COVID-19 Analysis

Matthew Helke

5/14/2020

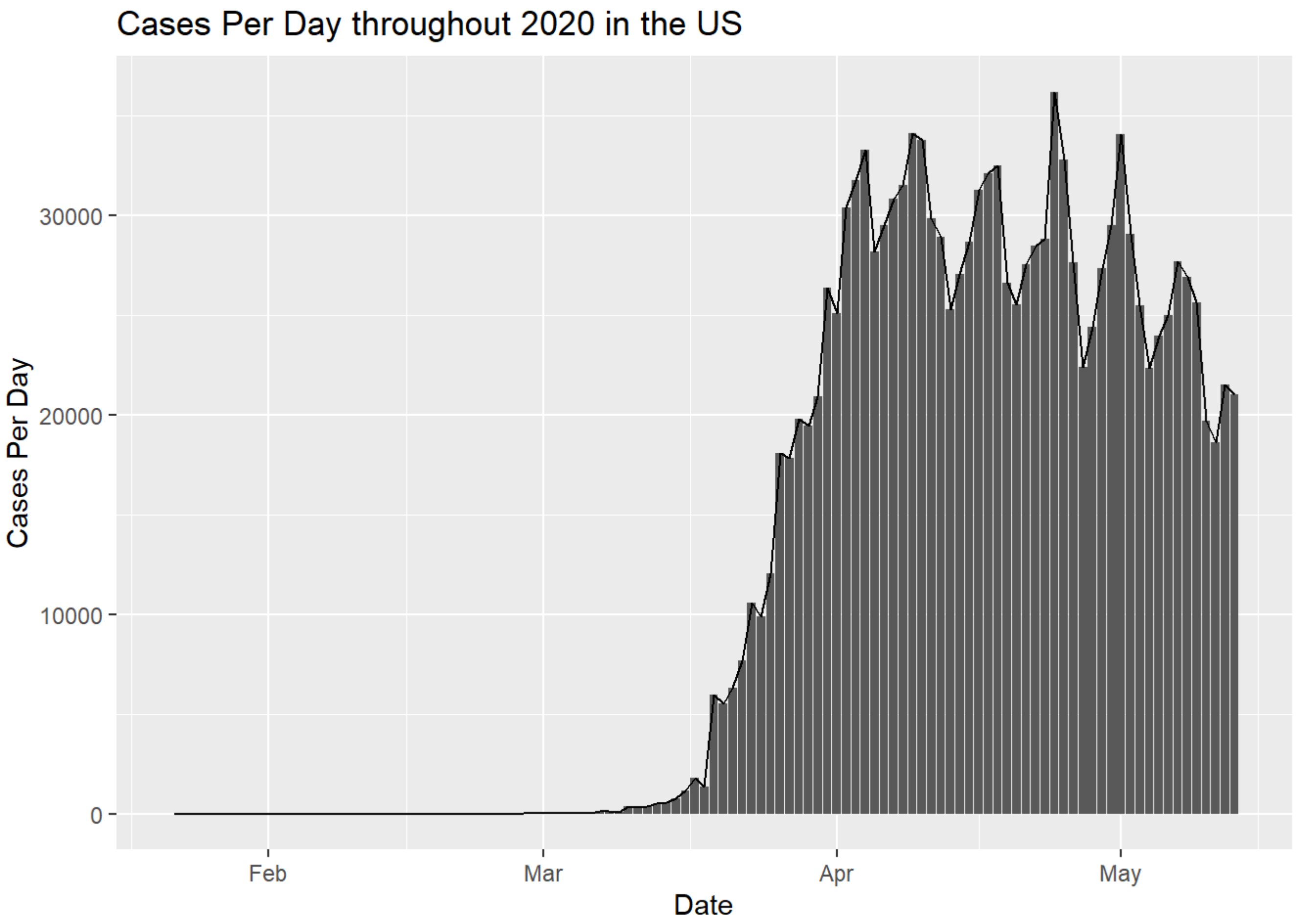
The goal of this project is to examine trends of COVID-19 across the world, and the US.

Analysis of Global Cases

(Data from John Hopkins Univeristy Cener for System Science and Engineering)

Here is a chart of report cases in the US per day

```
global_confirmed %>%
  rename(province = "Province/State", country_region = "Country/Region") %>%
  pivot_longer(-c(province, country_region, Lat, Long), names_to = "Date",
               values_to = "cumulative_cases") %>%
  mutate(Date = mdy(Date)) %>%
  filter(country_region == "US") %>%
  arrange(province, Date) %>%
  group_by(province) %>%
  mutate(cases_per_day = c(0, diff(cumulative_cases))) %>%
  ungroup() %>% select(-c(country_region, Lat, Long, cumulative_cases)) %>%
  ggplot(aes(x = Date, y = cases_per_day)) +
  geom_col() +
  geom_line() +
  xlab("Date") +
  ylab("Cases Per Day") +
  ggtitle("Cases Per Day throughout 2020 in the US")
```



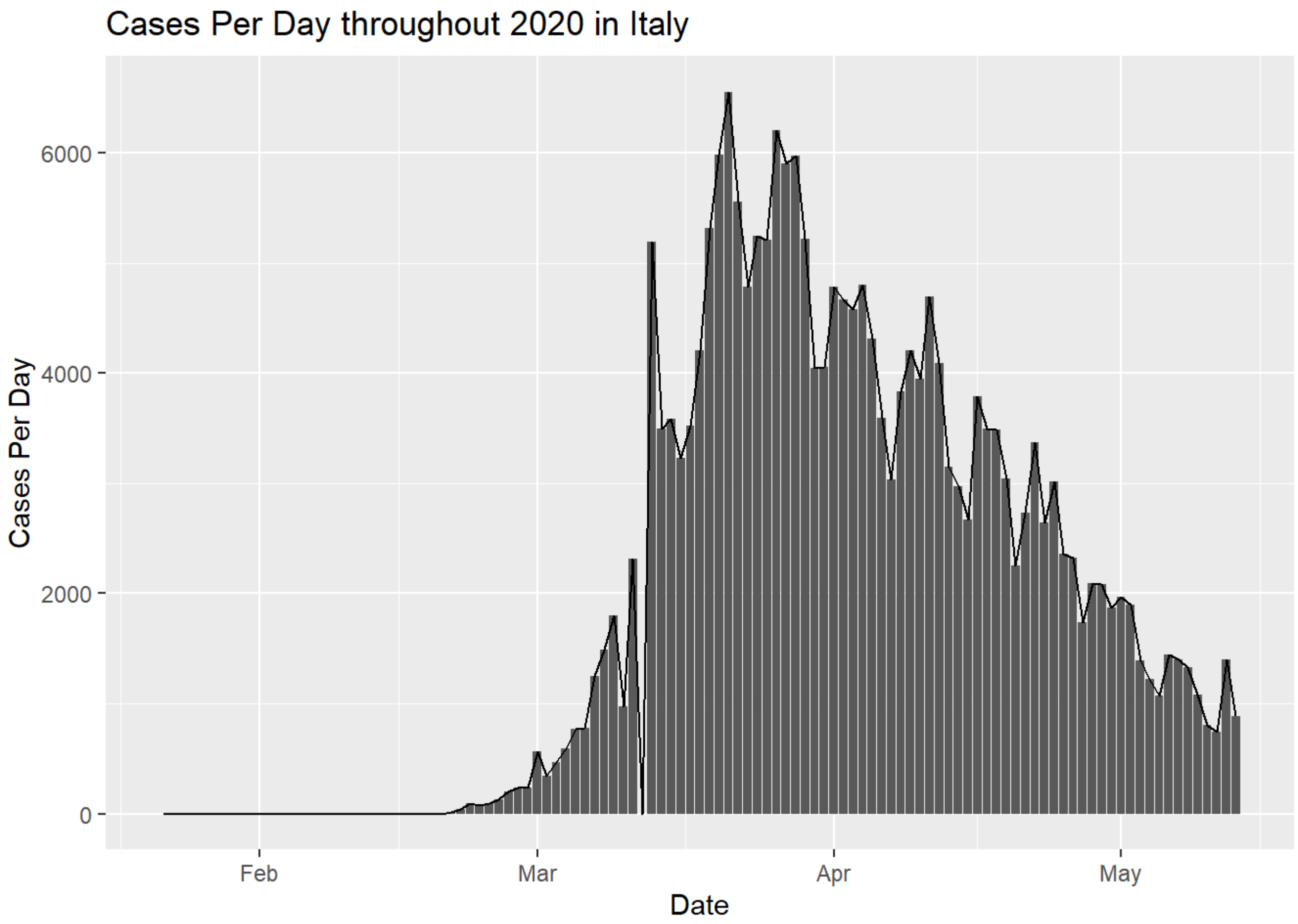
The graph shows that reported cases have been averaging between 20,000 and 35,000 perday since late March. This graph also shows that the spread of the Novel Coronavirus is steady and fast in the US

We can examine what China’s graph looks like too.

It appears that China has been controlling the spread well since early March, but there have been concerns over whether their reporting is trustworthy and accurate.

Italy was a country that peaked in terms of deaths from COVID-19, but looking at cases you can see the spread is not as large as a country like the US.

```
global_confirmed %>%
  rename(province = "Province/State", country_region = "Country/Region") %>%
  pivot_longer(-c(province, country_region, Lat, Long), names_to = "Date",
               values_to = "cumulative_cases") %>%
  mutate(Date = mdy(Date)) %>%
  filter(country_region == "Italy") %>%
  arrange(province, Date) %>%
  group_by(province) %>%
  mutate(cases_per_day = c(0, diff(cumulative_cases))) %>%
  ungroup() %>% select(-c(country_region, Lat, Long, cumulative_cases)) %>%
  ggplot(aes(x = Date, y = cases_per_day)) +
  geom_col() +
  geom_line() +
  xlab("Date") +
  ylab("Cases Per Day") +
  ggtitle("Cases Per Day throughout 2020 in Italy")
```



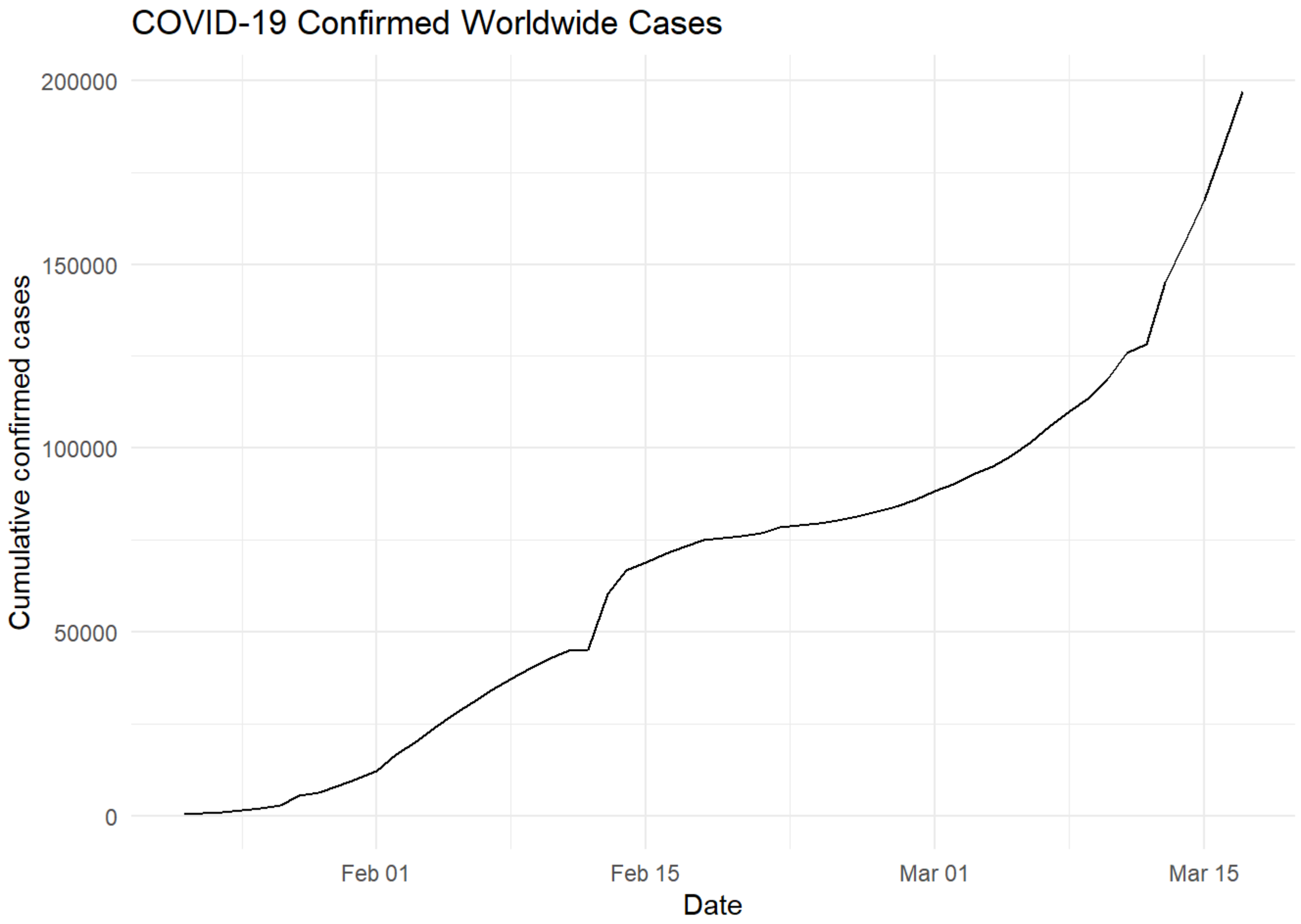
The US has the most cases per day out of many other countries. This trend seems to be constant for the past couple months.

(Data from John Hopkins Univeristy Cener for System Science and Engineering, WHO, and CDC)

Trends of COVID-19

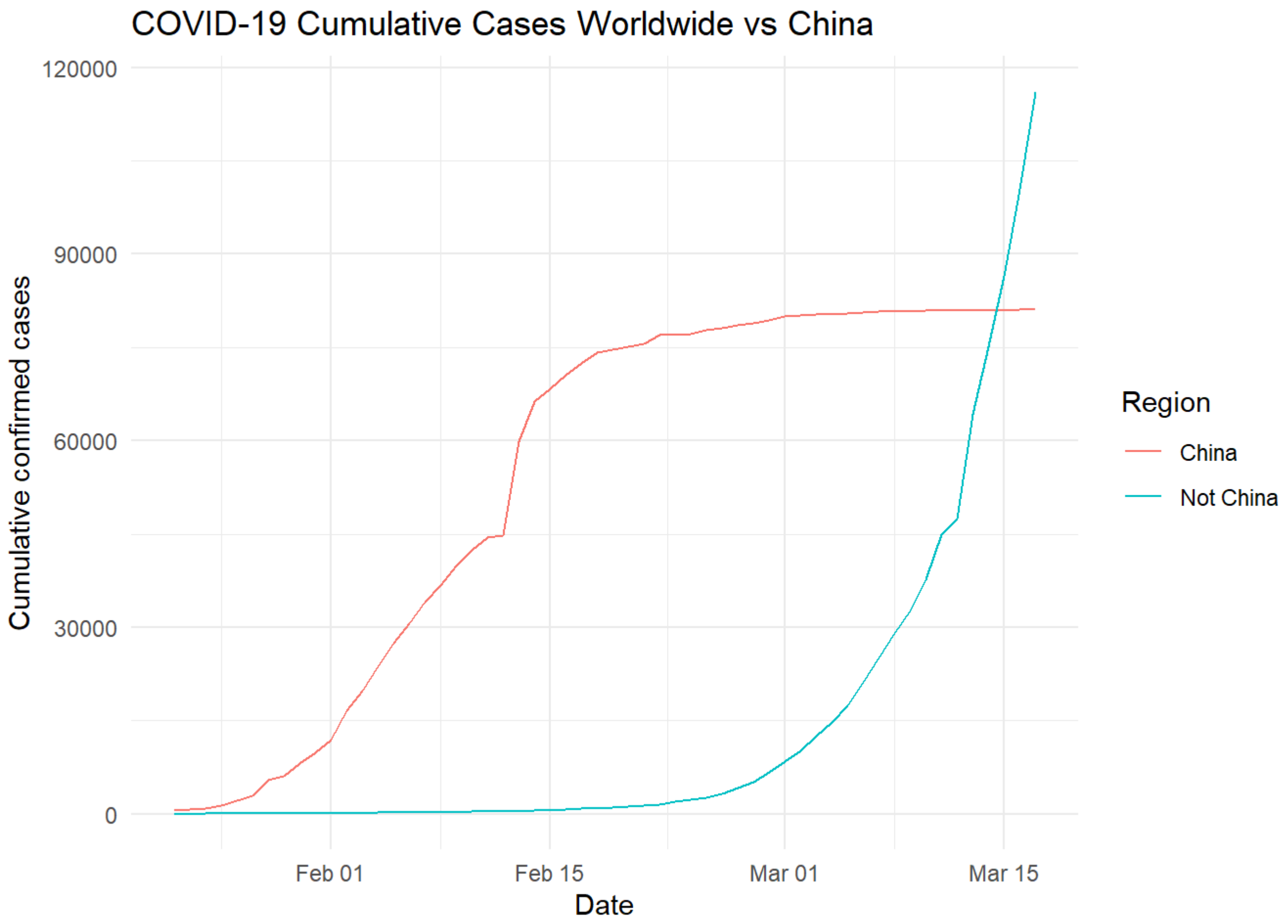
COVID-19 is spreading rapidly. This graph shows the cumulativereported cases gloablly. This data shows the earlier stages of COVID-19, through mid March

```
confirmed_cases_worldwide %>%
ggplot(aes(x = date, y = cum_cases)) +
  geom_line() +
  theme_minimal() +
  ylab("Cumulative confirmed cases") +
  xlab("Date") +
  ggtitle("COVID-19 Confirmed Worldwide Cases")
```



How does the world look when compared to china where COVID-19 originated from?

