

# COVID-19 Cases in Toronto\*

Exploring epidemic impacts among different age groups

Yingying Zhou

31 January 2021

## Abstract

The novel coronavirus has been around for over a year; 2020 witnessed an ongoing global epidemic that has cost countless lives and jobs and systematically disrupted the development of human society in almost every aspect. Taking the Toronto data as one sampled data point from many of the global epicenters, this paper intends to conduct a case study on COVID-19. More specifically, the study investigates the epidemic impacts of COVID-19 on people from different age groups in order to inform the general public of its risk to further educate them for better precautions in disease prevention.

## 1 Introduction

When lockdowns and quarantine orders become part of the city regulation rules, masks worn as one of the outfit essentials, and vaccines deemed worth lining for, Torontonians know their public health system and personal safety are being challenged by the COVID-19 virus, just like almost any other cities in the world. The following section will anatomize Toronto COVID-19 data from the aspect of structure and content. Description on data source, bias, and dataset outlines the problem framework, followed by analysis of epidemic impacts of COVID-19 on different age groups in transmission sources and illness risks (hospitalization and fatality).

## 2 Data

### 2.1 Source of data

The dataset **COVID-19 Cases in Toronto** (“COVID-19 Cases in Toronto” 2020) is retrieved from Open Data Toronto via the API using the package `opendatatatoronto` (Gelfand 2020). It is published by Toronto Public Health and refreshed on a weekly basis.

### 2.2 Bias in data

The dataset might have ethical and statistical problems. On one hand, sensitive features such as gender, age and FSA postal code are included in the dataset. Potential threats would be posed on personal privacy once those sensitive features are reverse engineered to infer individual private attributes. On the other hand, the case numbers might be undercounted. It is possible that asymptomatic carriers did not get tested because they do not feel unwell or show any suspected symptoms. Human bias in respondent’s answer and lags in data entry also cause some answers in source of infection missing and outcome data for resolved cases to be underreported.

---

\*Code and data are available at: [https://github.com/StephaininZ/open\\_data\\_toronto\\_covid19](https://github.com/StephaininZ/open_data_toronto_covid19).

Table 1: Toronto COVID-19 Cases by year for 2020-2021

Year	classification	Number
2020	CONFIRMED	60876
2020	PROBABLE	2232
2021	CONFIRMED	14397
2021	PROBABLE	367

### 2.3 Snapshot of COVID-19 Situation in Toronto

The original dataset contains 77,872 observations of COVID-19 cases in Toronto from the inception of the pandemic outbreak, Jan 23rd, 2020 to present. There are 18 features including demographic and geographic information (case ID, patient’s age group, gender, neighbourhood, FSA postal code, etc.), case-related diagnosis (source of infection, diagnosis classification, episode date, reported date, case outcome) and records of hospitalization indicating whether the patient is or has ever been hospitalized/in ICU/intubated.

1. **\_id**: Unique row identifier for Open Data database
2. **Assigned\_ID**: A unique ID assigned to cases
3. **Outbreak Associated**: Outbreak associated cases are associated with outbreaks of COVID-19 in Toronto healthcare institutions and healthcare settings and congregate settings.
4. **Age Group**
5. **Neighbourhood Name**
6. **FSA**: first three characters of postal code
7. **Source of Infection**: Travel; Close contact with a case; Institutional; Healthcare; Community; Pending; Unknown/missing; N/A
8. **Classification**: categorize the cases as confirmed or probable
9. **Episode Date**: best estimates when the disease was acquired
10. **Reported Date**: The date on which the case was reported to Toronto Public Health
11. **Client Gender**
12. **Outcome**: Fatal; Resolved; Active
13. **Currently Hospitalized**
14. **Currently in ICU**
15. **Currently Intubated**
16. **Ever Hospitalized**: Cases that were hospitalized, admitted to ICU, intubated related to their COVID-19 infection (includes cases that are currently hospitalized and those that have been discharged or are deceased)
17. **Ever in ICU**
18. **Ever Intubated**

For the purpose of exploration of the novel coronavirus effect on different age groups, only the variables **Age Group**, **Classification** (filtered to be **CONFIRMED**), **Source of Infection**, **Reported Date**, **Ever\_Hospitalized**, and **Outcome** are selected, cleaned with package **janitor** (Firke 2021) and **tidyverse** (Wickham et al. 2019), and constructed into a new dataframe called **covid\_toronto** using package **kableExtra** (Zhu 2020). Graphs are plotted with **ggplot2** (Wickham 2016).

