Statistical Analysis of COVID-19 Cases In Toronto (Rough Draft)

Woolim Kim

09/12/2020

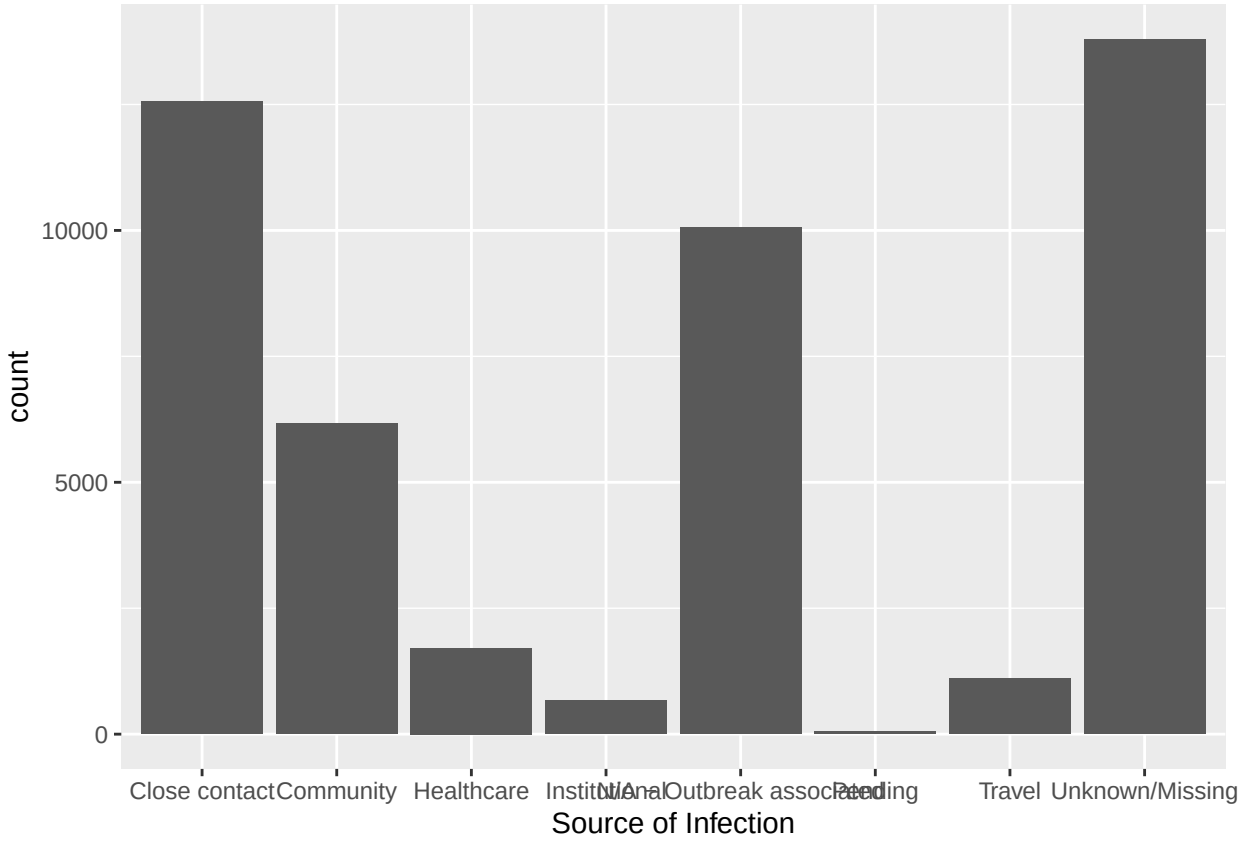
Introduction

Description Using 'About COVID-19 Cases in Toronto' data from the 'Toronto Open Data' website. The data contains gender, age group, source of infection, classification, etc. I want to create an analysis of this data to see how people mostly got infected, and what age group is most exposed to the viruses. I am also planning to find relationships between variable some of the variables that is important and interesting here but have not figure out how to do it. Most of the data is categorical variable so I will have to convert the variables to quantitative values.

```
library(dplyr)
library(tidyverse)
Data <- read_csv("COVID19 cases.csv")
head(Data)
```

```
## # A tibble: 6 x 18
##    `_id` Assigned_ID `Outbreak Assoc~ `Age Group` `Neighbourhood ~ FSA
##    <dbl>       <dbl> <chr>            <chr>       <chr>            <chr>
## 1 526909           1 Sporadic         50 to 59 Y~ Willowdale East  M2N
## 2 526910           2 Sporadic         50 to 59 Y~ Willowdale East  M2N
## 3 526911           3 Sporadic         20 to 29 Y~ Parkwoods-Donal~ M3A
## 4 526912           4 Sporadic         60 to 69 Y~ Church-Yonge Co~ M4W
## 5 526913           5 Sporadic         60 to 69 Y~ Church-Yonge Co~ M4W
## 6 526914           6 Sporadic         50 to 59 Y~ Newtonbrook West M2R
## # ... with 12 more variables: `Source of Infection` <chr>,
## #   Classification <chr>, `Episode Date` <date>, `Reported Date` <date>,
## #   `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>
```

```
Data %>%
  ggplot(mapping = aes(x = `Source of Infection`)) + geom_bar()
```



Citation

CSV Data: Open Data Dataset. (n.d.). Retrieved December 10, 2020, from https://open.toronto.ca/dataset/covid-19-cases-in-toronto/