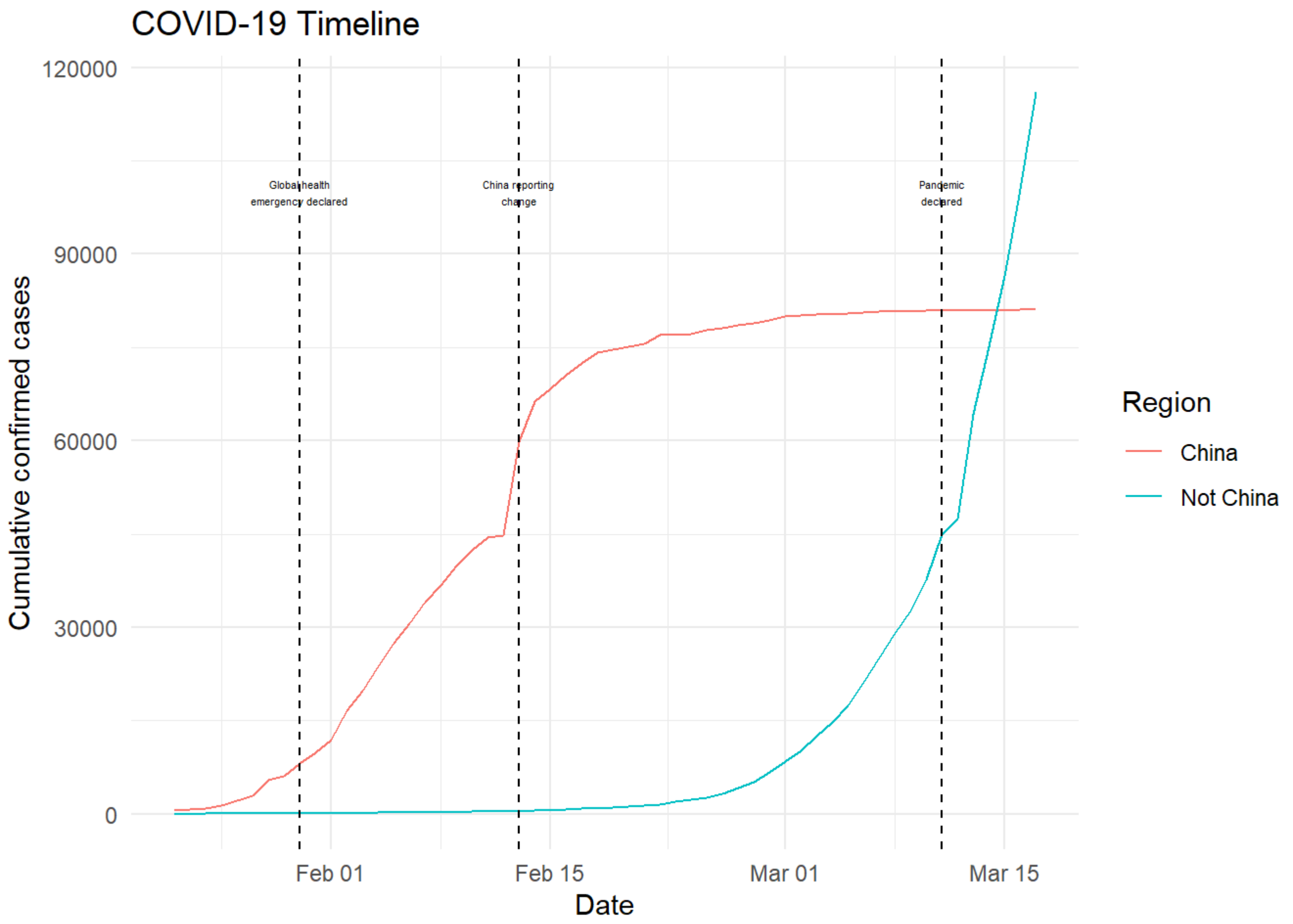
```
confirmed_cases_china_vs_world %>%
  rename(Region = is_china) %>%
  ggplot() +
  geom_line(aes(x = date, y = cum_cases, group = Region, color = Region)) +
  theme_minimal() +
  ylab("Cumulative confirmed cases") +
  xlab("Date") +
  ggtitle("COVID-19 Cumulative Cases Worldwide vs China")
```



Looking at the early trends of the virus we can see it took off outside of China by early March.

Here is a look at the timeline of how things went

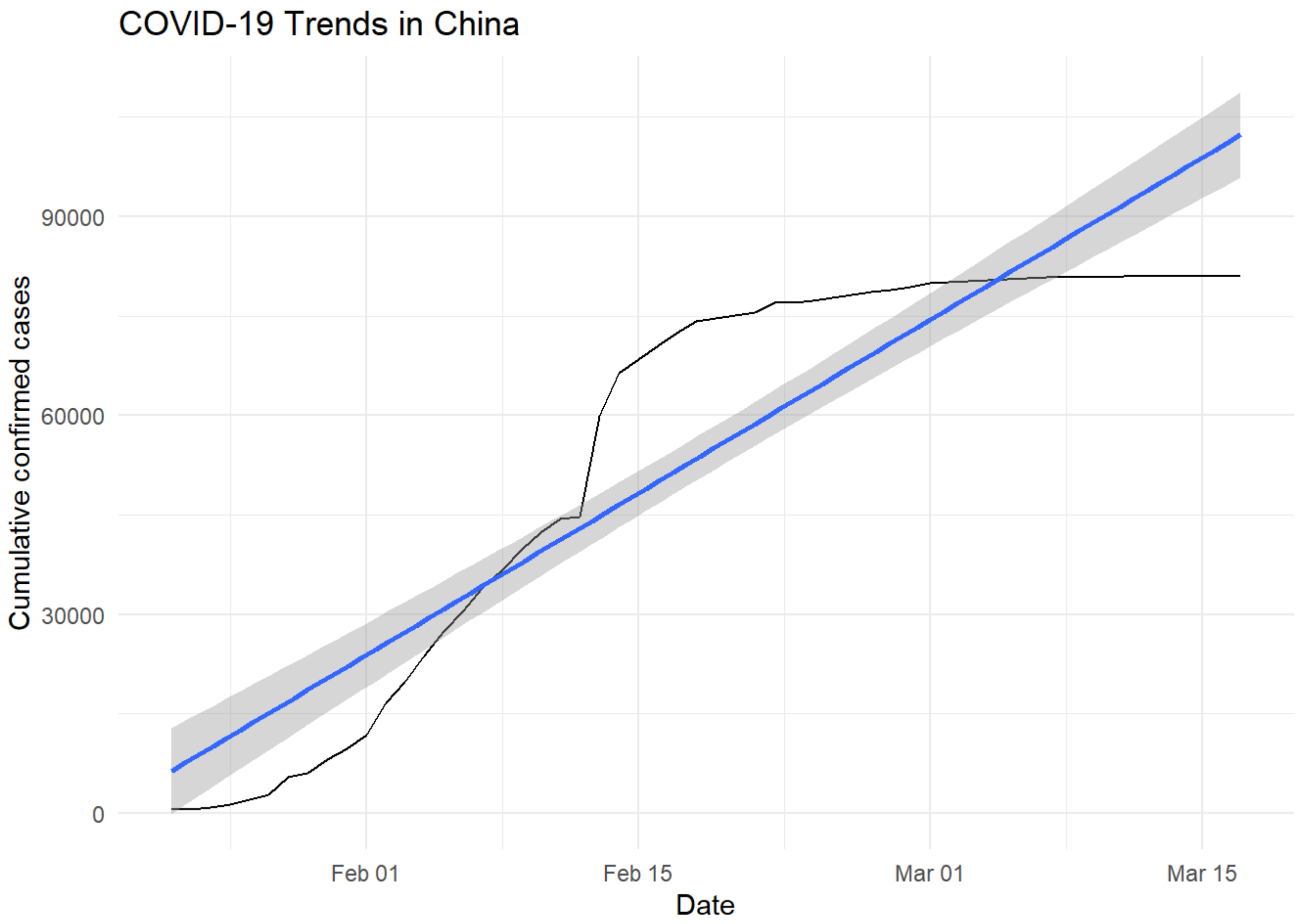
```
WHO_reports <- tribble(
  ~ date, ~ event,
  "2020-01-30", "Global health\nemergency declared",
  "2020-03-11", "Pandemic\ndeclared",
  "2020-02-13", "China reporting\nchange"
) %>%
  mutate(date = as.Date(date))

confirmed_cases_china_vs_world %>%
  rename(Region = is_china) %>%
  ggplot() +
  geom_line(aes(x = date, y = cum_cases, group = Region, color = Region)) +
  geom_vline(aes(xintercept = date), data = WHO_reports, linetype = "dashed") +
  geom_text(aes(x = date, label = event), data = WHO_reports, y = 1e5, lwd = 1.5) +
  theme_minimal() +
  ylab("Cumulative confirmed cases") +
  xlab("Date") +
  ggtitle("COVID-19 Timeline")
```



The pandemic was declared after cases hit roughly 45,000. At that point China was starting to flatten its curve.

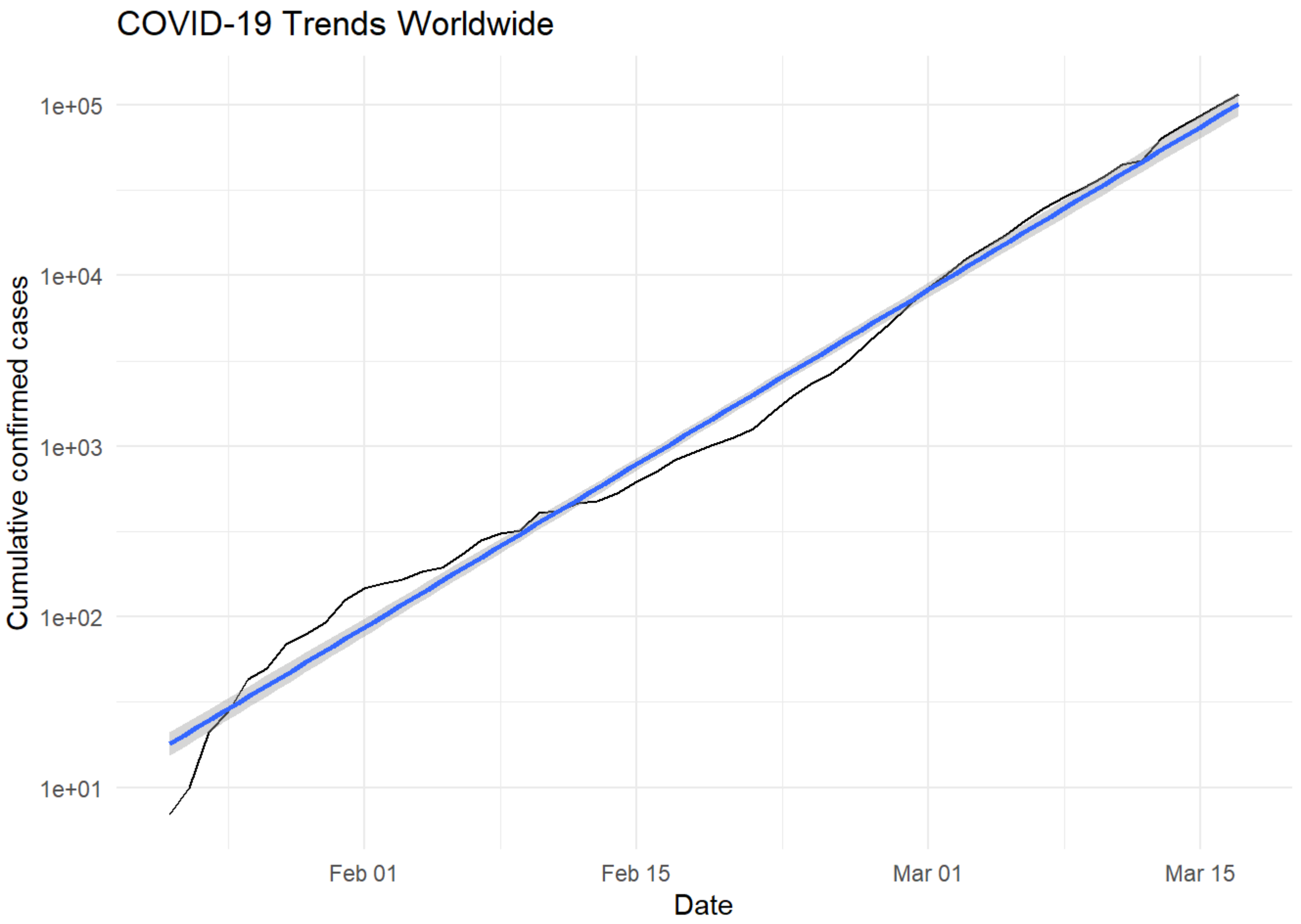
Here is a closer look at the early trends from China. They started slowing down with new reports in March.

```
confirmed_cases_china_vs_world %>%
  filter(is_china == "China") %>%
  ggplot(aes(date, cum_cases)) +
  geom_line() +
  geom_smooth(method = "lm", se = TRUE) +
  theme_minimal() +
  ylab("Cumulative confirmed cases") +
  xlab("Date") +
  ggtitle("COVID-19 Trends in China")
```



