

# Varying Demographics and Geography of the Toronto Population leads to differences of COVID-19 Cases in Toronto\*

Jacob Yoke Hong Si

06 February 2022

## Abstract

Toronto COVID-19 data was obtained from the City of Toronto Open Portal to determine how different age groups and neighbourhoods in Toronto affected the number of COVID-19 cases. From our data analysis, we determined that there is an obvious difference in the number of covid cases in these variables where specific age groups and neighbourhoods tend to have higher covid cases. With the information obtained, we could use it to implement tighter lockdown measures in these areas to curb the impacts of Covid.

## 1 Introduction

Since 2019, the global pandemic known as COVID-19 has affected the lives of Canadians and the world, leading to a recession in the economy, unemployment, poverty, hospitalizations and deaths. With recent COVID cases rising in Toronto, I hope to identify the location and age groups that are susceptible to COVID and curb its effect. A search on the National Center for Biotechnology Information and The New England Journal of Medicine yielded three research papers titled “The impact of vaccination on COVID-19 outbreaks in the United States” (Moghadas 2021), the “Effect of Vaccination on Household Transmission of SARS-CoV-2 in England” (Harris 2021) and the “Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine” (Polack 2020). These papers indicated that vaccination rates reduced the infection rate of COVID-19 in their samples. Furthermore, research (Yang 2021) has also proved that implementing lockdown measures would help reduce the spread of COVID-19. Hence, I hope to highlight the part of the population that would benefit most from reducing the spread of COVID-19.

In the paper, I will use the Covid cases in Toronto from the City of Toronto to explore the age groups and neighbourhoods that is impacted most by COVID-19. This will be done through first providing summary statistics of our dataset, then creating tables and plots that depicts the number of Covid cases with respect to the neighbourhoods in Toronto and age groups. This would provide us with a clear picture of the demographics and geography of Covid in Toronto.

The remainder of this paper is structured as follows: Section 2 discusses the data.

## 2 Data

In order to analyse the data of Covid cases in Toronto over the pandemic such as the Covid Cases by Age Groups and Neighbourhoods, I obtained the COVID-19 Cases in Toronto dataset from the City of Toronto Open Data Portal using the `opendatatoronto` package. The data is collected by Toronto Public Health

---

\*Code and data are available at: [https://github.com/jacobyokehongsi/COVID-19\\_Cases\\_in\\_Toronto](https://github.com/jacobyokehongsi/COVID-19_Cases_in_Toronto).

Table 1: First ten rows of a dataset that shows date of episode, neighbourhood, age group and gender of an individual with Covid

Episode Date	Neighbourhood	Age Group	Gender
2021-12-27	Rouge	30 to 39 Years	MALE
2021-12-25	Scarborough Village	30 to 39 Years	MALE
2021-12-24	Bay Street Corridor	40 to 49 Years	FEMALE
2021-12-28	Bay Street Corridor	50 to 59 Years	FEMALE
2021-12-22	Wexford/Maryvale	20 to 29 Years	FEMALE
2021-12-25	Cliffcrest	20 to 29 Years	FEMALE
2021-12-25	Scarborough Village	19 and younger	MALE
2021-12-24	Humber Heights-Westmount	60 to 69 Years	FEMALE
2021-12-20	NA	30 to 39 Years	FEMALE
2021-12-26	High Park-Swansea	20 to 29 Years	MALE

Table 2: Number of Covid Cases in Different Age Groups

Age Group	Number of Covid Cases
19 and younger	41599
20 to 29 Years	57571
30 to 39 Years	50774
40 to 49 Years	39179
50 to 59 Years	36055
60 to 69 Years	22164
70 to 79 Years	10966
80 to 89 Years	8168
90 and older	4486

and contains demographical and geographical information regarding the covid outbreaks which was last updated on the 26th of January 2022. A potential weakness of the source is that there could be individuals who contracted COVID but do not report to clinics or hospitals thus, leading to an underrepresentation of our data. This would also bias the data since we underestimate the number of people who contracted covid due to the unreported cases. The data obtained contain age groups ranging from 19 and younger to 90 and older, as well as 140 different neighbourhoods in Toronto where Covid outbreaks occurred. Using the statistical programming language R (R Core Team 2020), as well as packages including `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016) and `janitor` (Firke 2021). I conducted data cleansing and data visualisation to depict the data in appropriate tables and plots.

