COVID-19 and Its Impact on the World

W200 Project 2

Nate Leech, Mohamed Gesalla, Mikayla Pugel

The COVID-19 pandemic has transformed the way the people are living their lives and how people perform their daily activities. For most people, this is the first pandemic they have lived through and there is so much as a society that we can learn to become better prepared in the chances of another pandemic. Throughout this report we will clean and analyze COVID-19 data from many different countries around the world. Through this analysis we hope to gain a better understanding of how different countries were affected, and explore some of the demographic and geographical features of different countries, and how these could have affected their number of cases, spread rate, and death rate. The COVID-19 data set we used is from the website OurWorldInData.org. The dataset itself did not require much cleaning other than removing any empty variables for certain countries. Many of our sanity checks we completed in the first few graphs below as we saw that the COVID-19 data for many countries trended as we expected it would.

The worldwide toll of COVID-19 thus far has been devastating. In *Figure 1*, below, we see the current total number of cases is nearing the 60 million mark and the current total number of deaths is nearing 1.4 million. These have been trending upwards since the start of the pandemic with only the hope of a vaccine to show us an end in sight.

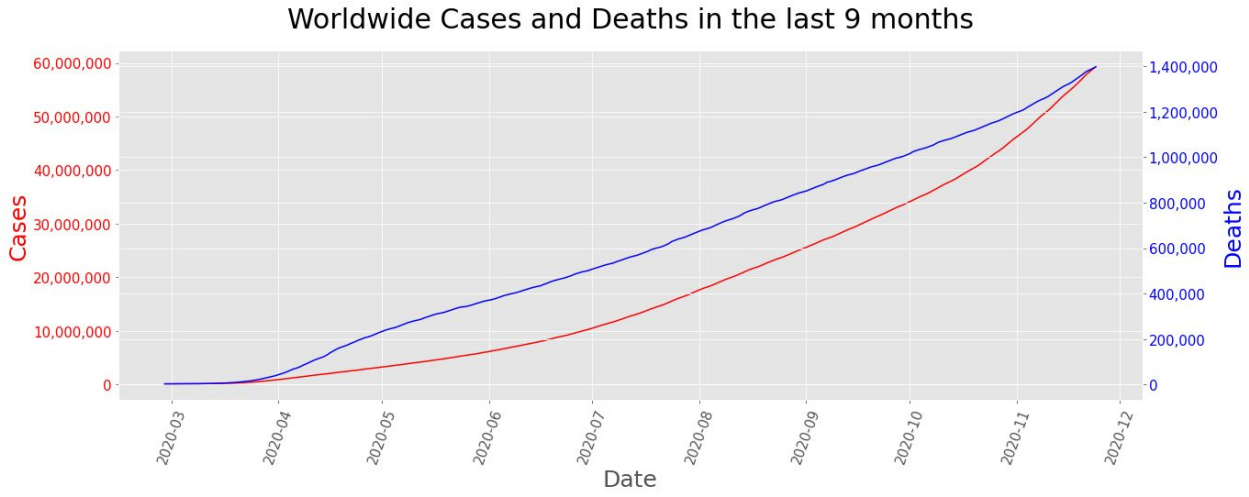

Figure 1

For *Figure 2* and *Figure 3*, we see data from countries most affected by COVID-19 in terms of cases and hospitalizations, respectively. Case data follows a trend we expect to see based on the information coming in from several news sources. We see very few cases in the beginning of March, 2020 with a steady rise in cases throughout the year. We chose to base these figures off of cases per million in order to more fairly compare all countries to one another.

Figure 2 is particularly interesting in that we can see specific spikes in hospitalization over the lifespan of the pandemic. An initial spike is seen at the beginning of April, a guess could be made as to why and it could be that the virus was very new and as a matter of caution, more patients were admitted than actually needed to be. We see a second spike beginning at the onset of Fall, coinciding with the possible start of a school year.