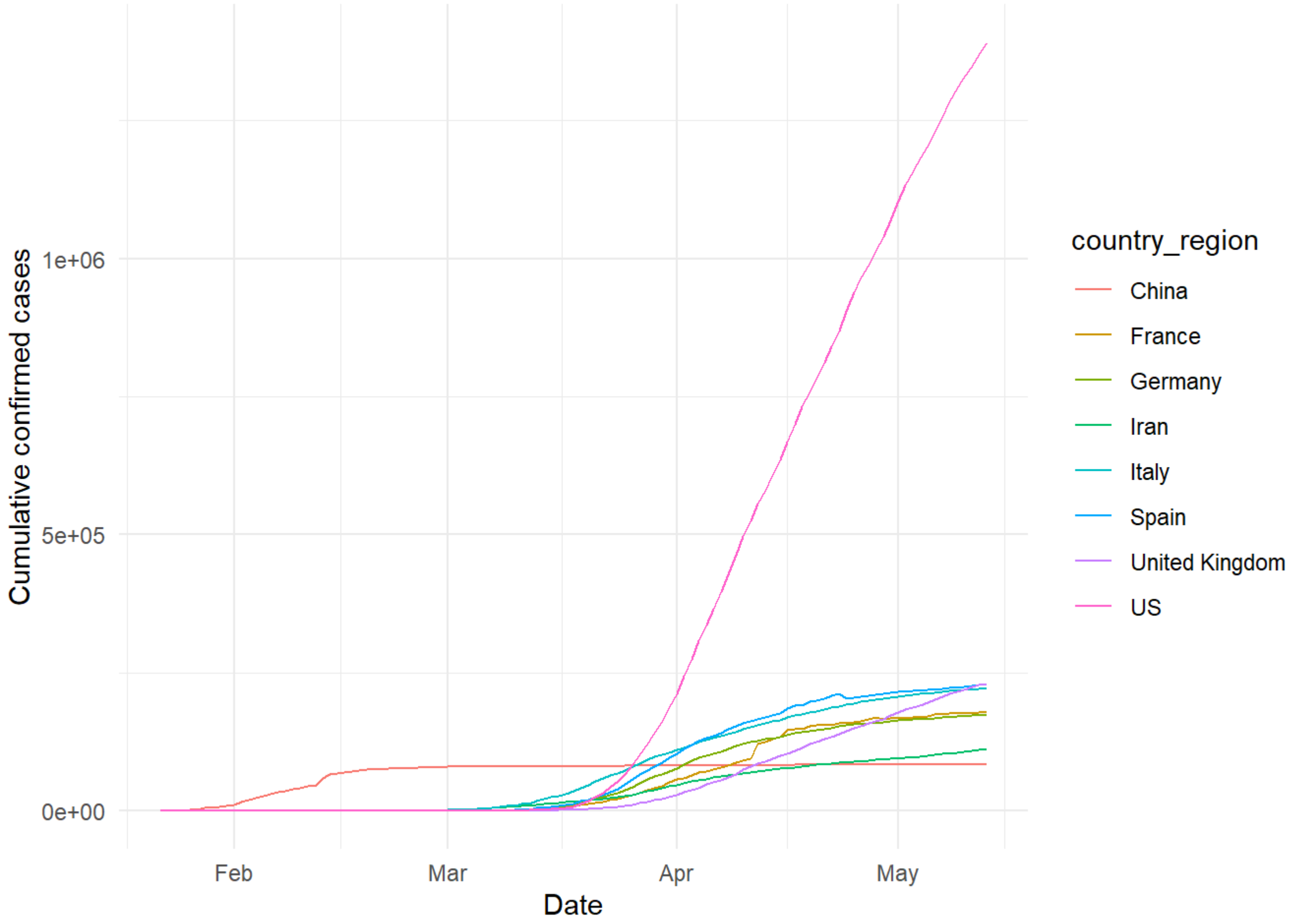
Here is what the trends are looking like outside of China.

```
confirmed_cases_china_vs_world %>%
  filter(is_china == "Not China") %>%
  ggplot(aes(date, cum_cases)) +
  geom_line() +
  geom_smooth(method = "lm", se = TRUE) +
  theme_minimal() +
  ylab("Cumulative confirmed cases") +
  xlab("Date") +
  ggtitle("COVID-19 Trends Worldwide") +
  scale_y_log10()
```



The projections only show it getting worse. This graph confirms that new cases are still on the rise and the US is far above other countries when it comes to total reported cases.

```
global_confirmed %>%
  rename(province = "Province/State", country_region = "Country/Region") %>%
  pivot_longer(-c(province, country_region, Lat, Long), names_to = "Date",
               values_to = "cumulative_cases") %>%
  mutate(Date = mdy(Date)) %>%
  arrange(Date) %>%
  filter(country_region == "US" |
         country_region == "China" |
         country_region == "United Kingdom" |
         country_region == "Italy" |
         country_region == "France" |
         country_region == "South Korea" |
         country_region == "Germany" |
         country_region == "Iran" |
         country_region == "Spain") %>%
  group_by(country_region, Date) %>%
  summarise(cum_cases = sum(cumulative_cases))%>%
  ungroup %>%
  ggplot(aes(Date, cum_cases, color = country_region , group = country_region)) +
  geom_line() +
  theme_minimal() +
  ylab("Cumulative confirmed cases") +
  xlab("Date")
```



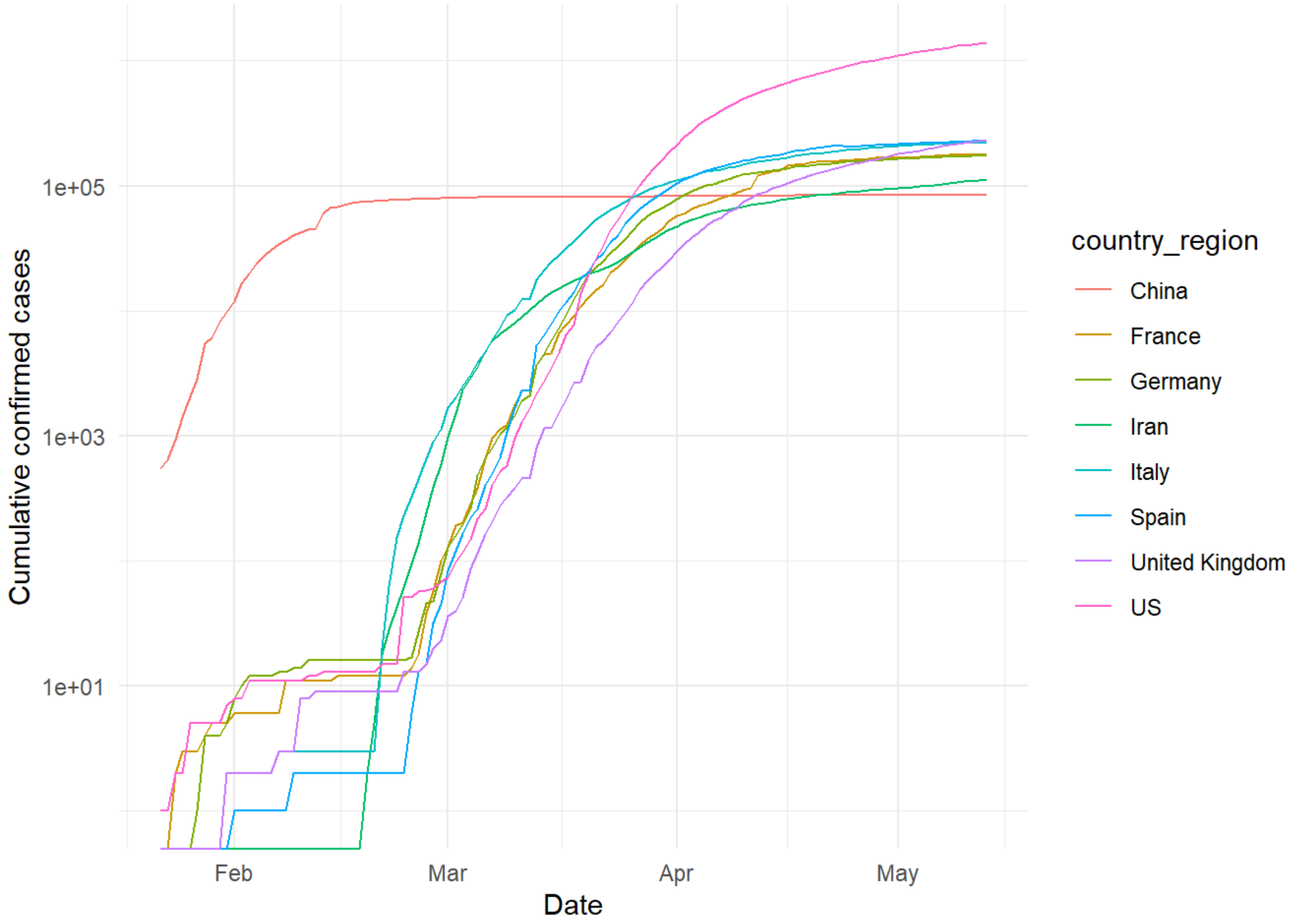
Here are the top 7 countries with the most cases.

```
top_countries_by_total_cases <- global_confirmed %>%
  rename(province = "Province/State", country_region = "Country/Region") %>%
  pivot_longer(-c(province, country_region, Lat, Long), names_to = "Date",
               values_to = "cumulative_cases") %>%
  mutate(Date = mdy(Date)) %>%
  arrange(Date) %>%
  group_by(country_region) %>%
  summarize(total_cases = max(cumulative_cases)) %>%
  top_n(7, total_cases) %>%
  arrange(desc(total_cases))

top_countries_by_total_cases
```

```
## # A tibble: 7 x 2
##   country_region total_cases
##   <chr>                <dbl>
## 1 US                 1390406
## 2 Russia              242271
## 3 United Kingdom      229705
## 4 Spain               228691
## 5 Italy               222104
## 6 Brazil              190137
## 7 France              176207
```

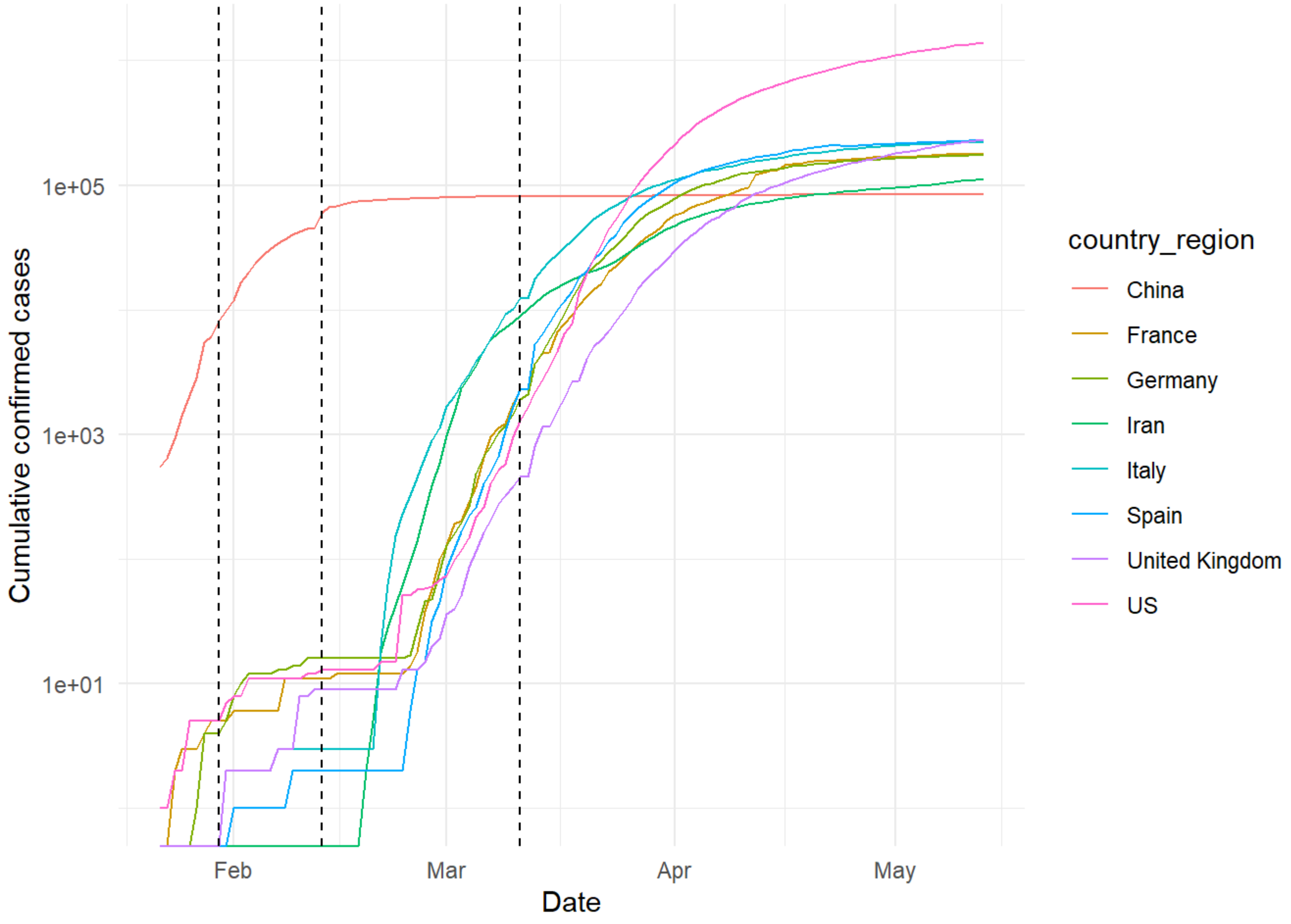
This is a popular graph the media is showing using a logarithmic scale. It appears the curve is flattening, but in reality this graph shows that the virus is spreading exponentially which means it is actually spreading faster than this graph makes it look. Because it is a logarithmic scale, the “flattening” of the curve does not mean things are getting better. Cases are still rising at a fast pace. The last graph shown, is actually this same graph WITHOUT the scale! Notice the difference? The US in particular in the graph below looks to have a flattening curve, but if you look at the previous graph, the US is still rising sharply.

```
global_confirmed %>%
  rename(province = "Province/State", country_region = "Country/Region") %>%
  pivot_longer(-c(province, country_region, Lat, Long), names_to = "Date",
               values_to = "cumulative_cases") %>%
  mutate(Date = mdy(Date)) %>%
  arrange(Date) %>%
  filter(country_region == "US" |
         country_region == "China" |
         country_region == "United Kingdom" |
         country_region == "Italy" |
         country_region == "France" |
         country_region == "South Korea" |
         country_region == "Germany" |
         country_region == "Iran" |
         country_region == "Spain") %>%
  group_by(country_region, Date) %>%
  summarise(cum_cases = sum(cumulative_cases))%>%
  ungroup %>%
  ggplot(aes(Date, cum_cases, color = country_region, group = country_region)) +
  geom_line() +
  theme_minimal() +
  ylab("Cumulative confirmed cases") +
  xlab("Date") +
  scale_y_log10()
```



Here is the same graph, with the time line included. The dashed line on the right shows when WHO formally declared the coronavirus outbreak to be a pandemic.

