Who were most affected by COVID-19 in Toronto*

Lantao Cui

07 February 2022

Abstract

The first COVID-19 case was reported on 31 December 2019, exactly two years from now. At first, everyone thought that this is nothing but a seasonal flu and it will end in less than a month, however, it turned out to be a global pandemic and has taken numerous people's lives. In this paper, We obtain and analyze all reported COVID-19 cases in the city of Toronto. We found out that COVID-19 is more fatal toward elders and hits hardest on communities with low income or high population density. Our findings will help the city of Toronto to better help those people at needs.

1 Introduction

Ever since the dawn of civilization, humanity had been fighting against different plagues. The earliest recorded pandemic happened during the Peloponnesian war, which was almost 3000 years from now. This plague wiped through all of Libya, Ethiopia and Egypt, killed almost two-thirds of the entire population. Another well known plague is the black death, which lasted for decades and was responsible for the death of nearly one-thirds of the World Population. But after the industrial revolution, the technology advanced so fast and people developed the illusion that we should be able to defeat any kinds of plague. Therefore, nobody cared at all when COVID-19 first started, until it started to take away millions of people's lives.

To find out who was most affected, we obtained all of COVID-19 case reports from the City of Toronto's data portal. After some investigation, we found out that as the patient's age increases, the mortality and hospitalization rate also increases exponentially. The mortality rate of kids under 19 years old is 0.00007 and this number increases to 0.23 for people above 90 years old, by a factor of 3000. On top of that, low income or high density neighbourhood are hit the hardest during this pandemic. Among 140 different communities, the top 10 communities with the most COVID-19 cases have contributed 20% of Toronto's total COVID-19 cases.

Our findings have lots of implications, it identifies the groups of people and neighborhood that are most affected by COVID-19, this information will allow the government to better support them and come up with new proceudres that will slow down the spread of COVID-19.

2 Data

2.1 Data Source

This dataset contains all of the data reported to Toronto Public Health and are gathered through an R package called "opendatatoronto" (Gelfand 2020). This dataset contains all of the demographic, geographic, and severity information for all of the confirmed and potential cases of COVID-19. It's updated on a weekly basis and was last updated on Feb 2, 2022.

2.2 Data Collection and Methodology

All of COVID-19 cases are reported to Toronto Public Health by local testing laboratories, then Open data toronto collect those data from the provincial Case & Contact Management System. The strength of this

*Code and data are available at: https://github.com/cuilantao/Impace-of-COVID19-on-Toronto.

dataset is that it is very detailed and contains a lot of valuable information, for instance, for every reported case, this dataset will track it and include the final result. This can help us to analyze the mortality rate of COVID-19 on different age groups. However, although the dataset contains all of the reported cases, it can not be viewed as an accurate measure of the actual COVID-19 cases because of various reasons. The most notable problem is that you have to get tested to be officially diagnosed with COVID-19. This leads to several problems, first of all, COVID-19 has similar symptoms as a seasonal flu or a common cold, so some people will refuse to get tested. Another problem is the shortage of testing kits, according to the Ontario Government, due to the fast spread of the Omicron variant, PCR tests are only available to "high-risk individuals and individuals who work at high-risk settings." Therefore, those numbers should only be interpreted as "an underestimate of the true number of individuals with COVID-19 in Toronto."

2.3 Data overview

This dataset contains all of the demographic, geographic, and severity data of the reported COVID-19 cases in the City of Toronto. It has a total of 272017 cases and has 18 attributes. The first two attributes serves as a distinct identifier and are removed in the data pre-processing step. The next 5 attributes contain information about the age of the patient, their source of infection and the neighborhood they live in. The next 5 columns are about the gender of the patient, whether case is probable or confirmed, and the outcome. The final 6 columns help to assess the severity of the case. To get a better understanding of the actual data, let us take a look at Figure 1, which is a sample view of the data and is produced by R (R Core Team 2020);

```
## # A tibble: 6 x 16
##   `Outbreak Associated` `Age Group`    `Neighbourhood N~` FSA   `Source of Inf~`
##   <chr>                 <chr>          <chr>              <chr> <chr>
## 1 Sporadic              50 to 59 Years Willowdale East    M2N   Travel
## 2 Sporadic              50 to 59 Years Willowdale East    M2N   Travel
## 3 Sporadic              20 to 29 Years Parkwoods-Donalda  M3A   Travel
## 4 Sporadic              60 to 69 Years Church-Yonge Corr~ M4W   Travel
## 5 Sporadic              60 to 69 Years Church-Yonge Corr~ M4W   Travel
## 6 Sporadic              50 to 59 Years Newtonbrook West   M2R   Travel
## # ... with 11 more variables: Classification <chr>, `Episode Date` <date>,
## #   `Reported Date` <date>, `Client Gender` <chr>, Outcome <chr>,
## #   `Currently Hospitalized` <chr>, `Currently in ICU` <chr>,
## #   `Currently Intubated` <chr>, `Ever Hospitalized` <chr>,
## #   `Ever in ICU` <chr>, `Ever Intubated` <chr>
```

