DSCI310 Final Project

Nisi Mohan Kuniyil 300321388

09/12/2020

Problem Statement

Analysis of COVID-19 and the impact of COVID on US Stock Market Indices

The novel coronavirus is taking a toll on people's lives every day. The pandemic's impact on the global economy has been significant. Countries going into complete lockdowns in order to reduce the spread of the virus has had a major adverse effect on the economy overall. Analyzing COVID factors such as Infection rate, the mortality rate would give an overall idea about the trend of the disease over the past months. The intend of this project is to analyze the spread of the disease in the US compared to a few other major countries where a serious spread was observed. Also, analyze the impact of various factors related to COVID on the US stock market by studying a few indices.

Solution

Question1:

First, We study the COVID curve in the US comparing with Spain, Japan, Italy, and China. The below curve shows the cumulative confirmed cases of all the countries in the study between January 2020 and December 2020. From this, we can see that there is an exponential growth in the number of cases in the US compared to other countries. While the curve has started to flatten in other countries, in the US, the number of cases is still rising.

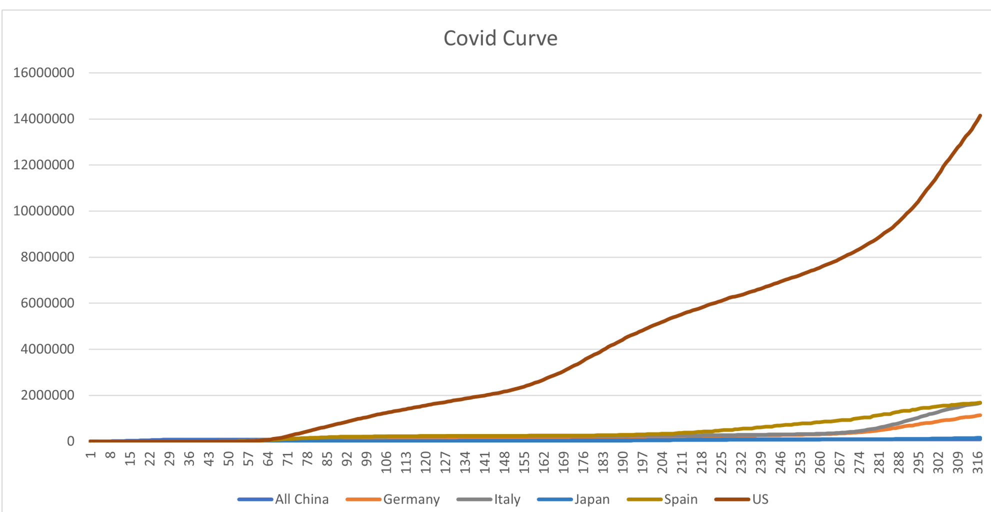


Figure 1: COVID Curve

This significant difference in the US can be clearly observed by visualizing the "Daily Confirmed Cases" in all the countries. The US daily confirmed cases are still rising exponentially while in the other countries it has significantly come down.

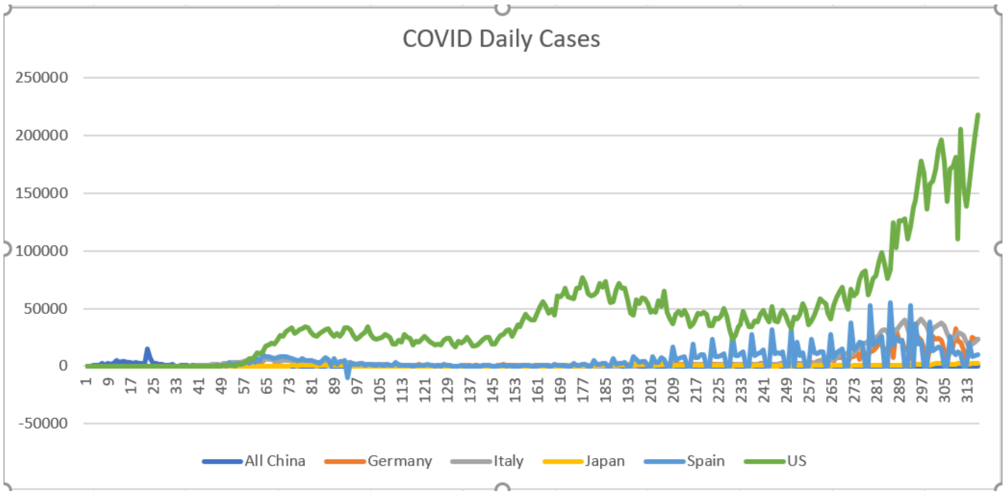


Figure 2: COVID Daily Cases

Furthermore, we can analyze the rate of spread by visualizing the daily percentage change in confirmed cases. The trend here shows that in terms of percentage the spread of the disease was significantly high at the beginning of the period.

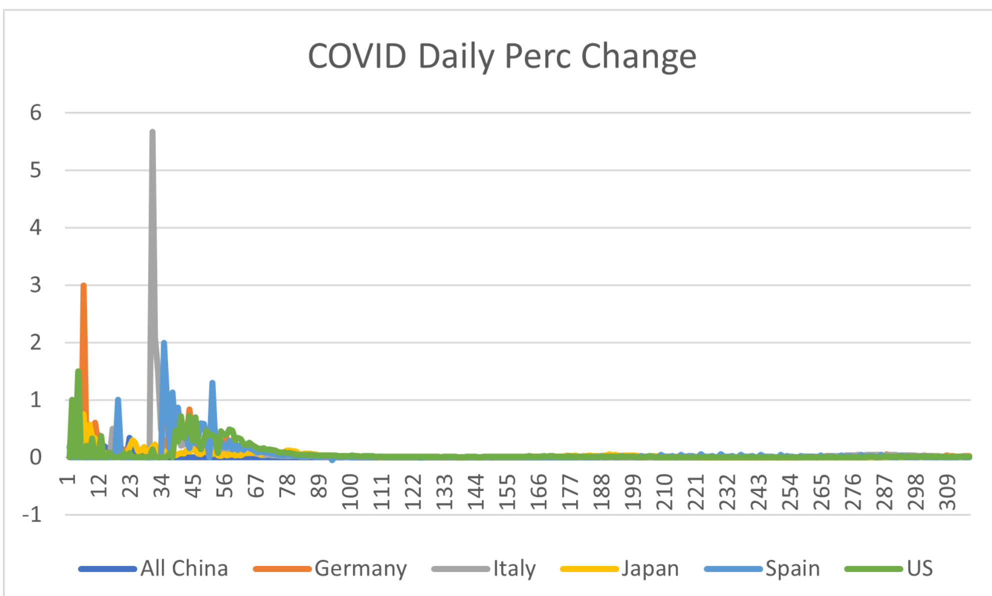


Figure 3: COVID Case Perc Change Curve

Next, we analyze the COVID death curve. Overall, this curve also follows the same trend as the total cases. The US is having a large number of deaths when taking the overall period from January 2020 to December 2020. Other countries have very less death reported due to COVID.

Below is a plot of the percentage change in death and the pattern is different than the percentage change in cases. Here percentage change in death is recorded more in the US during the initial period, whereas in the case of the percentage change in confirmed cases, Italy had the highest peak for percentage change.

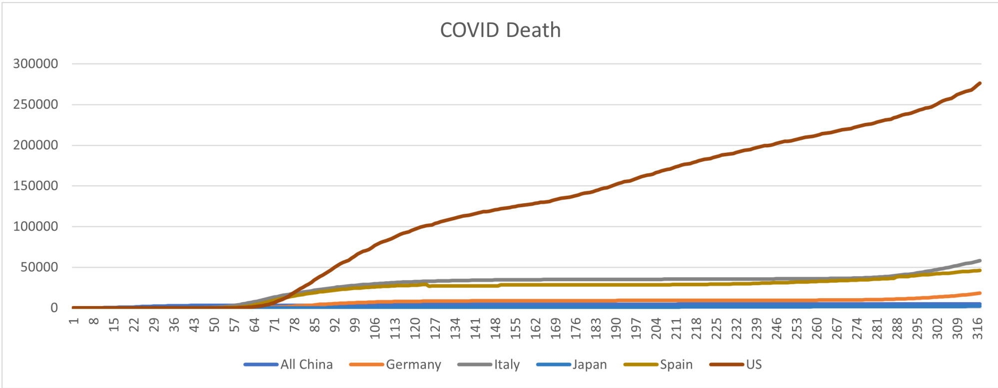


Figure 4: COVID Death Curve

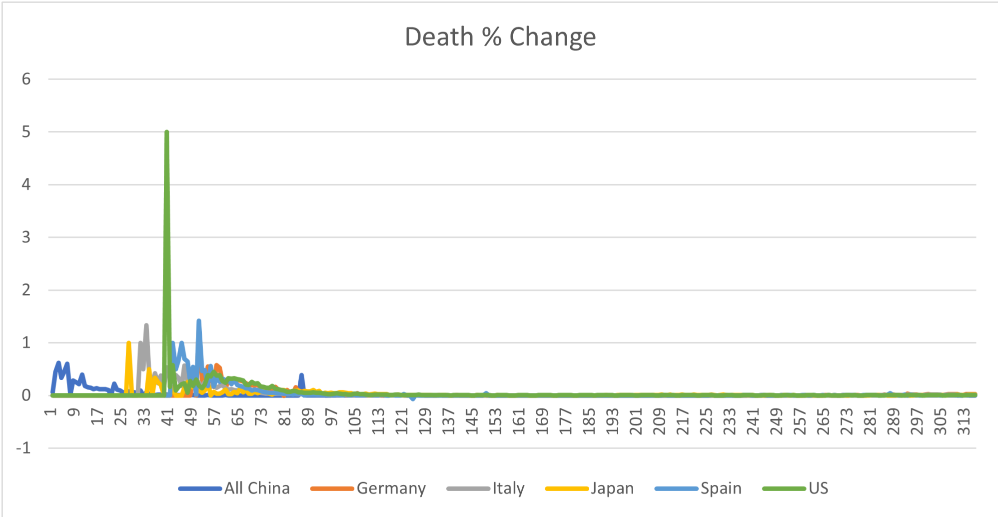


Figure 5: COVID Death Perc Change Curve

Next, we analyze the death rate of infected cases. From the below curve we could say that even though there was an initial peak in the US, Overall, Italy and Spain had the largest death rate recorded. The trend observed in all countries is similar. That is the death rate has come down significantly towards the end of the period.

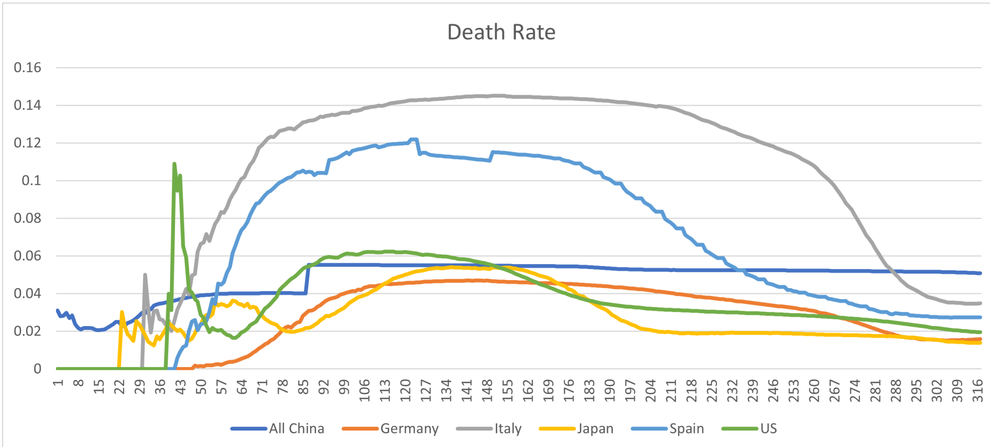
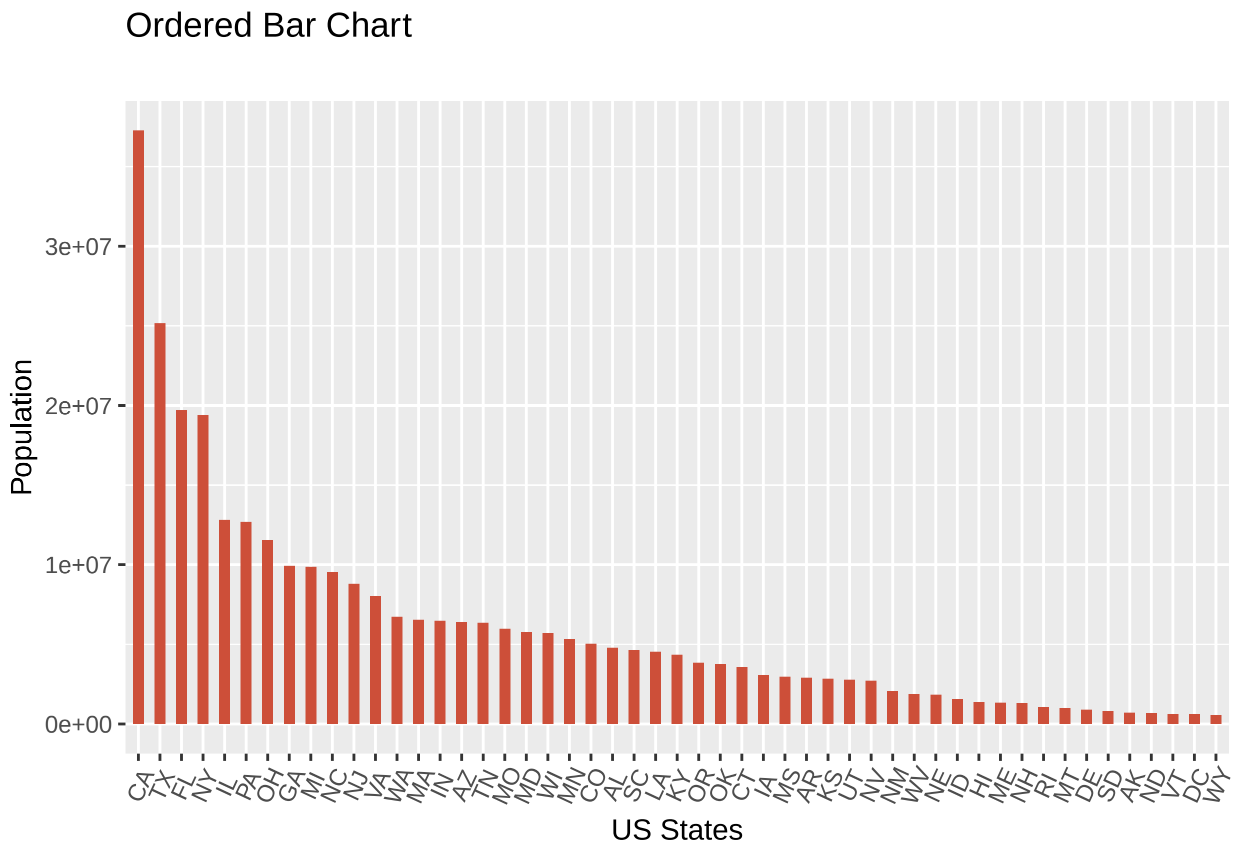


Figure 6: COVID Death Rate Curve

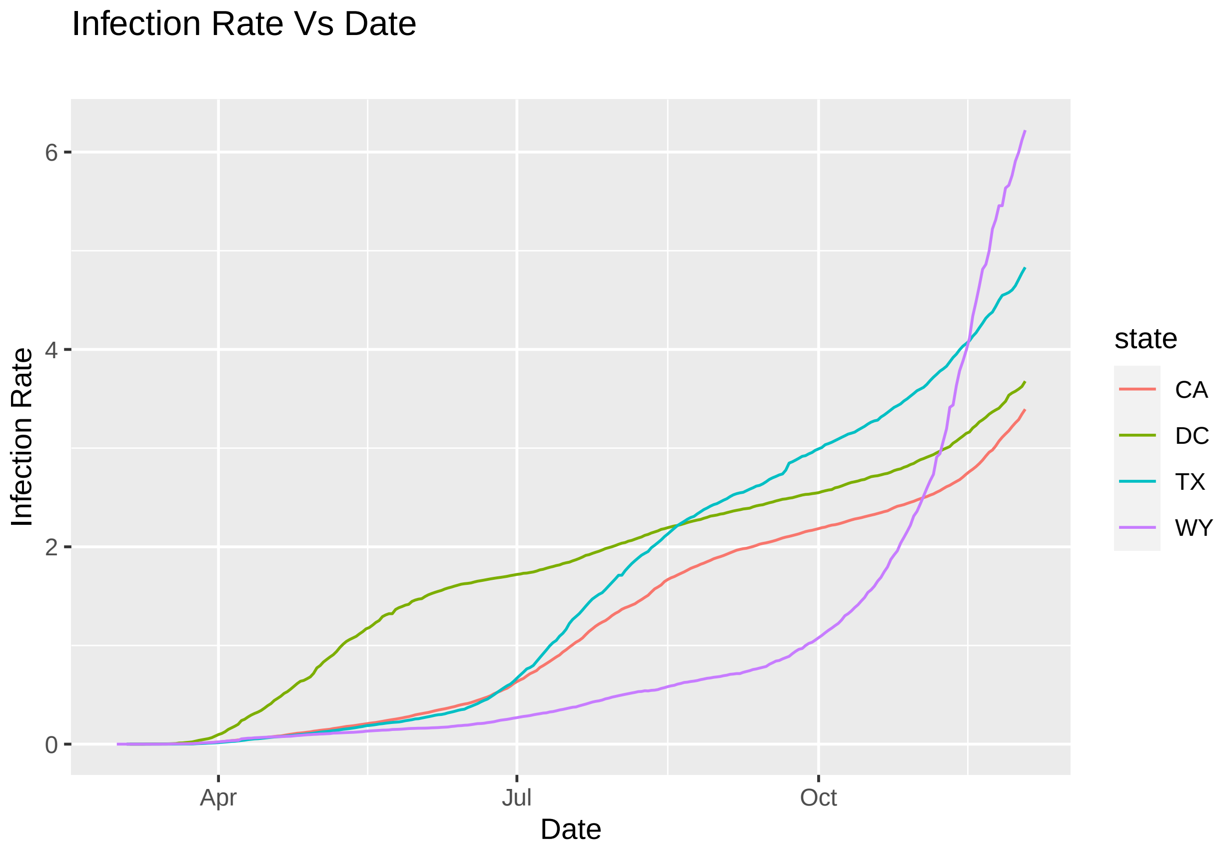
Question2:

Here we analyze different COVID factors state-wise. We also study the potential relationship between the population and the spread of the disease.

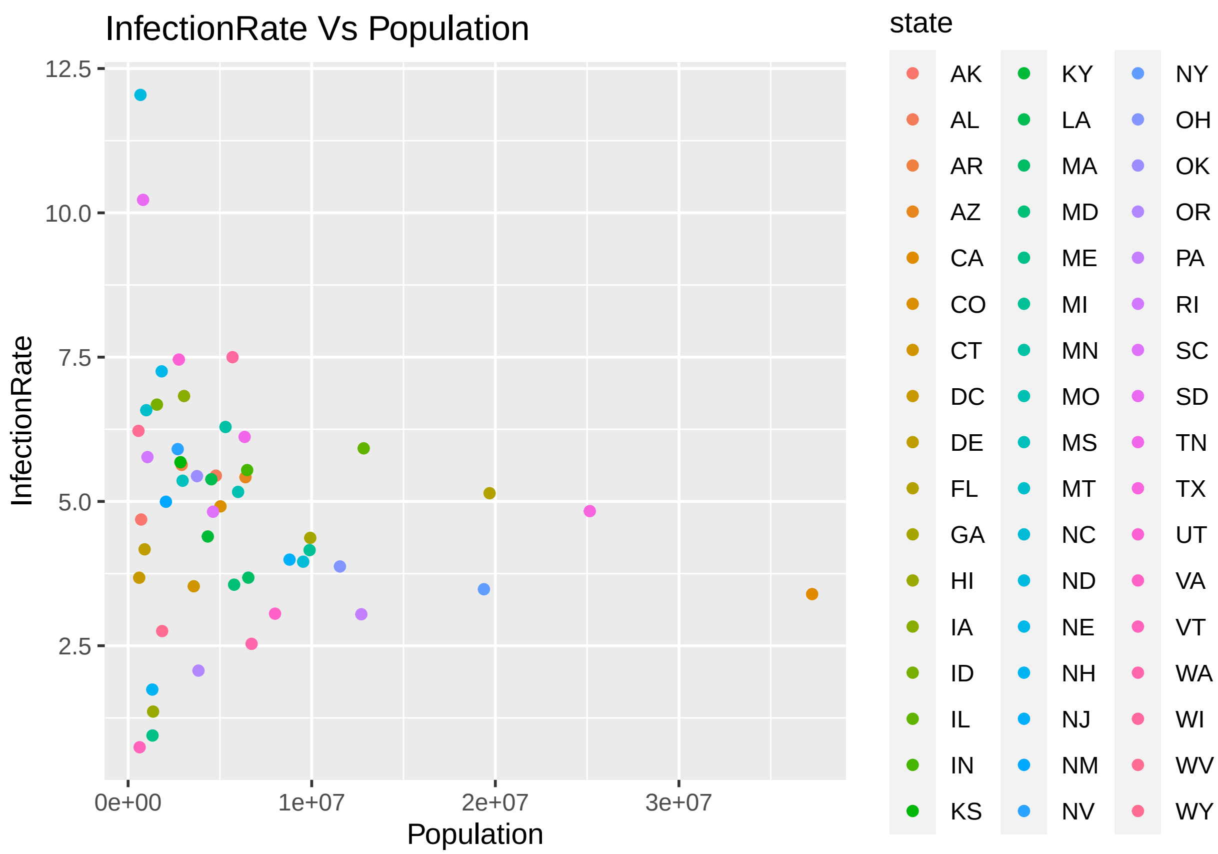
We start off by understanding the distribution of state-wise population in the US. From the below bar chart we can identify the highest and lowest population in states. To check if any relationship exists between infection rate and population, we choose California and Texas, which have the highest population, and DC and Wyoming which have the lowest population.



We then visualize the infection rates of the 4 states. From the line chart, we can only observe that the state with the lowest population size, Wyoming had a very low spread in the initial period but exponentially rose towards the end. From this chart, it is unable to come to the conclusion that any correlation exists between population and infection rate.