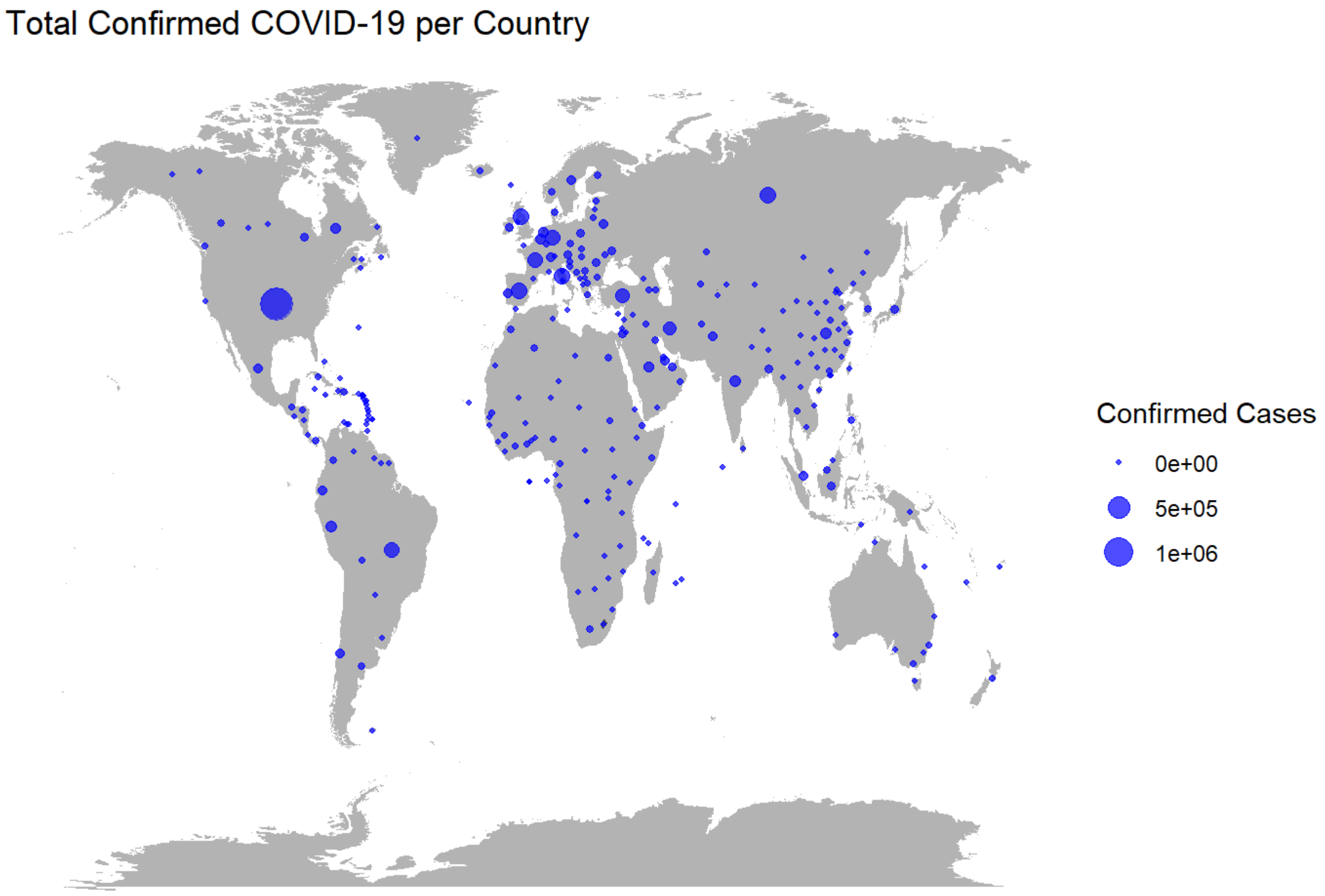
```
global_confirmed %>%
  rename(province = "Province/State", country_region = "Country/Region") %>%
  pivot_longer(-c(province, country_region, Lat, Long), names_to = "Date",
               values_to = "cumulative_cases") %>%
  mutate(Date = mdy(Date)) %>%
  arrange(Date) %>%
  filter(country_region == "US" |
         country_region == "China" |
         country_region == "United Kingdom" |
         country_region == "Italy" |
         country_region == "France" |
         country_region == "South Korea" |
         country_region == "Germany" |
         country_region == "Iran" |
         country_region == "Spain") %>%
  group_by(country_region, Date) %>%
  summarise(cum_cases = sum(cumulative_cases))%>%
  ungroup %>%
  ggplot(aes(Date, cum_cases, color = country_region, group = country_region)) +
  geom_line() +
  theme_minimal() +
  scale_y_log10() +
  ylab("Cumulative confirmed cases") +
  xlab("Date") +
  geom_vline(aes(xintercept = date), data = WHO_reports, linetype = "dashed")
```



World Maps of COVID-19 Cases

This map gives you an idea of where the coronavirus is. This is a rough visual before looking at more descriptive maps. It gives you a good idea of the data before digging in deeper.

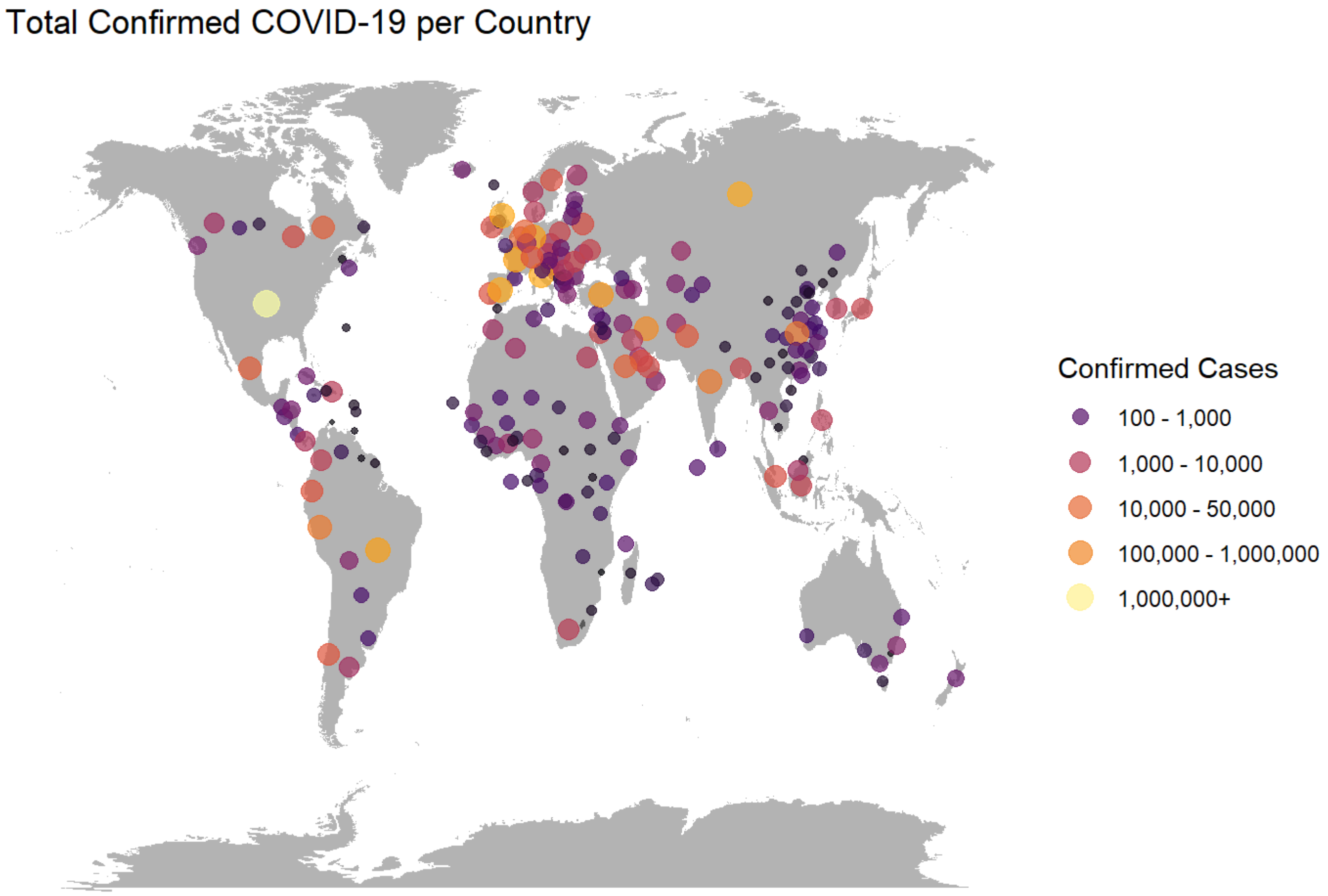
```
ggplot() +
  geom_polygon(data = world_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = global_confirmed_ts, aes(Long, Lat, size = `5/12/20`),
             stroke=F, alpha = 0.7, color = "blue") +
  theme_void() +
  guides(size = guide_legend()) +
  ggtitle("Total Confirmed COVID-19 per Country") +
  labs(size = "Confirmed Cases")
```



Looking at the data, there are 72 regions with under 100 cases. These places have been minimally affected so I will not be including them on the maps as a way of reducing clutter.

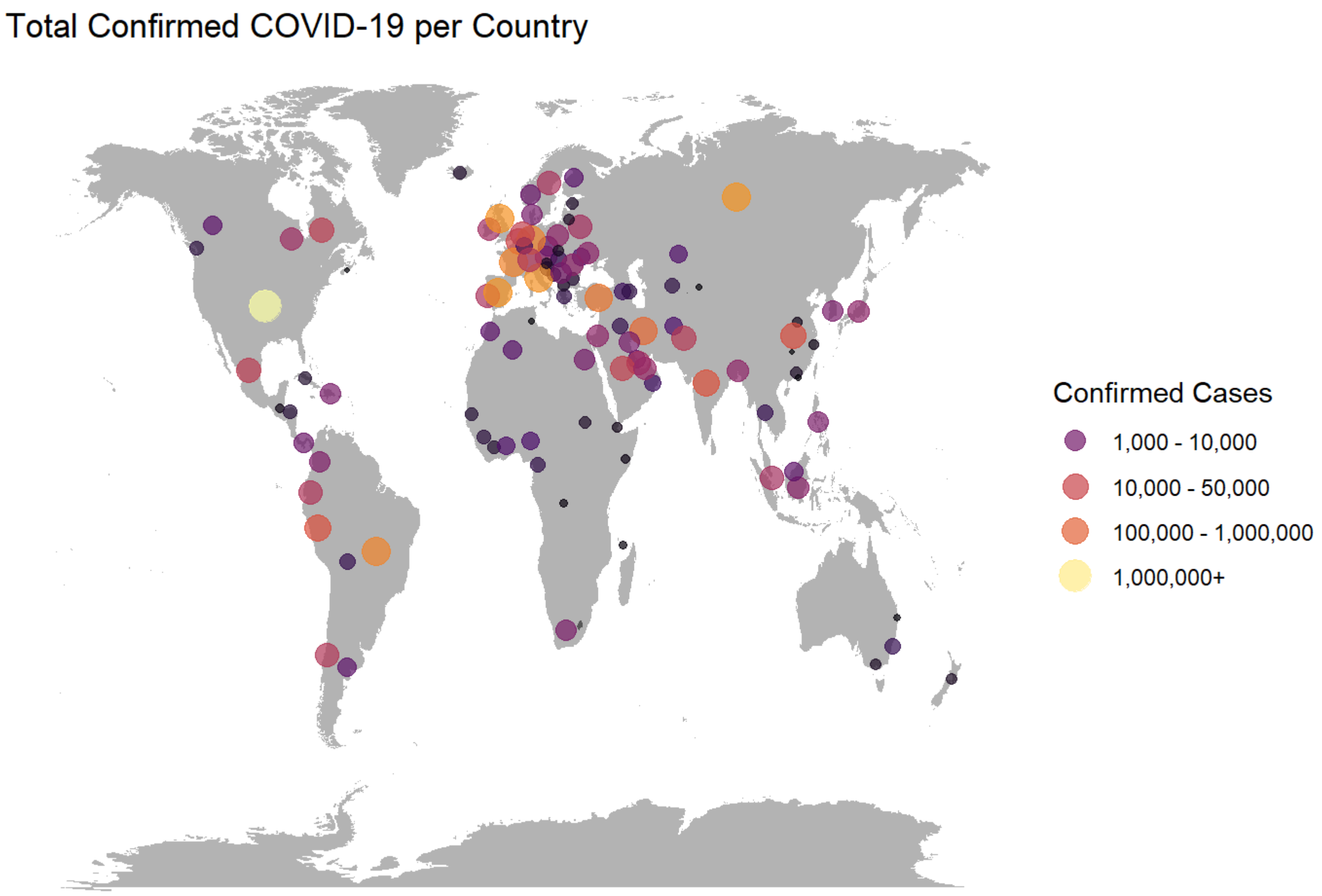
Here are the countries with over 100 cases. The points are given colors and sizes based on how many cases have been reported in the country.

```
breaks = c(1000,10000,50000,100000,1000000)

ggplot() +
  geom_polygon(data = world_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = global_confirmed_ts_over_100, aes(Long, Lat, size = `5/12/20`, color = `5/1
2/20`),
             stroke=F, alpha = 0.7) +
  theme_void() +
  guides(color = guide_legend()) +
  ggtitle("Total Confirmed COVID-19 per Country") +
  labs(color = "Confirmed Cases") +
  scale_size_continuous(name = "Confirmed Cases", trans = "log", range = c(1,5), breaks = break
s,
                        labels = c("100 - 1,000", "1,000 - 10,000",
                                   "10,000 - 50,000", "100,000 - 1,000,000",
                                   "1,000,000+")) +
  scale_color_viridis_c(name = "Confirmed Cases", option = "inferno", trans = "log",
                        breaks = breaks, labels =
                          c("100 - 1,000", "1,000 - 10,000",
                            "10,000 - 50,000", "100,000 - 1,000,000",
                            "1,000,000+"))
```



In order to reduce even more clutter, this next map shows countries that have over 1,000 cases. Compared to many places in the world having 1,000 cases seems like nothing. By doing this you can see many African, Asian, and Caribbean countries have been removed. These are more remote places in the world with far less global traffic compared to places like the US, China, etc.

```
global_confirmed_ts_over_1000 <- global_confirmed_ts %>%
  filter(`5/12/20` >= 1000)

breaks = c(10000,50000,100000,1000000)

ggplot() +
  geom_polygon(data = world_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = global_confirmed_ts_over_1000, aes(Long, Lat, size = `5/12/20`, color = `5/1
2/20`),
             stroke=F, alpha = 0.7) +
  theme_void() +
  guides(color = guide_legend()) +
  ggtitle("Total Confirmed COVID-19 per Country") +
  labs(color = "Confirmed Cases") +
  scale_size_continuous(name = "Confirmed Cases", trans = "log", range = c(1,6), breaks = break
s,
                        labels = c("1,000 - 10,000",
                                   "10,000 - 50,000", "100,000 - 1,000,000",
                                   "1,000,000+")) +
  scale_color_viridis_c(name = "Confirmed Cases", option = "inferno", trans = "log",
                        breaks = breaks, labels =
                          c("1,000 - 10,000",
                            "10,000 - 50,000", "100,000 - 1,000,000",
                            "1,000,000+"))
```

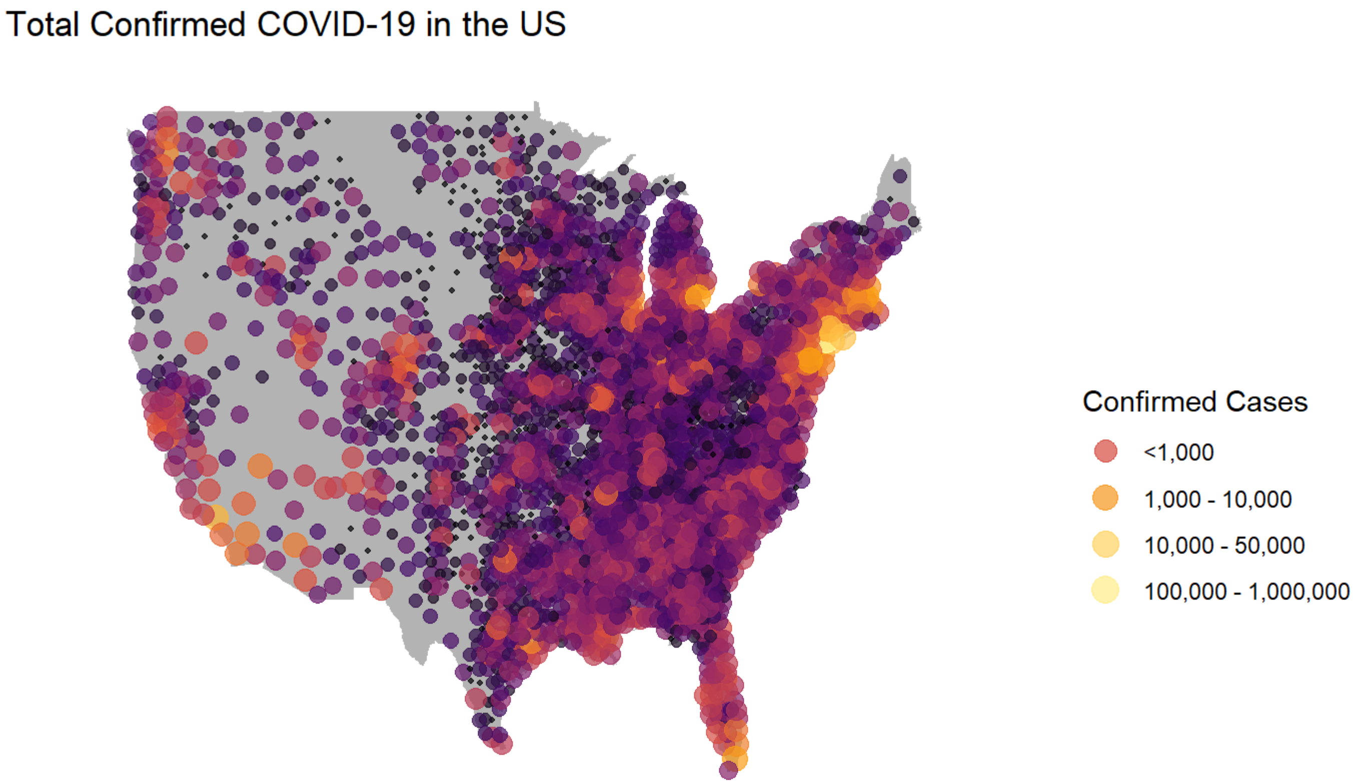


US Maps of COVID-19 Cases

I am deciding to do only the mainland US in order to keep it simple.

