Covid Data Notebook

Goal

The purpose of this notebook is to show the analysis of Covid data from the Virginia Department of Health

```
library(ggplot2)
```

Preparing the data

Read in raw data as a data frame

```
rawData <- read.csv("VDH-COVID-19-PublicUseDataset-EventDate.csv", stringsAsFactors = FALSE)

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

Now we build two new data frames: one with confirmed cases and one with probable cases and then merge them together. Before we merge them we change the column headings of the 2 dataframes. We also need to replace all of the NA values with 0's.

```
# Build a data frame with just the confirmed cases
confirmedCases <- subset(rawData, Case.Status == "Confirmed")

# Build a data frame with just the probable cases
probableCases <- subset(rawData, Case.Status == "Probable")

# Change the column headings to reflect the two data sets
col_headings          <- c('Event.Date',
                           'Health.Planning.Region',
                           'Case.Status - Confirmed',
                           'Cases-confirmed',
                           'Hospitalizations-confirmed',
                           'Deaths-confirmed',
                           'FixedDate')
col_headings2         <- c('Event.Date',
                           'Health.Planning.Region',
                           'Case.Status - Probable',
                           'Cases-probable',
                           'Hospitalizations-probable',
                           'Deaths-probable',
                           'FixedDate')
names(confirmedCases) <- col_headings
names(probableCases)  <- col_headings2

# Merge the two dataframes and build the new columns combining the data
mergedData            <- merge(confirmedCases,
                               probableCases,
                               by = c("FixedDate","Health.Planning.Region"),
                               all.x = TRUE,
```

```
                                all.y = TRUE)
mergedData[is.na(mergedData)] <- 0 # Replace all of the NA's with 0
```

Here we build a new column with the total of confirmed and probable cases, deaths and hospitalizations.

```
mergedData$totalCases         <- mergedData$'Cases-confirmed' +
  mergedData$'Cases-probable'
mergedData$totalHospitalized  <- mergedData$`Hospitalizations-confirmed` +
  mergedData$`Hospitalizations-probable`
mergedData$totalDeaths        <- mergedData$`Deaths-confirmed` +
  mergedData$`Deaths-probable`
```

Build 3 new dataframes for the different regions in Virginia:

```
casesEastern   <- subset(mergedData, mergedData$Health.Planning.Region == "Eastern")
casesCentral   <- subset(mergedData, mergedData$Health.Planning.Region == "Central")
casesNorthern  <- subset(mergedData, mergedData$Health.Planning.Region == "Northern")
```

Calculating the Totals for the State

