Evaluating the factors pose the greatest threat to the lives and health of different groups of homeless people during 2017-2023*

Juliana Zhu

September 27, 2024

The paper evaluates the factors posing the greatest threats to the lives and heath of various homeless people between 2017 and 2023. The dataset represents key determinants such as accident, drug toxicity and cancer that have impacted mortality rates among homeless people. By examining data, it emphasizes how these factors affect different groups of homeless people, including those based on age, gender, and year of death. The findings underscore the need for targeted interventions to reduce mortality of homeless and improve overall heath conditions .

1 Introduction

The homeless are one of the most vulnerable groups in society, and there are many factors that seriously affect their health and well-being. Discussing and studying the health and life threats of homeless can help urban planners consider the special needs of these people when designing and managing cities, thus creating a more inclusive and safer urban environment. At the same time, such research can prompt more social attention and support for homeless to promote the overall progress of society.

In this paper, I will analyze the homeless deaths by cause data from Open Data Toronto, focusing on the distribution and the relationships between the homeless deaths count, year of death, age groups, gender and cause of data. From the data, the main causes of death among the homeless from 2017 to 2023 are Accident, Cancer, Cardiovascular Disease, COVID-19, Drug toxicity, Homicide, Infection, Pneumonia, and Suicide. By analyzing the data on the

---

causes of death, we can better understand which factors are most harming the health and well-being of the homeless so that we can call on governments and society to take measures.

I use 'R' to do all the statistical analysis and data visualization (R Core Team (2023)) and use 'tidyverse' to plot the data (Wickham et al. (2019)). At the same time, I use 'opendatatoronto' to access and download Toronto City's open data of homeless deaths by cause (Gelfand (2022)). Besides of that, I use 'here' package to simplify the file paths (Müller (2020)) and use 'readr' package to read the date from the file (Wickham, Hester, and Bryan (2024)).

The remainder of this paper is structured as follows: in Section 2 I present the data and the variables used for the analysis. Then I made the plots about relationships between other variables and the count of deaths and compare them to get the final finding.

2 Data

2.1 The background of the data

The dataset is from Toronto Public Health (TPH), strarting in January 2017 to track the deaths of people experiencing homelessness to get a more accurate estimate of the number of deaths and their causes. For this data collection initiative, homelessness is defined as "the situation of an individual or family without stable, permanent, appropriate housing, or the immediate prospect, means and ability of acquiring it".

2.2 Measurement of variables from the data

The data has 5 variables: "year_of_death": the year that the death occurred; "cause_of_death": the reasons of homeless deaths; "age_group": the age of homeless people; "gender": the gender of homeless people; "count": the total number of homeless deaths.

2.3 Limitations of the data (from opendatatoronto)

-This data reflects only deaths that are reported to TPH by SSHA, community partners and the Coroner's Office.

-Cause of death is unknown or pending in approximately 25% of the reported deaths.

-In 2019, TPH ceased reporting on location of death as this information is often unknown or unverified.

-The data does not identify Indigenous status as this is reported as unknown or missing in 70% of the reported cases. With this high a percentage of missing information, accurate conclusions

cannot be drawn, as such, it is standard epidemiological practice to suppress the release of data.

-Cause of death for transgender people not shown due to small counts. -Cause of death is unknown in approximately 25% of the reports. To protect privacy, causes of death with less than 2% of the cases are included in the Other category.

