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 (Omstead (2023)). By the end of 2024, 24 shelter hotels will no longer be available due to expiring contracts, exacerbating an already overcrowded shelter system (Omstead (2023)). 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. For example, Maureen Crane and Louise Joly have studied the growing number of shelter residents and the issue of aging in Western countries (Crane and Joly (2014)). Similarly, in their article, John M. Quigley, Steven Raphael, and Eugene Smolensky pointed out that the number of shelter residents in the United States is increasing, along with the underlying causes contributing to this rise (Quigley, Raphael, and Smolensky (2001)). 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 population numbers, which are often difficult to obtain and track (Rossi et al. (1987)).

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. This project also examines anomalies in the data by comparing yearly death numbers, 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 in Toronto 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 (Gelfand (2022)). 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 (Gelfand (2022)). The dataset is continuously updated. Table 1 contains descriptions of these variables.

Table 1: Column descriptions of the original data

Column	Description
_id	Unique row identifier for Open Data database.
Year	The calendar year being reported on.
Month	The month being reported on.
Total decedents	The total number of shelter residents who died in the reported month/year.
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 (Gelfand (2022)). 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 in the results.

2.2 Data Cleaning

During the data processing phase, tasks such as downloading, cleaning, simulating, analyzing, and code testing were all performed using R code (R Core Team (2023)). Additionally, the project utilized packages such as Tidyverse (Wickham et al. (2019)), Dplyr (Wickham et al. (2023)), Lubridate (Grolemund and Wickham (2011)), and Knitr (Xie (2024)) for data handling. These tools were instrumental in performing linear regression, creating visualizations, generating tables, and formatting data for analysis.

The data cleaning process involved several key steps, each designed to improve the dataset's suitability for analysis. First, the year and month variables were combined, with the month converted to numerical format to facilitate time series analysis. A new variable was then created to classify each month into one of the four seasons: March, April, and May as spring; June, July, and August as summer; September, October, and November as autumn; and December, January, and February as winter, enabling seasonal analysis. To streamline the dataset, unnecessary variables, such as ID and the three gender categories, were removed. Additionally, to ensure that the seasonal analysis of average shelter resident deaths covered equal time intervals for each season, certain samples from the dataset's beginning and end were removed, resulting in a total number of samples divisible by four. Finally, after reviewing

the dataset, it was confirmed that there were no missing values, so no further treatment was necessary. Upon completion, the cleaned dataset was reduced to three variables: Time, Total Deaths, and Season, as outlined 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
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. Additionally, we can infer that the data is somewhat skewed, as the highest number of deaths in a single month is 19, while the average and median number of deaths are 3 and 3.83, respectively.

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 depicts the monthly variation in the number of deaths over time. The overall trend reveals a significant increase in the total number of shelter resident deaths in Toronto, particularly after 2015. Prior to 2015, the number of deaths remained relatively stable, rarely exceeding 5 per month. However, from 2015 onwards, there is not only a noticeable upward trend in deaths but also increased volatility. Spikes in the data become more frequent, with several months recording over 10 deaths, and some months reaching as high as 15 or more.
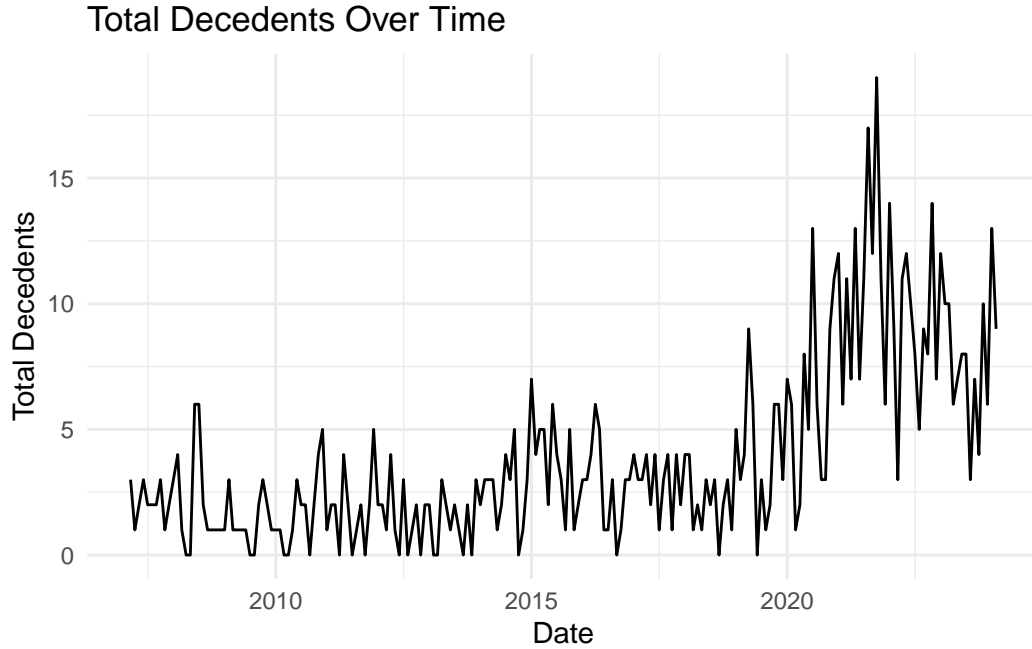
Figure 1: Total Decedents Over time

