Mid-Term Project Report

STAT 515 – 006
Group 3
Koustav Roy, Indu Priya Sharma, and Carter Forry

Project Purpose

In this report we pose to redesign the COVID data presented in Figure 1 created by Ref [1]. The graphic is readable; however, we contest that the "Total cases" is not properly displayed to convey the relationship between total cases and cases per one million. Humans are not adept to making comparisons when given only a list of numbers such as the "Total cases" shown in Figure 1. To make a holistic comparison, there should be two units of measure on the same plot such as a scatter plot.

Figure 1 also shows only a snapshot in time; therefore, it would be more insightful to display the total cases and cases per one million for given population over time. This may not be achievable in print media, but with today's digital world, there are plenty of options to convey how multivariable data changes over time such as animations. Therefore, the purpose is to convert Figure 1 into a more digestible graphic showing both Total cases and cases per million in a more digestible format as well as showing how the day changes over a given time.

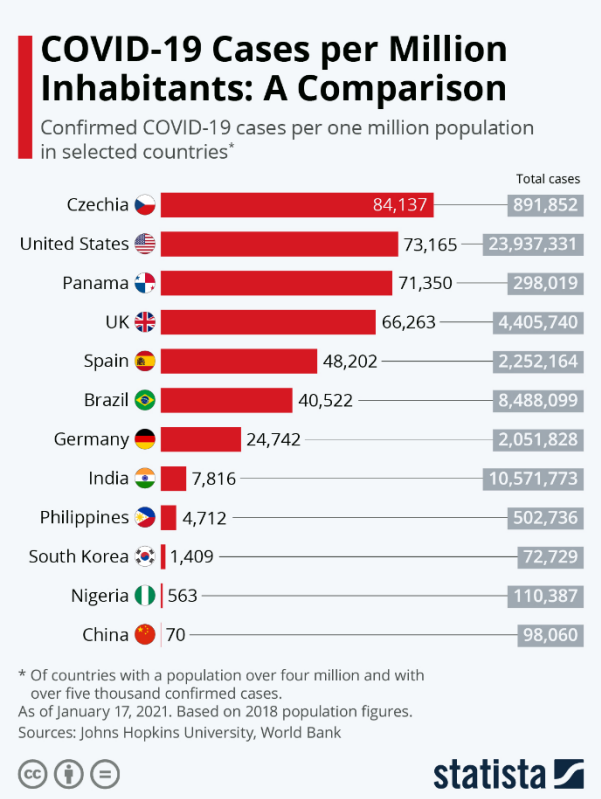

Figure 1: Original COVID-19 graphic [Ref 1]

COVID-19 Scatter Plot Representation

The first objective is to take Figure 1 and convert it into a scatter plot to show the relationship between Total cases and cases per million inhabitants. Figure 2 shows how the COVID cases per million relates to the Total amount of COVID cases in each country. The scatter plot shown in Figure 2 better illustrates how larger countries in terms of population have varying infection rates per million people. For example, the United States has a comparable infection rate per one million people as Panama; however, the United States has significantly more total infections. Figure 1 generated by Ref [1] only shows a bar chart of the infection rate and not the total number of cases. Using the below scatter plot allows the reader to further understand the relationship between total cases and the infections rate per one million citizens. What this allows the reader to do is interoperate better the how countries of varying size have handled COVID-19.

