COVID-19 Data Analysis

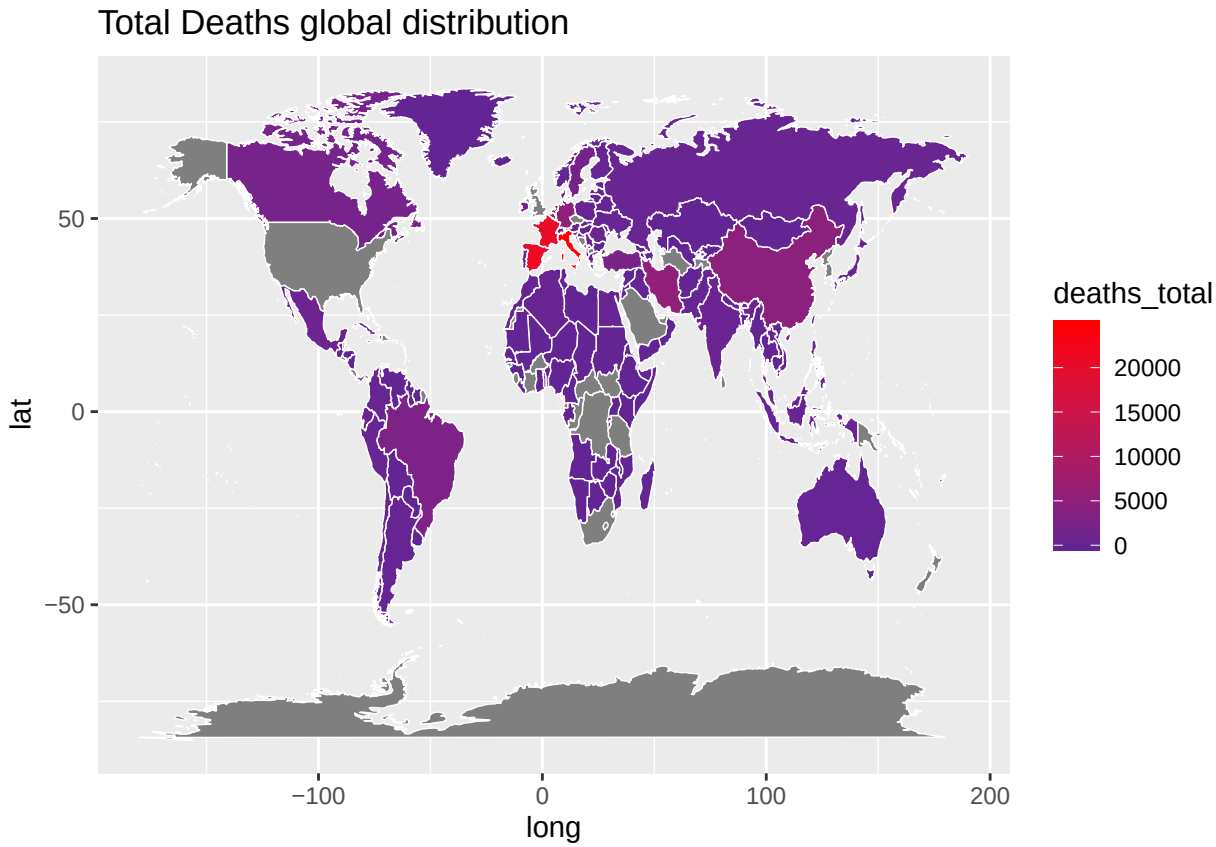
Dagmawe Haileslassie and Michael Tiede

4/22/2020

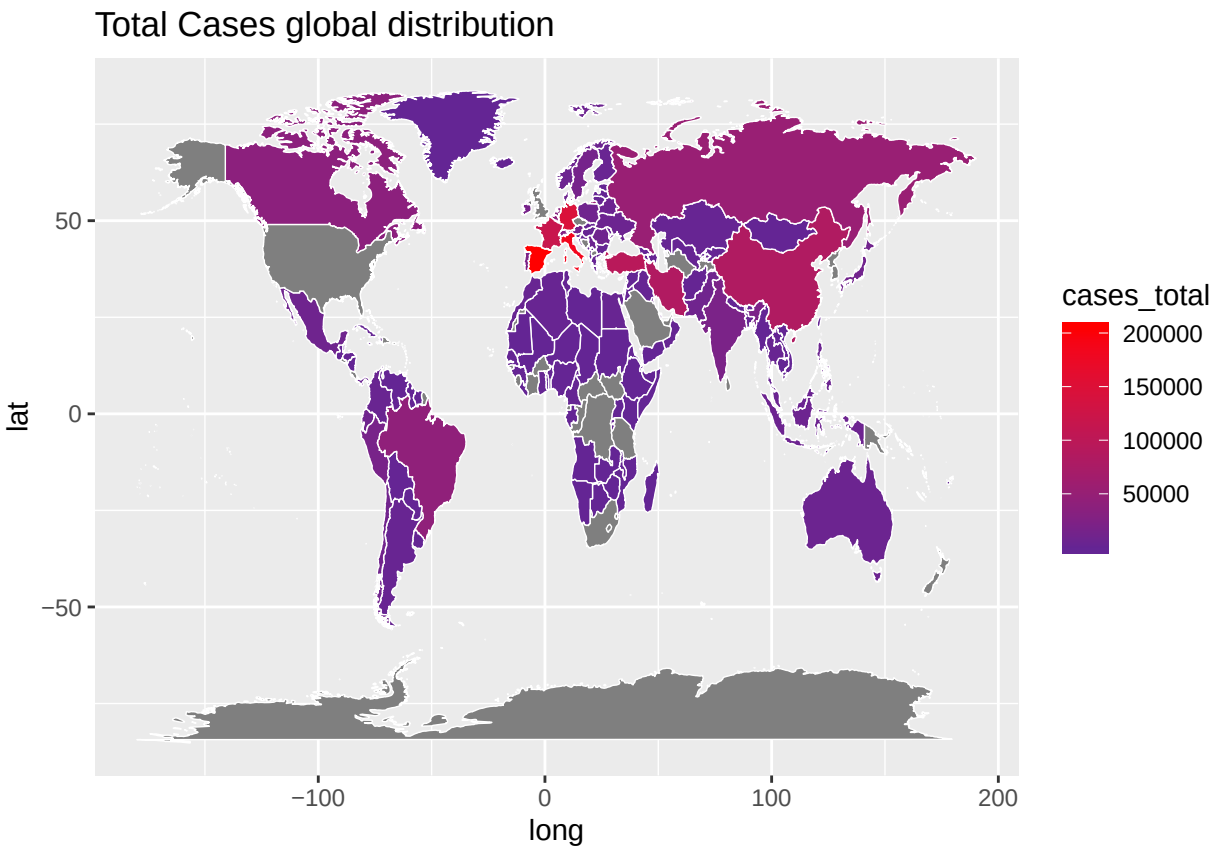
Global Data

Before we start looking at specific data for the relationship bettween COVID-19 and different variables like (testing, temprature, Mobility/footprint, number of airports etc..) we should look at the distribution of the total cases for each country.

```
map.world <- map_data("world")
COVID_TD <- COVID.19.geographic.disbtribution.worldwide %>%
  group_by(countriesAndTerritories, continentExp) %>%
  summarise(cases_total=as.double(sum(cases)),
            deaths_total=sum(deaths))%>%
  ungroup()
map.world_joined <- left_join(map.world, COVID_TD, by = c('region' = 'countriesAndTerritories'))