3 Result

I used the total number of shelter resident deaths per month as the target variable to be predicted, with time and season as predictors, and performed a linear regression analysis. Since time and season are evidently highly correlated, I decided to retain only time as the final predictor for the model. After completing the simple linear regression, I obtained a positive slope, represented by a straight line (Figure 2), indicating that the number of shelter resident deaths in Toronto has increased over time. A comparison with the actual data revealed that the real values closely align with the predictions from the SLR model, demonstrating the reliability of my model.

Figure 3 displays the total number of shelter resident deaths per year, along with their respective proportions. It is important to note that I excluded the data for 2007 and 2024, as these are incomplete years and including them could mislead the reader. From the chart, it is clear 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, where the number of deaths exceeded 125, representing 17.8% of the total.

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%,
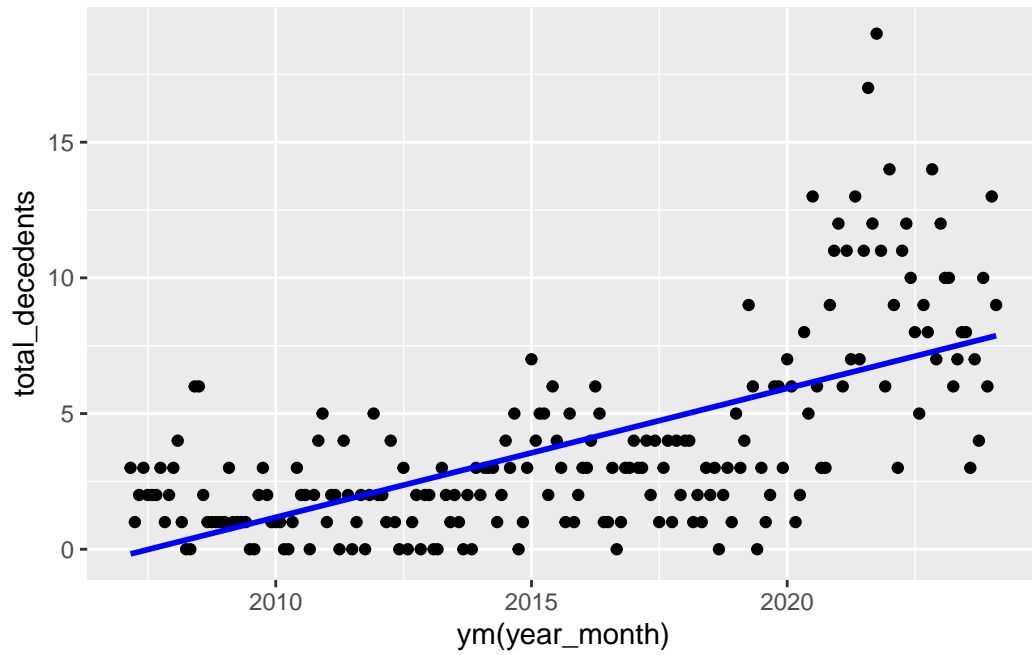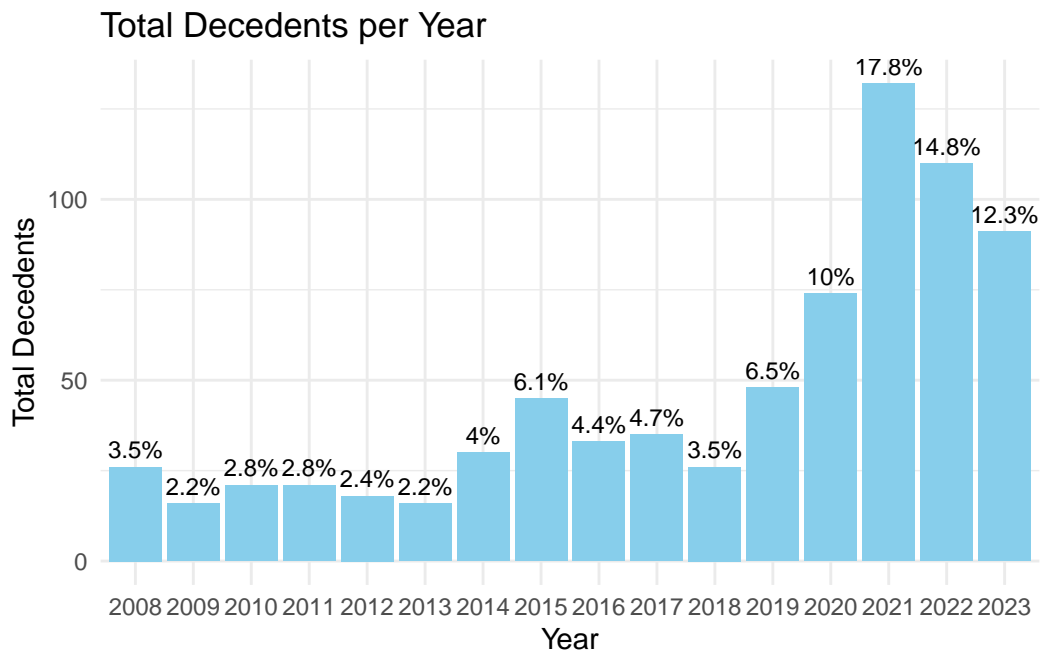
Figure 2: Simple linear regression