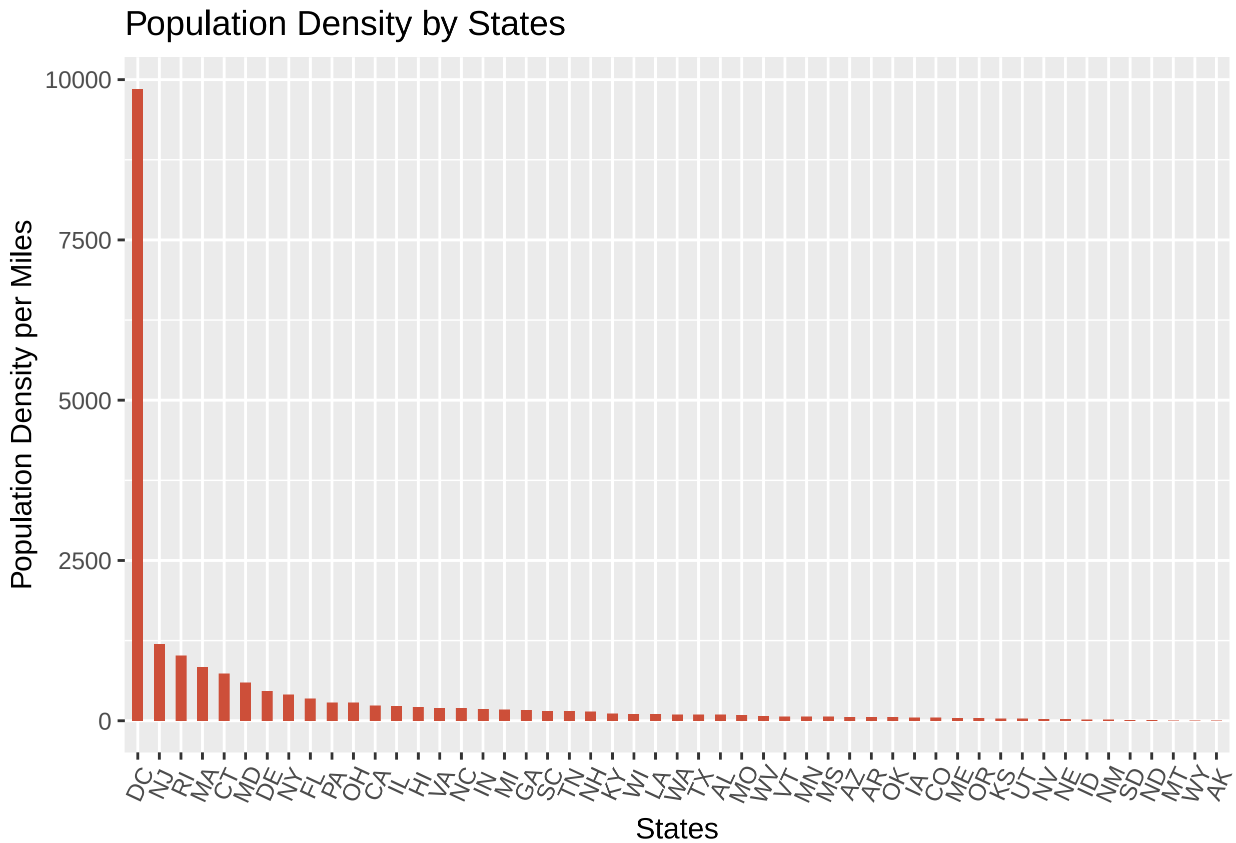


Furthermore, to get more idea about the relationship between infection rate and population we plot a scatter plot by taking the cumulative infection rate. We expect that when the population increases the infection rate will also increase. However, the below chart does not show strong evidence that supports our hypothesis.

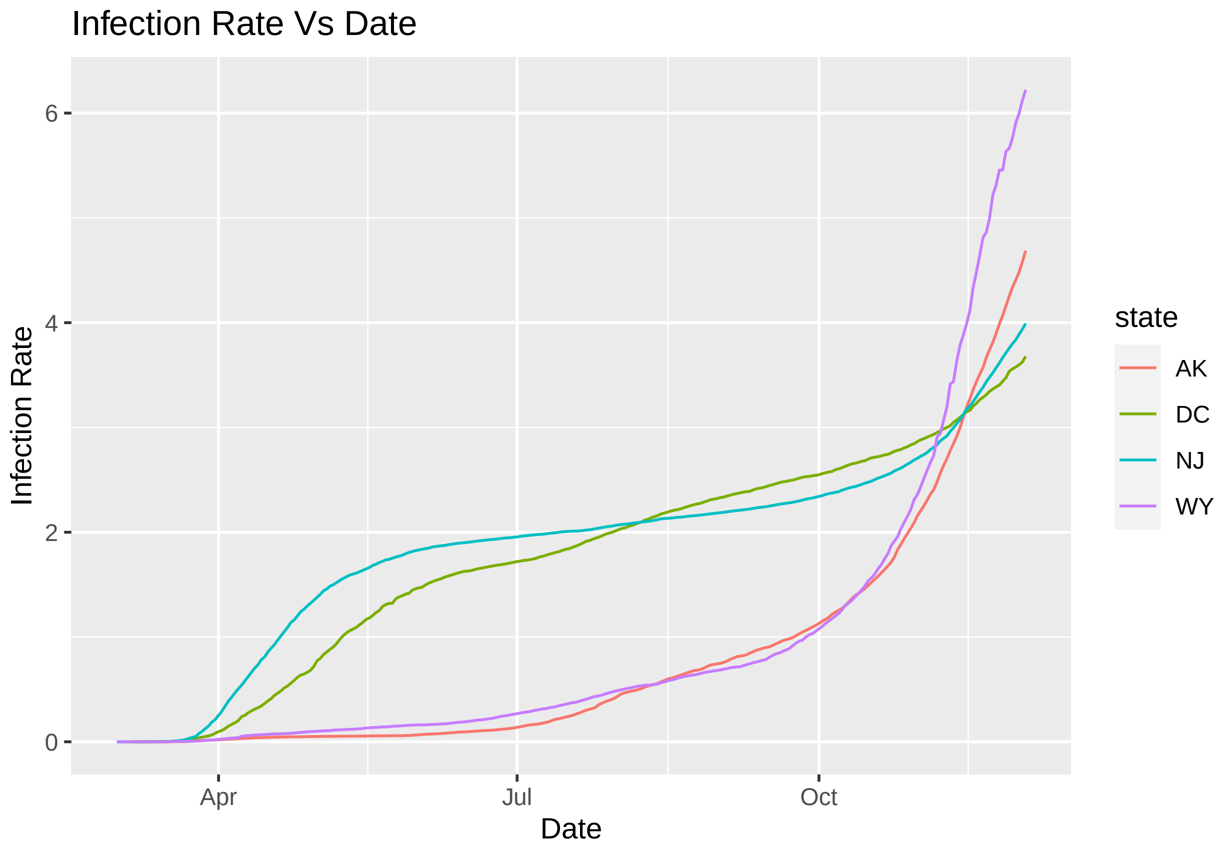


Population Density and Daily Infection Rate Analysis

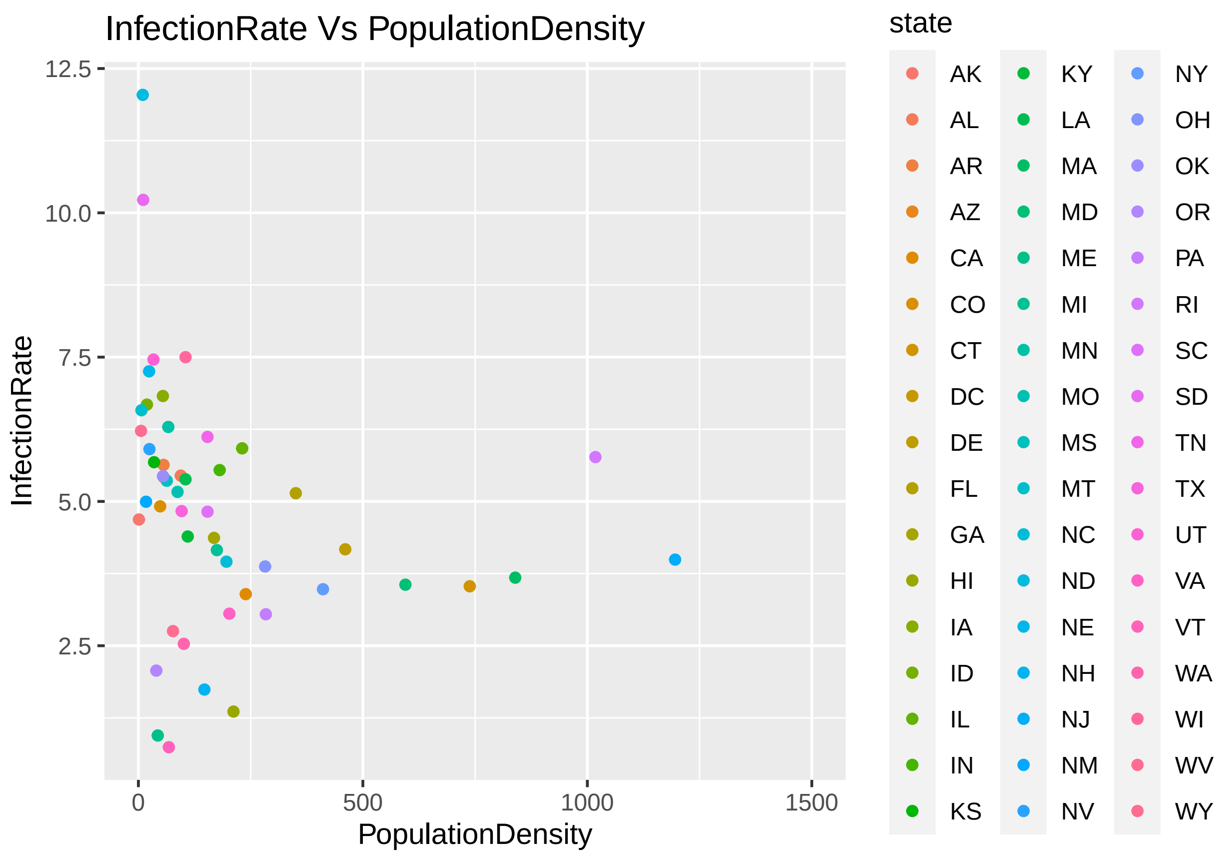
Since population density seems like a more accurate way to study the relationship between population and disease, we now compare the population density of all the states with infection rate. The below bar chart shows the distribution of population density by each state in the US. From this chart, we can identify that DC and New Jersey have the highest, and Wyoming and Alaska have the lowest population density.



A line chart is plotted to see if there is any relationship between population density and infection rate. We see the same trend that we had seen previously for the population chart here. Wyoming and Alaska are the least densely populated states had a very low infection rate during the initial stages, but later observed an exponential rise.



In addition, to get more idea about the relationship between infection rate and population density we plot a scatter plot by taking the cumulative infection rate. One might think that if the place is densely populated then the chance of infection rate will also be more. But, the chart shows no such relationship to conclude our assumption.

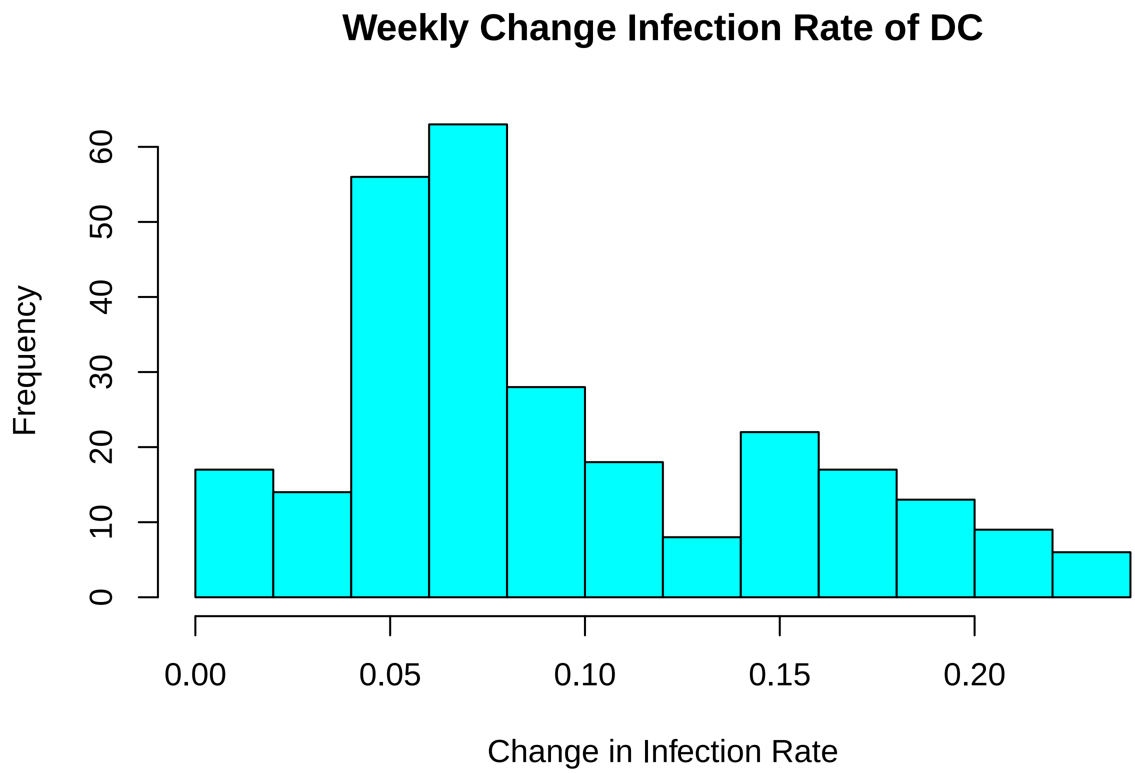


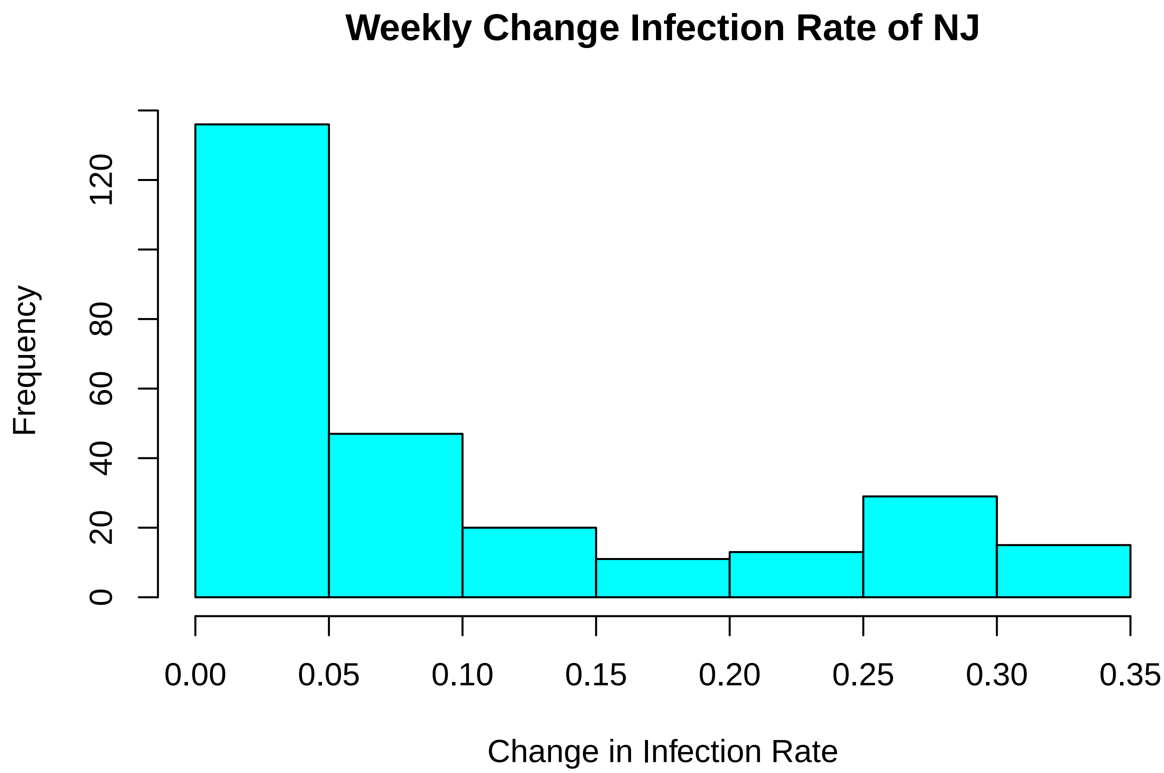
Infection Rate weekly change

From the time series analysis above we could not conclude any definitive relationship between infection rate and population, population density. So, here we will analyze the weekly change in infection rate which can also be considered as to how the disease is spreading week over week, and see if we can find a pattern between states with high population density and states with low population density.

1. DC & New Jersey - Highest Population Density

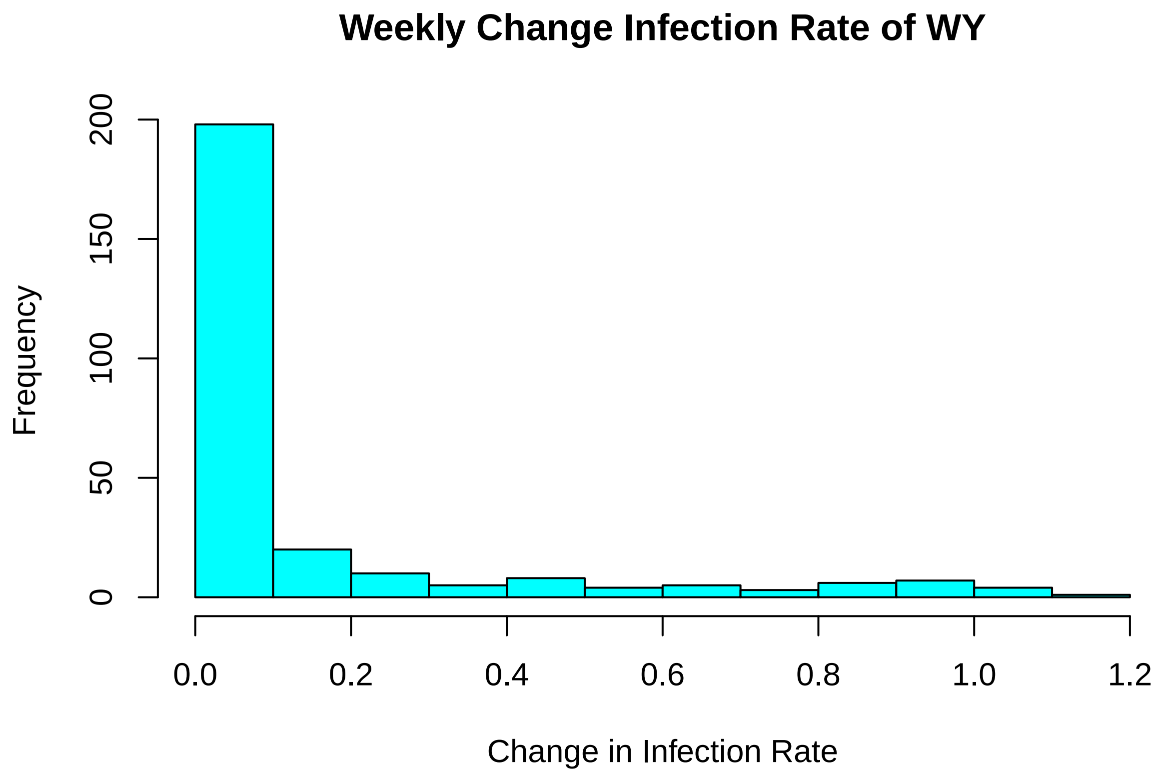
We can see the change of infection rate week over week is high. This indicates a faster spread where population density is higher.


Weekly Change Infection Rate of DC
Frequency
0
10
20
30
40
50
60
0.00
0.05
0.10
0.15
0.20
Change in Infection Rate

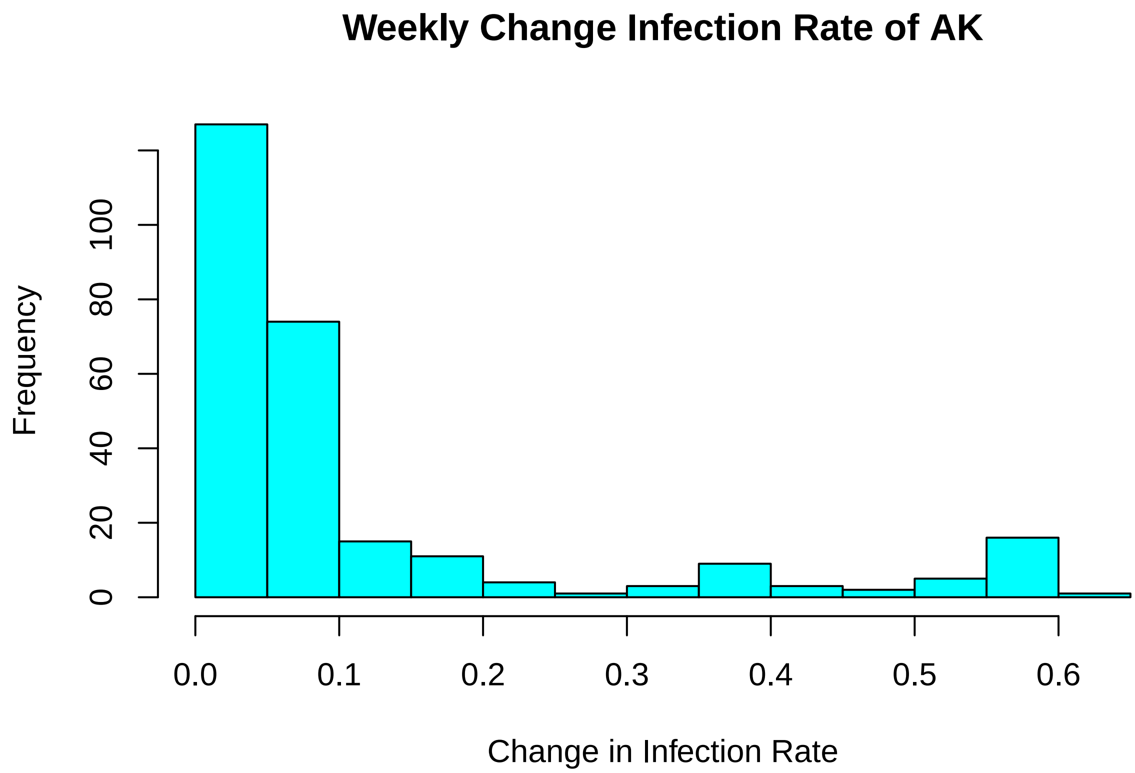


2. Wyoming & Alaska- Lowest Population Density

We can see the change of infection rate week over week is low mostly concentrated towards 0. This indicates a slower spread where population density is lower.