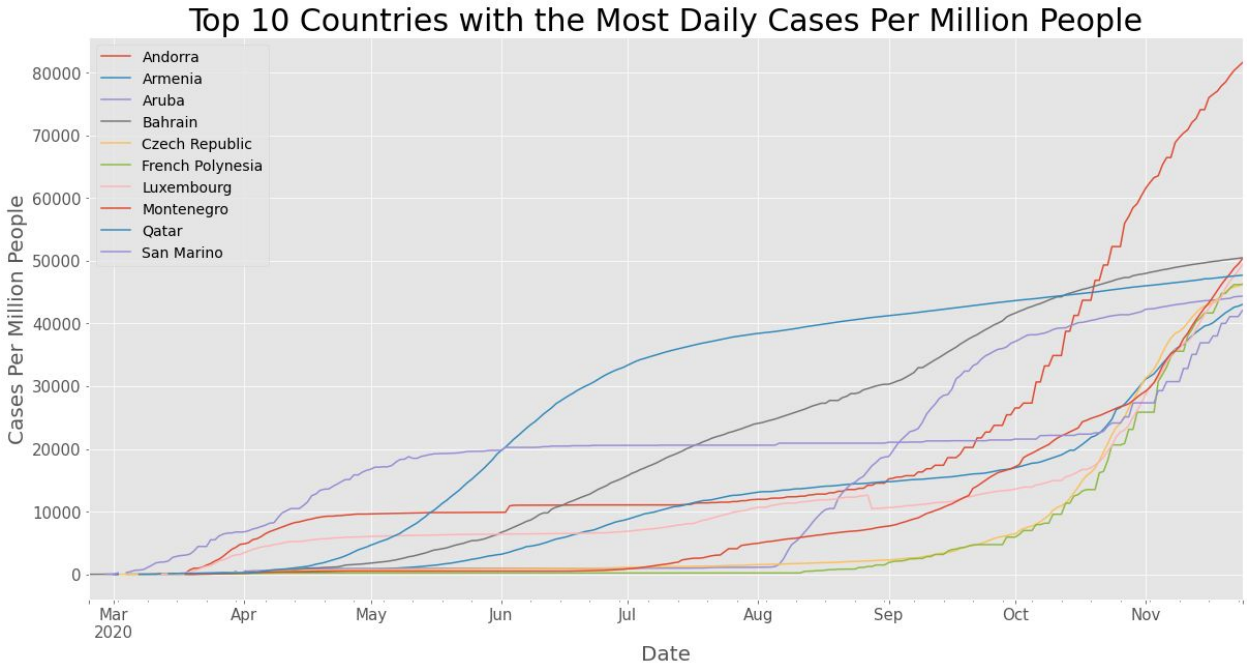

Figure 2

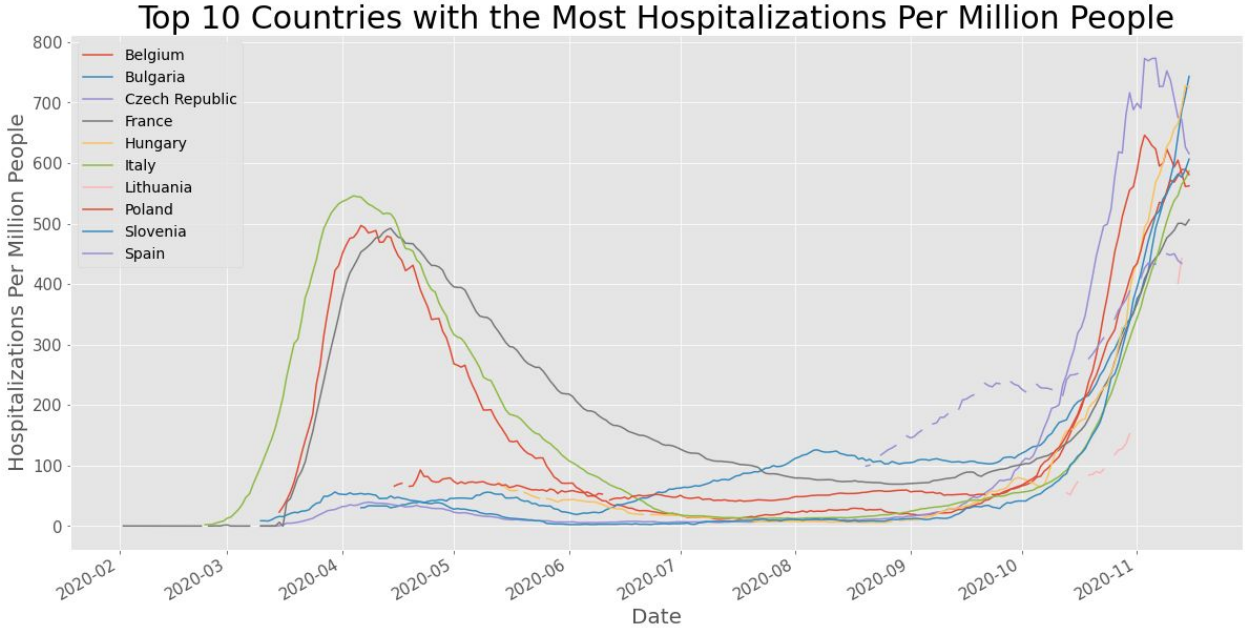

Figure 3

The next plot, *Figure 4,* shows the top 10 countries with populations over 100 million that show the highest rate of increasing cases. We initially plotted the top 10 highest rates of spread overall, but that was giving a confusing looking plot where a country with 1 case initially, and 4 cases, seven days later would show a 400% increase. While true, the number of cases seemed too small to be able to gain any insights from. The use of countries with larger populations seemed to make more sense and gives the