COVID19

Ji

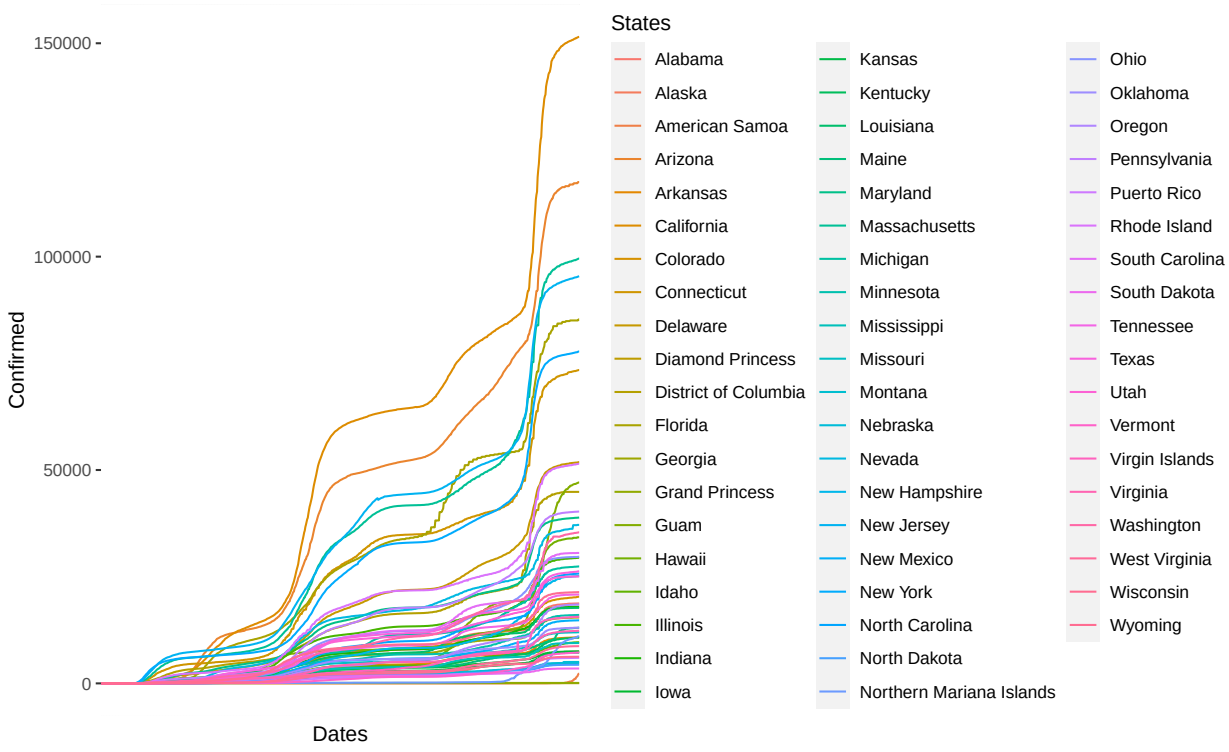
3/27/2022

Firstly, I imported the confirmed cases of covid19 data (01/22/2020-03/25/2022). Then I looked at the trends of mean confirmed cases in different states. Finally, I checked the percentage of confirmed cases among all states along that period of time.

