Differences in COVID Fatality Across Various Populations within Toronto*

Tara Chakkithara

March 16, 2024

Abstract

Throughout history, human societies have confronted many communicable diseases but the COVID-19 pandemic stands out significantly, as being the first pandemic met with a structured and technologically advanced response. This paper examines the varying prevalence of the illness among different populations in Toronto. Demographics such as age, neighbourhood, and medical attention were considered. A linear model was constructed to infer COVID-19 fatality based on this data where varaibles such as age over 90 and hospitalization were especially significant in predicting fatality.These results highlight target areas that could be used to improve public health policies in the future.

Table of contents

1 Introduction

One of the most famous occurrences of quarantine was in Eyam, England, during the Bubonic plague pandemic. The villagers of this small town decided to isolate within their community rather than escape elsewhere to avoid contagion, hoping to prevent the spread of the disease. Their sacrifice succeeded in containing the disease, but unfortunately, most infected individuals in the town passed away (Ross 2022).

In the modern era, COVID-19 was the first pandemic equipped with mass communication and innovative technology, and the outcomes of this pandemic were far better than pandemics of the past. However, there were discrepancies in COVID-19 fatality and prevalence among different demographics. By identifying

---

*Code and data available at https://github.com/pixel-echo/covid-trends

vulnerable groups and locations, public health facilities can continue to make better and more informed policy decisions.

This paper presents a case study of Toronto and analyzes the relationships that age, neighborhood, and access to medical attention have with COVID-19 fatality. The analysis examines each variable individually and its relationship to fatality. Additionally, a model is constructed to estimate the risk of dying if one contracts COVID-19.

2 Data

The data for this paper was sourced from Open Data Toronto and published by Toronto Public Health. It contains records of COVID-19 from January 2020 to January 2024 within the Toronto region, consisting of variables such as gender, age group, neighborhood, health outcomes, etc. This data set was selected over other similar data sets due to the credibility and extensive outreach of Toronto Public Health.

Raw data was cleaned, truncated, and visualized with the help of multiple R packages (R Core Team 2023). A clean data set focusing on variables of interest such as year, age group, neighborhood, outcomes, and health history indicators like hospitalization, intubation, and ICU stay was generated using the `tidyverse` package (Wickham et al. 2019). The cleaning process involved mutating data entries for enhanced clarity, removing unconfirmed COVID-19 cases, and filtering out entries outside of the 2020-2023 time period. No additional variables were constructed during this process.

Following the cleaning process, the distribution and trends of each variable were visualized using graphs, tables, and maps. Graphs were created using the `ggplot2` package in R (Wickham 2016), tables were created using `knitr` (Xie 2023), and maps were created using `sf` (Pebesma and Bivand 2023). These visualizations were then assessed and explained, taking into account broader contexts such as demographics and pandemic measures.
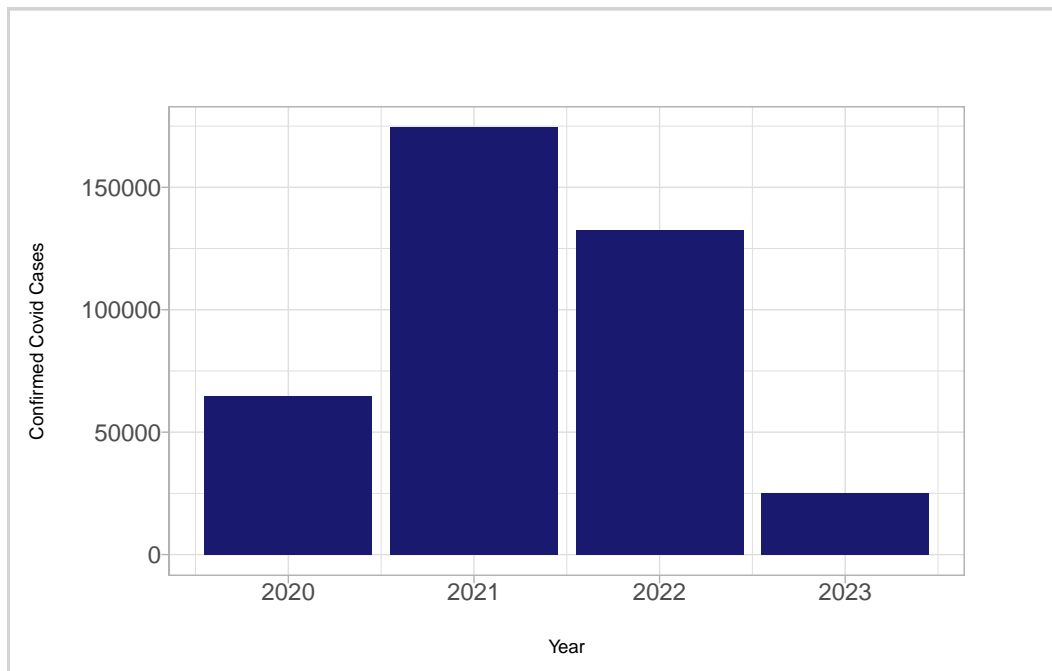