reader more familiar countries to compare against. For countries with this high of a population, it seems fascinating to see as high as a 12% increase of cases in a country like Japan.

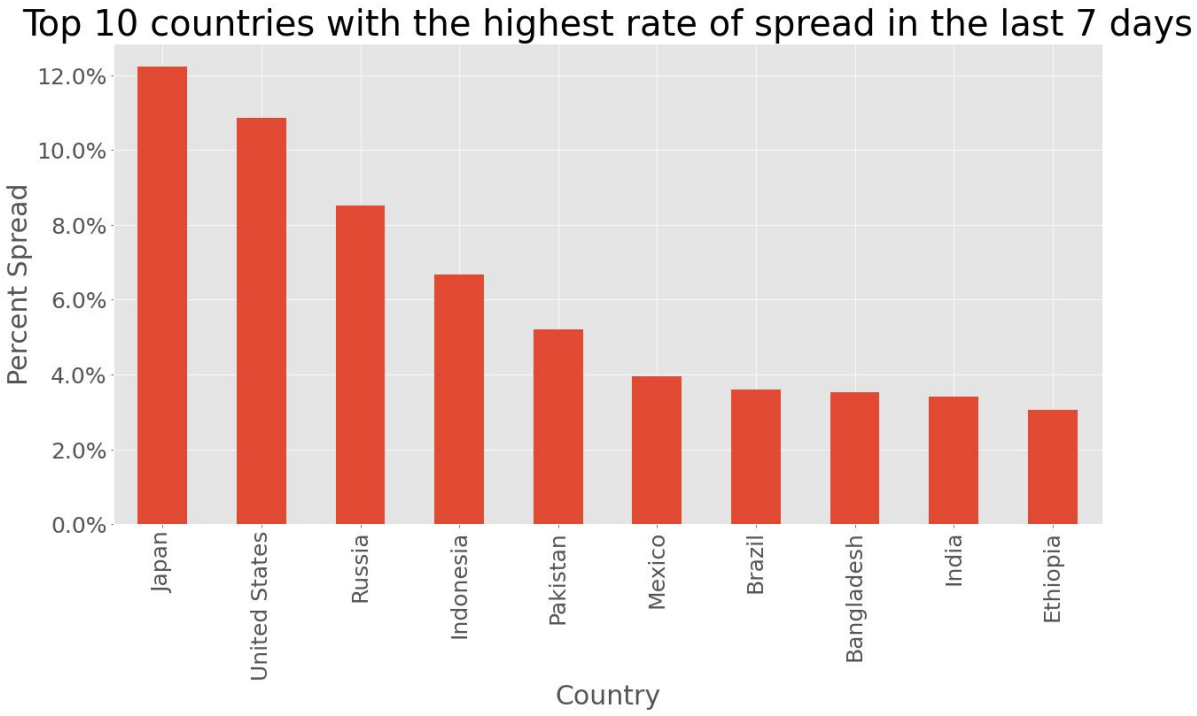

Figure 4

As just discussed, many countries were affected by COVID-19 in varying severity. These differences can be attributed to a wide variety of differences among the countries. One group of attributes of particular interest are the demographic characteristics. We will now explore how the different demographic features of these countries could have impacted the number of cases and deaths.