Starting with the raw data, I first conducted data cleansing by cleaning the names of the columns. I then selected columns in the dataset that are relevant to our analysis which includes the covid episode date, neighbourhood, age group and gender. Table 1 presents the data which outlines the variables I have selected.

As I investigate the relationship between the number of covid cases with the age groups and neighbourhoods, I grouped the data with respect to these two variables and obtained a count for the covid cases. Lastly, since there are many neighbourhoods (140), I decided to select the ten neighbourhoods with the highest covid cases to conduct my analysis. Table 2 represents the relationship between age groups and covid cases whereas Table 3 represents the relationship between neighbourhoods and covid cases.

Table 3: Number of Covid Cases in Different Neighbourhoods

Neighbourhood	Number of Covid Cases
Woburn	6448
Downsview-Roding-CFB	6117
Waterfront Communities-The Island	6046
Malvern	5554
Rouge	5492
West Humber-Clairville	5457
Mount Olive-Silverstone-Jamestown	5268
Glenfield-Jane Heights	4729
York University Heights	4411
Islington-City Centre West	4130

Utilizing the data, I will generate bar charts to depict the covid cases at different age groups and neighbourhoods.

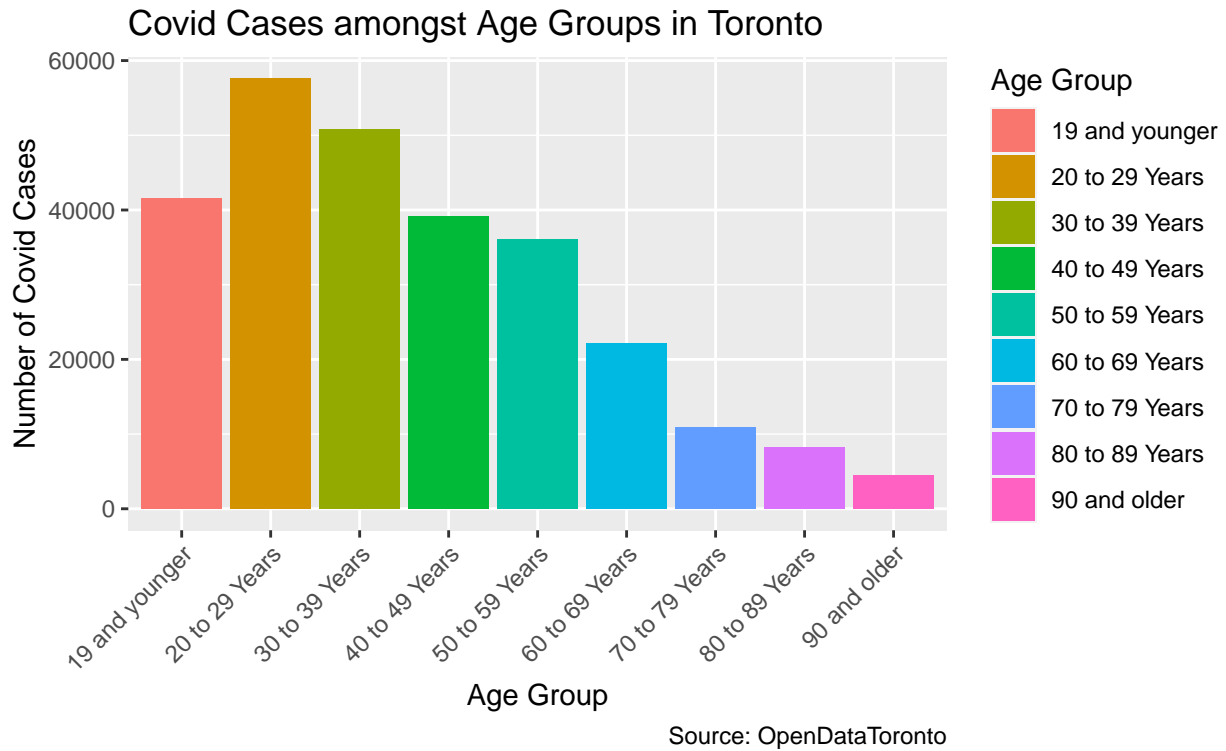


Figure 1: Covid Cases amongst Age Groups in Toronto

Figure 1 indicates that the age group between 20 to 29 Years have the highest number of covid cases whereas the age group between 90 and older to have the least. This would be due to the fact that despite covid restrictions, young adults are less wary of it and take lesser precautions against covid compared to the older adults. Additionally, the older population also tend to stay home compared to young adults, causing the young adults to be more susceptible to covid.

