Toronto Death Registry*

Yunzhao Li

2024-01-22

In Toronto, there are many people died each year. The data of death registry supports the City's operational requirements and business functions. In this project we will make a table of number of death in Toronto for each month in 2023. By analyzing this, we can find months with high number of death and make conjectures with factors like weather, poicies or incidents of that particular month.

Table of contents

*Code and data are available at: https://github.com/yunzhaol/Toronto_Death_Registry.git.

1 Introduction

Death is a topic that people are not often willing to talk about, but it is closely related to municipal management and social resource allocation, so it is necessary to understand the data of death registration and do some basic statistical analysis of it. The data source in this study was obtained by opendatatoronto, from which the data of the Toronto area in 2023 were selected for research and analysis, hoping to produce a table reflecting the number of deaths registered in the Toronto area in each month of 2023, and correlate the data with related factors. Speculation may help in the formulation of future policies, such as whether the death toll is higher in the colder months, then the government can increase investment in the construction of shelters, if it is because of the epidemic or epidemic in certain months, then it can be prevented according to the relevant month and transmission trend, formulate a mask-related policy or appropriately restrict travel, or because of some large-scale accident that caused a large number of new deaths. Which months have a low number of deaths can also be related to the input of some public resources, the increase in per capita income, and the improvement of social welfare, which in turn can be used to judge whether these policies and improvements are significant, whether people's quality of life and income are stable, and whether social security is good. The government can regulate these factors to reduce the number of deaths caused by non-aging, which is the long-term significance of this study. This study will end with these numbers documenting the lives that have accompanied the city and ending their stories. The remainder of this paper is structured as follows. **?@sec-data**....

2 Data

2.1 Read raw data

2.2 Clean data

2.3 Tests

3 Results

The table we were looking for at the beginning of our research was modeled with a dataset, the raw data was read, the clean data was cleaned, and the data was accurately presented and consistent with the sketch. The number of deaths in each month is…. We found that the highest number of deaths was in May and June, with a high of around 2,000, and the lowest months were in February, July and December, with around 500. The death toll in the remaining months was concentrated around 1,000. Combined with the ideas introduced at the beginning of the article, this data can be combined with other relevant data to analyze, and perhaps unexpected discoveries can be obtained. This is the direction of future research that can be further explored, and this data can be used as a basis. You can also drill down into other years' data, or trends in total deaths over several years, to create a line chart.

The output of that code is Table 1

Table 1: Number of death in Toronto for each month in 2023

Month	Number of death
January	1083
February	485
March	1012
April	1677
May	2008
June	1928
July	417
August	912
September	1268
October	1182
November	1458
December	611

4 Conclusion

Overall, this study successfully plotted the number of deaths registered in Toronto by month in 2023, and achieved the expected results. The highest number of deaths was concentrated in May and June, with a high of around 2,000, and the lowest months were in February, July and December, with around 500. The death toll in the remaining months was concentrated around 1,000. It is believed that data can be used to assist in government management, policy formulation, and improvement of people's livelihood.

Appendix

A Additional data details

A.1 Simulation

Table 2: Dataset Simulation

time period	place_of_death	death_number
2023-01	Toronto	51
2023-02	Toronto	46
2023-03	Toronto	48
2023-04	Toronto	64
2023-05	Toronto	69
2023-06	Toronto	53

A.2 Read raw data

Table 3: Raw Death Registry

_id	CIVIC_CENTRE	DEATH_LICENSES	PLACE_OF_DEATH	TIME_PERIOD
19435	ET	69	Outside City Limits	2011-01
19436	ET	341	Toronto	2011-01
19437	NY	141	Outside City Limits	2011-01
19438	NY	540	Toronto	2011-01
19439	SC	129	Outside City Limits	2011-01
19440	SC	545	Toronto	2011-01

A.3 Clean data

Table 4: Cleaned Death Registry

time_period	death_licenses
2023-01-01	20
2023-01-01	1015
2023-01-01	48
2023-02-01	12
2023-02-01	446
2023-02-01	27

time_period	death_licenses

A.4 Tests

I wrote down 3 tests to check whether the cleaned death registry data suits our expectation. Separately test time_period's value, time_period's class and death_licenses's class.

B References

R Core Team (2022) Alexander (2024) Gelfand (2022) Wickham et al. (2019) Xie (2022) Grolemund and Wickham (2011) Firke (2023) Ushey and Wickham (2023) Hester et al. (2023) Müller and Walthert (2023) Angelis (2023)

Alexander, Rohan. 2024. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com.

Angelis, Inessa De. 2023. "Do More Subdivisions in Wards Increase Voter Turnout? An Analysis of the 2022 Toronto Municipal Election∗." https://github.com/InessaDeAngelis/Toronto_Elections.

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

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://sharlagelfand.github.io/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/.

Hester, Jim, Florent Angly, Russ Hyde, Michael Chirico, Kun Ren, Alexander Rosenstock, and Indrajeet Patil. 2023. *Lintr: A 'Linter' for r Code*. https://github.com/r-lib/lintr.

Müller, Kirill, and Lorenz Walthert. 2023. *Styler: Non-Invasive Pretty Printing of r Code*. https://github.com/r-lib/styler.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ushey, Kevin, and Hadley Wickham. 2023. *Renv: Project Environments*. https://rstudio.github.io/renv/.

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.

Xie, Yihui. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://cran.r-project.org/package=knitr.