As shown in the table (Table 1), 2020 has witnessed 63,208 COVID-19 confirmed or probable cases in Toronto, but January 2021 alone reports 14764 cases, accounting for about 23.4% of total cases in 2020.

As for the fatality rate (Table 2), April has the highest rate of 36.79%, followed by November (15.86%) and May (14.67%).

Among the confirmed cases (Figure 1), young people aged between 20 to 39 years old are the two majority groups, totalling around 35% every month. At first sight, this finding seems counter-intuitive since according to the World Health Organization (WHO) (“Statement – Older People Are at Highest Risk from Covid-19,

Table 2: Toronto COVID-19 Fatality Rate by Month

Month	Number	Fatality Rate
01	75	0.0342779
03	50	0.0228519
04	805	0.3679159
05	321	0.1467093
06	82	0.0374771
07	40	0.0182815
08	8	0.0036563
09	43	0.0196527
10	182	0.0831810
11	347	0.1585923
12	235	0.1074040

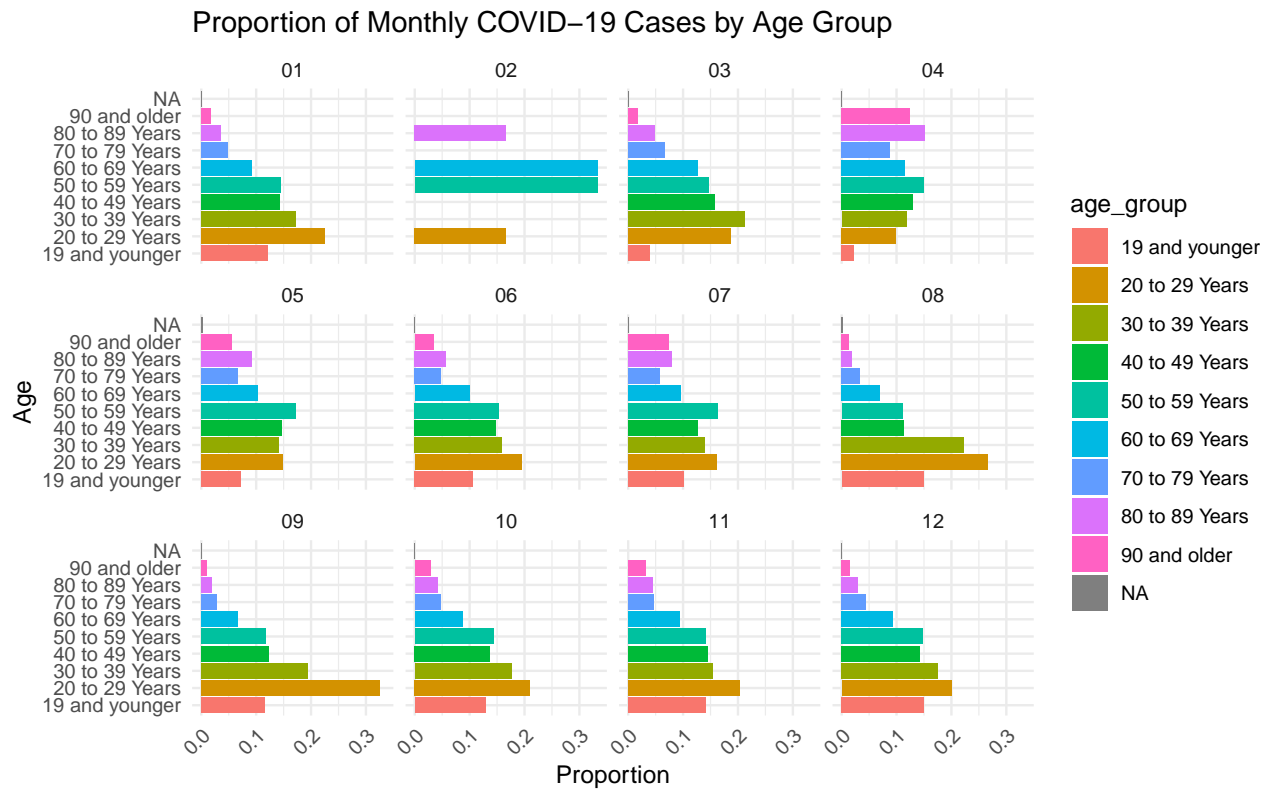