Figure 3: Total Decedents per Year

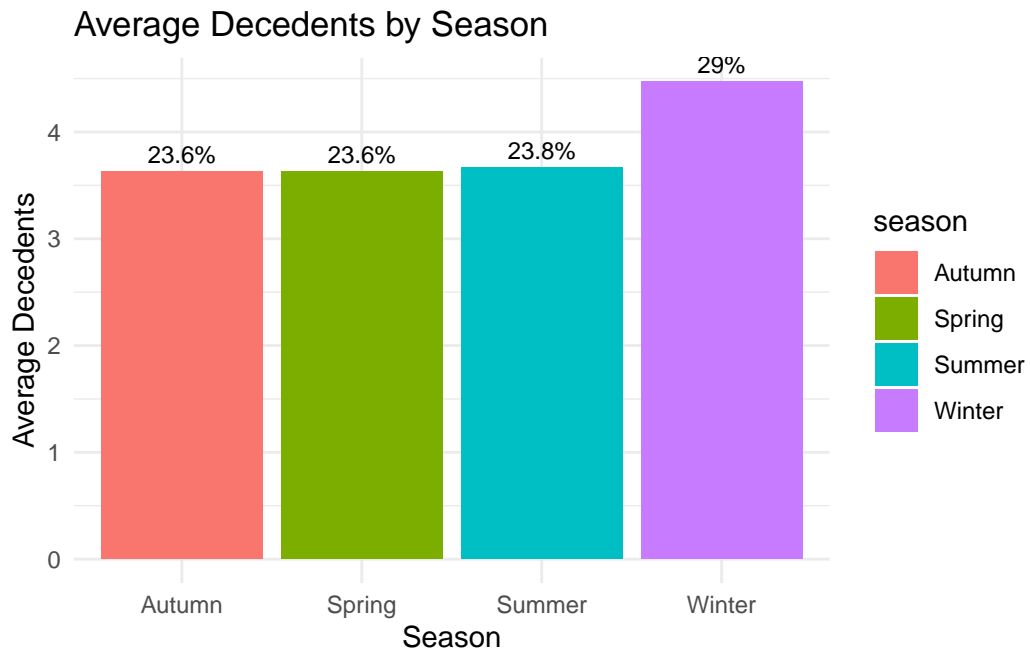notably surpassing the other three seasons.

Average Decedents by Season

Figure 4: Average Decedents by Season

4 Discussion

Appendix

References

Crane, Maureen, and Louise Joly. 2014. "Older Homeless People: Increasing Numbers and Changing Needs." *Reviews in Clinical Gerontology* 24 (4): 255–68. https://doi.org/10.1017/S095925981400015X.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Omstead, Jordan. 2023. *Residents Demand Action from Toronto Mayor Olivia Chow as Another Shelter Hotel to Close.* globalnews.ca. https://globalnews.ca/news/9835385/toronto-shelter-hotel-to-close-residents-demand-action/.

Quigley, John M., Steven Raphael, and Eugene Smolensky. 2001. "Homeless in America, Homeless in California." *The Review of Economics and Statistics* 83 (1): 37–51. https://doi.org/10.1162/003465301750160027.

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

Rossi, Peter H., James D. Wright, Gene A. Fisher, and Georgianna Willis. 1987. "The Urban Homeless: Estimating Composition and Size." *Science* 235 (4794): 1336–41. https://doi.org/10.1126/science.2950592.

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, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.