Understanding and preventing COVID-19 of Toronto

Shi Yingxuan

05/02/2022

Abstract

Currently, COVID-19 pandemic is a serious issue all over the world, and therefore the data analysis of COVID-19 cases is important because it gives valuable insights, provides public understanding, develops prevention and aid with planning and so on. This report utilizes data from COVID-19 cases in Toronto to analyze the relation between the number of cases and age group, gender, geographical divisions, source of infection. In the analysis, the 20-29 years age group gets the most COVID-19 cases so young people should realize the severity of COVID-19. More importantly, Scarborough and Rexdale have the most COVID-19 cases because they are low-income neighbourhoods and areas with a higher proportion of essential workers. The underlying logic of society affects the number of COVID-19 cases.

Introduction

What is COVID-19? It is coronavirus, SARS-CoV-2, identified in 2019, has caused a pandemic of respiratory disease, called COVID-19. Coroncirus is spread by droplets and virus particles released into the air, and therefore, close contact is a key transmission of COVID-19 (Johns Hopkins Medicine 2021). Worldwide pandemic of COVID-19 have been in Canada because of an individual who had returned to Toronto from Wuhan, Hubei, China, tested positive and most cases of the pandemic have been in Ontario, British Columbia and Alberta. In Ontario, Toronto is affected by COVID-19 most seriously and it has the longest continous COVID-19 lockdown because Toronto is the most populous city in Canada (Wikipedia 2022).

COVID-19 cases statistics play an important role for decreasing the negative influence of COVID-19. Data from COVID-19 cases in Toronto are often used to notice the safe places, the reasons for infection and who needs to care more. National Statistician Professor, Sir Ian Diamond said that statistics help give important insights, provide public understanding, develop prevention and aid with planning and provision of services (SAGE Publishing 2020). Therefore, I will analyze the dataset of COVID-19 cases in the Toronto and give several suggestions about prevention, understanding and action.

For this report, I will use open-access data from the City of Toronto's Open Data Portal to understand COVID-19 cases in Toronto, and then, I will analyze the number of COVID-19 cases related to age group, gender. Moreover, I discuss that different types of transmission of COVID-19 affect different age groups. Finally, with reference to Toronto's FSA map, I find the relation between the number of COVID-19 cases and geographical divisions.

Data

3.1 Data source

This report utilizes data about COVID-19 cases in Toronto that were obtained from the Toronto Public Health (TPH). Since 1883, the TPH has focused on protecting and promoting the health of Toronto residents. It reports the health and well-being of residents to the Board of Health (Toronto Public Health, 2022). COVID-19 cases in Toronto are published by Toronto Public Health on the City of Toronto's Open Data Portal (Gelfand 2020). Thus, the dataset of COVID-19 cases was downloaded in csv format from the City of Toronto's Open Data Portal. The dataset was refreshed on February 2nd, 2022.

3.2 Methodology and Data Collection

R statistical programming language is used for this analysis (R Core Team 2020), and more specifically, the tidyverse package (Wickham et al. 2019), dplyr(Wickham et al. 2021). The raw dataset has 272017 observations that contains information on COVID-19 cases in Toronto. This data is extracted from the provincial Case & Contact Management System (CCM). CCM is a central data repository that contains COVID-19 cases and contact management in Ontario (HealthSTATS, 2020). Although this dataset contains demographic, geographic, and severity information of all confirmed and probable case and includes cases, that are outbreak-associated, since the first case was reported in January 2020, this dataset still is not accurate and exists biases.

The actual number of cases of COVID-19 is definitely more than the numbers in CCM. The biggest reason causing the bias is that not all people who have COVID-19 symptoms seek medical treatment or testing, and then, these cases cannot be recorded in CCM. The case number of homeless people is alarming because these people are not looking for medical treatment or testing. Thus, counting the COVID-19 cases of homeless people is difficult work (Muriel 2021). Currently, most people who get COVID-19 just experience mild illness and can recover at home, and therefore, they do not need to seek testing or medical treatment (MAYO clinic 2022). The cases of homeless people and people who get mild illness cannot be collected so the actual case number is much more than the case number of the dataset.

