Yang's Personal Project, Toronto Covid-19 Studies

Yang Shang, 1002603765

27/09/2020

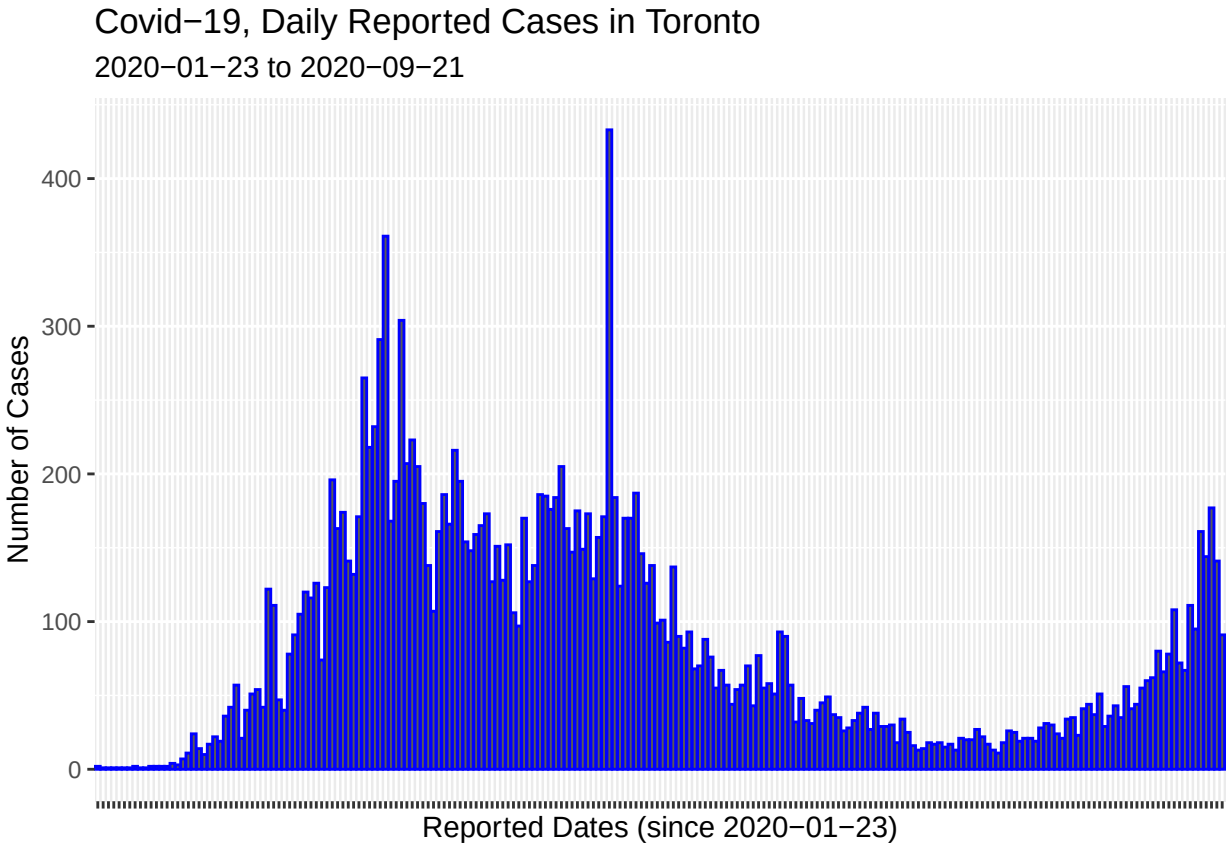
```
library(opendatatoronto)
library(ggplot2)
packages <- show_package("64b54586-6180-4485-83eb-81e8fae3b8fe")
resources <- list_package_resources("64b54586-6180-4485-83eb-81e8fae3b8fe")
covid_resources <- filter(resources, tolower(format) %in% c('csv', 'geojson'))
covid_case_report <- filter(covid_resources, row_number()==1) %>% get_resource()
covid_case_report
```

```
## # A tibble: 17,872 x 18
##     `_id` Assigned_ID `Outbreak Assoc~ `Age Group` `Neighbourhood ~ FSA
##     <int>       <int> <chr>            <chr>       <chr>            <chr>
##  1 143647           1 Sporadic         50 to 59 Y~ Willowdale East  M2N
##  2 143648           2 Sporadic         50 to 59 Y~ Willowdale East  M2N
##  3 143649           3 Sporadic         20 to 29 Y~ Parkwoods-Donal~ M3A
##  4 143650           4 Sporadic         60 to 69 Y~ Church-Yonge Co~ M4W
##  5 143651           5 Sporadic         60 to 69 Y~ Church-Yonge Co~ M4W
##  6 143652           6 Sporadic         50 to 59 Y~ Newtonbrook West M2R
##  7 143653           7 Sporadic         80 to 89 Y~ Milliken         M1V
##  8 143654           8 Sporadic         60 to 69 Y~ Willowdale West  M2N
##  9 143655           9 Sporadic         50 to 59 Y~ Willowdale East  M2N
## 10 143656          10 Sporadic         60 to 69 Y~ Henry Farm       M2J
## # ... with 17,862 more rows, and 12 more variables: `Source of
## #   Infection` <chr>, Classification <chr>, `Episode Date` <chr>, `Reported
## #   Date` <chr>, `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>
```

This data set collects all reported Covid-19 cases in Toronto, On, Canada. Until September 11^{th}, 2020, 17872 cases were reported. Each case is assigned a unique ID. Patients were labeled by their age groups, area of living (Neighborhood Name), source of infection, episode date, reported date, gender, outcome, and states of treatments.