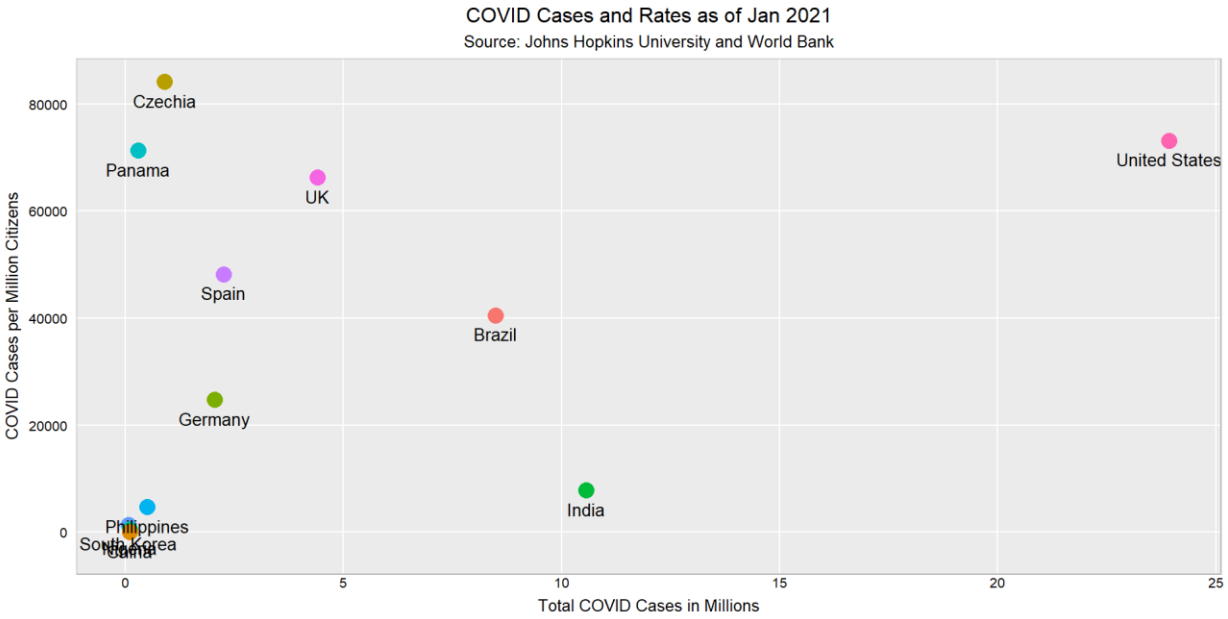

Figure 2: COVID Cases per Million vs Total COVID Cases (data derived from Ref [1])

Figure 2 has some limitations, particularly viewing the countries in low total cases and low infection rates. This is due to the data only representing one instance in addition to the data being significantly different (ie. United Stated versus Philippines). Therefore, Figure 2 better shows the relationship between countries compared to the original graphic (Figure 1), but some countries with very similar data (China, South Korea, and Philippines) Figure 2 is not appropriate to use. What would be better to show is how the COVID data changes over time or a geographic representation.

COVID-19 Geographical Map

The original graphic (Figure 1) simply lists the countries in descending order of COVID infection rates, so using a map yields insight to how COVID data varies globally (Figure 3).

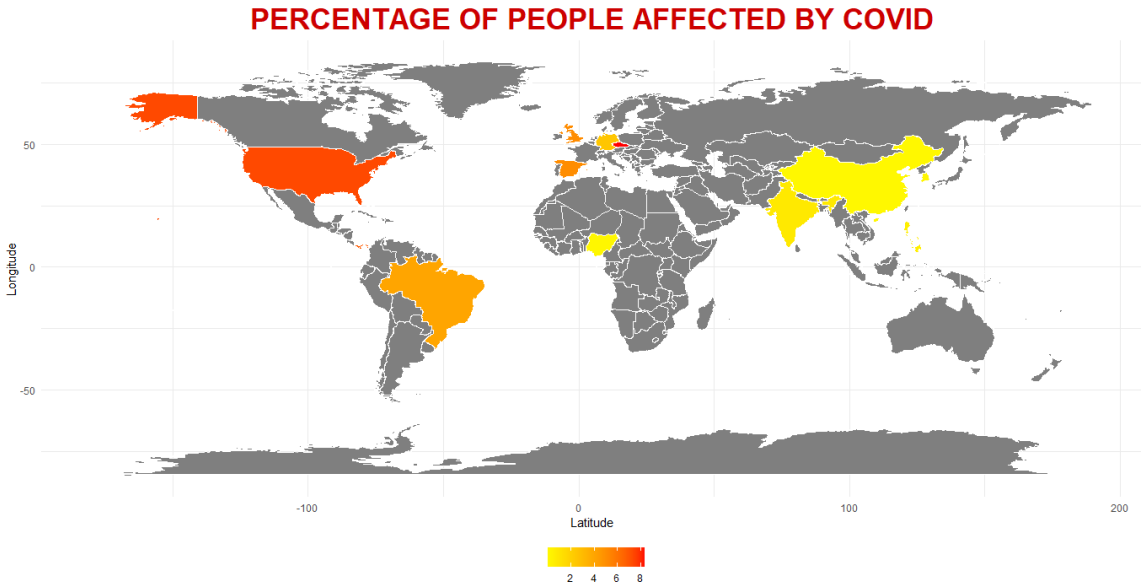

Figure 3: Percent Population Infected with COVID 17 Jan 2021 (Ref [1] data)

Overlaying the data in Figure 1 on a world map is not too insightful as only a few countries are listed. Therefore, when the original data is used to show geographical differences in COVID cases and COVID rate, it is optimal to use as much country data as possible. Using data from Ref [2], more of the global map can be filled in to show differences across the globe (Figure 4).