3.3 Data Characteristic

The dataset obtains all confirmed and probable cases which contain demographic, geographic, and severity information, since the first case was reported in January 2020 (Toronto's Open Data Portal, 2022). There were 272017 observations in the dataset and 18 variables. I filtered 4 variables, Age group, Client Gender, FSA, Source of Infection and Episode Date, they describe almost important information, and then I mutated a new attribute for integer number of age group (Wickham et al. 2019). Age is age group at time of illness; Client Gender is self-reported biological sex; FSA is first three characters of postal code; Source of Infection is that distinct reasons cause the infection. Age group and gender are personal information. Source of Infection and geographic division are details of COVID-19 cases. These variables all are worth analyzing. A sample view of dataset is displayed blow.

```
## # A tibble: 6 x 4
##   `Age Group`    `Client Gender` FSA   `Source of Infection`
##   <chr>          <chr>           <chr> <chr>
## 1 50 to 59 Years FEMALE          M2N   Travel
## 2 50 to 59 Years MALE            M2N   Travel
## 3 20 to 29 Years FEMALE          M3A   Travel
## 4 60 to 69 Years FEMALE          M4W   Travel
## 5 60 to 69 Years MALE            M4W   Travel
## 6 50 to 59 Years MALE            M2R   Travel
```

3.3.1 Age group and Gender

According to information provided by the Toronto Public Health, Age Group and Client Gender are the basic information of patients. Age Group includes 19 and younger, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90 and older. Figure 1 displays that the total number of COVID-19 cases distribute in different age group in Toronto. Client Gender just collects two options, male and female. Figure 2 displays that the total number of COVID-19 male and female cases in Toronto.

