

Abstract

This paper presents some analysis and conclusions for different types of questions related to the COVID-19 pandemic. It utilizes data sets that have been provided on the website Kaggle, where the user, 'SRK,' has been updating the sources daily. The pandemic has been an ongoing problem globally and so the research for this issue has been made available for free to interested researchers. However, since the pandemic is still not over, the research is still in the process of going through refinement and further improvement. This project will attempt to analyze six data visualizations based on data that tracks the spread of the pandemic over time throughout different countries. By analyzing patterns in the spread of the pandemic, different insights will be delivered to help make clearer the current situation regarding COVID-19. The paper examines how the developed world has gone through the most severe stages of the pandemic and are entering the late stage where China is apparently given their plateauing number of new infections. The implication then is that there are still billions of people in the developing world (e.g., India, Africa, South America) whom are susceptible to the pandemic but lack the same means to handle the virus due to poor healthcare infrastructures. The conclusion then is that what has happened over the past months is unfortunate, but the real next steps are for the world to come together and help those who are truly vulnerable right now.

1. Introduction

Currently, with the pandemic first being discovered in Wuhan, China, it has since been found to have spread globally, reaching Europe and North America. There's a great fear that once it reaches the least developed parts of the world (e.g., South America, Africa, South East Asia) that there will be even more devastating consequences. Over a period of just a few months over three million have been infected with tens of thousands of deaths being related to COVID-19. The pandemic has made it such that there's now a global effort for citizens of all nations to work together globally to fight the war against this virus.

The virus itself is genetically like the seasonal flu virus that appears annually. The difference about COVID-19 is that it's especially infectious while having devastating effects for those that suffer serious symptoms. Like the normal flu, patients may experience pneumonia, the difference is that the pneumonia experienced by COVID-19 patients is particularly deadly and has led to thousands of deaths thus far throughout much of the developed world. The most common victims of COVID-19 are elderly patients with weaker systems and are unable to fight off the symptoms. At the same time there are many people who face no symptoms and overcome the virus on their own by developing antibodies.

Given the infectious nature of COVID-19, it's become law throughout nations around the world for citizens to remain in a state of quarantine until the virus passes. The quarantine stage makes it such that the number of patients needing medical care goes down and so hospitals are not flooded with patients and thus overwhelming the healthcare systems of the respective countries. In this paper, the goal is to use data visualization techniques to investigate the dataset and drive insights to better understand the current situation regarding the pandemic's progress.

2. Background

As mentioned previously, the COVID-19 pandemic is an ongoing problem for the world. Therefore, the work done on understanding how to overcome the virus is still a hot topic in the research world. Medical communities globally are lowering barriers for getting access to research related to COVID-19. Furthermore, medical teams globally are working together to find solutions to different problems that are arising that are caused by the pandemic.

To deal with this issue in a practical way, research related to past pandemics will be resourced to help understand how epidemiological techniques can be utilized. It's worth noting that the COVID-19 pandemic is significant in terms of its effect on humanity and has been called a once in a lifetime event (the analogy has been made to the Spanish Flu in 1918 where millions died).

Another resource is the YouTube user Grant Sanderson who is famous for his '3Brown1Blue' channel that uses visualization techniques to understand mathematical concepts. A popular video of his recently seeks to understand the exponential growth seen within epidemics by using data based on COVID-19. Since Sanderson has already created some interesting visualizing methods to explore the current pandemic, time will be spent reproducing some of his work to try and perform analysis on further questions.

Additionally, from this Data Visualization class, many different techniques have been learned that can help to visualize datasets such as the COVID-19 dataset. The one that will be used in the end is the sunburst visualization. This method breaks down the data into hierarchies which is useful for the geographic data.

3. Approach

In this section, the different data visualization techniques will be explained and briefly shown. Six different visualizations were created and so this section shall explain what each of these methodologies are and how they work. The different visualizations are used to drive new insights about the dataset from various angles. Therefore, by understanding this section it'll be possible to then understand the Results section where the different methods are implemented, and insights are derived from the data.

3.1 First Visualization

The first data visualization is a geospatial graph that shows the number of confirmed cases daily for each country. The goal of this visualization is to gain an understanding of the spread of the pandemic over land. The virus was first discovered to have been in the city of Wuhan. This city is in Hubei province, a province within the People's Republic of China. The origins of the virus are still uncertain, but this is the location where cases were first confirmed through testing that had been developed within the country. Trying to understand the spread of the virus globally in terms of confirmed cases is a logical way to understand the severity of the pandemic over time.

Below in Figure 1, it's possible to see the bubble map that was created in Python using the visualization package Plot.ly. The data was first changed from a long format to a wide format so that the dates would move from columns to being rows. Initially, the Bubble Map made it such that each country had a unique color and marker. With over a hundred countries this became cumbersome, therefore a new column was created to indicate which continent each country is from. This makes it possible to see more clearly how the outbreak began in Asia and spread west

towards Europe and eventually North America. As time has gone by, the pandemic has also begun to move southwards towards the developing nations in South America and Africa.

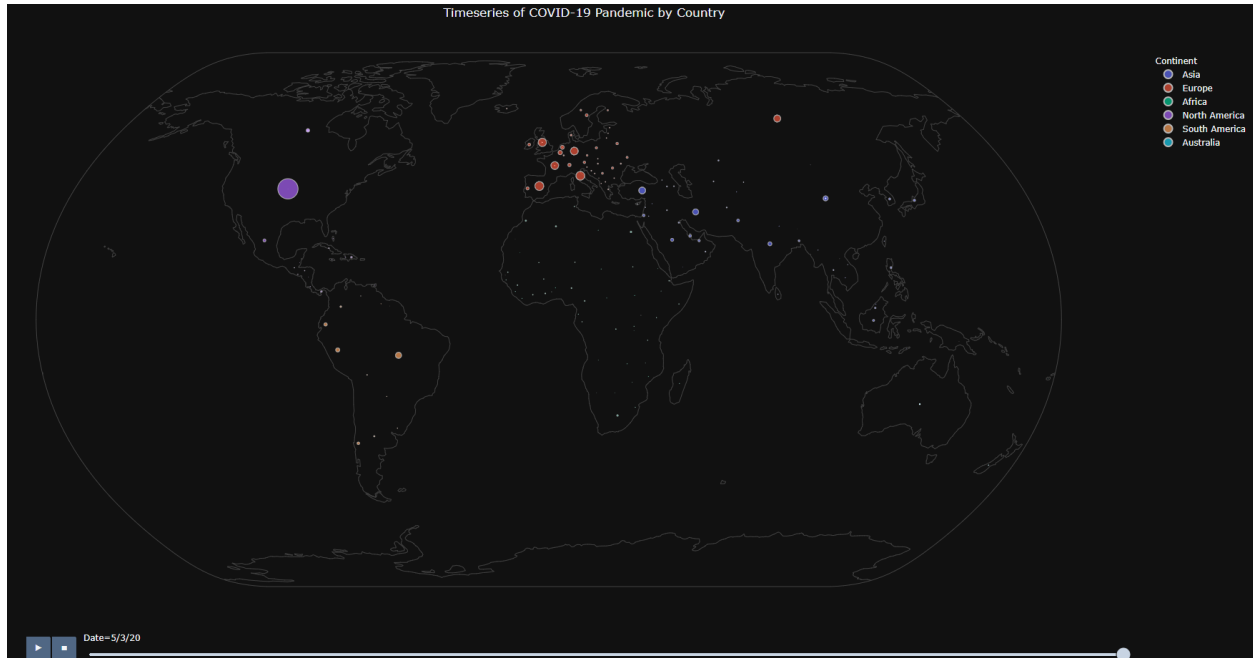


Figure 1 Bubble map of daily confirmed cases by country. The geospatial graph colors countries according to their respective continent. The visualization is an animation where the size of each bubble varies by the number of confirmed cases on a given day.

3.2 Second Visualization

Another way to analyze the growth of the spread of the pandemic is to plot the growth in number of cases over time in a line chart. It would also be helpful to compare side-by-side the growth in number of cases for different countries. Due to the large number of countries, a technique to manage this is to subset the countries so that only countries with a minimum number of cases are included. The value for this minimum number of cases is 50,000. As of May fourth, the dataset pulled from Kaggle includes twelve countries that fit into this category. Another important factor for comparing these countries side-by-side on a line chart is to consider how the spread of the virus to each of the countries had different start dates. Furthermore, at times the number of cases could be quite low, for example less than ten for a long period of time

before testing starts to consider all the cases that are unknown. Therefore, it also makes sense to include a lag of for example ten days after the first confirmed case.

Below in Figure 2 is a line chart created in Python using the visualization package Plot.ly. It shows the number of confirmed cases against the lag days for countries with at least 50,000 confirmed cases. The lag days represents number of days that there have been confirmed cases in the country, minus ten (the number of days to lag the count). For example, China has had confirmed cases since the first day, so they've had confirmed cases for a total of 113 days. Subtracting ten, gives 93 lag days, the maximum number of days in the below line chart. Another country, Italy, has a maximum lag day of 84, so therefore the line for that country isn't as long. This number implies that Italy's first confirmed case began 94 days before. The inspiration for this idea of creating lags comes from online sources, for example Johns Hopkins University has not only a dashboard for the number of cases on a geospatial map, but also a page that shows numerous different Plot.ly visualizations (including one similar to Figure 2).

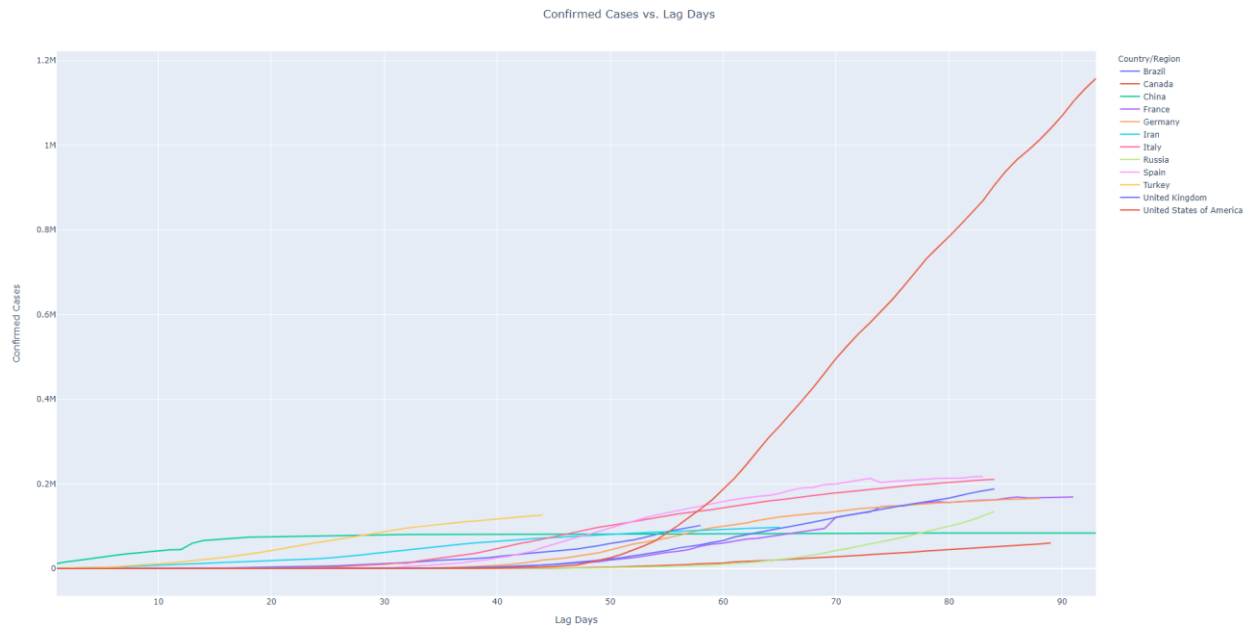


Figure 2 A line chart of the number of confirmed cases against lag days (a feature explained above). The countries with at least 50,000 confirmed cases have been subset from the original dataset to avoid flooding the visualization with too much information.

The country with the largest number of cases at almost 1.2 million is the United States of America. This large value makes it difficult to see in a proper perspective the other countries with fewer cases. However, Plot.ly's chart allows users to zoom in on any section of the chart so that this problem is no longer an issue. Below in Figure 3, this feature is highlighted where the other countries are zoomed in on so that it's possible to compare the growth more easily in number of cases daily.

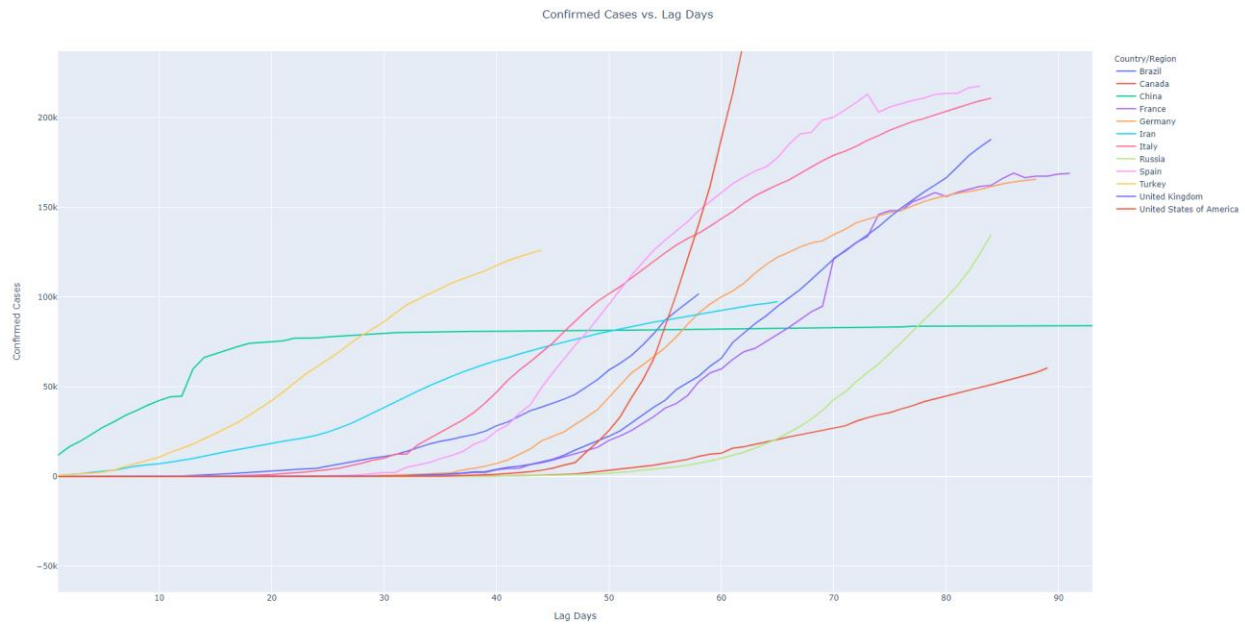


Figure 3 The line chart allows for zooming in on areas and so this image shows all the countries up close after the line representing the United States of America is left out of the image. This feature makes it possible to compare more realistically the other countries and their growth in number of cases.