-Data are subject to change should TPH receive additional reports or further information on existing reports. Previously published data may change due to delays in reporting. The most recently published data should be considered the most complete.
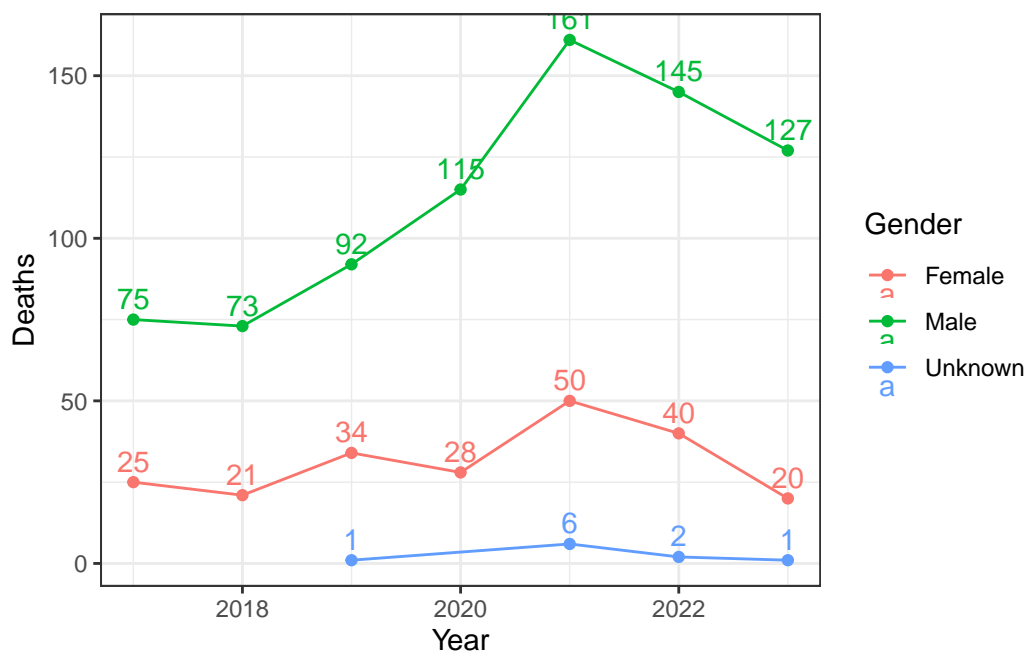
2.4 Visualization of the data

Figure 1: Number of deaths by year for gender groups

Talk way more about it.

Figure 2 presents….. Figure 2a

Figure 2b presents

Figure 2c presents

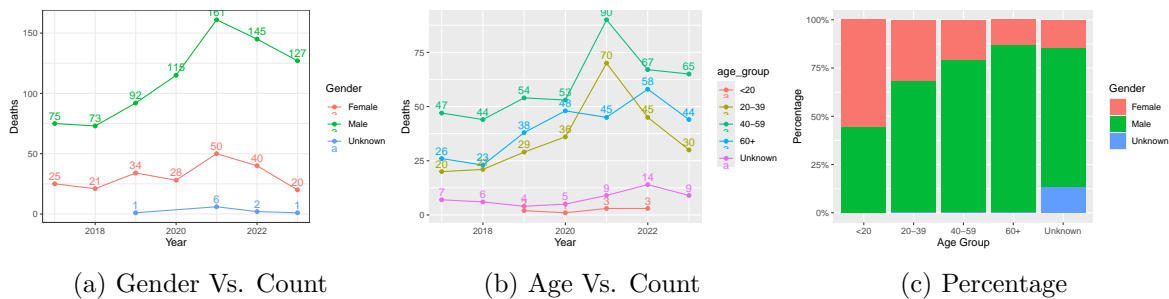(a) Gender Vs. Count   (b) Age Vs. Count   (c) Percentage