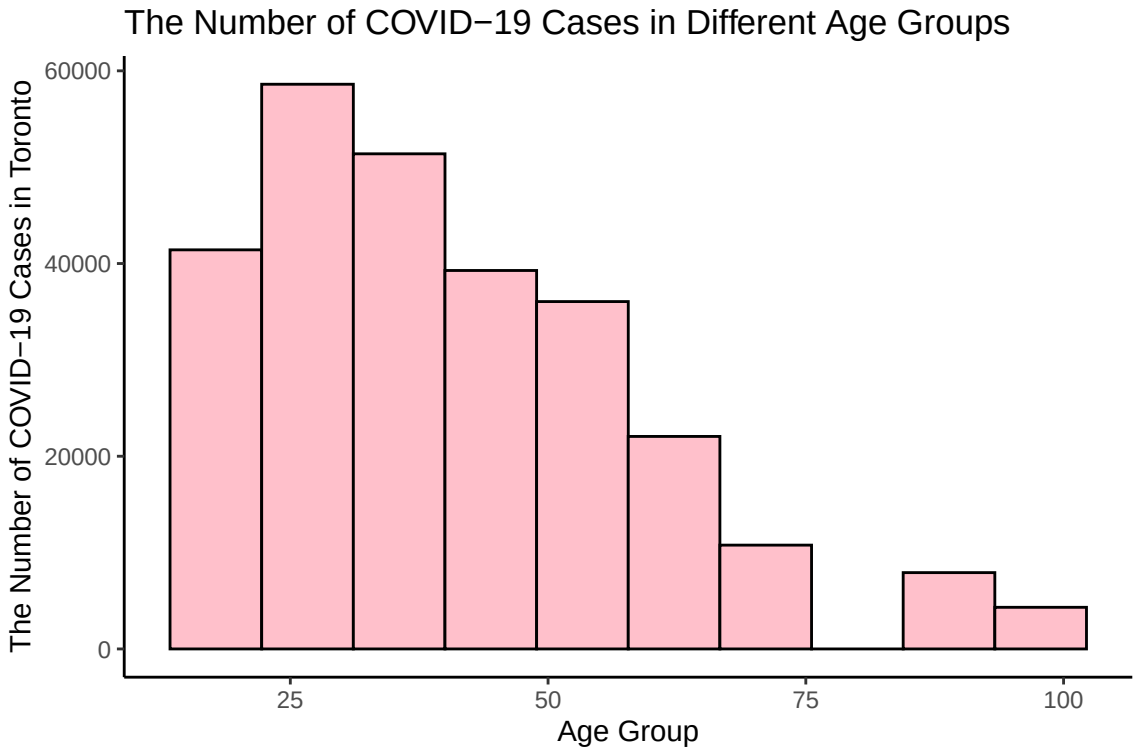

Figure 1: The number of Toronro COVID-19 cases in different age groups

Based on figure 1, we can see a gradual decrease in the number of COVID-19 cases in Toronto from age 19 and younger to age 90 and older. It is right skewness and its center is at 20 to 29 years. It means that the most number of COVID-19 cases in Toronto happen in the age group 20 to 29 years. Obviously, the age group 90 and older has the least number of cases. From a COVID-19 report in Ontario, experts count the cumulative COVID-19 cases of February 3, 2022. It shows that the age group from 20 to 39 has the most cases (402,702) and the age group 80 and over has the least number of cases (44,339). Age from 40 to 59 and Age from 60 to 79 have 291,139 cases and 126,326 cases respectively Public Health Ontario 2022). From the report, it matches the data that I analyze significantly. Infectious disease specialist Dr. Abdu Sharkawy explains that the second and third waves of the COVID-19 pandemic infect more people who are under the age of 40 (Brooklyn 2021). More young people who are ages between 20 to 39 were inflicted than other age groups.

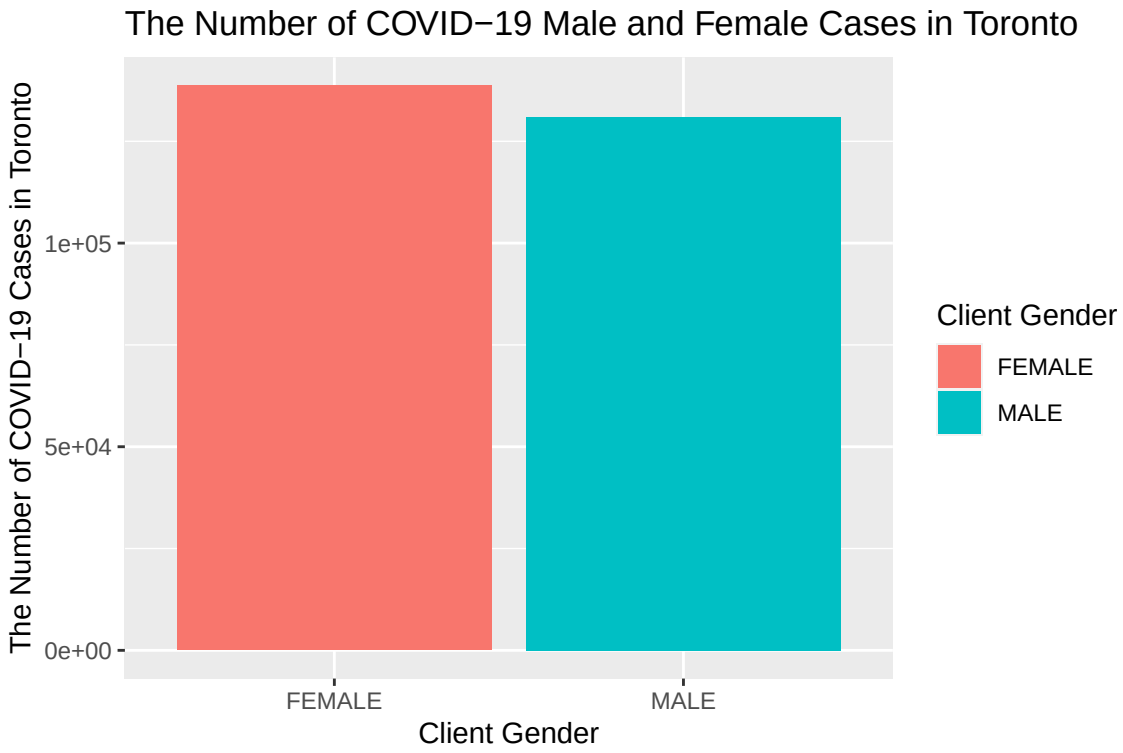

Figure 2: The number of Toronro COVID-19 cases in male and female

Based on figure 2, we can see a two-bar chart that shows an obvious comparison between female and male. The number of female COVID-19 cases in Toronto is slightly more than the number of male cases. From a COVID-19 report in Ontario, experts count the cumulative COVID-19 female and male cases of February 3, 2022. Female is 540,869 cases and male is 501,587 (Public Health Ontario 2022). The number of female COVID-19 cases is approximately 40000 more than male's. According to the latest data from the Public Health Agency of Canada, 55 percent of COVID-19 cases are women and 45 percent of COVID-19 cases in Canada (Olivia 2020). More women than men have been diagnosed with COVID-19.