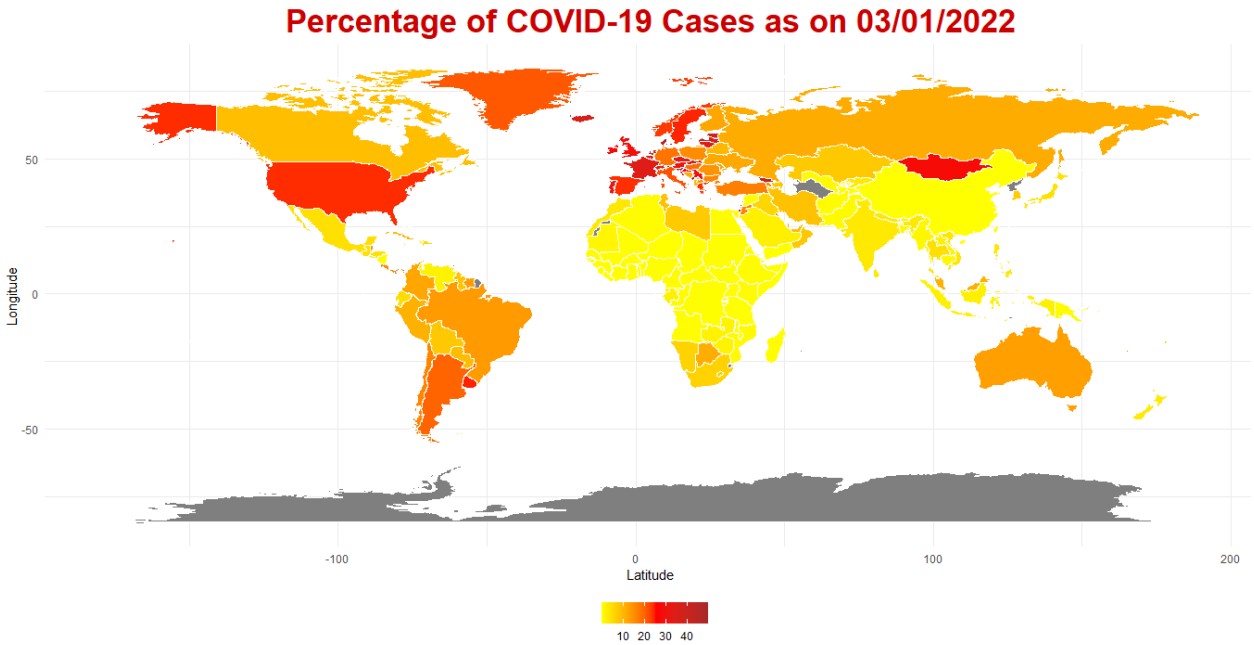

Figure 4: Percent Population Infected with COVID 01 Mar 2022 (Ref [2] data)

Overall, to illustrate geographical differences in COVID cases and infection rate, it is best to show a map versus a bar chart such as the original graphic in Figure 1.

COVID-19 Changing in Time

The original COVID graphic (Figure 1) only shows a data recorded on 17 January 2022; however, it is more insightful for the reader to visualize how COVID-19 cases and infection rate change over time in relation to each other. Data was found (Ref [2]) from the start of the global pandemic (Feb 2020) through March 2022 to illustrate how COVID data has changed for the indicated countries. Figure 5 shows the total cases over time and Figure 6 shows the infection rate over the same time duration.

Showing how the total COVID cases and infection rate change over time yields further insight into particular COVID strands such as Omicron in late 2021 and early 2022. This can be seen as a rapid rise in total cases and an increase in infection rate. This trend would have never been seen in the original data presented in Figure 1.