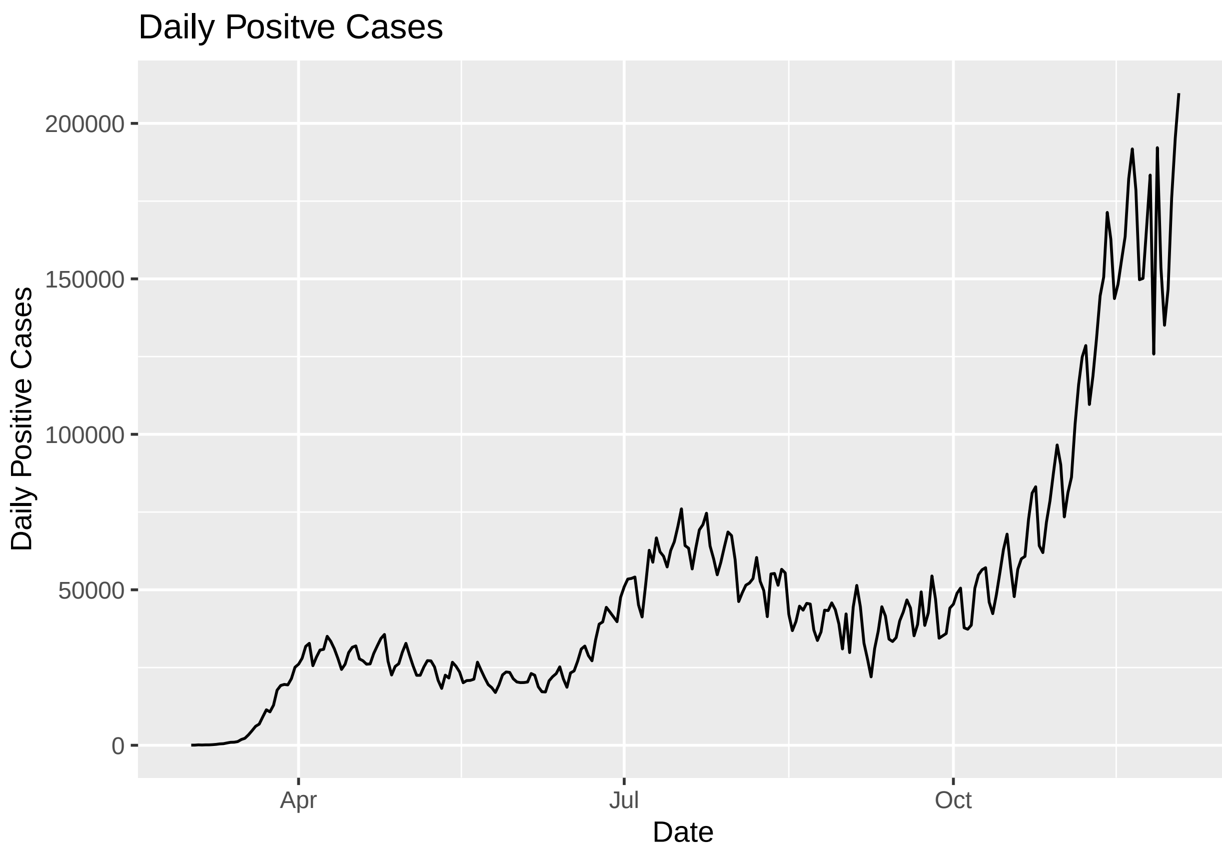
Weekly Change Infection Rate of WY
Frequency
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Change in Infection Rate

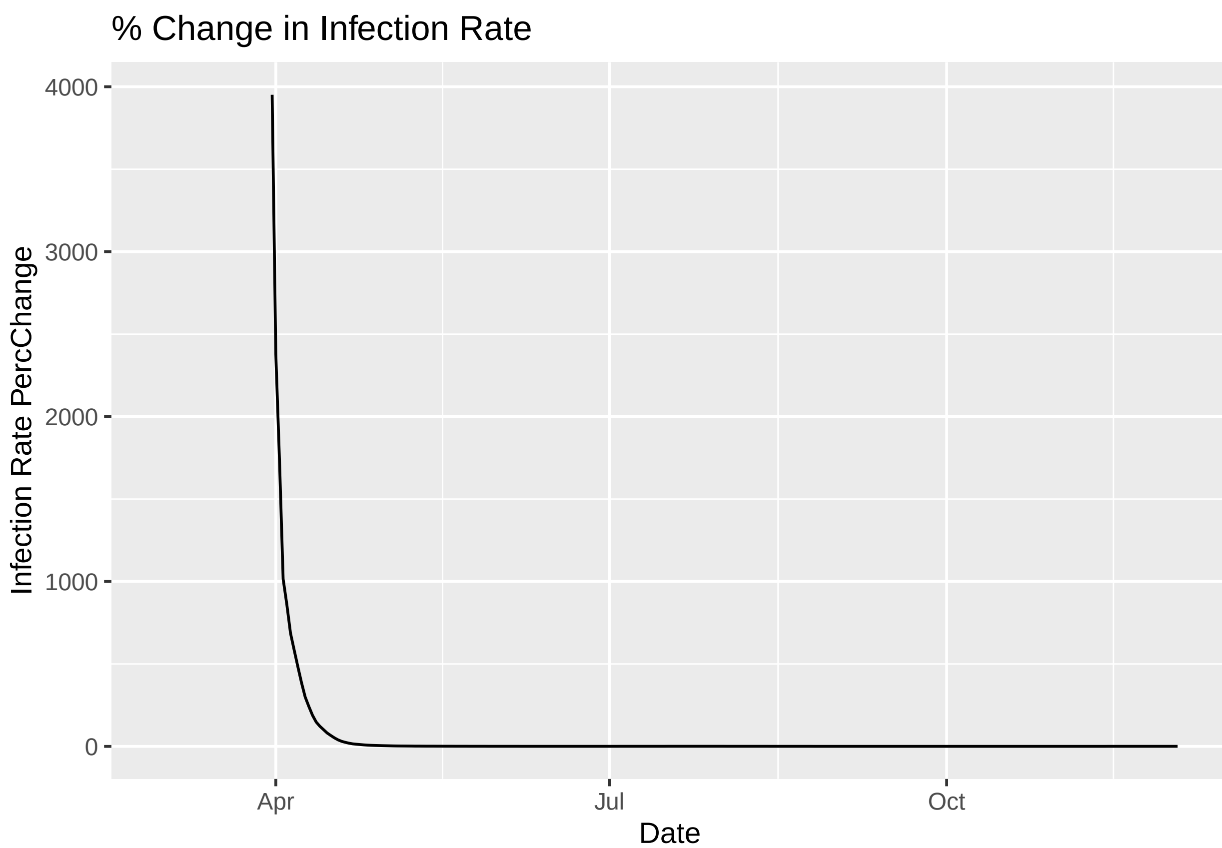

Weekly Change Infection Rate of AK
Frequency
0
20
40
60
80
100
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Change in Infection Rate

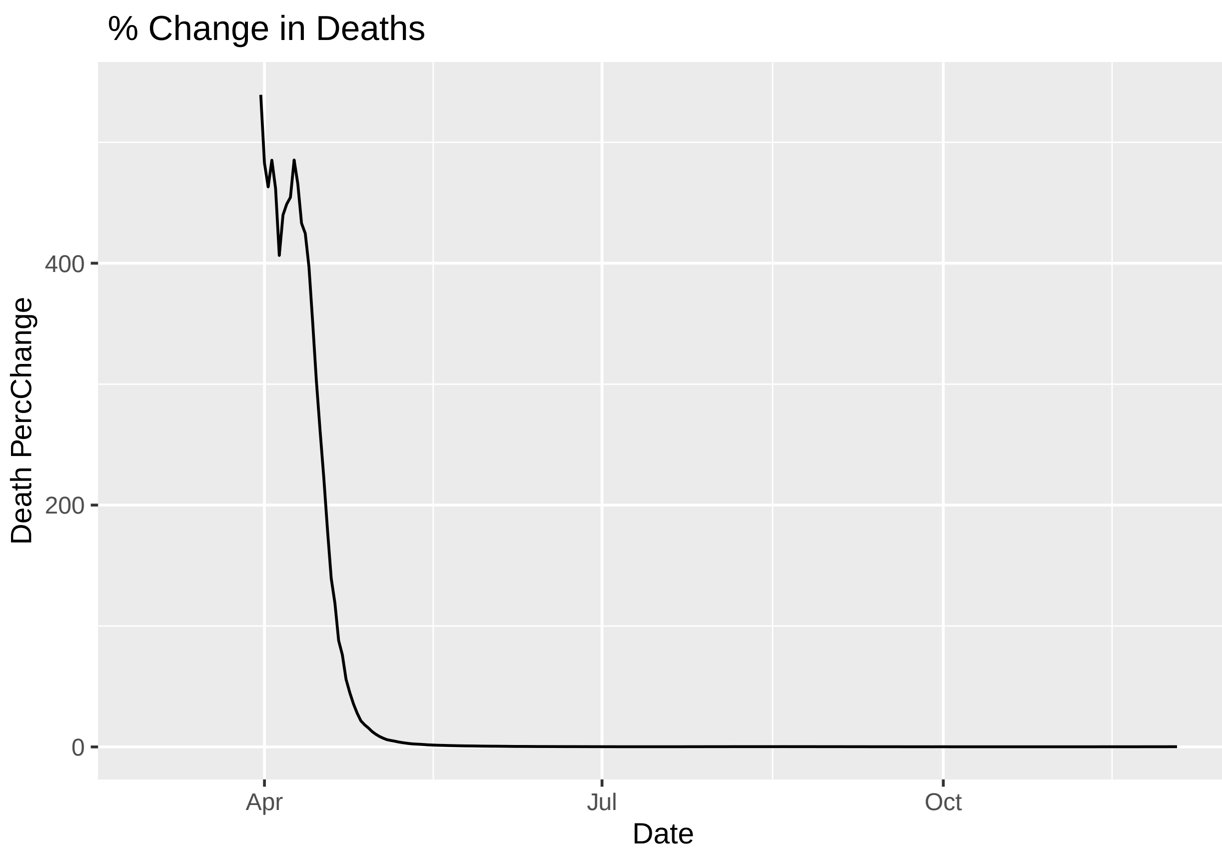
Question3:

Next, we analyze the impact of COVID-19 on the US stock exchange. For this analysis, we have taken the S&P 500, the NASDAQ, the Russel 2000, and Crude oil futures indices. We will study if any correlation exists between the COVID infection rate time series and the returns time series of the four assets.

First, we visualize the COVID-19 infection rate time series by plotting daily cases, the percentage change in cases, and the percent change in deaths for the period of March 2020 to December 2020.




% Change in Infection Rate
Infection Rate PercChange
4000
3000
2000
1000
0
Apr
Jul
Oct
Date

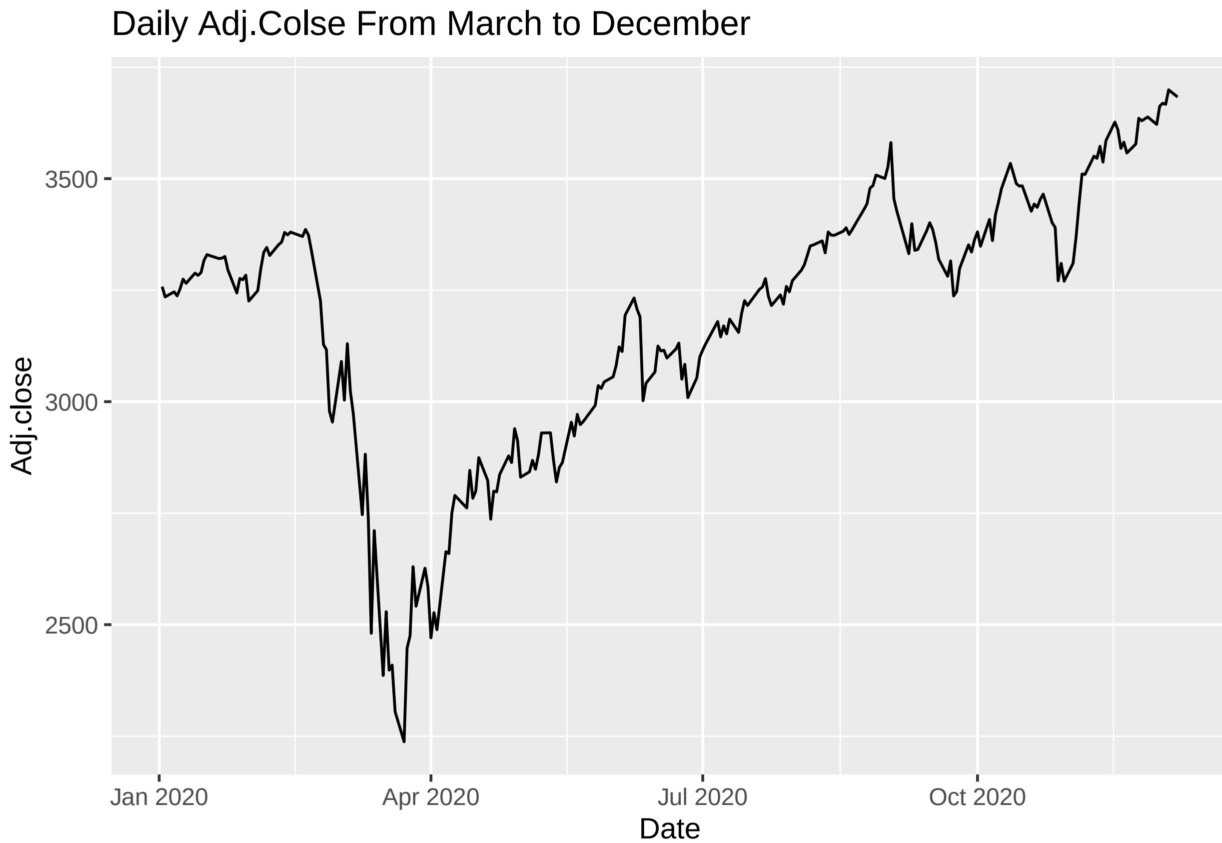


All of the above charts show a similar trend. It can be seen that the first peak is around April. We can say that it's because the first wave of COVID happened around this time. Also, the positive case chart is based on daily cases and the infection rate and death plots are thirty day percentage changes.

S&P 500

To see if COVID has any impact on the S&P 500 market index we plot the Adj.Close during the period from March 2020 to December 2020. The line chart shows a dip around April. We have already observed from the daily positive case plot that the first peak was around this time. So we could say that there was a market crash when the first wave hit. This is probably because of the lockdown that was imposed in almost every part of the USA during the first wave of COVID-19. However, the market seems to have recovered afterward.

To dig in deep, correlation of monthly percentage changes in infection rates and returns is computed. We have taken correlation of infectionrates percentage change and returns from March to December and we get correlation metric as **-0.5811911**. This value tells us there is a negative correlation between those two.

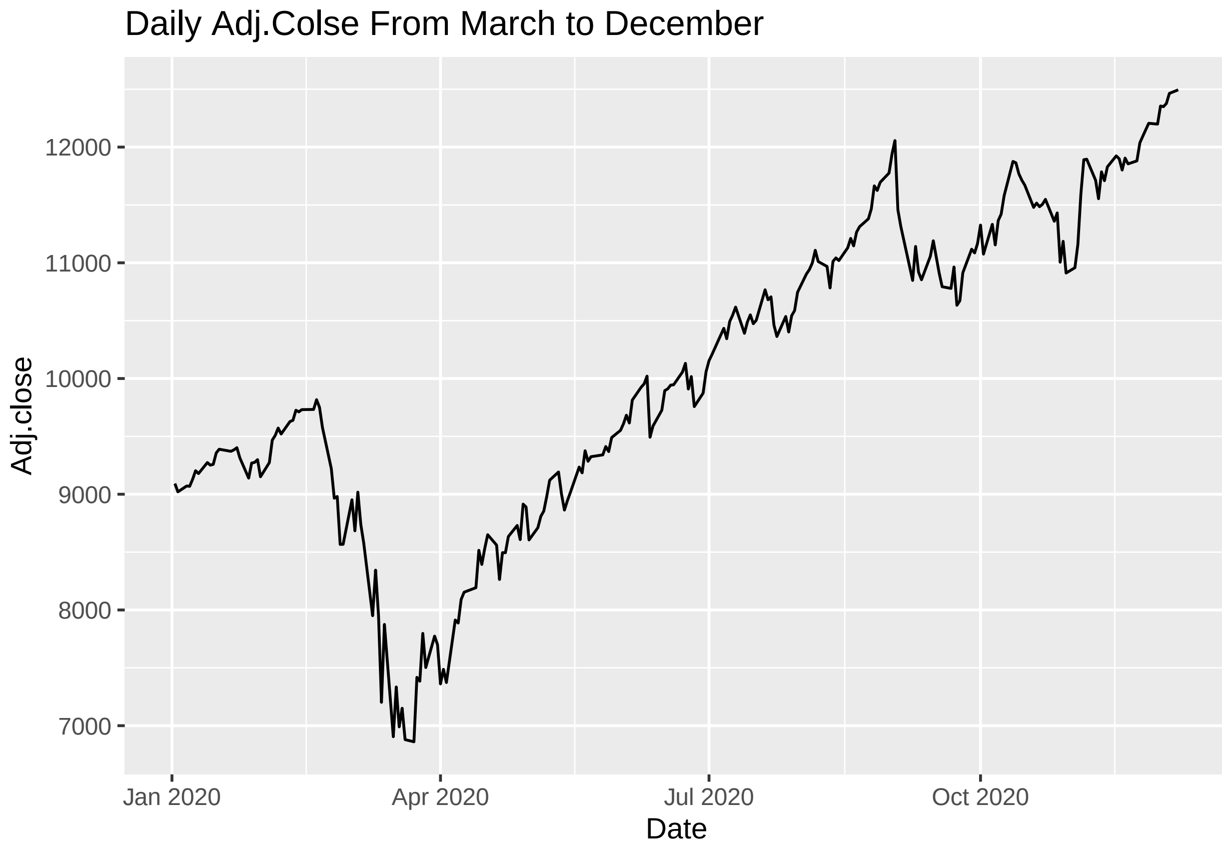


Correlation between S&P 500 Daily Returns and Daily Infection Rate Percentage Change.

```
## [1] -0.5811911
```

NASDAQ Composite

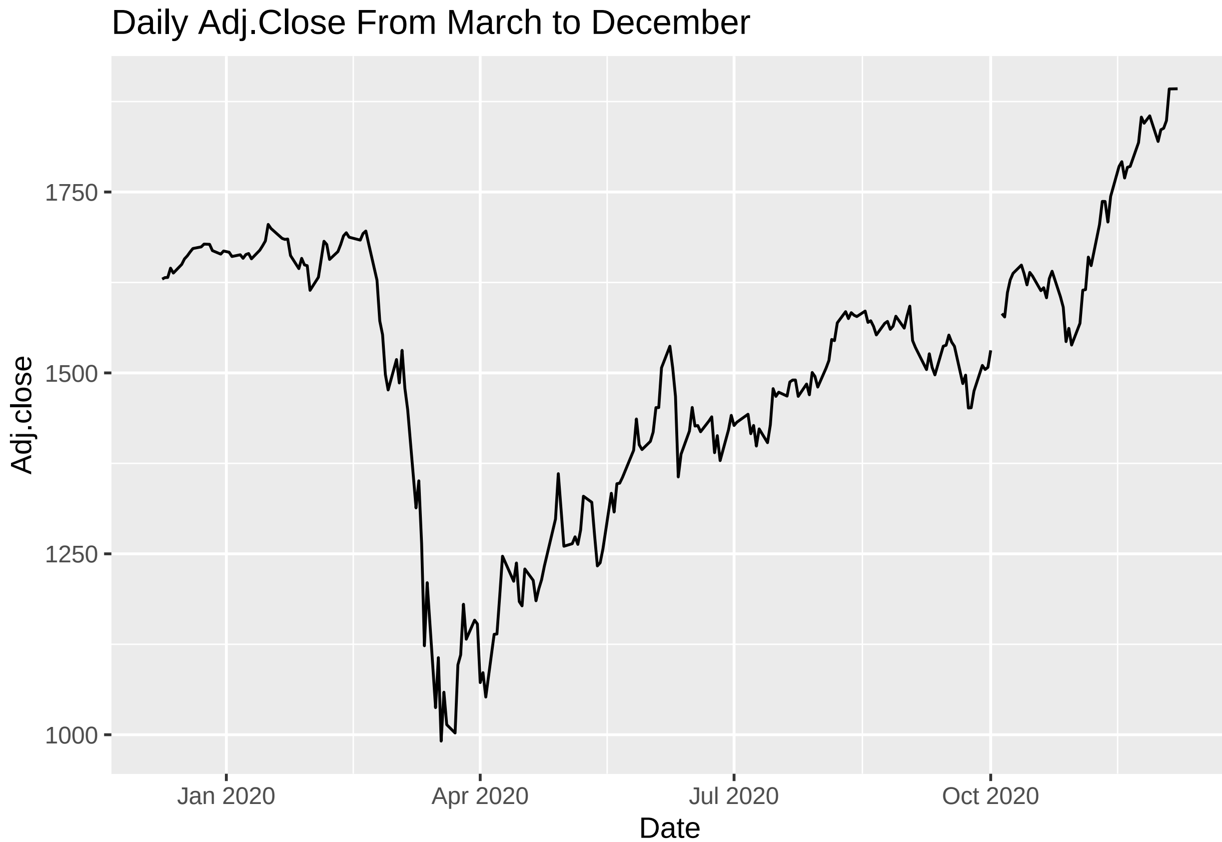
Similarly, we will do the same analysis on NASDAQ data and we get a pretty similar trend. It can be seen that there is a dip around April here as well. The overall correlation value is **-0.5325548** which is almost the same as the S&P 500.