3.3.2 Source of Infection

Reported COVID-19 cases are broken up into five main kinds of infection. Household contact: case who acquired infection from a confirmed or probable COVID-19 case, such as roommate or family member. Close contact: case who acquired infection from a confirmed or probable COVID-19 case, such as co-worker or client. Travel: case that travelled outside of Ontario. Outbreaks: case is related to confirmed COVID-19 outbreaks in some places. Community: other sources of inflection except household contact, travel, close contact and outbreaks. Therefore, Close contact is the most frequent way to transmit virus and most transmission occurs indoors (Government of Canada 2021). Figure 3 displays that five main reasons of COVID-19 infection are distributed in different age groups.

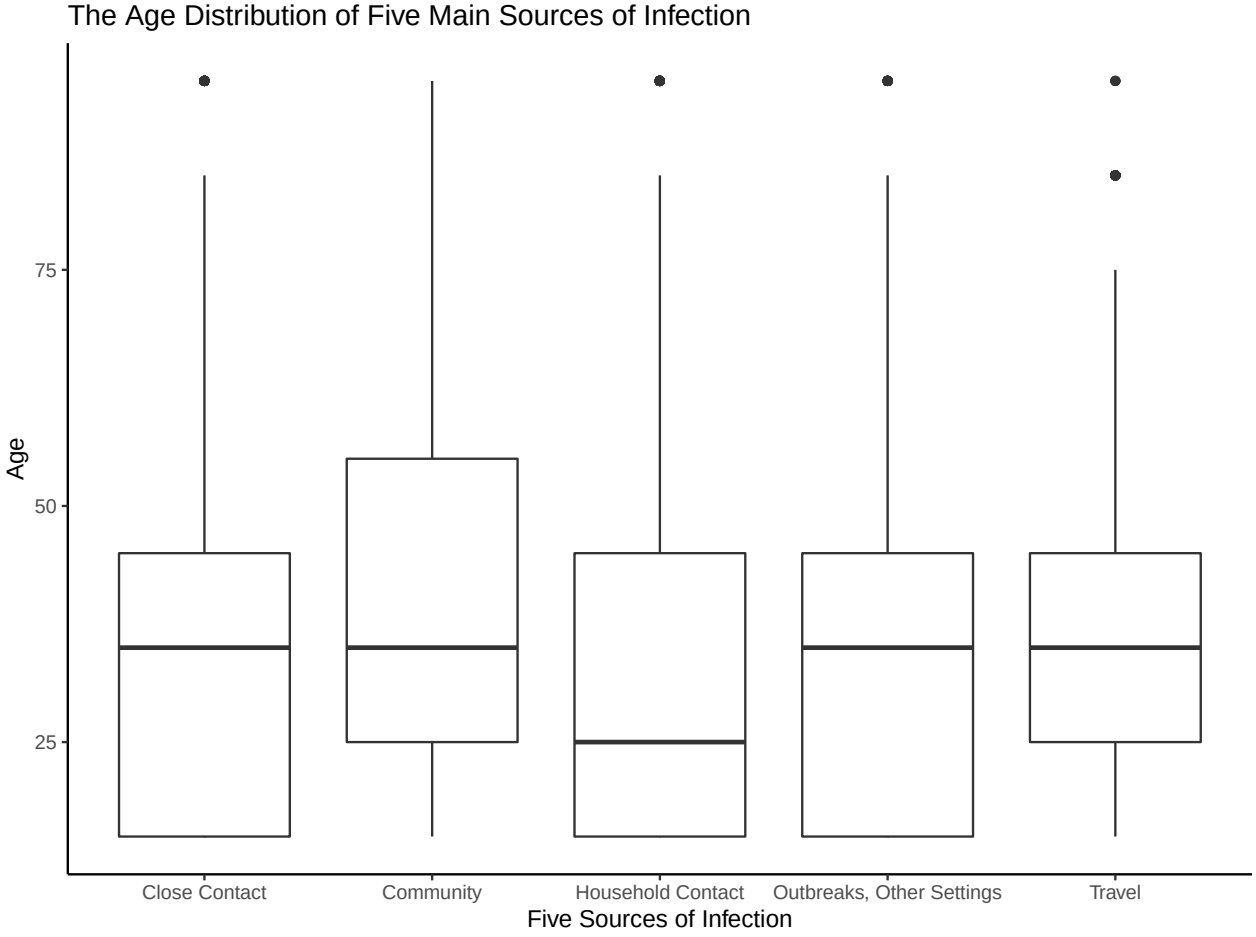

Figure 3: The distribution of five main rources of infection

Base on figure 3, we can see the five main sources of infection on a box plot. The median values of Travel, Outbreaks, Community and Close Contact are almost equaled at 30 years old. Infectious disease specialist Dr. Abdu Sharkawy explains that the second and third waves of the COVID-19 pandemic infect more people who are under the age of 40 (Brooklyn 2021). The box plot proves that the number of young people in this dataset is more than the number of other age groups significantly. However, the median value of Household Contact is smaller than others. In COVID-19 Pandemic, schools were closed and government recommended youth, younge adults and students stay at home (Government of Canada 2022). Therefore, it increases the rate of infection that is caused by household contact greatly and more younger people get COVID-19 because of household contact.

3.3.3 FSA

A forward sortation area (FSA) is the first three characters in a Canadian postal code to identify a geographical unit. The first letter M means Metropolitan Toronto (Government of Canada 2015). Therefoe, FSA of the dataset describes the number of COVID-19 cases in every geographical unit of Toronto. Figure 4 shows every area pictured geographically in Toronto. Table 1 shows the number of cases in each FSA.

