Mortality Trends Among Shelter Residents in Toronto: A Longitudinal Study from 2007 to 2024*

Fangning Zhang

September 26, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019a).

The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

The dataset used in this study is titled "Deaths of Shelter Residents". The dataset is published by the Toronto Shelter and Support Services Department and is obtained from OpenData-Toronto (Gelfand (2022)). It covers the deaths of shelter residents from 2007 to the present and is still updated monthly. Variables such as the total number of decedents and the breakdown of deaths by gender (male, female, transgender/non-binary/two-spirit) are included. All death counts are reported on a monthly basis. For privacy and ethical reasons, the dataset does not contain any personally identifiable information.

A dataset titled "Death Registry Statistics" is also found on OpenDataToronto (Gelfand (2022)). The dataset is related to the study, but the variables are too general. For example, information on the number of deaths among vulnerable populations cannot be found in

*Code and data are available at: https://github.com/FangningZhang81/Deaths-of-Shelter-Residents.

this dataset. The dataset titled "Fatal and non-fatal suspected opioid overdoses in the shelter system" also found on OpenDataToronto (Gelfand (2022)) can be a helper dataset. Since it only covers one specific cause of death, it cannot serve as the primary dataset for the study.

A Table 1 with first few rows is shown below, which can give an overview of the raw data and variables.

Table 1: First few rows of the raw dataset

_id	Year	Month	Total decedents	Male	Female	Transgender/Non-binary/Two-Spirit
1	2007	Jan	0	0	0	n/a
2	2007	Feb	3	3	0	n/a
3	2007	Mar	3	2	1	n/a
4	2007	Apr	1	1	0	n/a
5	2007	May	2	2	0	n/a
6	2007	Jun	3	3	0	n/a

2.2 Data Tools

The whole study is done with R programming language (R Core Team (2023)). During processing and analyzing the data, multiple of packages are used. For example, The data was extracted, cleaned and visualize using R programming language and the whole coding process is done with R (R Core Team (2023)). Some packages are used throughout the paper, including tidyverse (Wickham et al. (2019b)), ggplot2 (Wickham (2016)), janitor (Firke (2023)), OpenDataToronto (Gelfand (2022)), knitr (Xie (2024)), and KableExtra (Zhu (2024)).

2.3 Data Cleaning

The raw data was first imported from the file and the column names were standardized using the `janitor` package(Firke (2023)) to ensure consistency and ease of use throughout the analysis. The next step involved converting the month abbreviations into numeric values to increase the ease of use and analysis. For example, "01" for January, "02" for February.

A new column named `year_month` was generated by concatenating the `year` and `month` columns making it easier to perform time-series analyses and visualizations.

The dataset also included a column named `transgender_non_binary_two_spirit`, which some rows contained non-numeric values (n/a). All `NA` values in this column were replaced with 0 to ensure the data was suitable for statistical analysis and visualization.

Finally, the cleaned dataset was saved to a new file named `analysis_data.csv` in the `analysis_data` directory. It includes the following variables:

- `year`: The year in which the death occurred.
- `month`: The month in which the death occurred, converted to numeric values for ease of analysis.
- `total_decedents`: The total number of deaths recorded in that month.
- `male`: The number of male decedents.
- `female`: The number of female decedents.
- `transgender_non_binary_two_spirit`: The number of decedents identified as transgender, non-binary, or two-spirit. In cases where this information is unavailable, the value is recorded as zero.
- `year_month`: A constructed variable representing the year and month combined, formatted as "year_month" (e.g., 2007_01 for January 2007).

The Table 2 below provides a summary of the cleaned data, showing key statistics (maximum, minimum, mean, median, and variance) for the total number of deaths, as well as for male and female decedents. It highlights that the average number of deaths per month is 4, with males averaging 3 deaths and females 1 death.

Table 2: Summary of the cleaned data

Statistic	Total	Male	Female
Max	19	16	5
Min	0	0	0
Mean	4	3	1
Median	3	2	0
Variance	13	8	1

2.4 Data Visualization and Result

Graphical representations, including line plot, pie chart and bar charts, are used to visualize the trends in the data.These visualizations are critical for understanding the distribution of deaths across different demographic groups and time periods.

This bar chart(Figure 1) displays the total number of deaths among shelter residents in Toronto from 2007 to 2023. The data shows a significant increase in deaths starting from 2019, with a peak in 2021 (132 people). The upward trend continues into 2022 and 2023 (110 and 91 people), though slightly lower than 2021. The data for 2024 has been excluded from this chart as the dataset only covers information up to August 2024.

