My title*

My subtitle if needed

Colin Sihan Yang

September 26, 2024

This study explores the seasonal patterns and the deaths of the homeless people who are under the roof of the shelter system in Toronto. The research encompasses the deaths among the shelter residents and conducts analysis for 2007-2024 basing on open data material received from Open Data Toronto. It does have a particular diligence when winter months regime with death increase is noted while also establishing the effect of the COVID-19 pandemic on the ever rising death rates between the years 2019 and 2024. The results underscore the need for targeted health strategies particularly related to medical, and psychiatric services and shelter improvements in vulnerable periods.

Table of contents

*Code and data are available at: https://github.com/hoodiexxx/deaths_shelter.

1 Introduction

The increasing number of homeless people deaths in the City of Toronto has emerged as a core public health concern which portrays the harsh situation of the unaccommodated people. According to the studies that were recently published by the Toronto Public Health (Weekly 2023),79 homeless persons were reported to die in the first six months of 2023, many of the deaths due to diease, flu, and epidemic that has plagued the city (Relations 2023). This mortality not only implicates the health consequences of diease, but it also indicates systemic deficits in provision of safe facilities and healthcare services.

The increasing rates of homelessness coupled with the complexity of health issues faced by this population present a critical challenge for urban centers worldwide. Toronto, Canada's largest city, exemplifies this challenge, with a significant homeless population reliant on the shelter system. Recent trends and emerging data suggest that the mortality rate among these individuals is influenced by both environmental factors, such as harsh winter conditions, and systemic pressures including public health crises like the COVID-19 pandemic, drawing on open data from the period 2007 to 2024 provided by Open Data Toronto.(Weekly 2023). It also demonstrates that there is an apparent gap when it comes to devising public health policies to manage the risks posed by drug misuse in a matter of urgency and address the key requirements of housing and Epidemiological sanitary conditions in the long term(Weekly 2023).

A mix of plans like the Toronto Overdose Action Plan or the Integrated Prevention and Harm Reduction Initiative (iPHARE) aimed at minimization of these risks have been adopted by the city. Collectively, they have provided some harm reduction services and addressed some mental health needs of their clients as well. Although, the high mortality rates have remained persistent and there is a need to contextualize these strategies and more efforts from the provincial and federal governments to protect both the measures used in the city and those who need the city most.

The study presented in this paper investigates the association between seasonal trends and the number of deaths recorded within Toronto's shelter system. This seasonal cycle could have some impact on the rate of deaths. The analysis implies that the government of Toronto could enhance its strategies and assistance of the homeless population based on these trends during certain months or seasons which are likely perceived to be problem new. Particularly, providing more medical care and better shelter conditions in the winter months will likely help to minimize deaths from the elements or some illnesses. In a similar way, increasing the availability of COVID-19 testing, vaccinations, and treatment during periods of heightened transmission, as seen in time series data, could significantly help in mitigating the spread

of the virus. This method would allow for adjusting public health interventions to peak transmission times, ultimately reducing the infection rate and lowering the general health risks for vulnerable populations, such as the elderly and those with pre-existing conditions. By strategically focusing resources on critical moments, we could dramatically lessen the overall impact of the pandemic, both in terms of hospitalizations and mortality rates.

The remainder of this paper is structured as follows. Section 2 discuss the raw data cleaning process, and offers a insight at the underlying distribution of data through tabular and graphical representations of the observations.

2 Data

2.1 Raw Data

The data used in this paper is sourced from Open Data Toronto and is load into this paper through the opendatatoronto library (Gelfand 2022). The particular data set is used to analyze the deaths of Toronto shelter residents(Support Services 2024). All the data analysis was done through R (R Core Team 2023) with the aid of the following packages: tidyverse (Wickham et al. 2019), here (Müller 2020), tinytex (Xie 2019), dplyr (Wickham et al. 2023), tibble (Müller and Wickham 2023), janitor(Firke 2023), ggplot2 (Wickham 2016), and knitr (Xie 2014).