```
totalCases            = sum(rawData$'Number.of.Cases')
totalHospitalizations = sum(rawData$'Number.of.Hospitalizations')
totalDeaths           = sum(rawData$'Number.of.Deaths')
```

Totals for the State:

- Cases: 67964
- Hospitalizations: 6223
- Deaths: 1896

Calculate heard immunity rate of Eastern Virginia

According to CDC nationwide IFR is 0.26%

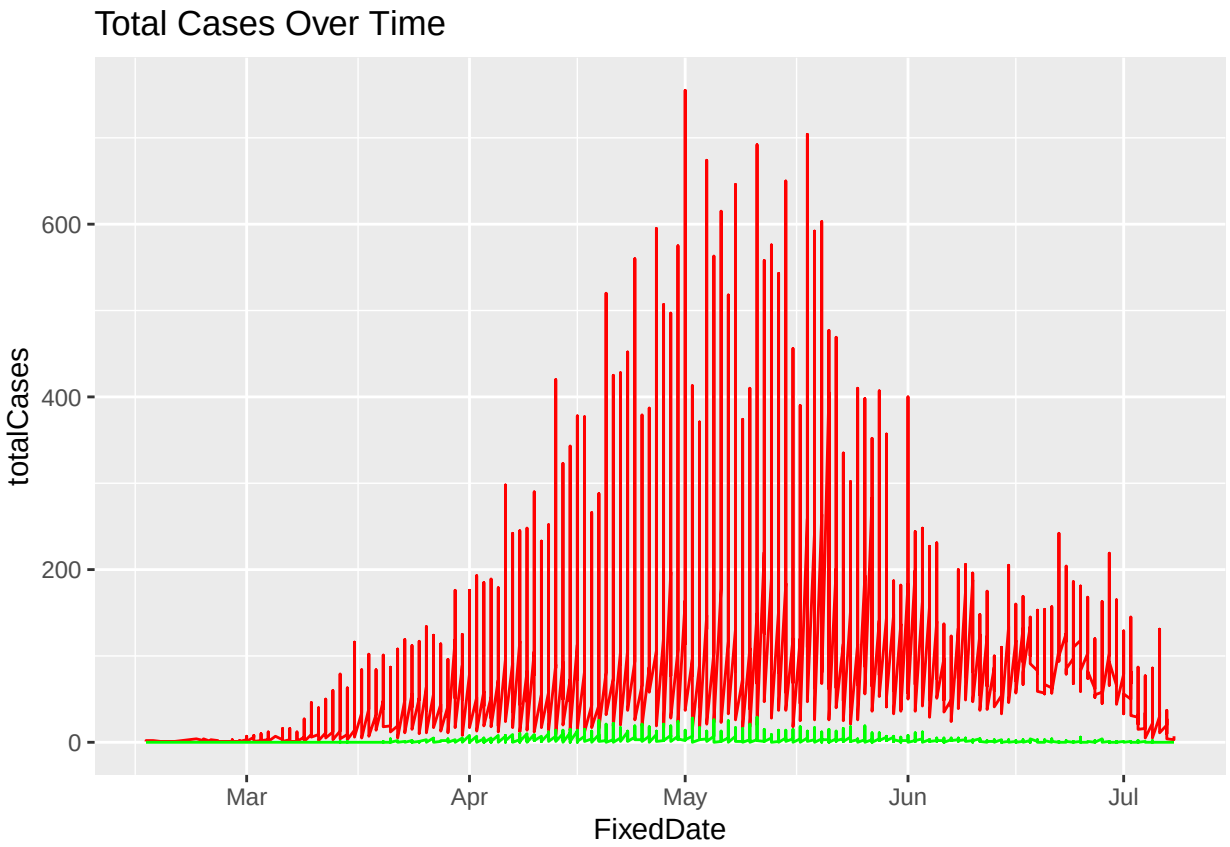
```
numberHadit = sum(casesEastern$'totalDeaths') / .0026
```

Individuals who have had Covid: 89230.77

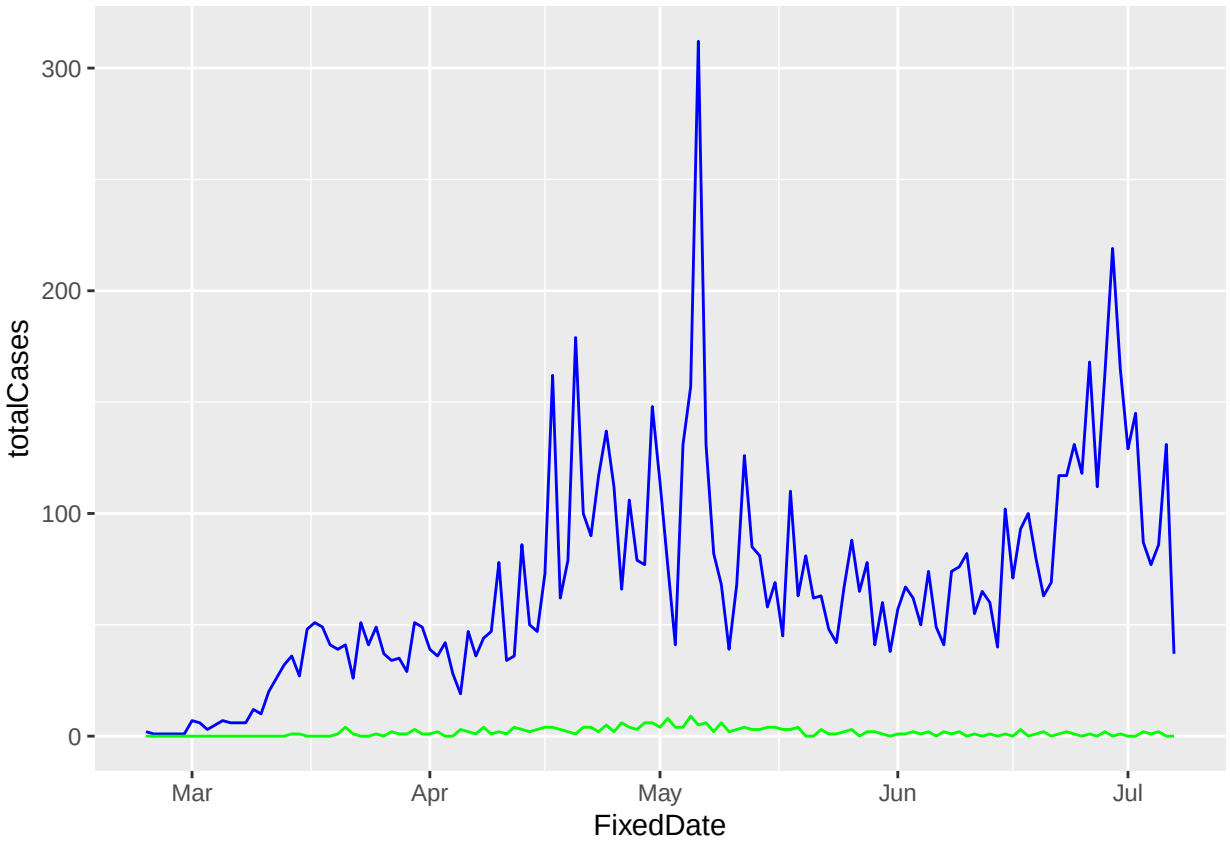
Total Population in the Eastern Planning Region: 1,854,806. Percentage of individuals who have had Covid: 0.0481079

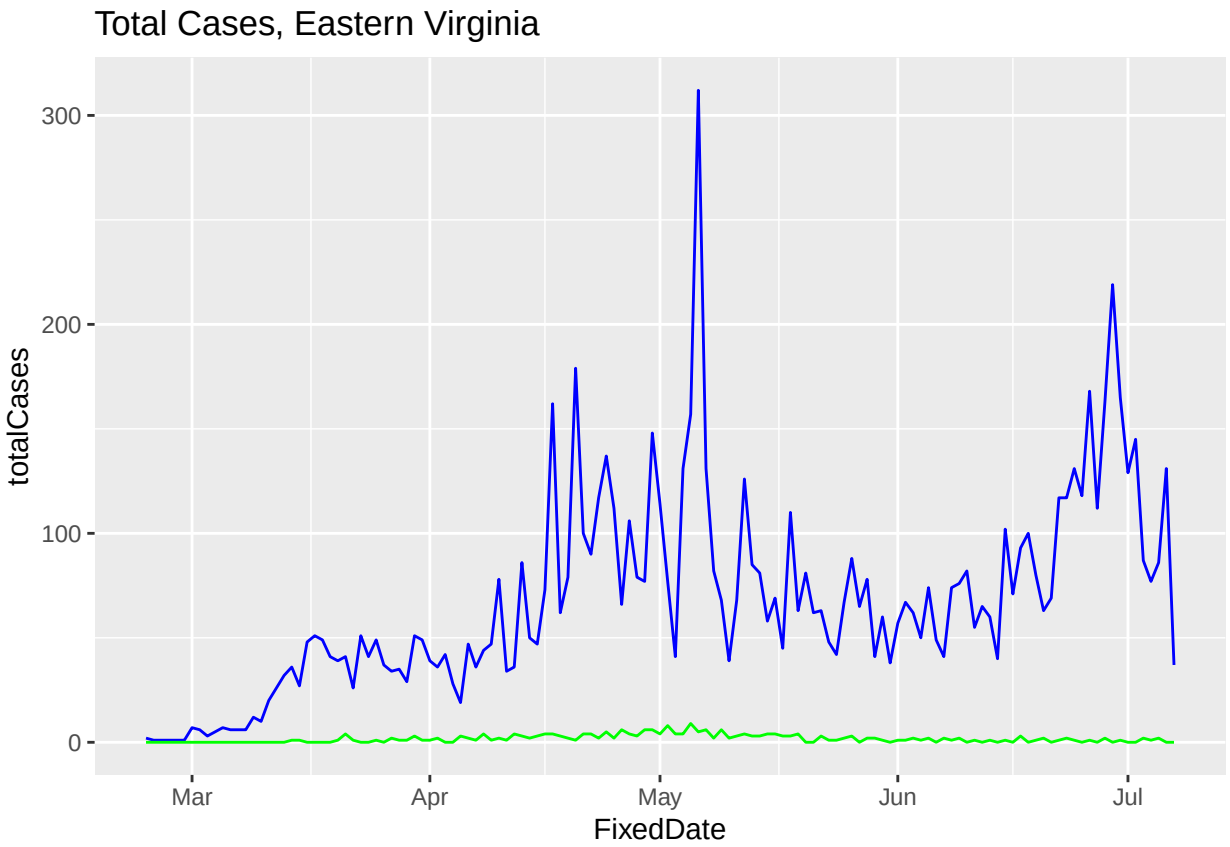
Total Cases Over Time