Correlation between NASDAQ Daily Returns and Daily Infection Rate Percentage Change.

```
## [1] -0.5325548
```

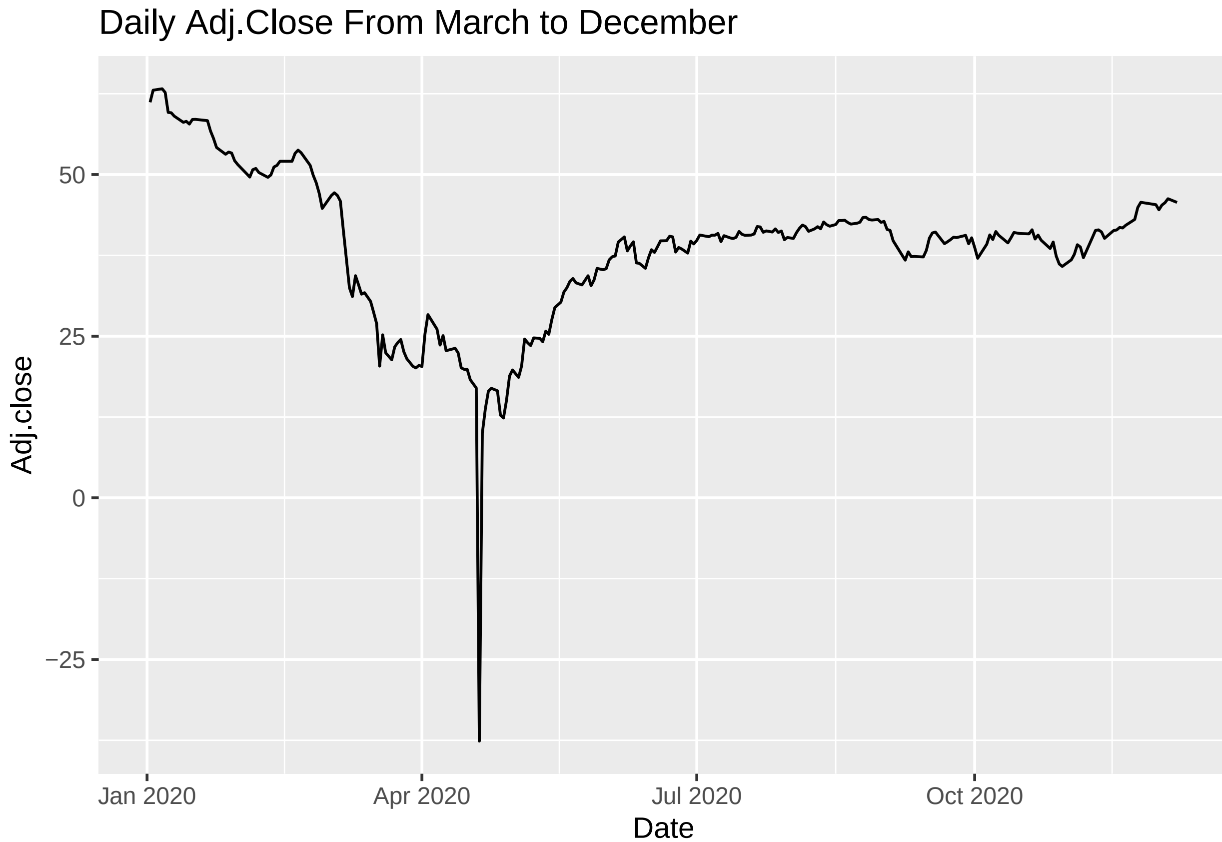
Russell 2000



Correlation between Russel Daily Returns and Daily Infection Rate Percentage Change.

```
## [1] -0.5418861
```

Crude Oil Futures



Correlation between Crude Oil Futures Daily Returns and Daily Infection Rate Percentage Change.

```
## [1] -0.2134725
```

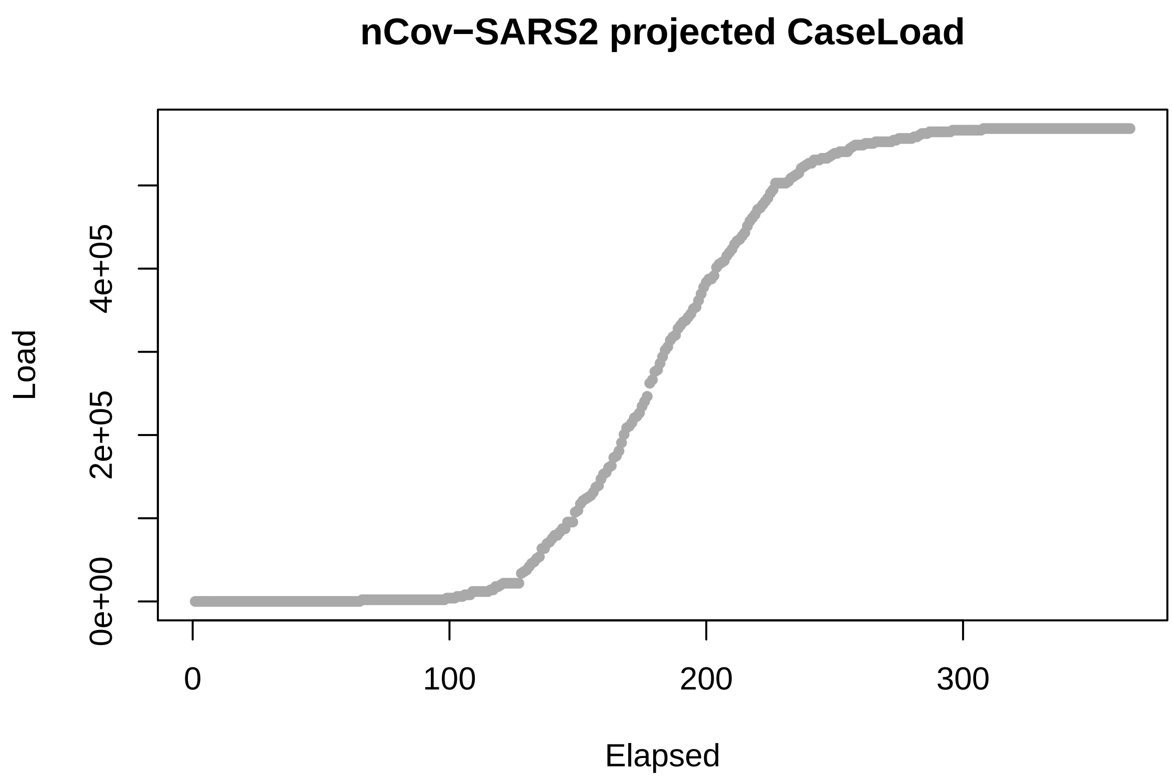
Overall, we can observe that all of the US Stock indices monthly returns are negatively correlated with the monthly percentage change in infection rates for the period March 2020 to December 2020.

Question4:

The client's simulation model tries to project the COVID load by sampling from a set of Poisson distributions.

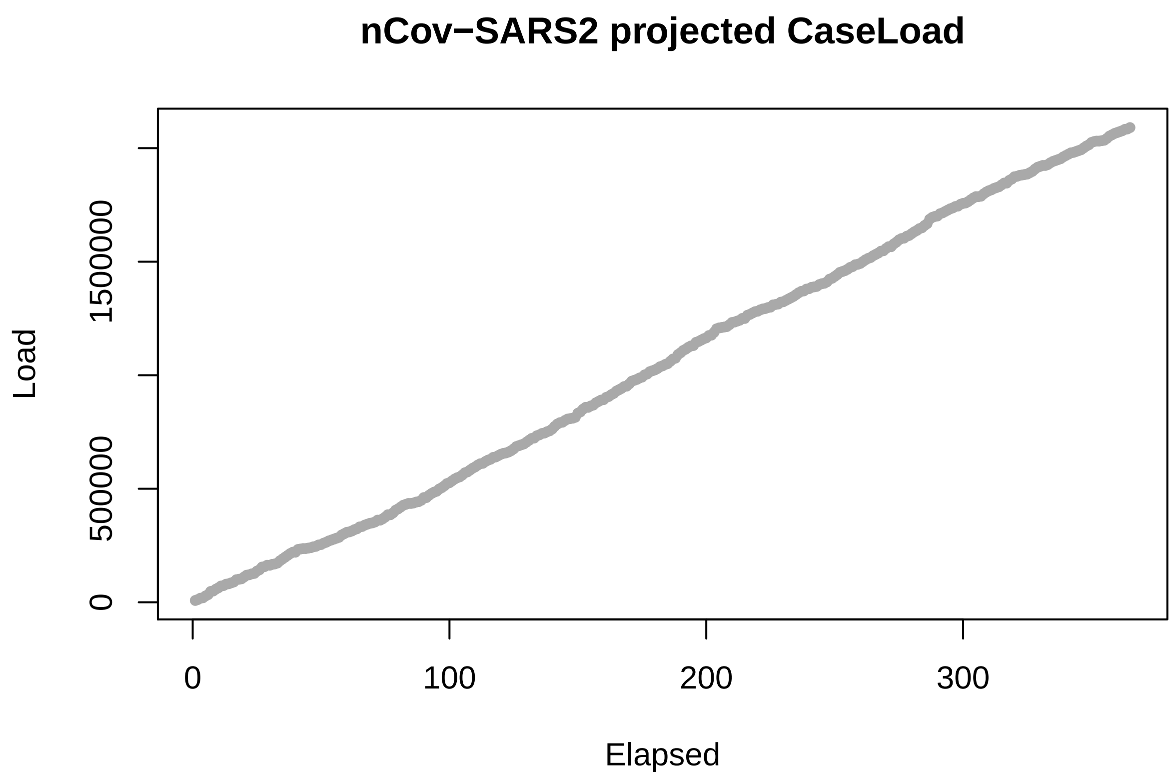
- In the first step, a set of means is computed using an exponential function which will further be used as "lambdas" in the sampling process from Poisson distributions.
- The means are calculated in such a way that, for the early days of the projection the means of the Poisson distributions are close to 0 indicating the slow start of the COVID spread. as the number of days reaches the middle, the means are close to 1, indicating the highest increase in COVID spread. Towards the end, means are close to 0 again so that the curve flattens.
- After the means are calculated, daily COVID cases are simulated by sampling from a series of Poisson distributions using the lambda's calculated above.
- The sampled values are multiplied by 1987.32 which looks like the average number of daily cases from the historical data.

```
num_days = 365
days <- 1:num_days
lambda_sim <- exp(-0.92*((days-182)^2/51.6^2))


W <- 1987.32*rpois(num_days,pi*lambda_sim)

plot(cumsum(W), xlab = "Elapsed", ylab = "Load",
main = 'nCov-SARS2 projected CaseLoad', pch = 16, cex = 0.75, col = "darkgrey")
```

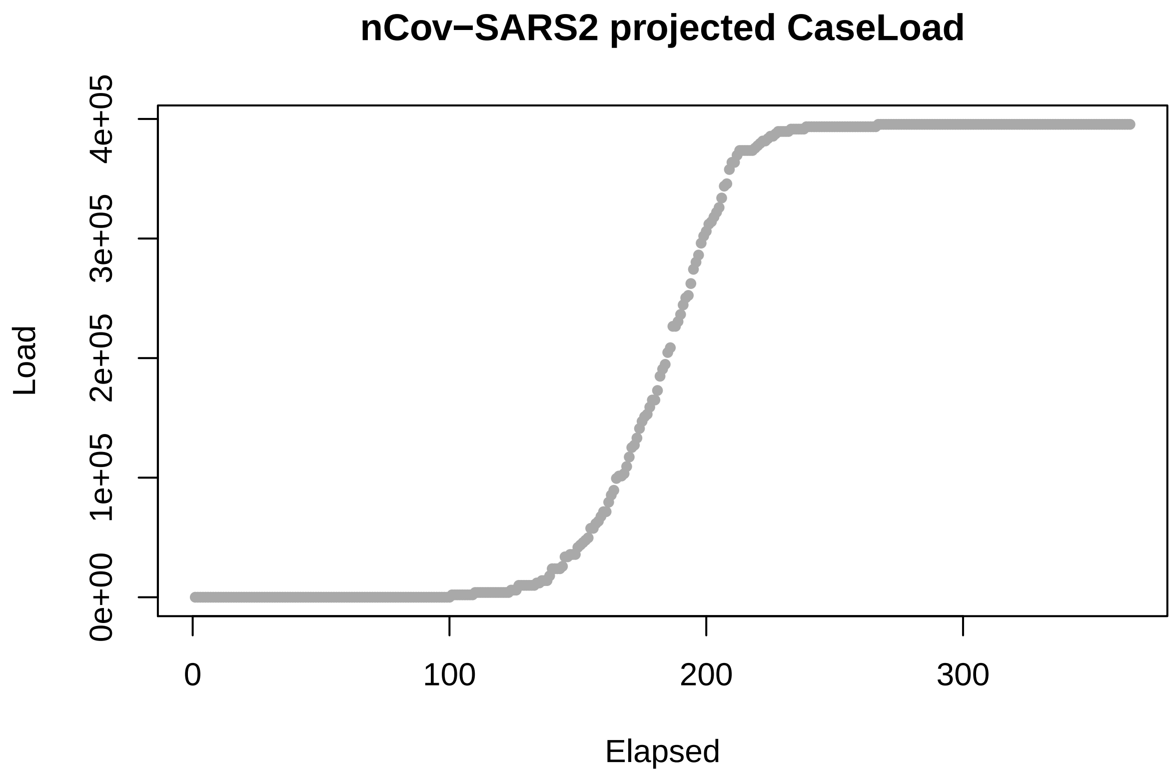


Sensitivity Analysis

Slightly changing the "-0.32" coefficient of the exponential function, changes the trend of the projection to a great extent. Bringing that value close to 0 results in a linear trend in the projection. This coefficient determines the means of the Poisson distributions used to sample daily cases. Therefore, it decides the projected rate of the spread of the disease from one day to the next. Although this model captures the exponential growth of the virus, it incorrectly projects that the spread of the virus levels off, and the number of daily cases comes close to 0 towards the end of the 365 days which is not observed from the historical data we analyzed.



Changing that coefficient to a value further away from 0 brings the trend of the projection to a steep exponential curve.


nCov−SARS2 projected CaseLoad
Load
0e+00
1e+05
2e+05
3e+05
4e+05
0
100
200
300
Elapsed

Appendix 1 : Methodology

Question2: US sate-wise COVID data

```
USCovid <- read.csv("all-states-history.csv")
UScovid_df <- USCovid[,c("date","state","death","hospitalized","positive","totalTestResults")]
UScovid_df$date <- as.Date(UScovid_df$date, format="%d-%m-%Y")
UScovid_df <- filter(UScovid_df, date >= as.Date("2020-03-01"))
head(UScovid_df)
```

```
##         date state death hospitalized positive totalTestResults
## 1 2020-12-03    AK   130          783    33291          1040505
## 2 2020-12-03    AL  3776        26062   260359          1615948
## 3 2020-12-03    AR  2555         9206   164310          1719401
## 4 2020-12-03    AS     0           NA        0             1988
## 5 2020-12-03    AZ  6821        27456   346421          2305084
## 6 2020-12-03    CA 19437           NA  1264539         24474642
```

Population Data

```
perfcarsales<- read.csv("PerfCarSales.csv")
pop<- perfcarsales[,c("ï..State","Abb","Population")]
pop <- pop %>% dplyr::rename(state = Abb)
head(pop)
```

```
##     ï..State state Population
## 1    Alabama    AL    4780558
## 2     Alaska    AK     710350
## 3    Arizona    AZ    6392123
## 4   Arkansas    AR    2916808
## 5 California    CA   37254238
## 6   Colorado    CO    5030018
```

```
PopDensity <- read.csv("PopulationDensity.csv")
PopDensity <- PopDensity %>% dplyr::rename(ï..State = GEO.display.label,  PopDensityPerMiles = Density.p
PopDensity<- PopDensity[, c("ï..State", "PopDensityPerMiles")]

head(PopDensity)
```

```
##     ï..State PopDensityPerMiles
## 1    Alabama               94.4
## 2     Alaska                1.2
## 3    Arizona               56.3
## 4   Arkansas               56.0
## 5 California              239.1
## 6   Colorado               48.5
```

```
population_df <- plyr::join(pop, PopDensity, by="ï..State")
head(population_df)
```

```
##     ï..State state Population PopDensityPerMiles
## 1    Alabama    AL    4780558               94.4
## 2     Alaska    AK     710350                1.2
## 3    Arizona    AZ    6392123               56.3
## 4   Arkansas    AR    2916808               56.0
## 5 California    CA   37254238              239.1
## 6   Colorado    CO    5030018               48.5
```

Join the population in perfcarsales to USCovid_df.

```r
UScovid_df <- plyr::join(UScovid_df, population_df, by = "state")
head(UScovid_df)
```

```
##         date state death hospitalized positive totalTestResults   ï..State
## 1 2020-12-03    AK   130          783    33291          1040505     Alaska
## 2 2020-12-03    AL  3776        26062   260359          1615948    Alabama
## 3 2020-12-03    AR  2555         9206   164310          1719401   Arkansas
## 4 2020-12-03    AS     0           NA        0             1988       <NA>
## 5 2020-12-03    AZ  6821        27456   346421          2305084    Arizona
## 6 2020-12-03    CA 19437           NA  1264539         24474642 California
##   Population PopDensityPerMiles
## 1     710350                1.2
## 2    4780558               94.4
## 3    2916808               56.0
## 4         NA                 NA
## 5    6392123               56.3
## 6   37254238              239.1
```

Removing AS from the dataset since there are no covid cases reported.

```r
UScovid_df <- filter(UScovid_df, !(state %in% c("AS", "MP", "GU", "PR", "VI")))


head(UScovid_df)
```

```
##         date state death hospitalized positive totalTestResults   ï..State
## 1 2020-12-03    AK   130          783    33291          1040505     Alaska
## 2 2020-12-03    AL  3776        26062   260359          1615948    Alabama
## 3 2020-12-03    AR  2555         9206   164310          1719401   Arkansas
## 4 2020-12-03    AZ  6821        27456   346421          2305084    Arizona
## 5 2020-12-03    CA 19437           NA  1264539         24474642 California
## 6 2020-12-03    CO  2716        14579   247209          3343095   Colorado
##   Population PopDensityPerMiles
## 1     710350                1.2
## 2    4780558               94.4
## 3    2916808               56.0
## 4    6392123               56.3
## 5   37254238              239.1
## 6    5030018               48.5
```

```r
UScovid_df <- UScovid_df[,-7]
head(UScovid_df)
```

```
##         date state death hospitalized positive totalTestResults Population
## 1 2020-12-03    AK   130          783    33291          1040505     710350
## 2 2020-12-03    AL  3776        26062   260359          1615948    4780558
## 3 2020-12-03    AR  2555         9206   164310          1719401    2916808
## 4 2020-12-03    AZ  6821        27456   346421          2305084    6392123
## 5 2020-12-03    CA 19437           NA  1264539         24474642   37254238
## 6 2020-12-03    CO  2716        14579   247209          3343095    5030018
##   PopDensityPerMiles
## 1                1.2
## 2               94.4
## 3               56.0
## 4               56.3
```

```
## 5              239.1
## 6               48.5
```

Compute infection rate, postive test rate and death rate.

```
UScovid_df$InfectionRate <- (UScovid_df$positive/UScovid_df$Population)*100

UScovid_df$PosTestRate <- (UScovid_df$positive/ UScovid_df$totalTestResults)*100

UScovid_df$DeathRate <- (UScovid_df$death/ UScovid_df$positive)*100
```

Infection Rate vs Population

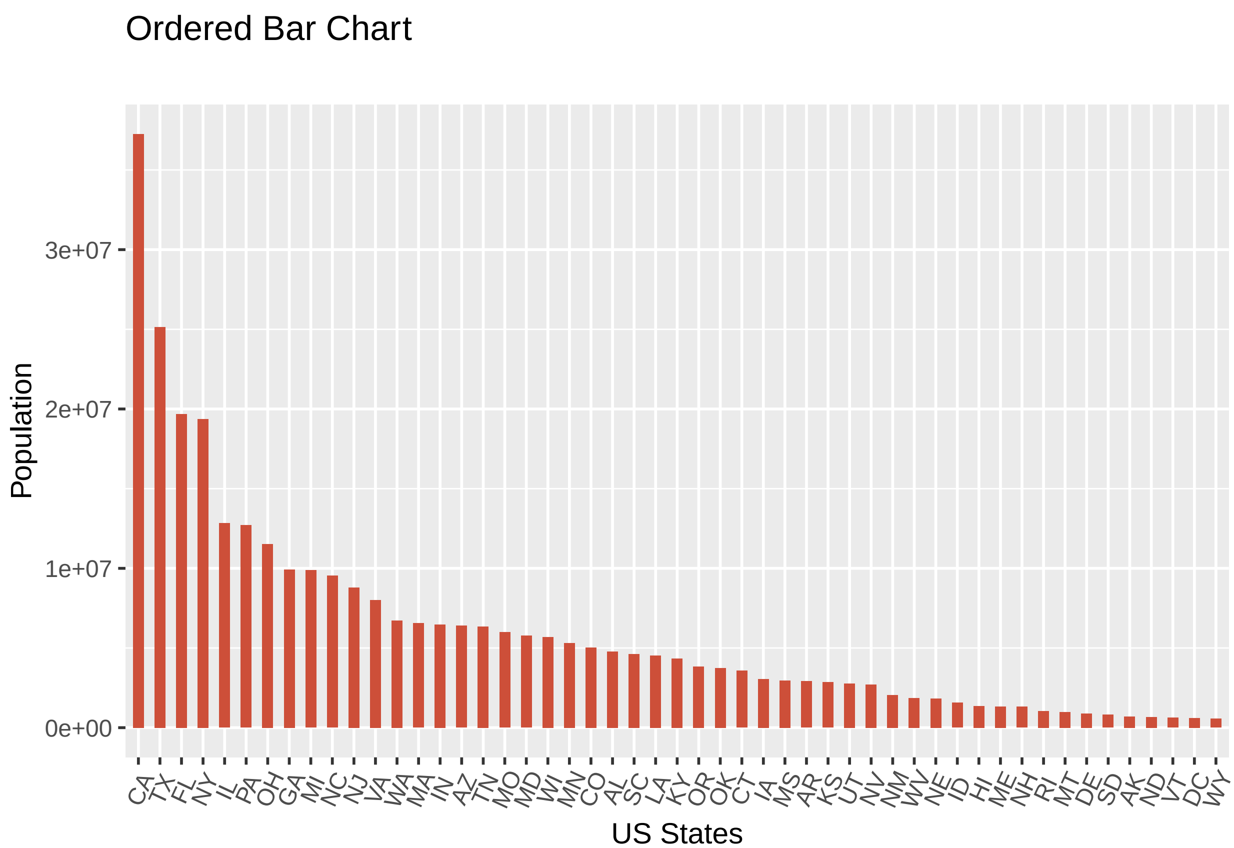
```
infection_rate_df <- UScovid_df[,c("date", "state", "InfectionRate")]
infection_rate_df <- infection_rate_df[order(as.numeric(rownames(infection_rate_df)), decreasing = TRUE
infection_rate_df[is.na(infection_rate_df)] <- 0
head(infection_rate_df)
```

```
##             date state InfectionRate
## 13999 2020-03-01    WY  0.0000000000
## 13998 2020-03-01    WA  0.0005055521
## 13997 2020-03-01    VA  0.0000000000
## 13996 2020-03-01    RI  0.0001899929
## 13995 2020-03-01    NJ  0.0000000000
## 13994 2020-03-01    NE  0.0000000000
```

```
sorted_population_df <- population_df[order(population_df$Population, decreasing=TRUE),]
sorted_population_df$state <- factor(sorted_population_df$state, levels=sorted_population_df$state)
head(sorted_population_df)
```

```
##        ï..State state Population PopDensityPerMiles
## 5    California    CA   37254238              239.1
## 44        Texas    TX   25145933               96.3
## 10      Florida    FL   19688112              350.6
## 33     New York    NY   19378245              411.2
## 14     Illinois    IL   12831383              231.1
## 39 Pennsylvania    PA   12702690              283.9
```

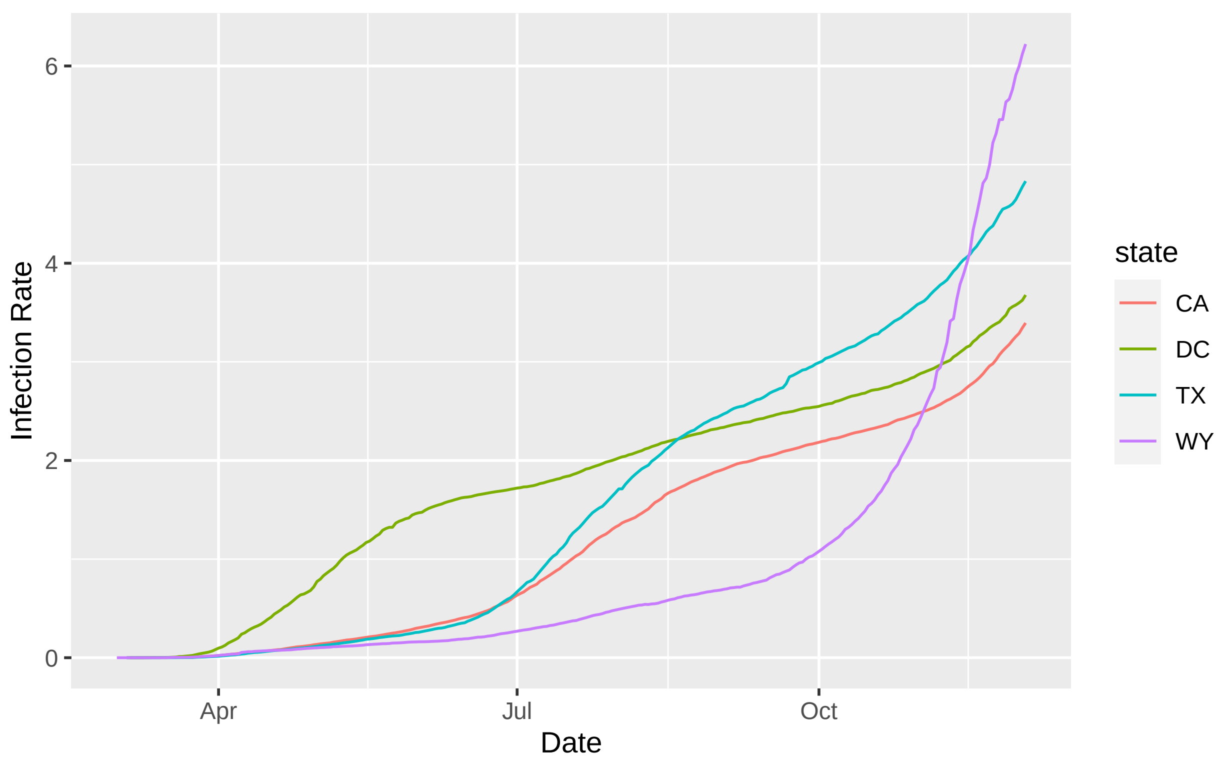
```
# Draw plot
ggplot(sorted_population_df, aes(x=state, y=Population)) +
  geom_bar(stat="identity", width=.5, fill="tomato3") +
  labs(
    title="Ordered Bar Chart",
    subtitle="",
    x="US States",
    y="Population") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))
```