ggplot() +
  geom_polygon(data=map.world_joined, aes(x=long, y=lat, group=group, fill = deaths_total),  color="whi
  scale_fill_gradient2(low = "blue",
                       mid = scales::muted("purple"),
                       high = "red")+
   labs(title='Total Deaths global distribution', label = "Total cases")
```

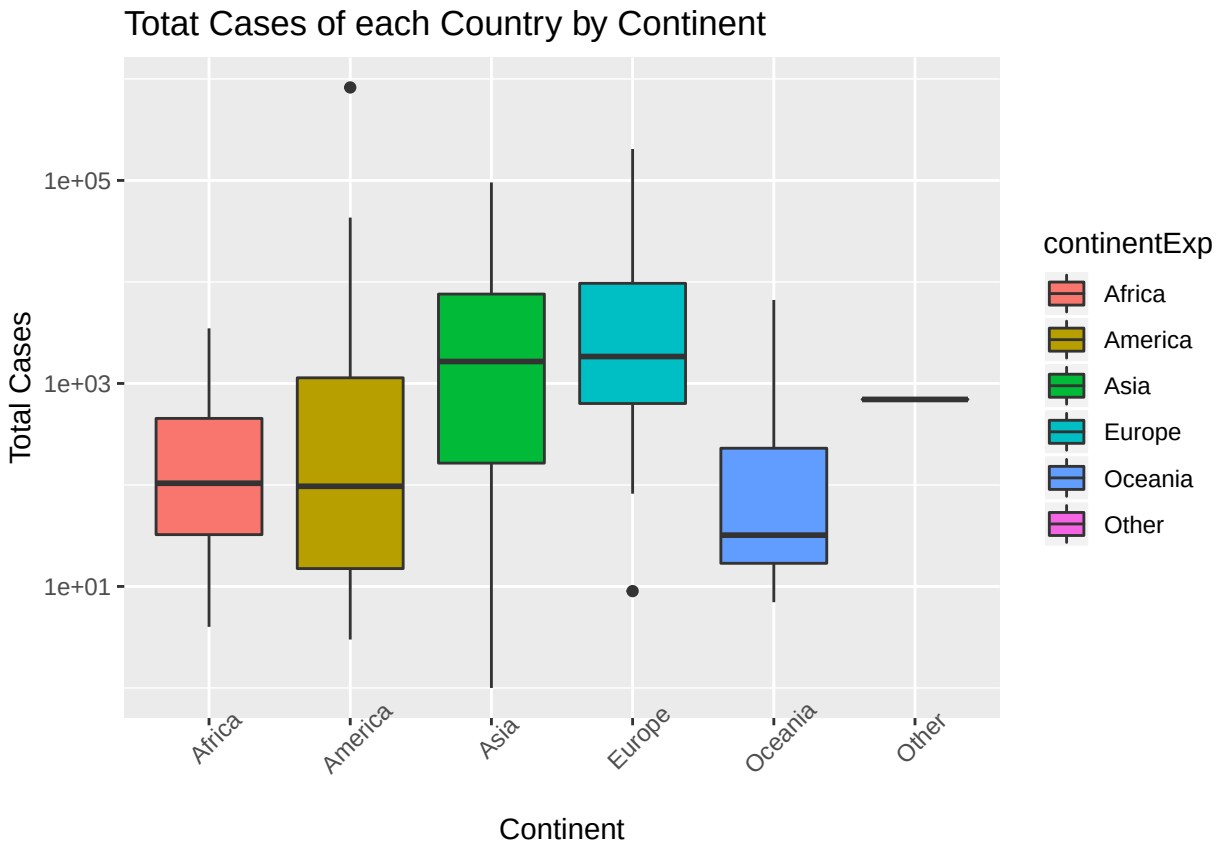


```
ggplot() +
  geom_polygon(data=map.world_joined, aes(x=long, y=lat, group=group, fill = cases_total),  color="white
  scale_fill_gradient2(low = "blue",
                       mid = scales::muted("purple"),
                       high = "red")+
   labs(title='Total Cases global distribution', label = "Total cases")
```