Here is the case data in the US per county (Minimum 100 cases):

```
us_confirmed_main <- us_confirmed %>%
  filter(iso2  == "US")  %>%
  filter(Province_State != "Alaska") %>%
  filter(Province_State != "Hawaii")

US_map <- map_data("usa")

breaks = c(1000,10000,50000,100000,1000000)

ggplot() +
  geom_polygon(data = US_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = us_confirmed_main, aes(Long_, Lat, size = `4/30/20`, color = `4/30/20`),
             stroke=F, alpha = 0.7) +
  theme_void() +
  guides(color = guide_legend()) +
  ggtitle("Total Confirmed COVID-19 in the US") +
  labs(color = "Confirmed Cases") +
  scale_size_continuous(name = "Confirmed Cases", trans = "log", range = c(1,5), breaks = break
s,
                        labels = c("<1,000", "1,000 - 10,000",
                                   "10,000 - 50,000", "100,000 - 1,000,000",
                                   "1,000,000+")) +
  scale_color_viridis_c(name = "Confirmed Cases", option = "inferno", trans = "log",
                        breaks = breaks, labels =
                          c("<1,000", "1,000 - 10,000",
                            "10,000 - 50,000", "100,000 - 1,000,000",
                            "1,000,000+")) +
  ylim(20, 50) +
  xlim(-130, -60)
```



This is a very messy map since there are so many counties with over 100 cases. You can tell where the vast majority of the population is though just by looking at this map. Every major populated area of the US is reporting over 100 cases. For a better look, we can view the total cases by state.

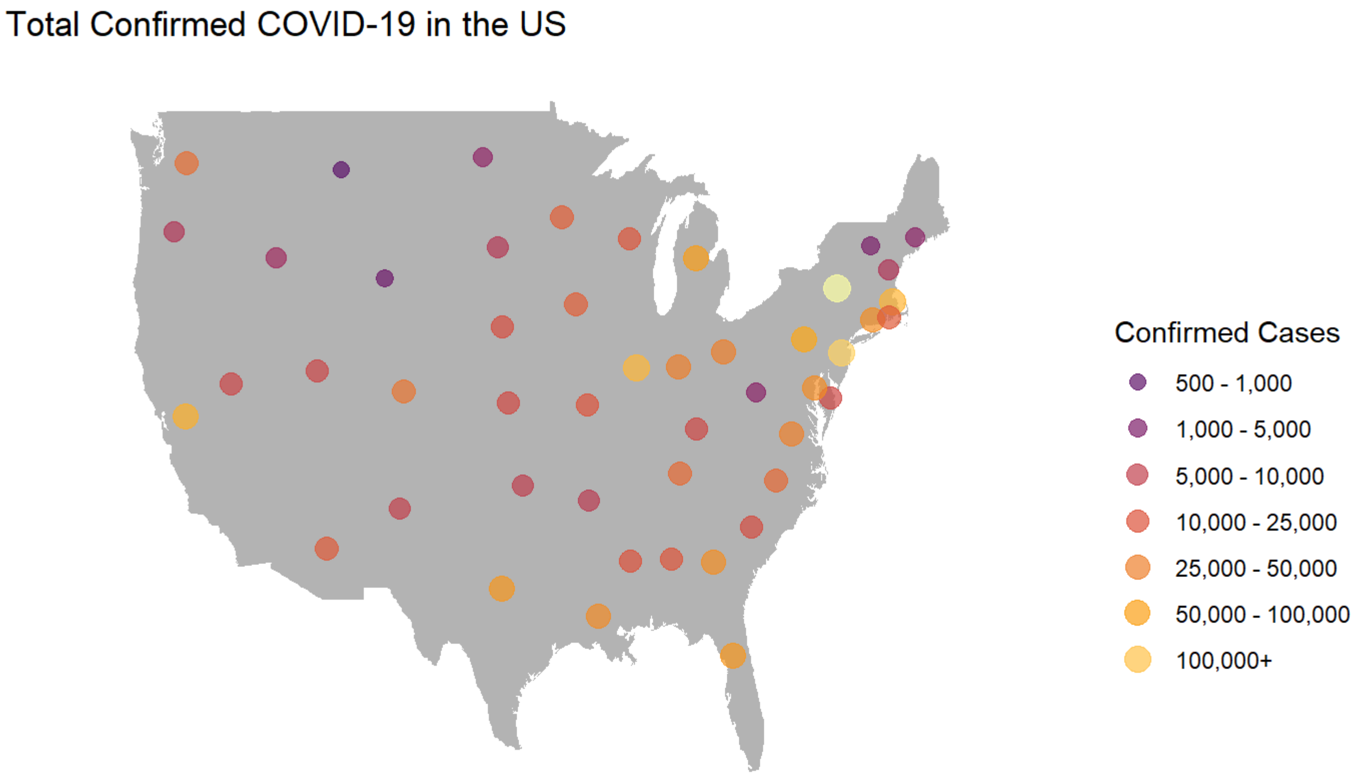
Here is a look at the number of total cases in each mainland state.

```
us_confirmed_main %>%
  select(Province_State, Lat, Long_, `5/12/20`) %>%
  group_by(Province_State) %>%
  summarise(total_cases = sum(`5/12/20`)) %>%
  ungroup() %>%
  filter(Province_State != "Diamond Princess") %>%
  filter(Province_State != "Grand Princess") %>%
  arrange(desc(total_cases))
```

```
## # A tibble: 49 x 2
##    Province_State total_cases
##    <chr>                <dbl>
##  1 New York            338485
##  2 New Jersey          140917
##  3 Illinois             83021
##  4 Massachusetts        79332
##  5 California           70978
##  6 Pennsylvania         61310
##  7 Michigan             48021
##  8 Florida              41923
##  9 Texas                41432
## 10 Georgia              34924
## # ... with 39 more rows
```

From this data, we can create a map.

```
us_confirmed_main_by_state <- us_confirmed_main %>%
  select(Province_State, Lat, Long_, `5/12/20`) %>%
  group_by(Province_State) %>%
  summarise(cases = sum(`5/12/20`), lat = median(Lat), long = median(Long_))

US_map <- map_data("usa")

breaks = c(500, 1000, 5000, 10000, 25000, 50000, 100000)


ggplot() +
  geom_polygon(data = US_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = us_confirmed_main_by_state, aes(long, lat, size = cases   , color = cases
 ),
             stroke=F, alpha = 0.7) +
  theme_void() +
  guides(color = guide_legend()) +
  ggtitle("Total Confirmed COVID-19 in the US") +
  labs(color = "Confirmed Cases") +
  scale_size_continuous(name = "Confirmed Cases", trans = "log", range = c(1,5), breaks = break
s,
                        labels = c("500 - 1,000",
                                   "1,000 - 5,000", "5,000 - 10,000",
                                   "10,000 - 25,000", "25,000 - 50,000", "50,000 - 100,000", "10
0,000+"))+
  scale_color_viridis_c(name = "Confirmed Cases", option = "inferno", trans = "log",
                        breaks = breaks, labels =
                          c("500 - 1,000",
                            "1,000 - 5,000", "5,000 - 10,000",
                            "10,000 - 25,000", "25,000 - 50,000", "50,000 - 100,000", "100,000+"
)) +
  ylim(20, 50) +
  xlim(-130, -60)
```



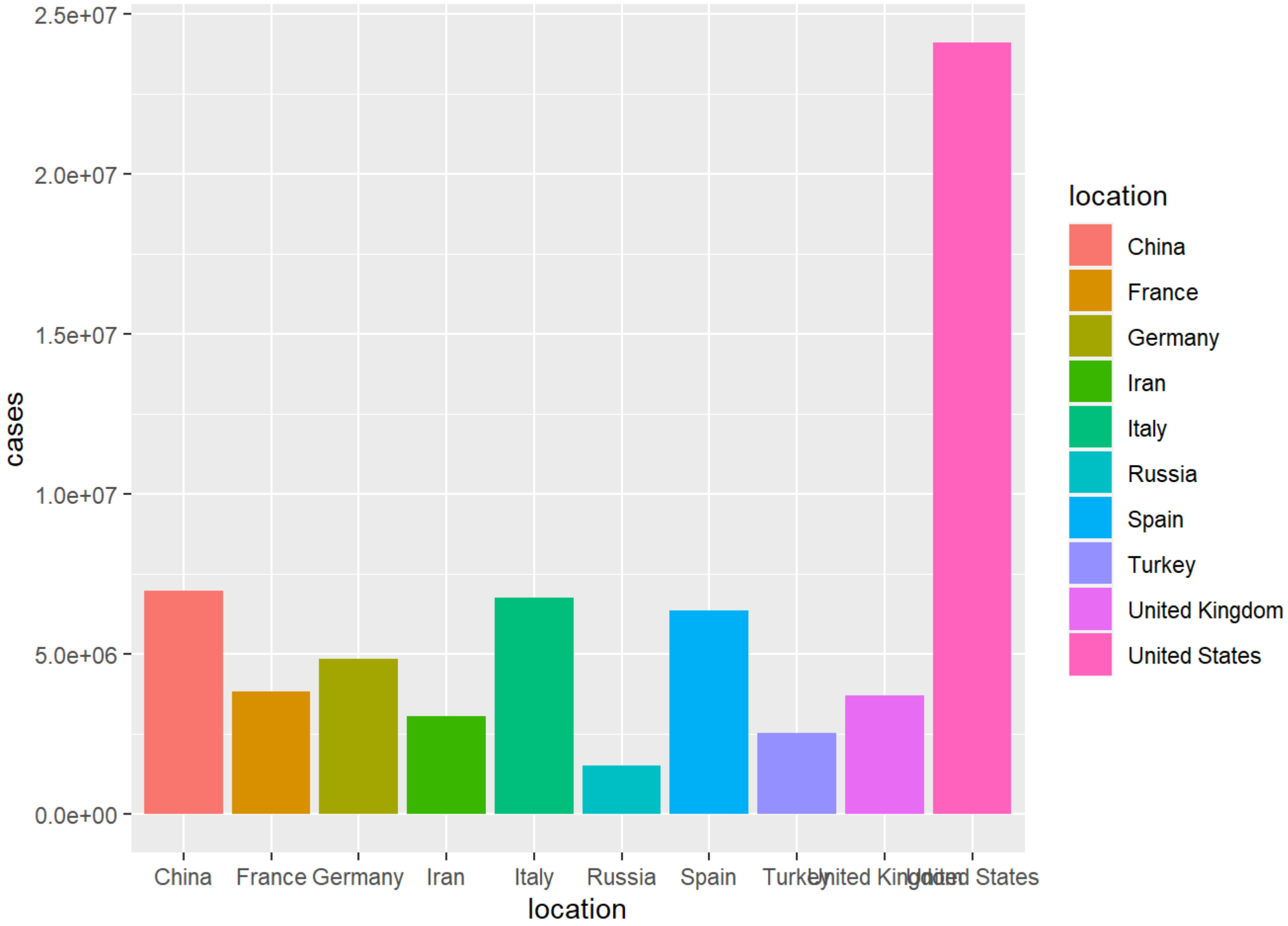
This map also shows that the states affected the most are located on the East coast. The map also shows that the virus is spreading from the East Coast toward the middle of the country to places like the Mid West, and the South.

Most affected countries

(Data from https://ourworldindata.org/ (https://ourworldindata.org/))

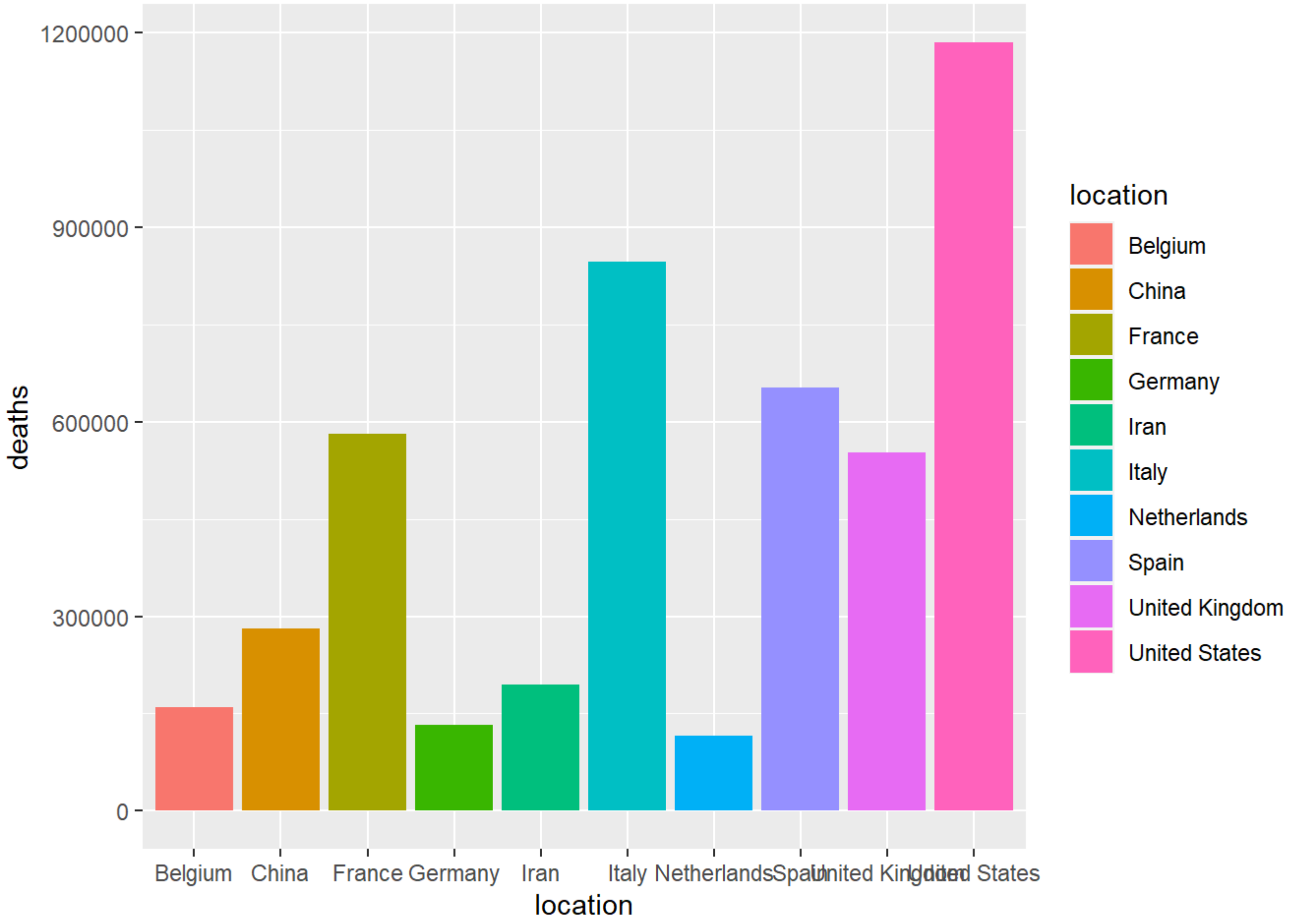
Worst Countries based on total cases

```
covid_data %>%
  select(location, total_cases, total_deaths) %>%
  filter(location != "World") %>%
  group_by(location) %>%
  summarise(cases = sum(total_cases)) %>%
  top_n(10, cases) %>%
  ggplot(aes(x = location, y = cases, fill = location)) +
  geom_col()
```