Figure 1: Distribution of Covid Cases by Year

The distribution of COVID-19 cases from 2020 to 2023 offer us insights into general trends in COVID-19 activity within the Toronto region. The first case in Toronto was reported on January 23 2020. In the subsequent year the pandemic reached its peak with a total of 174,465 cases confirmed by Toronto Public Health. By 2023 the pandemic had reached its lowest point with only 25,199 confirmed cases. This distribution follows a bell-shaped curve as depicted in Figure 1. These trends are a reflection of Toronto's pandemic measures. COVID vaccination in Ontario started in at the end of 2020 and vaccination rates reached 80% by August 2022 (Lysyk, n.d.), helping drive COVID-19 incidence down in 2023.
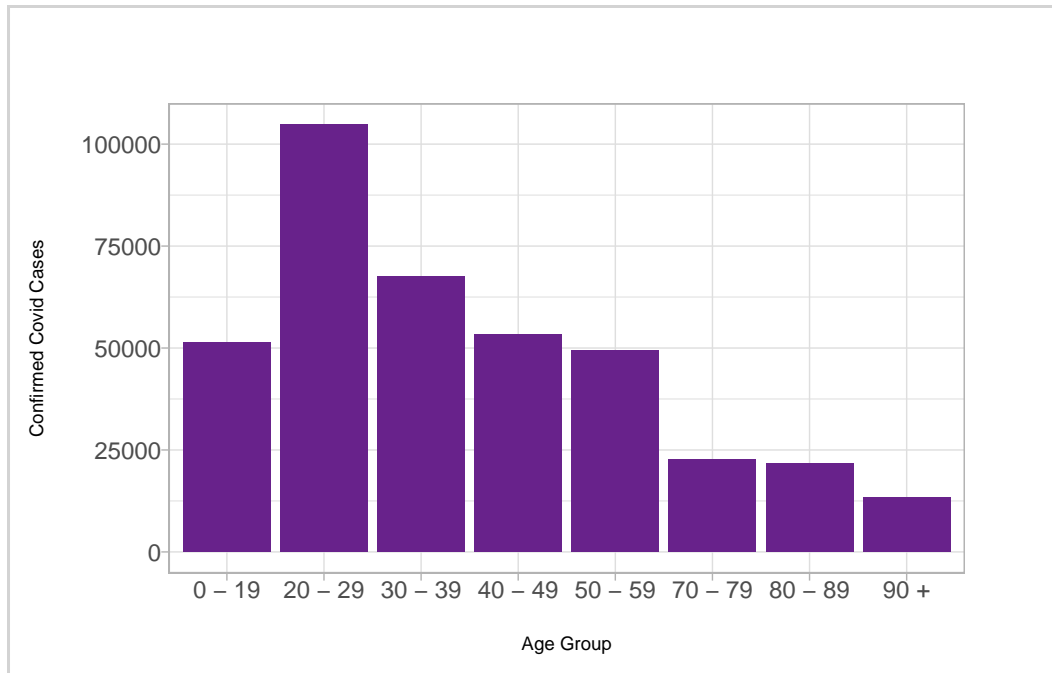

Figure 2: Distribiution Covid Cases

The distribution of age group is visualized in Figure 2. During the pandemic the age group 20-29 faced the highest rates of COVID incidence. This group had a total number of 108, 953 COVID cases from 2020 - 2023. While the age group with the least incidence was the 90+ group. This may be due to Toronto's demographics. In 2021 only 2.6% of the population was over the age of 85 while 40% of the population has an age from 25 to 40 (n.d.).