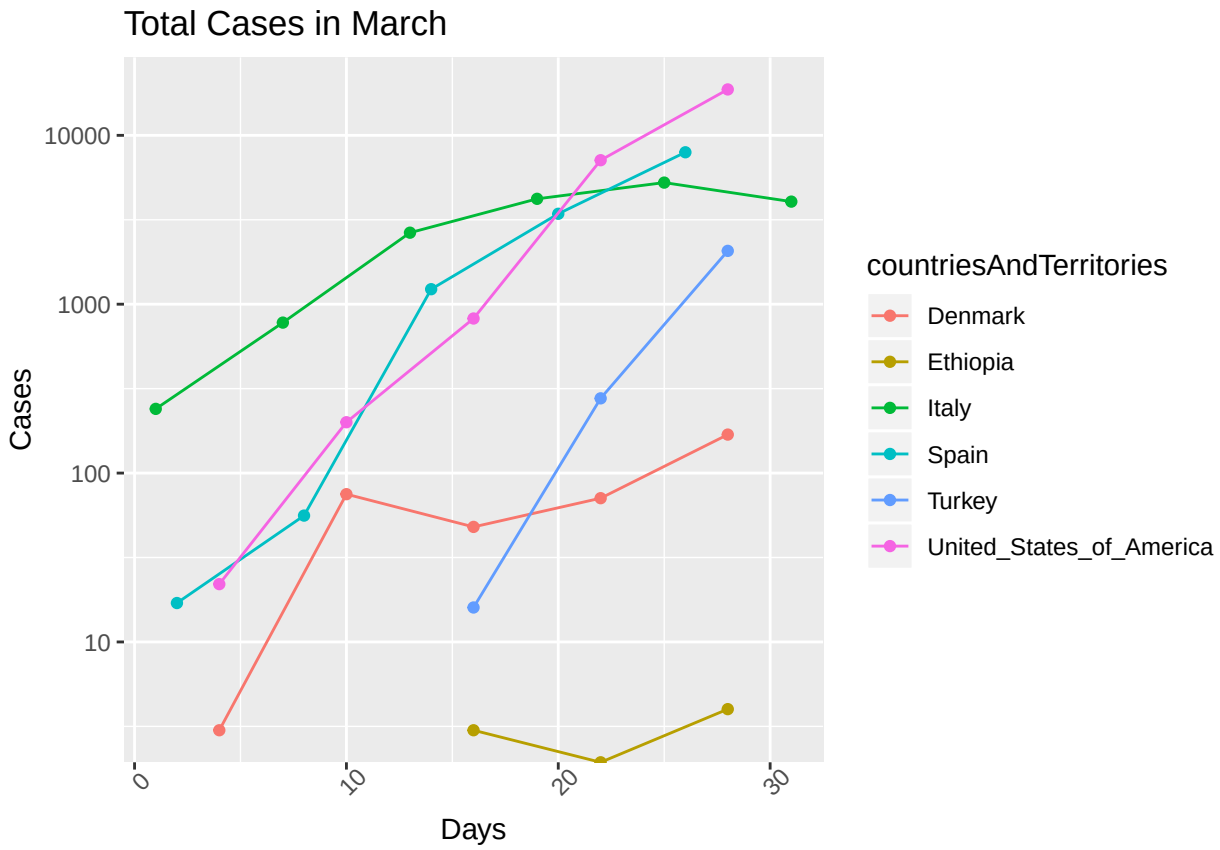


Looking at the distribution of total number of cases between continents, we want to note that this is total cases and not specific to population density. The total population of each country within these continents decide the outcome.