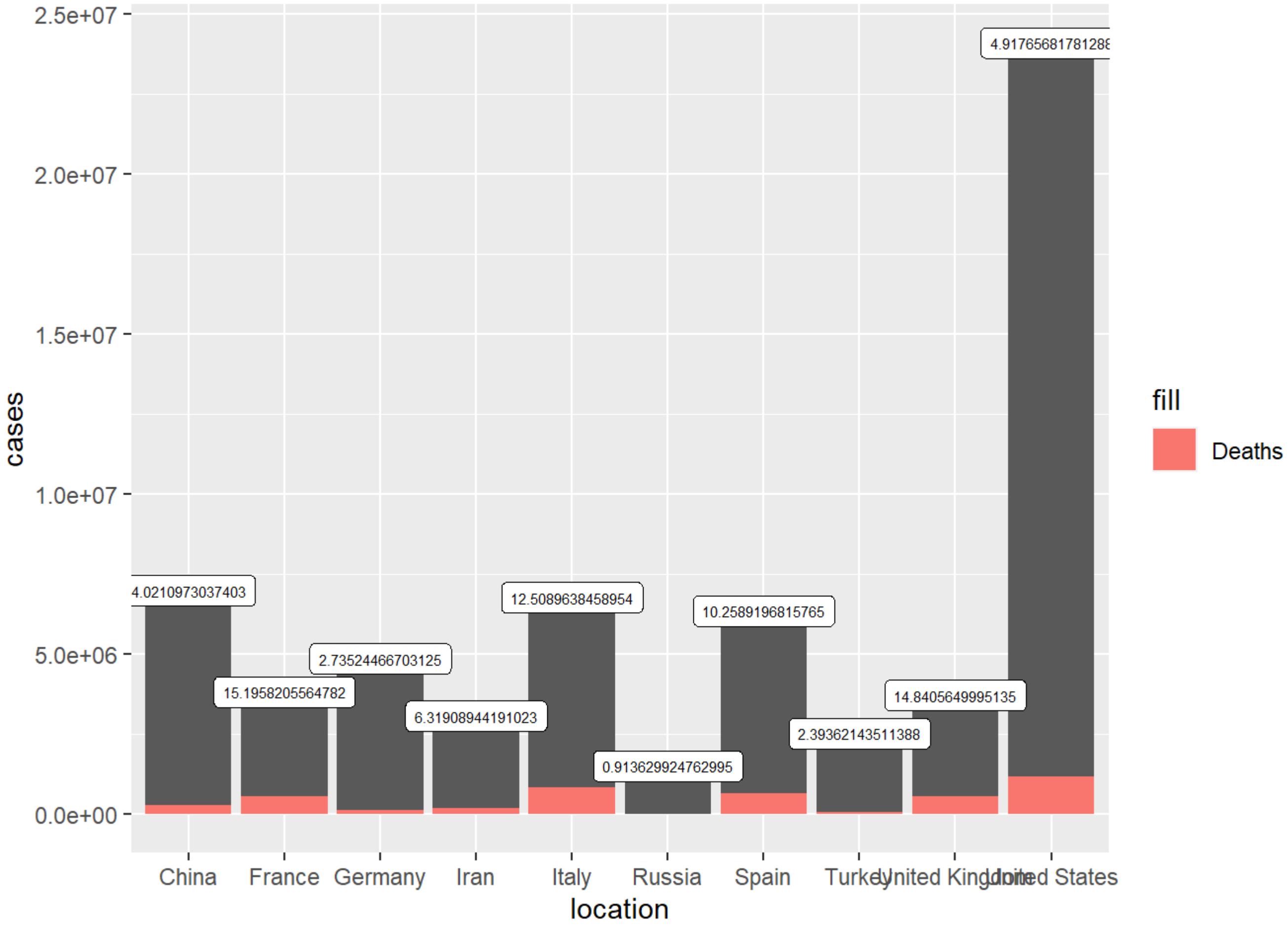
Worst Countries based on Deaths

```
covid_data %>%
  select(location, total_cases, total_deaths) %>%
  group_by(location) %>%
  summarise(deaths = sum(total_deaths)) %>%
  filter(location != "World") %>%
  top_n(10, deaths) %>%
  ggplot(aes(x = location, y = deaths, fill = location)) +
  geom_col()
```



Compare the top 10 countries with the most cases and calculate their death rates.

```
covid_data %>%
  select(location, total_cases, total_deaths) %>%
  group_by(location) %>%
  summarise(deaths = sum(total_deaths), cases = sum(total_cases)) %>%
  filter(location != "World") %>%
  top_n(10, cases) %>%
  ungroup() %>%
  ggplot() +
  geom_col(aes(x = location, y = cases)) +
  geom_col(aes(x = location, y = deaths, fill = "Deaths")) +
  geom_label(aes(x = location, y = cases, label = (deaths/cases)*100), lwd = 2)
```



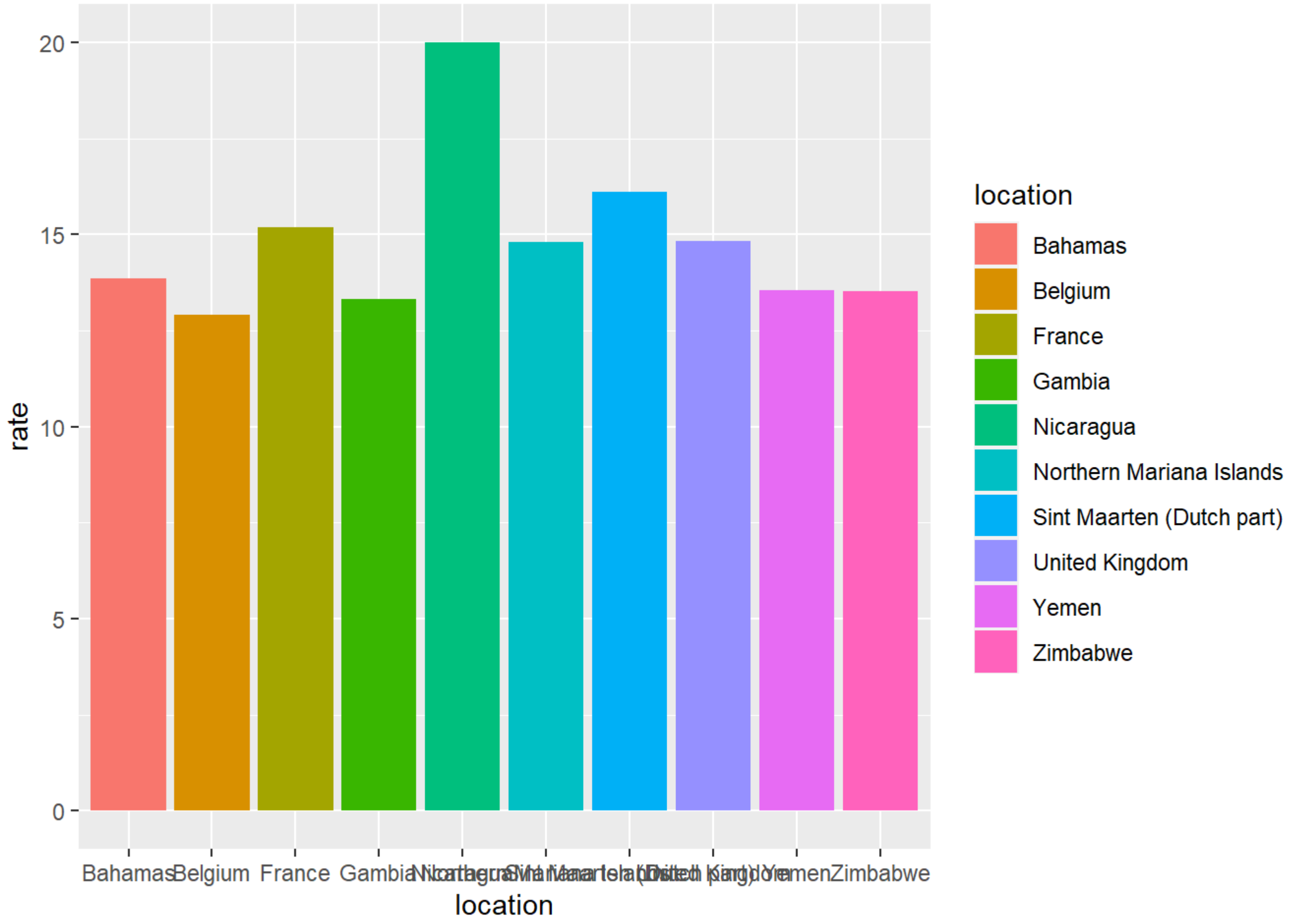
Here are the top 10 countries based on the death rate previously calculated. Some countries on this list have not shown up in any other graph yet. There are some 3rd world countries on this list that probably havea lack of heath care which is probably causing their death rates to spike. The UK and France are shocking to see on this list! I am surprised the US and Italy are not on there; I am sure that it is due to the higher amounts of cases in those two countries which makes the death rate fall.

```
covid_data %>%
  select(location, total_cases, total_deaths) %>%
  group_by(location) %>%
  summarise(deaths = sum(total_deaths), cases = sum(total_cases),
            rate = (deaths/cases)*100) %>%
  filter(location != "World") %>%
  top_n(10, rate) %>%
  ungroup() %>%
  arrange(desc(rate))
```

```
## # A tibble: 10 x 4
##    location                  deaths   cases  rate
##    <chr>                      <dbl>   <dbl> <dbl>
##  1 Nicaragua                     68     340  20  
##  2 Sint Maarten (Dutch part)    302    1873  16.1
##  3 France                    581183 3824624  15.2
##  4 United Kingdom            552080 3720074  14.8
##  5 Northern Mariana Islands      60     405  14.8
##  6 Bahamas                      272    1962  13.9
##  7 Yemen                          8      59  13.6
##  8 Zimbabwe                     103     762  13.5
##  9 Gambia                        42     315  13.3
## 10 Belgium                   159116 1231216  12.9
```

Here is a graph of that data:

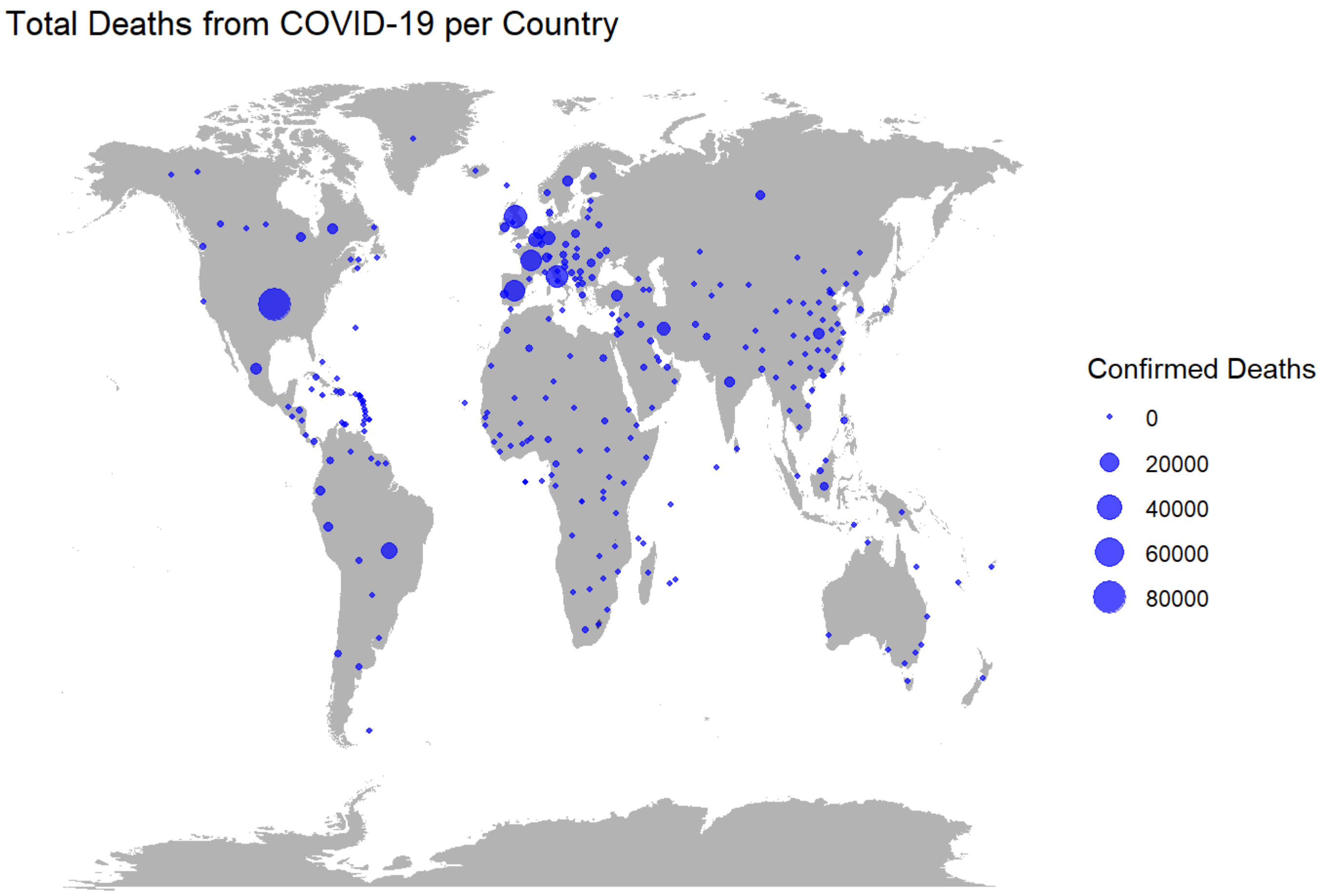
```
covid_data %>%
  select(location, total_cases, total_deaths) %>%
  group_by(location) %>%
  summarise(deaths = sum(total_deaths), cases = sum(total_cases),
            rate = (deaths/cases)*100) %>%
  filter(location != "World") %>%
  top_n(10, rate) %>%
  ungroup() %>%
  ggplot(aes(x = location, y = rate, fill = location)) +
  geom_col()
```