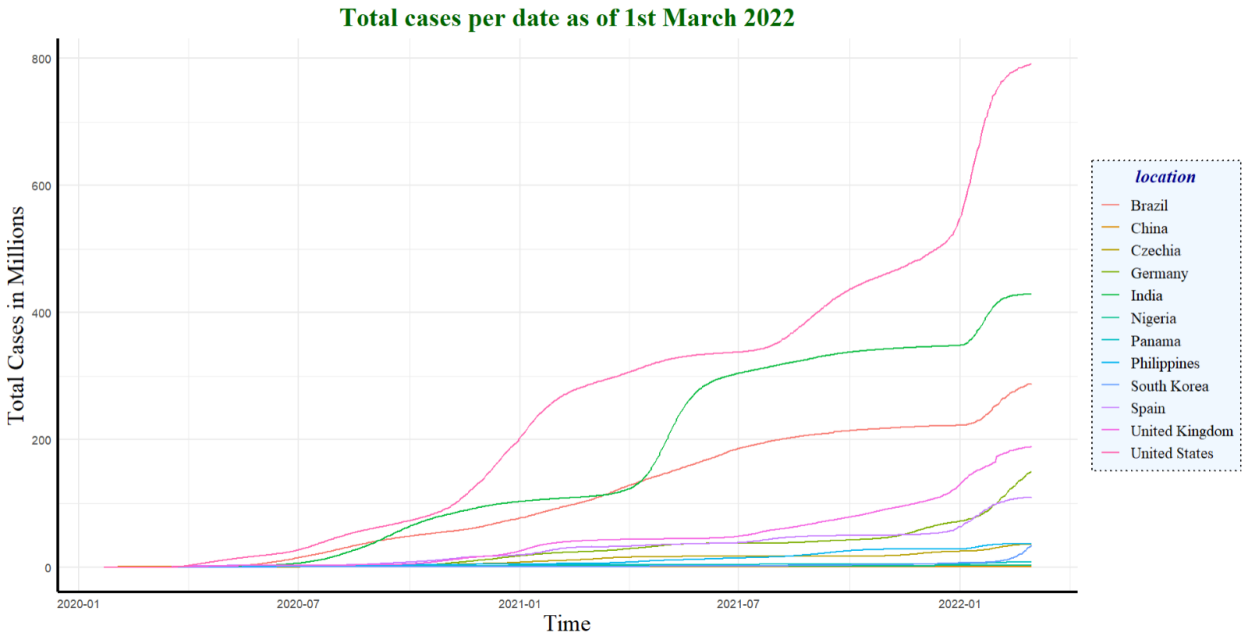

Figure 5: Total COVID Cases During the Pandemic (Ref [2] data)