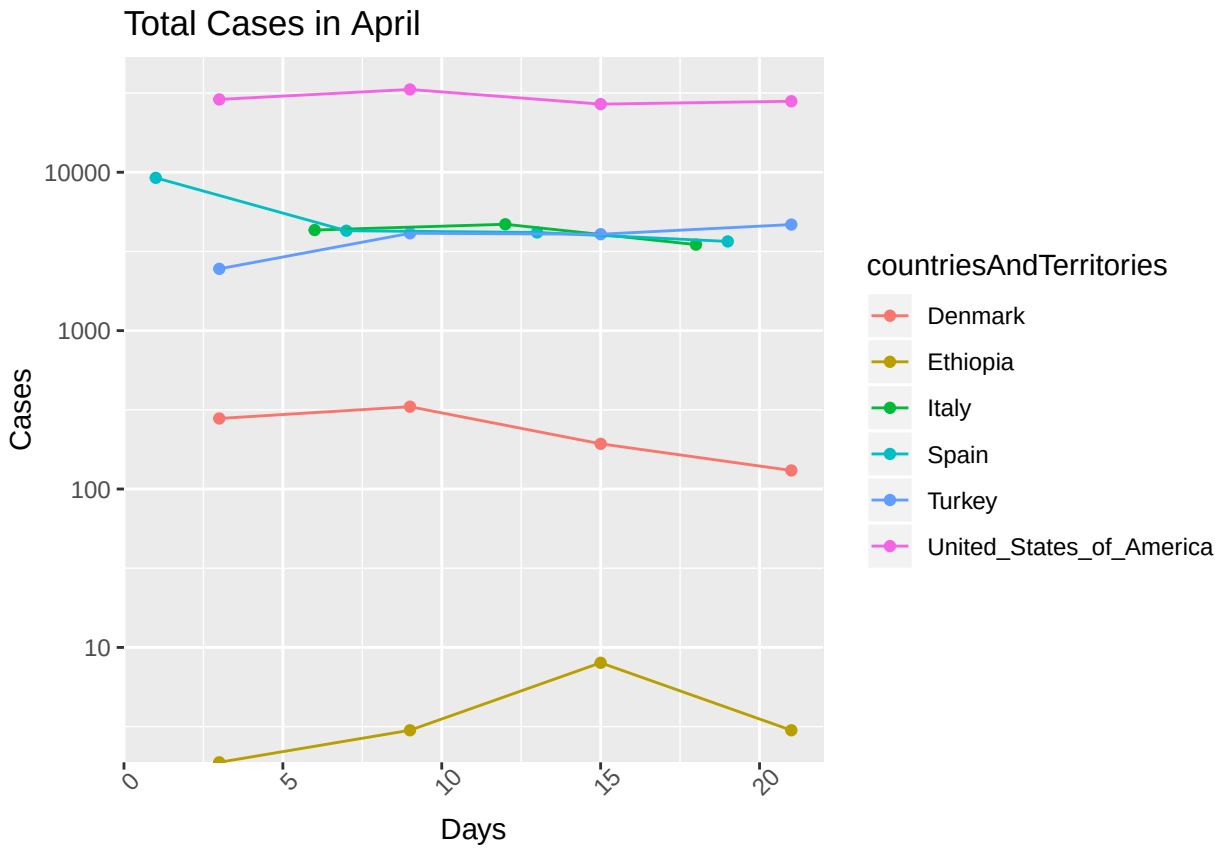
```
#Total cases for each month and every continent
COVID_TD_a <- COVID.19.geographic.disbtribution.worldwide %>%
  group_by(countriesAndTerritories, continentExp) %>%
  summarise(cases_total=as.double(sum(cases)),
            cases_mean=mean(cases),
            cases_sd=sd(cases))%>%
  ungroup()

p = ggplot(data=COVID_TD_a, aes(y=cases_total, x=continentExp, fill = continentExp ))
p + geom_boxplot() + theme(axis.text.x = element_text(angle = 45)) + scale_y_log10()+
  labs(x="Continent", y="Total Cases", title="Totat Cases of each Country by Continent")
```




```
#Does Social Distancing work?
#COVID_Specific_Countries
COVID_SC <- COVID.19.geographic.disbtribution.worldwide %>%
  filter(countriesAndTerritories==c("Italy", "Spain", "United_States_of_America", "Denmark", "Turkey",

COVID_SC$year <- year(mdy(COVID_SC$date))
COVID_SC$month <- month(mdy(COVID_SC$date))
COVID_SC$day <- day(mdy(COVID_SC$date))
COVID_SC_3 <- COVID_SC %>%
  filter(month == c(3))
COVID_SC_4 <- COVID_SC %>%
  filter(month == c(4))

p = ggplot(data=COVID_SC_3, aes(x=day, y=cases, color = countriesAndTerritories))
p + geom_point()+ geom_line() + theme(axis.text.x = element_text(angle = 45)) + scale_y_log10()+
  labs(x="Days", y="Cases", title="Total Cases in March")
```




```
p = ggplot(data=COVID_SC_4, aes(x=day, y=cases, color = countriesAndTerritories))
p + geom_point()+ geom_line() + theme(axis.text.x = element_text(angle = 45)) + scale_y_log10()+
  labs(x="Days", y="Cases", title="Total Cases in April")
```

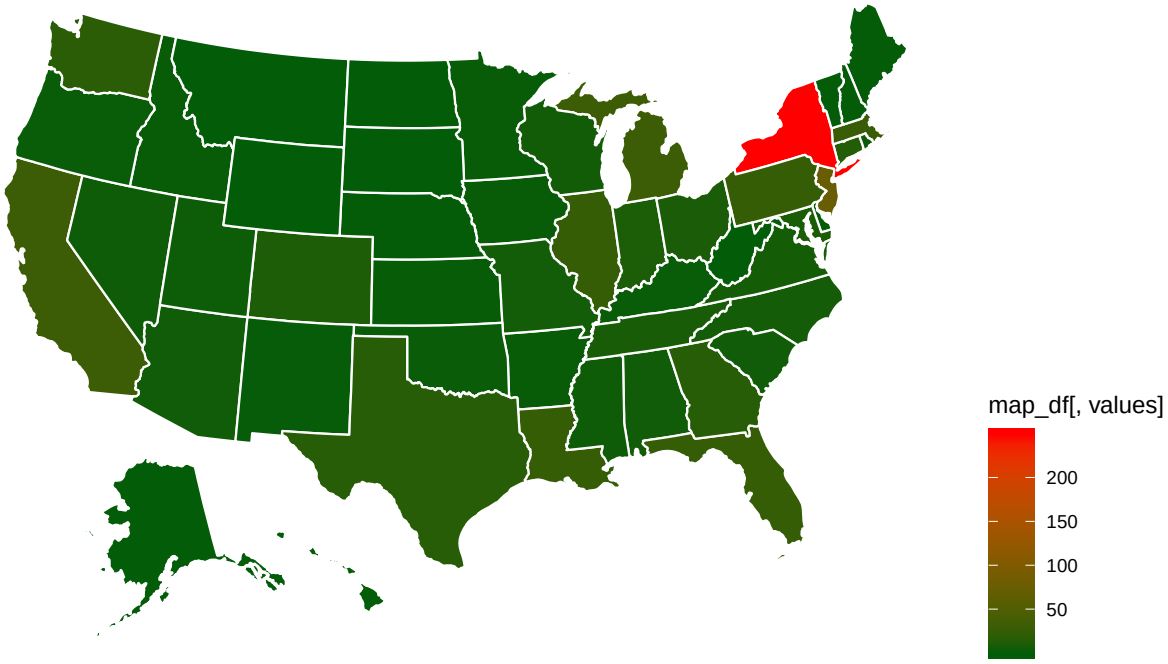


US

Specific Data Looking at total case distribution for the United States.

```
#Does Social Distancing work?
#COVID_US_Specific

usdata<-us.states%>%
   group_by(state,fips) %>%
  summarise(cases_total=(sum(cases)/10000),
            deaths_total=sum(deaths))%>%
  ungroup()

plot_usmap(data = usdata, values = "cases_total", color = "white")+
  theme(legend.position = "right")+labs(title = "Total Cases")+
  scale_fill_gradient2(low = "blue",
                       mid = scales::muted("white"),
                       high = "red")+
   labs(title='Total Cases per 10,000 in the USA', label = "Total cases per 10,000")
```