Figure 2 displays the death trend of each category among shelter residents in Toronto from 2007 to 2023. The data shows a significant increase in deaths starting from 2019 of all three lines, while the yellow line remains at a relatively low level. The total number of deaths equals the sum of male, female, and other gender deaths.The males death rate has consistently been higher than females and the differences is becoming increasingly obvious. The pie chart

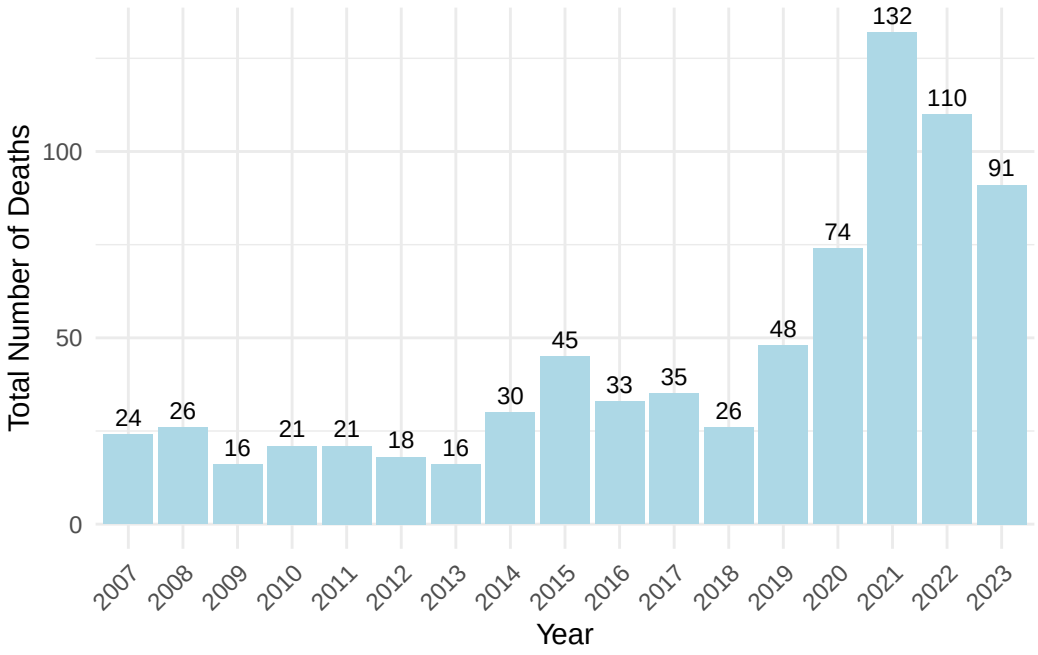

Figure 1: Total Number of Deaths by Year (2007-2023)