Global deaths caused by COVID-19

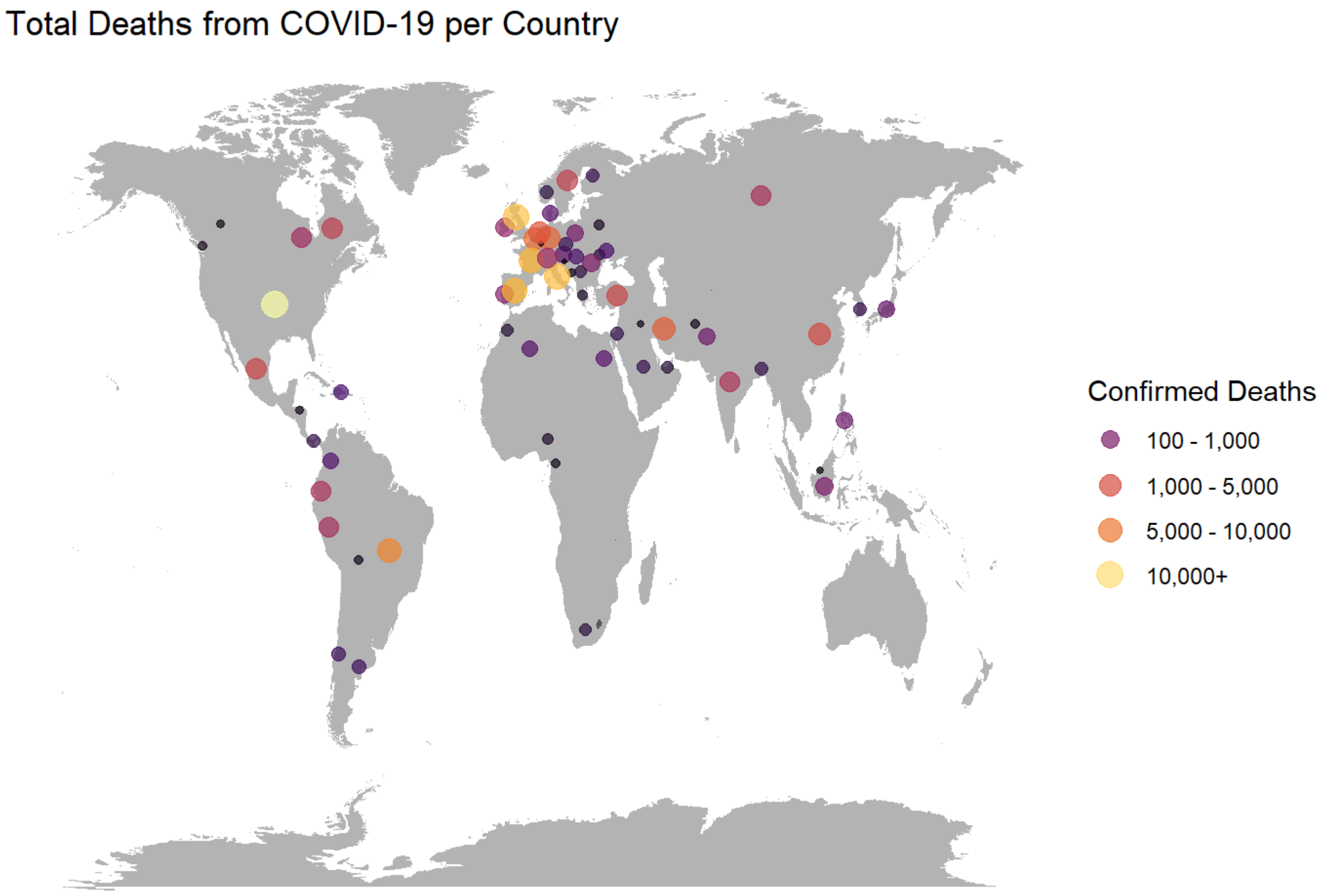
((Data from John Hopkins Univeristy Cener for System Science and Engineering))

```
global_deaths <- read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse
_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv")

global_deaths <- global_deaths %>%
  rename(province = "Province/State", region = "Country/Region") %>%
  arrange(region)

world_map <- map_data("world")


ggplot() +
  geom_polygon(data = world_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = global_deaths, aes(Long, Lat, size = `5/12/20`),
           stroke=F, alpha = 0.7, color = "blue") +
  theme_void() +
  guides(size = guide_legend()) +
  ggtitle("Total Deaths from COVID-19 per Country") +
  labs(size = "Confirmed Deaths")
```



That was a quick look at the deaths around the world. A lot of places stand out right away. Here is an even mroe in depth graph. This graph shows countries with over 100 deaths

```
global_deaths_over_100 <- global_deaths %>%
  filter(`5/12/20` >= 100)

breaks = c(100,1000,5000,10000,50000)

ggplot() +
  geom_polygon(data = world_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = global_deaths_over_100, aes(Long, Lat, size = `5/12/20`, color = `5/12/20`),
             stroke=F, alpha = 0.7) +
  theme_void() +
  guides(color = guide_legend()) +
  ggtitle("Total Deaths from COVID-19 per Country") +
  labs(color = "Confirmed Deaths") +
  scale_size_continuous(name = "Confirmed Deaths", trans = "log", range = c(1,5), breaks = break
s,
                        labels = c("<100", "100 - 1,000",
                                   "1,000 - 5,000", "5,000 - 10,000",
                                   "10,000+")) +
  scale_color_viridis_c(name = "Confirmed Deaths", option = "inferno", trans = "log",
                        breaks = breaks, labels = c("<100", "100 - 1,000",
                                    "1,000 - 5,000", "5,000 - 10,000",
                                    "10,000+"))
```



Total Deaths from COVID-19 per Country

You can see the US and Western Europe have very high death tolls, most likely due to their population which gives the virus many opporunities to spread quickly.

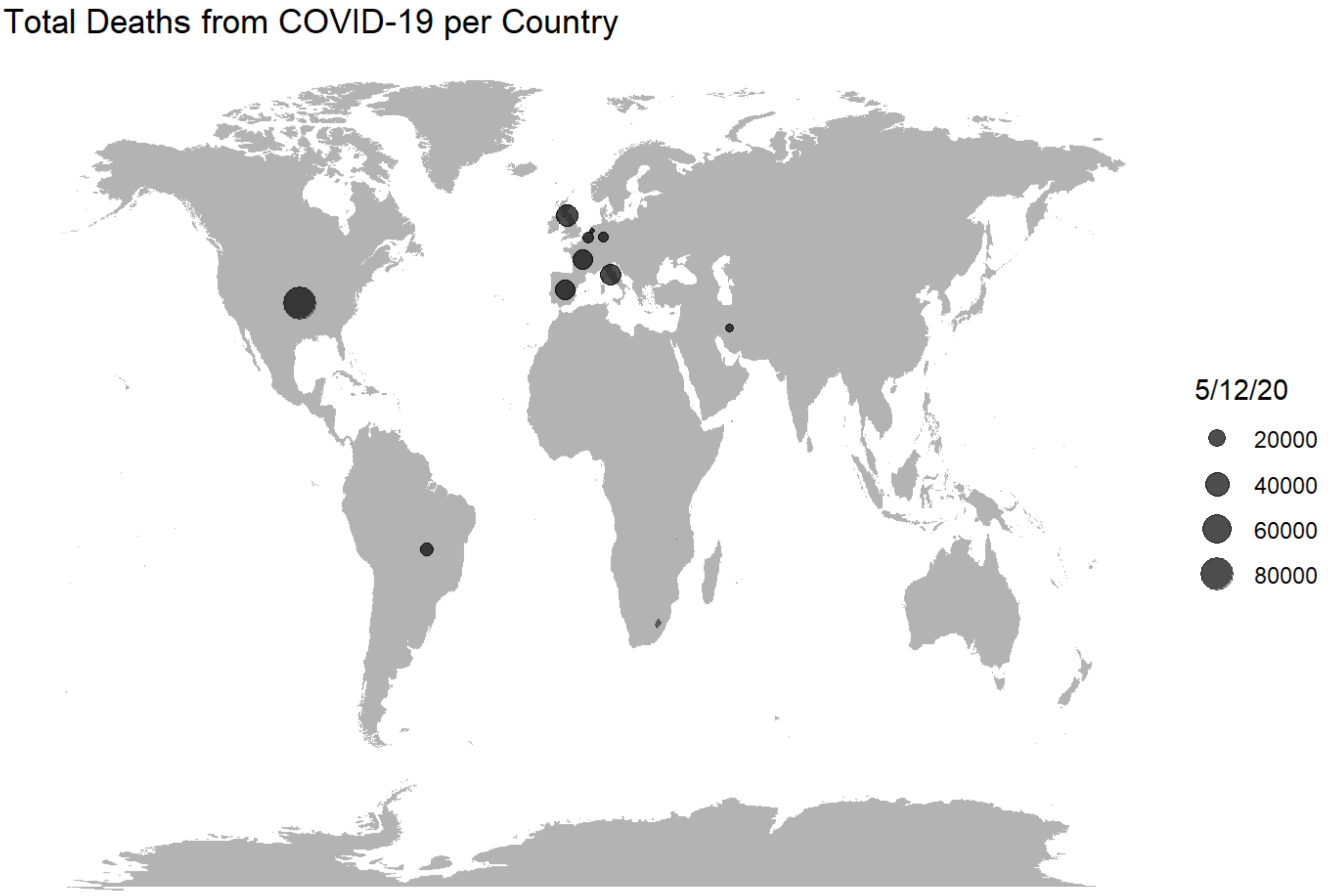
These are the top 10 countries when it comes to deaths:

```
global_deaths_top_10 <- global_deaths %>%
  select(region, Lat, Long, `5/12/20`) %>%
  arrange(desc(`5/12/20`)) %>%
  slice(1:10) %>%
  rename(Total_Deaths = "5/12/20")
global_deaths_top_10
```

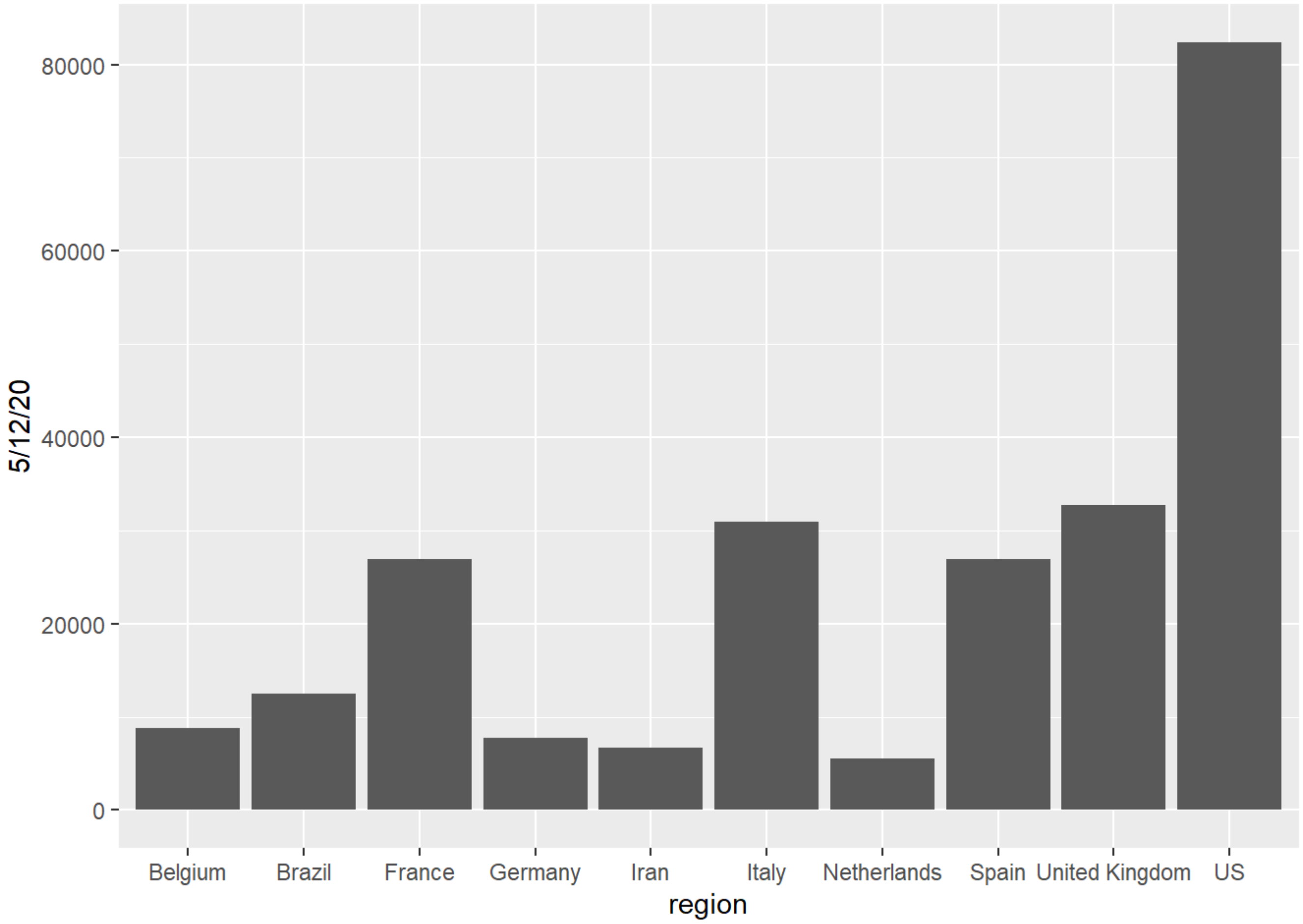
```
## # A tibble: 10 x 4
##    region           Lat   Long Total_Deaths
##    <chr>          <dbl>  <dbl>        <dbl>
##  1 US              37.1 -95.7         82356
##  2 United Kingdom  55.4  -3.44        32692
##  3 Italy           43    12           30911
##  4 France          46.2   2.21        26951
##  5 Spain           40    -4           26920
##  6 Brazil         -14.2 -51.9         12461
##  7 Belgium         50.8   4            8761
##  8 Germany         51     9            7738
##  9 Iran            32    53            6733
## 10 Netherlands     52.1   5.29         5510
```

Here are the top 10 countries with the highest death toll:

```
global_deaths_top_10 <- global_deaths %>%
  select(region, Lat, Long, `5/12/20`) %>%
  arrange(desc(`5/12/20`)) %>%
  slice(1:10)
ggplot() +
  geom_polygon(data = world_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = global_deaths_top_10, aes(Long, Lat, size = `5/12/20`),
             stroke=F, alpha = 0.7) +
  theme_void() +
  guides(color = guide_legend()) +
  ggtitle("Total Deaths from COVID-19 per Country") +
  labs(color = "Confirmed Deaths")
```

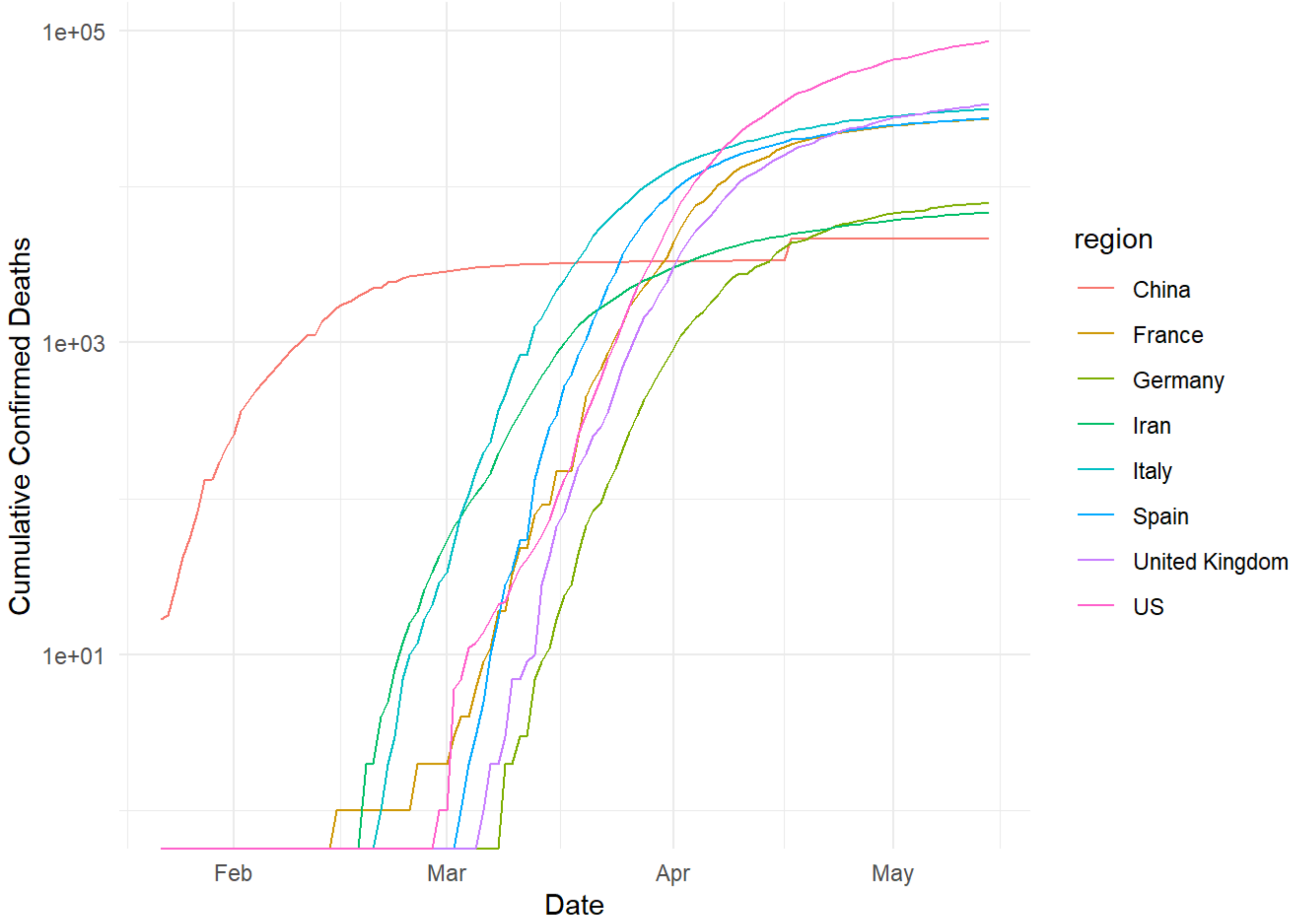


```
global_deaths_top_10 %>%
ggplot(aes(x = region, y = `5/12/20`)) +
  geom_col()
```



