Group 2

Anna Zou, Jen Lo, Navin Kheth, Xavier Torres

**Final Report: The Impact of COVID-19 Cases**

Outline:

* **Problem:** COVID-19 is a problem that all nations currently face. We want to look at how our historical data of world-wide COVID-19 can aid in the decision-making for nations with the goal of decreasing COVID-19 cases.
* **Data Source:** [Novel Coronavirus 2019 Dataset covering years 2020 - 2021](https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset), obtained from Kaggle. The data is collected and organized by WHO (World Health Organization). The data contains multiple datasets that detail the COVID-19 death rates, recovery rates, and cases. The datasets each contain 493 records that represent a day from January 22nd 2020 to May 29th 2021. The dataset contains the variables Province/State, Country/Region, Confirmed, Deaths, Lat, Long, and Date.
* **Goal:** To forecast the COVID-19 rates to determine specific actions that nations can take to mitigate their COVID-19 cases.
* **Visualization:** Created Tableau graphs to provide visual insights into trends on COVID-19 in China, United States, India, Australia, Thailand, and the United Kingdom.
* **Time series techniques:** Due to the lack of seasonality and a noticeable trend in the dataset, we decided to use the following four techniques:
  + **Box-Jenkins Methodology:** The Box-Jenkins methodology is a set of procedures for identifying, fitting, and checking ARIMA models with time series data.
  + **Drift Method:** A trend-adjusted naive method used to capture the trend of the data in future forecasts. It does not take seasonality into consideration.
  + **Exponential Smoothing (SES):** This method produces forecasts that are weighted averages of past observations where the weights of older observations exponentially decrease.
  + **Holt Winters:** The method extends the exponential smoothing method to allow trend. The Holt method has no seasonal component.

The Problem:

COVID-19 is a continuous challenge that the world is facing right now. From the beginning of 2020, countries have responded differently towards this pandemic and have used different methods to prevent the spread of disease. The current state of COVID-19 in each country is different at this point in time because of the different actions they have taken thus far. The goal of this project is to forecast the deaths and confirmed cases of COVID-19 in China, United States, India, Australia, Thailand, and the United Kingdom [Exhibit 7]. We hope that by forecasting the data from these nations we can help nations in controlling and decreasing COVID-19 rates. Therefore, the aim of this project is to assess each country’s performance during the COVID-19 pandemic. We will evaluate their actions and compare it to the data we have gathered on the number of cases and deaths to determine the ideal future actions to take. In addition, we intend to assess the effectiveness of each type of policy responding to the pandemic and to help countries perform better during the pandemic in the future by comparing the countries that performed best and worst.

The Data Source:

The data source that we used is called *Novel Corona Virus 2019 Dataset*. We retrieved this data from a website called Kaggle. Kaggle is a website that contains published data sets that can be used to explore different areas of interest. The dataset contains multiple files that covers the daily confirmed COVID-19 positive cases, deaths and recoveries from a variety of nations from 2020 - 2021. We decided to work only with two datasets, one with data on the daily deaths and another with the confirmed COVID-19 cases. There are 493 records that each represent a day from January 22nd 2020 to May 29th 2021. The dataset contains the variables Province/State, Country/Region, Confirmed, Deaths, Lat, Long, and Date.

Major Goals:

The major goal of our project is to aid countries with their decision-making for the future, specifically in application to COVID-19, but furthermore for future pandemic diseases. In order to do so, a goal is to forecast the real life data of COVID-19 cases and deaths to determine the future rates in nations. In this way, we hope to utilize the historical data and forecast analysis to move ahead to stop the growth of COVID-19 and potentially help prevent future similar diseases. As we forecasted the data, we looked for specific patterns in the dataset that correlated with real efforts nations were taking to control the spread of COVID-19. Ultimately, we hope to develop actionable options that nations can take away from our forecasted data.

Data preprocessing:

Originally, the dataset had a few issues, so we began by ‘cleaning’ the dataset. We conducted all of the cleaning in R script. We took the original three datasets we had planned to use:

|  |  |
| --- | --- |
| time\_series\_covid\_19\_confirmed.csv | Cumulative number of confirmed COVID-19 cases |
| time\_series\_covid\_19\_recovered.csv | Cumulative number of individuals recovered from COVID-19 |
| time\_series\_covid\_19\_deaths.csv | Cumulative number of deaths from COVID-19 |

We then imported the three csv files onto a new R script. We loaded the packages “fpp2”, “dplyr”, and “janitor”. We renamed the csvs as global.cases, global.deaths, and global.recovered. For the global.cases, we began by removing the columns region, latitude, and longitude since they provided duplicate information on location which we did not need. Then we combined the different regional data under single country variables. Since the data was formatted as each date was a column rather than a row we needed to transpose it to show the true number of records. In order to do this, we converted the data frame into a matrix to transpose them into columns. We then re-converted global.cases back into a data frame once it was transposed. We converted the data frame from string to numeric. Finally, we defined the ts object with just the data from Australia, China, Thailand, United States, United Kingdom, and India. Once the dataset was cleaned we checked to see if there were any missing values and there were none for global.cases. We followed this same procedure for global.recovered and global.deaths.