With knowing how this disease is spread, the first demographic characteristic that came to mind that may affect the severity of COVID-19 was the population density. Since the disease can be transmitted quickly in large groups and from being close to one another, the initial assumption is that countries with a large population density are more likely to spread the disease faster, and therefore have a higher number of cases and deaths, and also lead to a higher positive rate. From *Figure 5* below, we can see that countries with some of the highest positive rates have a low population density. It is obvious from the graph below that population density is not a large determinant of if a country will have a high positive rate. The largest outlier from our initial assumption is Singapore at the bottom right hand corner of the figure. Singapore, with a population density of 7,915 people per square kilometer had one of the lowest positive rates. This could be for many reasons, for example countries with a really high population density may have provided many tests to people which would reduce the positive rate from tests. Countries with high population density could also be hyper aware of how easily it could be spread in their community and therefore take the social distancing and quarantining very seriously. From this analysis, our initial assumption was shown to be incorrect.

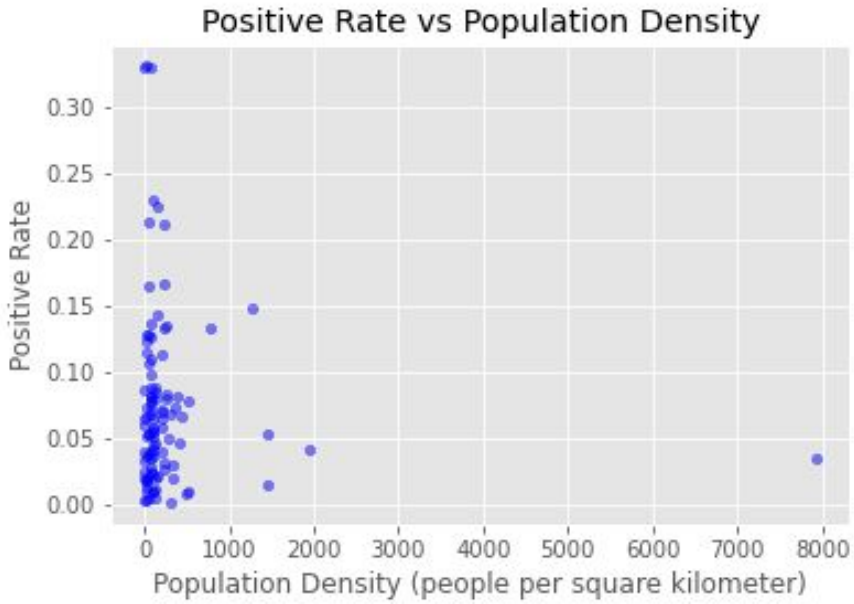

Figure 5

After analyzing population density, it occurred to us that the case count is only going to be as high as the number of tests that the country is able to provide. This led us to our next analysis question. How does GDP affect the number of tests that a country provides, and how does the availability of these tests affect the case count. We initially assumed that the higher a country's GDP, the more tests they could provide, and more tests could then lead to a higher number of cases. From *Figure 6* below, it can be seen that there is a decently strong correlation between a country's GDP and the number of tests that are available per person. From the second figure, *Figure 7,* we can see that there is a weak correlation between the total number of tests per thousand and the number of cases per thousand in a country. From these graphs we can see that our initial assumption is generally correct, however there are some possible risks with many of the assumptions we made. One assumption is that the countries with low GDP have less tests because they cannot afford it, not because they have less people with COVID-19 symptoms. Countries with higher GDP could have more people who think they have COVID, possibly due to these countries having tourism and travel, and that is why they have more tests. With the second figure below, we assumed that countries with low number of tests, most likely have a low number of cases. This assumption is a valid assumption, as you cannot know if someone is sick without a test. However, we cannot assume that a country's cases are low, just because they are unable to test, but rather they may not have many sick people. From these assumptions, we can say that GDP does have a positive relationship with the number of tests performed in a country, and that this has a positive effect on the total cases, however most likely not as strong as the graphs below make it seem.