Infection rate vs Population Size

```
infection_rate_df_filtered <- filter(infection_rate_df, state %in% c("CA", "TX", "DC", "WY"))
ggplot(data=infection_rate_df_filtered, aes(x=date, y=InfectionRate, colour=state)) +
  labs(
    title="Infection Rate Vs Date",
    subtitle="",
    x="Date",
    y="Infection Rate")+
  geom_line()
```

Infection Rate Vs Date

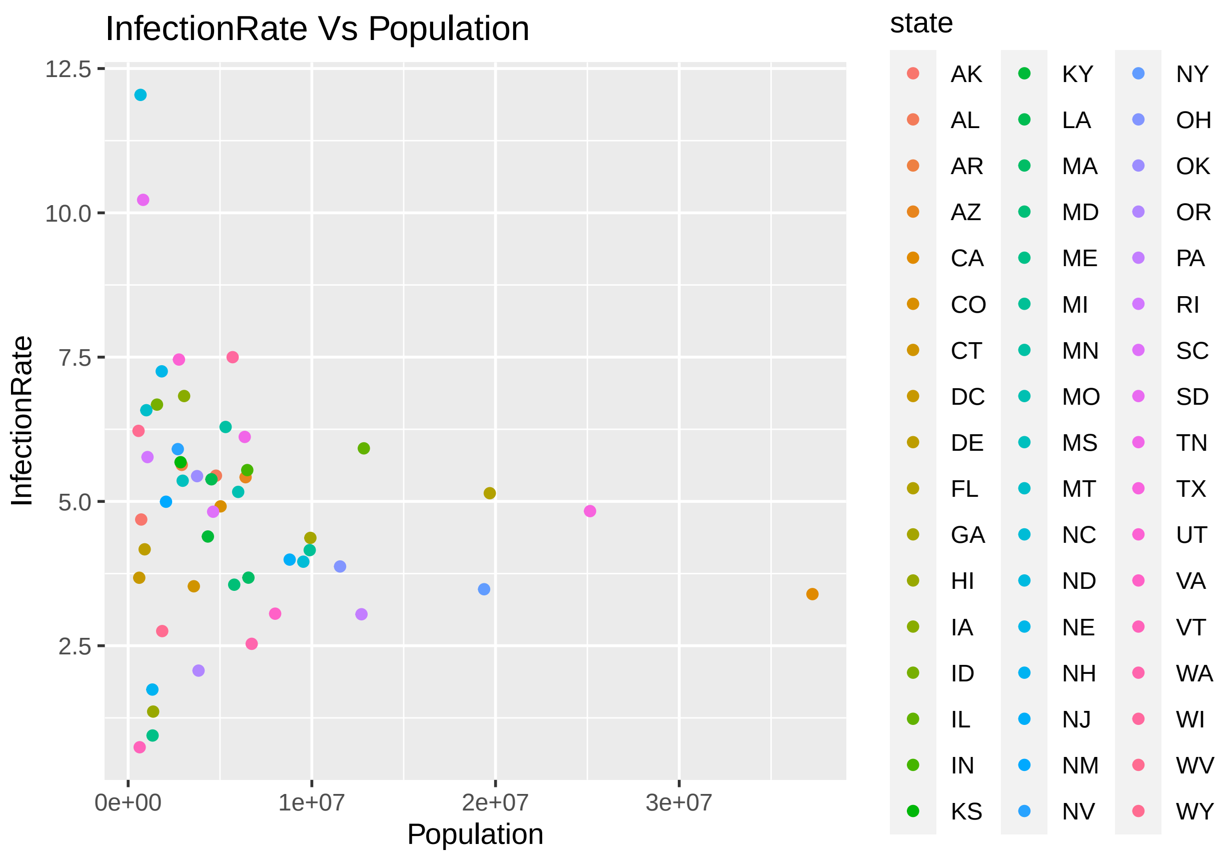


Cumulative Infection Rate Vs Population Size

```r
infection_rate_with_pop <- UScovid_df[,c("date", "state", "InfectionRate", "Population")]
cumulative_infection_rate <- filter(infection_rate_with_pop, date == as.Date("2020-12-03"))
head(cumulative_infection_rate)
```

```
##         date state InfectionRate Population
## 1 2020-12-03    AK      4.686563     710350
## 2 2020-12-03    AL      5.446205    4780558
## 3 2020-12-03    AR      5.633213    2916808
## 4 2020-12-03    AZ      5.419498    6392123
## 5 2020-12-03    CA      3.394349   37254238
## 6 2020-12-03    CO      4.914674    5030018
```

```r
ggplot(cumulative_infection_rate)  +
  geom_point(aes(x=Population, y=InfectionRate, col= state)) +   # draw points
  labs(
    y="InfectionRate",
    x="Population",
    title="InfectionRate Vs Population" )
```

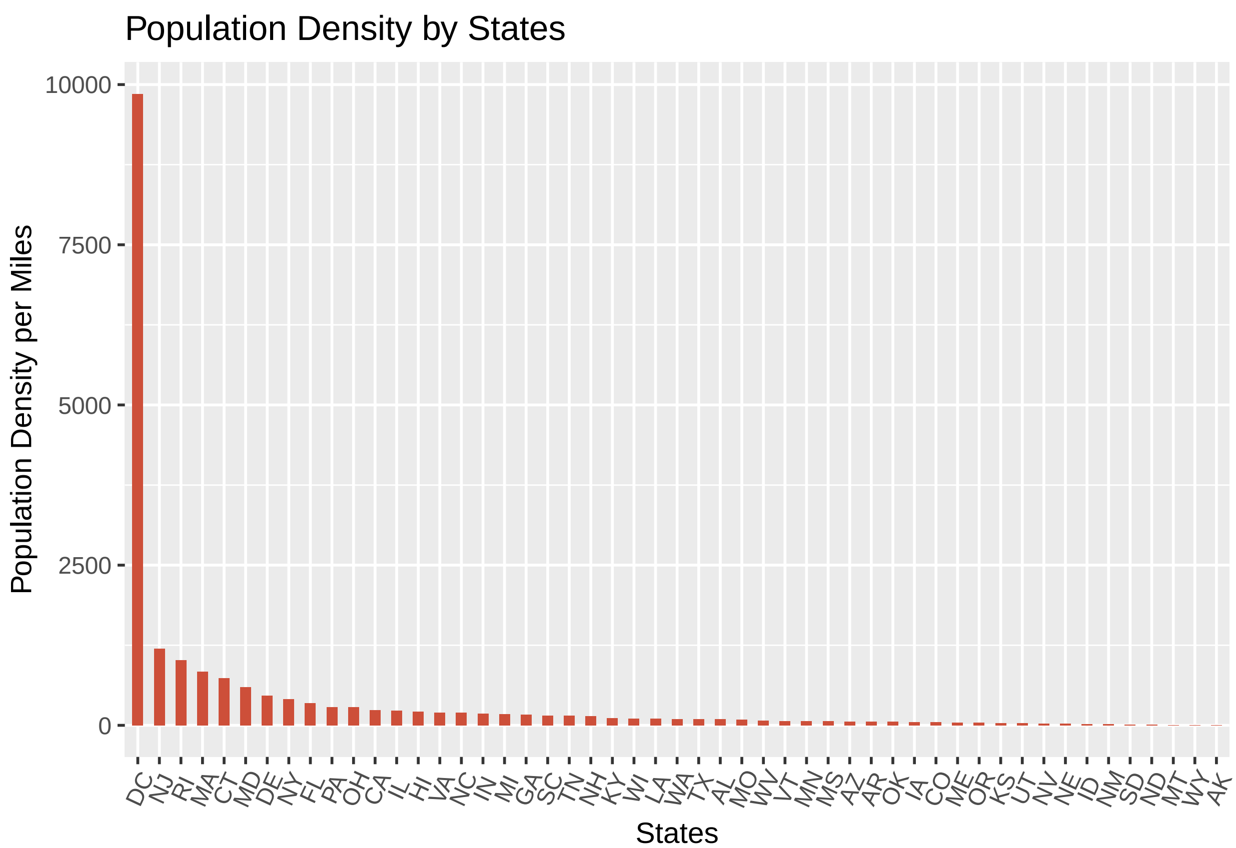


Infection Rate vs Population Density

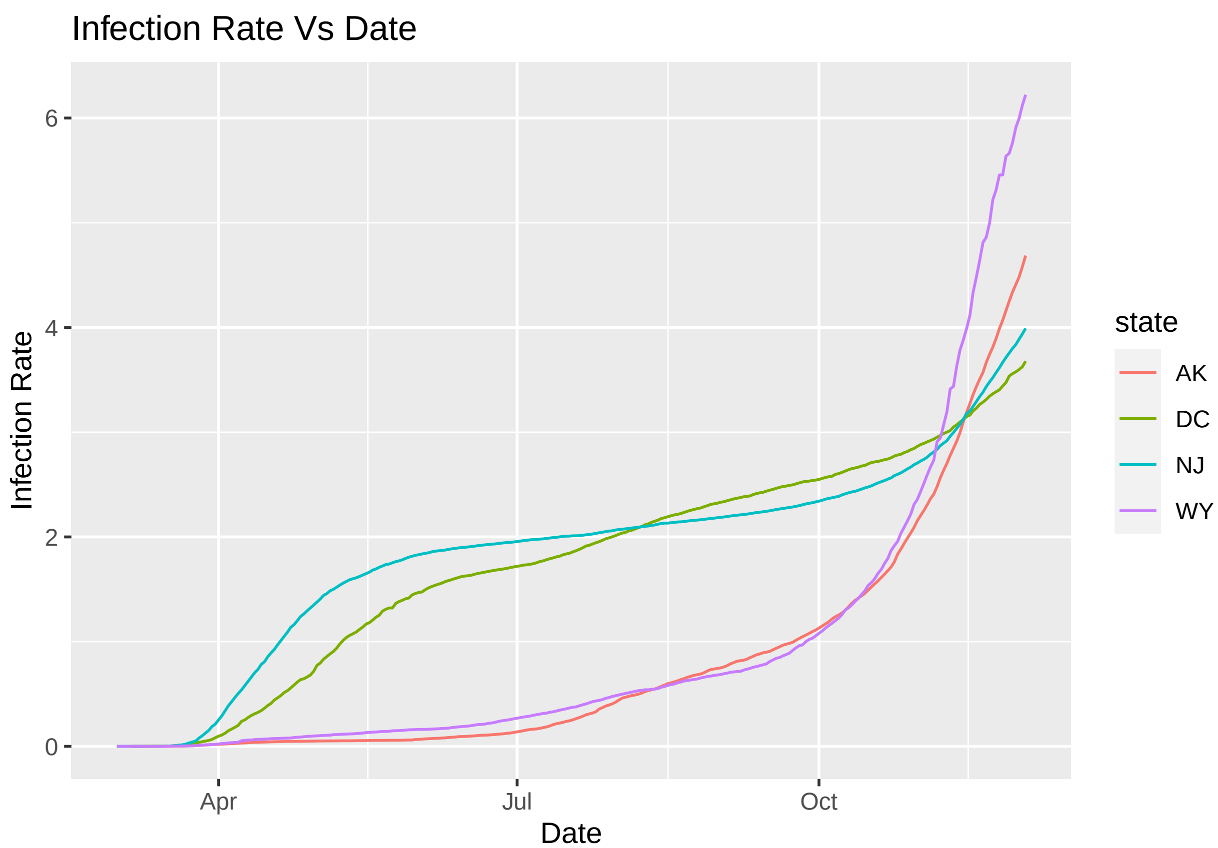
```
sorted_pop_density_df <- population_df[order(population_df$PopDensityPerMiles, decreasing = TRUE),]
sorted_pop_density_df$state <- factor(sorted_pop_density_df$state, levels=sorted_pop_density_df$state)
head(sorted_pop_density_df)
```

```
##                 ï..State state Population PopDensityPerMiles
## 9  District of Columbia    DC     602628             9856.5
## 31           New Jersey    NJ    8792421             1195.5
## 40         Rhode Island    RI    1052671             1018.1
## 22        Massachusetts    MA    6547704              839.4
## 7           Connecticut    CT    3574453              738.1
## 21             Maryland    MD    5773853              594.8
```

```
ggplot(sorted_pop_density_df, aes(x=state, y=PopDensityPerMiles)) +
  geom_bar(stat="identity", width=.5, fill="tomato3") +
  labs(
    title="Population Density by States",
    x="States",
    y="Population Density per Miles") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))
```




```
infection_rate_df_filtered <- filter(infection_rate_df, state %in% c("DC", "NJ", "WY", "AK"))
ggplot(data=infection_rate_df_filtered, aes(x=date, y=InfectionRate, colour=state)) +
   labs(
    title="Infection Rate Vs Date",
    x="Date",
    y="Infection Rate")+
  geom_line()
```

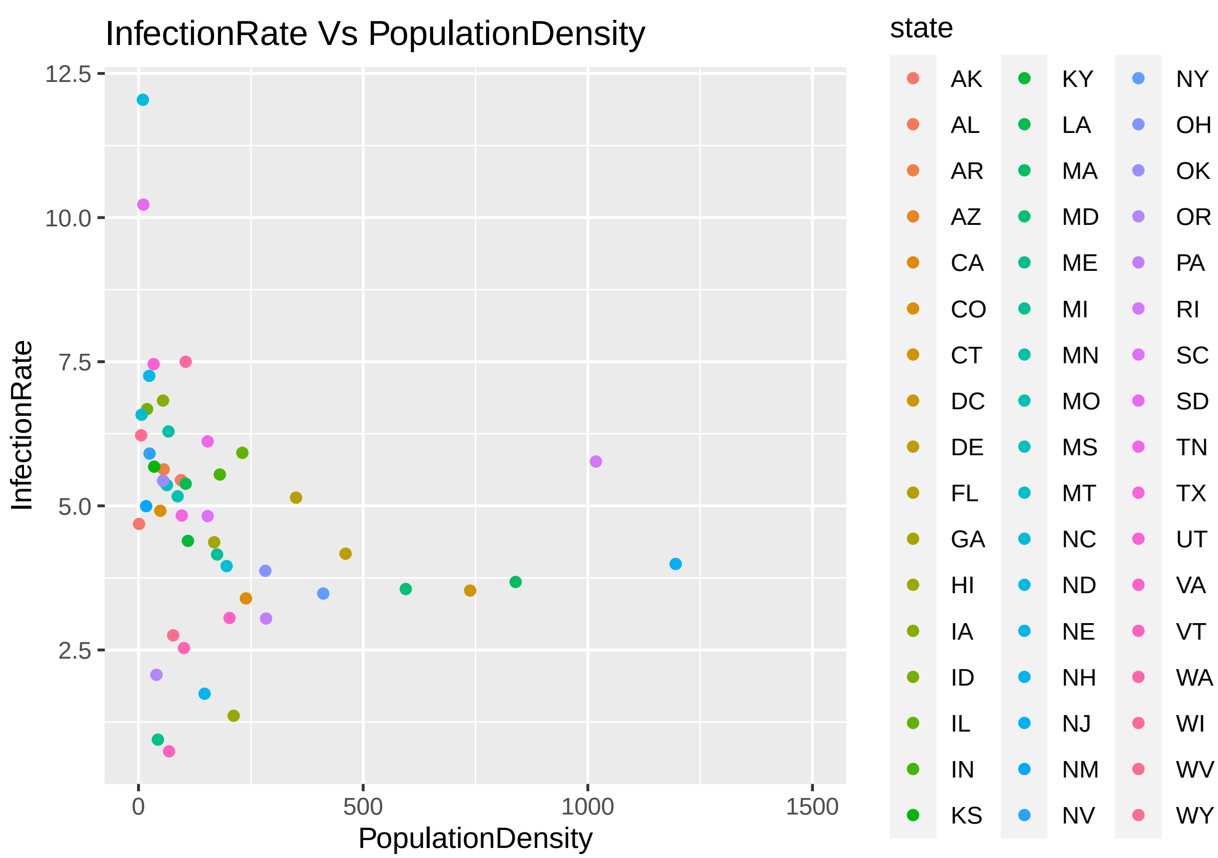


Cumulative Infection Rate vs Population Density

```
infection_rate_with_pop_density <- UScovid_df[,c("date", "state", "InfectionRate", "PopDensityPerMiles")
cumulative_infection_rate <- filter(infection_rate_with_pop_density, date == as.Date("2020-12-03"))
head(cumulative_infection_rate)
```

```
##         date state InfectionRate PopDensityPerMiles
## 1 2020-12-03    AK      4.686563                1.2
## 2 2020-12-03    AL      5.446205               94.4
## 3 2020-12-03    AR      5.633213               56.0
## 4 2020-12-03    AZ      5.419498               56.3
## 5 2020-12-03    CA      3.394349              239.1
## 6 2020-12-03    CO      4.914674               48.5
```

```
# Plot
ggplot(cumulative_infection_rate)  +
  geom_point(aes(x=PopDensityPerMiles, y=InfectionRate, col=state)) +   # draw points
  xlim(0, 1500) +
  labs(
    title="InfectionRate Vs PopulationDensity",
    y="InfectionRate",
    x="PopulationDensity"
    )
```