```
totalCases_vs_Deaths_Plot = ggplot() + geom_line(data = mergedData, aes(x = FixedDate, y = totalCases),
print(totalCases_vs_Deaths_Plot + ggtitle("Total Cases Over Time"))
```

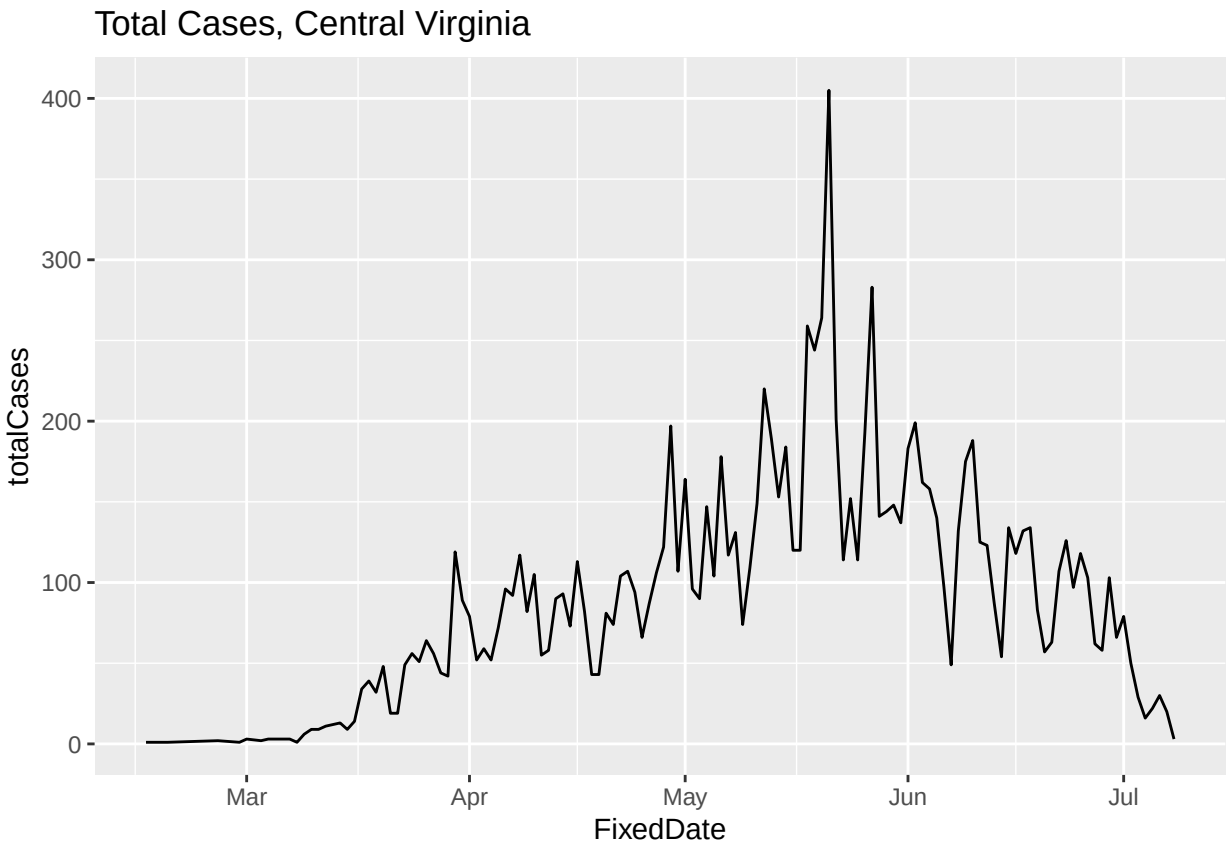