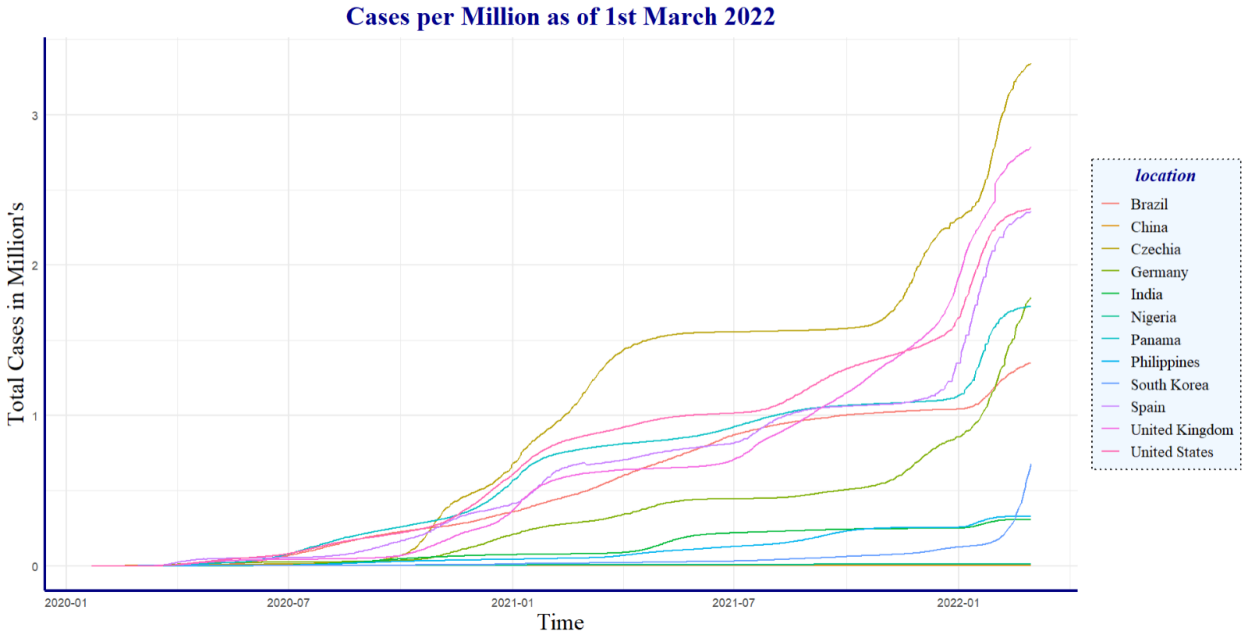

Figure 6: COVID Cases Per Million Citizens During the Pandemic (Ref [2] data)

Figures 5 and 6 would be appropriate to use in print media and on digital platforms. However, since a large portion of the population digests news via digital means such as Twitter and Facebook, it would be an attractive and insightful to implement an animated plot to show the COVID data. Figures 5 and 6 can be converted into digitally-animated plots since animations tend to be more attractive to views.