We ended up with 493 records which were associated with each day from January 22nd 2020 to May 29th 2021. We noticed some issues after concluding with our cleaning process. For the global.recovered when we checked for missing data it returned that there were no missing data in R, however, when we scanned the dataset time\_series\_covid\_19\_recovered.csv by hand we noticed that for the United States after the date of 12/14/2020, all of the cells had "0" for the recovery amount. Since there was data inputted in the cell, it was not recognized as missing data. However, since it was “0” the data was inaccurate since there have since been recoveries in the United States. Essentially, the recovery amount is the number of COVID-19 cases minus the deaths, and because of this discrepancy in the data we decided to remove the dataset of time\_series\_covid\_19\_recovered.csv from our evaluations. We also discovered that the datasets were providing numbers that were cumulative up to date rather than new daily cases and deaths. We needed the daily amounts in order to conduct a time series forecast. To resolve the issue, we decided to difference the entire datasets by lag 1 so that each time period represented the new cases or deaths recorded for that particular day. This also shifted our started date by one day to January 23rd.

Time Series Application:

We moved forward with the death rates and confirmed cases as our time series variables. We first ran the correlogram to see if there was any trend or seasonality in the original data. We identified trends but there was a lack of conclusive evidence to support seasonality. Especially since our data set only covered one years worth of data, so it was difficult to note significant seasonality peaks or dips. Therefore, we decided to move forward with techniques that focused on trends. The techniques we used:

1) Box-Jenkins Methodology: We ran an exhaustive and rough search for ARIMA where we evaluated the AIC. After running each one we checked to see if the model was adequate or not adequate. We then predicted with the models we developed and selected the best model based on the RMSE (root mean squared error) and MAPE (mean absolute percentage error) as they were our key performance metrics.

2) Drift Method: We ran rwf to account for the trend in the data. For this we checked to see if the models were adequate or not. Then we selected the best model based on the RMSE (root mean squared error) and MAPE (mean absolute percentage error) as they were our key performance metrics.

3) Exponential Smoothing (SES): We decided to use exponential smoothing as a control test since it does not handle trend or seasonality well. We wanted to see what the outcomes of the models would be since we did not identify strong time variables in the data set. We ran ses tes to see if the models were adequate or not. Then we selected the best model based on the RMSE (root mean squared error) and MAPE (mean absolute percentage error) as they were our key performance metrics.

4)Holt Winters: We ran Holt to account for the trend in the data. For this we checked to see if the models were adequate or not. Then we selected the best model based on the RMSE (root mean squared error) and MAPE (mean absolute percentage error) as they were our key performance metrics.

Visualizations:

We conducted our visualizations on Tableau with cleaned datasets from the original CSV files. The datasets were cleaned to only contain information on China, United States, India, Australia, Thailand, and the United Kingdom. We separated the nations into two separate groupings: (1) India, United States, and United Kingdom (2) Australia, Thailand and China. This was done so that our visualizations would accurately represent the numbers graphically since the combination of all the nations skewed the graph representation. As shown in Exhibits 1 and 2 the COVID-19 cases trend line for these two groupings of nations. In Exhibit 1, we can see that the United States and the United Kingdom followed a similar pattern of COVID-19 cases, but the United States on a much larger scale. From the end of February both nations begin to show cases of COVID-19 and the United Kingdom maintains a steady level of cases until September as it increases. The United States during this time had a more rapid case growth and occasional spikes. For both nations, the bulk of their cases skyrocketed from September until January. Afterwards, both nations COVID-19 cases decreased up until our data ends in May 2021. India on the other hand had a completely different pattern of COVID-19 cases. They did not begin experiencing COVID-19 cases until almost April 2020 and once cases began they grew at a steady rate reaching a high in September 2020, after which the number of cases fell for a few months until March when the number of cases spiked exponentially to a collective high of 8,729,831 cases in May 2021 [Exhibit 1]. This is due to the early restrictions being lifted in March 2021, that is when “many Indians stopped taking precautions. Large gatherings, including political rallies and religious festivals, resumed and drew millions of people” (“*Coronavirus in India: What to Know - the New York Times”).* The reaction of the people is a big reason for the dips and jumps in COVID-19 numbers, how they choose to respond to the protocols and the lack of them heavily impact what the nation will look like in the pandemic landscape.