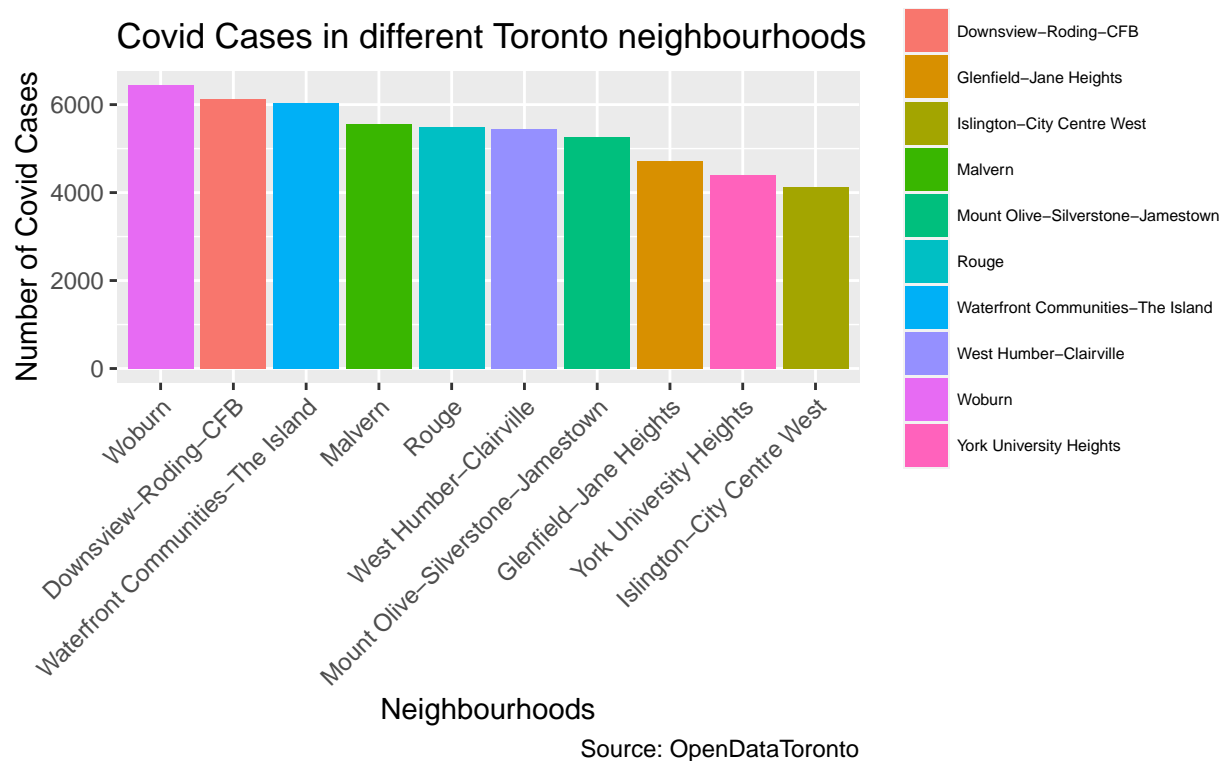


Figure 2: Covid Cases in different Toronto neighbourhoods

On the other hand covid cases in the neighbourhoods depicted in Figure 2 indicates the neighbourhoods with the most number of covid cases in Toronto. Neighbourhoods such as Woburn have total covid cases of up to approximately 6500 which is very high. Thus, in order to curb the effects of covid, we should enforce tighter restrictions in these areas to prevent further covid outbreaks.

## References

- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Harris. 2021. *Effect of Vaccination on Household Transmission of SARS-CoV-2 in England*. <https://www.nejm.org/doi/full/10.1056/nejmc2107717>.
- Moghadas. 2021. *The Impact of Vaccination on COVID-19 Outbreaks in the United States*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/>.
- Polack. 2020. *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine*. <https://www.nejm.org/doi/full/10.1056/nejmoa2034577>.
- 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, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Yang. 2021. *Does City Lockdown Prevent the Spread of COVID-19? New Evidence from the Synthetic Control Method*. <https://ghrp.biomedcentral.com/articles/10.1186/s41256-021-00204-4>.