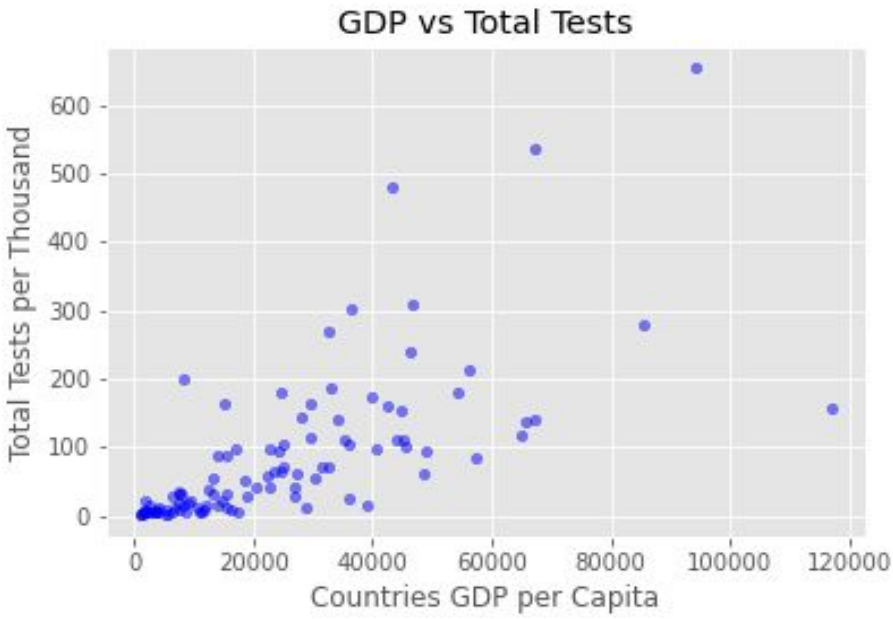

Figure 6

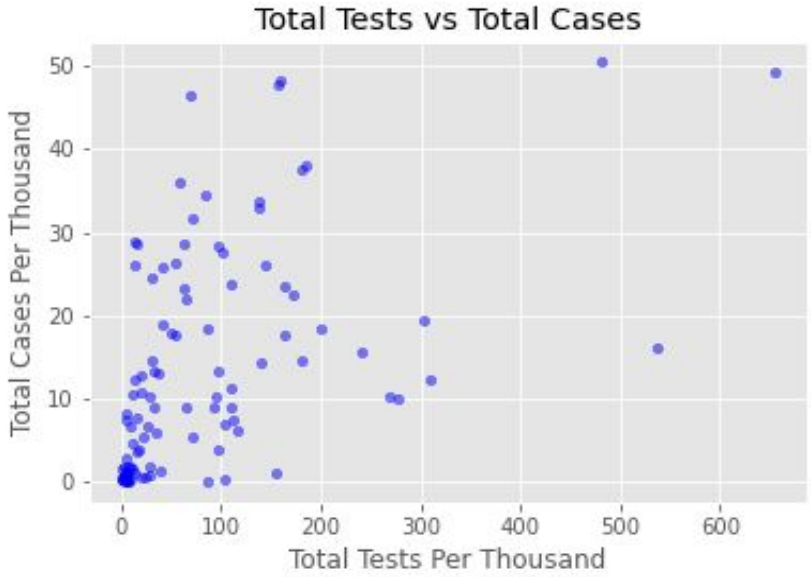

Figure 7

For the second part of exploring the demographics of each country, we wanted to see if there were any characteristics that impacted the death rate of COVID-19. Since COVID-19 is a respiratory disease, therefore impacts people's lungs and their ability to breath, we assumed that countries with a high percentage of smokers would have a higher death rate among their COVID-19 patients (Galiatsatos, 2020). *Figure 8* below displays the rate of women and men smokers within a country in correlation with the country's deaths per million. From the graphs we can see that there is not a strong correlation with the number of smokers and the number of deaths, however we have made many assumptions in this comparison. One assumption is that a country's infected population has the same fraction of people who

smoke as the overall country. It is possible that a certain country's population who got COVID-19 had a smaller fraction of smoking individuals and vice versa. This question seems inconclusive with the data and information we have from this data set at this time.