The nations Australia, China, and Thailand followed a different pattern of COVID-19 cases compared to Exhibit 1. In the case of China, they were the nation that was first exposed to COVID-19 so they showed the earliest cases in January 2020 [Exhibit 2]. Immediately following the outbreak the spike in cases occurred in February. However, since the initial outbreak, China shutdown most activities and enforced a strict quarantine lockdown which led to a decrease in cases. Since March 2020, they have been able to control their cases with heavy monitoring and limitations on unnecessary travel. Australia has been able to maintain the lowest COVID-19 case numbers in all the nations we evaluated. Part of this is due to the geographic isolation and population density they have (Haseltine). It was also their fast reaction to the COVID-19 spread, their cases growing up until March which is when the Federal government responded by “closing international borders and implementing a mandatory home isolation program for returning Australian citizens” (Haseltine). Once the government took action their cases decreased and were able to be maintained at a low rate. They followed their COVID-19 protocols strictly even with quarantine hotel rooms being “guarded by police or military” to ensure that citizens were following the protocols (Haseltine). Thailand was quick to respond to the outbreak as well. When Thailand first heard news of the COVID-19 cases they had set up airports to screen visitors since “Thailand is among the top destinations for travelers from Wuhan” (Arunrugstichai). They got their first case on January 13, 2020. However, Thailand could not “afford mass screening so contact tracing was deemed more effective” and so COVID-19 testing units were available for free for people who were at risk for contact tracing. One of the big reasons for their ability to maintain a low number of cases from the beginning to early 2021 was also the response by the people of Thailand. The citizens cooperated very well with the government and wore masks, especially since masks were common prior to COVID-19 because of the bad air pollution from their heavy traffic (Arunrugstichai). This heavily impacted the slow spread of the disease. It wasn’t until April 2021 when issues began as Thailand opened its borders for tourism again. They needed to open borders due to financial losses in their lack of tourism from the previous year. Once borders opened up again the spread of COVID-19 cases increased rapidly as Thailand hit an all time high in May 2021 with 86,689 cases [Exhibit 2].

We also created a histogram chart in Tableau to demonstrate the difference in number of COVID-19 cases from 2020 compared to 2021 [Exhibit 3]. The nations United States, United Kingdom, China, and Australia had a decrease in their cumulative number of cases. However, this could have been due to the fact that for 2021 we only had data from January to May as opposed to the whole year for 2020. The nations India and Thailand had more cases in 2021 compared to 2020, for Thailand this was due to the lifting of border restrictions and for India this was because of the removal of COVID-19 restrictions. This demonstrates that the specific actions nations choose to take can have drastic ramifications across the nation that can heavily impact their COVID-19 rates. The countries that chose to stay diligent about their restrictions and protocols ended up decreasing their COVID-19 case numbers in 2021.

We also graphed the trend lines for death cases in all 6 nations. We grouped them the same as earlier. We have Australia, China, and Thailands trend lines for the number of deaths shown in Exhibit 4. Essentially, the death rates mirror the number of COVID-19 cases, when there is a spike in cases there is a simultaneous spike in deaths during the same timeframe. Additionally, there may sometimes be a lagged response for the number of deaths as some COVID-19 cases result in death later on. This is demonstrated by China as they have a spike in deaths during the initial outbreak in February 2020 resulting in 2,624 deaths and later on a spike on a smaller scale with 1,328 deaths in April 2020. The mirrored pattern is also displayed in Exhibit 5, as the nations India, United States, and United Kingdom spikes in death correlate with the spikes in COVID-19 cases. This effect is demonstrated as well in the histogram comparing the number of deaths in 2020 to 2021. The same nations: United States, United Kingdom, China, and Australia that had a decrease in COVID-19 cases also had a decrease in the number of deaths. Only the nations Thailand and India saw an increase in the number of deaths in 2021 which correlates with their increase in the number of COVID-19 cases in 2021.

Models/Methods Developed:

During the initial phase of our modeling process, we ran a total of 90 models, 5 models per country for 6 countries, on 3 different datasets. The 6 countries include Australia, Thailand, India, China, the United Kingdom, and the United States [Exhibit 7]. The 5 models we ran for each country include drift (rwf), exponential smoothing (SES), Holt-Winters (non-seasonal), and non-seasonal ARIMA (both exhaustive and rough search). As we studied the quality of our data and forecasts, we decided that the data set for recovery was too inconsistent to provide accurate and valuable forecasts for this particular project. After cutting out the recovery dataset, we were left with 60 models on the two final datasets and we used these 60 models to assess their performance to determine what model we should use to forecast confirmed COVID-19 cases and deaths.

Prior to building the models, we converted the data frame into a time series object and partitioned the two datasets into training and testing data. We chose the frequency of the time series to be 365 because the time increment was daily. The training datasets included the earliest 80% data points of confirmed cases and deaths respectively. The testing data consisted of the latest 20% of data points. We decided that the forecasting horizon for our project would be 30 days or about a month, which would provide enough time for actionable solutions to combat the potential rising number of cases or deaths. We also felt that a month's forecast would be more accurate since we did not factor in outside variables to our dataset such as vaccination rates and new variants of COVID-19. We then fitted our models on the training data to check for model adequacy and finally evaluated the forecasting performance of each model on the testing data set. To test for fit, we studied the residuals of each model to see whether the model eliminated time series elements from the residuals as well as whether the residuals were normally distributed with a mean of 0. As for forecasting performance, we mainly looked at the RMSE (root mean squared error) and MAPE (mean absolute percentage error) as our key performance metrics.

