Does the Time of Year Affect Death Tolls Among the Homeless Population? An Analysis of Deaths of People Experiencing Homelessness*

Faiza Imam

January 25, 2024

Toronto's homeless crisis has been a persistent issue for years which prompted Toronto Public Health to begin collecting data on the death tolls among this population back in 2017. Using the data acquired from Open Data Toronto, this paper investigated the relationship between the time of year and its impact on the death toll of the homeless population and the findings suggest that there is no correlation.

Table of contents

*Code and data from this analysis are available at: https://github.com/fazim07/Homeless_Death_Toll

1 Introduction

Toronto's homeless crisis has been an ongoing issue for the city, in recent years the number of people facing homelessness increased significantly. It is estimated that over 10,000 people in the city are homeless on any given night (Health (2023)). For the course of this paper, the term 'homelessness' is described as a situation of an individual or family without stable or permanent housing or any means of acquiring housing. (Health (2023)). This is a result of systemic barriers such as the lack of affordable and appropriate housing, and the financial, mental, cognitive, behavioural, racial or physical challenges faced by the individual or individuals (Toronto (n.d.)). It is important to note that people do not choose to become homeless and the experience of being homeless is unpleasant, unsafe, stressful and distressing for those who are facing this challenge currently (Victor (n.d.)). In 2017, TPH (Toronto Public Health) began collecting data regarding the deaths within the population facing homelessness. The intention was and still is to understand the trends, gain an accurate estimation of the demographics affected and understand the cause of these deaths.

A study conducted on Poland's homeless population found that homeless people are more susceptible to higher mortality rates due to prolonged exposure to environmental conditions. (Jerzy Romaszko (2017)). Atmospheric conditions such as weather, freezing temperatures and heat waves can cause an increased mortality rate among this population (Jerzy Romaszko (2017)). The result of the study found that Poland's homeless population is more at risk of cold-weather-related deaths than those in the heat. Another report from the University of Chicago saw an overwhelming number of recorded deaths among the homeless population in 2022 due to harsh weather conditions. Shelters in the city of Chicago were at maximum capacity and had to turn away individuals (Rios (2023)). Particularly in the winter season, shelters will not open their doors unless the temperature drops below -30 degrees (Rios (2023)). Similarly, in Toronto, shelters in the GTA will not open doors until the temperatures reach below -30 degrees and colder (Freeman (2023)). There has been little research and data on Toronto's homeless population regarding whether the time of the year has an impact on the death toll. This paper looks to analyze that same question to find a trend or pattern to contribute to research.

The purpose of this paper is to identify and analyze the collective monthly death tolls of the homeless population in Toronto from 2017 to 2023 to find a common pattern to understand whether the time of year (month) affects the death tolls of the homeless population. To examine the patterns and their implications, a graphical analysis was conducted to visualize the data collected from the past seven years. Findings suggest that there is no relation between the time of month and death tolls among the homeless population.

This paper is divided into the following sections: Data, Results, Discussion and Conclusion. Each section is written to reflect and analyze the data obtained and collected from Open Data

Table 1: Table of Yearly Death Tolls

Year	Death Tolls
2017	101
2018	94
2019	127
2020	145
2021	219
2022	189
2023	79

Toronto. In the Data section, I discuss the data set acquired and the reason for conducting an analysis using the programming language R. The Result section discusses the trends and patterns observed in the data, followed by the Discussion section, which further evaluates the observations made in the previous sections. Lastly, the Conclusion summarizes the paper and future steps to take regarding this topic of interest.

2 Data

The data utilized in this paper was gathered from the Open-DataToronto data reserve. The data set is entitled 'Deaths of People Experiencing Homelessness'(Health (2023)). Data was obtained and analyzed through the statistical programming language R (R Core Team (2023)). Additional packages used are from tidyverse (Wickham et al. (2019)), ggplot (Wickham (2016)), janitor (Firke (2023)), tibble (rTibb), and dyplr (Wickham et al. (2023)). More details regarding analysis and data collection are discussed further in this paper. The data chart is separated into three main categories, 'death tolls by months', 'death tolls by cause' and 'death tolls by demographics'. For this paper, the first data set is used and analyzed. This data consisted of four columns, 'ID', 'year_of_death', 'month_of_death', and 'count'(Health (2023)).

2.0.1 Death Tolls by Year

This portion looked over the yearly death toll from the beginning of the research, 2017, to the current year (2023). I used the count function to tally up the total death tolls for each year and grouped them into their respective years (See Table 1).