3.3 Third Visualization

The third visualization is like the previous visualization. The only difference is that the number of confirmed cases has been transformed with the natural logarithm function. When analyzing data such as the COVID-19 confirmed cases, it's helpful to transform the values so that they're easier to visualize. Data such as this tends to follow an exponential growth rate and so taking the natural logarithm helps to scale the data so that it appears more linear. Later, this relationship between exponential growth will be examined further.

Below in Figure 4 it shows a graph like the previous two graphs, the only difference is that the confirmed cases variable has been transformed via natural log transformation. The concept related to lag days are identical and so those variables in the plot below remain the same. Without further analysis, it's interesting to see the data scaled in such a manner, but it's not

particularly intuitive to most people what this implies. However, in later visualizations there will be more similar plots except that there will be deeper analysis of the scaled data.

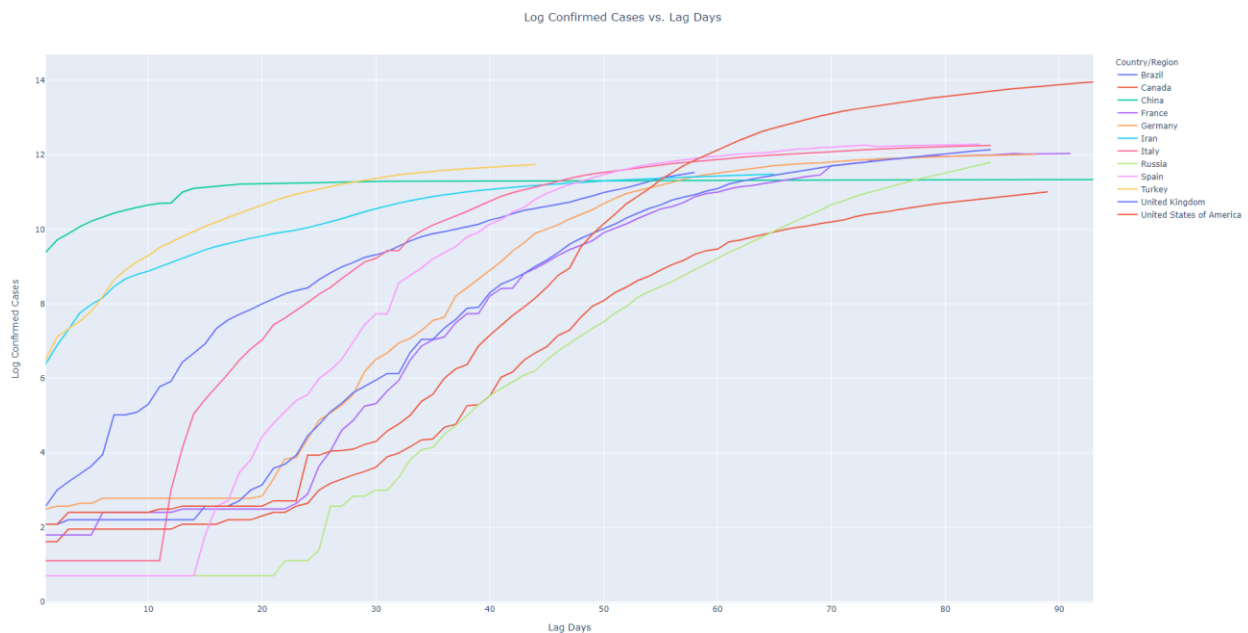


Figure 4 The graph shows the confirmed cases after being transformed by the natural logarithm against the lag days. The countries have been subset so that only those with at least 50,000 current cases are shown.

3.4 Fourth Visualization

The fourth visualization is based on some concepts explained by Grant Sanderson from 3Blue1Brown. The following few visualizations will revolve around ideas that he elaborates on regarding exponential growth and the pandemic. In the third visualization, the number of confirmed cases was transformed using the natural logarithm. In this presentation, it will become more apparent the reason why this was done.

In Sanderson's YouTube video, he explains how the number of cases of COVID-19 patients increases with an exponential growth. This growth is explained where the number of new cases increases by some value slightly over 1. In other words, for every person that is infected, they

infect $1.5x$ new people, where $x > 0$. This is a natural phenomenon seen in different environments and therefore it makes sense to perform some modeling based on this characteristic. To be more specific, imagine that the growth rate is,

$$\exp(1.5x + 3),$$

where x is the n' th day of infections. Then on day 1 there are 4.5 cases, on day 2 there are 6 cases, on day 30 there are 48 cases, etc. By transforming this growth rate with the natural logarithm, it becomes $1.5x + 3$, the formula for a linear equation. Using simple linear regression, it's possible to see how well this fits a straight line by analyzing the R^2 value. Admittedly, this is an ad-hoc trick and there are more complex methods for determining what sort of exponential growth that the data is going through.

Below in Figure 5 is a grid of the 12 countries with at least 50,000 as of May 4th, 2020. Furthermore, the confirmed cases have been transformed through log transformation and the days have been changed to lag days. Lag days in this case is the same as the previous visualization with 10 lag days being used since the first confirmed case for each country. On the bottom of each plot it shows the Adjusted R^2 values for each of the fitted lines derived using simple linear regression. Values that are high (≥ 0.9) indicate that the relationship between the log number of cases and lag days is highly linear. Examples of this can be seen in Russia, United Kingdom, United States of America, and Canada with Adjusted R^2 values of 0.9692, 0.9462, 0.9453, and 0.9548.

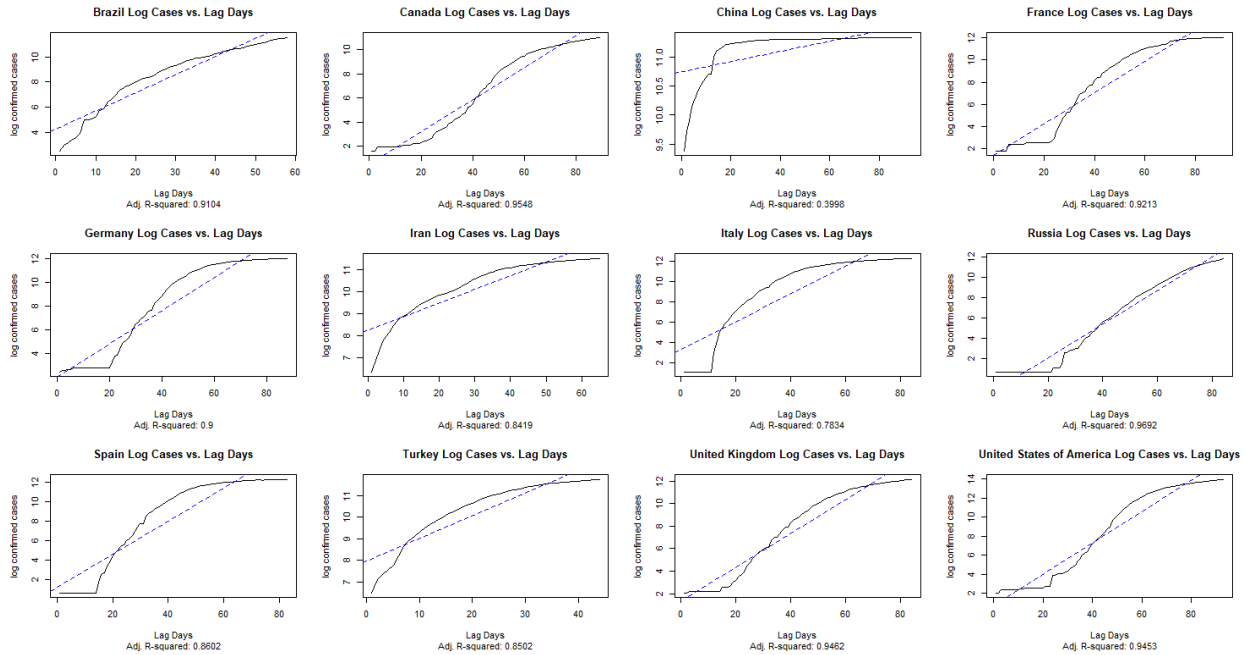


Figure 5 A 3x4 grid of the 12 countries with at least 50,000 confirmed cases. The grid plots the log number of cases against the lag days. A linear regression line is fitted to each of the plots and their corresponding Adj. R-squared values are shown at the bottom of each plot.

An example of how this method can be effective is to look at an example where it seems to not work. In this case, China has an Adj. R^2 of 0.3998, an extremely low value indicating a non-linear pattern in the data. Below in Figure 6 on the right-hand side shows a subset of China's data of only the first 20 days since infections began. In this plot, the Adj. R^2 is 0.9459 and so the data is much more linear than before. Over time, as the growth ceases to continue exponentially, the data will begin to appear more sigmoidal rather than exponential. This sigmoidal appearance is evident on the left-hand side of Figure 6 that shows China's plot from Figure 5. This concept is elaborated on in the next visualization. However, this contrast shows how the spread of the pandemic can be understood to be in exponential terms for some time before beginning to plateau towards the next stage which is sigmoidal.

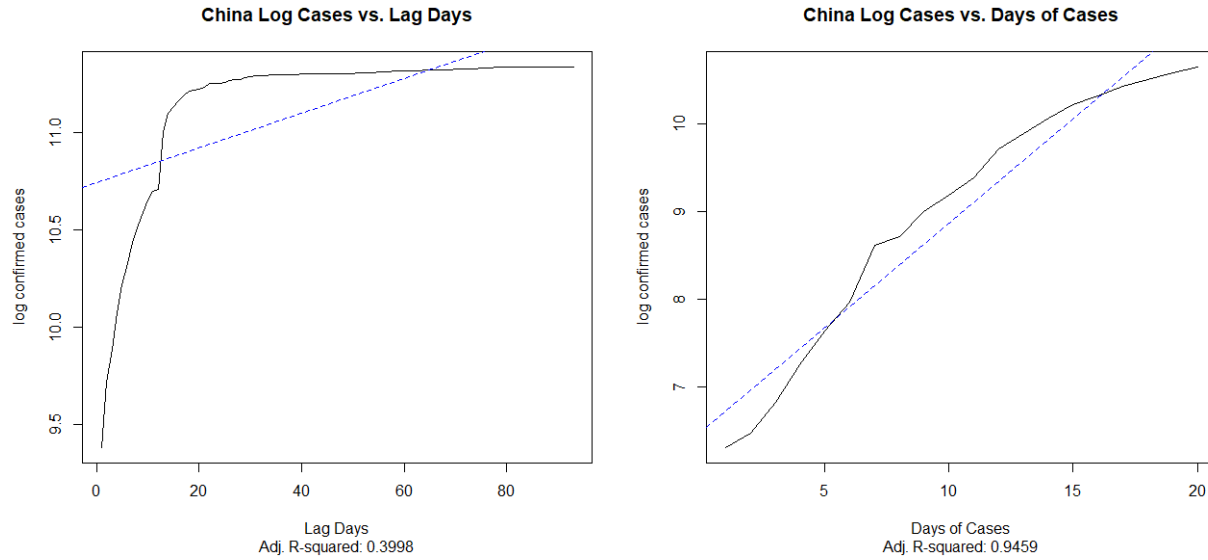


Figure 6 On the left-hand side is the plot of China's log confirmed cases with 10 lag days from Figure 5. On the right-hand side is a subset of China's data showing only the first 20 days of confirmed cases. Applying simple linear regression to this subset shows a much more linear appearance in the data.

3.5 Fifth Visualization

The fifth visualization expands on the idea of the exponential growth. In exponential growth, the growth rate continues to be growing at a rate of $1.x$, where $x > 0$. However, after some period the value of x should start to shrink until the total rate becomes less than 1. When modeling pandemics, the following formula is used (in the Figure 8 plot it's called the growth rate),

$$\text{Growth factor} = \frac{\text{New cases on current day}}{\text{New cases on previous day}}.$$

During the beginning of the pandemic, this growth rate starts at 1 and increases by some decimal points until it reaches the inflection point of the sigmoidal curve and the growth rate starts to go back to 1. After the inflection point, the growth rate starts to decrease lower than 1 as the growth

decreases. This sigmoidal curve has an “S”-shape and the inflection point is in the middle. Below in Figure 7 is an image of the sigmoidal curve that models the growth in number of cases.

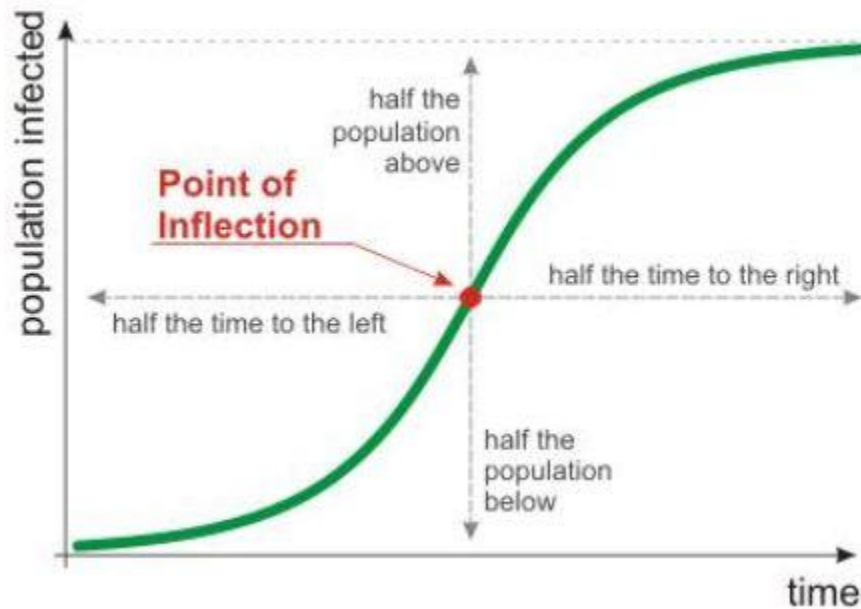


Figure 7 Image of sigmoidal curve that models the spread of a pandemic. The focus is on the growth rate which begins to decrease at the point of inflection. (Source: nctm.org)

The idea behind the fifth visualization is to create a plot of the growth factor to try and monitor the growth rate for each country. By plotting the growth rate as a vertical line and the days on the x -axis, it's possible to see the growth factor for each day in a country. The red lines indicate when the previous day had 0 new cases, which defaults the growth rate to infinity since it's dividing by 0 in RStudio. Also, on certain days the growth factor is too large, and the plots lose perspective, therefore the growth rate is capped at 5 for each of the plots.