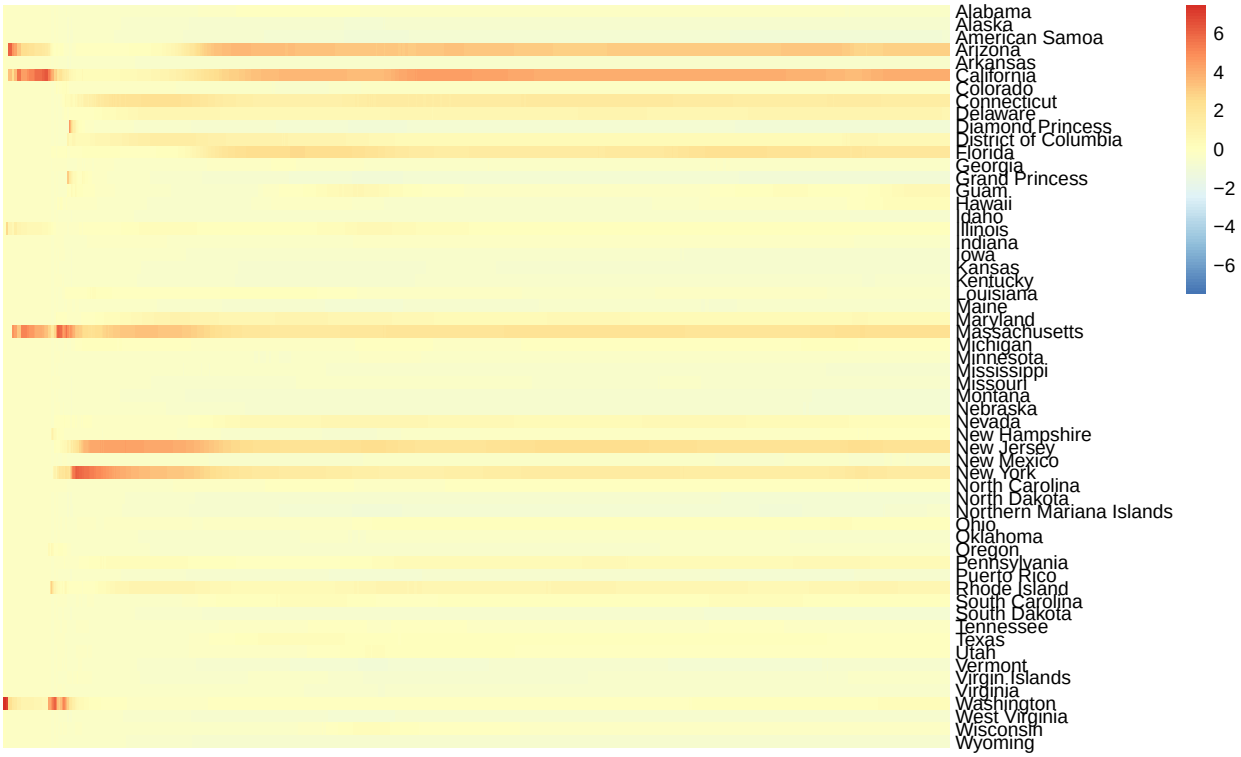
```
library(ggplot2)
library(stringr)

df = read.csv("time_series_covid19_confirmed_US.csv",header = T)
df1 = df[,c(7,12:806)]
df1$Province_State = factor(df1$Province_State)

df2 = aggregate(df1, by=list(df$Province_State), FUN=mean)[,-2]
df2 = reshape::melt(df2)
```

```
## Using Group.1 as id variables
```

```
colnames(df2) = c("States", "Dates","Confirmed")

ggplot(data = df2, aes_string(x="Dates",y="Confirmed",color="States", group="States"))+
  geom_line()+
  theme(axis.ticks.x = element_blank(),
        axis.text.x = element_blank())
```

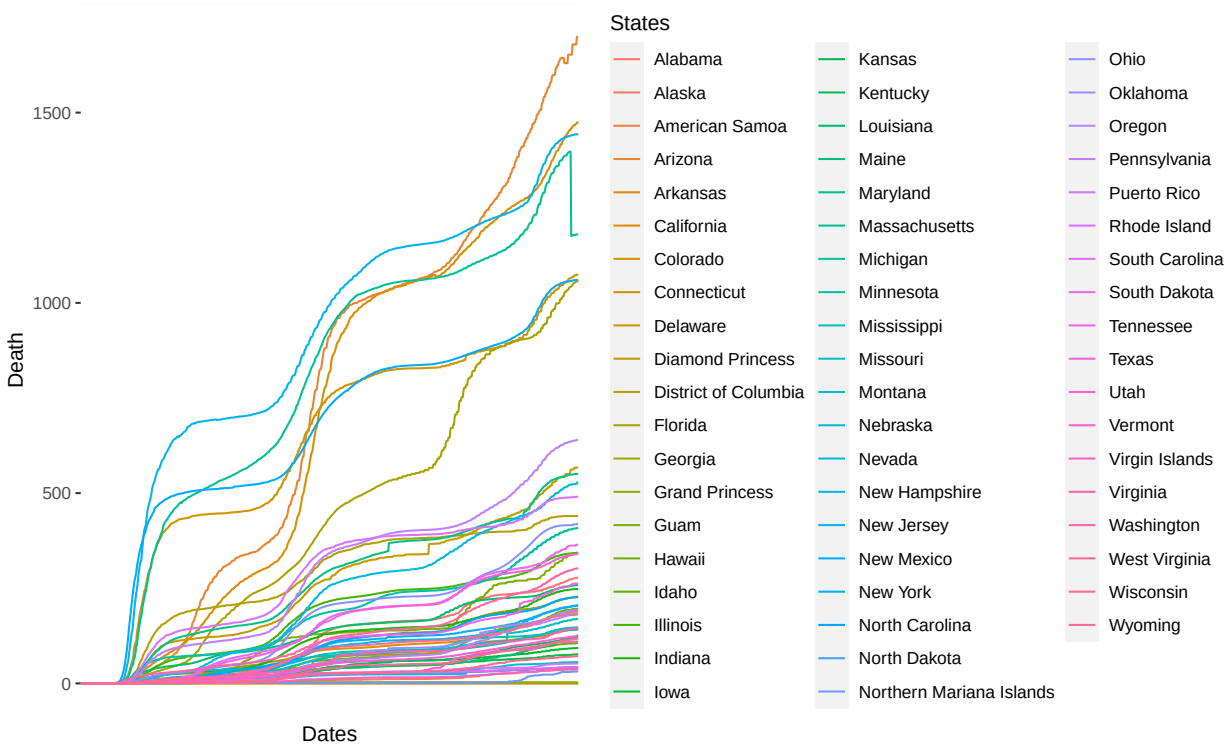



```
library(pheatmap)
df = read.csv("time_series_covid19_confirmed_US.csv",header = T)
df1 = df[,c(7,12:806)]
df1$Province_State = factor(df1$Province_State)

df2 = aggregate(df1, by=list(df$Province_State), FUN=mean)[,-2]
rownames(df2) = df2$Group.1
df2 = df2[,-1]
pheatmap(df2, cluster_cols = F, cluster_rows = F, show_colnames = F, scale = "column")
```