Table 2: Sample of Monthly Deaths

Month	Death Tolls
January	106
February	80
March	64
April	72
May	102

Table 3: Sample of Year-Monthly Death Tolls

Year	Month	Death Tolls
2017	January	9
2017	February	11
2017	March	7
2017	April	8
2017	May	7

2.0.2 Death Tolls by Months

This portion required more cleaning than the original data set had the months organized and grouped by their respective years. The data in this portion also contained 'Unknown'/NA variables. To ensure that there are no outliers within the data set that may affect the overall analysis of the data. I used the factor () function to order the months in chronological order, then once again used count() to tally up the death toll from all the years by month (See Table 2). This set will aid in comparing a monthly death toll analysis, discussed further in the paper.

2.0.3 Death Tolls by Month and Year

This data consisted of four columns, 'ID', 'year_of_death', 'month_of_death', and 'count'(Health (2023)). After doing first rounds of clean up, I selected; 'year_of_death', 'month_of_death', and 'count'ensuring that they are represented in their respective classes and names (See Table 3).

3 Results

3.0.1 Yearly Death Tolls

Over 7 years since the data collection project had begun, 954 deaths were recorded. The year, 2021 recorded the most deaths, sitting at 219. While 2023 reported about 79 deaths in total. From 2018 and onwards there was been a steady increase in death tolls before a drop near 2023. Figure 1, displays the yearly death tolls from 2017-2023.

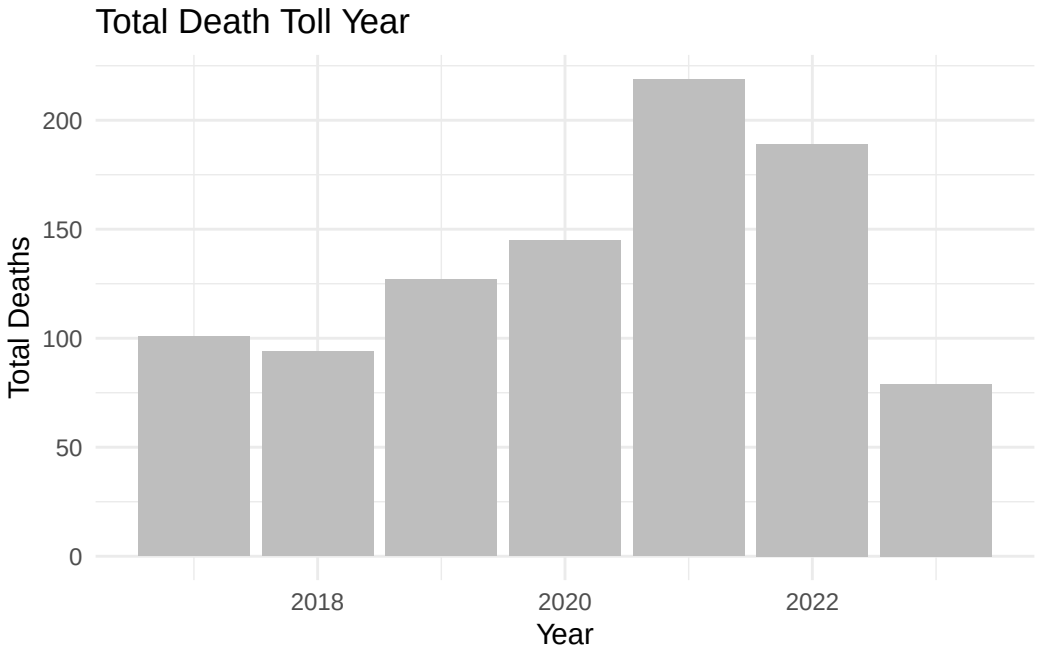

Figure 1: Trend of Yearly Deaths from 2017-2023

3.0.2 Collective Monthly Death Tolls

The collective monthly death tolls show a gradual fluctuation in total calculated death tools from all the years. January and May marked the highest death toll records with 106 and 102 respectively. Figure 2 , displays the collective monthly death tolls.