Total Cases per 10,000 in the USA




```
#library(maps)
#us_states <- map_data("state")
#map.us_joined <- left_join(us_states, us.states, by = c('region' = 'state'))
```

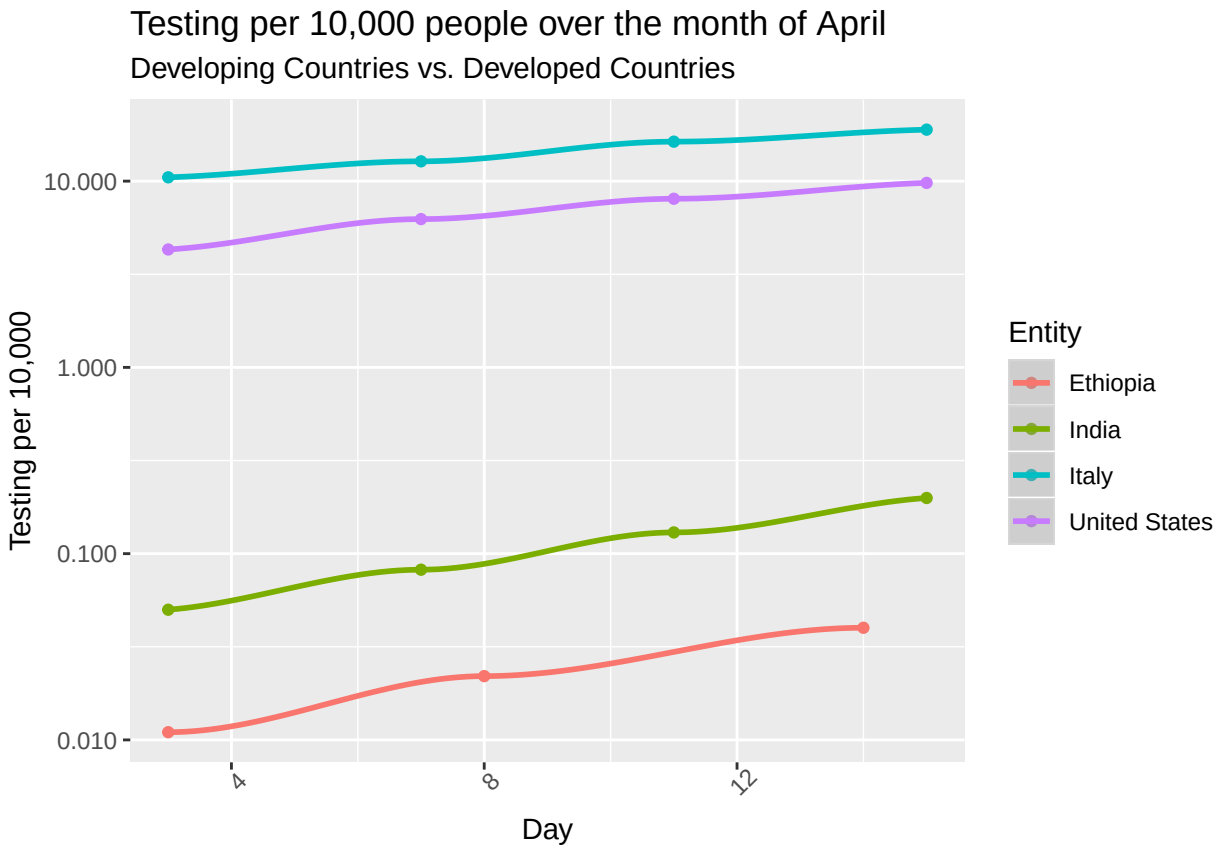
Does Social Distancing work in the US? Has there been any changes?

```
COVID_US <- us.states %>%
  filter(state==c("New York", "Minnesota", "Washington"))

COVID_US$year <- year(mdy(COVID_US$date))
COVID_US$month <- month(mdy(COVID_US$date))
COVID_US$day <- day(mdy(COVID_US$date))
COVID_US_3 <- COVID_US %>%
  filter(month == c(3))
COVID_US_4 <- COVID_US %>%
  filter(month == c(4))

p = ggplot(data=COVID_US_3, aes(x=day, y=cases, color = state))
p + geom_point()+ geom_smooth()+ theme(axis.text.x = element_text(angle = 45)) + scale_y_log10(labels =
  labs(x='Day', y='Cases per Thousand', title='March: Cases in MN, NY, WS over time per 1,000 people')
```


```
p = ggplot(data=COVID_US_4, aes(x=day, y=cases, color = state))
p + geom_point()+ geom_smooth()+ theme(axis.text.x = element_text(angle = 45)) + scale_y_log10(labels =
   labs(x='Day', y='Cases per Thousand', title='April: Cases in MN, NY, WS over time per 1,000 people')
```