Figure 1: Proportion of Monthly COVID-19 Cases by Age Group

but All Must Act to Prevent Community Spread” 2020) that “Older people are at highest risk from COVID-19”, especially aged above 85.

But a closer analysis on source of infection, hospitalization and fatality cases would verify the correctness of WHO’s statement.

### 2.3.1 Source of Infection

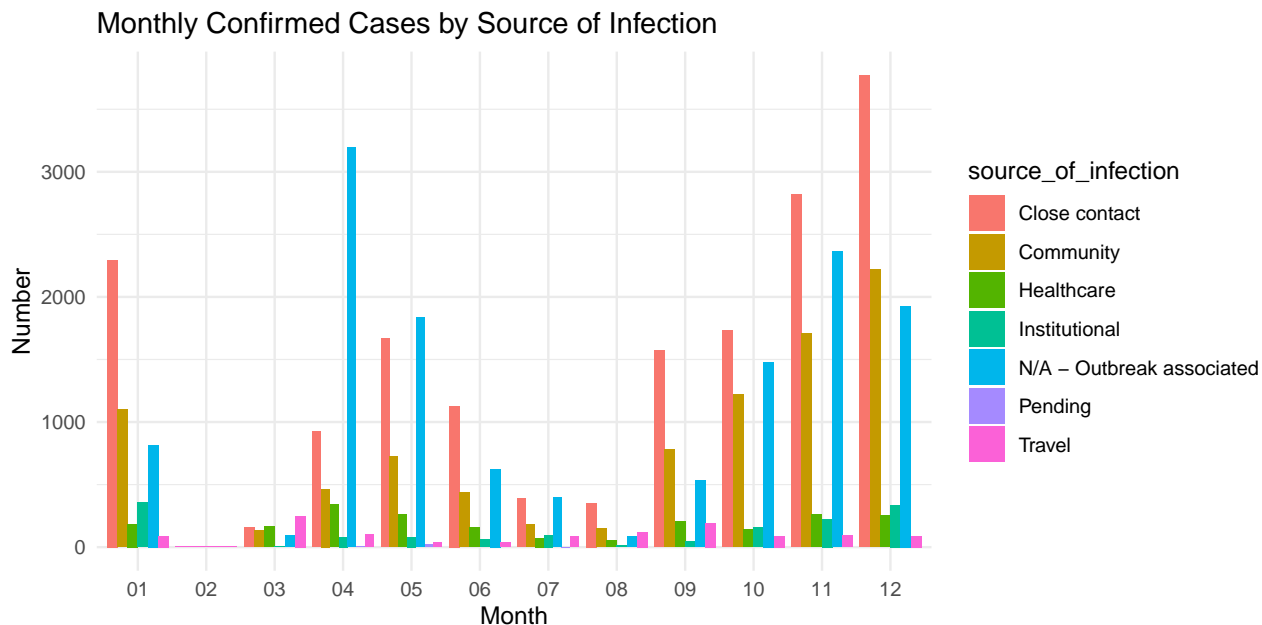


Figure 2: Monthly Confirmed Cases by Source of Infection

Except for February when all 6 cases resulted from overseas traveling, the top 2 frequent causes of infection (Figure 2) are close contact and community, excluding unknown or missing sources. Young people tend to engage in more social activities such as partying and clubbing. Assuming same level of susceptibility, their frequent practice in social gatherings expose them to higher chances of virus infections. Thereby, it largely accounts for why young people are the majority group in COVID-19 cases.