The data used is published by Toronto Shelter & Support Services, providing information about the deaths of shelter residents. The data used in this paper is up-to-date as of Sept 19, 2024, the raw data set features the total decedents with 212 observations and 7 attributes. The dataset includes the id, Year, Month, Total Decedents, Male, Female, Transgender/Non-Binary/two_spirit. A sample of the cleaned data can be seen in Table 1 and a bar grapgh showcasing every observation, by date of collection, can be seen in **?@fig-lead-samples-over-time**.

2.2 Clean Data

Measurement: First, pipe the data into the janitor to unify the column attributes name, to avoid Uppercases and other special symbol that are hard to deal with. Secondly, extract the specific attributes we focus in this paper by using “select” method. Finally, write down the clean dataset from the RAM into SSD by using “write_csv” for the future usage.

the clean data has 212 rows x 3 columns(year, month, total_decedents)

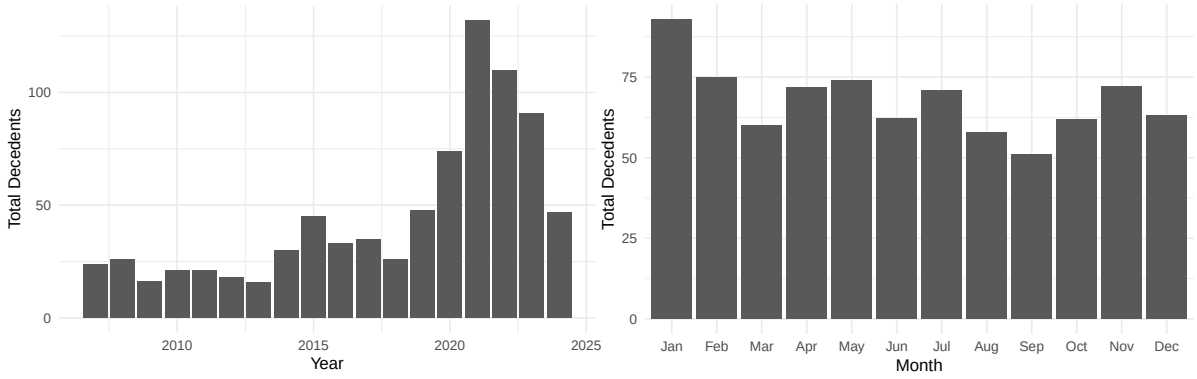
Table 1: Sample of cleaned lead data

year	month	total_decedents
2007	Jan	0
2007	Feb	3
2007	Mar	3
2007	Apr	1
2007	May	2
2007	Jun	3

2.3 Basic Summary Statistics of the Data

Figure 1a the bar graph that plot the total decendents in Toronto shelter vs the year between 2007 to 2024. We could observe a strong skew to the right pattern in the bar graph. The number of deaths dramatically increase and reach to the peak during 2019-2024 that is strongly overlap with the covid period.

Figure 1b the bar graph that show the total decedents in Toronto shelter vs the months from January to the December. it is worthy to mention that the number of death increase during winter season, for more specific, January and Feburary which the months that toronto usual have coldest weather.



(a) total decedents vs year

(b) total decedents vs month

Figure 1: the total decedents vs year in shelter and the total decendets vs month in shelter

Table 2 the table shows the number of observation about each year toronto shelter deaths and we could see that 2007-2023 got 12 months observations while the 2024 only got 8 observations that is because the data used in this paper is up-to-date as of Sept 19, 2024.

Table 2: Number of observations by year

Year of Sample Collection	Number of Observations
2007	12
2008	12
2009	12
2010	12
2011	12
2012	12
2013	12
2014	12
2015	12
2016	12
2017	12
2018	12
2019	12
2020	12
2021	12
2022	12
2023	12
2024	8

Table 3 the table show the Mean of response variable total decedents and the standard deviation of response variable total decedents.