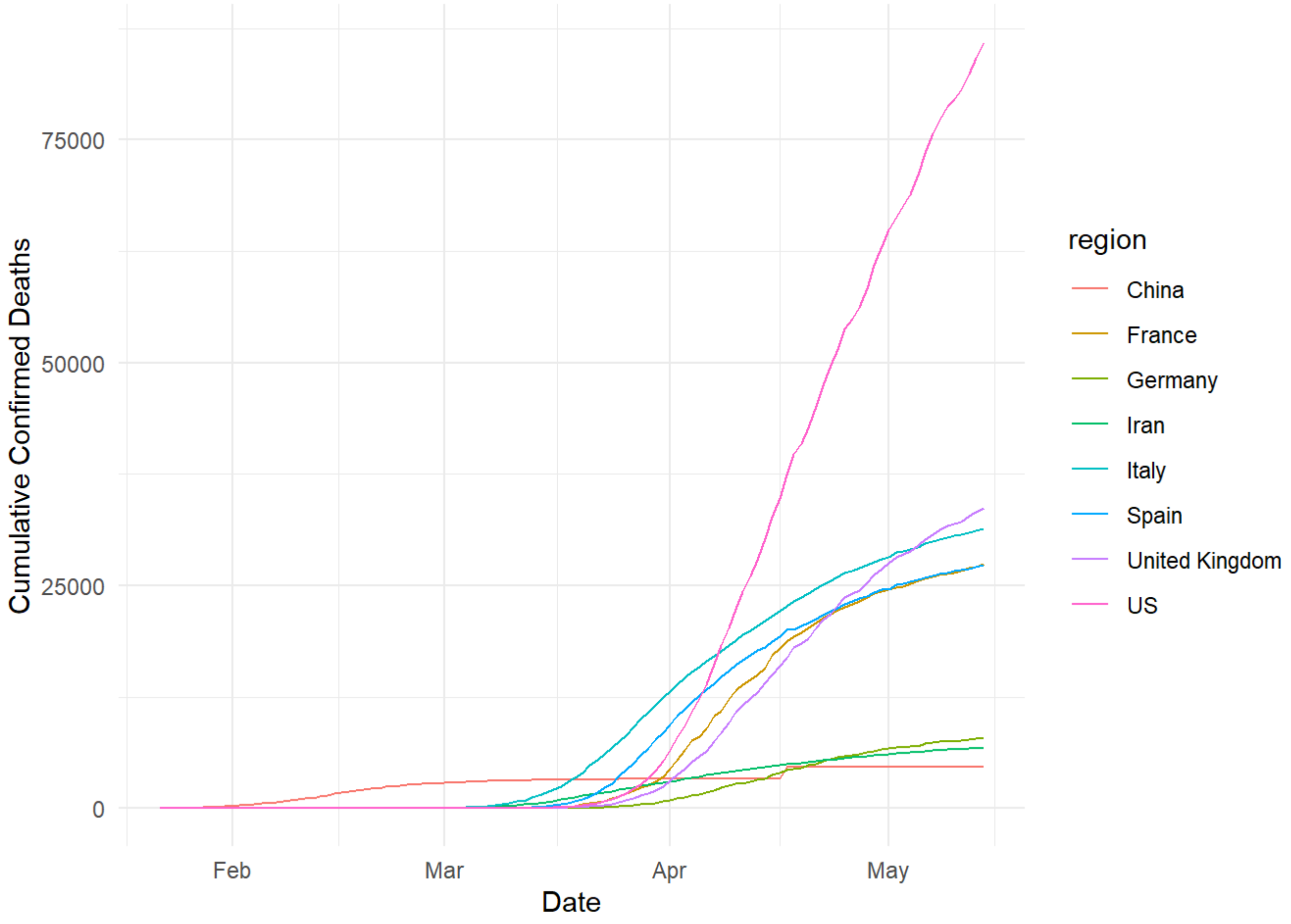
Here are the trends. You can see that deaths are on the rise still

```
global_deaths %>%
  pivot_longer(-c(province, region, Lat, Long), names_to = "Date",
               values_to = "cumulative_deaths") %>%
  mutate(Date = mdy(Date)) %>%
  arrange(Date) %>%
  filter(region == "US" |
         region == "China" |
         region == "United Kingdom" |
         region == "Italy" |
         region == "France" |
         region == "South Korea" |
         region == "Germany" |
         region == "Iran" |
         region == "Spain") %>%
  group_by(region, Date) %>%
  summarise(cum_deaths = sum(cumulative_deaths)) %>%
  ungroup() %>%
  ggplot(aes(Date, cum_deaths, color = region, group = region))+
  geom_line() +
  theme_minimal() +
  ylab("Cumulative Confirmed Deaths") +
  xlab("Date") +
  scale_y_log10()
```



Here is that same graph without the logarithmic transformation just to show that the curve is not necessarily flattening yet, especially in the US. Other countries have experienced some slight flattening

```
global_deaths %>%
  pivot_longer(-c(province, region, Lat, Long), names_to = "Date",
               values_to = "cumulative_deaths") %>%
  mutate(Date = mdy(Date)) %>%
  arrange(Date) %>%
  filter(region == "US" |
         region == "China" |
         region == "United Kingdom" |
         region == "Italy" |
         region == "France" |
         region == "South Korea" |
         region == "Germany" |
         region == "Iran" |
         region == "Spain") %>%
  group_by(region, Date) %>%
  summarise(cum_deaths = sum(cumulative_deaths)) %>%
  ungroup() %>%
  ggplot(aes(Date, cum_deaths, color = region, group = region))+
  geom_line() +
  theme_minimal() +
  ylab("Cumulative Confirmed Deaths") +
  xlab("Date")
```

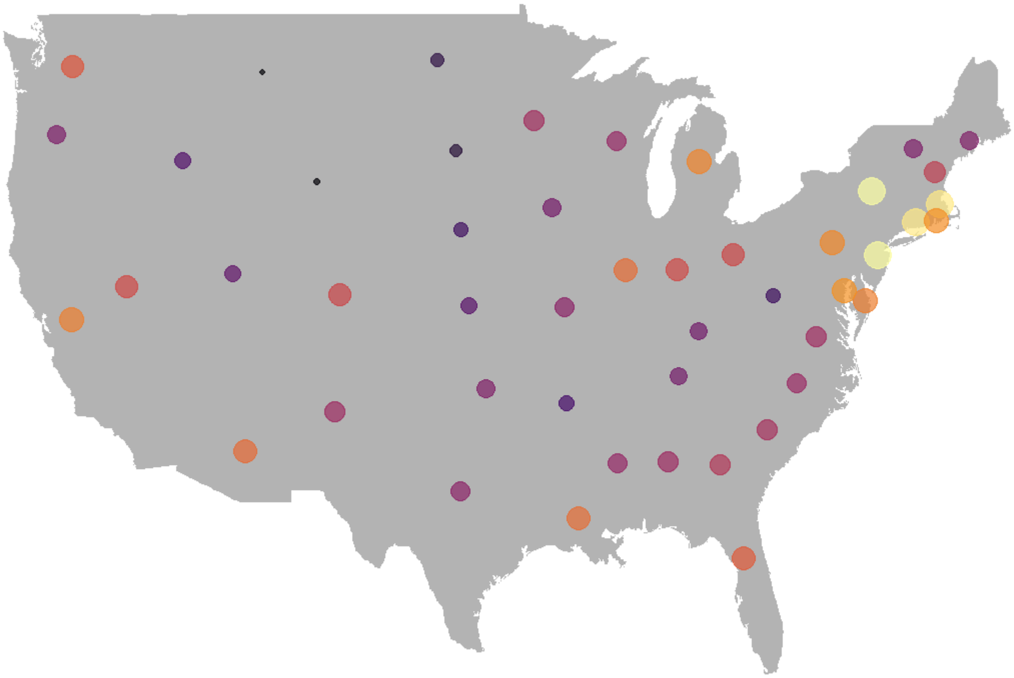


Deaths in the US caused by COVID-19

(Data from John Hopkins Univeristy Cener for System Science and Engineering)

```
us_deaths <- read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_cov
id_19_data/csse_covid_19_time_series/time_series_covid19_deaths_US.csv")

# Only main land US like the graph to make it easier

us_deaths_main <- us_deaths %>%
  filter(iso2 == "US") %>%
  filter(Province_State != "Alaska",
         Province_State != "Hawaii",
         Province_State != "Diamond Princess",
         Province_State != "District of Columbia",
         Province_State != "Grand Princess") %>%
  select(Province_State, Lat, Long_, `5/12/20`) %>%
  mutate(mean = sum(`5/12/20`)) %>%
  group_by(Province_State) %>%
  summarise(deaths = mean(`5/12/20`), lat = median(Lat), long = median(Long_))

breaks = c(500, 1000, 5000, 10000, 25000, 50000, 100000)

ggplot() +
  geom_polygon(data = US_map, aes(long, lat, group = group), fill="black", alpha = 0.3) +
  geom_point(data = us_deaths_main, aes(long, lat, size = deaths, color = deaths),
             stroke=F, alpha = 0.7) +
  theme_void() +
  guides(color = guide_legend()) +
  ggtitle("Total Deaths by COVID-19 in the US") +
  labs(color = "Deaths") +
  scale_size_continuous(name = "Deaths", trans = "log", range = c(1,5), breaks = breaks,
                        labels = c("500 - 1,000", "1,000-5,000", "5,000-10,000", "10,000 - 25,00
0",
                                   "25,000-50,000", "50,000-100,000", "100,000+")) +
  scale_color_viridis_c(name = "Deaths", option = "inferno", trans = "log",
                        breaks = breaks, labels =
                          c("500 - 1,000", "1,000-5,000", "5,000-10,000", "10,000 - 25,000",
                            "25,000-50,000", "50,000-100,000", "100,000+")) +
  ylim(20, 50) +
  xlim(-130, -60)
```

Total Deaths by COVID-19 in the US



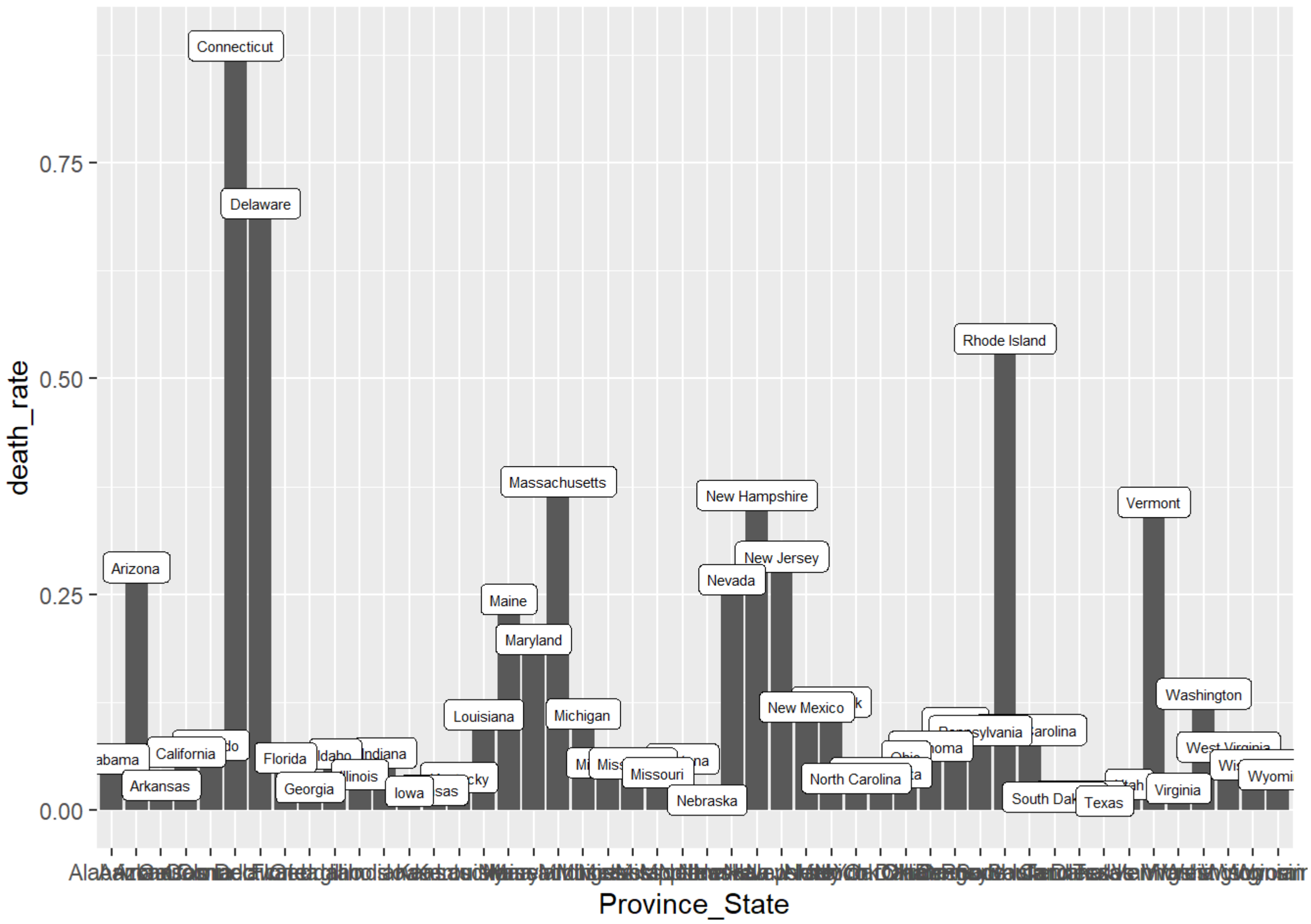
Here are the states with the highest detah rates:

```
us_deaths <- us_deaths_main %>%
  select(-c(lat, long)) %>%
  inner_join(us_confirmed_main_by_state, by = "Province_State") %>%
  select(-c(lat, long)) %>%
  mutate(death_rate = (deaths / cases)*100) %>%
  arrange(desc(death_rate))

us_deaths
```

```
## # A tibble: 48 x 4
##    Province_State deaths  cases death_rate
##    <chr>           <dbl>  <dbl>      <dbl>
##  1 Connecticut    304.    34333      0.886
##  2 Delaware        47.4    6741      0.703
##  3 Rhode Island    63.4   11614      0.546
##  4 Massachusetts  302.    79332      0.381
##  5 New Hampshire   11.8    3239      0.365
##  6 Vermont          3.31    927      0.357
##  7 New Jersey     414.   140917      0.294
##  8 Arizona         33.1   11736      0.282
##  9 Nevada          16.9    6313      0.268
## 10 Maine            3.61   1477      0.244
## # ... with 38 more rows
```

Here is a graph of the worst states. This gives a visual at how some states are doing worse based on the death rate. From this you can see that the death rate is very high in Eastern states with smaller populations. New York is one of the hardest hit states in the country, but their death rate is far smaller becasue they have a large population. For this reason, death rate should not be looked at too closely

```
us_deaths_main %>%
  select(-c(lat, long)) %>%
  inner_join(us_confirmed_main_by_state, by = "Province_State") %>%
  select(-c(lat, long)) %>%
  mutate(death_rate = (deaths / cases)*100) %>%
  arrange(desc(death_rate)) %>%
  ggplot(aes(x = Province_State, y = death_rate)) +
  geom_col() +
  geom_label(aes(label = Province_State), lwd = 2)
```



In conclusion, the data clearly shows that coronavirus cases are still rapidly rising, especially in the US. Deaths from COVID-19 are also still increasing. It looks like the pandemic will still be continuing for a few more months at least.