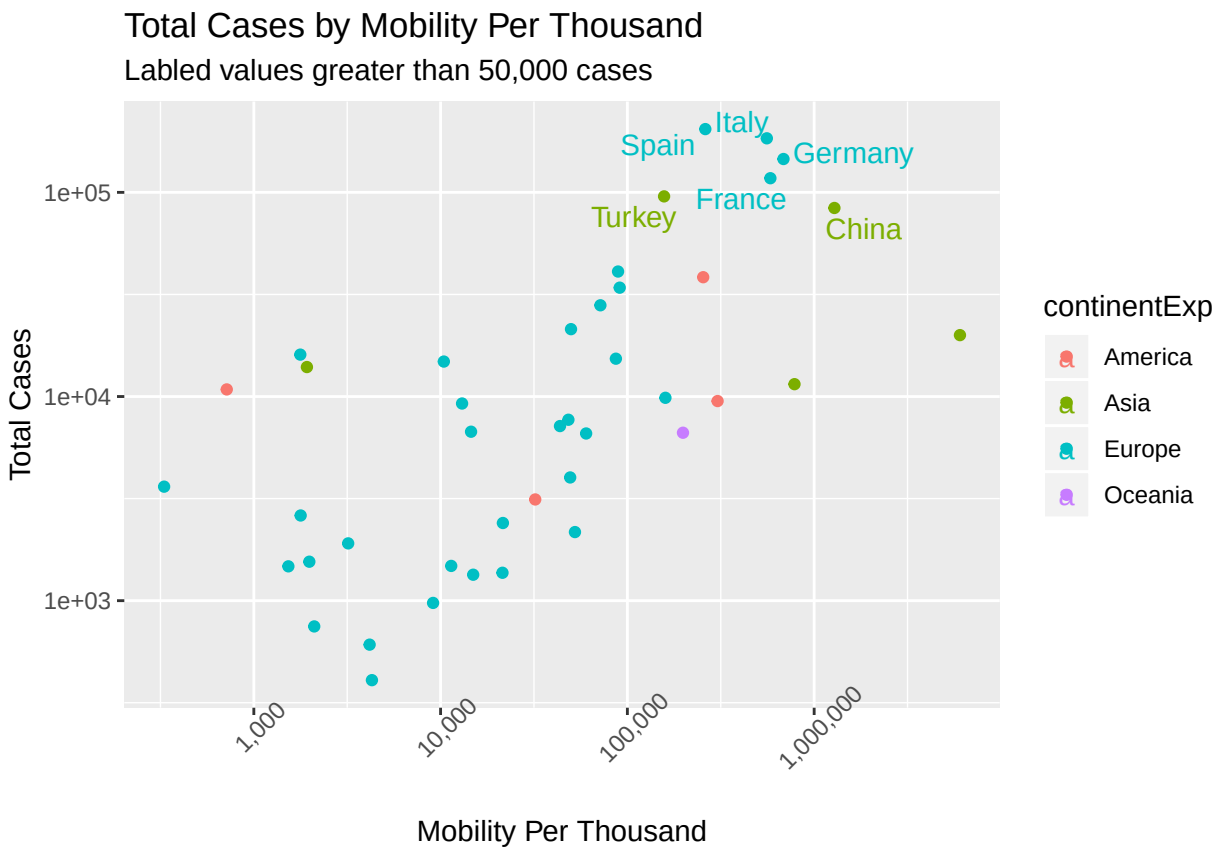
Testing Data In Developing Countries vs Developed Countries and What is the reason for the big disparity? Do developed countries have a large number of positive cases because of rampant testing.

```
#Testing rate comparison amongst countries
tests.of.covid.19.per.thousand.people.vs.gdp.per.capita$GDP.per.capita..international... <- NULL
Testing <- tests.of.covid.19.per.thousand.people.vs.gdp.per.capita
Testing_SC <- Testing %>%
  na.omit(Testing) %>%
  filter(Entity ==c("United States", "Italy", "Ethiopia", "India"))

Testing_SC$year <- year(dmy(Testing_SC$Date))
Testing_SC$month <- month(dmy(Testing_SC$Date))
Testing_SC$day <- day(dmy(Testing_SC$Date))

Testing_SC_4 <- Testing_SC %>%
  filter(month == c(4))

p = ggplot(data=Testing_SC_4, aes(x=day, y=Total.tests.per.thousand, color = Entity))
p + geom_point()+ geom_smooth()+ theme(axis.text.x = element_text(angle = 45)) + scale_y_log10(labels =
   labs(x='Day', y='Testing per 10,000', title='Testing per 10,000 people over the month of April', sub
```