3.0.3 Monthly and Yearly Breakdown of Death Tolls

Figure 3, displays the monthly death tolls grouped by year. The months of January and May once again have high recorded death tolls throughout the years, with a mean death toll of 17 each year compared to the other months, averaging about 12.5 death tolls. However, in 2021,

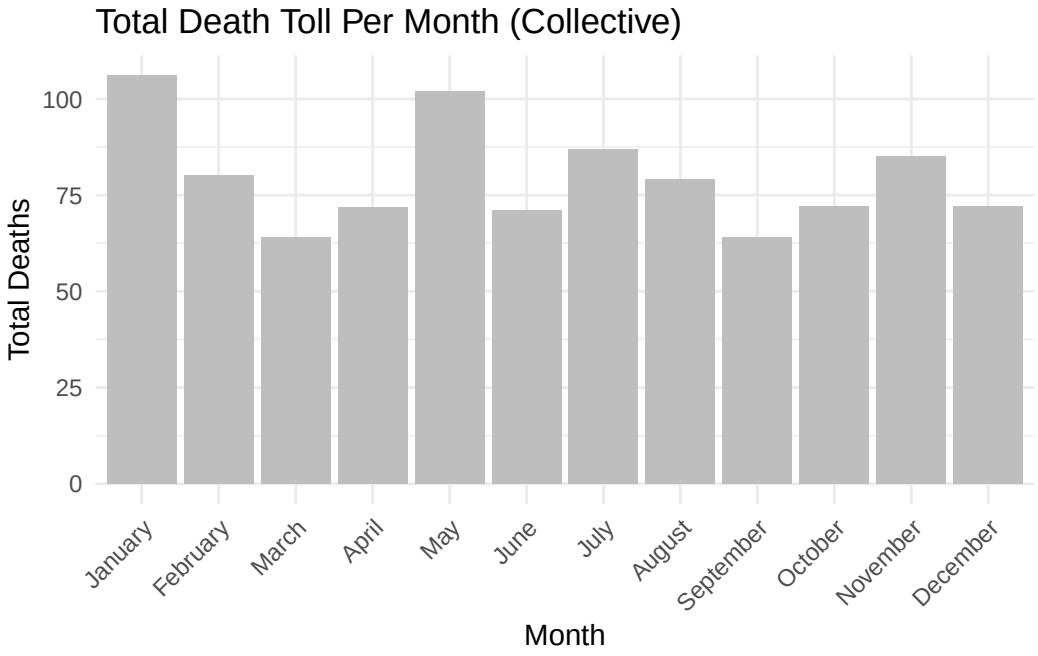

Figure 2: Trend of Deaths Tolls Monthly (Collective)

the months of August to November have higher death tolls compared to the previous year's records.

4 Discussion

Based on the results, it is revealed that there is no correlation between the time of the years (months) and their effects on death tolls. As projected in Figure 3, there is no seasonal or weather-related trend to suggest that there is an impact on death tolls within the homeless population based on the time of the year. In Figure 2, the collective monthly death dolls showed that January and May had the highest death tolls recorded throughout the years, however, when comparing that to Figure 3, there is no trend to show the specific patterns present in all the years to establish a trend.

The data set used for analysis may have potential errors which may have affected the overall results. One of the errors may have been from my conclusion to omit "NA" variables. In the month's column of the set, there were a few "NA" entries and I decided to omit them during the cleanup process to avoid any outliers.

It is important to remember that the data collection for this particular set began in early 2017, pre-pandemic. As the pandemic hit, social distancing and lockdowns were enforced, making any sort of social interaction difficult and tedious. As a result, the effects of COVID-19 may