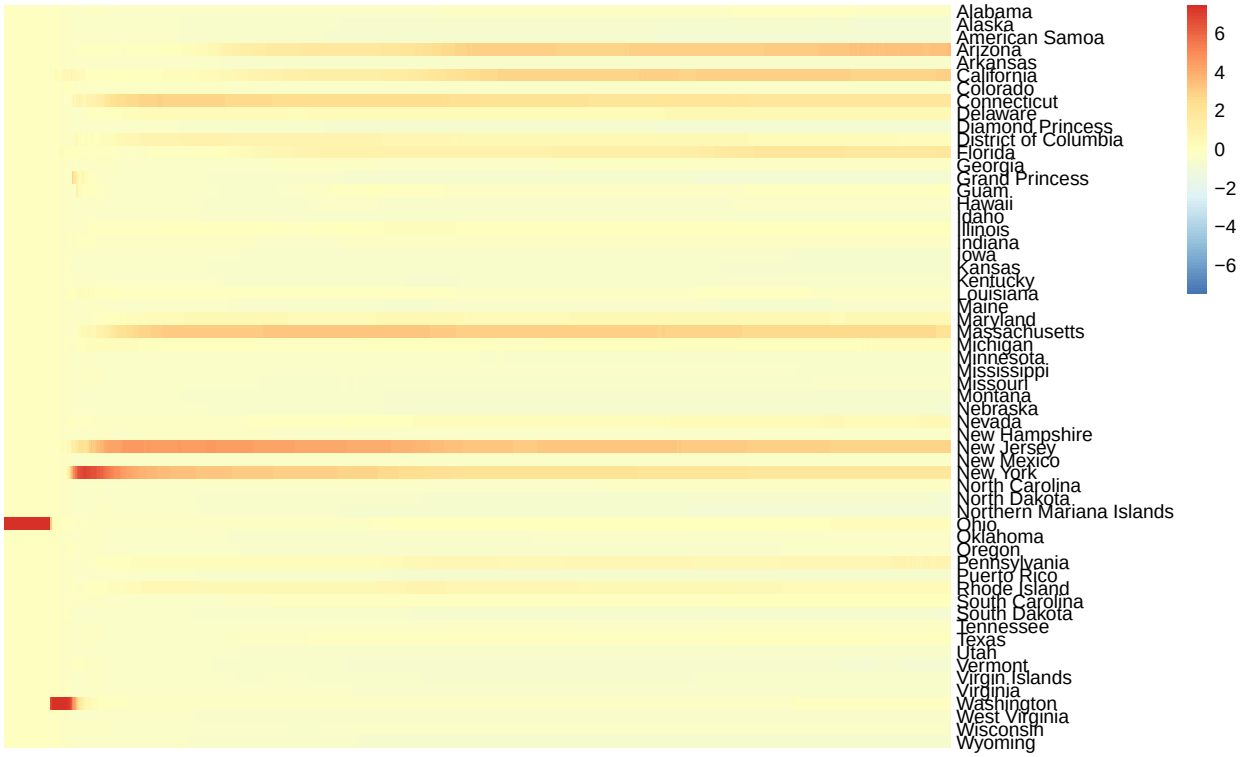


Firstly, I imported the death cases of covid19 data (01/22/2020-03/25/2022). Then I looked at the trends of mean death cases in different states. Finally, I checked the percentage of death cases among all states along that period of time.

```
df3 = read.csv("time_series_covid19_deaths_US.csv",header = T)
df4 = df3[,c(7,13:806)]
df4$Province_State = factor(df4$Province_State)

df5 = aggregate(df4, by=list(df$Province_State), FUN=mean)[,-2]
df5 = reshape::melt(df5)


## Using Group.1 as id variables

colnames(df5) = c("States", "Dates","Death")

ggplot(data = df5, aes_string(x="Dates",y="Death",color="States", group="States"))+
  geom_line()+
  theme(axis.ticks.x = element_blank(),
        axis.text.x = element_blank())
```




```
library(pheatmap)
df = read.csv("time_series_covid19_deaths_US.csv",header = T)
df1 = df[,c(7,13:806)]
df1$Province_State = factor(df1$Province_State)

df2 = aggregate(df1, by=list(df$Province_State), FUN=mean)[,-2]
rownames(df2) = df2$Group.1
df2 = df2[,-1]
pheatmap(df2, cluster_cols = F, cluster_rows = F, show_colnames = F, scale = "column")
```


Alabama
Alaska
American Samoa
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Diamond Princess
District of Columbia
Florida
Georgia
Grand Princess
Guam
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Northern Mariana Islands
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virgin Islands
Virginia
Washington
West Virginia
Wisconsin
Wyoming
6
4
2
0
−2
−4
−6