Infection rate change vs Population

```
head(infection_rate_df)
```

```
##             date state InfectionRate
## 13999 2020-03-01    WY  0.0000000000
## 13998 2020-03-01    WA  0.0005055521
## 13997 2020-03-01    VA  0.0000000000
## 13996 2020-03-01    RI  0.0001899929
## 13995 2020-03-01    NJ  0.0000000000
## 13994 2020-03-01    NE  0.0000000000
```

```
infection_rate_df_wide <- reshape(infection_rate_df, timevar="state", idvar="date", direction="wide")
infection_rate_df_wide[is.na(infection_rate_df_wide)] <- 0
head(infection_rate_df_wide)
```

```
##             date InfectionRate.WY InfectionRate.WA InfectionRate.VA
## 13999 2020-03-01                0     0.0005055521                0
## 13989 2020-03-02                0     0.0009664966                0
## 13978 2020-03-03                0     0.0013382261                0
## 13965 2020-03-04                0     0.0019032549                0
## 13941 2020-03-05                0     0.0026615830                0
## 13911 2020-03-06                0     0.0032414810                0
##       InfectionRate.RI InfectionRate.NJ InfectionRate.NE InfectionRate.MI
## 13999     0.0001899929     0.000000e+00                0     0.0001416419
## 13989     0.0001899929     0.000000e+00                0     0.0002731665
## 13978     0.0001899929     0.000000e+00                0     0.0005058639
## 13965     0.0001899929     0.000000e+00                0     0.0007486786
```

```
## 13941     0.0002849893     1.137343e-05                0     0.0010117278
## 13911     0.0002849893     1.137343e-05                0     0.0014164190
##        InfectionRate.MA InfectionRate.IN InfectionRate.FL InfectionRate.NY
## 13999                 0     0.000000e+00     0.000000e+00     0.000000e+00
## 13989                 0     0.000000e+00     0.000000e+00     0.000000e+00
## 13978                 0     0.000000e+00     1.015841e-05     5.160426e-06
## 13965                 0     0.000000e+00     1.523762e-05     5.160426e-06
## 13941                 0     0.000000e+00     3.555445e-05     1.548128e-05
## 13911                 0     1.542198e-05     4.063366e-05     1.290107e-04
##        InfectionRate.WI InfectionRate.VT InfectionRate.TX InfectionRate.SC
## 13999     0.000000e+00                0     0.000000e+00                0
## 13989     0.000000e+00                0     0.000000e+00                0
## 13978     0.000000e+00                0     0.000000e+00                0
## 13965     1.758169e-05                0     3.976786e-06                0
## 13941     1.758169e-05                0     3.976786e-06                0
## 13911     1.758169e-05                0     1.988393e-05                0
##        InfectionRate.OR InfectionRate.NH InfectionRate.NC InfectionRate.IL
## 13999     0.000000e+00     0.0000000000     0.000000e+00     0.000000e+00
## 13989     0.000000e+00     0.0000000000     0.000000e+00     0.000000e+00
## 13978     0.000000e+00     0.0000000000     0.000000e+00     0.000000e+00
## 13965     7.829603e-05     0.0001518509     1.048707e-05     3.117357e-05
## 13941     7.829603e-05     0.0001518509     1.048707e-05     3.896696e-05
## 13911     7.829603e-05     0.0001518509     2.097413e-05     3.896696e-05
##        InfectionRate.HI InfectionRate.GA InfectionRate.CO InfectionRate.CA
## 13999                 0     0.000000e+00     0.000000e+00     0.0000000000
## 13989                 0     0.000000e+00     0.000000e+00     0.0000000000
## 13978                 0     0.000000e+00     0.000000e+00     0.0000000000
## 13965                 0     2.015967e-05     3.976129e-05     0.0001422657
## 13941                 0     2.015967e-05     3.976129e-05     0.0001422657
## 13911                 0     2.015967e-05     1.590452e-04     0.0001610555
##        InfectionRate.AZ InfectionRate.TN InfectionRate.OH InfectionRate.NV
## 13999     0.000000e+00     0.000000e+00                0     0.000000e+00
## 13989     0.000000e+00     0.000000e+00                0     0.000000e+00
## 13978     0.000000e+00     0.000000e+00                0     0.000000e+00
## 13965     3.128851e-05     0.000000e+00                0     0.000000e+00
## 13941     3.128851e-05     1.575703e-05                0     3.702146e-05
## 13911     4.693276e-05     1.575703e-05                0     1.110644e-04
##        InfectionRate.NM InfectionRate.MD InfectionRate.DC InfectionRate.WV
## 13999                 0     0.000000e+00                0                0
## 13989                 0     0.000000e+00                0                0
## 13978                 0     0.000000e+00                0                0
## 13965                 0     0.000000e+00                0                0
## 13941                 0     0.000000e+00                0                0
## 13911                 0     5.195837e-05                0                0
##        InfectionRate.PA InfectionRate.MN InfectionRate.KY InfectionRate.KS
## 13999      0.00000e+00     0.000000e+00                0                0
## 13989      0.00000e+00     0.000000e+00                0                0
## 13978      0.00000e+00     0.000000e+00                0                0
## 13965      0.00000e+00     0.000000e+00                0                0
## 13941      0.00000e+00     0.000000e+00                0                0
## 13911      1.57447e-05     1.885378e-05                0                0
##        InfectionRate.IA InfectionRate.DE InfectionRate.AR InfectionRate.AK
## 13999                 0                0                0                0
## 13989                 0                0                0                0
```

```
## 13978                 0                 0                 0                 0
## 13965                 0                 0                 0                 0
## 13941                 0                 0                 0                 0
## 13911                 0                 0                 0                 0
##       InfectionRate.UT InfectionRate.SD InfectionRate.OK InfectionRate.ND
## 13999                 0                 0                 0                 0
## 13989                 0                 0                 0                 0
## 13978                 0                 0                 0                 0
## 13965                 0                 0                 0                 0
## 13941                 0                 0                 0                 0
## 13911                 0                 0                 0                 0
##       InfectionRate.MT InfectionRate.MS InfectionRate.MO InfectionRate.ME
## 13999                 0                 0                 0                 0
## 13989                 0                 0                 0                 0
## 13978                 0                 0                 0                 0
## 13965                 0                 0                 0                 0
## 13941                 0                 0                 0                 0
## 13911                 0                 0                 0                 0
##       InfectionRate.LA InfectionRate.ID InfectionRate.CT InfectionRate.AL
## 13999                 0                 0                 0                 0
## 13989                 0                 0                 0                 0
## 13978                 0                 0                 0                 0
## 13965                 0                 0                 0                 0
## 13941                 0                 0                 0                 0
## 13911                 0                 0                 0                 0
```

```r
infection_rate_weekly_change <- data.frame(diff(as.matrix(infection_rate_df_wide[, -1]), lag=7))
head(infection_rate_weekly_change)
```

```
##       InfectionRate.WY InfectionRate.WA InfectionRate.VA InfectionRate.RI
## 13821     0.0000000000      0.004207978     2.499370e-05     9.499644e-05
## 13770     0.0000000000      0.004654053     3.749056e-05     1.899929e-04
## 13719     0.0000000000      0.005917933     9.997482e-05     2.849893e-04
## 13668     0.0000000000      0.007345374     1.124717e-04     2.849893e-04
## 13617     0.0001773855      0.009248629     2.124465e-04     9.499644e-04
## 13566     0.0001773855      0.011895343     3.749056e-04     1.519943e-03
##       InfectionRate.NJ InfectionRate.NE InfectionRate.MI InfectionRate.MA
## 13821     6.824059e-05     5.473493e-05      0.002397795     0.0000000000
## 13770     1.251077e-04     1.642048e-04      0.003490461     0.0000000000
## 13719     1.706015e-04     1.642048e-04      0.004836059     0.0000000000
## 13668     2.729624e-04     2.736746e-04      0.006606583     0.0000000000
## 13617     3.298295e-04     5.473493e-04      0.009075199     0.0001221802
## 13566     5.572982e-04     7.115541e-04      0.012353197     0.0003512682
##       InfectionRate.IN InfectionRate.FL InfectionRate.NY InfectionRate.WI
## 13821     3.084397e-05     6.095049e-05     0.0003096256     1.758169e-05
## 13770     3.084397e-05     6.602969e-05     0.0004541175     1.758169e-05
## 13719     9.253190e-05     8.126732e-05     0.0007740639     3.516338e-05
## 13668     1.542198e-04     8.634652e-05     0.0010011227     3.516338e-05
## 13617     1.850638e-04     9.650494e-05     0.0012797857     1.406535e-04
## 13566     1.696418e-04     1.168218e-04     0.0016926197     3.867971e-04
##       InfectionRate.VT InfectionRate.TX InfectionRate.SC InfectionRate.OR
## 13821     0.0001596531     3.181429e-05     4.323113e-05     0.0003653815
## 13770     0.0001596531     4.772143e-05     1.513090e-04     0.0003653815
## 13719     0.0001596531     5.169822e-05     1.513090e-04     0.0003914801
## 13668     0.0003193062     7.953572e-05     1.945401e-04     0.0004175788
```

```
## 13617      0.0003193062      8.748930e-05      2.161556e-04      0.0004175788
## 13566      0.0007982655      1.352107e-04      2.593868e-04      0.0007046643
##        InfectionRate.NH InfectionRate.NC InfectionRate.IL InfectionRate.HI
## 13821      0.0003037019      2.097413e-05      4.676035e-05      7.349839e-05
## 13770      0.0003037019      2.097413e-05      5.455375e-05      1.469968e-04
## 13719      0.0003037019      7.340947e-05      1.480745e-04      1.469968e-04
## 13668      0.0001518509      6.292240e-05      1.169009e-04      1.469968e-04
## 13617      0.0003037019      1.153577e-04      1.558678e-04      1.469968e-04
## 13566      0.0003037019      1.363319e-04      2.104216e-04      1.469968e-04
##        InfectionRate.GA InfectionRate.CO InfectionRate.CA InfectionRate.AZ
## 13821      7.055886e-05      0.0001789258      0.0002362147      7.822127e-05
## 13770      1.209580e-04      0.0002385677      0.0003060055      7.822127e-05
## 13719      1.713572e-04      0.0005566580      0.0003570064      9.386553e-05
## 13668      2.015967e-04      0.0008548677      0.0002791629      1.095098e-04
## 13617      2.923153e-04      0.0013916451      0.0003999545      1.095098e-04
## 13566      4.031935e-04      0.0018488999      0.0003811647      9.386553e-05
##        InfectionRate.TN InfectionRate.OH InfectionRate.NV InfectionRate.NM
## 13821      4.727109e-05      0.000000e+00      0.0001480858      0.0000000000
## 13770      4.727109e-05      2.600264e-05      0.0001851073      0.0000000000
## 13719      1.102992e-04      2.600264e-05      0.0001851073      0.0001456660
## 13668      1.102992e-04      3.467019e-05      0.0003331931      0.0002427767
## 13617      2.678695e-04      4.333774e-05      0.0004072361      0.0004855533
## 13566      3.939257e-04      1.126781e-04      0.0007034078      0.0004855533
##        InfectionRate.MD InfectionRate.DC InfectionRate.WV InfectionRate.PA
## 13821      5.195837e-05      0.0001659399                0      4.723409e-05
## 13770      8.659729e-05      0.0001659399                0      7.872348e-05
## 13719      1.039167e-04      0.0008296993                0      9.446818e-05
## 13668      1.558751e-04      0.0008296993                0      1.259576e-04
## 13617      2.078335e-04      0.0016593985                0      1.731917e-04
## 13566      2.424724e-04      0.0016593985                0      3.070216e-04
##        InfectionRate.MN InfectionRate.KY InfectionRate.KS InfectionRate.IA
## 13821      3.770755e-05      2.304433e-05      3.504614e-05      0.0000000000
## 13770      9.426889e-05      9.217730e-05      3.504614e-05      0.0000984691
## 13719      2.073915e-04      1.382660e-04      3.504614e-05      0.0002625843
## 13668      3.959293e-04      1.843546e-04      3.504614e-05      0.0004266994
## 13617      8.107124e-04      1.843546e-04      1.401846e-04      0.0004595225
## 13566      1.150080e-03      2.534876e-04      2.102769e-04      0.0005251685
##        InfectionRate.DE InfectionRate.AR InfectionRate.AK InfectionRate.UT
## 13821      0.0000000000      0.0000000000                0      3.617623e-05
## 13770      0.0000000000      0.0000000000                0      3.617623e-05
## 13719      0.0000000000      0.0000000000                0      7.235246e-05
## 13668      0.0001113354      0.0000000000                0      7.235246e-05
## 13617      0.0004453416      0.0002057043                0      1.447049e-04
## 13566      0.0004453416      0.0003085565                0      2.170574e-04
##        InfectionRate.SD InfectionRate.OK InfectionRate.ND InfectionRate.MT
## 13821      0.0000000000      2.665334e-05      0.0000000000      0.0000000000
## 13770      0.0000000000      2.665334e-05      0.0000000000      0.0000000000
## 13719      0.0000000000      5.330668e-05      0.0000000000      0.0000000000
## 13668      0.0006133946      5.330668e-05      0.0000000000      0.0000000000
## 13617      0.0009814313      7.996002e-05      0.0001484935      0.0001010458
## 13566      0.0011041102      7.996002e-05      0.0001484935      0.0001010458
##        InfectionRate.MS InfectionRate.MO InfectionRate.ME InfectionRate.LA
## 13821      0.0000000000      1.669543e-05      0.000000e+00      0.000000e+00
## 13770      0.0000000000      1.669543e-05      0.000000e+00      2.205564e-05
```

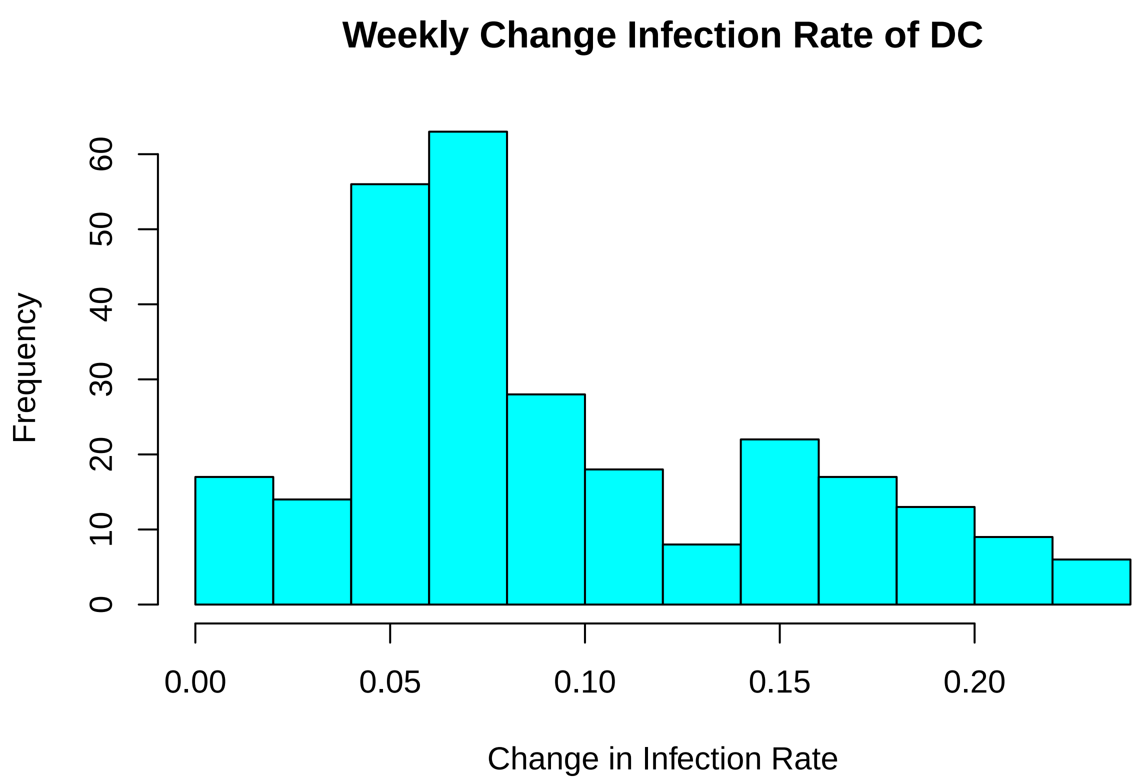
```
## 13719     0.0000000000     1.669543e-05     0.000000e+00     2.205564e-05
## 13668     0.0000000000     1.669543e-05     0.000000e+00     1.323338e-04
## 13617     0.0000336937     1.669543e-05     7.522994e-05     3.087789e-04
## 13566     0.0001347748     3.339086e-05     2.256898e-04     7.940030e-04
##       InfectionRate.ID InfectionRate.CT InfectionRate.AL
## 13821                0     2.797631e-05     0.000000e+00
## 13770                0     2.797631e-05     0.000000e+00
## 13719                0     5.595262e-05     0.000000e+00
## 13668                0     8.392893e-05     0.000000e+00
## 13617                0     1.678579e-04     0.000000e+00
## 13566                0     1.678579e-04     2.091806e-05
```

Lets analyse the difference in weekly change in infection rates between states with high population density and low population density.

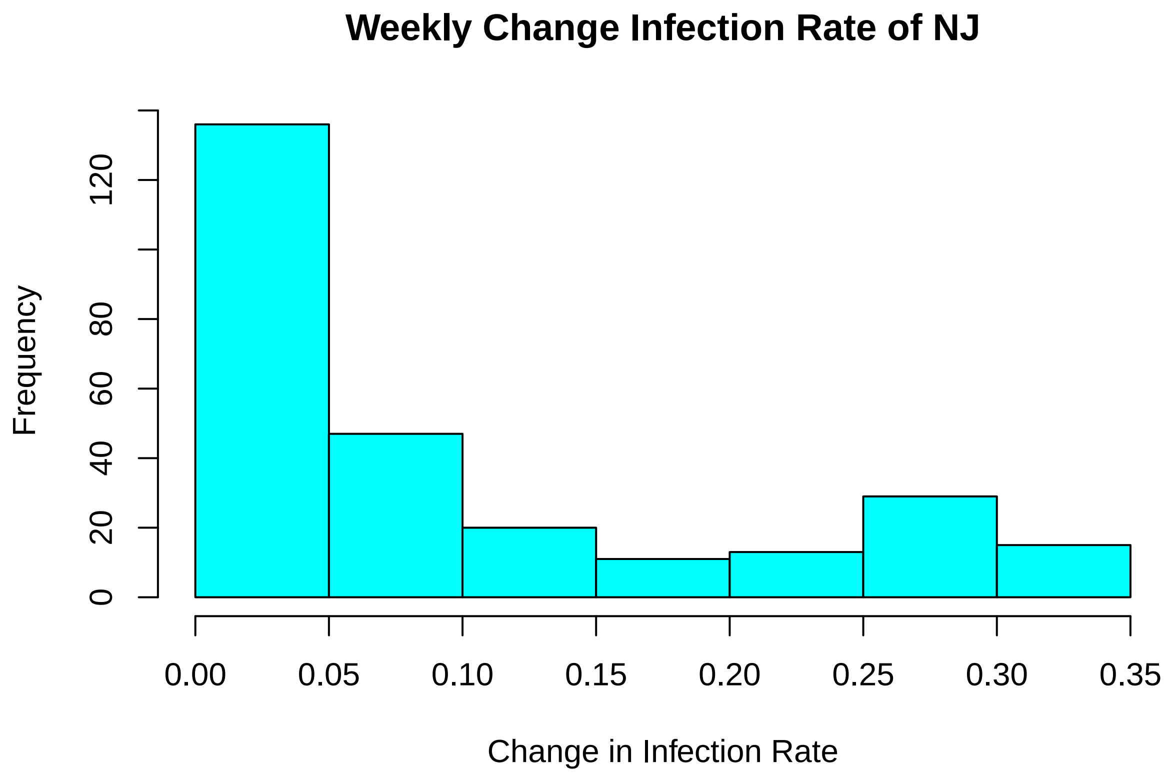
1. DC - Highest Popuation Density

We can see the change of infection rate week over week is very high. This indicates a faster spread where population density is higher.

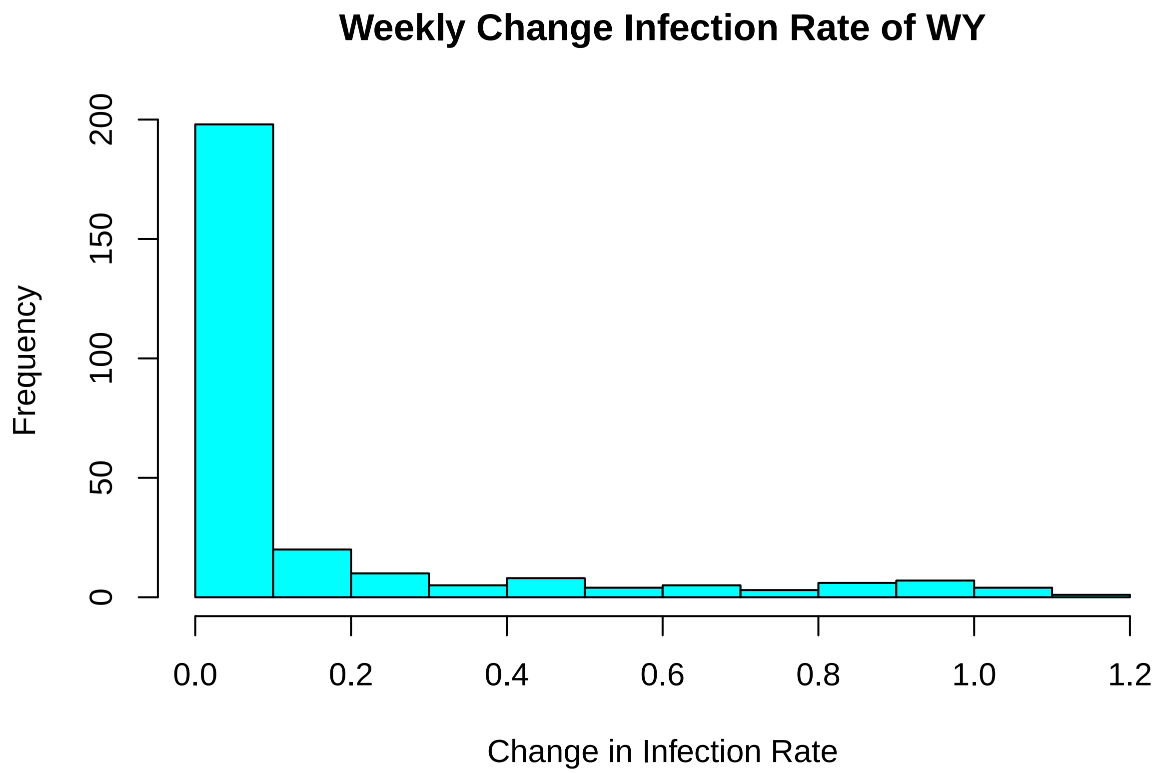
```r
hist(infection_rate_weekly_change$InfectionRate.DC, col= "cyan", xlab="Change in Infection Rate",
        main="Weekly Change Infection Rate of DC")
```




```r
hist(infection_rate_weekly_change$InfectionRate.NJ, col= "cyan", xlab="Change in Infection Rate",
        main="Weekly Change Infection Rate of NJ")
```

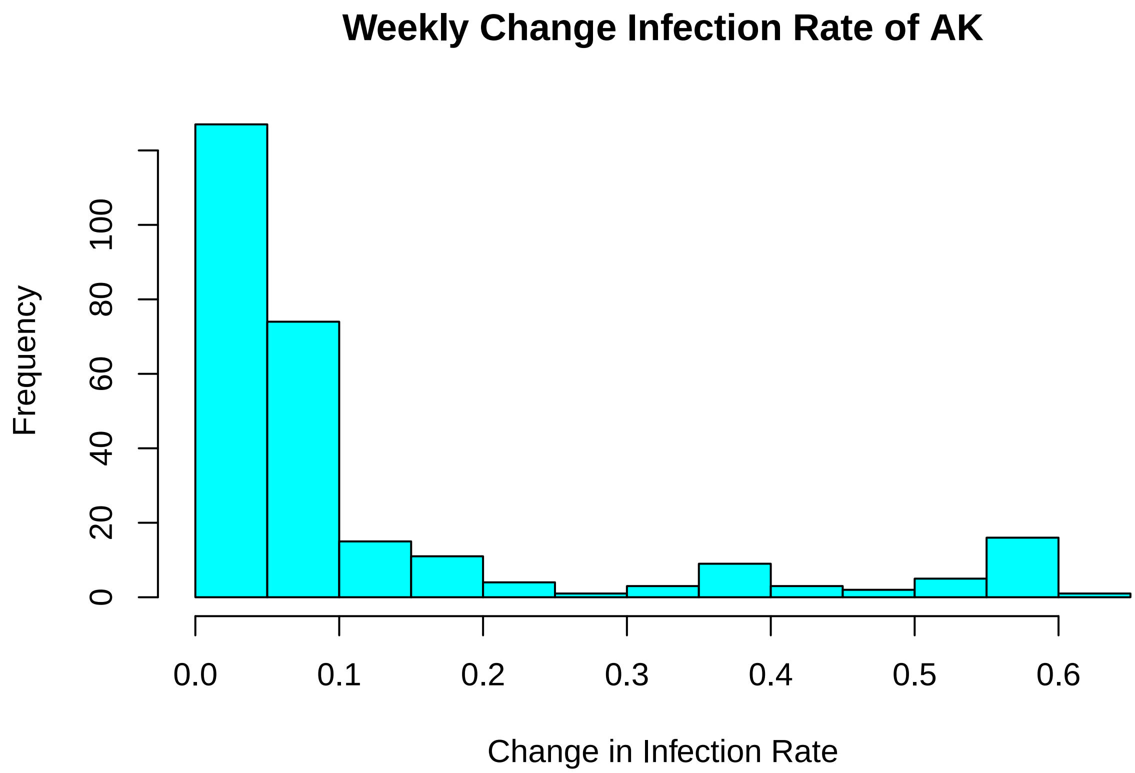



```r
hist(infection_rate_weekly_change$InfectionRate.WY, col= "cyan", xlab="Change in Infection Rate",
        main="Weekly Change Infection Rate of WY")
```



On the other hand in states with low population density, the histogram shows that the week over week change infection rate is lower.

```
hist(infection_rate_weekly_change$InfectionRate.AK, col= "cyan", xlab="Change in Infection Rate",
        main="Weekly Change Infection Rate of AK")
```


Weekly Change Infection Rate of AK
Frequency
0
20
40
60
80
100
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Change in Infection Rate