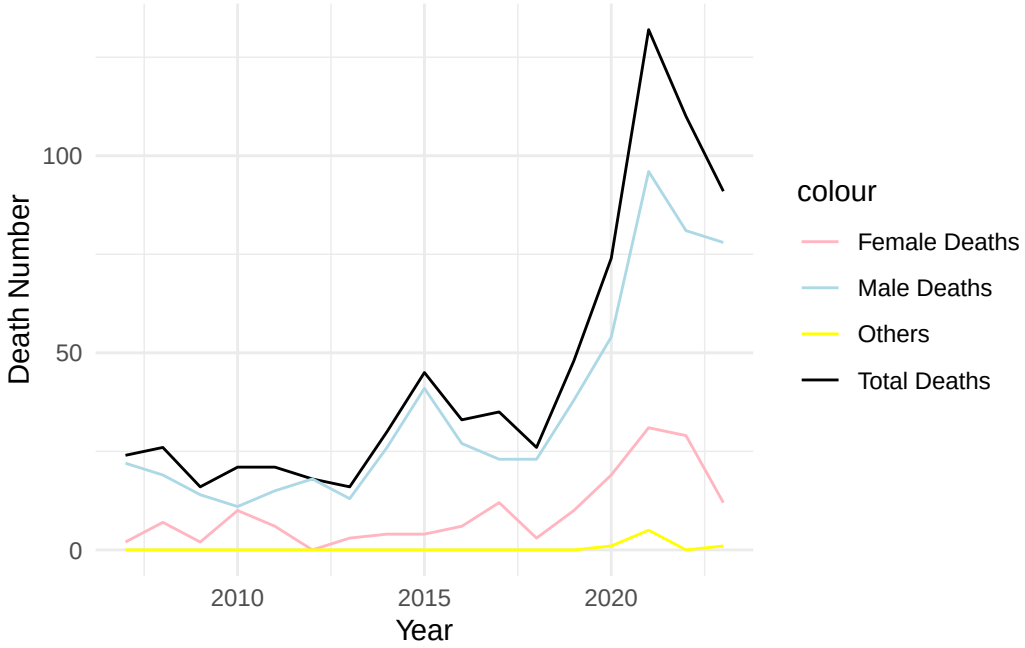

Figure 2: Death Numbers by Year and Category

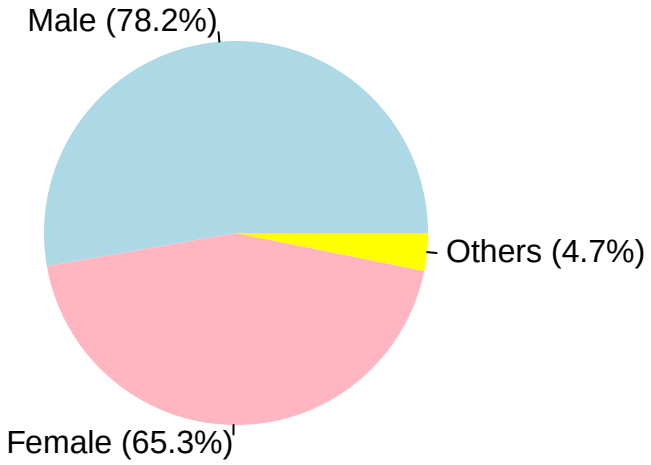

Figure 3: Gender Distribution of Deaths (2007-2023)

(Figure 3) shows the gender distribution of the deaths. 78.2% of the total deaths are from males, while females account for 65.3%. In contrast, individuals of other genders make up only 4.7%.

Figure 4 illustrates the total number of deaths by month among shelter residents. The highest number of deaths occurred in January which is 80 people, followed by relatively high numbers in November which is 72 people. Conversely, the lowest number of deaths occurred in March and September which is 53 and 51 separatly. There is a slight increase in deaths during the winter months (December to February).

2.5 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

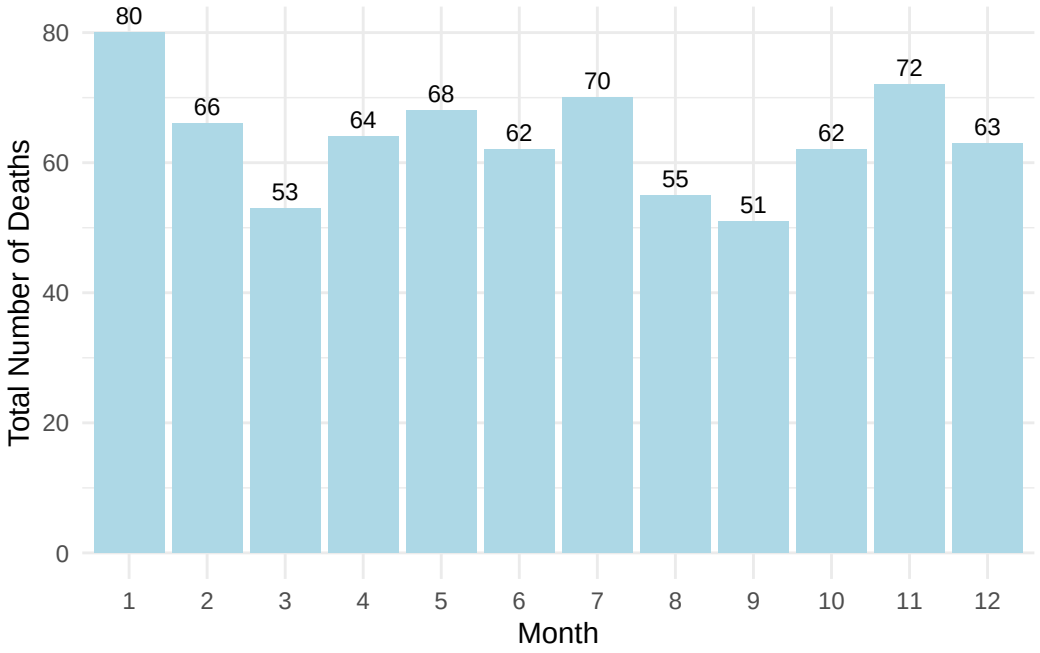

Figure 4: Total Number of Deaths by Month (2007-2023)

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

2.5.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

3 Results

Our results are summarized in **?@tbl-modelresults**.

4 Discussion

4.1 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.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

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.

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

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. 2019b. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

———, et al. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

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

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.