```
totalCases_Eastern_Plot = ggplot() + geom_line(data = casesEastern, aes(x = FixedDate, y = totalCases),
print(totalCases_Eastern_Plot) + ggtitle("Total Cases, Eastern Virginia")
```

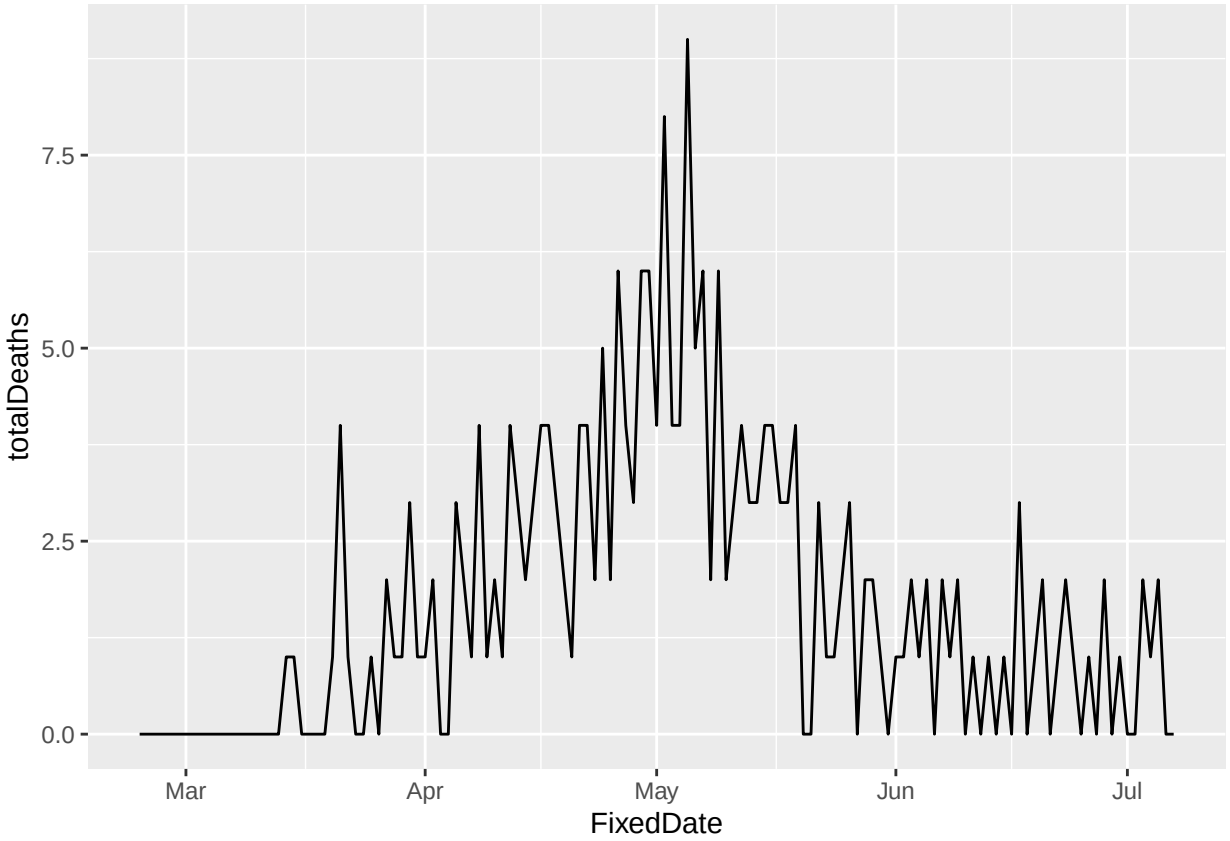

300
200
100
0
totalCases
Mar
Apr
May
Jun
Jul
FixedDate

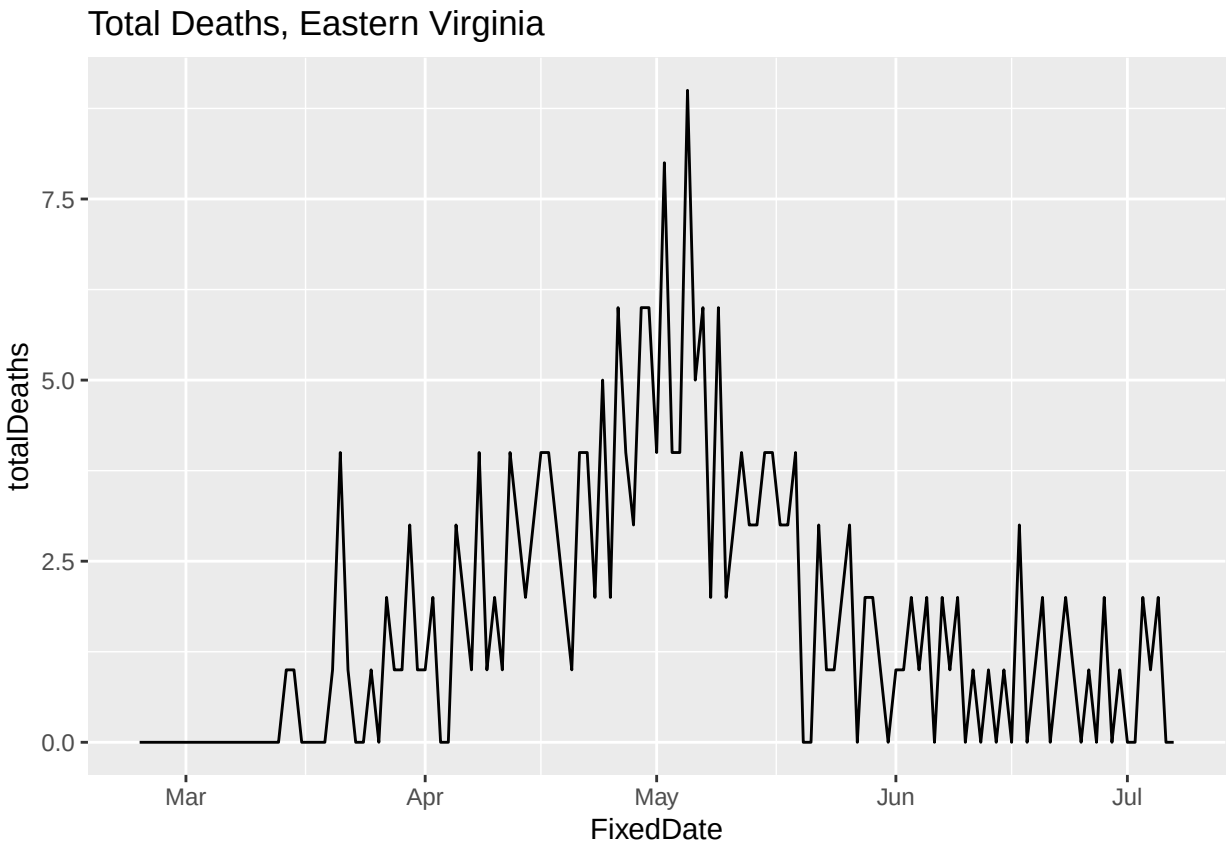