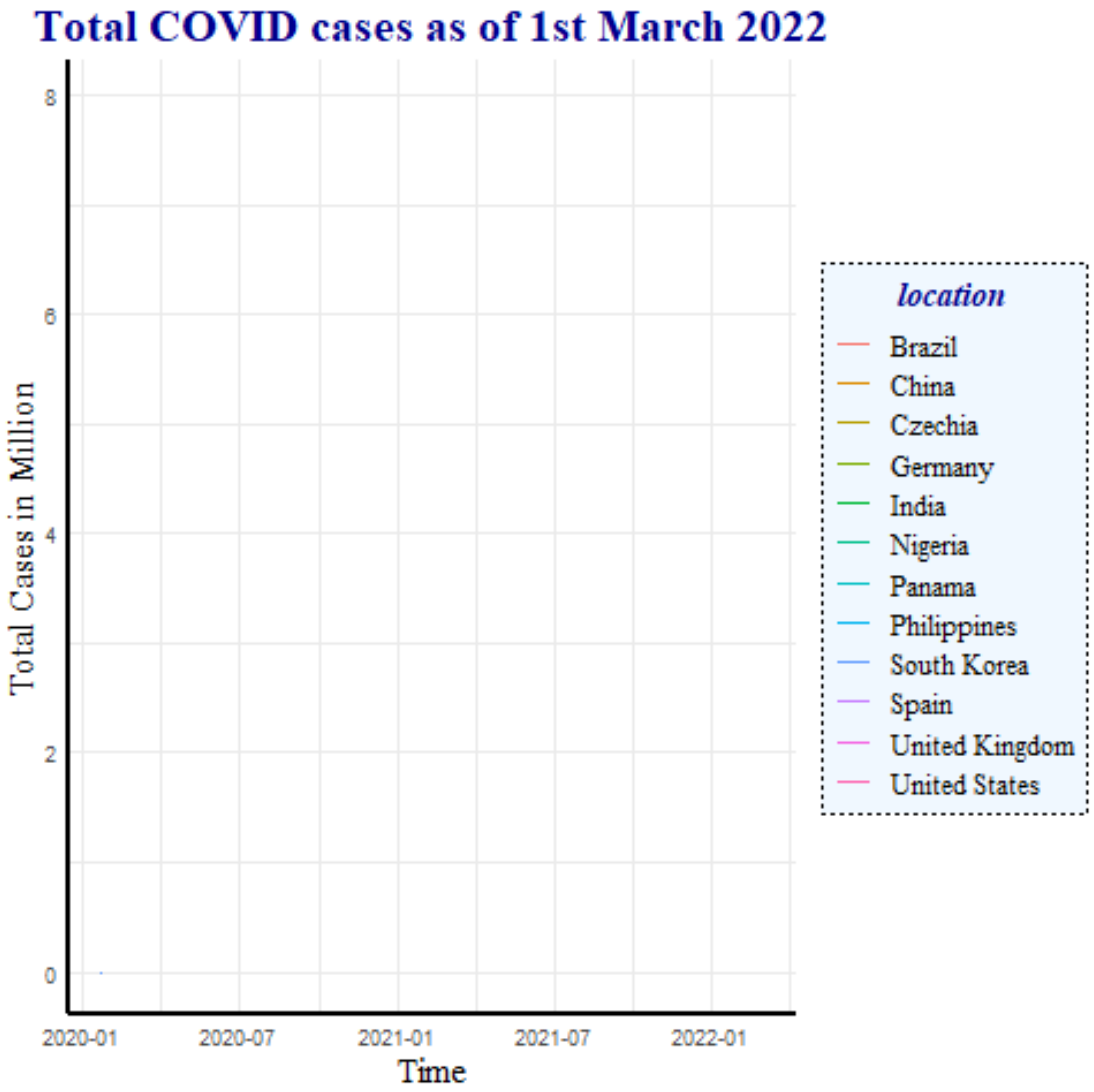

Figure 7: Total COVID Cases During the Pandemic - Animated (Ref [2] data)