Below in Figure 8, it shows the fifth visualization that plots the growth rate against the days of cases (without the lag). It shows in a grid for all twelve countries with at least 50,000 confirmed cases and their growth rate plot. For many of the countries, their growth in number of cases didn't start for some time until after China first experienced cases. Therefore, they have a

growth rate of 0 for long periods of time. The plots show that countries will experience a growth in cases that is above 1 for some period. This shows that for each consecutive day, the total number of new cases is slightly greater than the previous day's number of new cases. Relating this to Figure 7, this indicates that they're still in the first half of the pandemic. Once the number of cases start to level out at 1, it implies that they've past the inflection point and the number of new cases is beginning to slow down.

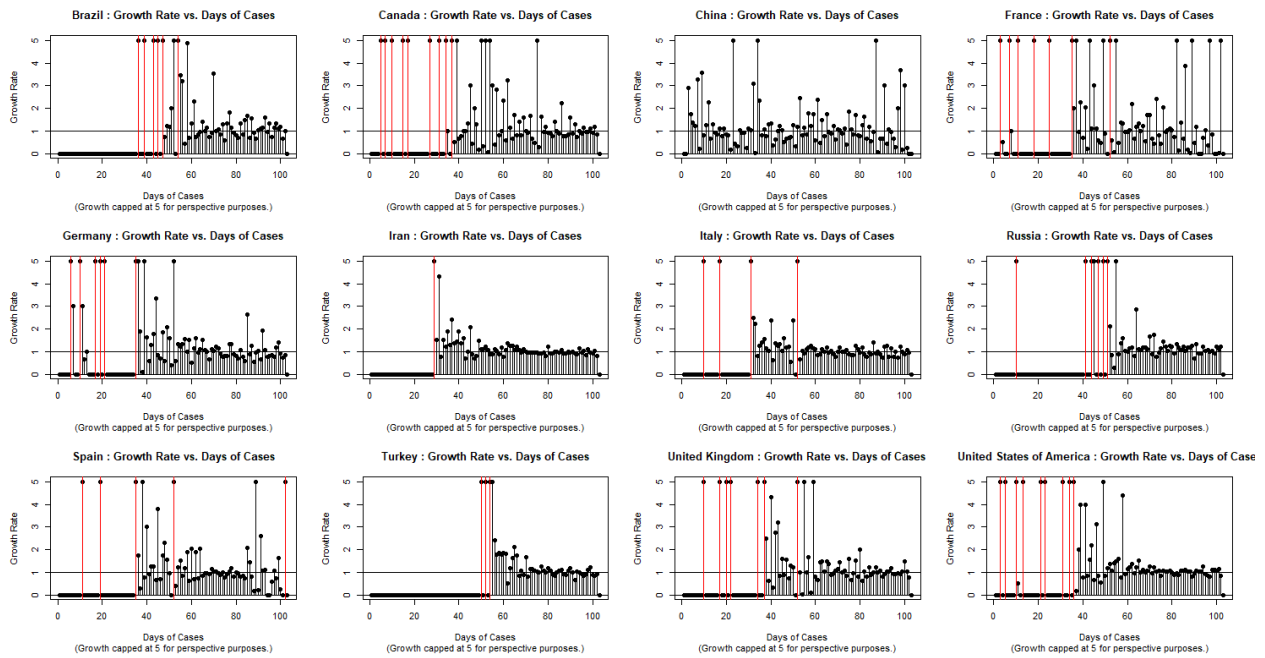


Figure 8 A grid of the twelve countries with over 50,000 confirmed cases and their growth rates. The red lines signify that the growth rate is infinite due to a division by 0 calculation in RStudio. The plot is also capped at 5 so that the perspective is not too heavily skewed by certain outlier dates.

3.6 Sixth Visualization

The sixth visualization is a sunburst that allows for better understanding of the hierarchical data structure of the dataset. The idea is to understand the spread in number of cases based on the

geographic region. The first visualization performed a similar task, but it wasn't simple to determine the relative number of cases by continent since the bubbles in the bubble map are spread far apart. The sunburst visualization allows for a radial view of the number of confirmed cases, breaking down the data's hierarchy into from continent to country or region. This simpler view of the geographic spread allows for interpretation such as how the spread is largely confined to developed countries such as European countries, the United States of America, and some Asian countries (whom may be developing but still have a comparatively more developed infrastructure).

The sunburst visualization of the dataset can be seen below in Figure 9. As mentioned before, it's apparent that the developed countries with high levels of people movement across borders are the ones with the most infected. However, poorer countries where the population is largely poor and there is less movement across borders have much fewer cases. The areas of Africa and South America are of concern, due to their poorer healthcare infrastructure and a potential inability for the governments to handle the pandemic in a manner that more developed countries have.

Sunburst Hierarchy of Confirmed Cases

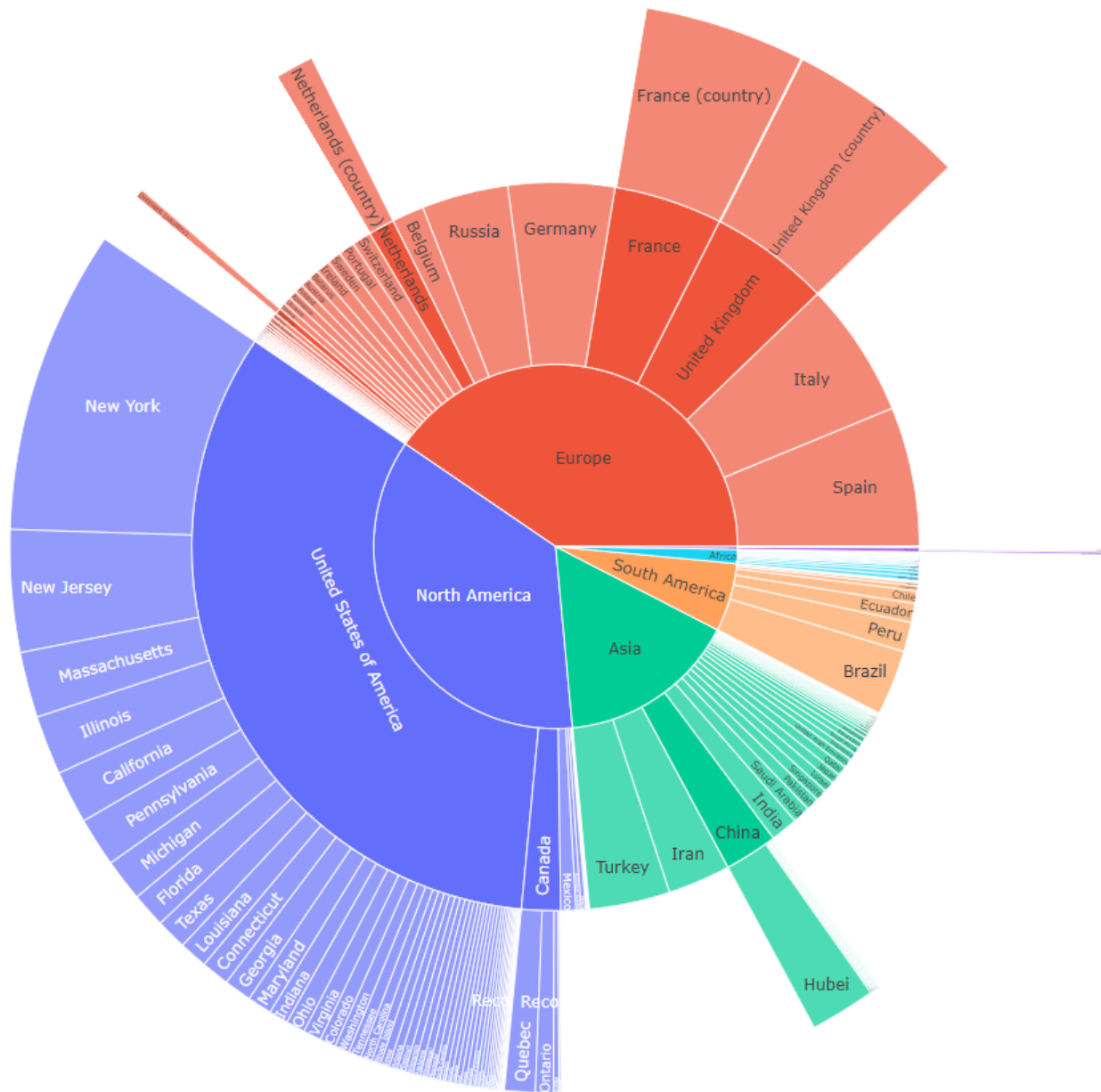


Figure 9 A sunburst visualization of the confirmed cases. The hierarchy of the sunburst is to go from Continent to Country / Region to Province / State.

Looking closely at Figure 9, it's apparent that there are many small slices of the visualization that are difficult to analyze. This visualization however has a feature that makes it possible to

click on a level within the hierarchy, for example, North America, and it will allow the user to zoom in to see the different pieces more clearly. This can be seen below in Figure 10.

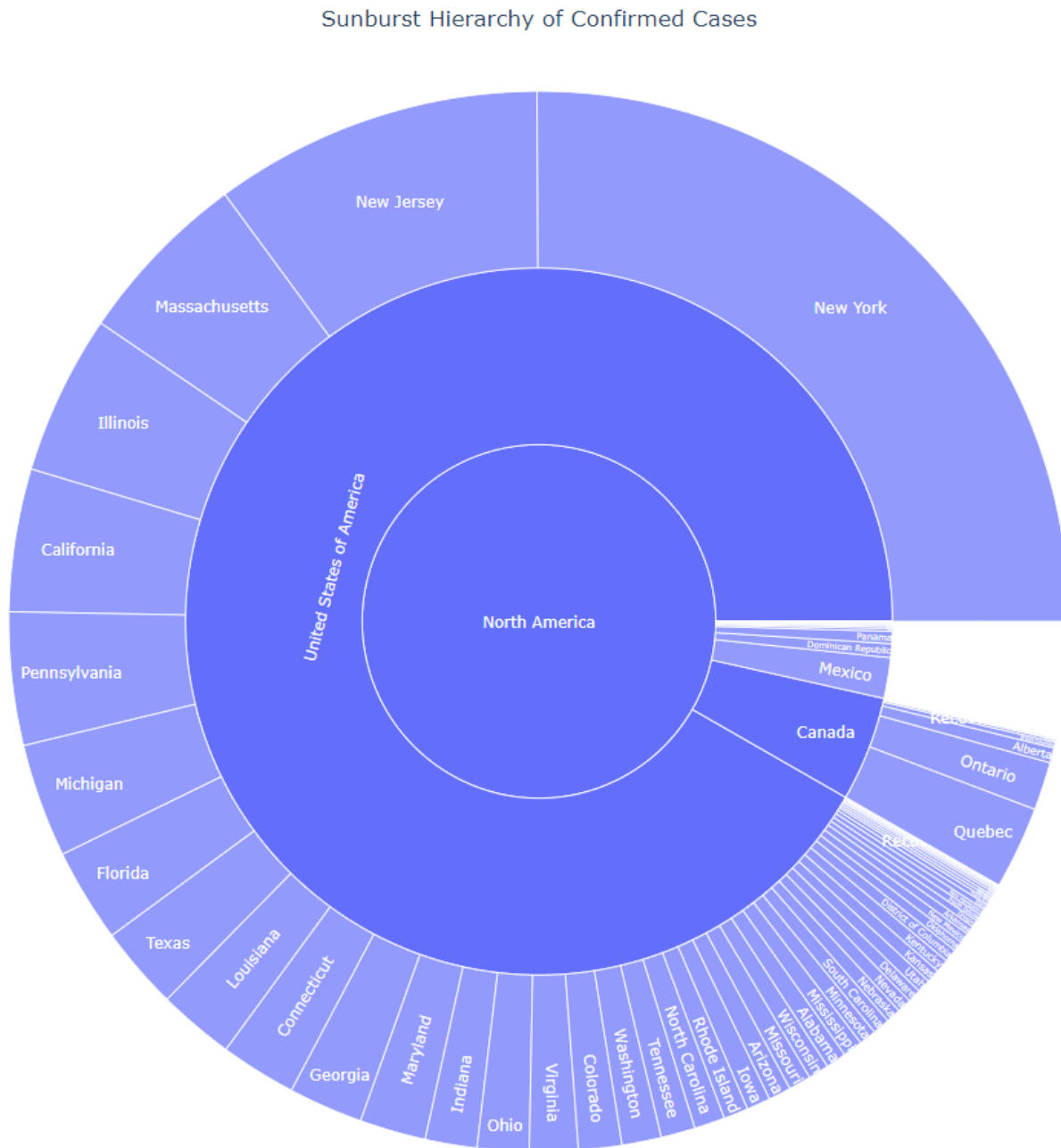


Figure 10 The above shows the same visualization as in Figure 9, except the 'North America' level of the data's hierarchy is selected. This allows the user to zoom in on successive levels of the dataset to see more clearly the pieces that maybe too small to notice in comparison due to the scaling of the numbers in relation to the size of the slices.

The sunburst visualization was created in Python using Plot.ly. Furthermore, it used an alternate dataset from the series of datasets that were part of the database on Kaggle. This dataset had better details related to the variables of: Province/State, Country/Region, and Confirmed. This dataset is called 'covid_19_data.csv,' while previously a modified version of 'time_series_covid_19_confirmed.csv' was used.

4. Results

4.1 Analysis of First Visualization

The goal of the first visualization is to analyze the spread of the pandemic through time and space. In other words, the idea is to find out where and when the virus has been since it was first discovered in the city of Wuhan within Hubei, China. Looking below in Figure 11, it's apparent that up until 3/6/2020, the virus was restricted to mainly the borders of China. Less noticeable are the countries of South Korea, Iran, and Italy with between 4,000 and 7,000 cases each. It's quite logical why South Korea would be so susceptible since they're nearby neighbors. However, a possible reason for why Iran and Italy also large numbers of infected are because they have cross-border travel between the country. It's possible that some infected people had traveled back and forth from Wuhan to Iran and Italy.



Figure 11 The screenshot shows the bubble map of confirmed cases on 3/6/2020. At this point, the spread of the virus is largely contained within China. There are other countries with notable amounts of cases such as South Korea, Iran, and Italy. These countries are highlighted using arrows and text in the image.

Below in Figure 12 it shows the bubble map on 3/23/2020, roughly 3 weeks after the previous screenshot. By this time, the spread has gotten more serious in Europe with major countries like the United States, United Kingdom, France, Spain, and Germany experiencing tens of thousands of cases. The spread in the previously infected countries are also increasing. However, it's interesting to see that Iran has increased much more rapidly, reaching over 23,000 infected, while South Korea has only reached about 9,000 infected. Some logical reasons for the difference in the increase of cases is that South Korea has a significantly more developed economy. Furthermore, besides the government's ability to implement useful prevention measures, the culture is quick to practice habits such as wearing masks and social distancing to help slow the spread of infections.

Although the number of cases is still comparatively low, Japan and Malaysia have broken the threshold of one thousand cases. These are both countries that are roughly nearby to China. Of special concern is Malaysia, which is a country with a less sophisticated healthcare system and thus is likely less prepared to handle the effects of the pandemic spreading throughout their

society. It's logical to think that from Italy, the pandemic could've easily spread to other European countries due to the heavy traffic flow between the states within the European Union. The spike in number of cases for the United States is also worrisome due to the similar danger of heavy traffic flow from major economic hubs across large distances (e.g., infected people traveling from New York to California).