Performance of the techniques:

Before evaluating the forecasting performance of the models, we attempted to evaluate the models by studying its residual to check for adequacy. However, none of the models were adequate as the ACF’s of the residuals were above the confidence band for all models. Since none of the models were adequate, we could not decide which models to forecast with based on adequacy. To pick the best model, we then forecasted each model of the 60 models with a forecasting horizon of the length of the testing data and then compared the forecasted numbers against the actual to produce RMSE and MAPE. All the RMSE and MAPE are listed in the tables below.

**Table 1 - Confirmed Cases RMSE:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Australia | China | Thailand | USA | UK | India |
| Drift | 11.83 | 14.41 | 1898.54 | 4.3x10^4 | 1.02x10^4 | 2.07x10^5 |
| Holt | 10.89 | 9.94 | 1919.25 | 2.75x10^4 | 8.54x10^3 | 2.1x10^5 |
| SES | 52.43 | 95.07 | 1891.84 | 1.55x10^5 | 8.59x10^3 | 2.02X10^5 |
| ARIMA-E | 10.46 | 10.17 | 1843.53 | 3.18x10^4 | 8.31x10^3 | 2.08x10^5 |
| ARIMA-R | 10.46 | 9.94 | 1843.53 | 2.75x10^4 | 8.54x10^3 | 2.08x10^5 |

**Table 2 - Confirmed Cases MAPE:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Australia | China | Thailand | USA | UK | India |
| Drift | 74.03 | 46.48 | Infinite | 1.08x10^2 | 3.42x10^2 | 71.08 |
| Holt | 63.10 | 36.58 | Infinite | 68.80 | 2.84x10^2 | 73.93 |
| SES | 505.36 | 471.33 | Infinite | 3.57x10^2 | 2.86x10^2 | 68.86 |
| ARIMA-E | 59.37 | 43.82 | Infinite | 78.41 | 2.76x10^2 | 71.73 |
| ARIMA-R | 59.37 | 37.20 | Infinite | 68.81 | 2.84x10^2 | 71.73 |

**Table 3 - Deaths RMSE:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Australia | China | Thailand | USA | UK | India |
| Drift | 0.10 | 0.44 | 15.57 | 2.12x10^3 | 5.41x10^2 | 2.30x10^3 |
| Holt | 0.10 | 0.93 | 15.36 | 1.75x10^3 | 4.97x10^2 | 2.32x10^3 |
| SES | 0.10 | 3.30 | 15.31 | 1.76x10^3 | 1.69x10^2 | 2.45x10^3 |
| ARIMA-E | 0.10 | 0.39 | 15.36 | 1.14x10^3 | 4.53x10^2 | 2.32x10^3 |
| ARIMA-R | 0.10 | 0.91 | 15.36 | 1.03x10^3 | 4.22x10^2 | 2.32x10^3 |

**Table 4 - Deaths MAPE:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Australia | China | Thailand | USA | UK | India |
| Drift | 100 | Infinite | 100 | 3.28x10^2 | 4.66x10^3 | 67.7 |
| Holt | Infinite | Infinite | Infinite | 2.71x10^2 | 4.13x10^3 | 69.57 |
| SES | Infinite | Infinite | Infinite | 2.73x10^2 | 1.40x10^3 | 87.9 |
| ARIMA-E | Infinite | Infinite | Infinite | 1.79x10^2 | 3.80x10^3 | 70.15 |
| ARIMA-R | Infinite | Infinite | Infinite | 1.62x10^2 | 3.55x10^3 | 70.15 |

While evaluating the performance of the models, we ran into the issue where the MAPE outputs infinite for many of the death forecasts. We examined the data to figure the cause and found that many of the actual death counts for countries such as Australia, China, and Thailand were 0, but our forecasts were above 0. This meant that when calculating the percentage error, the error was divided by 0’s causing the percentage to approach infinity.

When comparing the RMSE and MAPE of the 60 models, we found that the ARIMA models were consistently producing the most accurate forecasts since they usually had the lowest RMSE and MAPE. On the other hand, the exponential smoothing models (SES) were producing the worst forecasts. From the performance metrics, we decided to use the ARIMA model to predict the next 30 days following the entire data set. The (p, d, q) order of the models will be selected from both the rough and exhaustive search, depending on which one performed better against the test data. For confirmed cases [Table 1 and 2], we used the following (p, d, q) to forecast: Australia(1, 1, 3), China(2, 1, 1), Thailand(2, 1, 3), US(0, 1, 1), United Kingdom(3, 1, 2), and India(2, 1, 3). For deaths [Table 3 and 4], we used the following (p, d, q) to forecast: Australia(0,1,5), China(0, 1, 1), Thailand(2, 1, 1), US(3, 1, 2), United Kingdom(1, 1, 5), and India(1, 1, 3).