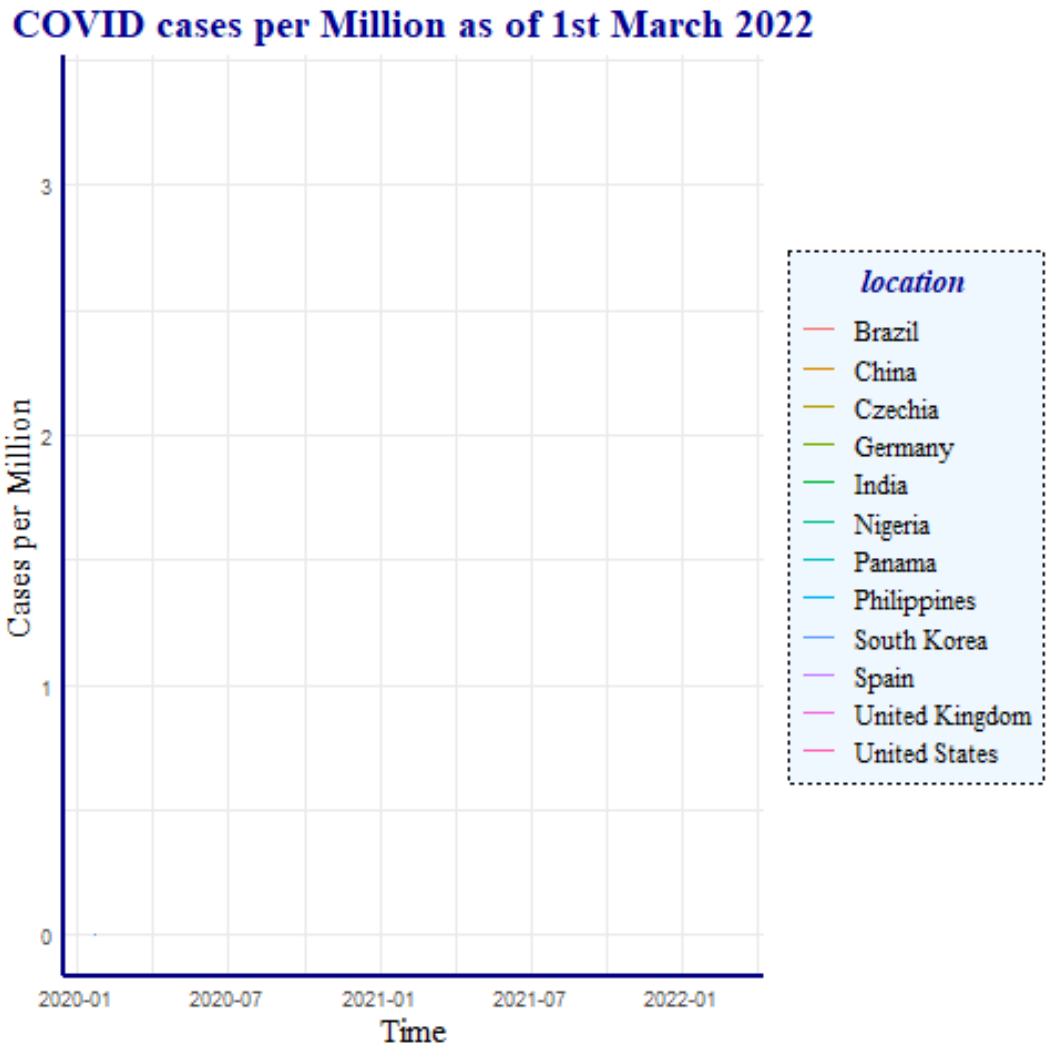

Figure 8: COVID Cases Per Million Citizens During the Pandemic – Animated (Ref [2])

With more than a few countries, line charts can become cluttered, so using R, the facet function can be used to separate out data based on country. Figure 9 takes the same data shown in Figure 6 and 8 and converts the measures into a line plot for each country taking advantage of facets. It is important to use the same y-axis scaling when faceting as to not show an unjust representation of the measured value across different countries. This same facet plot could be done for total cases as well. Figure 9 breaks down the infection rate per million people so the viewer can focus on one country and be able to accurately measure a point on the line. Figures 5-8 tend to show more of a comparison between countries at the expense of value resolution.

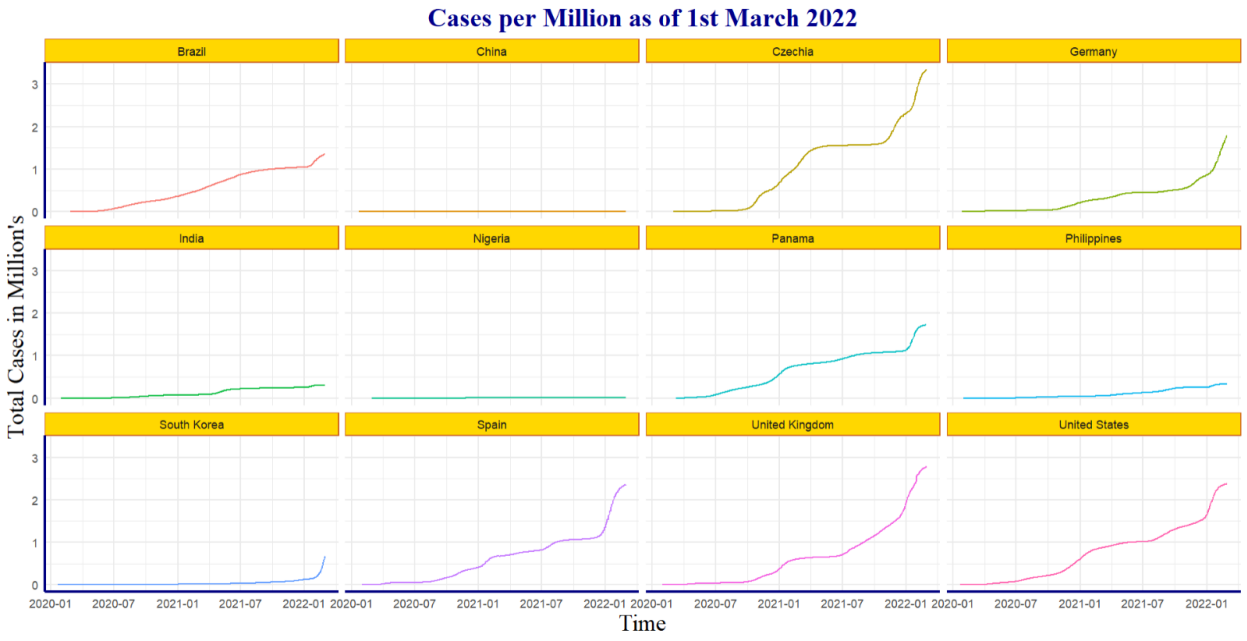

Figure 9: COVID Cases Per Million Citizens During the Pandemic – Facet (Ref [2] data)

Figures 5-9 only show how one variable changes in time, total cases and infection rate per million citizens respectively. However, using R, both measures are able to be shown simultaneously over time. Figure 10 shows the same data as Figures 5 and 6 as a scatter plot with time varying data.

