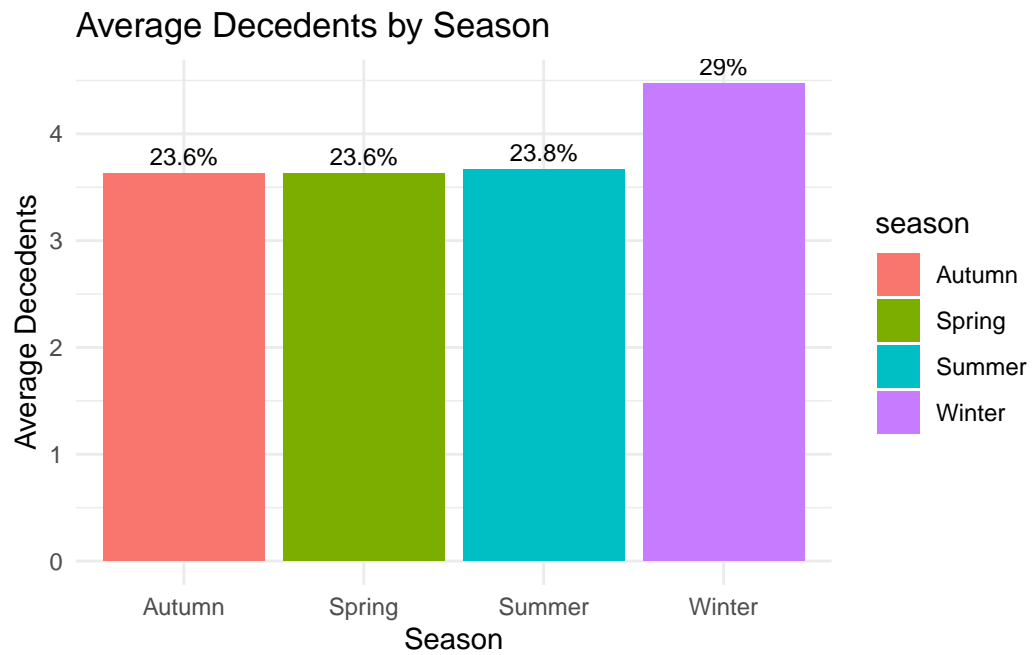


Figure 4: Average Decedents by Season

4 Discussion

Appendix

References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.