‘

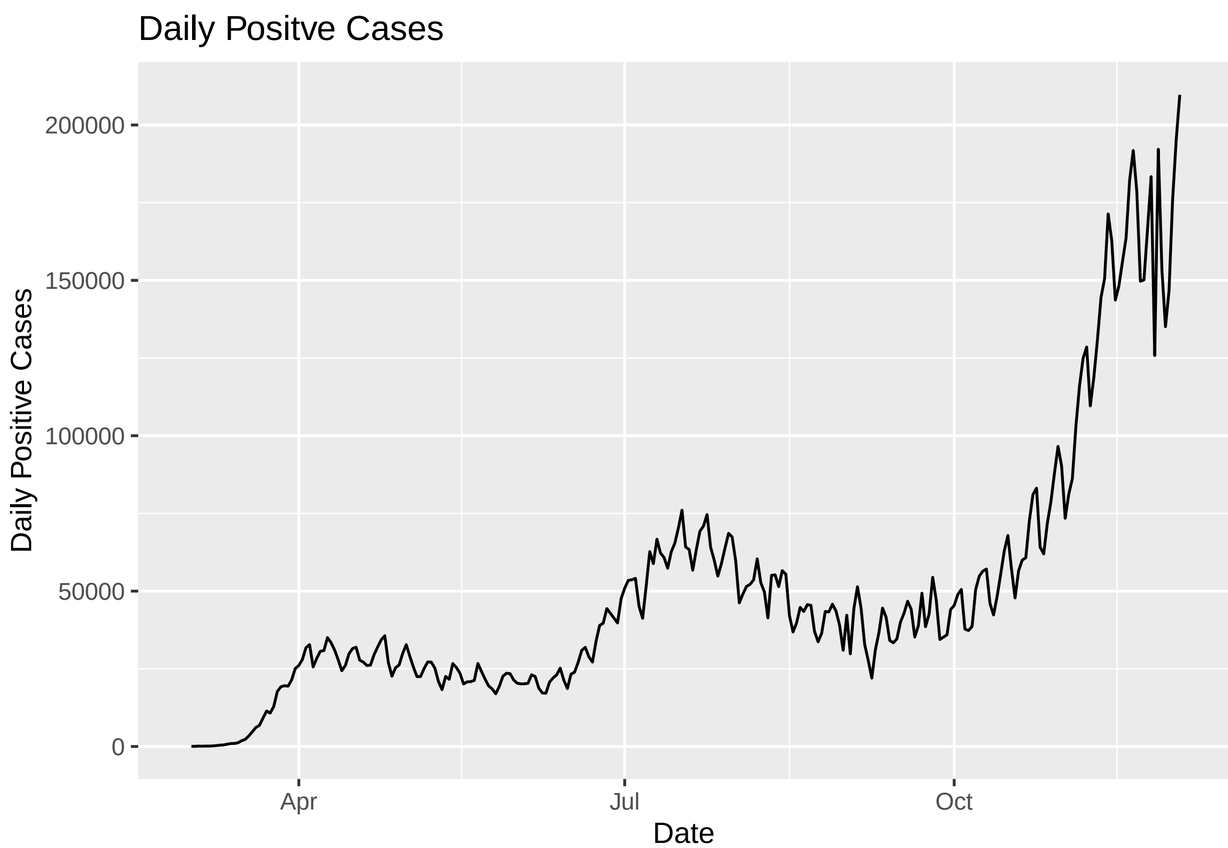
Question3: US Overall COVID Dataset

```
US_Covid_Overall <- UScovid_df %>%
  dplyr::group_by(date) %>%
  dplyr::summarise(
    Death = sum(na.fill(death, 0)),
    Positive = sum(na.fill(positive, 0)),
    TotalTestResults = sum(na.fill(totalTestResults, 0))
    )

US_Covid_Overall <- US_Covid_Overall[order(as.Date(US_Covid_Overall$date, format="%d-%m-%Y")),]

# Population
US_Population <- sum(population_df$Population)

# Infection Rate and Death Rate
US_Covid_Overall$InfectionRate <- US_Covid_Overall$Positive / US_Population
US_Covid_Overall$DeathRate <- na.fill(US_Covid_Overall$Death / US_Covid_Overall$Positive, 0)

# Handling "inf" values created because of division by 0
US_Covid_Overall[which(!is.finite(US_Covid_Overall$DeathRate)), "DeathRate"] <- 0

head(US_Covid_Overall)
```

```
## # A tibble: 6 x 6
##   date       Death Positive TotalTestResults InfectionRate DeathRate
##   <date>     <int>    <int>            <int>         <dbl>     <dbl>
## 1 2020-03-01     8       50             6661   0.000000161    0.16
## 2 2020-03-02    11       94             6873   0.000000303    0.117
## 3 2020-03-03    14      145             7165   0.000000468    0.0966
## 4 2020-03-04    16      279             8206   0.000000900    0.0573
## 5 2020-03-05    20      367             8863   0.00000118     0.0545
## 6 2020-03-06    26      497             9824   0.00000160     0.0523
```

Daily Positive Cases

```
US_Covid_Overall <- US_Covid_Overall %>% mutate(DailyPositiveCases = (Positive - lag(Positive, 1)))
ggplot(US_Covid_Overall, aes(x=as.Date(date, "%d-%m-%Y"), y=DailyPositiveCases)) +
  labs(
    title="Daily Positve Cases",
       x="Date",
    y="Daily Positive Cases")+
  geom_line()
```

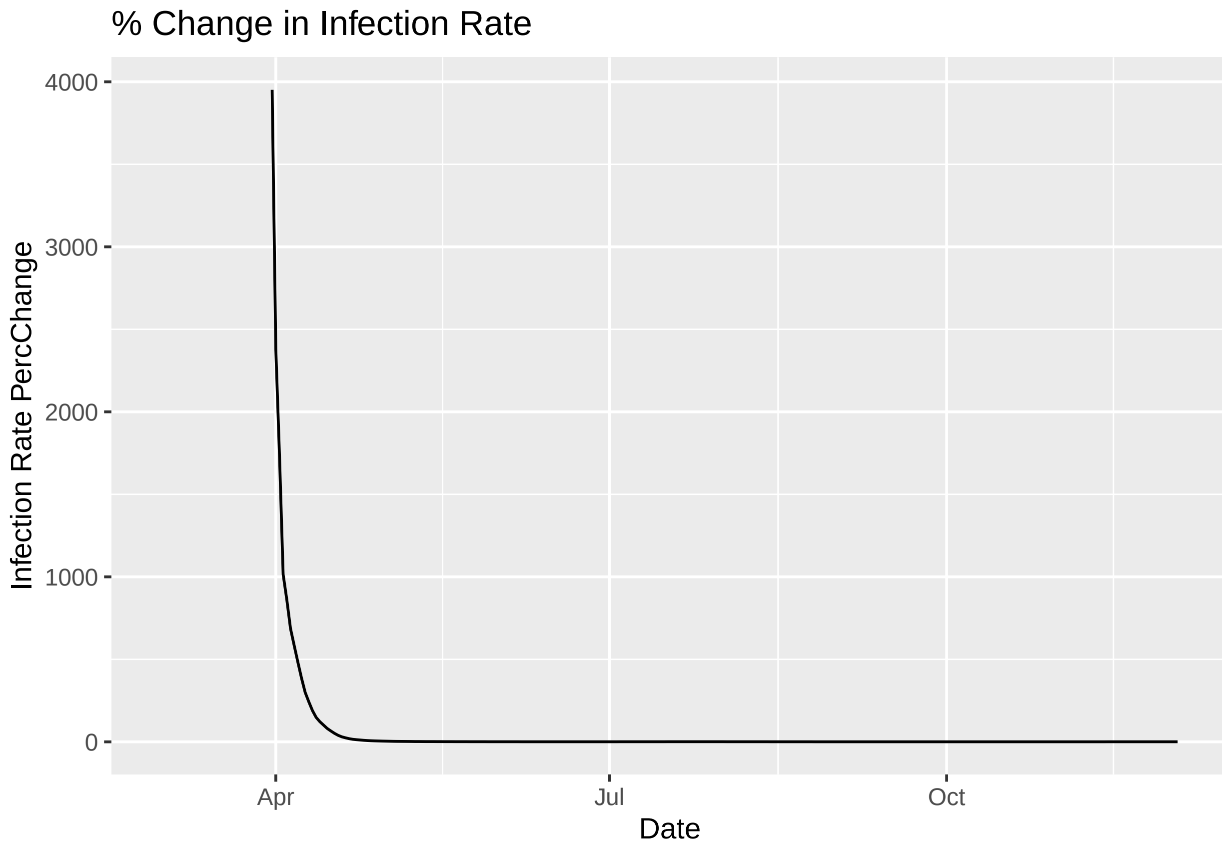


Daily Infection Rate Change in %

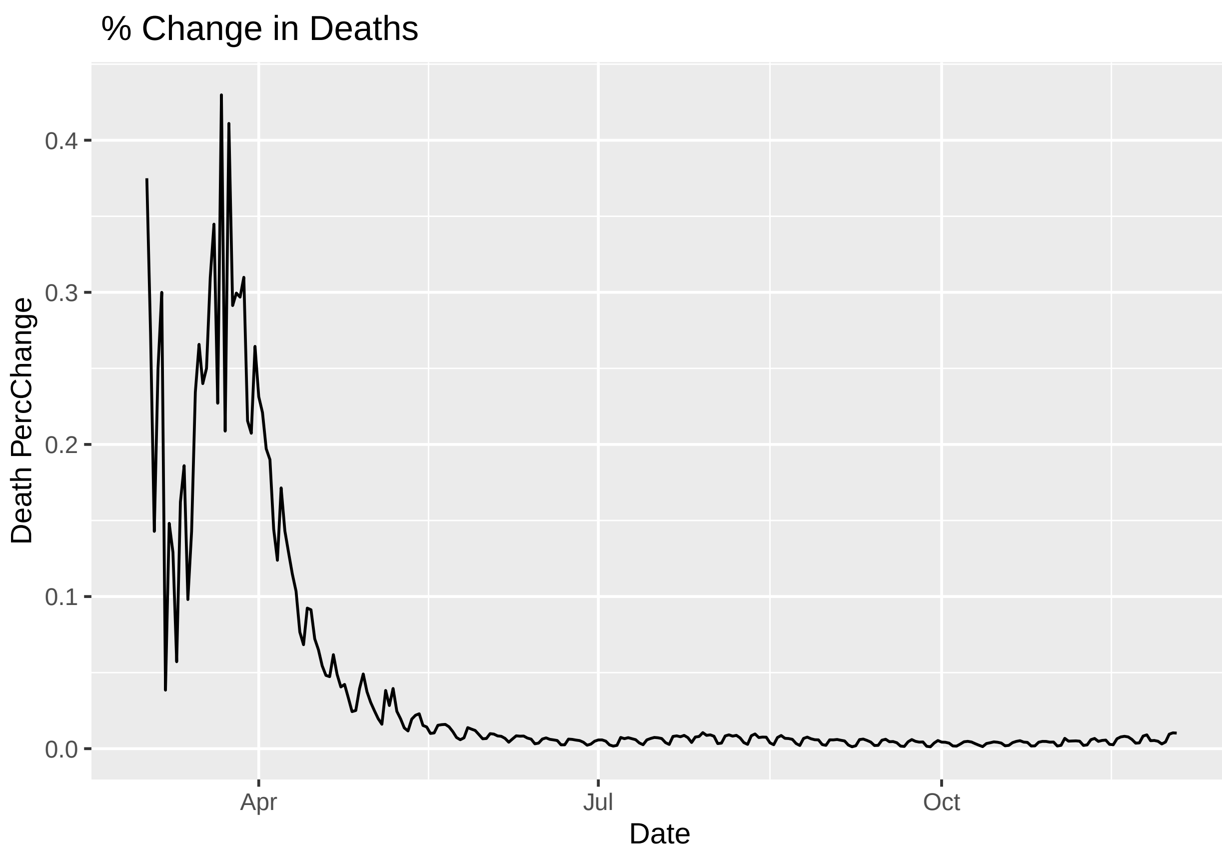
```
US_Covid_Overall <- US_Covid_Overall %>% mutate(InfectionRatePercChange = InfectionRate/lag(InfectionRa
US_Covid_Overall <- US_Covid_Overall %>% mutate(DeathPercChange = Death/lag(Death, 1) - 1)

head(US_Covid_Overall)
```

```
## # A tibble: 6 x 9
##   date       Death Positive TotalTestResults InfectionRate DeathRate
##   <date>     <int>    <int>            <int>         <dbl>     <dbl>
## 1 2020-03-01     8       50             6661   0.000000161    0.16
## 2 2020-03-02    11       94             6873   0.000000303    0.117
## 3 2020-03-03    14      145             7165   0.000000468    0.0966
## 4 2020-03-04    16      279             8206   0.000000900    0.0573
## 5 2020-03-05    20      367             8863   0.00000118     0.0545
## 6 2020-03-06    26      497             9824   0.00000160     0.0523
## # ... with 3 more variables: DailyPositiveCases <int>,
## #   InfectionRatePercChange <dbl>, DeathPercChange <dbl>
```

```
ggplot(US_Covid_Overall, aes(x=as.Date(date, "%d-%m-%Y"), y=InfectionRatePercChange)) +
  labs(
    title="% Change in Infection Rate",
       x="Date",
    y="Infection Rate PercChange")+
  geom_line()
```




```r
ggplot(US_Covid_Overall, aes(x=as.Date(date, "%d-%m-%Y"), y=DeathPercChange)) +
  labs(
    title=" % Change in Deaths",
       x="Date",
    y="Death PercChange")+
  geom_line()
```



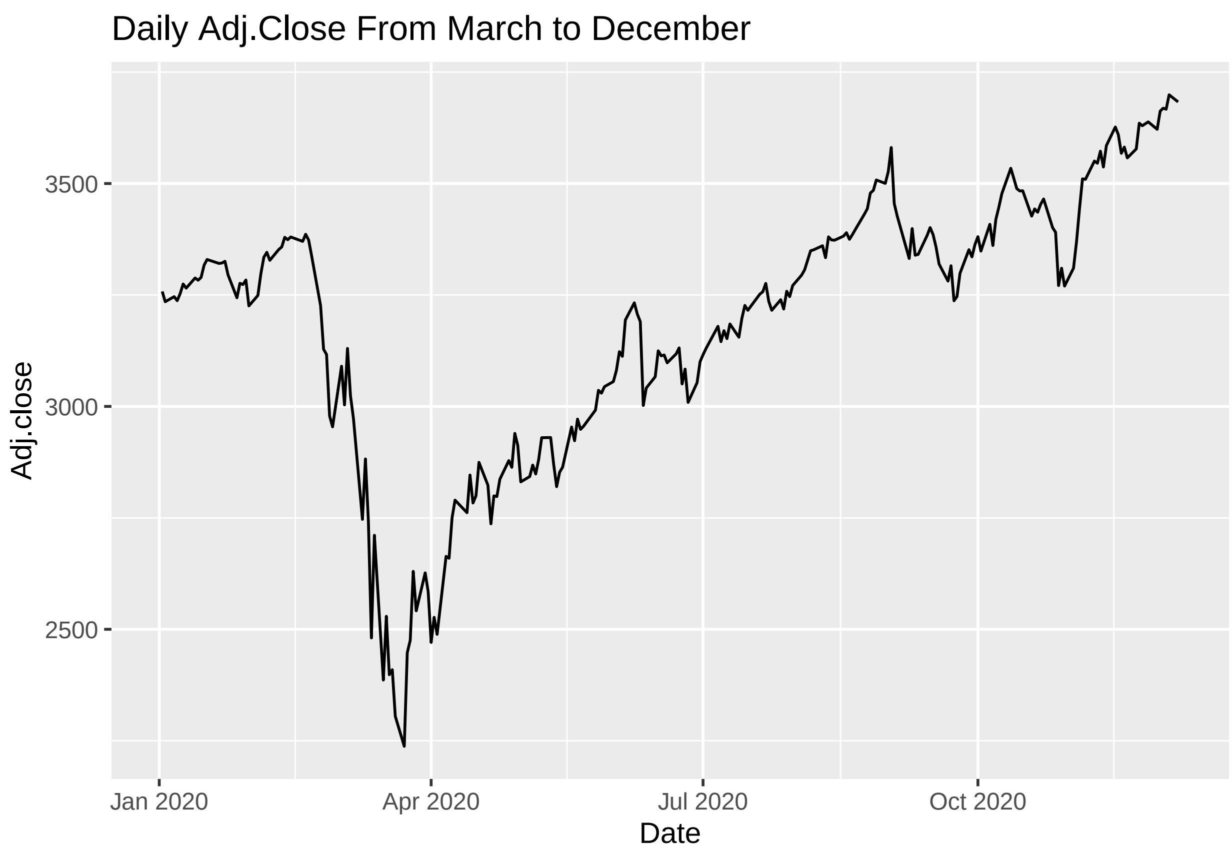
S&P 500

```
sp_500_df <- read.csv("SP500.csv")
sp_500_df$Date <- as.Date(sp_500_df$Date)
head(sp_500_df)
```

```
##         Date    Open    High     Low   Close Adj.Close     Volume
## 1 2020-01-02 3244.67 3258.14 3235.53 3257.85   3257.85 3458250000
## 2 2020-01-03 3226.36 3246.15 3222.34 3234.85   3234.85 3461290000
## 3 2020-01-06 3217.55 3246.84 3214.64 3246.28   3246.28 3674070000
## 4 2020-01-07 3241.86 3244.91 3232.43 3237.18   3237.18 3420380000
## 5 2020-01-08 3238.59 3267.07 3236.67 3253.05   3253.05 3720890000
## 6 2020-01-09 3266.03 3275.58 3263.67 3274.70   3274.70 3638390000
```

Visualize adjusted close

```
ggplot(sp_500_df, aes(x=as.Date(Date), y=Adj.Close)) +
  labs(
    title="Daily Adj.Close From March to December",
       x="Date",
    y="Adj.close")+
  geom_line()
```



Daily Returns

```
sp_500_df <- sp_500_df %>% mutate(Returns = Adj.Close/lag(Adj.Close, 30) - 1)
head(sp_500_df)
```

```
##         Date    Open    High     Low   Close Adj.Close     Volume Returns
## 1 2020-01-02 3244.67 3258.14 3235.53 3257.85   3257.85 3458250000      NA
## 2 2020-01-03 3226.36 3246.15 3222.34 3234.85   3234.85 3461290000      NA
## 3 2020-01-06 3217.55 3246.84 3214.64 3246.28   3246.28 3674070000      NA
## 4 2020-01-07 3241.86 3244.91 3232.43 3237.18   3237.18 3420380000      NA
## 5 2020-01-08 3238.59 3267.07 3236.67 3253.05   3253.05 3720890000      NA
## 6 2020-01-09 3266.03 3275.58 3263.67 3274.70   3274.70 3638390000      NA
```

```
compute_corr <- function(infection_rate_df, asset_returns_df, col, from, to){

 infection_rate_df <- filter(infection_rate_df, as.Date(Date) >= as.Date(from), as.Date(Date) <= as.Date
 asset_returns_df <- filter(asset_returns_df, as.Date(Date) >= as.Date(from), as.Date(Date) <= as.Date(t

 asset_returns_df <- plyr::join(asset_returns_df, infection_rate_df, by="Date")
 asset_returns_df <- na.omit(asset_returns_df)

 # Normalize
 # asset_returns_df[,col] = (asset_returns_df[,col] - mean(asset_returns_df[,col]))/sd(asset_returns_df

 correlation <- cor(asset_returns_df[,col], asset_returns_df$Returns)

 #print(asset_returns_df)
```

```r
  #plot(asset_returns_df[,col], asset_returns_df$Returns)

  return(correlation)
}
```

Correlation between S&P 500 Daily Returns and Daily Infection Rate Percentage Change.

```r
infection_rate_df <- US_Covid_Overall[,c("date", "InfectionRatePercChange", "DailyPositiveCases")]
infection_rate_df$Date <- as.Date(infection_rate_df$date, "%d-%m-%Y")
infection_rate_df <- infection_rate_df[,-1]

asset_returns_df <- sp_500_df[, c("Date", "Returns")]
asset_returns_df$Date <- as.Date(asset_returns_df$Date)

compute_corr(infection_rate_df, asset_returns_df, "InfectionRatePercChange", "2020-03-01", "2020-12-01")
```

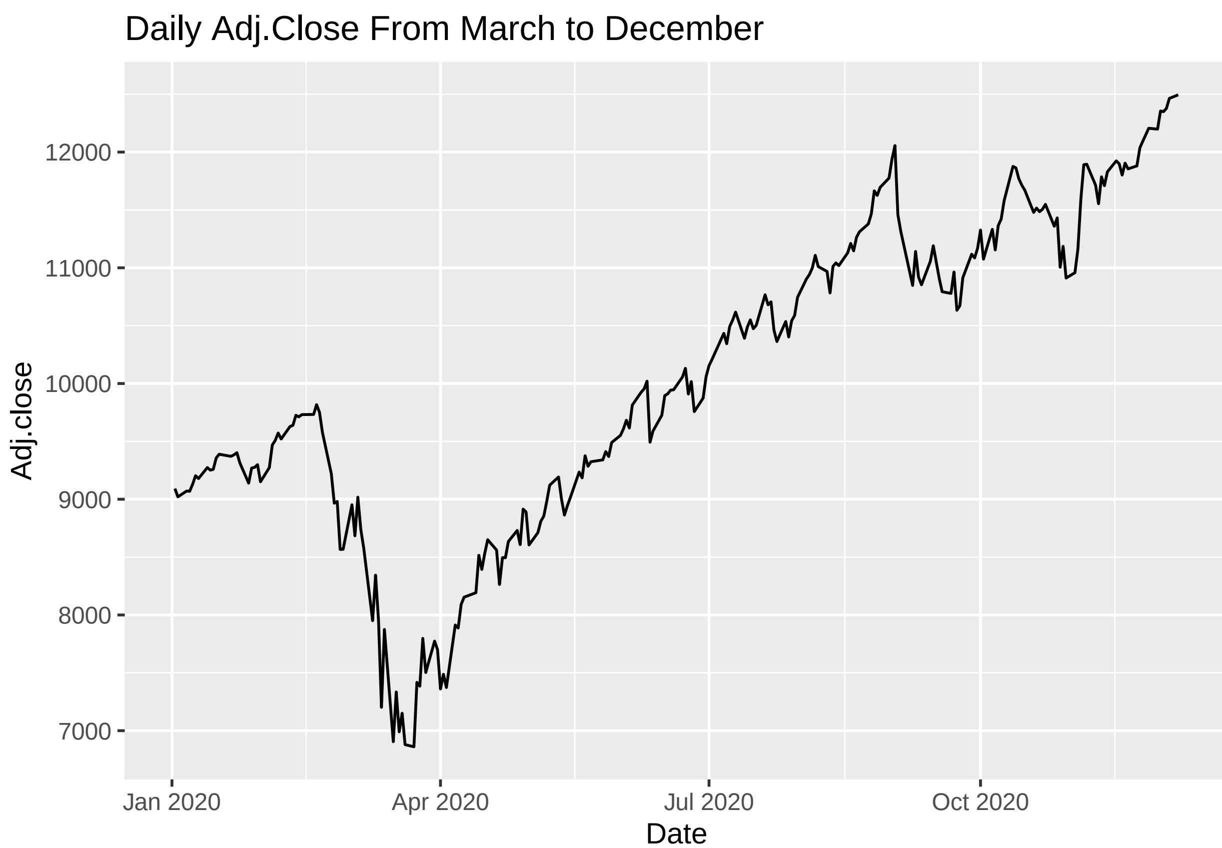
```
## [1] -0.5811911
```

NASDAQ Composite

```r
nasdaq_composite <- read.csv("NASDAQ Composite.csv")
head(nasdaq_composite)
```

```
##         Date    Open    High     Low   Close Adj.Close     Volume
## 1 2020-01-02 9039.46 9093.43 9010.89 9092.19   9092.19 2848370000
## 2 2020-01-03 8976.43 9065.76 8976.43 9020.77   9020.77 2567400000
## 3 2020-01-06 8943.50 9072.41 8943.50 9071.47   9071.47 2788120000
## 4 2020-01-07 9076.64 9091.93 9042.55 9068.58   9068.58 2352850000
## 5 2020-01-08 9068.03 9168.89 9059.38 9129.24   9129.24 2464090000
## 6 2020-01-09 9202.27 9215.95 9158.50 9203.43   9203.43 2534700000
```

```r
ggplot(nasdaq_composite, aes(x=as.Date(Date), y=Adj.Close)) +
  labs(
    title="Daily Adj.Close From March to December",
       x="Date",
    y="Adj.close")+
  geom_line()
```




```r
nasdaq_composite <- nasdaq_composite %>% mutate(Returns = Adj.Close/lag(Adj.Close, 30) - 1)
head(nasdaq_composite)
```

```
##         Date    Open    High     Low   Close Adj.Close     Volume Returns
## 1 2020-01-02 9039.46 9093.43 9010.89 9092.19   9092.19 2848370000      NA
## 2 2020-01-03 8976.43 9065.76 8976.43 9020.77   9020.77 2567400000      NA
## 3 2020-01-06 8943.50 9072.41 8943.50 9071.47   9071.47 2788120000      NA
## 4 2020-01-07 9076.64 9091.93 9042.55 9068.58   9068.58 2352850000      NA
## 5 2020-01-08 9068.03 9168.89 9059.38 9129.24   9129.24 2464090000      NA
## 6 2020-01-09 9202.27 9215.95 9158.50 9203.43   9203.43 2534700000      NA
```

