

Covid Multi-State Data Notebook

Goal

The purpose of this notebook is to show the analysis of Covid data using data from all of the states

Author: John Pratt

Data can be found here:

<https://covidtracking.com/data>

Preparing the data

First we need to prepare the data and read it in from a CSV as a dataframe in R.

```
rawData <- read.csv("daily.csv", stringsAsFactors = FALSE)

rawData$date <- as.character(rawData$date)
# Fix the date field
rawData$FixedDate <- as.Date(rawData$date, format = "%Y%m%d")
```

Now we build a dataframe for just the states we are interested in.

```
dataVirginia <- subset(rawData,
                        rawData$'state' == "VA")
dataFlorida <- subset(rawData,
                      rawData$'state' == "FL")
```

Calculating the Totals for the State

In order to get an idea of overall statistics for the state we will total the numbers in the raw data set.

```
VAtotalCases = head(dataVirginia$positive,1)
VAtotalHospitalizations = head(dataVirginia$hospitalizedCumulative,1)
VAtotalDeaths = head(dataVirginia$death,1)
FLtotalCases = head(dataFlorida$positive,1)
FLtotalHospitalizations = head(dataFlorida$hospitalizedCumulative,1)
FLtotalDeaths = head(dataFlorida$death,1)
```

Totals for the States:

Virginia

- Cases: 69782
- Hospitalizations: 10291
- Deaths: 1962

Florida

- Cases: 254511
- Hospitalizations: 18341
- Deaths: 4301

Calculate heard immunity rate of Eastern Virginia

The CDC has released the infection fatality rate (IFR) as 0.26%. Using the total number of deaths in specific states, we can calculate the total number of people who have had COVID in this region.

```
VANumberHadit = VAtotalDeaths / .0026  
FLnumberHadit = FLtotalDeaths / .0026
```

Individuals who have had Covid:

- Virginia: 754615.4
- Florida: 1654231

Using the number of individuals who have had COVID divided by the total population of the states, we can calculate the heard immunity threshold of the region.

Total Population in 2020:

- VA: 8,536,000
- FL: 21,480,000

Percentage of individuals who have had Covid:

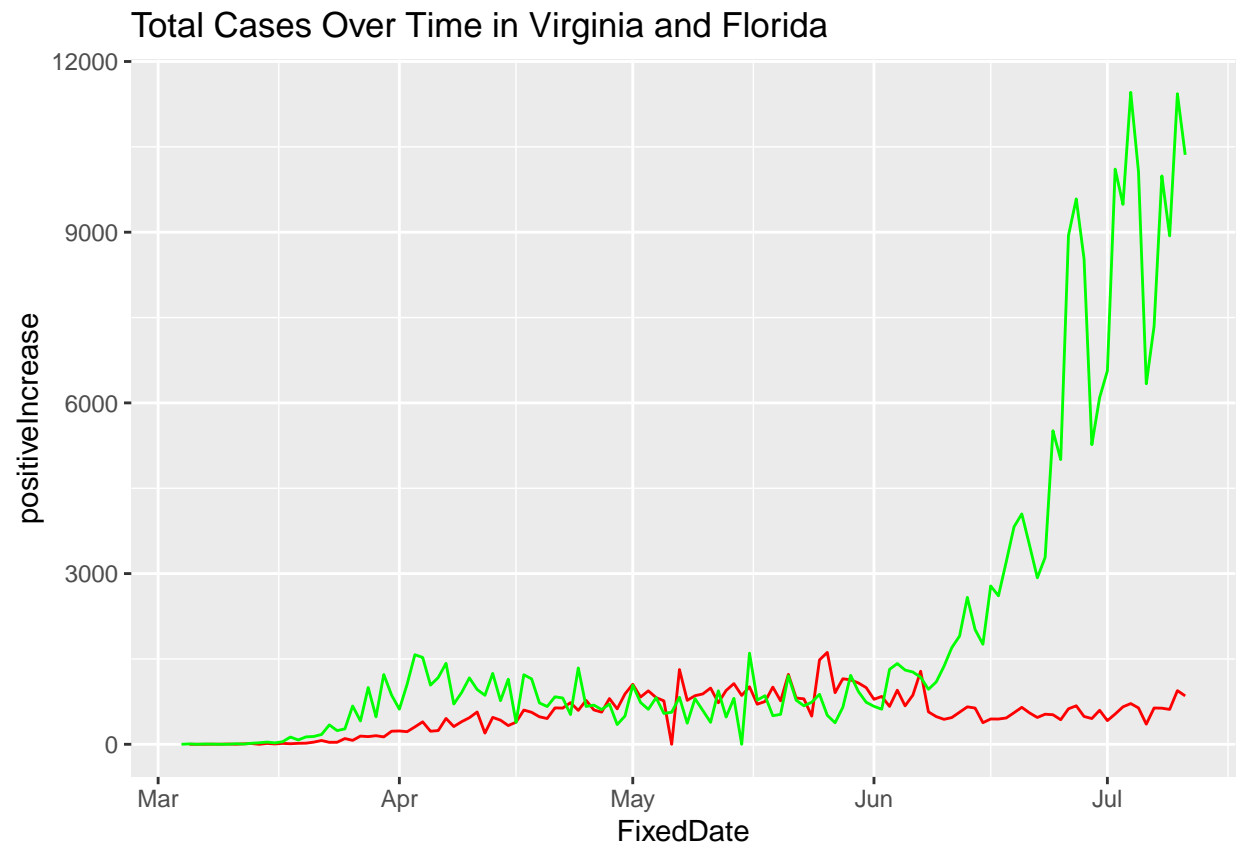
- Virginia: 8.8403864%
- Florida: 7.7012606%