2.4 The effect of COVID-19 on different age groups

People always have the illusion that younger people are less affected from disease because they have a stronger immune system. However, this is not always the case, during the 1918 flu, millions of healthy young people died, this is because the 1918 flu triggered a cytokine storm. Basically this will turn your own immune system against you, as a result, there were less deaths in older and middle-aged adults because they have a weaker immune system. To see if COVID-19 has a simular effect, Figure 1 made by ggplot2(Wickham 2016), Table 1 made by KableExtra (Zhu 2021) and Knitr (Xie 2014) shows total number of confirmed COVID-19 cases among each age group, and the severity of those COVID-19 cases.

From the figure and table, it's clear that there is some correlation between people's age and the effect of COVID-19. As people's age increase, the mortality and hospitalization rate grow almost exponentially, the mortality rate of kid under 19 years old is 0.00007, but this number increases to 0.23 when the person is older than 90, which is a factor of 3285. Study also shows that "adults over 65 years of age represent 80% of hospitalizations and have a 23-fold greater risk of death than those under 65" (Mueller, McNamara, and Sinclair 2020). The reason for this is that as people ages, their immune system function will gradually decline, which makes them more vulnerable to virus. Another reason is that older people are more likely to have underlying health conditions such as diabetes and high blood pressure. Those health condition will worsen the symptom of COVID-19 and make it more fatal.

Table 1: Effect of COVID-19 on different Age Group

Age Group	Total COVID-19 Cases	Fatal Cases	Mortality Rate	Hospitalized	Hospitalized Rate
19 and younger	40767	3	0.00007	261	0.00640
20 to 29 Years	56778	11	0.00019	456	0.00803
30 to 39 Years	49857	30	0.00060	738	0.01480
40 to 49 Years	38461	76	0.00198	1075	0.02795
50 to 59 Years	35422	201	0.00567	1865	0.05265
60 to 69 Years	21676	485	0.02237	2423	0.11178
70 to 79 Years	10628	765	0.07198	2491	0.23438
80 to 89 Years	7854	1315	0.16743	2477	0.31538
90 and older	4274	1017	0.23795	1117	0.26135

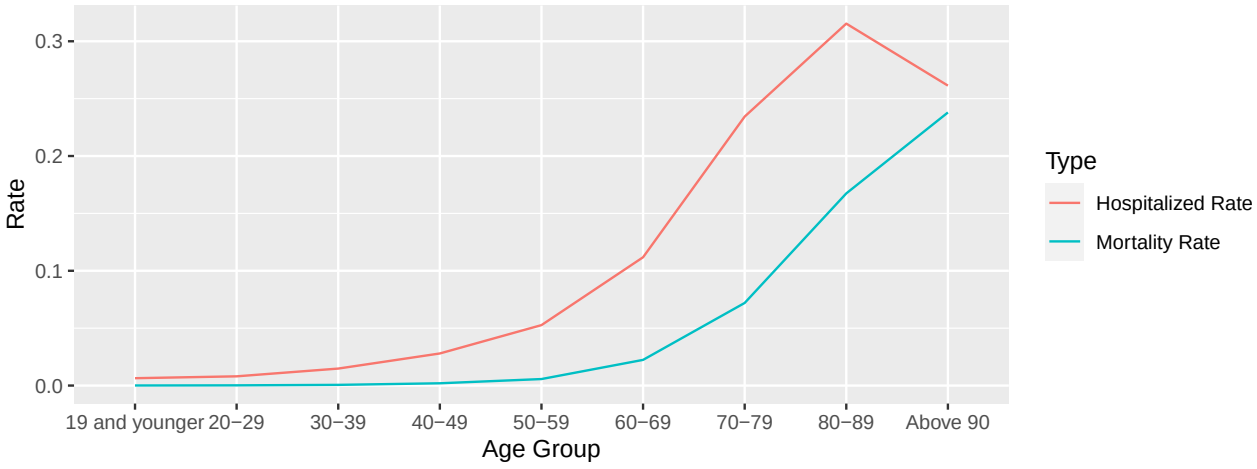

Figure 1: Effect of COVID-19 on different Age Groups

2.5 The effect of COVID-19 on different communities in Toronto

The next thing we will explore is the effect of COVID-19 on different communities in Toronto. After grouping by and sorting the data using R (R Core Team 2020), Figure 2 is produced with ggplot2 (Wickham 2016). The figure plots the top 10 communities with the most COVID-19 infections, a quick research shows those are mostly low-income communities or neighborhood that have a large population density. This is because most jobs with low income require the worker to have much more in-person interactions with other people, thus increasing the risk of COVID-19 infection. And a high density neighborhood usually has a lot of apartment buildings, and it's relatively easier for the virus to spread within an apartment building because people are sharing lots of things such as air vent and elevator. This finding shows that COVID-19 situations varies across different communities and it maybe beneficial if the government starts to enforce different procedures based on different communities.

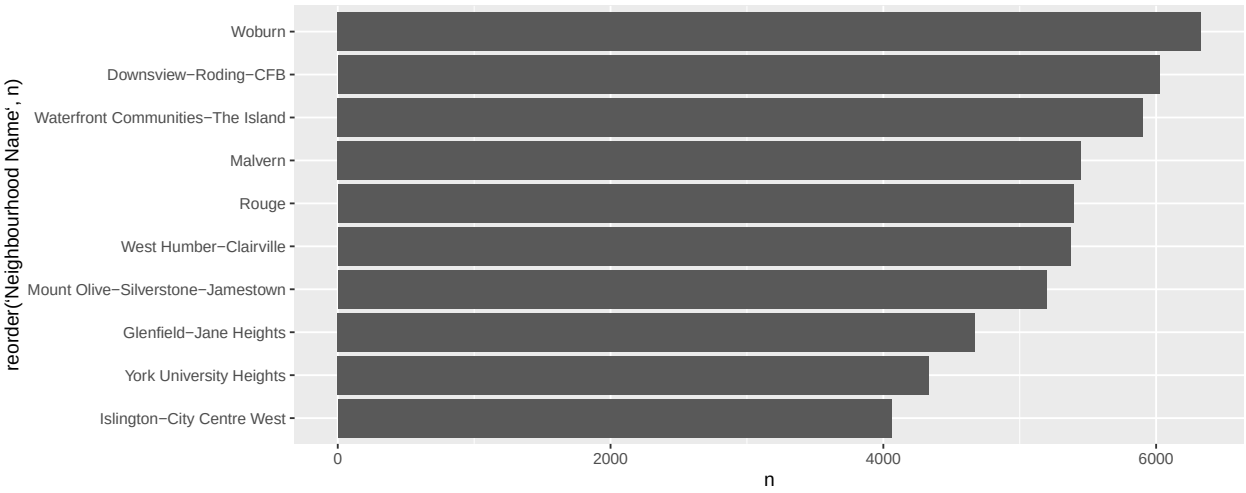

Figure 2: Effect of COVID-19 on different communities

2.6 Conclusion

Our findings clearly revealed that COVID-19 hits certain people and neighborhood harder and they require more attention from the government. Luckily, the City of Toronto is aware of this situation and have launched support programs targeting those neighborhood. For instance, the City of Toronto is offering free self-isolation sites for those people diagnosed with COVID-19 and is also giving seniors and people with underlying health condition top priority in the vaccination program.

References

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.

Mueller, Amber L, Maeve S McNamara, and David A Sinclair. 2020. "Why Does Covid-19 Disproportionately Affect Older People?" *Aging*. Impact Journals. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7288963/.

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.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.