

Contents

1	Introduction	1
2	Data	1
3	Model	5
4	Results	5
5	Discussion	5
5.1	First discussion point	5
5.2	Second discussion point	5
5.3	Third discussion point	5
5.4	Weaknesses and next steps	5
	Appendix	6
A	Additional details	6
B	References	7

title: “My title”

subtitle: “My subtitle if needed”

author:

- Arsh Lakhanpal

thanks: “Code and data are available at:

https://github.com/lakhan99/starter_folder-main”

date: “04 February 2022”

abstract: “The city of Toronto has been able to collect data pertaining to COVID-19 cases demographic and geographic va

output:

bookdown::pdf_document2

toc: FALSE

bibliography: references.bib

1 Introduction

On December 31, 2019 the world saw its first case of the coronavirus, officially referred to as SARS-CoV-2 or COVID-19, in Wuhan China and by January 25, 2020 Toronto had seen its first case. Ever since, the virus continues to spread in various forms like the Delta variant and the most recent being the Omicron variant. The province has taken many attempts at fighting the virus which include locking down the province and shutting down non-essential activities in hopes of limiting close contact between citizens.

You can and should cross-reference sections and sub-sections. For instance, Section 2. R Markdown automatically makes the sections lower case and adds a dash to spaces to generate labels, for instance, Section 5.1.

2 Data

```
covid_cases <- clean_names(covid_cases)
#covid_cases <- subset(covid_cases, select = -c(1,2,7,8))
```

The statistical programming language used to work with the data in this paper is R (add citation) along with packages like tidyverse, knitr. The “COVID-19 Cases in Toronto” dataset was obtained from Open Data

Table 2: Table 1: Number of covid cases from contact during holiday months

	September	November	December
2020	3012	5029	9047
2021	2806	1932	3975

Toronto which houses a broad number of datasets. This data set was published by Toronto Public Health and for this paper, this dataset was officially imported from the opendatatoronto package in R.

This data consists of confirmed and probable reported COVID-19 cases among Toronto residents with demographic, geographical and epidemiological variables.(citation) The data itself is claimed to have been extracted from the provincial Case and Contact Management System. It is important to note that as the pandemic is happening, this dataset continues to get updated therefore results of this dataset are based off data up till January 24,2022. The dataset consists of 272017 observations of 18 variables however, since the first two variables were simply just “id” and “Assigned_ID” these were removed as they held no importance the the focus of this paper.

COVID-19, like many other viruses thrives in colder drier climates and Figure ?? . To do this, two new variables named “month_of_diagnosis” and “year_of_diagnosis” were mutated using the tools from the dplyr package. The following two graphs show the number of covid cases for both 2020 and 2021 as those are the only two complete years that we have have to live with the virus.