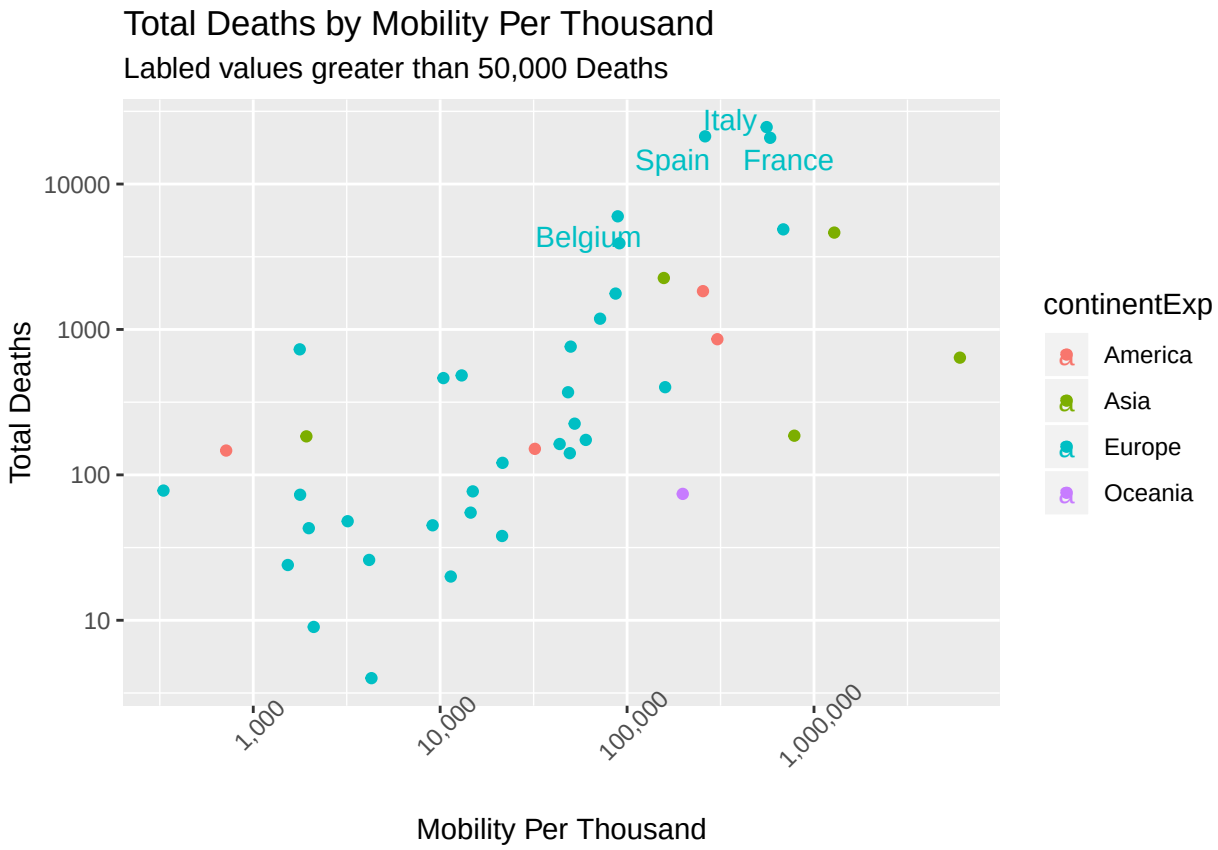


Does Movment/Mobility Affect the Spread of the Virus? Note: Most data points on footprint and mobility are calculated in europe.

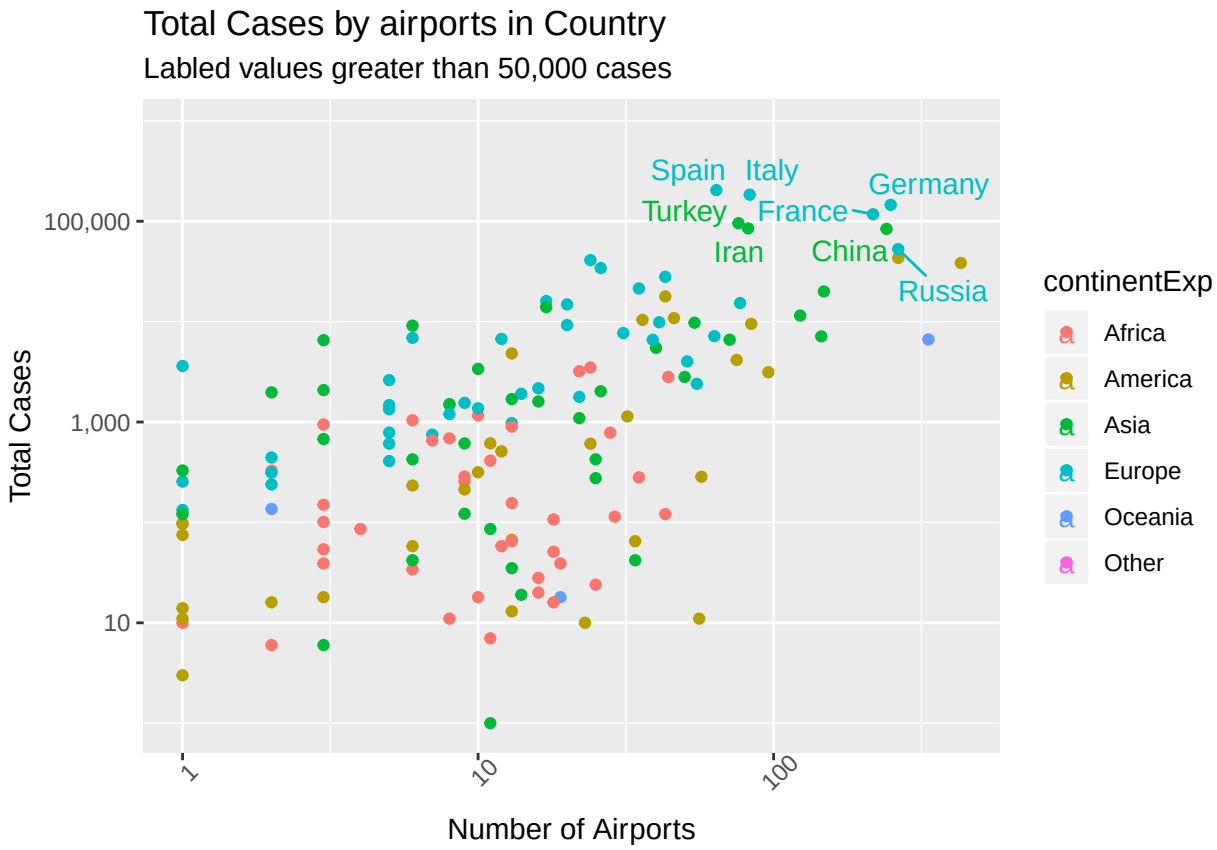
```
COVID_19tot <- COVID.19.geographic.disbtribution.worldwide %>%
  group_by(countriesAndTerritories, continentExp) %>%
  summarise(cases_total=sum(cases),
            deaths_total =sum(deaths))%>%
  ungroup()%>%
  rename(Country = countriesAndTerritories)
#COVID_19tot
Mobility <- Passanger.Report %>%
  group_by(Country) %>%
  summarise(value_mean=mean(Value))%>%
  ungroup()
#Mobility
MobilityvsCOVID_19tot <- left_join(COVID_19tot, Mobility, by = c("Country" = "Country"))
MobilityVSCases <- MobilityvsCOVID_19tot %>%
  na.omit(value_mean)

p = ggplot(data=MobilityVSCases, aes(x=value_mean, y=cases_total, color = continentExp))
p + geom_point()+ theme(axis.text.x = element_text(angle = 45)) + scale_x_log10(labels = scales::comma)+
  labs(x="Mobility Per Thousand", y="Total Cases", title="Total Cases by Mobility Per Thousand", subtitl
```




```
p = ggplot(data=MobilityVSCases, aes(x=value_mean, y=deaths_total, color = continentExp))
p + geom_point()+ theme(axis.text.x = element_text(angle = 45)) + scale_x_log10(labels = scales::comma)+
  labs(x="Mobility Per Thousand", y="Total Deaths", title="Total Deaths by Mobility Per Thousand", subti
```