Table 1: Medical Attention Needed During The Pandemic

Intubated	Hospitalized	ICU	Cases
No	No	No	378423
No	Yes	No	15662
No	Yes	Yes	1338
Yes	Yes	Yes	1735

The indicators variables intubated, hospitalized, and ICU tracks if a person required severe medical attention during their COVID case. Most cases in Toronto did not require severe medical attention but a total number of 18735 required some type of medical attention as seen in Table 1.
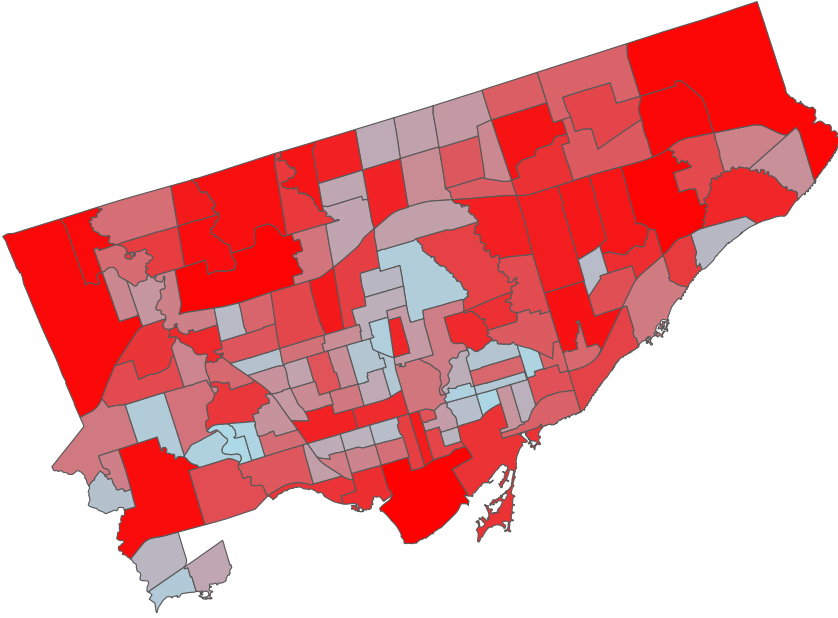
Different neighborhoods in Toronto saw different rates of incidence. Areas at the heart of Toronto experienced fewer cases of COVID while areas close to the outskirts of Toronto experienced a lot more. The three neighborhoods that had the highest rates of COVID-19 were Downsview, Woburn, and the Waterfront Communities. While the three neighborhoods that saw the lowest rate of incidence were Lambton Baby Point, Blake - Jones, and Runnymede.

Finally, the last variable studied was the outcome variables. It denotes the health outcomes of patients. Patient cases were either resolved, fatal, or they are still active. As seen in Table 2, most cases of COVID-19 were resolved but around 5089 cases were fatal.

Table 2: COVID Outcomes in Toronto From 2020 - 2023

Outcome	Count
Active	105
Fatal	5089
Resolved	391964

3 Measurement

Although Open Data Toronto and Toronto Public Health are both reliable sources no measurement is 100% accurate. However measurement within this data set is accurate since most variables are categorical variables that we can see. For example, its easy to measure whether or not a person receives additional medical attention, where they live, what year they catch COVID-19. This is the data set's strength. One way measurement may be lacking is due to under reporting. People may have COVID but that doesn't mean they will get tested and report their case to Toronto Public Health. So this data set may have a bias of reporting more serious cases of COVID compared to ones that were less serious.

4 Model

Logistic regression was used to model COVID fatality risk given factors such as elderly age, hospitalization, intubation, and ICU stay. Before training the logistic model a smaller data set was created by transforming the categorical variables into indicator variables. This step was essential in training the model.

The data set was then randomly split into 70% training data and 30% testing data to evaluate the model's performance. The logistic regression model was trained using the training data set, and its predictive performance was assessed on the testing data set. The logistic regression equation is as follows :

$$logit(x) = -6.45 + 2.43X_1 + 3.45X_2 + 1.621X_3 + 1.55X_4 + 1.91X_5 + 2.85X_6$$

$$X_1 \sim \text{Hospitalization, } X_2 \sim \text{Age Above 90, } X_3 \sim \text{ICU}$$

$$X_4 \sim \text{Intubated, } X_5 \sim \text{Age Group 70 to 79, } X_6 \sim \text{Intubated, Age Group 80 to 89}$$

A threshold of 0.5 was used to classify predicted probabilities into fatal and non-fatal cases of COVID-19. Probabilities greater than 0.5 were classified as fatal cases, while probabilities less than 0.5 were classified as non-fatal cases. This model had a test accuracy of 98.7%.