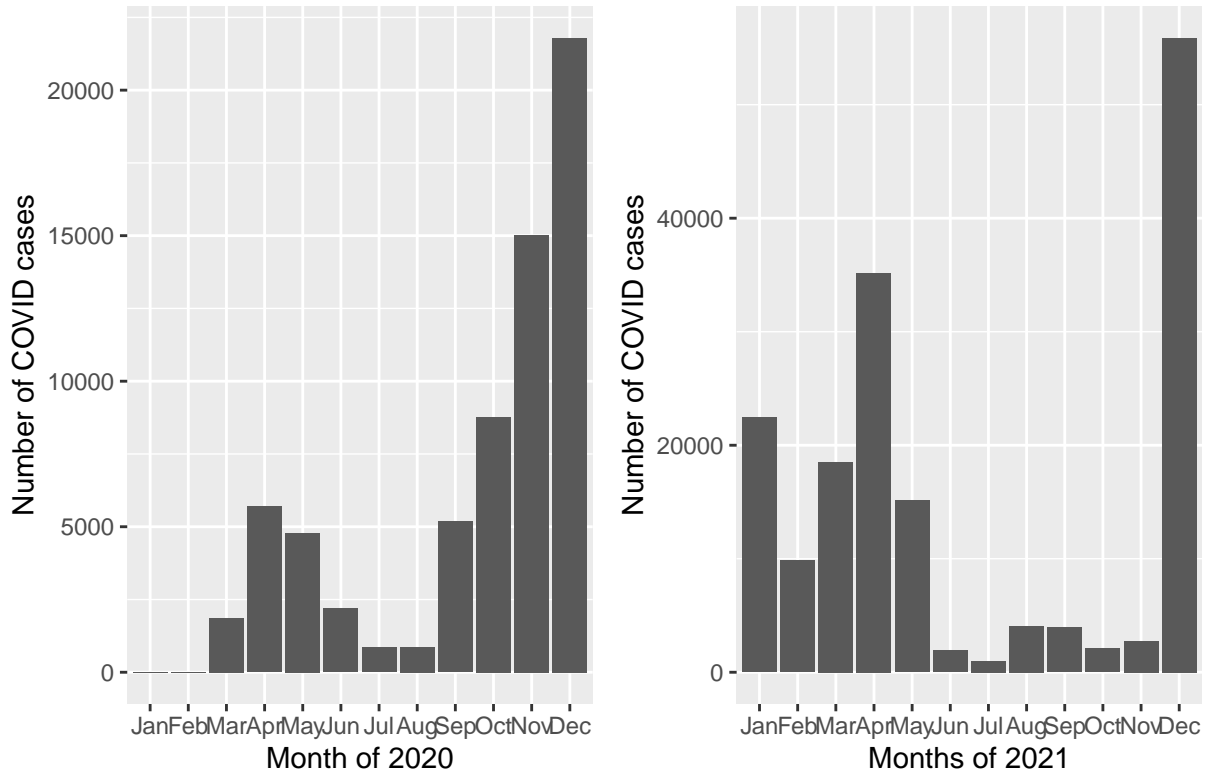


Figure 1: COVID-19 Cases for every month in 2020 and 2021

In 2020, we can see how the number of covid cases were highest during the fall and winter months and how that trend carries onto early 2021 which should be what is expected. The number of cases tend to increase near the winter 2020 and that trend continued in until April 2021 where the cases began to continuously decrease. One thing to notice is that in both years, there are a higher amount of COVID cases comparatively speaking to other months. One of the possible reasons explaining this rise in cases during this time is the fact that this is around the holidays when families tend to get together. Table 2 shows the count for COVID-19

cases that were transmitted through contact which includes close contact, household contact and community contact. For both years, we notice that the month of December had the highest number of COVID-19 cases, likely because of Christmas. Another thing to consider is that in December of 2020, there was a new and more contagious variant of the COVID-19 virus called the delta variant. Similarly, at the end of November 2021, we got the omicron variant, one that was even more contagious than the delta variant which also explains the high case counts in December of 2020 and 2021.