Figure 4: The FSA in Toronto map.

```
##
##  M0H  M16  M1I  M1O  M1Z  M2D  M2F  M2V  M2W  M2X  M2Y  M3D  M3n  M4I  M5O  M5U
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##  M5W  M69  M7A  M7J  M7P  M7R  M7V  M8H  M8L  M8P  M8S  M9F  M9O  M9X  M1Y  M2C
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    2    2
##  M2E  M2I  M2T  M2Z  M3I  M3O  M5D  M5Z  M7C  M8M  M8N  M9G  M9K  M9Y  M1D  M2S
##    2    2    2    2    2    2    2    2    2    2    2    2    2    2    3    3
##  M3V  M3X  M3Y  M3Z  M4D  M8A  M8E  M8R  M9D  M9S  M0M  M3T  M5L  M5X  M6Y  M8T
##    3    3    3    3    3    3    3    3    3    3    4    4    4    4    4    4
##  M9Z  M2A  M2B  M3S  M3W  M4Z  M5Y  M6T  M6Z  M9H  M8C  M2G  M5K  M1A  M6W  M9E
##    4    5    5    5    5    5    5    5    5    5    7    8    8    9    9    9
##  M3E  M3R  M3G  M8B  M6V  M5H  M5C  M2P  M4T  M5G  M3B  M8X  M4R  M2L  M5E  M4W
##   10   11   12   12   36  224  276  407  563  597  615  643  647  685  761  882
##  M5J  M4N  M4V  M5S  M3K  M5P  M4G  M5B  M5N  M2H  M2K  M6R  M4P  M5T  M4A  M4E
##  917  918 1046 1085 1113 1184 1188 1345 1370 1383 1414 1465 1498 1503 1566 1600
##  M4B  M5M  M8W  M8Z  M8Y  M5R  M6C  M9L  M1N  M4M  M4S  M1X  M4L  M6G  M6S  M4X
## 1652 1657 1695 1767 1836 1848 1853 1873 1888 1925 1925 2062 2172 2182 2195 2229
##  M4K  M9P  M1M  M4Y  M6P  M1S  M2M  M4H  M1H  M6J  M4J  M9B  M1T  M3A  M6A  M6L
## 2318 2337 2353 2454 2606 2649 2652 2660 2740 2745 2754 2868 2981 3028 3038 3041
##  M9A  M3L  M1R  M1C  M9C  M9N  M3M  M8V  M1V  M6B  M1W  M4C  M1G  M9M  M2N  M9R
## 3074 3262 3297 3333 3544 3554 3558 3587 3653 3726 3795 3840 3851 3931 3933 3949
##  M6H  M1L  M6K  M3J  M3C  M3H  M5A  M6E  M5V  M1J  M6N  M2J  M1E  M1P  M2R  M1K
## 4008 4090 4154 4157 4166 4293 4327 4415 4810 4815 5012 5036 5134 5223 5272 5477
##  M6M  M9W  M3N  M1B  M9V
## 5854 5858 6684 8117 9114
```