For all forecasts, see Exhibits 9 through 20. In general, the forecasts of the models made sense compared to the preceding data. For example, confirmed cases for Australia and China flattened out and stabilized at a few cases above 0 and the model forecasted that cases would continue to be stable. On the other hand, forecasts for Thailand showed a lot of fluctuations to reflect the fluctuations found in the historical data. The United States and the United Kingdom had similar COVID-19 trends, meaning that if cases went up in the US, then the United Kingdom also usually went up. The forecasts for both countries also were very similar. India peaked in terms of cases in early 2021, but the cases were going down at the end of the data and the model forecasted that the cases would continue to go down. The data and forecasts of deaths were highly correlated to the data and forecasts of confirmed cases. Australia and China stabilized to right above 0. Thailand’s death counts were fluctuating wildly toward the end of the data and the forecast predicted that the deaths would hover around 35. The model predicted that the United States death counts would fluctuate at around 1000 deaths, while the United Kingdom was forecasted to stabilize right above 0 deaths. India was once again predicted that deaths would continue to decline.

Future action and implementation to mitigate such diseases:

After forecasting the COVID-19 cases and deaths for the United States, China, Thailand, Australia, United Kingdom, and India we were able to see how each country would perform for the next month. We hope to be able to apply a similar forecasting model to other nations to provide accurate predictions so that they could be prepared to act in response to the spread of the virus. Providing early actions and gathering information as much as possible is important for preventative responses. It is essential to focus resources on eliminating the spread of the virus, which can be done through actions such as an early prohibition of gathering, travel restrictions at the beginning of the outbreak, or mandatory quarantine in the early stage (Grépin). The Worldwide Health Organization suggests that travel measures will allow governments to gain time, even if only a few days, to rapidly arrange effective and efficient preparedness measures (“Updated Who Recommendations for International Traffic in Relation to Covid-19 Outbreak”). The countries that did well in the early stages of COVID-19 were China and Thailand. Part of this was due to the stricter policies that they had in place. For instance, China had a strong stay in house policy at the very early stage in February 2020. The people were required to not leave the house with exceptions for daily exercise, grocery shopping, and “essential” trips at that time. It was a complete lock-down and China didn’t reopen until they had zero new infections a day (Davies). Its government did so to eliminate the spread of COVID-19 as much as possible. Thailand also had policies recommending its people to not leave their homes. Similarly, Thailand and China were also very strict on their school closures, which require schools at all levels to be closed: China first imposed it on January 21, 2020 and Thailand employed it on March 17th. In terms of international movement during the pandemic, both countries require the PCR COVID-19 test within 72 hours of departure immediately upon arrival. In addition, China also requires a serological (antibody) test to confirm the passenger is all cleared for the negative test results, while Thailand requires obtaining a certificate of entry and providing proof of an insurance policy that covers treatment for Covid-19 up to the cost of $100,000 (“Traveling to Thailand during Covid-19: What You Need to Know before You Go”). Aside from these requirements and preparation before traveling, these two countries also have an additional layer of protection in case of encountering the virus during the 72 hours gap after the test and before the flight - mandatory centralized quarantine upon arrival. All international travelers will be sent to an authorized and monitored hotel for 14 days (for China there is also an additional 7-14 days of home isolation depending on different city’s policy) (“Foreigners' Guide to Quarantine and Getting Vaccinated in China”). These actions helped the two countries in resisting the possible spreading or entrance of the virus from outside the country, and thus we recommend it to other countries as well for reduction of transmission.

In addition, effective and efficient allocation of health resources is also very essential, such as making vaccinations and testing accessible, convenient, and free. Such relevant support from the government will help to limit the disease's growth and spread. Laws and regulations for PPE (personal protective equipment) can also be very beneficial (Office of the Commissioner). Countries should enforce mandatory facial coverings on citizens and mandatory quarantine on international travelers. It is also important to continue these steps throughout the pandemic to prevent spikes in the spread. It is not enough to do it only early on, but to remain vigilant throughout the pandemic. PPE's effectiveness in minimizing the transmission, countries will be able to remain their border open, which will then offset the financial impact on their economies.

However, there are limitations to our analysis and recommendations. For example, Chinese and Thai citizens’ willingness to obey the government instructions plays a significant role in successfully controlling the pandemic. The mask culture is not new to the East, while many westerners were skeptical about it (Joung). Also in Asian cultures citizens are willing to wear face masks out of respect for other people’s safety. This type of collectivist thinking is common within Asian nations, while western nations are more individualistic. The effectiveness of government actions and policies is heavily dependent on the power of execution. In addition, the economic losses related to the policies are burdens for the countries. For countries that are already suffering from unemployment and deflation, a travel ban and restaurant closure would be huge damage to the country's economy. How can nations like this take actions to help mitigate the damage and limit the growth of COVID-19? We advise that they first gather accurate information from now on so that they can initiate tracking measures to contain COVID-19 numbers. More accurate and stricter tracing of COVID cases will not only help the countries prevent the spread of the virus, but it will provide us with more accurate data so that our models will perform better in predicting the future. This factor can be improved and implemented by actions such as pooled resources and collaboration of data gathering with other countries. This will allow validation of the data’s accuracy and quality, avoid ecological fallacy, and higher reliability of comparison between countries due to the difference in health recording system (e.g. US CDC may document data differently and have different measuring metrics than China’s CDC). Secondly, they can study the actions and results of comparable nations that have similar social and economic structures as them (comparable factors such as wealth, population, climate, healthcare, etc). Epidemiology is developed based on the idea of studying differences between populations (Pearce). By enabling comparison between countries will help unfold the pandemic, stimulate the progress of research experiments and medical developments, and evaluate the effectiveness of different approaches. In other words, international comparisons and learning different regulations play an important role in understanding the disease better and determining what works best in controlling COVID-19. They should then create policies and regulations that are supported by accurate and in-depth analysis of historical data to develop better strategies.

