Death in Toronto - Death Registry Statistics in Toronto Regions

Deaths in Toronto regions continues to rise due to global public health concerns

Yujun Jiao

5 February 2022

Abstract

The "Death Registry Statistics" data was drawn from the City of Toronto Open Portal to analyze the number of deaths in different regions in Toronto from January 2011 to January 2022. The data shows that the number of deaths has increased significantly since 2011, and a clear majority of the total deaths are from Etobicoke. This paper, along with other literary reviews, is to summarize the data, analyze the reasons for the increasing trend in the number of deaths and recommend further actions in improving public health.

Introduction[1]

Over the past few years, the constantly ongoing pandemic has evidently affected people's ordinary lives, increasing the threat to public health and global concerns. As people get more and more used to the isolating lifestyle, many have questioned whether it's necessary to take such strict health protocols. Over the past few years, people were often puzzled by the question: do people actually die as a result of Covid-19? In Open Data Toronto, the dataset called "Death Registry Statistics" offers the answer to this question. It provides real-life data about the number of deaths in four major Toronto regions in the past ten years and enables the public to get a clear sense of the current situation. By carefully studying the data, I obtained many key pieces of information to support my analysis.

Death statistics are essential in determining the effectiveness of the public health system. It shows the results of many government decisions such as vaccinations, hospitalization and preventive measures. By analyzing the results, people and the government can understand the severity of the issue and make appropriate adjustments in future plans. Therefore, the data allows the public to visualize the present health outcomes and helps society look at the trends in health over time. Using R studio (R Core Team 2020), this paper aims to describe the features in the dataset, analyze the dataset by providing meaningful graphs, and discuss the results. In a word,

---

[1] Code and data are available at https://github.com/bellajiao1999/Death_Toronto

this paper argues that COVID-19 is the major reason that leads to the increased number of total deaths.

Data

The dataset used in this paper is called "Death Registry Statistics," and it's obtained from the Open Data Toronto (Toronto Police Service 2020b). It records the death registration documented by four regions in Toronto: Etobicoke, North York, Scarborough and Toronto. It's created to support the Vital Statistics Act - a provincial legislation. The data is refreshed every month, and it was last updated on February 1st, 2022.

Originally, the dataset contained three independent variables and one outcome variable. In order to enhance the clarity of this report, I decided to perform data-cleaning using R Studio (R Core Team 2020). First, I acknowledged that the variable "PLACE_OF_DEATH" has only two outputs: Toronto and Outside City Limits. Since I want to analyze the health outcomes in Toronto regions, it's unnecessary to include data with "Out City Limits" because the results are hardly related to the Toronto health protocols. Also, as the data comes in pairs, the dataset includes both the number of deaths occurring within Toronto and deaths outside city limits for each individual civic center at each specific time. Thus, removing the rows where it's outside city control won't make my data incomplete. Then, I'd like to be less specific about the time period. Namely, I decided to remove the months in the time column and kept only the Year.

Now, the dataset contains four attributes: object id, civic center code, death licenses and Year. The object id is the numerical identifier. The civic center code is the categorical variable containing four categorical variables. The death licenses represent the count of deaths, and finally, the Year is also a categorical variable. The strength of this data is that it's very organized and easy to understand. Since it's updated every month, the dataset is also large, with 783 observations. Thus, it's easy to make graphs, plots and analyses using this dataset. Also, the two predictors I selected for this report are very responsive to the outcome variable - the number of deaths. Without a doubt, significant predictors are crucial in performing analysis and drawing conclusions. On the other hand, the data is not perfect because it has some data missing. Details about the missing data will be discussed later in the results section. A sample of data displayed in R studio is shown below.

```
##       X_id CIVIC_CENTRE DEATH_LICENSES Year
## 2    23669           ET            341 2011
## 4    23671           NY            540 2011
## 6    23673           SC            545 2011
## 7    23674           TO            297 2011
## 9    23676           ET            224 2011
## 11   23678           NY            307 2011
## 13   23680           SC            339 2011
## 14   23681           TO            223 2011
```