```
totalCases_Central_Plot = ggplot(data = casesCentral, aes(x = FixedDate, y = totalCases)) + geom_line()
print(totalCases_Central_Plot + ggtitle("Total Cases, Central Virginia"))
```

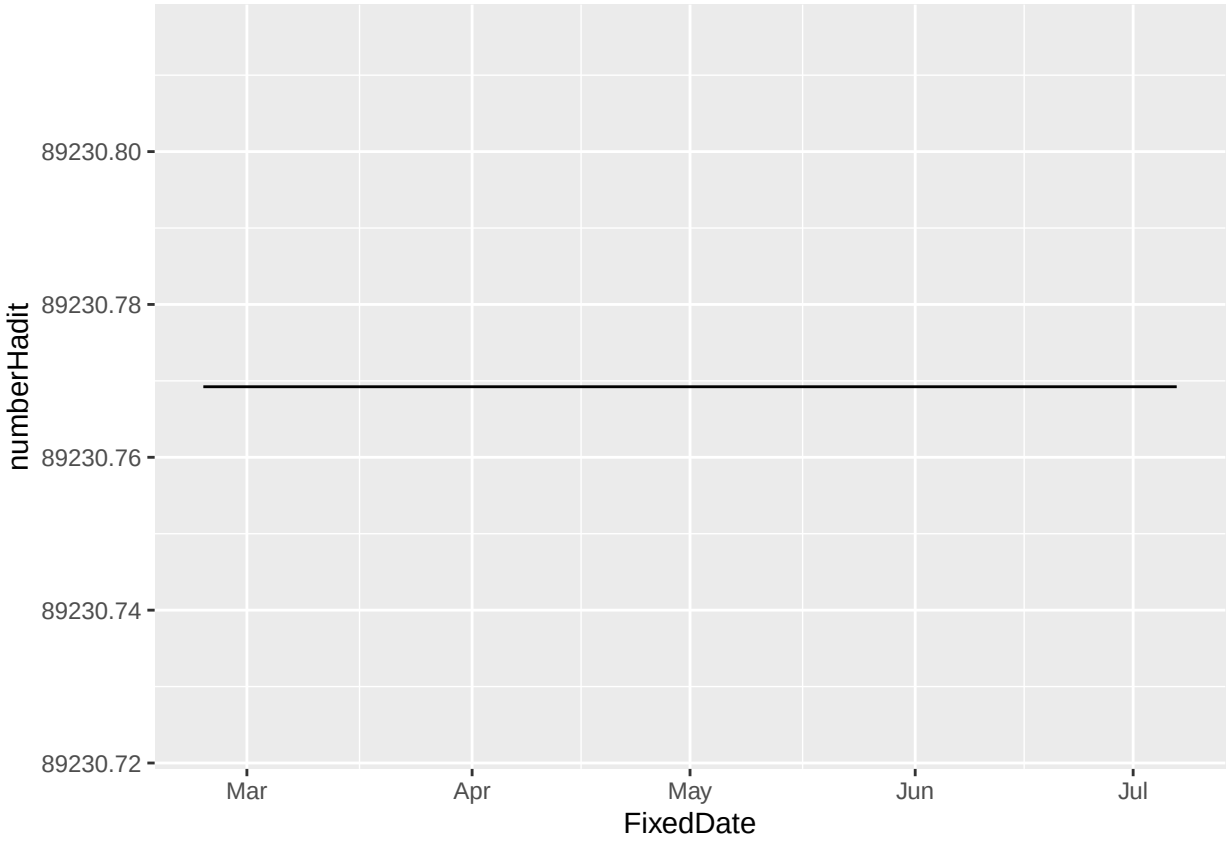


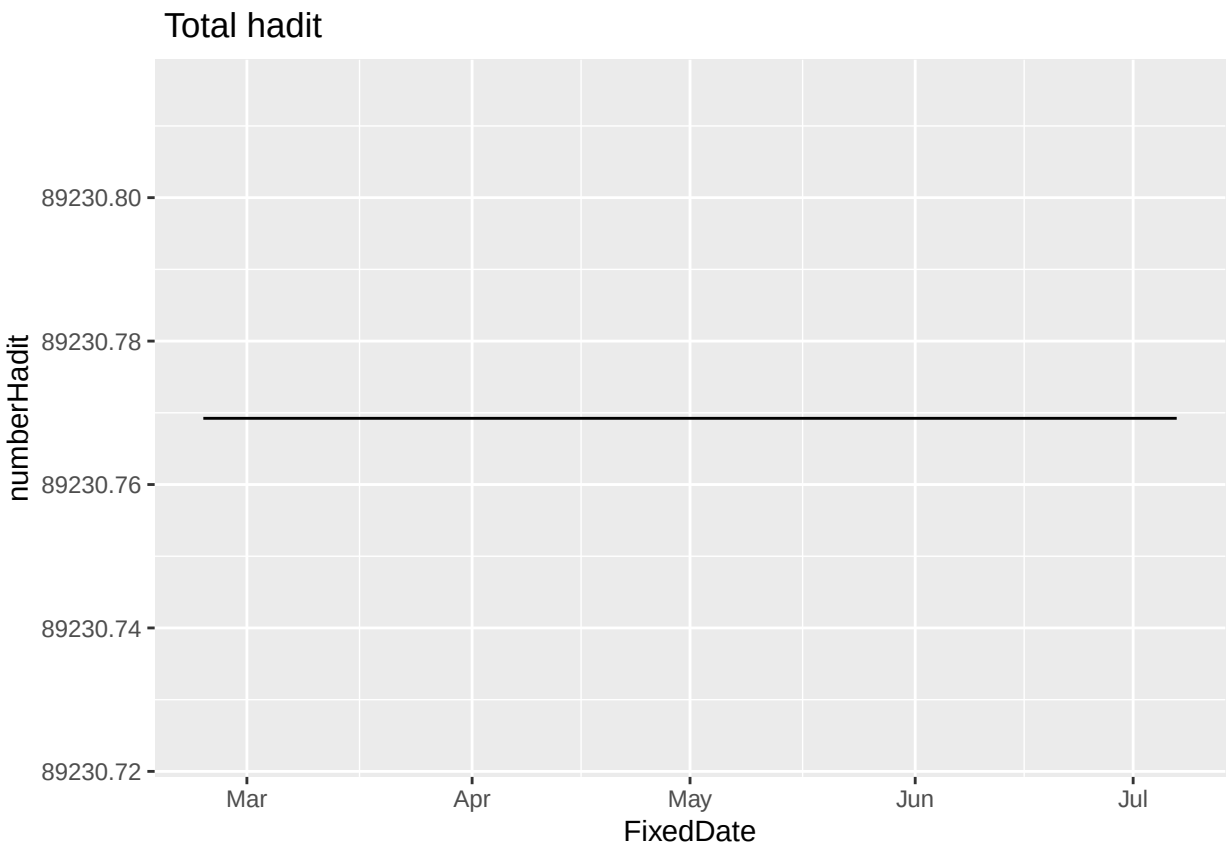

```
totalDeaths_Eastern_Plot = ggplot(data = casesEastern, aes(x = FixedDate, y = totalDeaths)) + geom_line
print(totalDeaths_Eastern_Plot) + ggtitle("Total Deaths, Eastern Virginia")
```


7.5
5.0
2.5
0.0
totalDeaths
Mar
Apr
May
Jun
Jul
FixedDate




```
percent_hadit_Eastern = ggplot(data = casesEastern, aes(x = FixedDate, y = numberHadit)) + geom_line()
print(percent_hadit_Eastern) + ggtitle(" Total hadit")
```


89230.80
89230.78
89230.76
89230.74
89230.72
numberHadit
Mar
Apr
May
Jun
Jul
FixedDate




```
# Calculate heard immunity rate of Eastern Virginia, according to CDC nationwide IFR is 0.26%
```