Summarized Results:

Our results were largely inconclusive of any huge or groundbreaking takeaways but we did find that the results of our models had extremely varied errors, such as the infinite MAPEs, based on the country that we predicted. This is probably due to variation in country response to the issue. All of our models did well in Australia and China for COVID-19 cases. However, due to different behaviors in the data, our models did substantially worse in countries where the rate of cases and deaths fluctuated more frequently or drastically as compared to countries with more stable rates. India, the United States, and the United Kingdom had some of the worst accuracies of all of our models. For example, the United Kingdom had an accuracy MAPE of 300% for deaths and for confirmed cases they had an accuracy MAPE of 280%. This follows our original assumptions since we went into this process expecting inaccurate conclusive data. This is in part due to the type of data that we are working with, since it is real life data, the data itself may not be completely accurate and thorough. All of our models were unsatisfactory, but we concluded that ARIMA was the strongest model and as such, we chose to use it as our final model. We found this model to be the most accurate in both cases and deaths. We found some problems with our models in China since they claimed zero deaths and as such, we found that our error metrics did not work well. MAPE specifically could not be used when the actual number of deaths would be zero or else it would become infinite. RMSE was a better metric for China for that reason. In addition, our models had large differences between countries. While our models had varied results it seems like for the countries that we did model on, in most of the countries our MAPE for cases ranged between 40% to 80%. An exception is the United Kingdom where the error rate goes through the roof. Our error rates of 40% to 80% are substandard. Cases were the main problem with our models, while our predictions on death were substantially better. For deaths we will use RMSE as it is better suited since the number of deaths of some countries can hover around zero. Our RMSE ranged between .10 to 2.32x10^3. These large ranges can be attributed to the differences between countries. We find that deaths are easier to predict, we are not exactly sure why. This seems to signal that there are more elements of seasonality in deaths rather than cases.We thought it would have to do with the timeline of symptoms that progress but we have no concrete evidence for that.

Summary:

Since we only used a time series analysis, there are many factors that were not taken into account. Some of these factors include:

* The existence of different variants
* The usage of vaccination
* The availability of hospital beds
* The density of countries
* The type of healthcare available
* The severity of restrictions put in place
* The overall health/age of the population

These factors are all useful for analyzing and predicting the stage of deaths and cases from each country. Each of these variables would vary in impact for each of the countries. Utilizing only time series elements to predict deaths and cases is not a method that produces results that are productive or conclusive. In further study we would like to research the strength of the stated variables in each country to find evidence for action items or recommendations that we could give to each country. This would allow us to recommend individual countries advice that is specifically tailored towards its situation. Our goal with the project was to identify trends in deaths and cases. This would allow us to create methods that would ease the pandemic in countries around the world. Using the models that we did, we didn't find any seasonality in cases but we did find seasonality in deaths for the U.S. using our models to predict that we trained on about 1 and a half years of data.

Takeaways:

Our group encountered some challenges while working with this dataset. Since the dataset involved real life data that was still being updated, we were faced with many inaccuracies. The datasets themselves had discrepancies in how they were recorded. We cannot be sure of how accurate the data is since most of it involved self-reporting from the nations themselves who could have skewed the information. In the future, it will be important to note that real life data may be inaccurate, so we must do the due diligence to locate accurate data and cross-check it across multiple verified sources. Overall, it was a good experience working with a large, uncleaned dataset. We gained more experience using the frameworks we learned in class to deal with these situations, as well as how to use the context of real-life scenarios to make those decisions.

Another challenge we faced was working on the same R code since each of us have different coding styles and only one of us could be working on the code at a time. It slowed us down in the coding process since it was more efficient for one person to write most of the code. We also ran into the issue of being unable to open and run the code on everyone else’s computers due to the way the source files were loaded in. To combat this issue, we had regular meetings to update everyone on the progress of our R code so that everyone could understand the logic behind our methods and results. We learned that good coordination and communication is necessary to complete a coding project in a group environment.

References

Arunrugstichai, Sirachai. “A Look inside Thailand, Which Prevented Coronavirus from Gaining a Foothold.” *History*, National Geographic, 18 June 2020, <https://www.nationalgeographic.com/history/article/look-inside-thailand-prevented-coronavirus-gaining-foothold>.

*Coronavirus in India: What to Know - the New York Times*. 17 Nov. 2021, <https://www.nytimes.com/article/india-coronavirus-cases-deaths.html>.