Figure 2: Number of deaths by year for different age groups and gender groups
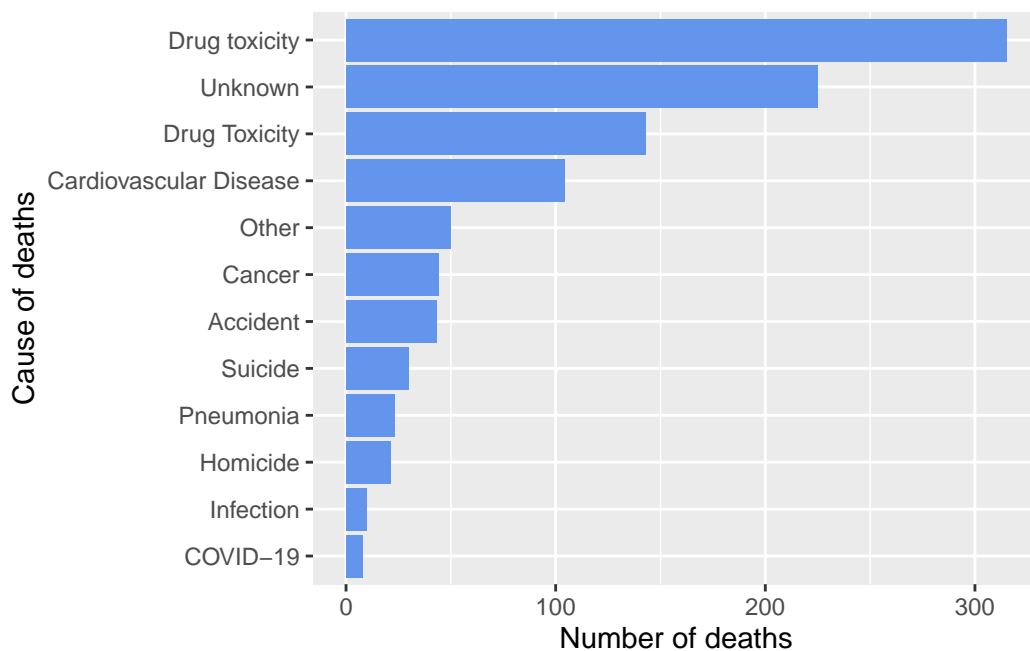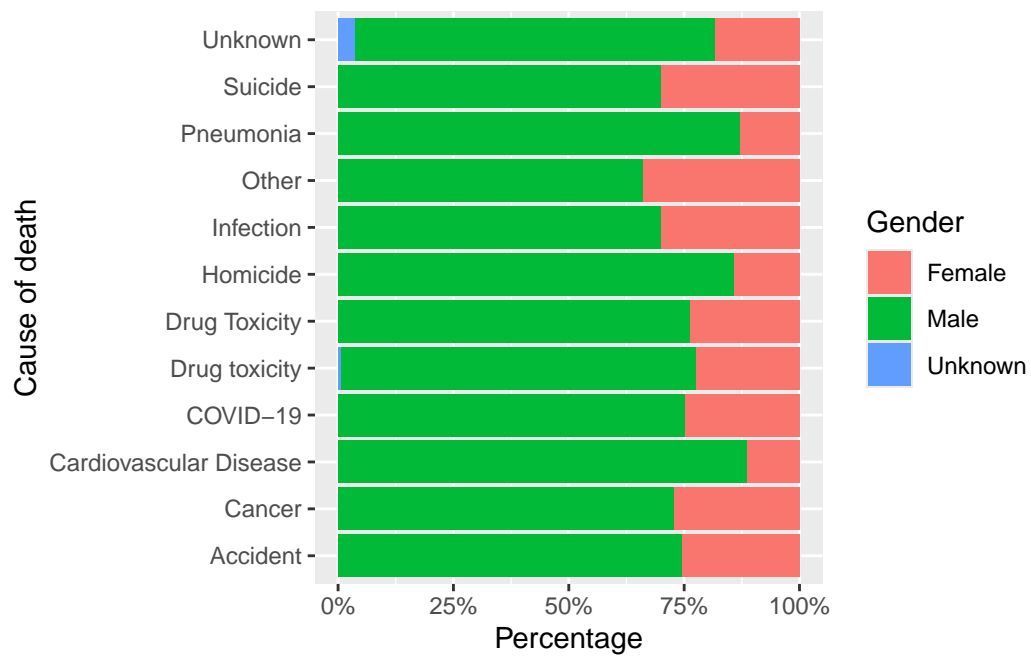

Figure 3: Number of deaths by different causes

Figure 4: Percentage of deaths by causes for different genders

References

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.

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, 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, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.