Figure 12 A screenshot of countries that are beginning to develop a larger number of confirmed cases. This screenshot is taken from the bubble map on 3/23/2020, roughly three weeks after the previous screenshot. It shows how other developed countries have become to have a more significant number of confirmed cases. The developed countries include: United States, United Kingdom, France, Spain, and Germany. Furthermore, other nearby countries to China, Japan and Malaysia, are starting to develop more cases.

The screenshot below in Figure 13 shows the number of cases per country on 4/1/2020. By this time, a week after the previous screenshot, the cases have now spread to most of the developed world. The major European countries are now near or over 100,000 confirmed cases. The United States has now become the new country with the quickly increasing number of cases at over 630,000. China has since peaked for some time and has remained at around 80,000 cases.

Evident now are other countries that are beginning to develop many cases. The countries of Brazil, Turkey, India, and Russia now have tens of thousands of cases. These countries are

poorer than the developed western countries of the United States, United Kingdom, France, Germany, etc. Therefore, there's now a significant concern over whether they can handle the combination of a large population and a highly infectious virus. Australia is now in the thousands and is starting to look like it'll grow quicker too. However, it's an island country and so it's difficult to know how the spread of the virus will take place there given its unique geographical characteristics. Another island nation with an increasing number of cases is the Philippines, however it has a comparatively weaker health infrastructure, so the concern is even greater. A major worry is South Africa with the number of confirmed cases in the 2,000 range. Egypt (not shown) also has a similar number of confirmed cases, however, it's quite close to Europe which has been having an exploding number of confirmed cases. The reason is that South Africa is possibly a sign that the rest of the continent is also not safe and there's a great risk that many extremely impoverished communities will become victim to the virus without means to overcome the pandemic.

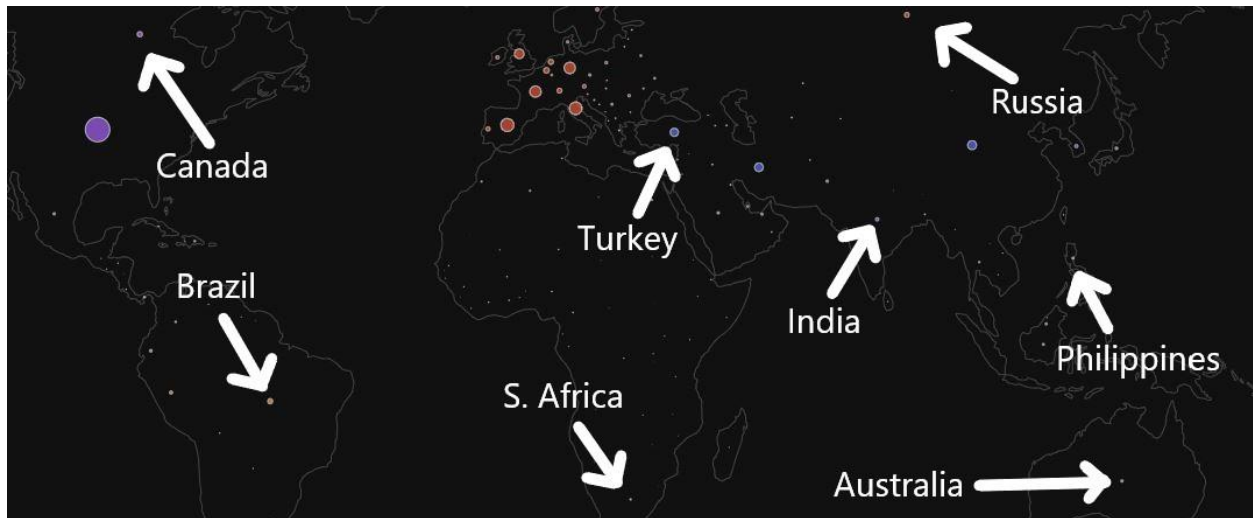


Figure 13 A screenshot of countries as the number of cases have not only spread, but are also growing to numbers in the hundreds of thousands. More distant parts of the world relative to China are now developing large numbers of cases. This screenshot represents the number of cases on 4/1/2020.

The final screenshot from the bubble map can be seen below in Figure 14. It shows the latest update for the spread of the pandemic for countries on 5/3/2020. As of this point in time, the number of confirmed cases in the developed western world (e.g., European countries and the United States) are skyrocketing. Cases in major European countries (e.g., U.K., France, Spain, etc.) are in the hundreds of thousands. The United States has over one million confirmed cases. Cases in Asian countries, with the major exception of China, are steadily increasing as of note is India. China and India are both countries with equally enormous populations and a developing infrastructure. However, China managed to stem the spread of the virus early on by imposing a lockdown on the epicenter. India on the other hand is in a position like the western world where they knew of the pandemic but were slow to act. With a population of over a billion citizens, there's the real danger that the current count in the 40,000 range could grow to incredibly high levels as such a dense population seems especially susceptible to the effects of a highly infectious virus.

In Figure 14, highlighted however are countries within the Caribbean, South America, and Africa. The number of cases in these areas are now growing to numbers in the thousands. This is the position that western countries previously saw themselves in and given the fate of those areas it's possible that these more remote parts of the world will face a similar future. The countries previously mentioned have had the means to obtain medical supplies on their own, but these countries are poor and will require the assistance of international organizations to help them in their battle against the pandemic. Furthermore, South America and Africa have a large population. Although the flow of people and goods maybe slower due to weaker infrastructure if the virus manages to reach every region there's a lack of means for these areas to overcome the virus. It's problematic if the developed world can overcome the pandemic, but the southern countries are left to be devastated by the virus running for years to come.



Figure 14 The latest screenshot from the dataset on 5/3/2020. It shows that as the number of cases in the developed world skyrockets into the hundreds of thousands, the pandemic is reaching the undeveloped countries. Currently, countries in the Caribbean, South America, and Africa are now seeing cases in the thousands and tens of thousands. This pattern was seen before in the developed world and if it continues the consequences can be devastating.

4.2 Analysis of Second Visualization

The second visualization is the line chart that includes a lag of 10 days after the first infection for each of the countries. Furthermore, the countries have been subset to only include those with at least 50,000 confirmed cases as of 5/3/2020. The first part of the analysis is to look at China, the country where the virus was first detected. Below in Figure 15, the arrow shows a large spike in the number of cases. However, this is associated with the number accounting for clinical diagnosis by doctors, where patients are confirmed cases through a doctor's observation, rather than through some sort of chemical test. Before this time, the number of confirmed cases were already jumping sharply to over 40,000 confirmed within 10 lag days. However, China is known to be the country that quickly implemented some of the strictest measures to stop the virus, including placing a lockdown on the city of Wuhan only days after the seriousness of the virus became known to authorities. Furthermore, medical personnel from across the country entered Wuhan to help manage the confirmed cases, and temporary hospitals were quickly set up to handle the situation. It could be said that the country had learned lessons from the previous SARS virus and so they were better prepared.

Quickly after this timely reaction to the virus by the Chinese government, the number of cases could be seen to begin a stage of plateauing. By around 30 lag days (40 days since the first confirmed case), the number of total confirmed cases hits 80,000 and the growth in number of cases rapidly slows down. It could then be said that when the clinically diagnosed cases were included into the confirmed cases at around 10 lag days, this is possibly the inflection point for when the virus spread through the country. By this point, half the population of infected had been infected, and the other half is left to be infected over a period (refer to Fig. 7). This pattern is highlighted using the gray S-shaped line that represents the sigmoid curve of a pandemic. The

only concern now for China then is to ensure that a second outbreak of the virus doesn't place the country back in danger.

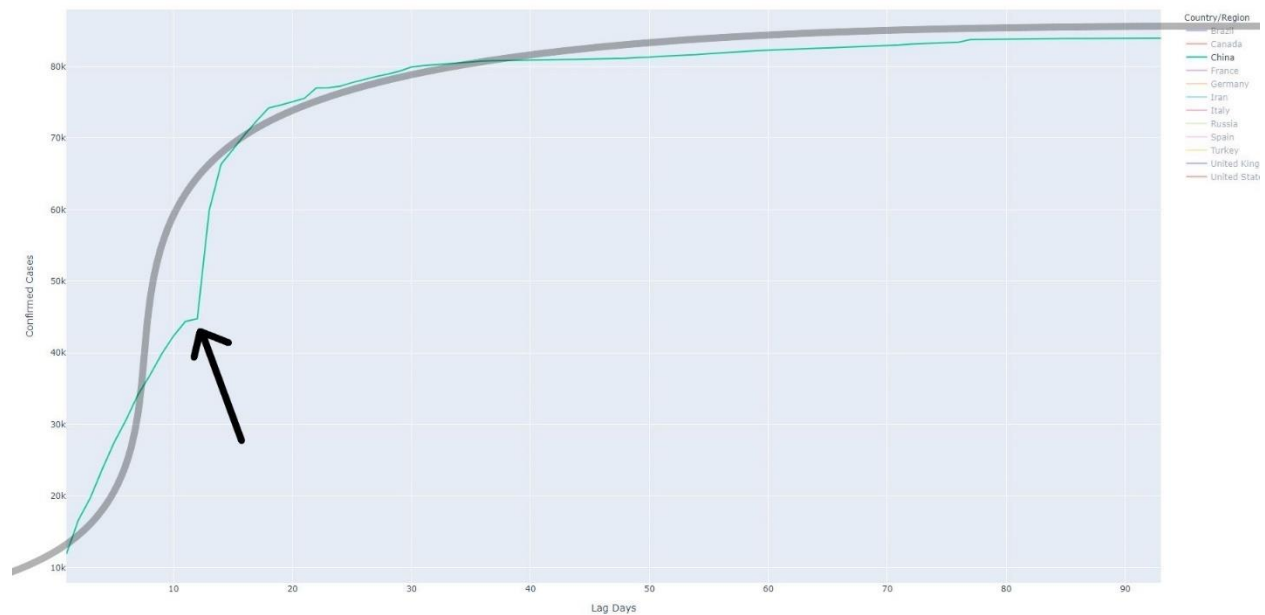


Figure 15 In the above screenshot, the country of China is highlighted out of the twelve countries. An arrow shows a spike in cases caused by the inclusion of clinically diagnosed cases. The gray S-shaped curve shows the sigmoid curve of a pandemic.

Based on the lag days, it's possible to compare countries that have a similar projection. Normally, there's be a significant lag between countries that have a similar trajectory since the virus first entered the countries on different days. By creating a lag of 10 days after the first confirmed case and counting from there, it's possible to see side-by-side the growth in number of cases in a logical manner. This example can be seen in Figure 16, where the following countries are shown: Spain, Italy, Brazil, United Kingdom, Germany, and France. The country of China is included as a reference point to maintain perspective on the growth in number of cases. It was shown previously that the S-shaped sigmoidal curve could be seen in China's growth pattern. The same half-opaque gray line can be seen in Figure 16. It shows that the European countries are following a similar pattern, and Brazil could also be on the same path. Two parallel red lines

show the channel where the countries enter this growth phase and eventually, they must exit the channel if they are to follow the S-shaped trend.

The plot shows that Italy and Spain, the two European countries most effected by the pandemic, are at the upper end of the channel but are still following the curve to eventually exit the channel. At the bottom of the channel are Germany and France. It seems then that they've had a comparatively slower spread of the virus and so they can begin plateauing at a lower level. It's worth noting that this is still about twice the level of where China began to plateau (160,000 compared to 80,000). In the center of the parallel trend channel is the United Kingdom. Looking carefully, it seems to be trending in a slightly different S-shaped pattern as the previously mentioned European countries. The other countries seem to be on a direction of plateauing, but the United Kingdom still seems to be moving upwards relative to the S-shaped sigmoid curve. It's possible that it could begin to plateau later than the others. A reference for the United Kingdom is France, where their trend began at a similar point in time (on the lag day scale). However, France has a less steep curve, while the United Kingdom has a slightly steeper curve. This difference has a greater effect over a longer period, and so it's worth noting that the United Kingdom is perhaps going to be a bit slower to enter the plateau phase.

To take a second look at Brazil, it's interesting that the country seems to be following the same S-shaped trajectory as the major European countries. Brazil on the other hand is a major country in South America. Also, in terms of the country's economic development, it's behind the European powers in terms of GDP per capita (\$8,000 range versus \$30-40,000 range, based on a Google search). Also, it could be said that the healthcare infrastructure there doesn't compare to the modernized hospitals of the European countries. So, although Brazil seems to be on a similar trend, this is perhaps an underestimate of the true numbers that Brazil may face in the future.

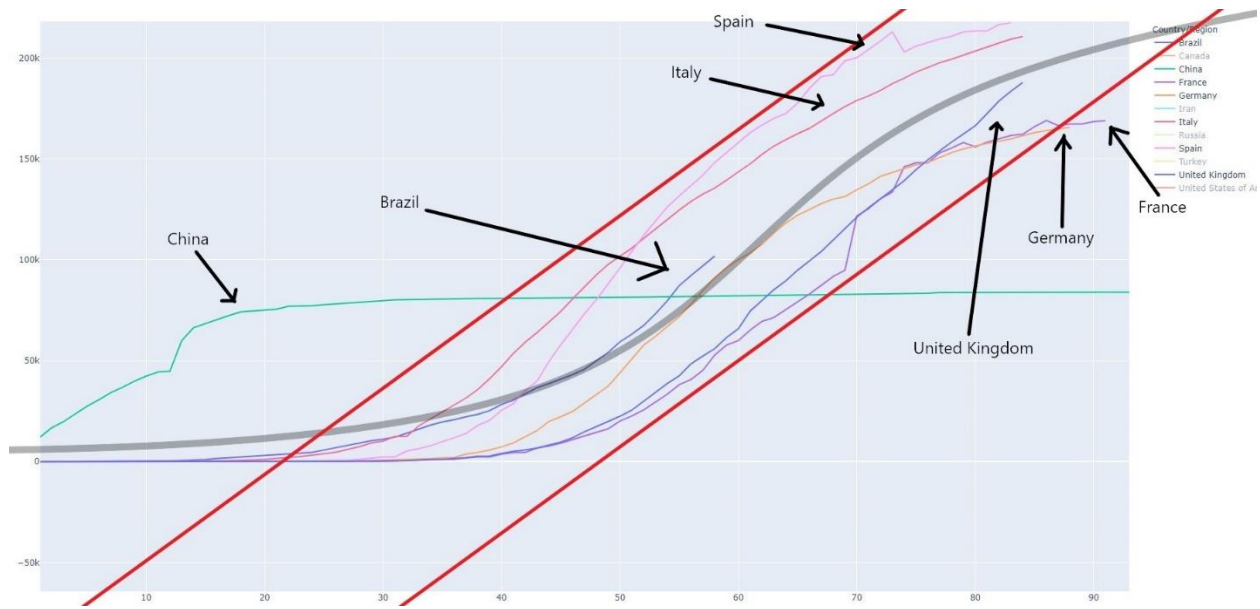


Figure 16 In the above screenshot, the countries with a similar S-shaped curve projection are highlighted. These countries include the European powers of Spain, Italy, United Kingdom, Germany, and France. The South American country of Brazil also seems to potentially be on a similar projection. The country of China is used as a reference point for the scale of the number of confirmed cases. The gray line represents the S-shaped sigmoid curve and the parallel red lines represent the channel for which countries would trend while moving along the sigmoid curve.