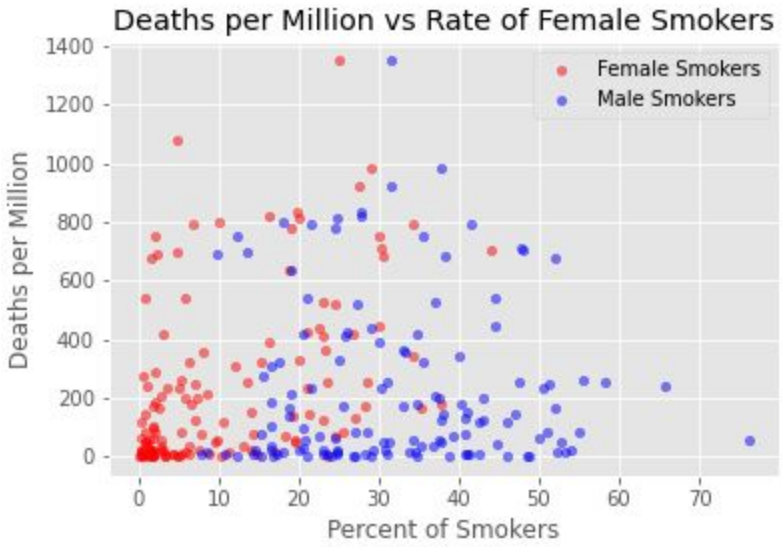

Figure 8

One last question we wanted to answer with the demographic data of the countries, is how does life expectancy impact the death rate of a country. Normally, we assume that countries with higher life expectancy have better general health than countries with lower life expectancy. Therefore, our initial assumption of this question would be that the higher the life expectancy, the lower the death rate, or at least a death rate consistent with countries with a lower life expectancy. From *Figure 9* below, we can see that our initial assumption was incorrect. There could be a couple reasons for this. First, COVID-19 has been seen to have an increased severity on the elderly population. The odds of dying from COVID-19 drastically increase from the age of 70 and up (“Older Adults”, 2020). Since countries with a low life expectancy have fewer people in the 70 and up age range, their total deaths per million is drastically lower than countries with life expectancies on the higher end. This is very interesting, as it might mean that COVID-19 impacts a healthy 70 year old more severely than an unhealthy 55 year old who is soon going to die. However, we cannot make that statement with confidence as life expectancy in different countries varies for many different reasons. For example, some countries may have a high infant mortality rate, which would skew the life expectancy. Overall, this graph shows an interesting depiction of how old age may be the best predictor of if you are going to live from COVID-19.

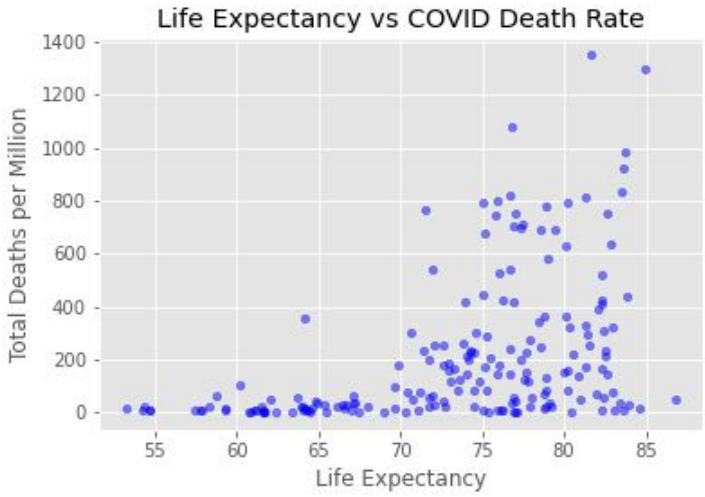

Figure 9

As we have seen in the past few months, COVID-19 case and death numbers varied widely from continent to continent. The two charts above depict stats from the beginning of the COVID-19 pandemic back in March to early November 2020. The assumptions we made in this section is that continents and countries with colder climates and larger and buiser airports will have a higher number of recorded COVID-19 cases.

Asia tends to have a warmer climate than Europe and North Africa and yet it recorded more cases and deaths as shown in *Figure 10* and *Figure 11* and that is most likely due to the very high population density. But when looking at the rest of the continents, we can see that colder continents such as North America and Europe which have much colder climate and less population density than African recorded much more cases and death. Also our assumption about the size and traffic of airports seem to follow the order of the weather conditions assumption, continents with larger and busier airports recorded more COVID-19 cases over the past 9 months.