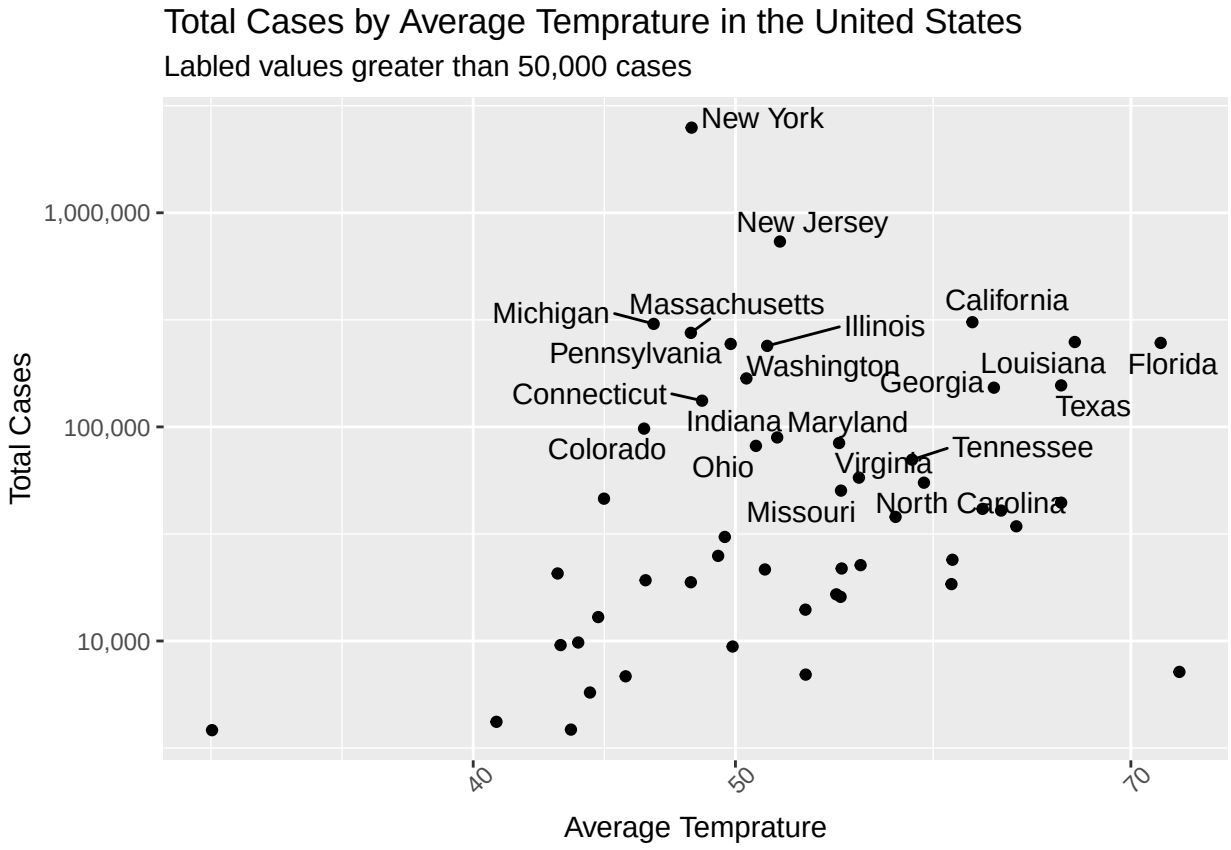


Is the number of Total Airports related to the spread of the virus? Do countries get more testing kits because of large airports.



Does Temprature Affect the Spread of the Virus in specifically the USA?

```
usdata<-us.states%>%
   group_by(state,fips) %>%
  summarise(cases_total=(sum(cases)),
            deaths_total=sum(deaths))%>%
  ungroup()
temp.us_join <- left_join(temprature_us, usdata, by = c('State' = 'state'))

p = ggplot(data=temp.us_join, aes(x=AvgTmp, y=cases_total))
p + geom_point()+ theme(axis.text.x = element_text(angle = 45)) + scale_x_log10() + scale_y_log10(labels
  labs(x="Average Temprature", y="Total Cases", title="Total Cases by Average Temprature in the United S
```


Total Cases by Average Temprature in the United States
Labled values greater than 50,000 cases
New York
New Jersey
Michigan
Massachusetts
California
Illinois
Pennsylvania
Washington
Louisiana
Florida
Georgia
Connecticut
Texas
Indiana
Maryland
Colorado
Tennessee
Ohio
Virginia
Missouri
North Carolina
1,000,000
100,000
10,000
Total Cases
40
50
70
Average Temprature