“Updated Who Recommendations for International Traffic in Relation to Covid-19 Outbreak.” *World Health Organization*, World Health Organization, 29 Feb. 2020, <https://www.who.int/news-room/articles-detail/updated-who-recommendations-for-international-traffic-in-relation-to-covid-19-outbreak>.

Davies, Dave. “Compared With China, U.S. Stay-At-Home Has Been 'Giant Garden Party,'

Journalist Says.” NPR, NPR, 29 Apr. 2020, [www.npr.org/sections/health-shots/2020/04/29/847755751/compared-to-china-u-s-stay-at-home-has-been-a-giant-garden-party-journalist-sa](http://www.npr.org/sections/health-shots/2020/04/29/847755751/compared-to-china-u-s-stay-at-home-has-been-a-giant-garden-party-journalist-says)ys.

Grépin, Dr Karen Ann. “Evidence of the Effectiveness of Travel-Related Measures during the Early Phase of the COVID-19 Pandemic: A Rapid Systematic Review.” *BMJ Global Health*, BMJ Specialist Journals, 1 Mar. 2021, <https://gh.bmj.com/content/6/3/e004537>.

Farr, Christina. “Will Masks Become the 'New Normal' Even after the Pandemic Has Passed? Some Americans Say So.” CNBC, CNBC, 24 Oct. 2020, <https://www.cnbc.com/2020/10/24/will-masks-become-the-new-normal-even-after-the-pandemic-has-passed-some-americans-say-so-.html>.

Haseltine, William A. “What Can We Learn from Australia's Covid-19 Response?” Forbes, Forbes Magazine, 24 Mar. 2021, <https://www.forbes.com/sites/williamhaseltine/2021/03/24/what-can-we-learn-from-australias-covid-19-response/?sh=2e2a05d3a01a>.

Joung, Madeline. “Face Mask Culture Common in East, New to West.” VOA, Face Mask Culture Common in East, New to West, 2 Apr. 2020, <https://www.voanews.com/a/science-health_coronavirus-outbreak_face-mask-culture-common-east-new-west/6186770.html>.

Office of the Commissioner. “FDA Insight: Personal Protective Equipment and Covid-19.” *U.S. Food and Drug Administration*, FDA, <https://www.fda.gov/news-events/fda-insight/fda-insight-personal-protective-equipment-and-covid-19>.

Pearce, Neil, et al. “Comparisons between Countries Are Essential for the Control of Covid-19.” *International Journal of Epidemiology*, Oxford University Press, 1 Aug. 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7337754/>.

“Traveling to Thailand during Covid-19: What You Need to Know before You Go.” CNN, Cable News Network, 30 Dec. 2020, <https://www.cnn.com/travel/article/thailand-travel-covid-19/index.html>

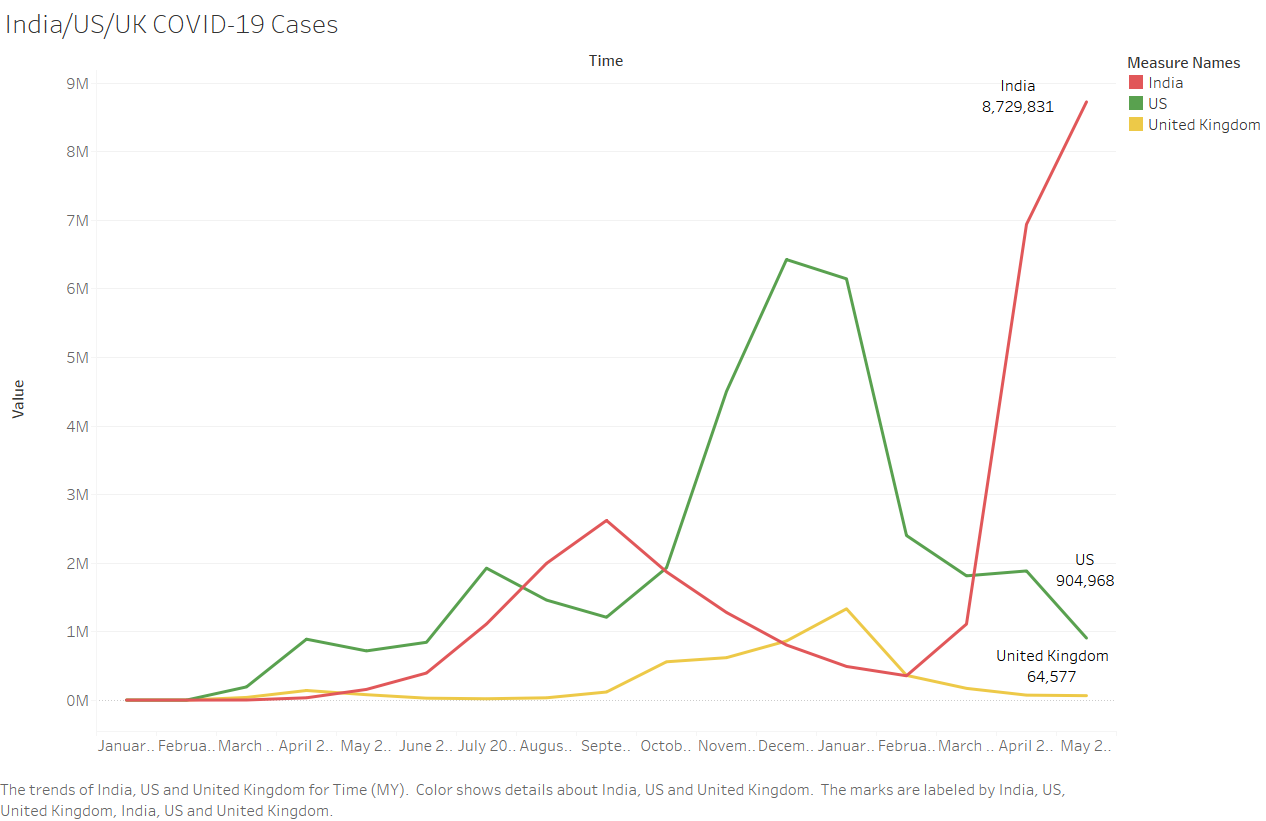
“Foreigners' Guide to Quarantine and Getting Vaccinated in China.” Nikkei Asia, Nikkei Asia, 1 Apr. 2021, <https://asia.nikkei.com/Spotlight/Caixin/Foreigners-guide-to-quarantine-and-getting-vaccinated-in-China>