The next screenshot shows the other countries not included in Figure 17 (apart from the United States). These countries follow a different trajectory than what was seen before with the European countries. These countries include: Turkey, Russia, Iran, and Canada. Like before, China is included as a reference point for the scale of confirmed cases. For all these countries, a gray line shows the potential S-shaped trajectory that they maybe are on.

Both Turkey and Iran seem to be on a steady trend to reach a plateau at some level possibly not too far from where they are currently. In other words, they may have already reached their inflection points and are on the second half of their journey through the pandemic. It seems that their experience with the pandemic is being controlled and they may meet a similar end as France and Germany where they plateau at a similar level of around 160,000 cases.

An interesting exception is Russia, a neighbor of China to the north. The trajectory of Russia seems to be sudden and steep. For many days, the country had few cases as an exception to what was seen in other European countries. However, geographically, Russia is quite different from what is seen in the other European countries. The western part of Russia in Moscow where the outbreak is serious is further east from the other countries such as: Germany, France, and Italy. Perhaps this distance could be a cause for why Russia is late to experience a growth in cases. However, of concern is the steep trajectory that the country is experiencing, if this growth continues without some sort of intervention, it could be dangerous for the country's healthcare infrastructure as there'd be concern over whether their hospitals would be flooded.

The country at the bottom is Canada and it seems that their growth is slow but steady. Canada is comparatively quite remote, located across the Pacific in North America and having large distances of land where there are many less dense areas makes possibly for a slow spreading pandemic. This is a good sign for their healthcare system as there's less of a chance of them being overwhelmed with cases. However, it's important for the country to maintain protective measures so that the number of cases plateaus eventually without dragging on for too long. Another worry for Canada is the fact that it's the neighbor of the United States to the north. The United States is currently the country with a skyrocketing number of confirmed cases and despite temporarily closing borders, if these measures are relaxed there's the risk that Canada would soon become overwhelmed.

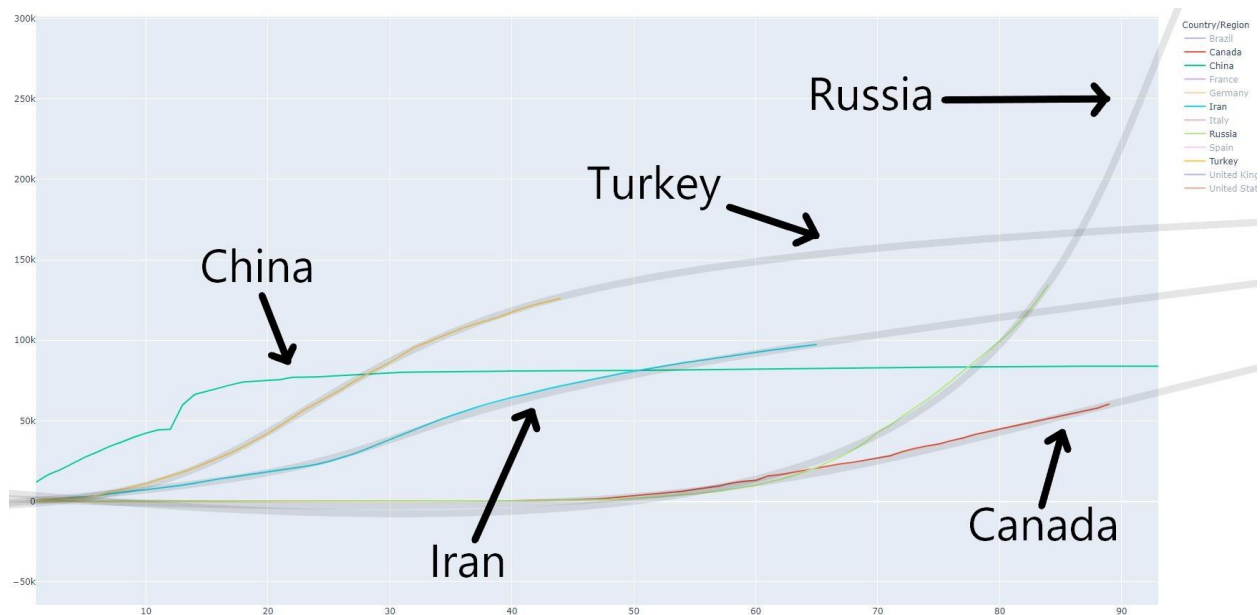


Figure 17 In the above screenshot it shows the countries of: Turkey, Russia, Iran, and Canada (with China for reference). Each of their prospective S-shaped sigmoid curve trajectories are shown.

The last screenshot for analysis from the line chart is regarding the country with the most cases, the United States of America. Below in Figure 18 it shows the comparison of confirmed cases against lag days for all twelve countries with at least 50,000 confirmed cases. A straight gray line is drawn through the line that represents the United States. It's interesting to note that while other countries tended to have some sort of S-shaped trajectory, the United States is the only one to have an obviously linear projection in terms of the number of cases. It's uncertain how this can be explained, but the number of confirmed cases doesn't represent the number of true cases, since there are likely many people who never get tested. Another factor is the healthcare system of the United States which is different from other countries. There are many in the country who lack healthcare and therefore can't afford treatment. In other words, they'd rather deal with the symptoms themselves rather than seek help from a hospital. To imagine the true number of cases, it'd make sense for the true current trajectory to be steeper. Therefore,

given that the true number of cases is steeper, then it could be imagined that a more accurate S-shaped sigmoidal curve could then be mapped on top of the line to project the path of the spread of the virus. The country of China is also highlighted to give reference to the scale of the number of confirmed cases.

On the other hand, another view could be that since the United States is one of the most developed countries in the world with the highest GDP, it could be reasoned that this is a good sign. For example, a linear growth in number of cases is much preferable to an exponential growth in the number of cases. Perhaps social distancing efforts and lockdowns by government officials has made it possible for a country like the United States with its large population to handle the influx of pandemic patients in a manageable manner. This is perhaps related to the concept of “flattening the curve,” where citizens are encouraged to practice safety measures to avoid hospitals being flooded with patients. Therefore, the S-shaped sigmoidal curve for this country is different from others in that it could be less steep, but quite sharp at the plateau.

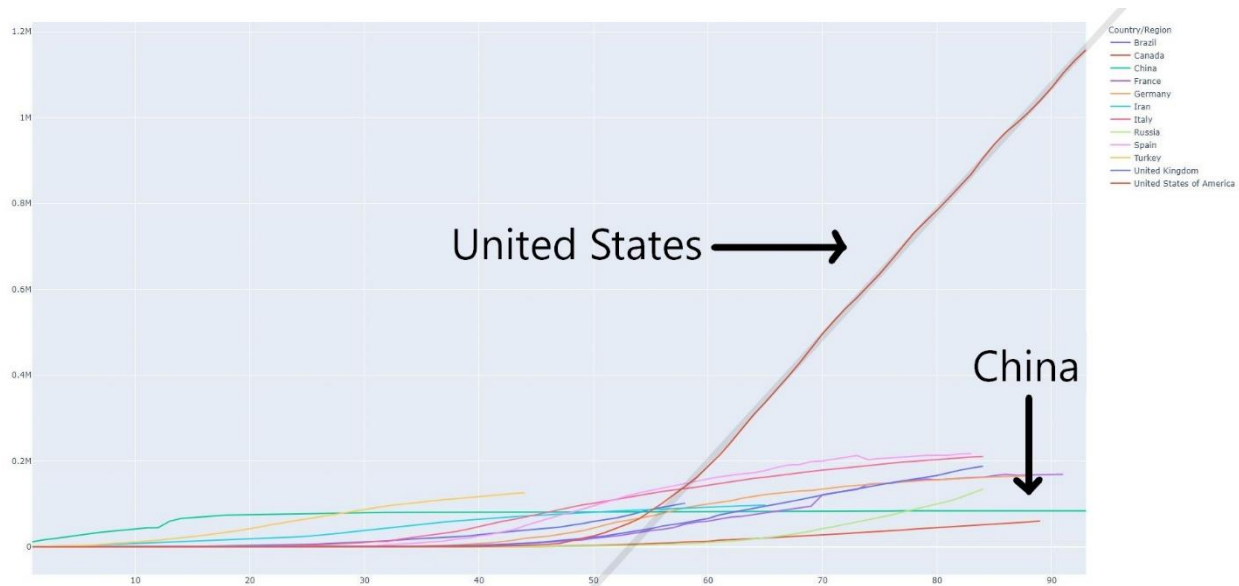


Figure 18 The above screenshot shows the number of confirmed cases for the United States of America. A straight gray line is drawn through the line representing the United States to indicate that the growth has been linear. The country of China is also highlighted as a reference for the scale in the number of cases.

4.3 Analysis of Third Visualization

The third visualization is the same as the second visualization, except for that the y-axis has been scaled using the natural log transformation. The first point to be made about this plot is by looking at the bottom left corner many countries are horizontal for a long period. This is highlighted in Figure 19 by a black box around those countries with a horizontal or near horizontal trend during this period. It means that many countries had a relatively few numbers of confirmed cases for almost three weeks past the first confirmed lag case. This is an interesting insight in that it possibly could imply that due to slow or late testing, the numbers were already spreading rapidly for three weeks, but these cases were amongst the population who were not yet aware of the existence of the virus within their community. The countries within this box include: Canada, France, Germany, Italy, Russia, Spain, United Kingdom, and United States. It's

interesting that these are all the European countries and their North American counterparts (i.e., Canada and the United States).

This pattern is possibly reflective of the response levels for western countries which are currently suffering an explosion in cases that double the levels that China had ever experienced at its peak. The data for the number of confirmed cases show that for over three weeks, these countries had data that only reflected the minor presence of the pandemic within their borders. It makes sense then why the government officials were so slow to react and hesitant to implement strict measures such as lockdown orders. Notably are the exceptions of Italy and Spain. These are the two hardest hit European countries and it shows that Italy was the first to exit this phase after about 10 lag days. Less than 5 lag days later, Spain would also exit this horizontal phase. Afterwards, their numbers would soon grow rapidly at an exponential rate as the pandemic would spread rapidly within their populations.

The other countries that were within this horizontal channel would also eventually all exit after 25 lag days. Eventually, they too would all experience the sudden exponential growth in the number of confirmed cases as their healthcare systems would begin to feel the strain of the pandemic running through their respective populations. This data then can show the reason why western government officials were both slow to react and surprised by the inevitable skyrocketing number of confirmed cases. They had no evidence other than news reports from China that the virus would become present in their communities. This lack of apparent data to support any decisive action on the part of the government led to the false notion that the virus was a problem for China and that they were content to act minimally. Once the confirmed cases of the virus spread within their population for a period of around 2-3 weeks, the hospitals would quickly come to the realization that the virus has infected and would continue to infect large

numbers of their population. At this point, the aspect of the virus' potential for exponential infection rates would become apparent. The virus would then continue to spread until the countries had over a hundred thousand infected and tens of thousands dead due to the decision to not take the measures required to halt its progress.

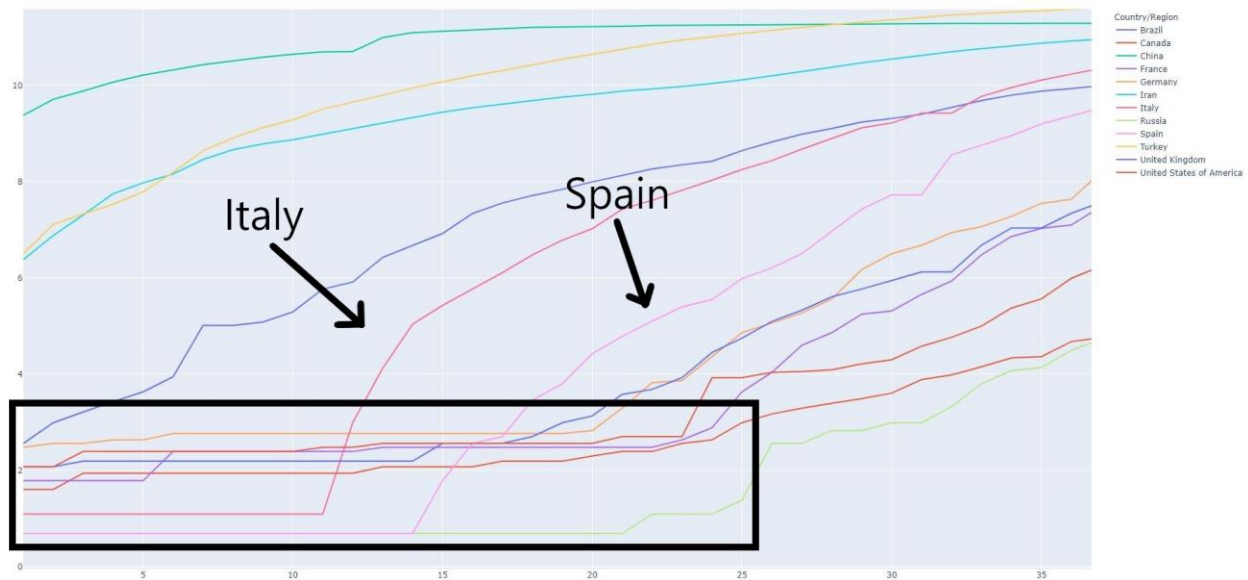


Figure 19 The above is a screenshot of the confirmed cases against lag days after logarithmic transformation. The horizontal box in black indicates that many European and western countries would remain in a phase where the confirmed cases was low and constant for a period. Only after they exited this channel would they experience the rapid exponential growth in the number of cases.

The last screenshot can be seen below in Figure 20. It shows all the countries in the graph of the logarithm of confirmed cases against lag days. Drawn on top of this chart are lines to indicate where and when the lines for each country are entering a trend phase. A linear trend on a logarithmic chart implies an exponential trend on the untransformed chart. The exceptions to this behavior can be seen in China and Turkey. China has already entered the plateau stage and so their number of cases are roughly constant, leading to a horizontal line. This is where all the countries are trying to eventually end up as the pandemic reaches its end. Turkey is an interesting

exception in that its trend is curved throughout, indicating that their growth level has never reached exponential growth levels.

The countries of Brazil and Italy are interesting for the purpose of understanding how the exponential growth trend lasts through time. In both cases, they seem to experience two different rates of exponential growth. This is indicated by two separate lines drawn through the lines representing the number of logarithmic confirmed cases. This may possibly imply that for both countries, they experience one type of exponential growth which was faster during the beginning of the pandemic. As time went on, factors such as government action and fewer potential people to become infected possibly caused the growth rate to alternate from one trend of exponential growth to another slow exponential growth trend.

The other countries all show that they too entered a period of roughly exponential growth as indicated by the straight lines drawn through their representative lines. As a note, the countries of Germany and United Kingdom are drawn together with a single line since they share some significant overlap in the logarithmic chart. The country of Russia is also different in that their growth rate is quite loosely being characterized as exponential. It's possibly related to how late Russia began to develop a growing number of cases. For that reason, no constant trend has had time to develop and take shape.

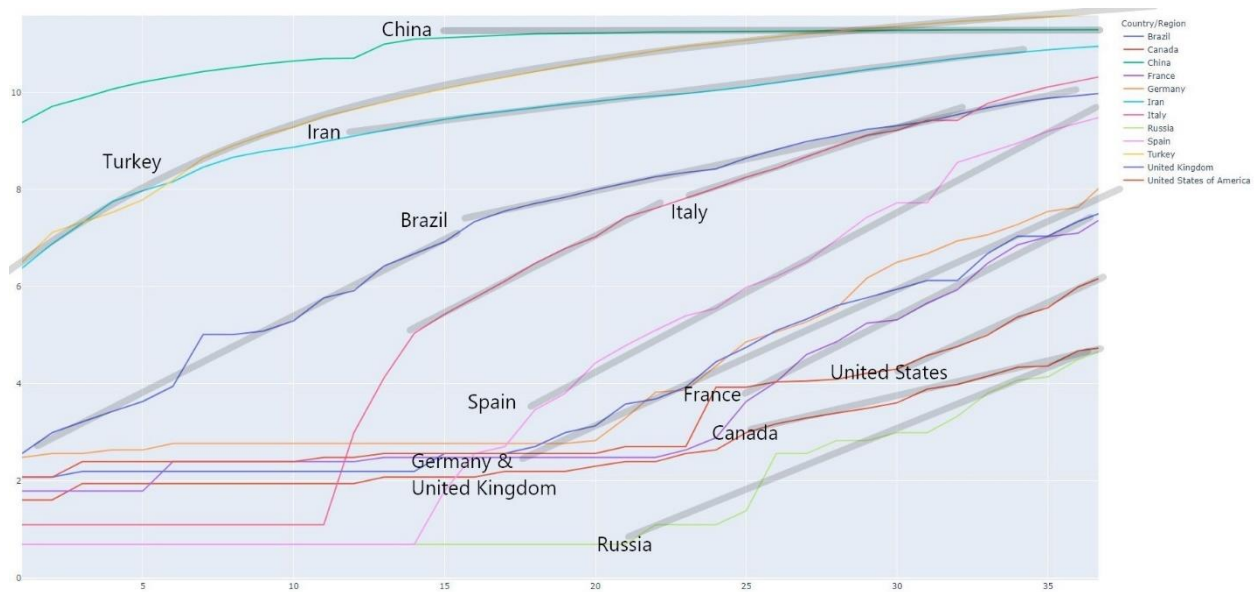


Figure 20 The above screenshot shows all the countries together and their graph of log confirmed cases against lag days. Straight lines have been drawn to signify what portions of their logarithmic trend appear linear. The implication is that a linear trend in a logarithmic chart equates with an exponential trend in the untransformed plot. The exceptions are China and Turkey. China has already reached the plateau stage and so their straight line is completely horizontal. Turkey's trend has a constant curve to it, indicating that their trend isn't exponential in growth.

4.4 Analysis of Fourth Visualization

The fourth visualization (see Figure 21) will be used to further expand on ideas elaborated on in the second and third visualizations. Those visualizations made it possible to view the number of both confirmed cases and log confirmed cases against lag days. Below in Figure 21 it shows the log cases against lag days for all twelve countries. Furthermore, for each country a linear regression model is fit to each set of data and a fitted line is drawn. The logic is that given that the growth is truly exponential, then the logarithm of the data should equate to a linear equation. Adding the fitted line using simple linear regression will model how well the transformed data is to a straight line. Of note is the fact that since the ultimate trend of the data is likely to represent a sigmoid function with an S-shape, then the level of how close the

transformed data fits a straight-line will change over time. Currently, as of 5/3/20, the data seems to signify that the data is beginning to plateau and so the time for the data to look linear rather than sigmoidal is beginning to end for the below twelve countries.

The grid of plots was created in RStudio using base R commands. The colored lines were added using Paint3D in Windows. The vertical red lines represent where the trend of the data crosses below the dotted line that represents the fit of the linear model. The red, yellow, and green highlights indicate the Adjusted R^2 of each of the plots. The color red indicates that the Adj. R^2 is at least 0.9. The color yellow indicates that the Adj. R^2 is larger than 0.75 and less than 0.9. The color green is for China whose Adj. R^2 is uniquely less than the other countries.

Looking first at China (row 1, column 3), it's apparent that the trend is no longer linear and so the country's growth is in no way linear. The value for the Adj. R^2 value is 0.3998, indicating that the trend doesn't fit a linear model. As seen in previous visualizations, the line at this point has become horizontal and so that explains the poor fit.

The countries that are colored yellow include: Iran, Italy, Spain, and Turkey. The pattern seems to be such that they were previously in a period of linear growth, but as the cases begin to plateau a linear model no longer properly fits the data. Therefore, they result in lower Adj. R^2 values. As the lines cross below the straight-line fit of the linear regression model, it seems to lead to the logic that they will continue to trend below those fitted lines. Since the number of new cases continues to increase at a much more gradual rate, there's no chance for the number of new cases to break upwards through the fitted line and potentially restart a linear trend (i.e., exponential trend when unscaled). The only way for these countries to resume that sort of linear growth is if the pandemic begins a second wave as the countries relax their lockdown measures.

If the second wave does occur, then the growth rates would move the lines in a vertical direction more rapidly, giving them the ability to start fitting an exponential growth curve.

The countries in red include: Brazil, Canada, France, Germany, Russia, United Kingdom, and United States. These countries all have an Adj. R^2 of at least 0.9, this indicates that their trend fits a linear model quite well at the current stage. However, as indicated by the vertical red lines, they're all breaking below the fitted line of the linear model. This indicates that they're all possibly on the track of plateauing. They've past their inflection points and are on the second half of their journey through the pandemic (referring to Fig. 7).

Looking carefully at Canada (row 1, column 2), it shows that this country has a more distinct S-shaped pattern than many of the other countries. The other countries tend to go from a flat line to straight to curve. Some countries also have a much more rounded appearance, like the graph of the logarithm function. The fact that Canada has a distinct S-shape while having a high Adj. R^2 value shows how the linear model will predict a strong linear fit when in fact there's some other possibly sigmoid pattern. The balance of the line above and below the fitted line is indicative of how the linear model can be tricked. It makes sense then to analyze Figure 21 in combination with Figure 20. To amend this, it makes sense to look back to Figure 20. From there, it can be remembered that the period of a straight line fitting the data occurs at the 15th - 25th lag day. Taking this into account, then it follows that if for certain plots 15-25 lag days are removed from the beginning, then an even stronger linear fit would be given (i.e., a higher Adj. R^2 value).

From what was mentioned, then the implications are that although some countries seem to more clearly be on a path of entering the plateau stage, others still need to be careful as their growth rates can still be subtly exponential, but at a different rate than what was experienced

initially. Looking at the infection numbers of being in the hundreds of thousands, there are in fact still millions of individuals that have yet to contract the virus. Therefore, governments need to be particularly cautious of reopening their countries and ending lockdowns. Despite the desire to restart their economies, it's apparent that the numbers are still growing exponentially. If the countries try to reopen in a manner that doesn't continue to halt the progress of the virus, then there's a change that the growth will again spike up. What would happen is that they would enter a new linear trend on the logarithmic chart, likely in a steeper direction. On the chart with the linear model, they'd start to point upwards at the right tip and start to move up above the fitted line. It makes sense then for health officials to be weary of any sudden changes in direction of the line representing number of confirmed cases. The most certain way seems then for them to have data which reflects what's seen now in China where the growth in cases is extremely low and remains low without any resurgences.

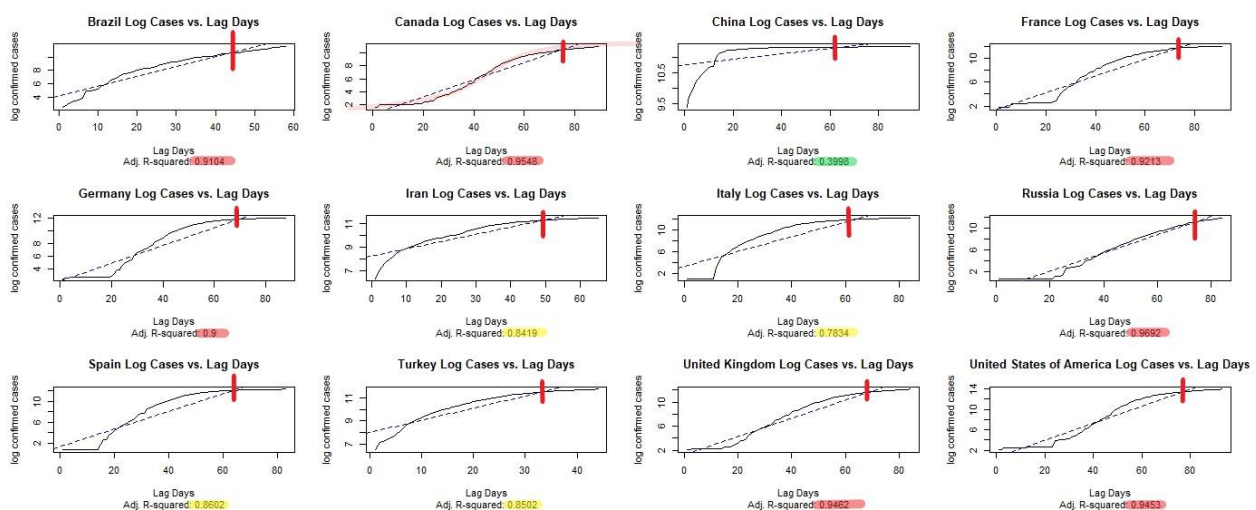


Figure 21 The above grid of plots show the comparison of logarithmic confirmed cases against lag days for the twelve countries with over 50,000 total confirmed cases. The vertical red lines indicate the point at which the log trend dips below the linear regression line. This possibly indicates that their growth is no longer exponential and is beginning to plateau. The numbers

highlighted are the Adj. R^2 values. The red color indicates that the values is ≥ 0.9 . The yellow color indicates that it's < 0.9 but ≥ 0.75 . The green color is for China with a significantly lower Adj. R^2 value of around 0.4.

4.5 Analysis of Fifth Visualization

The fifth visualization is somewhat more derived in comparison to the other examples in that it's based on the growth factor, which is based on an analysis of the sigmoidal curve. The concept is that at the beginning of the pandemic, the growth rate will steadily increase. Over time, the growth rate will begin to approach 1 for a period, then start to become smaller. Furthermore, once the data begins to enter the plateau phase, small spikes in the growth rate would be normal since the growth rate on a horizontal trend means that spikes should be normal. For example, if the previous two days had 10 and 5 new cases, then suddenly there are 30 cases, the spike would give a growth rate would suddenly jump from $\frac{1}{2}$ to 6. Therefore, it's important to be careful with the types of interpretations that are made based on this visualization.

The first chart to be analyzed is China. It's a good model for other countries to look towards when wondering what type of trend they may be interested in seeing within their own data. The reason is that China has already experienced the spike in cases and began a period of plateauing. Below in Figure 22 is the plot of growth rate against days of cases for China. The green lines are added in Pain3D. It shows many spikes above the line, but it also shows many points below the horizontal line of 1. This implies that there are many days in which the number of new cases is less than the previous day's number of new cases. The four green lines indicate that despite there being times when the growth rate spikes, it will continue to trend lower. This pattern would reflect that on certain days that the number of new cases would be higher than before, but this pattern isn't sustained. For example, there isn't a pattern of one day having 10 new cases, the next day having 20 new cases, and the next having 30 cases, etc. Furthermore,

besides there lacking that upward trend in growth rate, the health amount of points below the line means that it's typical for there to be fewer new cases than the day before. This is an important distinction as it will be seen in other countries that this pattern isn't yet picked up too commonly elsewhere.

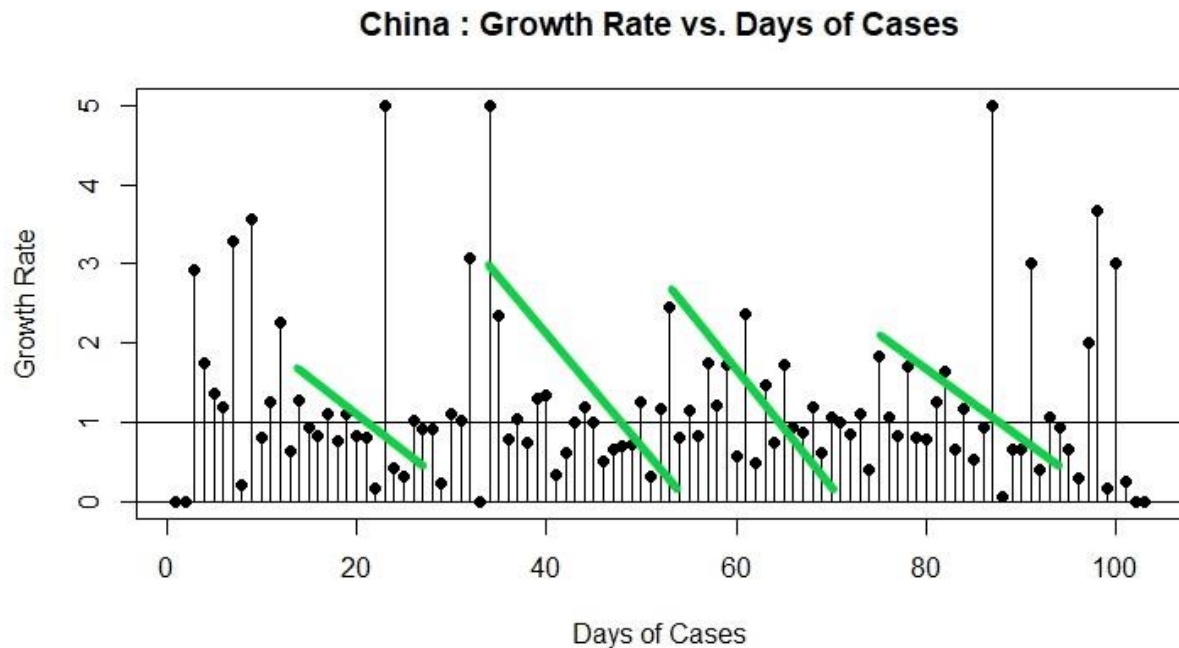


Figure 22 The above screen shot shows the growth rate (growth factor) against days of cases for China. The green lines indicate the trend of there being spikes in the growth rate followed by declines. There are also many days where the growth rate is below 1.

The next screen shot (see Figure 23) shows countries where the growth rate goes through periods of up and down fluctuation. The countries included are: Brazil, Canada, France, Italy, and Spain. Their growth factor plots are unique in that they all have curvy or up and down patterns. For example, looking at Brazil, Canada, and Spain show that they tend to trend slightly

up or down for over different periods. This implies that they will have periods where the growth in number of cases is increasing or decreasing regularly. In France and Italy, the trend is curvier in that it'll steadily trend up and down over a period time, oscillating between periods where the number of new cases is going up or down.

Brazil, Canada, and Italy also have periods where the growth factor is roughly horizontal. Looking at Brazil's recent growth factor, it's been steadily larger than 1 for some time, indicating that the number of new cases is regularly increasing daily. This makes sense given that Brazil's encounter with the pandemic is more recent and so there's still time for the pandemic to make its course before plateauing. Canada on the other hand has a recent horizontal period where the growth factor is slightly below 1. This indicates that the number of new cases is decreasing over time. However, recently it's also peaked slightly over 1, this could mean that there's also still some time before Canada reaches the plateau. In the case of Italy, although the growth factor has been roughly horizontal, it's been within a range that's both above and below 1. Therefore, it's possible that although there are new cases daily, it's not at a steadily increasing rate. This indicates that the trend in Italy is towards the plateau area.

An important point to look out for is whether they briefly have increasing numbers of new cases and later reach some flat stage where they hover around 1. This can be seen in Brazil, Canada, Italy, and Spain. In France however, this pattern is less certain as there's much more oscillation. Regardless, this pattern indicates that they've passed an inflection point and are at the time of decreasing daily cases.

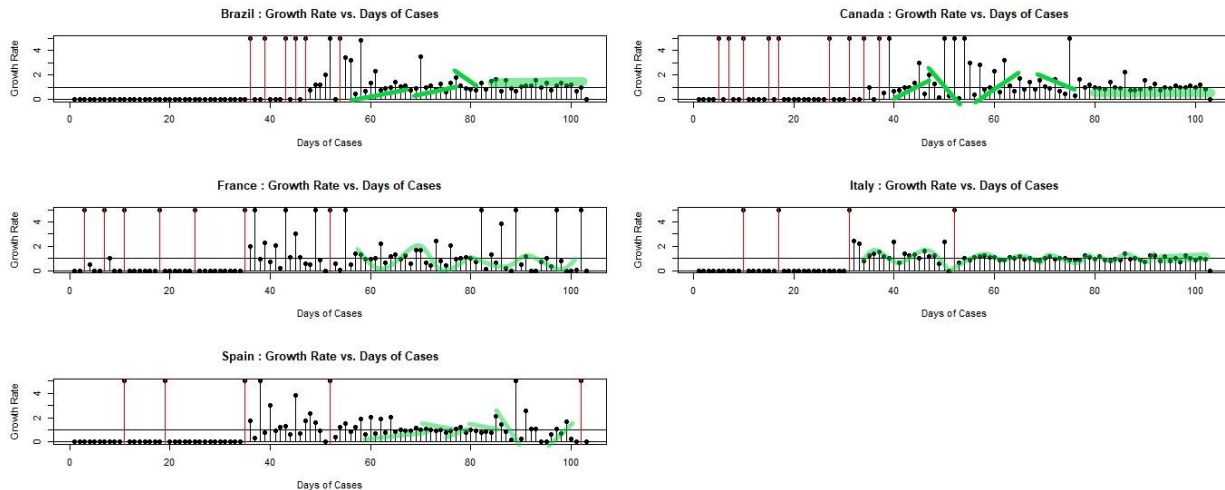


Figure 23 The screenshot above shows the countries of Brazil, Canada, France, Italy, and Spain. The graphs represent the growth rates against days of cases. Green lines are drawn to help indicate the trends and patterns of the growth rates over time.

The screenshot below (see Figure 24) includes the following countries: Germany, Iran, Russia, Turkey, United Kingdom, and United States. These countries are unique in that the plots of their growth rates against days of cases follows a significantly more horizontal pattern. Still, some differences can be made. For example, Germany is currently trending below the growth rate of 1. This indicates that the country is having fewer new cases regularly which is a good sign. The countries of Iran, Turkey, United States, and recently both Russia and United Kingdom have been trending at roughly 1. This indicates that they're possibly hovering around the inflection point and are yet to pass it but are on the way towards seeing a trend towards their plateaus. All these countries are also similar in that initially their growth rates were clearly above 1, but started to lower to become closer to 1 or move towards becoming below it. This makes sense given the sigmoid function where the growth rate is steadily above 1 for a period, before staying at 1, then starting to decrease.

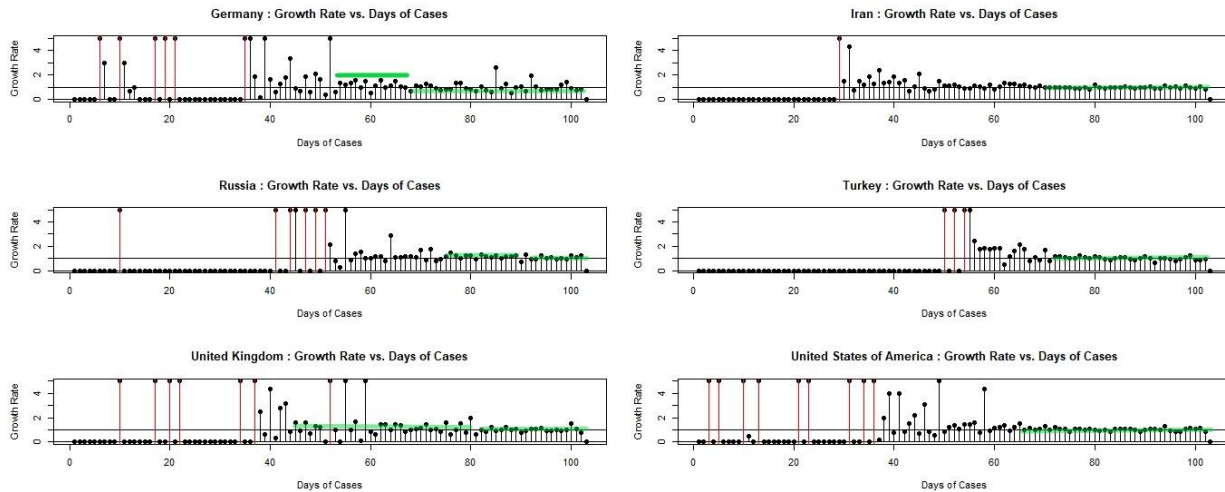


Figure 24 The above screenshot shows the growth rate against days of cases for the following countries: Germany, Iran, Russia, Turkey, United Kingdom, and United States. The green line shows horizontal trends for the growth rates. These countries are similar in that they all experience long periods of horizontal growth rate. This may reflect that they're nearing their inflection points and are around or passed the middle of the pandemic.

4.6 Analysis of Sixth Visualization

The last visualization is the sunburst chart that displays the number of confirmed cases by continent, then country, then province or state. The following table shows the corresponding number of confirmed cases by continent along with the respective percentages.

North America	Europe	Asia	South America	Africa	Australia	Total
1,262,686	1,419,373	559,528	212,500	44,307	8,335	3,506,729
(36.01%)	(40.48%)	(15.96%)	(6.06%)	(1.26%)	(0.24%)	

From the table, Europe has the largest number of cases followed closely by North America. The third most is Asia which is followed by South America. Africa and Australia have comparatively fewer cases as they're both still under 50,000 as of 5/30/2020.

The first step then will be to analyze the continent with the most cases of infections, Europe. Below in Figure 25 it shows the breakdown of cases for the continent. The hierarchy is such that the center is the continent of Europe, the next level are the countries, and the following level are provinces or states. The countries of United Kingdom, France, Netherlands, and Denmark all show a province or state region due to there being small areas that they are considered to have some sort of responsibility over according to the dataset. Examples of this include territories from colonial times that they still have domain over, or other places that they're legally supposed to provide with defense.

The sunburst chart shows that the order of most cases to the least is as follows: Spain, Italy, United Kingdom, France, Germany, and Russia. Between these countries there are between 130,000 and 218,000 confirmed cases as of 5/3/2020. It's worth noting that these are the largest countries within Europe and consist of the major powers of the continent. The other countries have less influence and have smaller economies. They're the top six in terms of both population and GDP (based on a Google search). It's not surprising then why these countries have the most confirmed cases. A logical abstraction then is that the most densely populated areas with the most movement due to economic activities are the most susceptible. For example, Ukraine has a population not too distant from Spain, but the number of cases is considerably fewer. Furthermore, Netherlands with a nominal GDP right behind Spain also has a comparatively high number of confirmed cases. By number of confirmed cases, behind Russia is Belgium then Netherlands. The population of the Netherlands however is less than half of Spain. It follows then that important factors for determining the spread of the virus include the combination of population and level of economic activity.

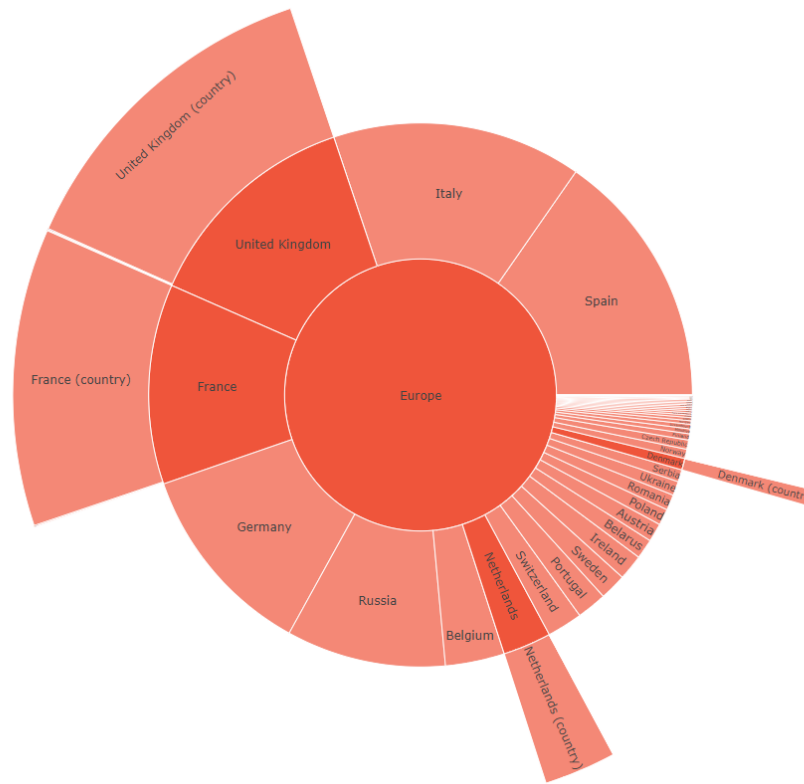


Figure 25 The above screenshot shows the sunburst visualization after selecting Europe. The plot shows the continent of Europe in the center and starting with Spain it shows from the most to the least number of confirmed cases. The third level of the hierarchy include small territories or other locations that the countries have some responsibility over.

The next screenshot shows the subset of North America after selecting it from the original sunburst visualization (Figure 9). The United States has a large majority of the confirmed cases for the continent. It's followed by Canada and Mexico, the other major countries in North America. It's interesting that Mexico which has over three times the population of Canada has fewer cases. However, from analyzing Europe it follows that the factors for having confirmed cases include the combination of population and economic activity. In terms of GDP per capita, Canada is far ahead of Mexico at nearly five times the amount.

Both the United States and Canada have third level values for the hierarchy that include provinces or states. In the United States, the state with the most cases is New York, followed by New Jersey, Massachusetts, Illinois, and California. It's worth noting that New Jersey and Massachusetts are both close neighbors to New York. There is also a large commute for people who live in New Jersey but work in New York City. It's interesting too that California, the most populous state has fewer cases than the other four. A possible reason is that the president closed borders to travelers that have been to China early in the pandemic. It's also been noted by researchers that most cases in New York come from infected people from Europe. Therefore, it follows that the spread of the pandemic comes from the east coast, whereas California is located far on the west coast. This is a possible reason why California as the state with both the largest economy and population of the country isn't number one in term of infections.

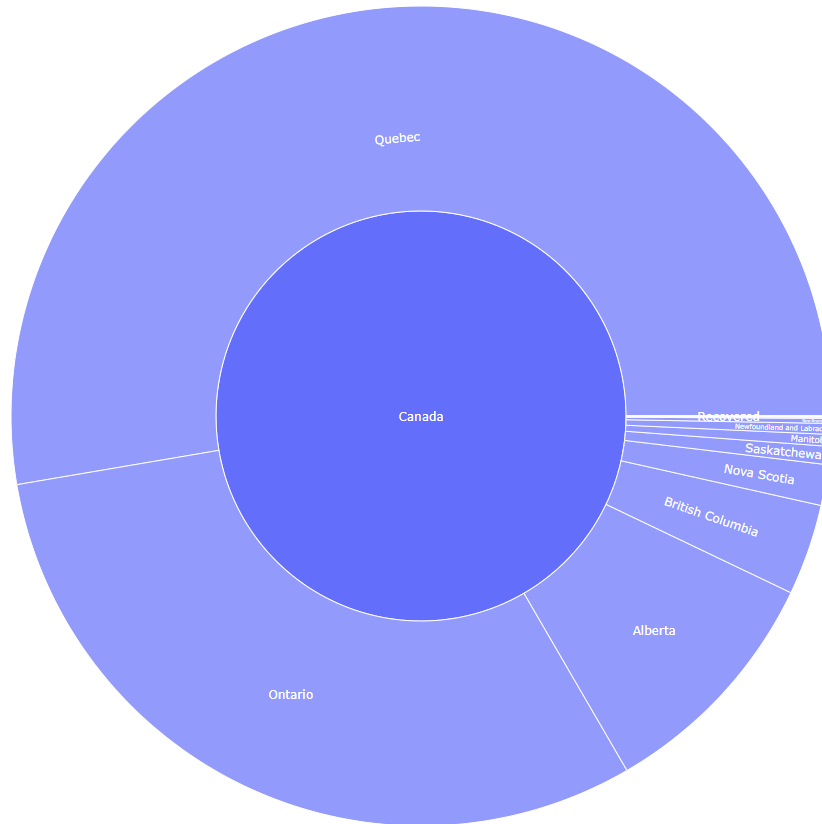


Figure 27 The above screenshot shows the sunburst after selecting Canada. It shows the country of Canada in the center followed by its provinces of Quebec, Ontario, Alberta, B.C., etc. in terms of the number of confirmed cases.

The next screenshot shows the continent of Asia. It's interesting to see that China is no longer the country with the most cases. The order from most to least are as follows: Turkey, Iran, China, and India. The third level hierarchy of the sunburst shows the provinces of China. The sunburst visualization makes clear how Hubei (the province of Wuhan) is the province with most cases. The most notable piece of this portion of the visualization is India with a population like China (over a billion). However, few countries have managed to stem the growth of the pandemic in a quick manner. The exceptions are South Korea and Singapore. Therefore, if India follows a similar path as some of the major European countries and United States, then it follows

that there could be hundreds of thousands infected and tens of thousands with a similar fate as the western counterparts.

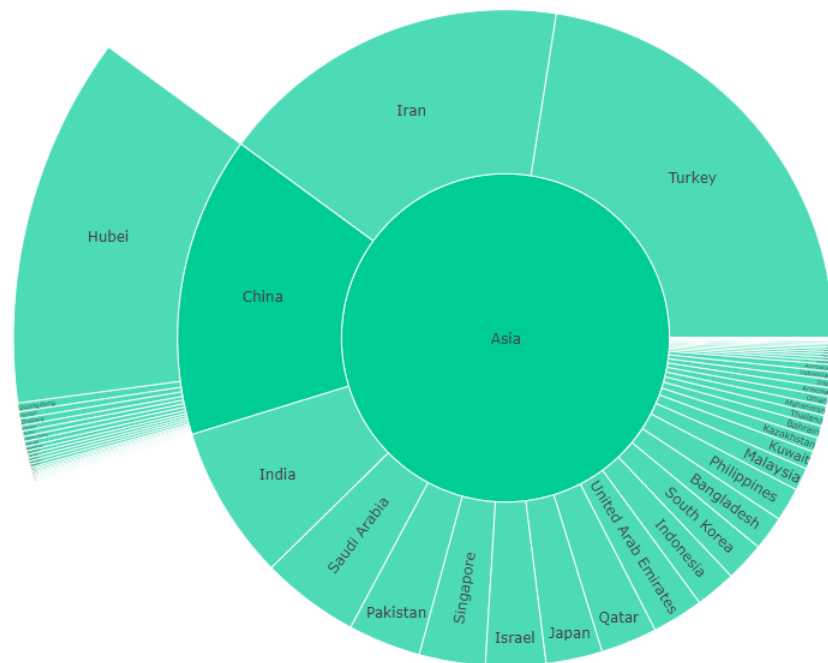


Figure 28 The above screenshot shows the sunburst visualization after selecting Asia. The plot shows the continent of Asia in the center and starting with Turkey it shows from the most to the least number of confirmed cases. The third level of the hierarchy include provinces from China.

The next screenshot shows South America where Brazil has almost half the total number of infected cases. It's followed by Peru, Ecuador, Chile, Colombia, and Argentina. Based on the idea of combining population and GDP, it makes sense why Brazil is the largest. It has both the largest population and highest GDP within the continent. Argentina and Colombia are the next two highest for those two factors, so it's interesting to see them behind Peru, Ecuador, and Chile. Another factor is that all these countries are neighbors or near neighbors to Brazil. Brazil is situated such that it's nearly the center of the continent with many neighboring countries.

However, Peru, Ecuador and Chile all have populations and GDP's that aren't too far away from the leaders.

Another possibility is that since Brazil is one of the most developed countries in the region, it has the best access to medical staff and supplies, including testing kits. Therefore, it's possible that the pandemic already exists in other regions of South America, but due to a lack of healthcare infrastructure it's going around unaccounted for. This is an issue that exists for many Asian, South American, and African countries. It makes sense then to be skeptical about the true numbers and to be unsure if the given sample is truly representative of the unknown population of confirmed cases.

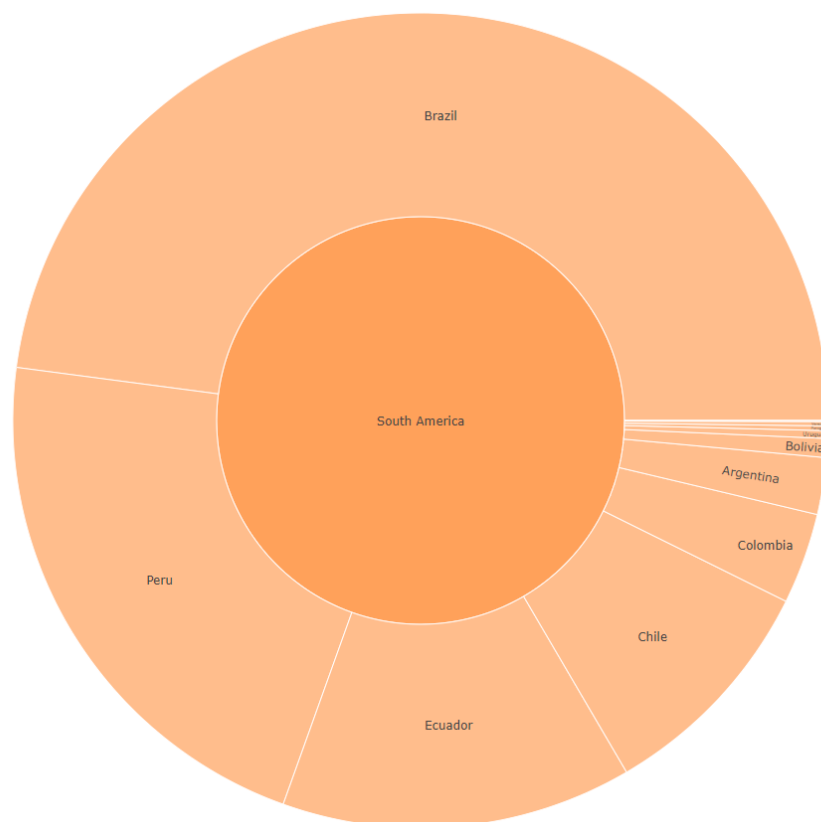


Figure 29 The above screenshot shows the sunburst visualization after selecting South America. The plot shows the continent of South America in the center and starting with Brazil it shows from the most to the least number of confirmed cases.

The next screenshot shows Africa, a continent with a population close to China and India at 1.2 billion. Therefore, the potential for the pandemic to spread is quite dangerous. Thinking back to the concept of exponential growth, the factor that slowed the spread for many countries isn't necessarily because of government action, but due to the way that environments experiencing exponential growth tend to use up all the fuel source until the growth is extinguished due to there being no more fuel. In this case, the fuel for the virus spreading is more uninfected people.

The country with the most confirmed cases is South Africa, followed by Egypt, Morocco, and Algeria. The top two countries have a similar number of cases at around 6,500. However, geographically speaking, Egypt is quite close to Europe where the outbreak has quickly exploded. South Africa is at the southern tip of the continent, further away from the pandemic. However, they both have significant population and GDP's and are both amongst the top 5 for both. Algeria and Morocco both have high GDP's for the continent while having populations that are slightly smaller than Egypt or South Africa. Therefore, it makes sense why there are also many confirmed cases there.

The concern for Africa however is related to the ability for each of the many countries to take care of themselves when the pandemic reaches their communities. In terms of GDP per capita, as a continent they're behind Asia, South America, Europe, and North America (in order from lowest to highest). In other words, they're the least prepared financially to help their population. It's then up to international organizations and multilateral institutions to help their African counterparts in their time of need. Now, the pandemic is ravaging through Europe and North America. It's possible that given that the pandemic is noticeable moving through time, African countries are maybe lucky that the virus will possibly spread first through the developed

world before reaching the continent. Once the developed countries are largely over handling the pandemic situation, they'll be more suited to assist other countries in need, such as developing countries in Africa, South America, and Asia.

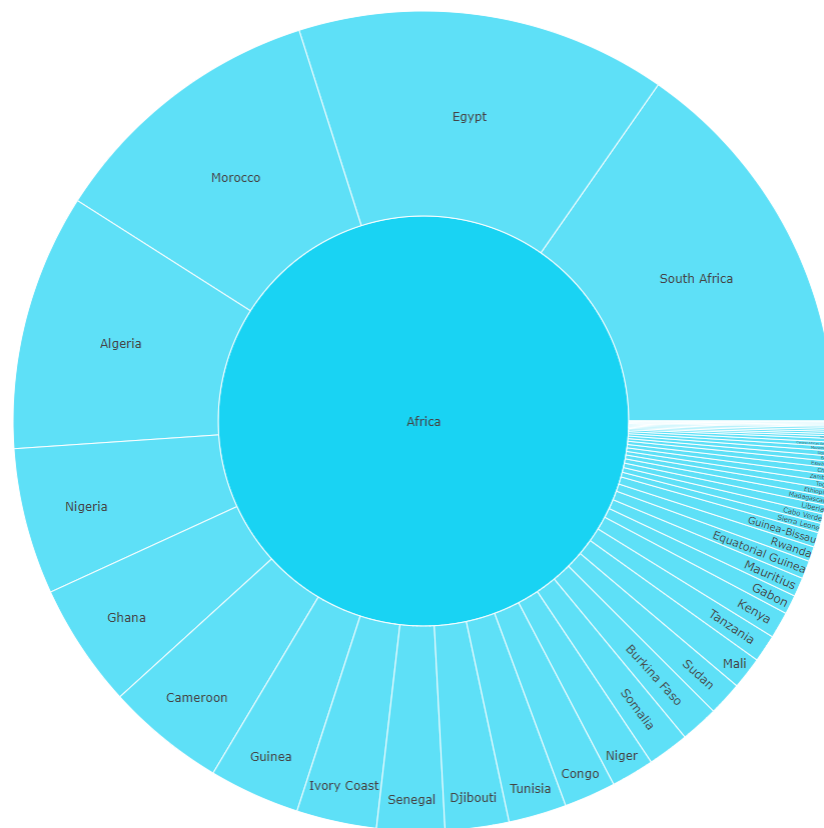


Figure 30 The above screenshot shows the sunburst visualization after selecting Africa. The plot shows the continent of Africa in the center and starting with South Africa it shows from the most to the least number of confirmed cases.

5. Conclusion

The six visualizations put into perspective the spread of the COVID-19 virus since it was first detected in the city of Wuhan. The first visualization showed from a timeline perspective how the pandemic spread globally. It began in China and traveled throughout Asia and Europe. It

was able to exist on continents across oceans and managed to infect much of the developed world with skyrocketing cases. An important theme from this visualization is that as of the date of data, 5/3/2020, there's still a looming potential for the pandemic to spread into the developing world and continue to infect the billions who live there.

The second visualization showed how the trajectory of the pandemic in terms of number of cases tends to follow a sigmoidal curve. At the beginning of the spread of the virus, countries experience an exponential growth in the number of cases. Over time as the rate of growth begins to slow, the countries are tending towards a plateauing stage. However, of the twelve countries highlighted, only China has secured itself as being in this end stage of the pandemic. Other countries are either still in the exponential phase or are trying to enter the end stage as their number of new cases start to level off.

The third visualization made apparent why western countries were hesitant to act despite their developed healthcare infrastructures. The countries had few cases for an extended period before they suddenly exploded with an exponentially growing number of cases. This was highlighted using straight lines on the log plot of the confirmed number of cases. It was further shown that many of the major European countries all experienced a similar growth pattern.

The fourth visualization expanded on the linear modeling of the logarithmic charts of the confirmed cases. It utilized linear regression to see how closely the countries' growth fit to a straight line, indicating exponential growth. Many of the countries had this trait still and it shows that they've yet to reach the end stage of the pandemic where the number of cases plateaus as in China. Therefore, countries need to be cautious unless they want to risk a second wave of the pandemic where the number of cases spikes and begins again a trajectory of exponential growth.

The fifth visualization shows the growth factor for different countries. This visualization made it apparent that the countries for the most part are still in the stage of staying around a growth factor of 1. They're trying to get past the inflection point until they make experience a pattern of where the growth factor is regularly below 1 as seen in China.

The sixth visualization makes it clear how the pandemic is as of now largely within the developed world (i.e., Europe and North America). However, from the previous visualizations it was apparent that despite the developed healthcare infrastructures of those modern countries, their populations were still highly susceptible to the spread of the virus. This visualization then makes more apparent what was seen in the first visualization which is that the virus can still have devastating effects on the world, but it's a matter of time for the pandemic to spread and to do damage. Furthermore, the developing countries have poor healthcare infrastructure that makes them less able to fend off the pandemic in the same manner that the developed world has.

Despite the devastating affects of the pandemic on European countries and the United States, it's apparent from the trends seen within this paper that the modern countries have gone through a severe part of the pandemic already. They seem to be on the way towards the end stage of the pandemic where the number of cases plateaus. At that point it's only up to the countries to be cautious to avoid a second outbreak as the world waits for a cure to the virus. The real problem then seems to be the fact that there are still billions of people in India, Africa, and South America that have yet to be infected. These are the vulnerable populations with poor citizens that won't be able to get help even if they seek it. Then it should be apparent that it's up to the commitment of international organization and multilateral institutions to develop mechanisms to help these poor countries as if the people were their own. Only until the virus is completely managed will the world be at peace. Otherwise, if the virus runs rampant throughout the developing world, then

it's only a matter of time until the virus returns to the countries who've fought hard to keep the pandemic from defeating their countries. If that happens, there would be no one else to help and the progress of the world could meet an unfortunate end due to the virus.

References

1. Google. (n.d.). Retrieved May 6, 2020, from <https://www.google.com/>
2. Johns Hopkins University. (n.d.). Maps & Trends Follow global cases and trends. Updated daily. Retrieved March 20, 2020, from <https://coronavirus.jhu.edu/data>
3. National Council of Teachers of Mathematics. (n.d.). Pandemics: How Are Viruses Spread? Retrieved March 20, 2020, from <https://www.nctm.org/Classroom-Resources/Illuminations/Interactives/Pandemics-How-Are-Viruses-Spread/>
4. Plotly Python Graphing Library. (n.d.). Retrieved May 20, 2020, from <https://plotly.com/python/>
5. Quaresima, V., Naldini, M. M., & Cirillo, D. M. (2020). The prospects for the SARS -CoV-2 pandemic in Africa. *EMBO Molecular Medicine*.
<https://doi.org/10.15252/emmm.202012488>
6. Rajkumar, S. R. (n.d.). Novel Corona Virus 2019 Dataset. Retrieved May 3, 2020, from <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>
7. Sanderson, G. S. (2020, March 8). Exponential Growth and Epidemics. Retrieved March 20, 2020, from <https://www.3blue1brown.com/videos-blog/exponential-growth-and-epidemics>
8. Stack Overflow - Where Developers Learn, Share, & Build Careers. (n.d.). Retrieved March 20, 2020, from <https://stackoverflow.com>

9. Stasko, J. S. (n.d.). SunBurst Page. Retrieved May 6, 2020, from <https://www.cc.gatech.edu/gvu/ii/sunburst/>
10. Wikipedia contributors. (n.d.). Wikipedia. Retrieved May 6, 2020, from <https://www.wikipedia.org/>