5 Results

Using our model results we can infer that elderly age and hospitalization are significant risk factors of COVID fatality. We can visualize the relationships between fatality and different demographics using the same packages as before.

When we analyzed the yearly distribution of COVID 19 it had a bell shaped curve, however COVID fatality decreased every year during the pandemic as seen in Figure 3.

Although Toronto has a higher population of Adults than Seniors they have the highest rates of COVID mortality as seen in Figure 4

Toronto's neighborhoods seem to have high fatality where there is high incidence as. The variable by itself wasn't that significant to COVID mortality.
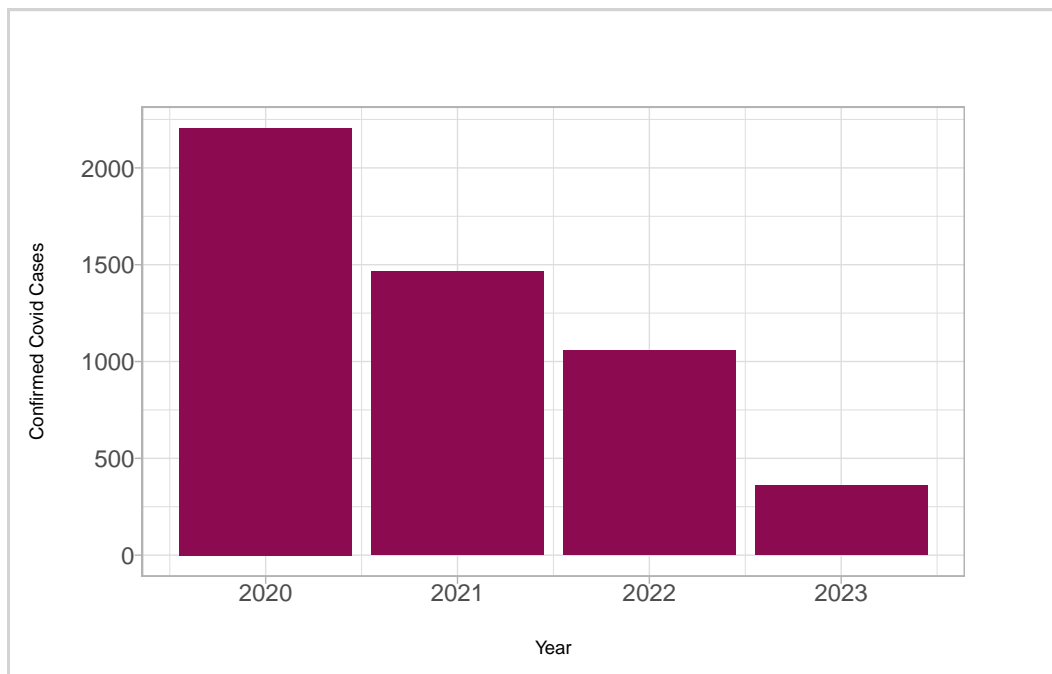
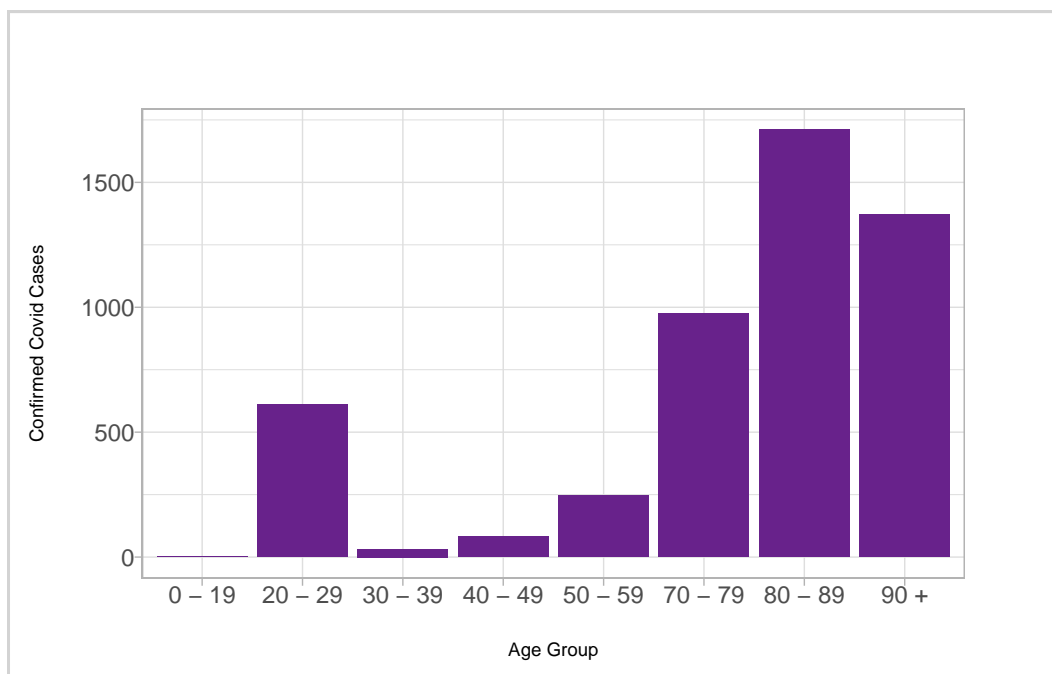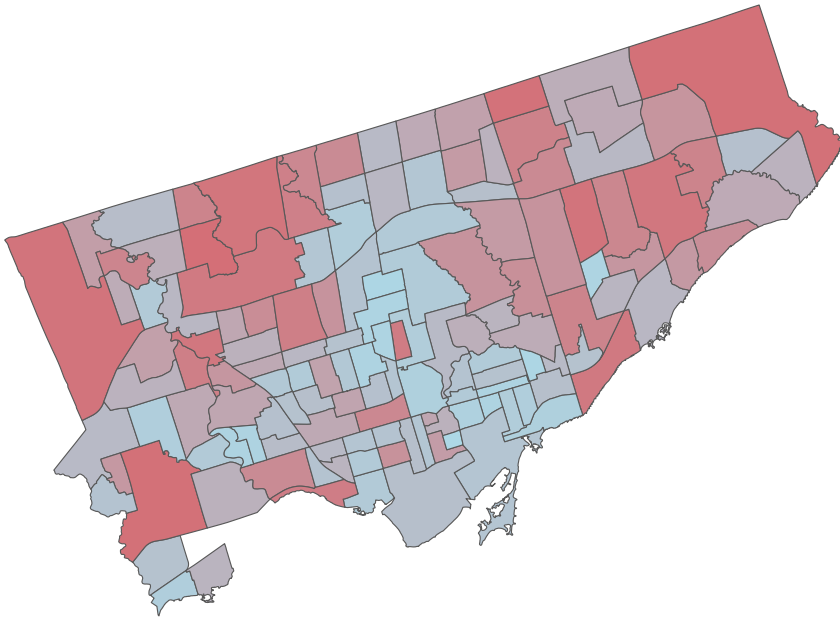Figure 3: COVID Mortality in Toronto From 2020 - 2023