### 2.3.2 Hospitalization and Fatality

Older people, especially aged above 80 top the list for hospitalization-fatality rates every month, together they take up more than 60% (Figure 3) of the death cases after hospitalization. It reveals a concerning fact that once older people are diagnosed with COVID-19, they might suffer from severe COVID-19 symptoms. In other words, even if they are hospitalized to receive medical treatment, their chances of recovery are still dismal.

## 3 Results

In conclusion, although young people are the major victims of the pandemic in terms of total cases in Toronto, it is the older people who bear the higher chance of serve illness once confirmed of COVID-19. As shown in the graph (Figure 4), The distribution of fatal cases concentrates around age group 60-69, 70-79, 80-89, and over 90s. From 2020 to present, there are more fatal cases among older Torontonians aged above 60 in total.

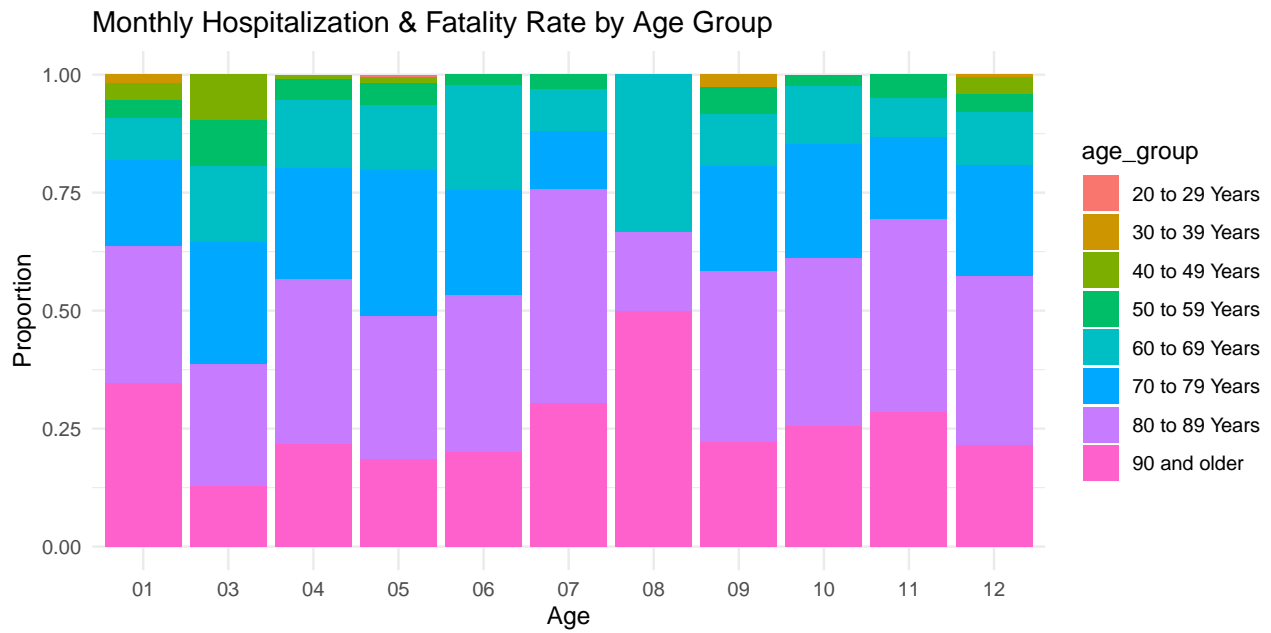


Figure 3: Monthly Hospitalization & Fatality Rate by Age Group

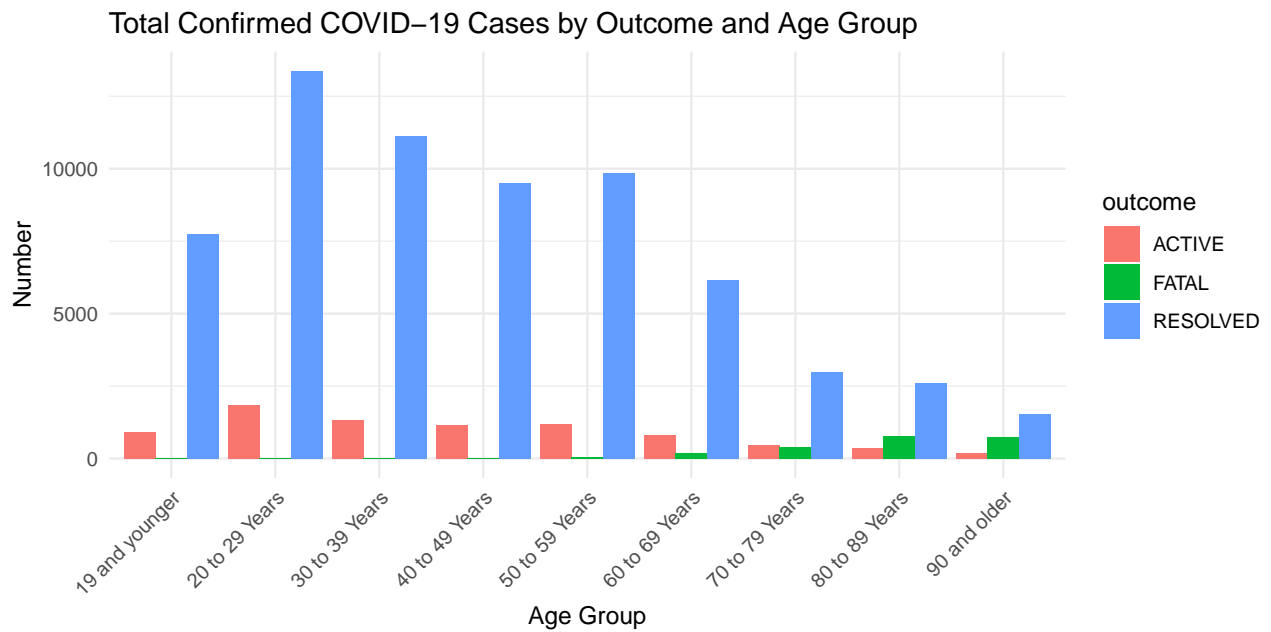


Figure 4: Total Confirmed COVID-19 Cases by Outcome and Age Group

Therefore, our findings echo with another CDC’s report that “older people are at a greater risk of requiring hospitalization or dying” (“Older Adults at Greater Risk of Requiring Hospitalization or Dying If Diagnosed with Covid-19” 2020). During this critical time of second wave, this paper urges that every resident in Toronto should perform their duty in obeying the ‘stay-home order’. Eventually, chances of COVID-19 infection and transmissions would be minimized and the curve will be flattened sooner.

## References

- “COVID-19 Cases in Toronto.” 2020. Open Data Toronto; <https://open.toronto.ca/dataset/covid-19-cases-in-toronto/>.
- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- “Older Adults at Greater Risk of Requiring Hospitalization or Dying If Diagnosed with Covid-19.” 2020. CDC; <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html>.
- “Statement – Older People Are at Highest Risk from Covid-19, but All Must Act to Prevent Community Spread.” 2020. WHO; <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/statements/statement-older-people-are-at-highest-risk-from-covid-19%2c-but-all-must-act-to-prevent-community-spread>.
- 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 Grolemond, 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>.