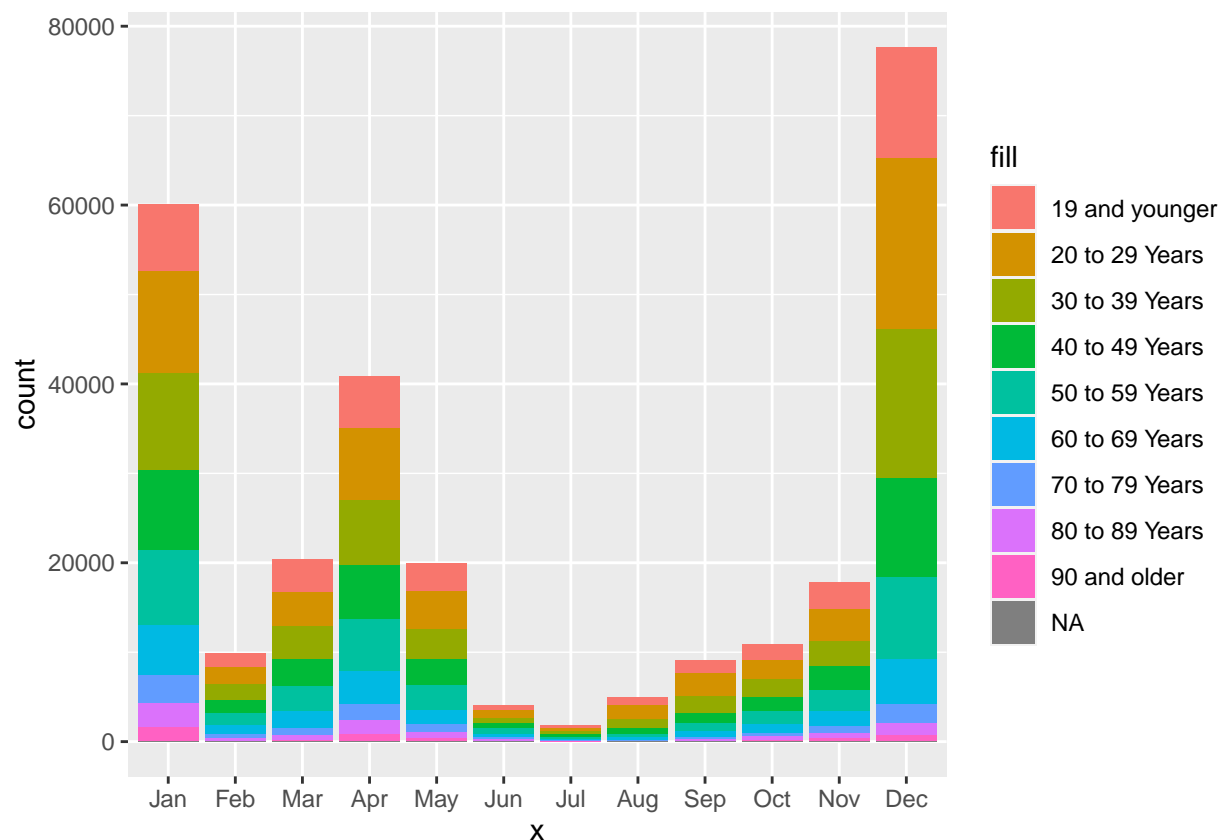
shows us

(talk about possible lockdown/vaccine trends)

We can further focus on the covid cases for every month and which age group seemed to be most affected by covid throughout this pandemic.

```
ggplot(data = covid_cases, mapping = aes(x = covid_cases$month_of_diagnosis, fill = covid_cases$age_group))
  geom_bar(position = 'stack')
```

Warning: Ignoring unknown parameters: position



```
#covid_cases %>% count(age_group)
```

"This is used to get a visual representation of the age groups being infected"

```
## [1] "This is used to get a visual representation of the age groups being infected"
```

```
#seperate this into 2020,2021 as well?
```

These bar charts..

```
#covid_cases2021 %>% ggplot(mapping = aes(x= age_group)) + geom_bar()
```

```
#fatal_2020 <- covid_cases %>%
```

```

#filter(year_of_diagnosis == "2020" & outcome == "FATAL") %>%
#select(year_of_diagnosis, outcome)

#resolved_2020 <- covid_cases %>%
#filter(year_of_diagnosis == "2020" & outcome == "RESOLVED") %>%
#select(year_of_diagnosis, outcome)

#fatal_2021 <- covid_cases %>%
#filter(year_of_diagnosis == "2021" & outcome == "FATAL") %>%
#select(year_of_diagnosis, outcome)

#resolved_2021 <- covid_cases %>%
#filter(year_of_diagnosis == "2021" & outcome == "RESOLVED") %>%
#select(year_of_diagnosis, outcome)

#covid_cases %>% count(year_of_diagnosis)

#covid_cases %>% count(month_of_diagnosis)
#covid_cases %>% count(outcome)
#propo <- covid_cases %>% count(source_of_infection)
#type <- unique(covid_cases[c("source_of_infection")])

#prop.table(propo)

#cases_per_year <- covid_cases %>% select(year_of_diagnosis, outcome)
#API

#knitr:: kable(table_c,
                #col.names = "January", "February", "March", "April", "May", "June", "July", "August", "S

#locations <- tibble(
  #postal = c(covid_cases$fsa)
  #)

#hopeful <- geocode(locations,
  #postalcode = "postal",
  #method = "osm")

#cities <- reugeocode(c(43.77977
#, -79.36618), output = "address")

#cities$locality

```

Our data is of penguins (Figure ??).

Talk more about it.

Also bills and their average (Figure ??). (Notice how you can change the height and width so they don't take the whole page?)

Talk way more about it.

3 Model

$$Pr(\theta|y) = \frac{Pr(y|\theta)Pr(\theta)}{Pr(y)} \tag{1}$$

Equation (1) seems useful, eh?

Here's a dumb example of how to use some references: In paper we run our analysis in **R** (**citeR?**). We also use the **tidyverse** which was written by (**thereferencecanbewhatever?**) If we were interested in baseball data then (**citeLahman?**) could be useful.

We can use maths by including latex between dollar signs, for instance θ .

4 Results

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional details

B References