Figure 4: COVID Mortality Across Different Neighbourhoods

6 Discussion

An interesting factor noticed during the creation of the linear model was that neighborhood data wasn't a significant predictor of fatality. Neighborhood data may indicate COVID prevalence but it didn't seem to affect fatality much. Some weaknesses of the model constructed are that only a few demographics are considered. Demographics such as race and gender were not considered. However we still learned about the significance of age and hospitalization when it comes to COVID fatality.

In the future further neighborhood analysis and other demographics may be analyzed to continually improve public health policy. Being prepared is better for everyone.

Appendix

n.d. *Toronto.ca.* City of Toronto. https://www.toronto.ca/wp-content/uploads/2022/04/9654-City-Planning-2021-Census-Backgrounder-Age-Sex-Gender-DwellingType.pdf.

Lysyk, Bonnie. n.d. "2022 Annual Report: Covid-19 Vaccination Program." *Auditor.on.ca.* Office of the Auditor General of Ontario. https://www.auditor.on.ca/en/content/annualreports/arreports/en22/AR__COVIDVaccination__en22.pdf.

Pebesma, Edzer, and Roger Bivand. 2023. *Spatial Data Science: With applications in R.* Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016.

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

Ross, Eleanor. 2022. "The Village That Stopped the Plague." *BBC News.* BBC. https://www.bbc.com/travel/article/20151026-the-sleepy-village-that-stopped-the-black-death.

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.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.