Table 3: Mean and standard deviation of all observed decedents

Mean total decedents	standard deviation of total decedents
3.83	3.59

Figure 2 This boxplot illustrates the distribution of total decedents observed at Toronto shelters. Here's a breakdown of its components:

- Median: The line inside the box represents the median number of decedents, which appears to be around 3 or 4.
- Interquartile Range (IQR): The box itself spans from the first quartile (Q1) to the third quartile (Q3), encompassing the middle 50% of the data. It appears to range from about 2 to 5 decedents.
- Whiskers: The whiskers extend from the edges of the box to the minimum and maximum values within 1.5 times the IQR from the first and third quartiles, respectively. Here,

the lower whisker appears to start at 0 (indicating the minimum), and the upper whisker extends up to around 7 or 8.

- Outliers: Points outside the range of the whiskers are considered outliers. These are shown as individual dots above the upper whisker. In this plot, there are several outliers, with the highest value around 15 decedents.

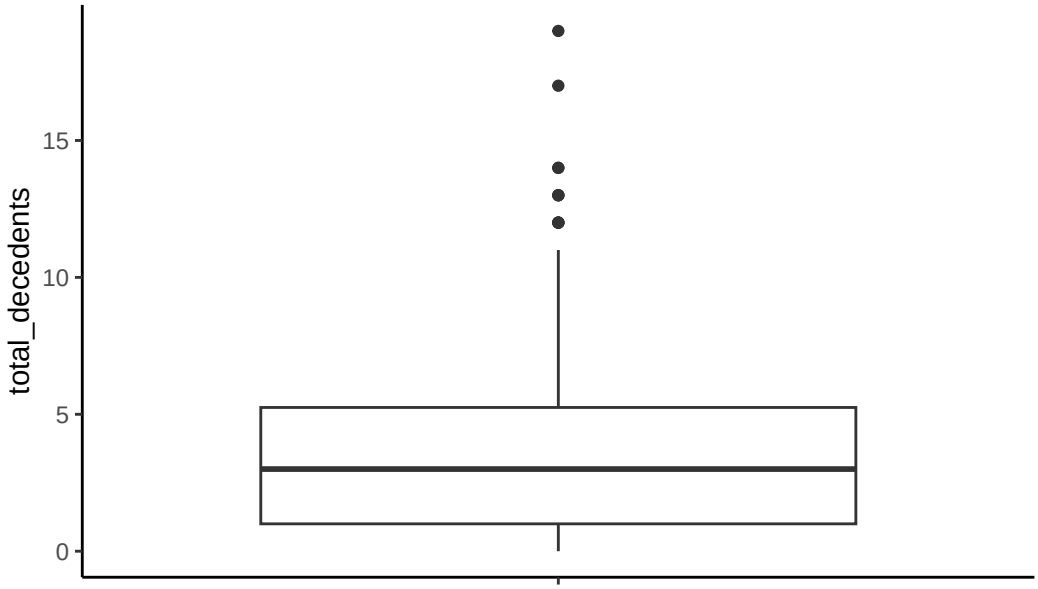


Figure 2: boxplot for the distribution of observed Toronto shelter decedents

3 Discussion

3.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

3.2 Second discussion point

3.3 Third discussion point

3.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

References

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

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

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

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

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

Relations, Toronto Public Health Media. 2023. "2023 Annual Review of Statistics on Deaths of Shelter Residents." https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/deaths-of-shelter-residents/annual-review-of-statistics-on-deaths-of-shelter-residents/.

Support Services, Toronto Shelter &. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.

Weekly, GTA. 2023. "Toronto Public Health Releases Mid-2023 Data for Deaths of People Experiencing Homelessness." https://www.gtaweekly.ca/toronto-public-health-releases-mid-2023-data-for-deaths-of-people-experiencing-homelessness/.

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

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

———. 2019. "TinyTeX: A Lightweight, Cross-Platform, and Easy-to-Maintain LaTeX Distribution Based on TeX Live." *TUGboat* 40 (1): 30–32. https://tug.org/TUGboat/Contents/contents40-1.html.