Results

As previously mentioned, the dataset shows an increasing trend in the number of deaths in Ontario regions. In order to prove it, I'll visualize the data and give the audience a clear sense of the trend. First, Figure 1 is a grouped scatterplot created using ggplot2 (Wickham 2016). It shows the number of deaths in each month and in each Toronto region from 2011 to early 2022. Thus, one little dot on the plot represents the number of deaths of a specific region in a specific month of the year. Since the plot doesn't give much information about the total number of deaths in that specific year, it's very general and simple. However, with this plot, we can still get a sense of the changes in the death registry in months. For instance, from figure 1, the number of deaths per month remains consistent from 2011 to 2019. Starting in 2020, the number of deaths per month starts to increase, and it reaches as high as 2007 in the year 2021. This number is unprecedented, and it shows an impact of the COVID-19 on the mortality of the population.

Figure 1: Number of Deaths from 2011 to 2022 in Toronto Regions
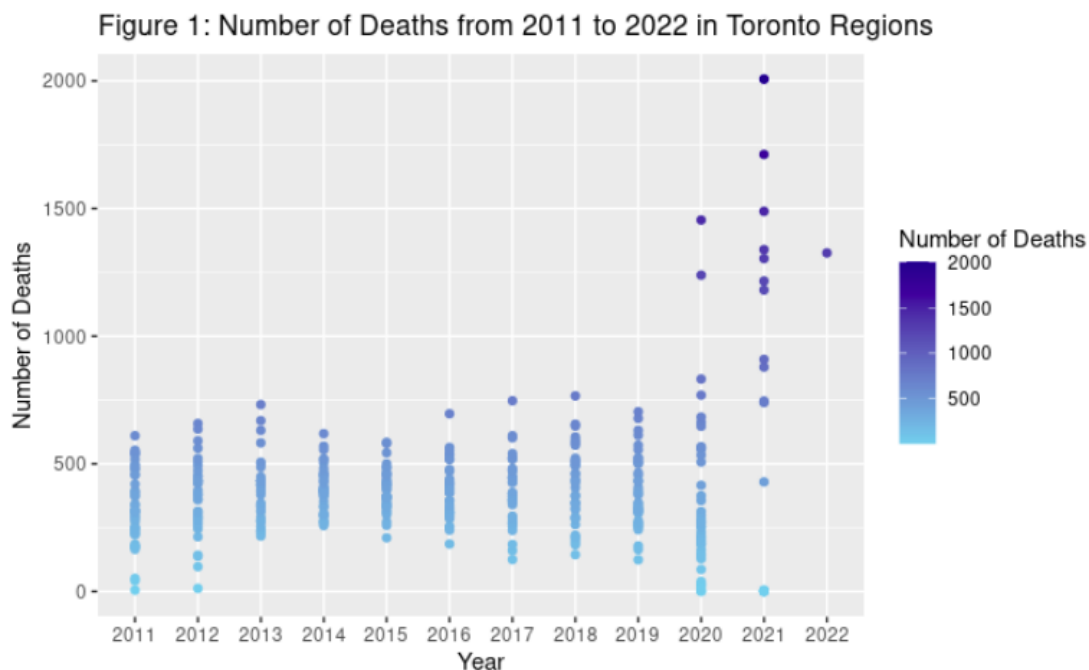
In regard to the detailed explanations of the relationship between the death registry statistics and public health concerns, I decided to make a table that shows an evident increasing trend. By employing the sum function in R, I calculated the total deaths in the regions from 2018 to 2021. Then, I created a table using the tidyverse (Wickham et al. 2019) and dplyr (Wickham et al. 2021) packages in R studio. As a result, it shows the number of deaths in four Toronto regions as well as the total number of deaths in the past ten years. The column "Total" lists the total number of deaths in that region from 2011 to 2022, and the Year 2018 displays the total number of deaths in 2018 in that region. For example, the total number of deaths in North York in the entire year of 2018 is 3607. I didn't include the year 2022 in the table because the data only contains the first month of 2022.