It looks like we are a couple of percentage points below the threshold needed for herd immunity.

Total Cases Over Time

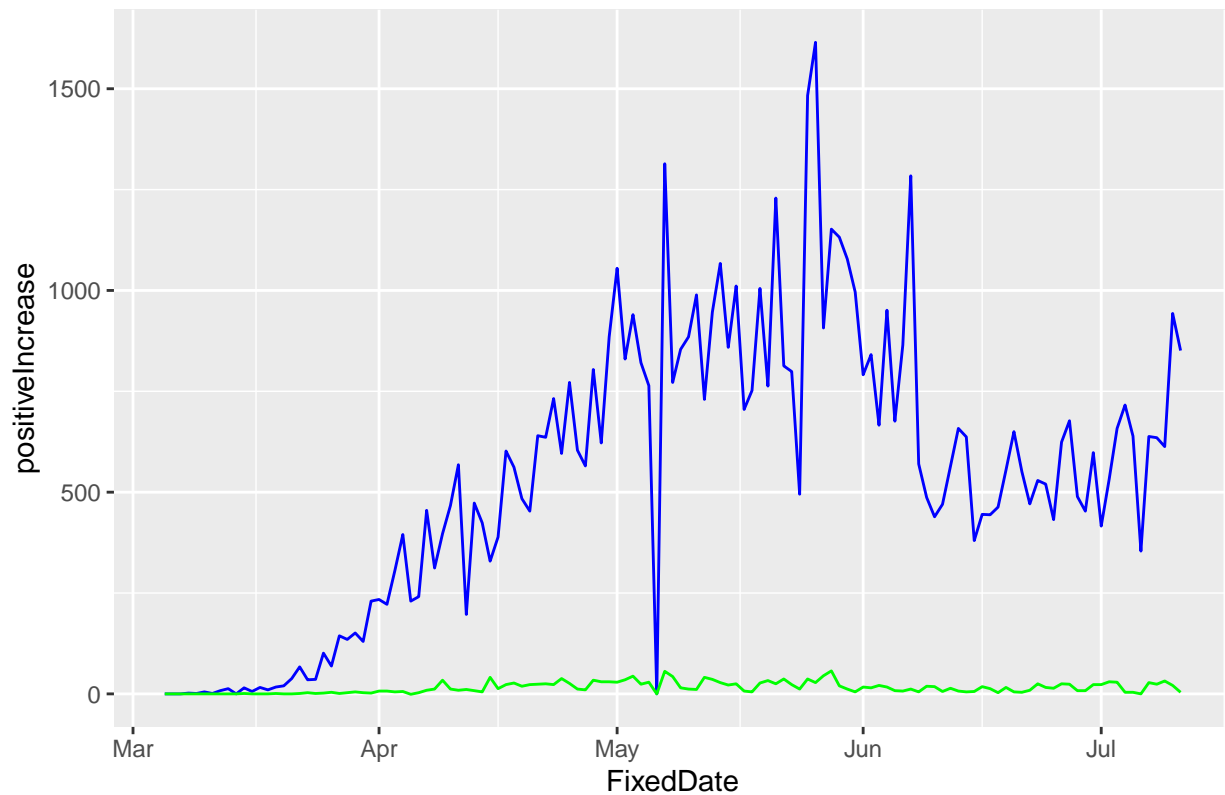
Some additional data gleaned from the data set:

```
totalCases_vs_Deaths_Plot = ggplot() +  
  geom_line(data = dataVirginia, aes(x = FixedDate, y = positiveIncrease), color='red') +  
  geom_line(data = dataFlorida, aes(x = FixedDate, y = positiveIncrease), color='green')  
print(totalCases_vs_Deaths_Plot +  
  ggtitle("Total Cases Over Time in Virginia and Florida"))
```



```
totalCases_Eastern_Plot <- ggplot() +
  geom_line(data = dataVirginia, aes(x = FixedDate, y = positiveIncrease), color='blue') +
  geom_line(data = dataVirginia, aes(x = FixedDate, y = deathIncrease), color = 'green')
print(totalCases_Eastern_Plot +
  ggtitle("Total Cases vs. Total Deaths, Virginia"))
```