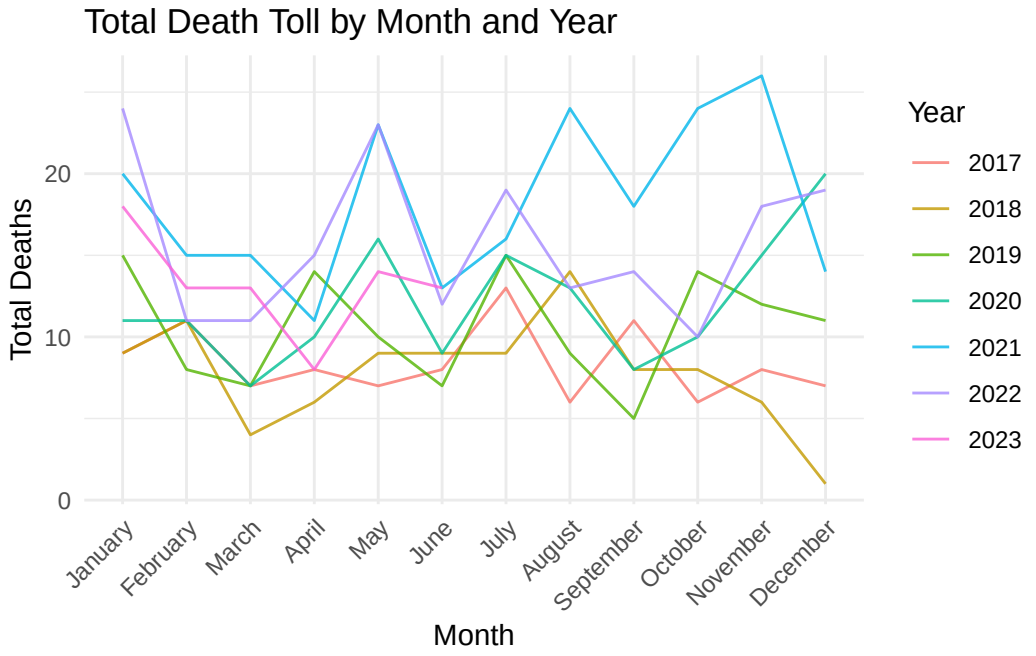

Figure 3: Monthly Death Tolls Every Year

have played a vital role in the data collection process as there may have been more deaths than accounted for. As we see in Figure 1, there is a jump in the death toll in 2021 in comparison to previous years. The number may not accurately represent the true number of death tolls in 2021.

Similarly, in Figure 3, there is an increase in recorded death tolls in 2021, contributing to the overall yearly increase, which was recorded during the peak of the pandemic. This may have affected the data as there had been an increase in recorded death tolls spanning from August to November 2021. This external factor may have masked any potential seasonal or weather-related trends. However, it is important to highlight that the death tolls reported for this specific data set may not reflect the true numbers, as there may be more homeless deaths that have not been accounted for. This requires improved data collection methods and a holistic plan to navigate existing and potential factors regarding death tolls among the homeless populations.

Future studies should aim to look at the cause of death among the homeless population and their effects on seasonal/weather-based death trends. To understand the broader implications of the homeless population death trends. This also includes looking at the demographics affected, the socioeconomic implications and the ages of these people. These studies can aid the city in understanding how to tackle the homeless crisis and how to allocate resources to support those who are facing homelessness.

5 Conclusion

This paper aimed to investigate the connection between the time of year (months) and its effects on the death toll of homeless people using recorded data for homeless death tolls in Toronto from the years 2017-2023. My results reveal that there is no correlation between the time of year and its impact on death tolls among the homeless population. Variables such as cause of death, situational implications, shelter capacity and records, location within the GTA, and gender implications should be analyzed and studied to further understand the trends of homeless death tolls in Toronto and to pinpoint any underlying factors.

The next data collection for 2024 is projected to be released near the end of the fourth quarter (Health (2023)). This new information may reveal additional information and implications regarding the death toll trends among the homeless population. As mentioned previously in the paper, homelessness is not a choice one makes but rather a situation in which individuals fall victim due to complex interconnected factors, often beyond their control. The municipal government must take action against the homelessness crisis as it is an ongoing issue that requires a multifaceted and comprehensive approach to finding an impactful and sustainable solution. Data analyzed in papers like this can aid in identifying patterns to contribute to finding a possible solution.

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Freeman, Joshua. 2023. *Toronto Says No to Immediate 24/7 Warming Centres, but Will Study the Idea Further*. https://toronto.ctvnews.ca/toronto-considering-expanding-warming-centre-hours-1.6264667.

Health, Toronto Public. 2023. *Opendatatoronto: Deaths of People Experiencing Homelessness*.

Jerzy Romaszko, Ewa Dragańska, Iwona Cymes. 2017. *Mortality Among the Homeless: Causes and Meteorological Relationships*. https://doi.org/10.1371/journal.pone.0189938.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. %3Chttps://www.R-project.org/.

Rios, Lisa. 2023. *Homeless Deaths Surged Across the US in 2022 Due to Extreme Weather, Economic Hardship*. https://www.wsws.org/en/articles/2023/01/03/bhjo-j03.html.

Toronto, Haven. n.d. *Why Spring Is the Worst Season to Be Homeless*. https://www.haventoronto.ca/single-post/2019/03/23/why-spring-is-the-worst-season-to-be-homeless#:~:text=The%20days%20are%20longer%20and,the%20homeless%20are%20especially%20susceptible.

Victor, Fred. n.d. *Facts about Homelessness in Toronto*. https://www.fredvictor.org/facts-about-homelessness-in-toronto/.

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. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.