I would like to first study about the trend of daily reported increasing cases. As "Date" is a discrete variable, I will use a bar graph. The x-axis is the reported date and y-axis is the number of reported cases.

```
ggplot(covid_case_report,aes(x=`Reported Date`))+geom_bar(colour="Blue")+
  theme(axis.text.x=element_blank())+
  labs(x="Reported Dates (since 2020-01-23)",y="Number of Cases",
       title="Covid-19, Daily Reported Cases in Toronto",
       subtitle="2020-01-23 to 2020-09-21")
```



As each column represents number of reported cases on a unique day, I am able to see the trend of the pandemic through increase and decrease in daily reported cases. The plot shows a Bi-modal shape, where the first climax was reached around March to May and the second one was reached in September.

By news from Toronto Government, sourced from https://www.toronto.ca/news/city-of-toronto-now-in-stage-3-reopening/, Toronto was reopened from pandemic on July 31^{st}, 2020. By the bar grouph, the day is approximately the local minimum of dayly reported cases. Since then, the reported cases started to increase again.

I also would like to study about the comprehensive response time of individuals and Toronto government towards covid-19. I will approximate the response time by calculating the differences between episode date and reported date.

```
date2 <- data.frame(Episode_Date=c(covid_case_report$`Episode Date`),
                    Reported_Date=c(covid_case_report$`Reported Date`))
date2$date_diff <- as.Date(as.character(date2$Reported_Date),format="%Y-%m-%d")-
  as.Date(as.character(date2$Episode_Date),format="%Y-%m-%d")
```

The mean of estimated response time is

```
mean(date2$date_diff)
```

```
## Time difference of 5.977171 days
```

with variance of

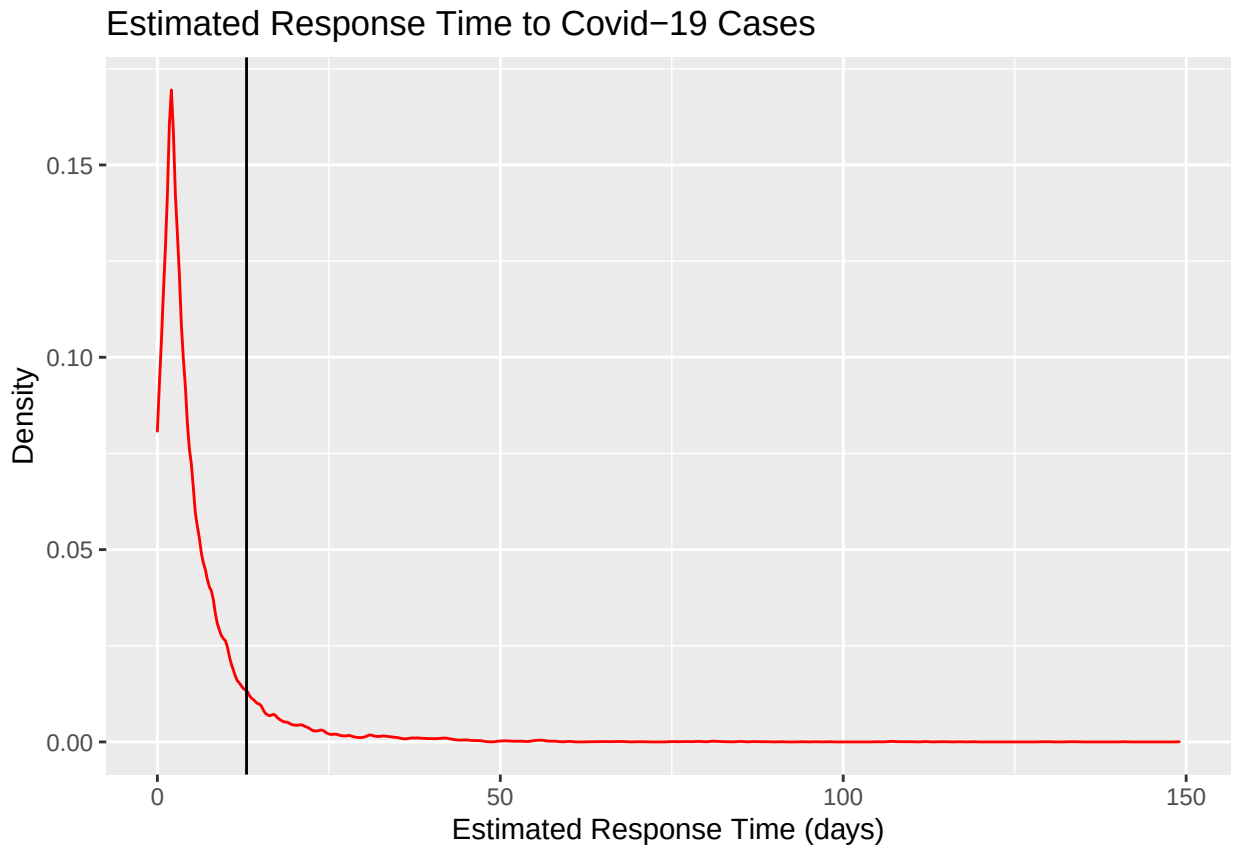
```
var(date2$date_diff)
```

```
## [1] 73.53627
```

Shown by a density plot, we obtain

```
ggplot(date2,aes(x=date_diff,y=..density..))+
  geom_density(colour="Red")+
  geom_vline(xintercept=quantile(date2$date_diff,0.90))+
  labs(x="Estimated Response Time (days)",y="Density",
       title="Estimated Response Time to Covid-19 Cases")
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



The density plot shows that most Covid-19 cases are responded around 5.977171 days, and 90 percent of the cases are responded under 13 days.