Figure 2: The Total Number of Deaths and the Number of Deaths per Year
In four Toronto regions

Civic Centre	Total	Year 2018	Year 2019	Year 2020	Year 2021
ET	63350	4811	4560	6128	13951
NY	38725	3607	3711	2026	6
SC	52109	6383	6166	3561	0
TO	3997	0	178	2621	13

More information is available at https://open.toronto.ca/dataset/death-registry-statistics/

From figure 2, it's obvious to see that the number of deaths has increased significantly since 2020. In 2018, when the pandemic had not yet started, the number of deaths in Etobicoke was 4811. After that, the augmentation of deaths is apparent: the number became as high as 13951 in 2021. However, many may have already noticed that the data for the Toronto region is weird: the number of deaths is 0 in 2018 and 178 in 2019. After I noticed this issue, I went back to the dataset and realized that the data for the Toronto region was missing in 2018 and 2019. This explains the unusually low numbers, and it's unpreventable when dealing with real-life data. Similarly, the data for Scarborough in 2021 is also absent. Hence, my table is not completely accurate since some data went missing.

Figure 3: Number of Death in Etobicoke in the Past Four Years
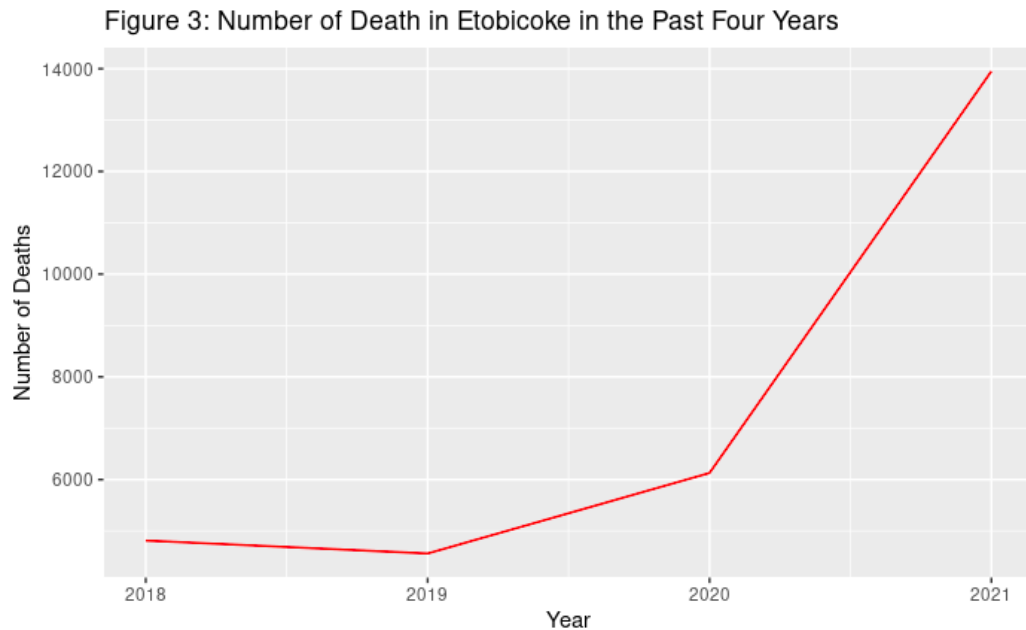
In figure 2, the table shows that the number of deaths has increased greatly in Etobicoke from 2018 to 2021. In order to visualize the trend, I made a line graph using the tidyverse package in R. From figure 3, Etobicoke experienced a dramatic increase in the number of deaths in 2021. This steep rise in the number of deaths proves that the pandemic, along with other public outcomes such as shortage of medical resources, mental illness and economic difficulties, have a huge impact on the mortality rate as well as the living standards.