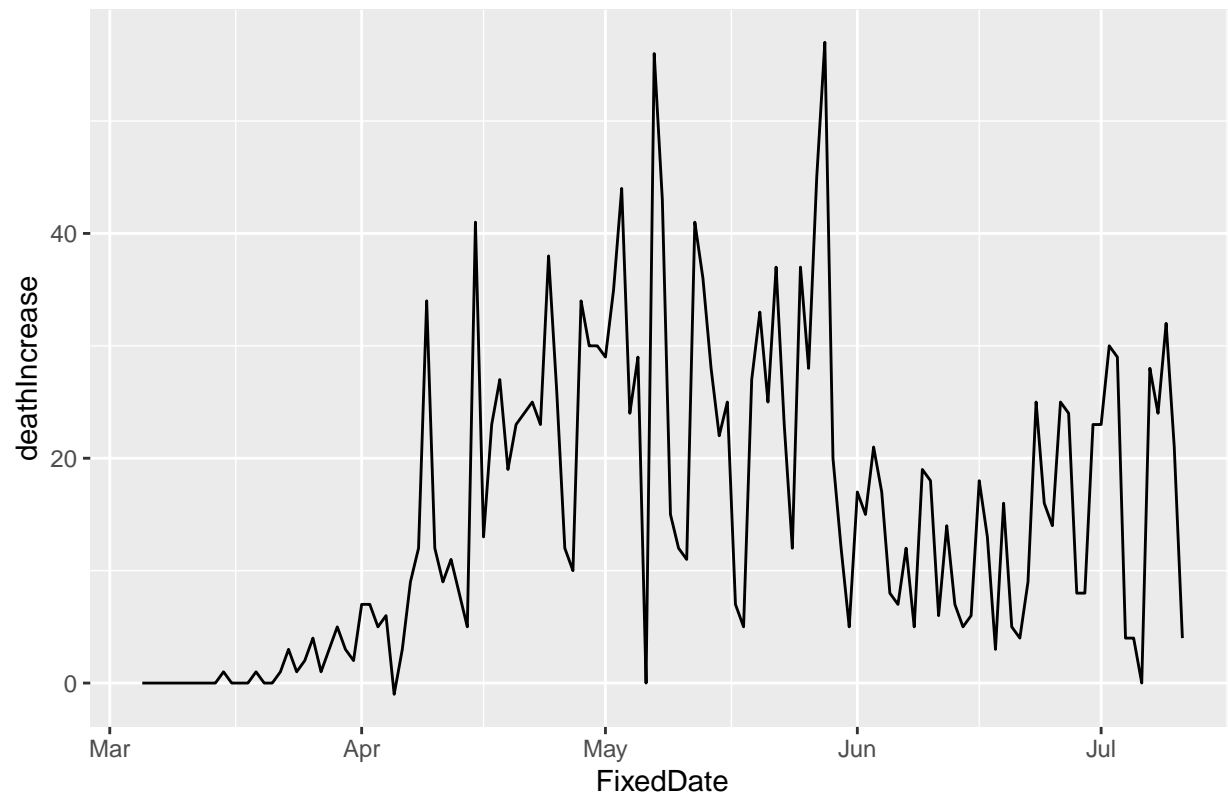
Total Cases vs. Total Deaths, Virginia



We can also separate out the graph of deaths related to COVID to more closely examine the trend. Although the number of cases is rising, the number of deaths is not.

```
totalDeaths_Eastern_Plot = ggplot(data = dataVirginia, aes(x = FixedDate, y=deathIncrease)) +  
  geom_line()  
print(totalDeaths_Eastern_Plot +  
  ggtitle("Total Deaths, Virginia"))
```

Total Deaths, Virginia



Some additional plots for comparison

```
totalCases_Central_Plot = ggplot(data = dataVirginia, aes(x = FixedDate, y = positiveIncrease)) +  
  geom_line()  
print(totalCases_Central_Plot +  
  ggtitle("Total Cases, Virginia"))
```

Total Cases, Virginia