Correlation between NASDAQ Daily Returns and Daily Infection Rate Percentage Change.

```r
asset_returns_df <- nasdaq_composite[, c("Date", "Returns")]
asset_returns_df$Date <- as.Date(asset_returns_df$Date)

compute_corr(infection_rate_df, asset_returns_df, "InfectionRatePercChange", "2020-04-01", "2020-12-01")
```

```
## [1] -0.5566532
```

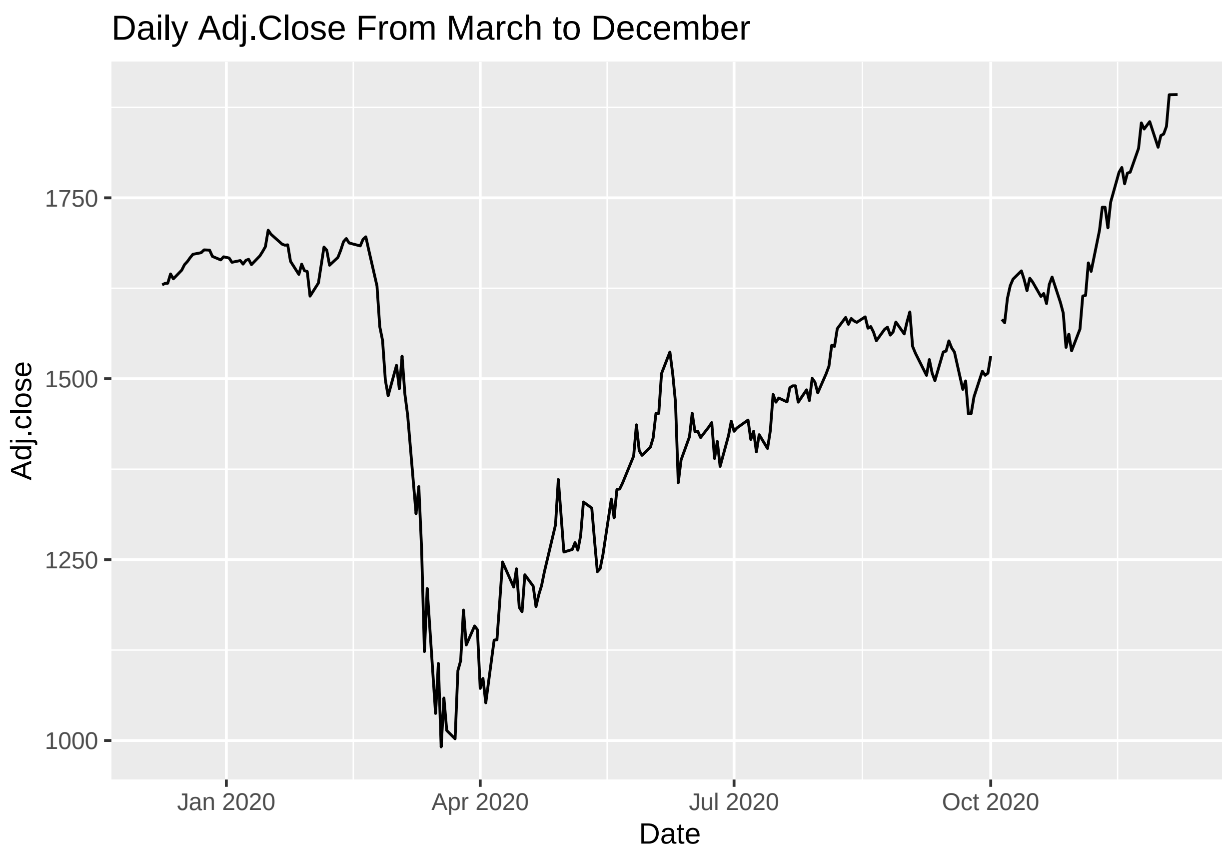
Russell 2000

```r
russel_2000 <- read.csv("Russel 2000.csv")
russel_2000$Date <- as.Date(russel_2000$Date)
russel_2000$Adj.Close <- as.numeric(russel_2000$Adj.Close)
head(russel_2000)
```

```
##         Date        Open        High         Low       Close Adj.Close   Volume
## 1 2019-12-09 1631.469971 1636.439941 1629.619995 1629.619995   1629.62 33459900
```

```
## 2 2019-12-10 1629.099976 1633.739990 1626.369995 1631.709961   1631.71 33437900
## 3 2019-12-11 1633.430054 1634.430054 1626.739990 1631.930054   1631.93 32525400
## 4 2019-12-12 1631.800049 1654.069946 1629.859985 1644.810059   1644.81 39906900
## 5 2019-12-13 1643.829956 1650.189941 1632.650024 1637.979980   1637.98 37368700
## 6 2019-12-16 1647.199951 1658.619995 1647.199951 1649.939941   1649.94 40517900
```

```
ggplot(russel_2000, aes(x=as.Date(Date), y=Adj.Close)) +
  labs(
    title="Daily Adj.Close From March to December",
       x="Date",
    y="Adj.close")+
  geom_line()
```




```
russel_2000 <- russel_2000 %>% mutate(Returns = Adj.Close/lag(Adj.Close, 30) - 1)
head(russel_2000)
```

```
##         Date        Open        High         Low       Close Adj.Close   Volume
## 1 2019-12-09 1631.469971 1636.439941 1629.619995 1629.619995   1629.62 33459900
## 2 2019-12-10 1629.099976 1633.739990 1626.369995 1631.709961   1631.71 33437900
## 3 2019-12-11 1633.430054 1634.430054 1626.739990 1631.930054   1631.93 32525400
## 4 2019-12-12 1631.800049 1654.069946 1629.859985 1644.810059   1644.81 39906900
## 5 2019-12-13 1643.829956 1650.189941 1632.650024 1637.979980   1637.98 37368700
## 6 2019-12-16 1647.199951 1658.619995 1647.199951 1649.939941   1649.94 40517900
##   Returns
## 1      NA
## 2      NA
## 3      NA
```

```
## 4      NA
## 5      NA
## 6      NA
```

```
asset_returns_df <- russel_2000[, c("Date", "Returns")]
asset_returns_df$Date <- as.Date(asset_returns_df$Date)

compute_corr(infection_rate_df, asset_returns_df, "InfectionRatePercChange", "2020-04-01", "2020-12-01")
```

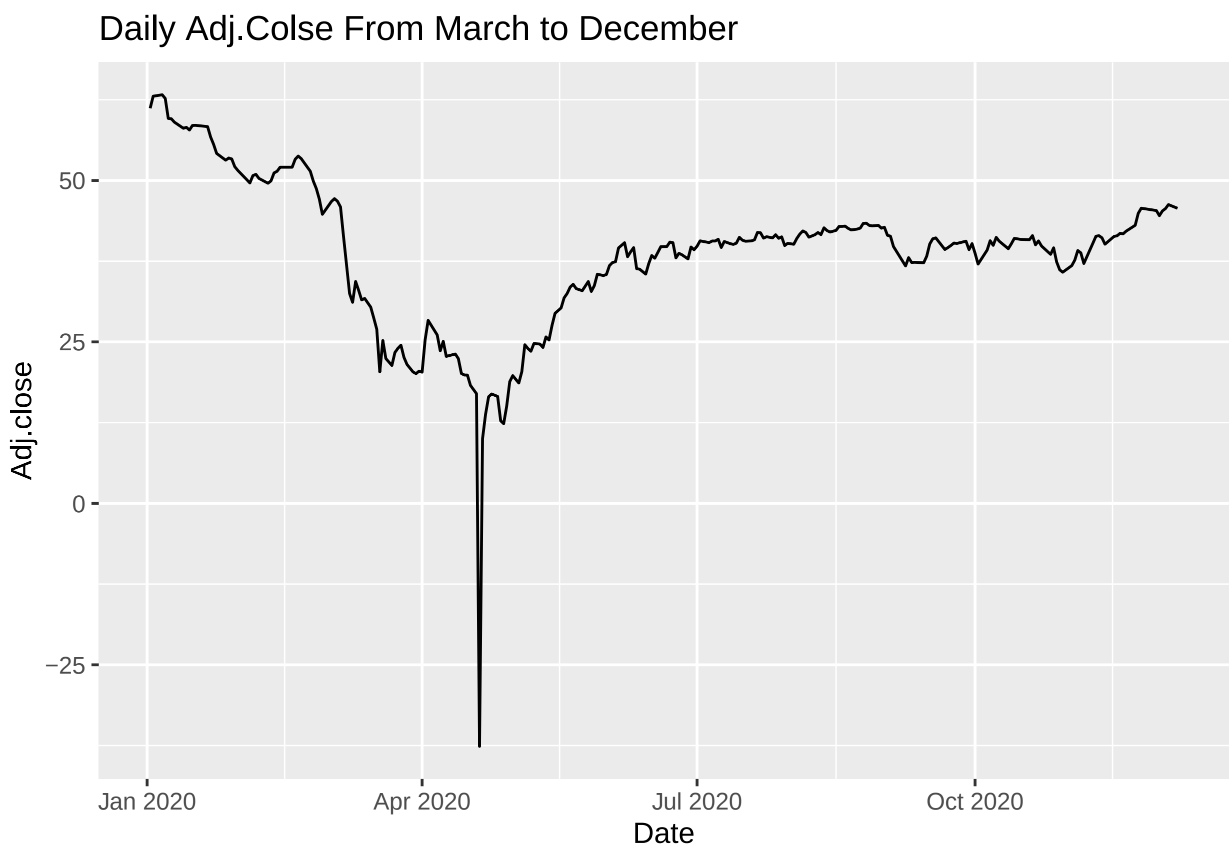
```
## [1] -0.5817983
```

Crude Oil Futures

```
crude_oil_futures <- read.csv("Crude Oil Futures.csv")
crude_oil_futures <- crude_oil_futures%>% filter(Open != "null")
crude_oil_futures$Adj.Close <- as.numeric(crude_oil_futures$Adj.Close)
head(crude_oil_futures)
```

```
##         Date      Open      High       Low     Close Adj.Close  Volume
## 1 2020-01-02 61.599998 61.599998 60.639999 61.180000     61.18  486873
## 2 2020-01-03 61.180000 64.089996 61.130001 63.049999     63.05  885861
## 3 2020-01-06 63.709999 64.720001 62.639999 63.270000     63.27  724236
## 4 2020-01-07 62.910000 63.150002 62.110001 62.700001     62.70  582649
## 5 2020-01-08 62.840000 65.650002 59.150002 59.610001     59.61 1205710
## 6 2020-01-09 59.990002 60.310001 58.660000 59.560001     59.56  750933
```

```
ggplot(crude_oil_futures, aes(x=as.Date(Date), y=Adj.Close)) +
  labs(
    title="Daily Adj.Colse From March to December",
       x="Date",
    y="Adj.close")+
  geom_line()
```




```
crude_oil_futures <- crude_oil_futures %>% mutate(Returns = Adj.Close/lag(Adj.Close, 30) - 1)
head(crude_oil_futures)
```

```
##         Date      Open      High       Low     Close Adj.Close  Volume Returns
## 1 2020-01-02 61.599998 61.599998 60.639999 61.180000     61.18  486873      NA
## 2 2020-01-03 61.180000 64.089996 61.130001 63.049999     63.05  885861      NA
## 3 2020-01-06 63.709999 64.720001 62.639999 63.270000     63.27  724236      NA
## 4 2020-01-07 62.910000 63.150002 62.110001 62.700001     62.70  582649      NA
## 5 2020-01-08 62.840000 65.650002 59.150002 59.610001     59.61 1205710      NA
## 6 2020-01-09 59.990002 60.310001 58.660000 59.560001     59.56  750933      NA
```

```
asset_returns_df <- crude_oil_futures[, c("Date", "Returns")]
asset_returns_df$Date <- as.Date(asset_returns_df$Date)

compute_corr(infection_rate_df, asset_returns_df, "InfectionRatePercChange", "2020-04-01", "2020-12-01")
```

```
## [1] -0.2114514
```

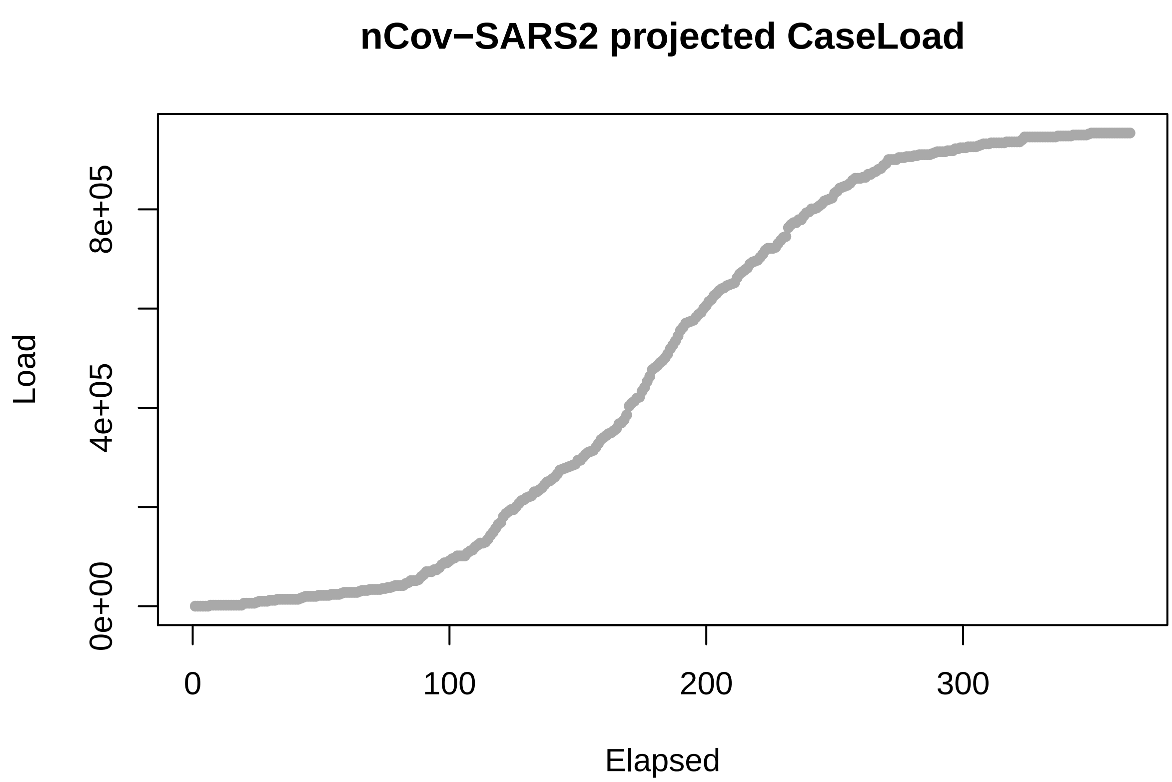
Question 4:

The clients' simulation model tries to project the COVID load by sampling from a collection of poisson distributions.

- In the first step, a set of means is computed using an exponential function which will further be used as "lambdas" in the poisson distributions.
- The means are calculated in such a way that, for the early days of the projection the means of the poisson distributions are close to 0 indicating the slow start of the COVID spread. as the number of days reach the middle the means are close to 1, indicating highest increase in COVID spread.

- After the means are calculated, daily COVID cases are simulated by sampling from a series of poission distributions using the lambda's calculated above.
- The sampled values are multiplied by 1987.32 which looks like the average number of daily cases from the data.

```
num_days = 365
days <- 1:num_days
lambda_sim <- exp(-0.32*((days-182)^2/51.6^2))
#plot(lambda_sim)

W <- 1987.32*rpois(num_days,pi*lambda_sim)

plot(cumsum(W), xlab = "Elapsed", ylab = "Load",
main = 'nCov-SARS2 projected CaseLoad', pch = 16, cex = 0.75, col = "darkgrey")
```

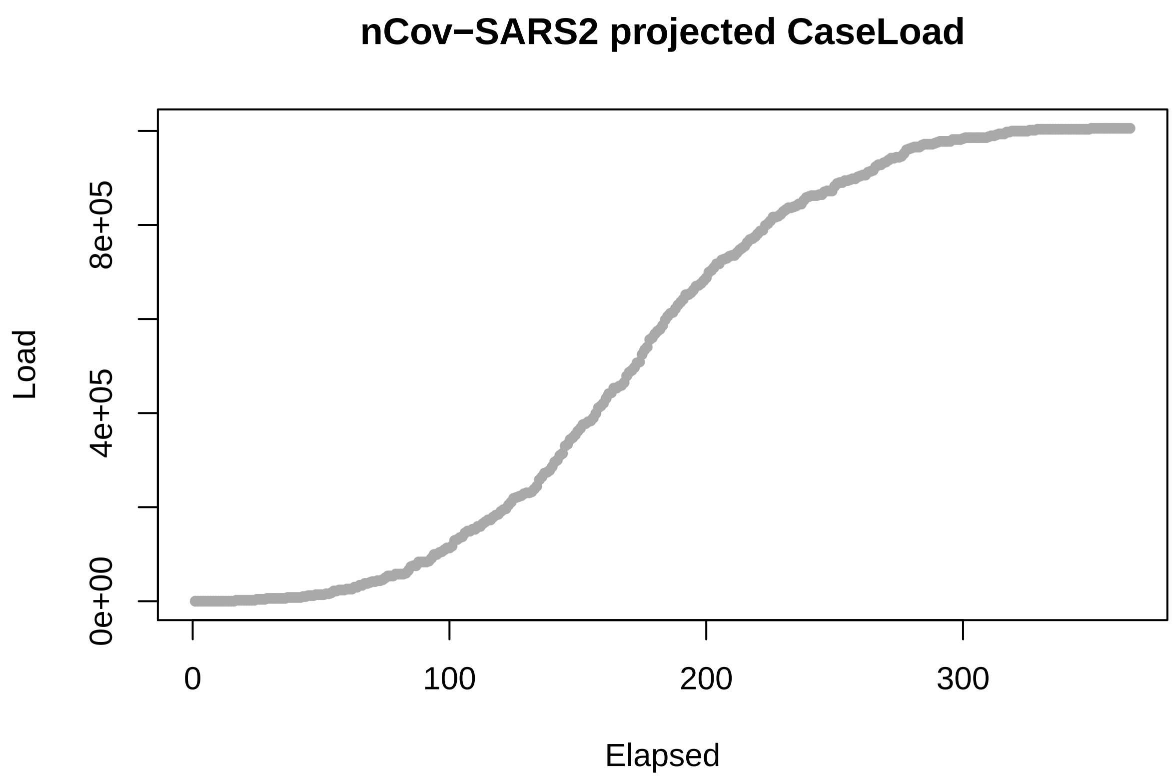


Sensitivity Analysis

Slightly changing the "-0.32" coefficient of the exponential function, changes the trend of the projection to a great extent. Bringing that value close to 0 results in a linear trend in the projection. This coefficient determines the means of the Poisson distributions used to sample daily cases. Therefore, it decides the projected rate of the spread of the disease from one day to the next. Although this model captures the exponential growth of the virus, it incorrectly projects that the spread of the virus levels off, and the number of daily cases comes close to 0 towards the end of the 365 days which is not observed from the historical data we analyzed.

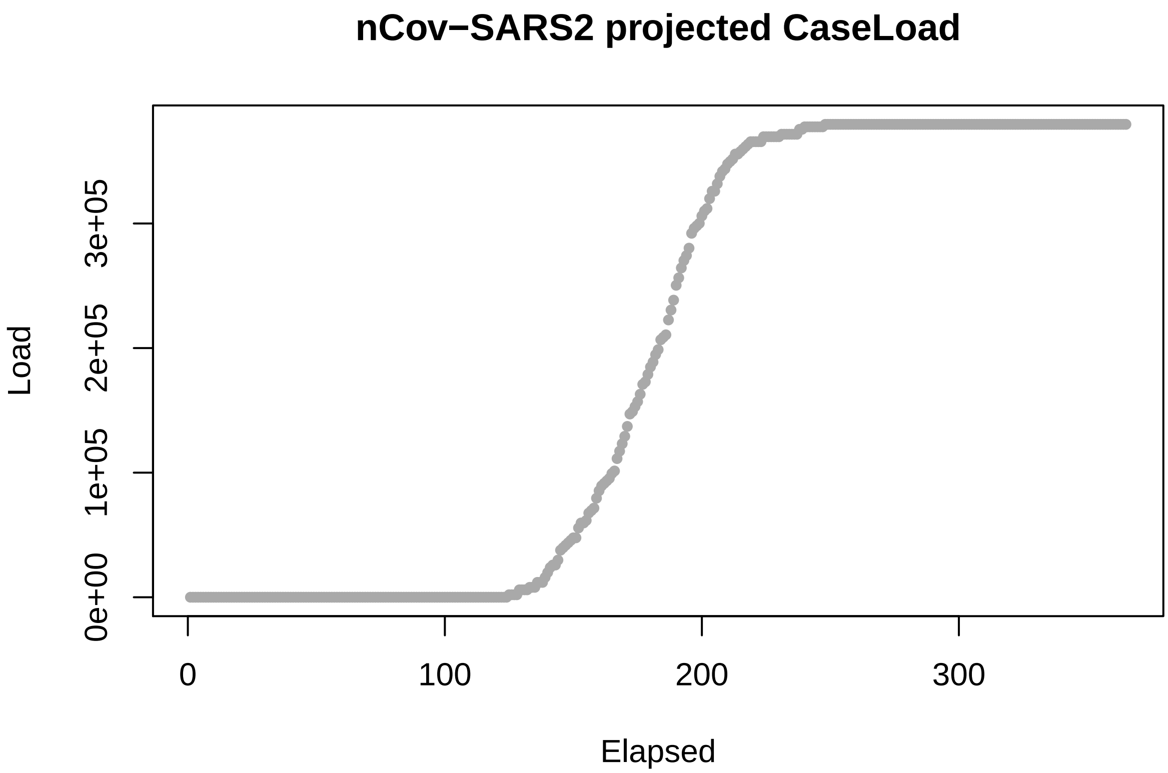
```
num_days = 365
days <- 1:num_days
lambda_sim <- exp(-0.32*((days-182)^2/51.6^2))
```

```r
W <- 1987.32*rpois(num_days,pi*lambda_sim)

plot(cumsum(W), xlab = "Elapsed", ylab = "Load",
main = 'nCov-SARS2 projected CaseLoad', pch = 16, cex = 0.75, col = "darkgrey")
```



Changing that coefficient to a value further away from 0, brings the trend of the projetion to a steep exponential curve.

```r
num_days = 365
days <- 1:num_days
lambda_sim <- exp(-2*((days-182)^2/51.6^2))


W <- 1987.32*rpois(num_days,pi*lambda_sim)

plot(cumsum(W), xlab = "Elapsed", ylab = "Load",
main = 'nCov-SARS2 projected CaseLoad', pch = 16, cex = 0.75, col = "darkgrey")
```


nCov−SARS2 projected CaseLoad
Load
0e+00
1e+05
2e+05
3e+05
0
100
200
300
Elapsed