In the last figure, I showed a sharp increase in the number of deaths in Etobicoke. Here's the question: do the other three regions in Toronto also experience a negative impact as a result of the pandemic? To answer this question, I want to visualize the differences in the number of deaths among different regions. In order to avoid the impreciseness in data, I decided to utilize only the data from 2020 because none of the data is missing in 2020. Figure 4 is a pie chart I made using the pie function in R. In the chart, it's evident to see that the portion coloured in brown occupies almost half of the pie. Thus, Etobicoke has the highest number of deaths among all four regions in 2020. Besides, Scarborough ranked 2nd, Toronto ranked 3rd, and North York ranked 4th in the chart.

Figure 4: Pie Chart for Number of Deaths in Different Regions in 2020
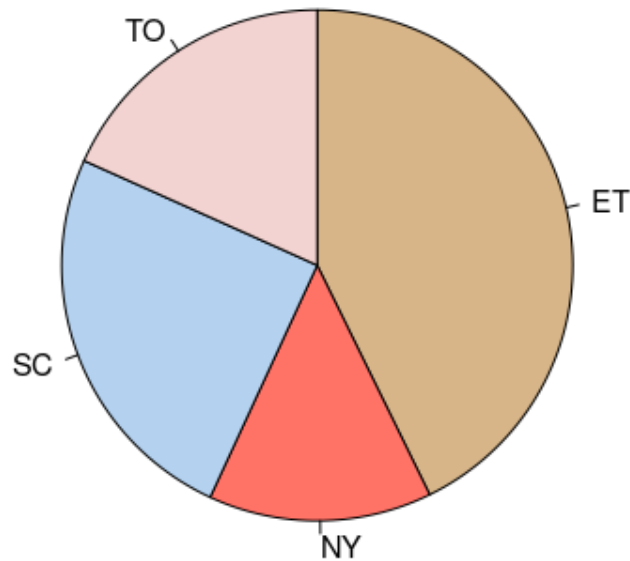
Discussion

With the figures shown in the previous section, the number of deaths has increased significantly since 2011. Among the four regions, Etobicoke contributes to most of the deaths, and it has shown a sharp increase in the number of deaths in 2020. From figure 1, it's obvious that the pandemic is the major cause of the rise in deaths registrations, as it has far exceeded the year-to-year variations in the past ten years. Another factor contributing to the augmentation in numbers is that the COVID-19 increases the mortality rate among the old population and raises the risk of death among younger generations. According to Dr. Ge, comorbidities were often associated with a higher risk of mortality in individuals under fifty years old, and "individuals with five or more comorbidities in the below fifty years age group had 395.44 times higher risk of mortality compared to those without. (Ge et al. 2021)." Therefore, as the whole population is threatened by the ongoing pandemic, the steep rise in the death numbers is well explained.

However, it's unwise to conclude that the dramatic trend is due to the disease alone from the limited information. Thus, there exist many indirect consequences of COVID-19 that affect all age groups. According to Statistics Canada, the indirect consequences are responsible for the excess mortality in the past few years, and there are 5535 more deaths than expected in 2021. These unintentional deaths are led by various reasons such as accidental poisonings, alcohol-induced mortality and mental and behavioural disorder.

Conclusion

In a word, the death data registry is one of the major pieces of evidence that scholars use to determine the effectiveness of current health protocols, government actions and preventive measures. This report argues that the number of deaths has increased dramatically in the past two years, resulting from both the disease itself and changes in people's daily routines. As we're still in early 2022, it's important to educate, motivate and impress the general public to take action.

References

Data, Toronto Open. 2020. "Death Registry Statistics."
https://open.toronto.ca/dataset/death-registry-statistics/.

Ge, Erjia, et al. 2021. "Association of Pre-Existing Comorbidities with Mortality and Disease Severity among 167,500 Individuals with Covid-19 in Canada: A Population-Based Cohort Study." https://doi.org/10.1371/journal.pone.0258154.

Government of Canada, Statistics Canada. 2021. "Provisional Death Counts and Excess Mortality, January 2020 to April 2021." *The Daily.*
https://www150.statcan.gc.ca/n1/daily-quotidien/210712/dq210712b-eng.htm.

R Core Team. 2020. 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, Romain François, Lionel Henry, and Kirill Müller. 2021. Dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr.