COVID-19 Cases in Toronto*

Examine All cases from January 2020 to January 2021

Xinyi Xu

27 January 2021

Abstract

COVID-19 has been recognized as a global threat; thus, numerous studies have been conducted in order to control and prevent this pandemic. I am going to use the dataset from Toronto Public Health to summarize and analyze data to find how the pandemic is evolving, and assess severity and risk in Toronto. It is crucial for the Ministry of Health and our local public health institution to have an in-depth understanding of the disease and develop the necessary interventions to prevent further spread.

1 Introduction

COVID-19, a worldwide pandemic, has reached to more than 150 nations, and is still spreading nowadays. This report aims to summarize the data and explore the trends among all COVID-19 cases reported in Toronto. I found that number of cases has risen dramatically from August to December, and finally had approximately 19,000 new cases in December. Also, I recognized the three most risky neighborhood area in Toronto. 20-29 years-old is the most vulnerable age group among all nine age-groups. There are three possible outcomes, and the recovery rate is 86.32%.

This paper includes two main parts which is data section and analysis section. In data section, I am going to describe the dataset from multiple aspects including source of the data, methodology of collections, ethical bias, as well as the detail dataset description. In analysis section, I will use the tidyverse (Wickham et al. 2019), kableExtra(Zhu 2020), and ggplot2 (Wickham 2016) to generate several charts and tables in the Rstudio environment (R Core Team 2020). The descriptive analysis enhances our understanding of how total number of cases has been fluctuated over one year, how COVID-19 impacts each neighborhood area, which age group is the most susceptible to infection.

2 Data

2.1 Source of the data

The dataset was published by Toronto Public Health. It was updated weekly, and it was last refreshed on January 20, 2021. Toronto Public Health takes responsibility of managing and tracking all confirmed and probable cases. The data are extracted from the provincial communicable disease reporting system (iPHIS) and Toronto's custom COVID-19 case management system (CORES) and combined for reporting. ("Open Data Dataset," n.d.)

*https://github.com/xuxinyi720/Assignment1.git

2.2 Methodology of data collection

Toronto Public Health collects data by receiving the reports of COVID-19 cases every day. To be specific, when a case that meets the provincial case definition("Case Definition – Coronavirus Disease (COVID-19)," n.d.), Toronto Public Health will immediately begin a detailed and careful investigation of each individual. In terms of how respondents are found, each individual will be directly collected from laboratory to local health department. All information will be summarized to show the trend of the pandemic.

2.3 Bias

One of the most significant ethical biases is that sex-disaggregated data on confirmed COVID-19 infections and deaths are incomplete. According to the review from Data2X and Open Data Watch, accurate gender data is considered necessary for a gender-sensitive response to COVID-19 (Lieberman 2020). The challenge is not only that people are not collecting gender data, but also that they are not uniformly collecting it.

2.4 Dataset Description

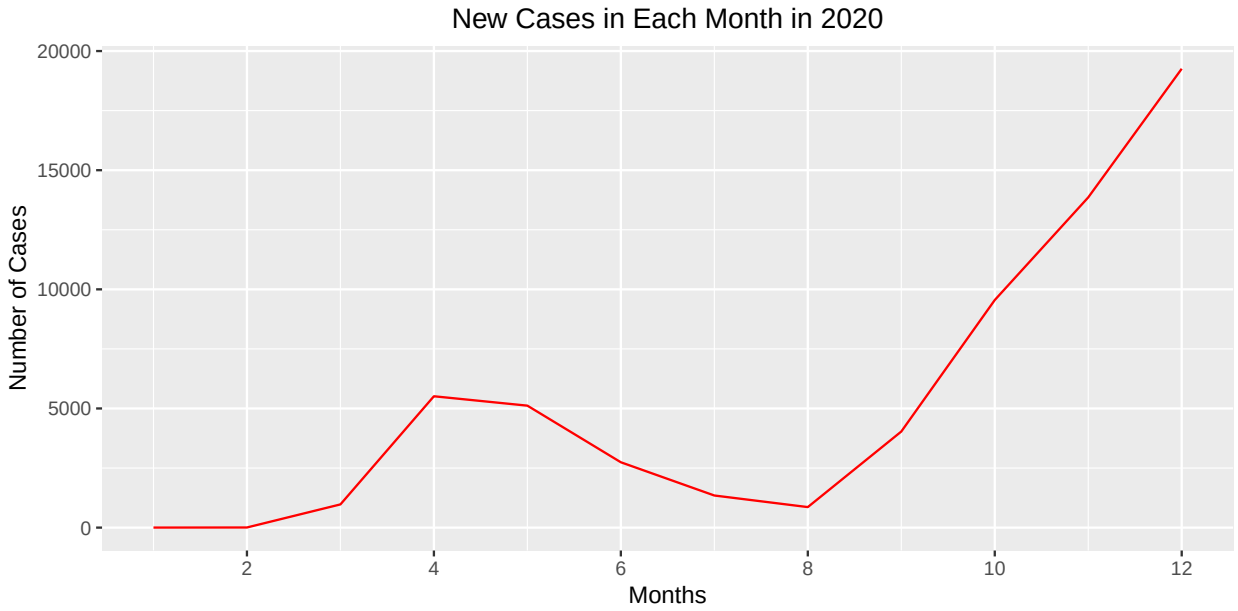
The dataset provides a list of all COVID-19 cases in Toronto from January 2020 to January 2021. There are 77,872 rows in total, as well as 18 variables describing each observation. Generally, the dataset contains demographic data, geographic data, and severity information for all confirmed and probable cases. To be specific, each case has its unique ID, episode date, reported date, and hospitalized record. Patients are coded into nine age groups and four gender categories. Also, according to the principles of epidemiology, each individual has been classified into either sporadic (occurring in the community) or outbreak-associated. For each sporadic case, it indicates the most likely way of infection, including travel, close contact with a case, institutional setting, healthcare setting, community, pending, and unknown. Also, there are three possible outcomes which is active, fatal, and resolved. In addition, Toronto is divided into 140 neighborhoods, and its corresponding FSA (first three characters of postal code) is recorded in the dataset for further local planning.

3 Descriptive analysis

First of all, we make a line chart to visualize the fluctuation in the total number of cases in 2020. We can see that the number of cases has hit the first peak to over 5000 cases in April, and has decreased from May to August. However, number of cases has increased significantly from August to December, and finally reached approximately 19,000 cases.(Figure **??**).

Table 1: Total Number of Three Outcomes

Month	Total Cases
ACTIVE	7830
FATAL	2330
RESOLVED	73314



In addition, by drawing the bar chart, we are able to gain insight into how this pandemic affect each age group. In other word, we can know 20 to 29 years-old group is the most susceptible to infection, so that we can carry out prevention planning for targeted age groups. (Figure 1)

Finally, the finding of the study enhances our understanding of how COVID-19 impacts each neighborhood areas differently in Toronto. According to the second bar chart, we can summarize the three most vulnerable areas by its FSA, which is M9V, M1B, and M3N. This information could also help the public health system to inform planning and equitable prioritization of public health resources. (Figure 2)

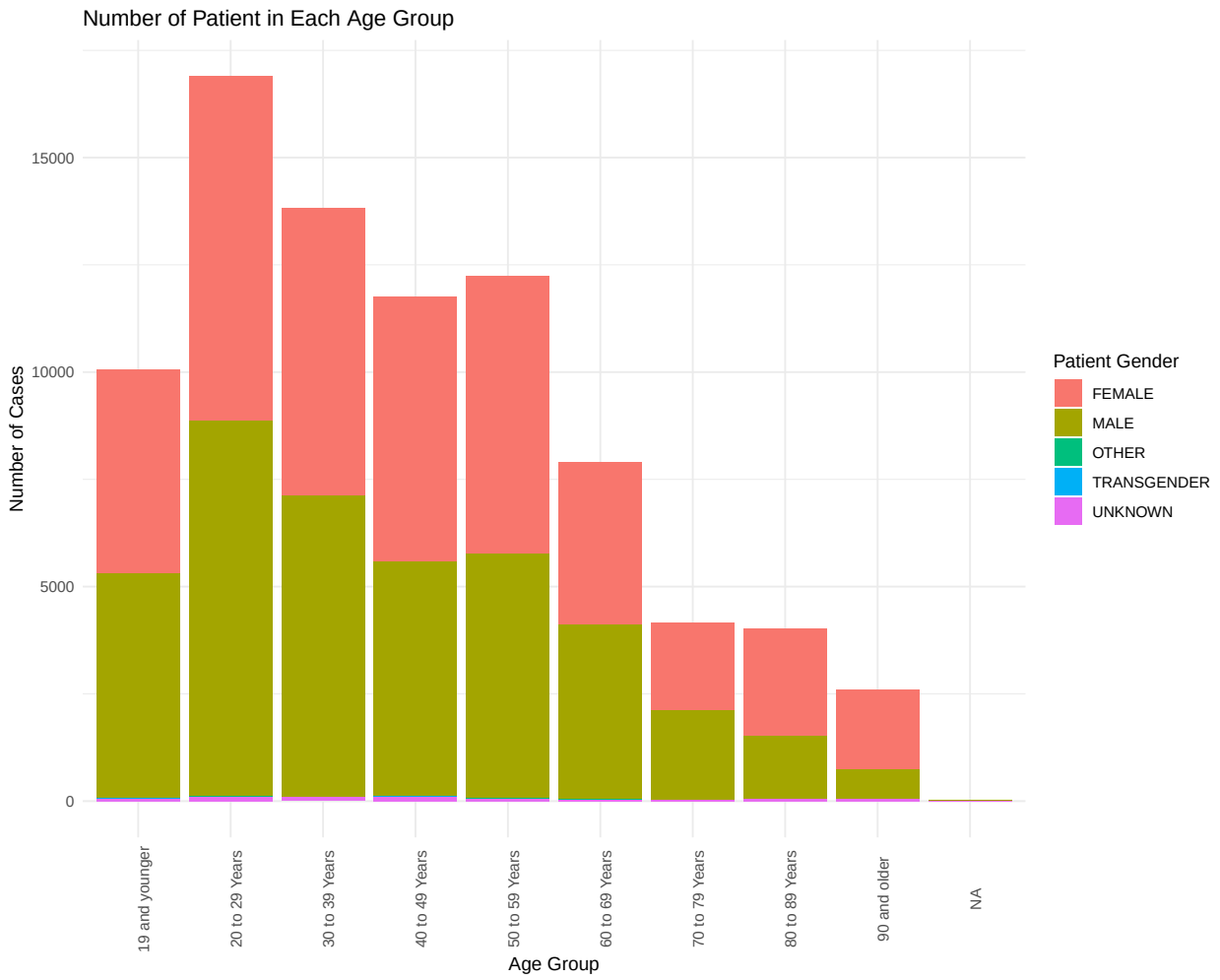


Figure 1: bar-chart-1

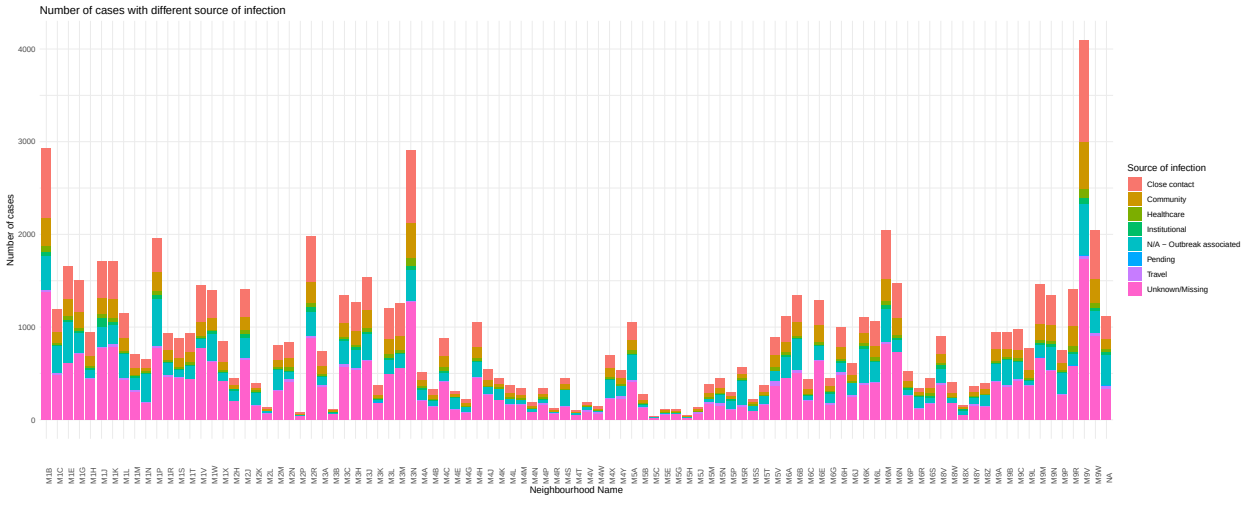


Figure 2: bar-chart-2

Table 2: Outcomes of All Age Group

Outcome	Age Group	Total
ACTIVE	19 and younger	882
ACTIVE	20 to 29 Years	1651
ACTIVE	30 to 39 Years	1286
ACTIVE	40 to 49 Years	1108
ACTIVE	50 to 59 Years	1205
ACTIVE	60 to 69 Years	836
ACTIVE	70 to 79 Years	431
ACTIVE	80 to 89 Years	298
ACTIVE	90 and older	131
ACTIVE	NA	2
FATAL	19 and younger	1
FATAL	20 to 29 Years	2
FATAL	30 to 39 Years	4
FATAL	40 to 49 Years	20
FATAL	50 to 59 Years	72
FATAL	60 to 69 Years	213
FATAL	70 to 79 Years	413
FATAL	80 to 89 Years	833
FATAL	90 and older	772
RESOLVED	19 and younger	9172
RESOLVED	20 to 29 Years	15242
RESOLVED	30 to 39 Years	12535
RESOLVED	40 to 49 Years	10626
RESOLVED	50 to 59 Years	10963
RESOLVED	60 to 69 Years	6844
RESOLVED	70 to 79 Years	3329
RESOLVED	80 to 89 Years	2883
RESOLVED	90 and older	1686
RESOLVED	NA	34

Table 3: Number of Cases in Each Month in 2020

Month	Total
1	2
2	6
3	974
4	5512
5	5119
6	2743
7	1348
8	861
9	4032
10	9551
11	13859
12	19257

References

"Case Definition – Coronavirus Disease (COVID-19)." n.d.

Lieberman, Amy. 2020. "Check Your Bias, Gender Data Experts Warn." https://www.devex.com/news/check-your-bias-gender-data-experts-warn-97366.

"Open Data Dataset." n.d. *City of Toronto Open Data Portal*. https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

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 Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.