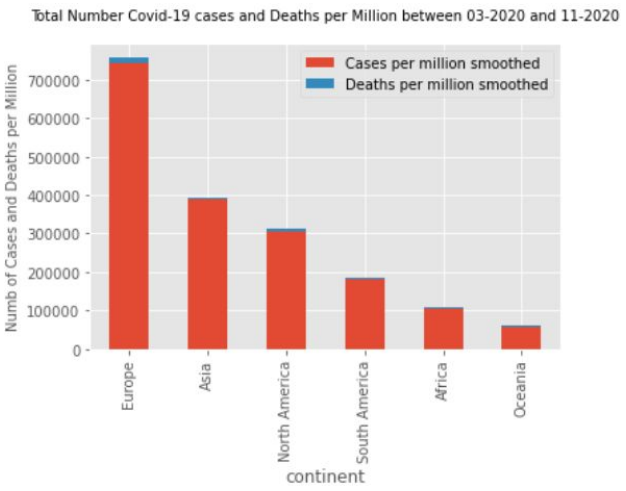

Figure 10

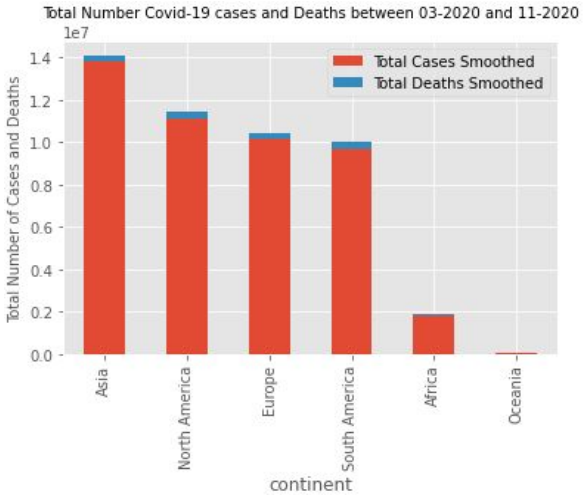

Figure 11

To take the analysis a step further, we explored data for North America as shown in *Figure 12* and *Figure 13* . When looking at the total number of cases, we can see that countries with the colder weather condition in North America like the United States and Canada have the highest number of cases but they are also the countries with the biggest population densities in the continent. To get a more accurate look and eliminate some of the impact that the population density has on the data.We looked at the per-million data. And we can see that it is not necessarily true that countries with colder climates recorded a higher number of cases per million. Even when comparing the number of cases per million in the United States and Canada. We can see that the United States is hit a lot harder with the virus despite having less colder weather than Canada.