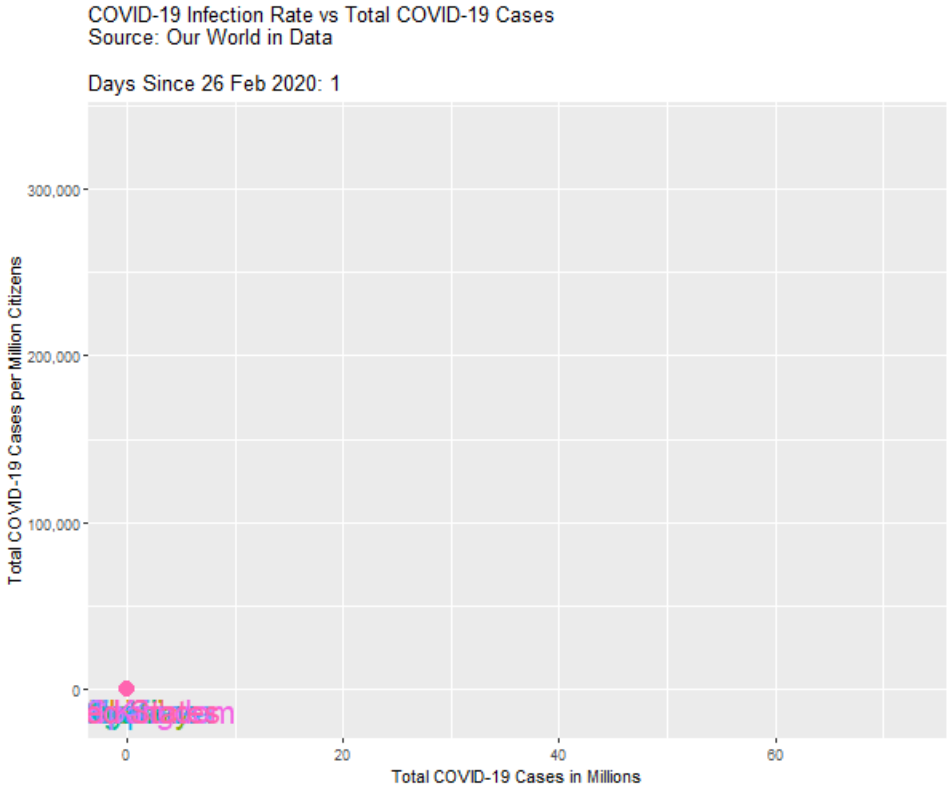

Figure 10: Total Cases and Infection Rate Per Million Citizens (Ref [2] data)

Summary

Figure 1 was limited in the data shown and listed total case numbers instead of showing a comparison between countries; therefore, it was decided that a redesign was justified. Two approaches were taken to accomplish the redesign. The first is a graphical representation to show geographical similarities. The data source in the original graphic only listed the COVID cases and infection rate for a few countries, but when using a global map, it is more meaningful to use data for all available countries. To accomplish this, more data was found and implemented into the redesign project.

The second option was to use the acquired additional data from Feb 2020 through March 2022 to show time varying data. Further consideration was taken how to represent time varying data based on the medium in which the graphic was presented. Line plots were first created both static (Figures 5, 6, 9) and animated figures (Figures 7, 8). The animated graphic can only be shown in digital format whereas the non-animated time varying data is able to be shown in both print and digital media and be understood while still showing how COVID data evolved. It was seen that line charts have their limitations when countries' lines overlap, so using R's facet function separated the countries to allow the viewer to make comparisons more accurately. Finally, the final animated graphic (Figure 10) represents two time varying variables (total COVID cases and infection rate per million people). The original graphic shows a bar plot comparing infection rate and lists the total number of cases. The final graphic was able to implement much more data to show changes in time over both variables in an animated format. Overall, the animated scatter plot (Figure 10), global map (Figure 4), and animated line plots (Figures 7 and 8) are better suited to convey global COVID data trends than the original data graphic.

References

[1] Buchholz, Katharina. "COVID-19 Cases per Million Inhabitants: A Comparison." *Statista.com*, 18 Jan. 2021. Accessed 6 Mar. 2022.

[2] "Data on COVID-19 (Coronavirus) by Our World in Data." 6 Mar. 2022, https://github.com/owid/covid-19-data/tree/master/public/data. Accessed 6 Mar. 2022.