From the table, M9V has the most number of COVID-19 cases (9114) and M1B ranks the second (8117). From the figure 4, we can see M9V is Rexdale that is the northwest of Toronto and M1B is Malvern that is close to Scarborough. Humber Heights-Westmount has the highest with 219 cases per 100,000 from November 2021 to December 2021 (Brooke 2021). Humber Heights-Westmount is in the M9V area so this area still had the most number of COVID-19 cases until December 2021. From DH news, low-income neighbourhoods and areas with higher proportion of essential workers have the most number of COVID-19 cases (Brooke 2021). The lowest incomes is found in Scarborough and Linkwood Lane Park which is the west of Toronto was in second (Better Dwelling 2017). It is an interesting finding because it is the reason why M9W and M1B have the most number of COVID-19 cases.

Reference

Brooke. 2021. "These Are the Toronto Neighbourhoods with the Highest Rates of COVID-19." *Daily Hive*. https://doi.org/https://dailyhive.com/toronto/toronto-neighbourhoods-highest-rates-covid-19.

Brooklyn. 2021. "More Young Canadians Getting Severe Covid-19, Being Hospitalized: Experts." *CTV News*. https://doi.org/https://www.ctvnews.ca/health/coronavirus/more-young-canadians-getting-severe-covid-19-being-hospitalized-experts-1.5364360.

CBCnews. 2021. "Toronto Reports COVID-19 Outbreaks at 20 Homeless Shelters, over 300 Cases, 1 Recent Death | CBC News." https://doi.org/https://www.cbc.ca/news/canada/toronto/toronto-shelters-in-outbreak-unhoused-people-one-death-covid-19-1.6003201.

Chris. 2020. "Panel: How Can Social Statistics Help Us Fight Covid-19?" https://doi.org/https://group.sagepub.com/blog/social-science-impact/panel-how-can-social-statistics-help-us-fight-covid-19.

Clinic, Mayo. 2022. "Tips for Treating Covid-19 at Home." *Mayo Foundation for Medical Education and Research*. https://doi.org/https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/treating-covid-19-at-home/art-20483273.

Daniel. 2017. "Here Is Toronto's Richest and Poorest Neighbourhoods." *Better Dwelling*. https://doi.org/https://betterdwelling.com/city/toronto/heres-torontos-richest-and-poorest-neighbourhoods-interactive/.

Medicine, Johns Hopkins. 2022. "What Is Coronavirus?" https://doi.org/https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus.

Olivia. 2020. "More Canadian Women Have Covid-19 and Are Dying as a Result. Here's Some Possible Reasons Why - National." *Global News*. https://doi.org/https://globalnews.ca/news/6920505/more-women-have-coronavirus/.

Ontario, Public Health. 2022. "Covid-19 in Ontario: January 15, 2020 to January 31, 2022." https://doi.org/https://www.publichealthontario.ca/-/media/documents/ncov/epi/covid-19-daily-epi-summary-report.pdf?sc_lang=en.

STATS, Health. 2022. "Case and Contact Management System (CCM)." https://doi.org/https://www.simcoemuskokahealthstats.org/resources/data-sources/ccm.

Toronto, City of. 2019. "Toronto Public Health." https://doi.org/https://www.toronto.ca/city-government/accountability-operations-customer-service/city-administration/staff-directory-divisions-and-customer-service/toronto-public-health/.

Toronto, Open Data. 2022. "COVID-19 Cases in Toronto." https://doi.org/https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.

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.

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

Wikipedia. 2022. "Covid-19 Pandemic in Toronto." https://doi.org/https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Toronto.