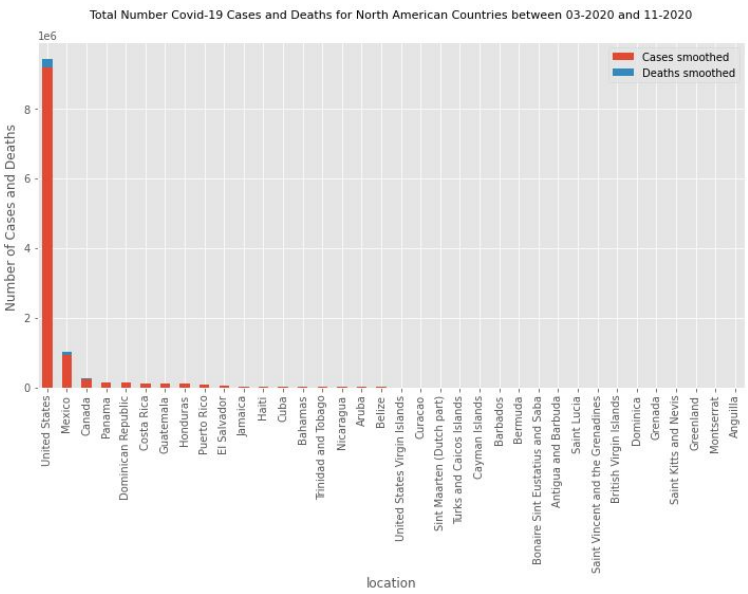

Figure 12

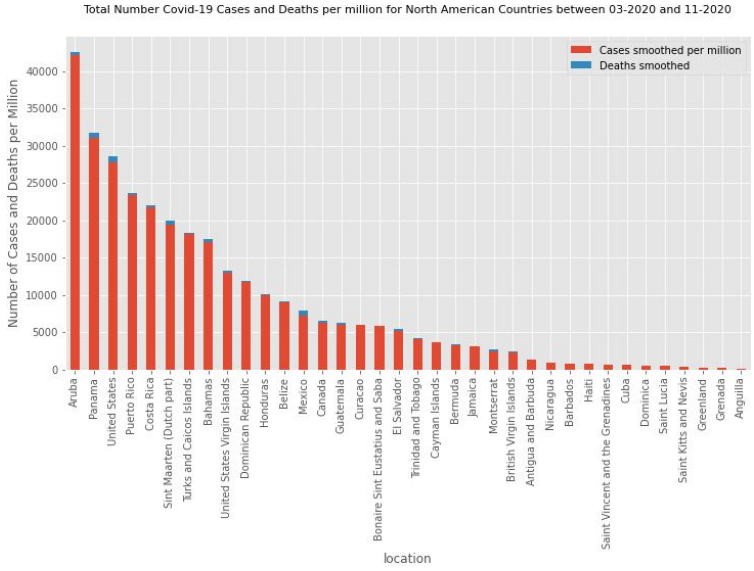

Figure 13

Below *Figure 14* is a time series plot of the number of COVID-19 cases over the first 9 months of the pandemic for the 4 countries with the highest number of cases in the world. Our assumption prior to exploring the data is that during colder seasons, there will be bigger recorded numbers of covid cases. When comparing these covid lines in the plot above against the weather charts for each of these four countries, we noticed that the movement of the COVID-19 line followed the weather chart for some countries but not all. For example, in South Africa, the most cases recorded were during the period from beginning of July to mid-August which is during the winter there. Also, in Russia we saw most of the spikes from October to about mid-April and that is exactly when the temperature dropped significantly.

However and interestingly enough, the United States and India did not show any correlation between the number of cases and the weather conditions. India maintained a temp between 26 and 30 °C through COVID-19 months. But it also maintained a steady rise in the

number of cases up until mid-September. Even though the weather conditions continued to be the same after mid-September there seem to be a decline in the number of cases.
In the United States. We noticed spikes throughout the year and during different weather conditions that includes the summer which is proven by the spike in mid-June.
Overall, from our data and analysis in weather conditions and COVID-19 data, we cannot conclude that there is a direct correlation between weather conditions and the number of COVID-19 cases.

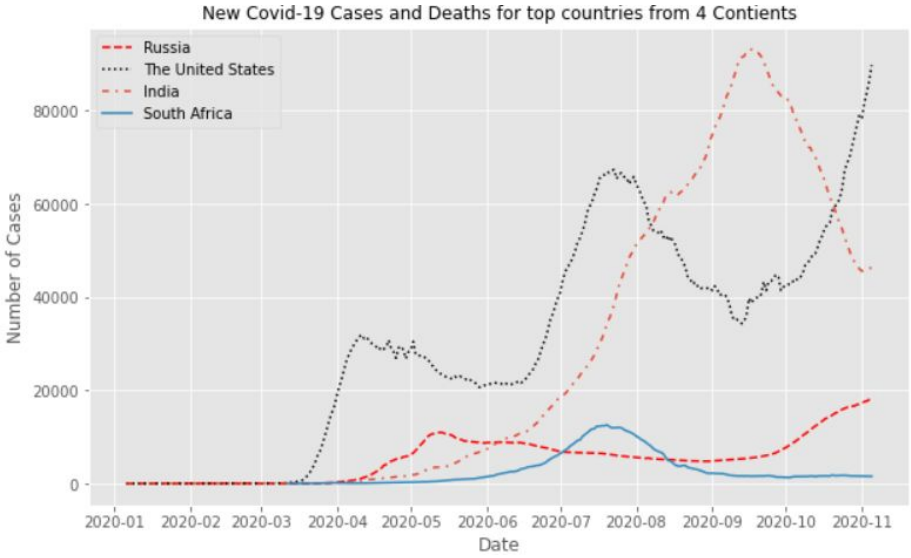

Figure 14

Throughout this analysis, we worked to understand the spread of COVID-19 throughout the globe, how it has impacted different groups of people, and then tried to explain why different places were affected so differently. From this analysis we learned that it is very difficult to analyze data that has a lot of variables, assumptions, and could have been collected in many different ways. However, we found many interesting insights different from many of our initial assumptions and questions. Some of these insights include how COVID-19 displays differently based on different demographic and geographic characteristics, such as the death rates in smokers and a person's life expectancy, as well as how temperature and geographic regions affected the spread. Overall, we were able to model the COVID-19 data in many different ways to explore and answer our research question.

References:

Galiatsatos, P. (2020, April 13). What Coronavirus Does to the Lungs. Retrieved December 04, 2020, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/what-coronavirus-does-to-the-lungs

Older Adults and COVID-19. (2020, November 27). Retrieved December 04, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html