<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>

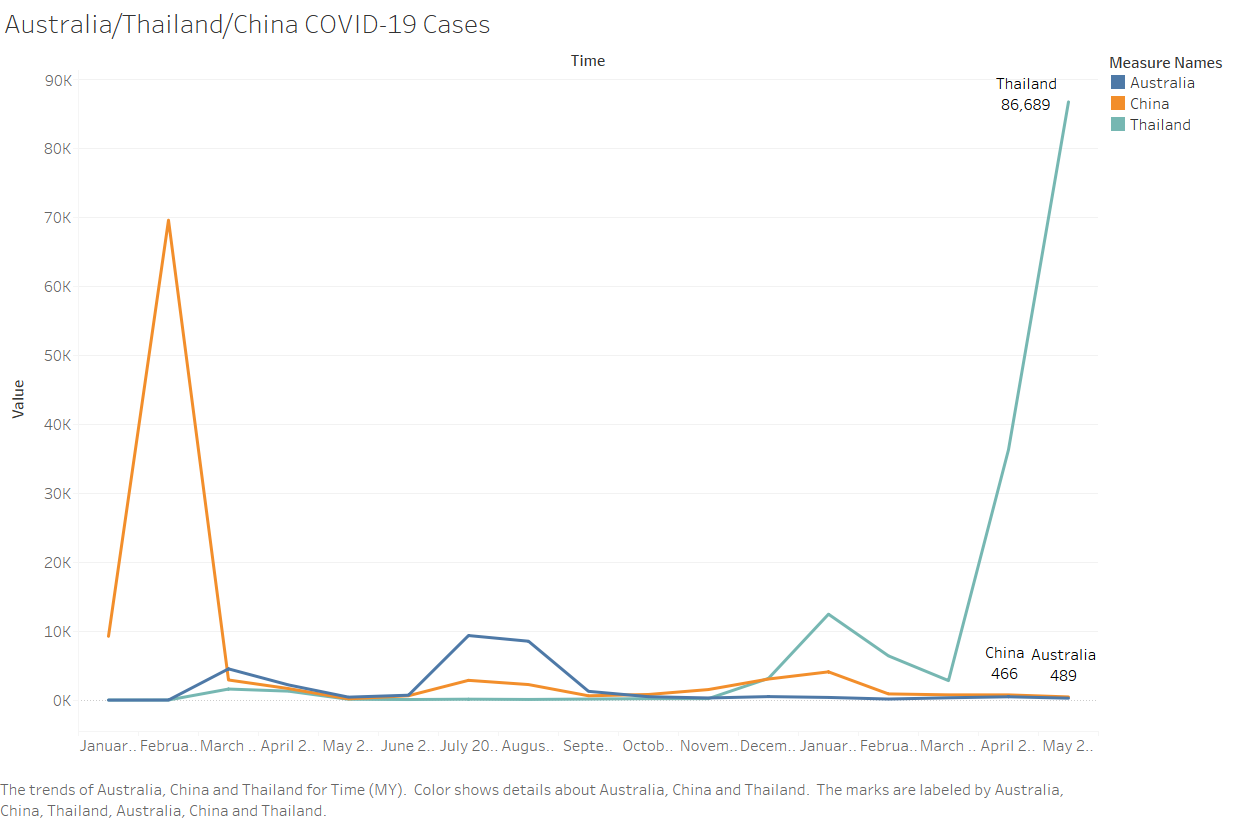
<https://www.cdc.gov/coronavirus/2019-ncov/your-health/quarantine-isolation.html>

Appendix

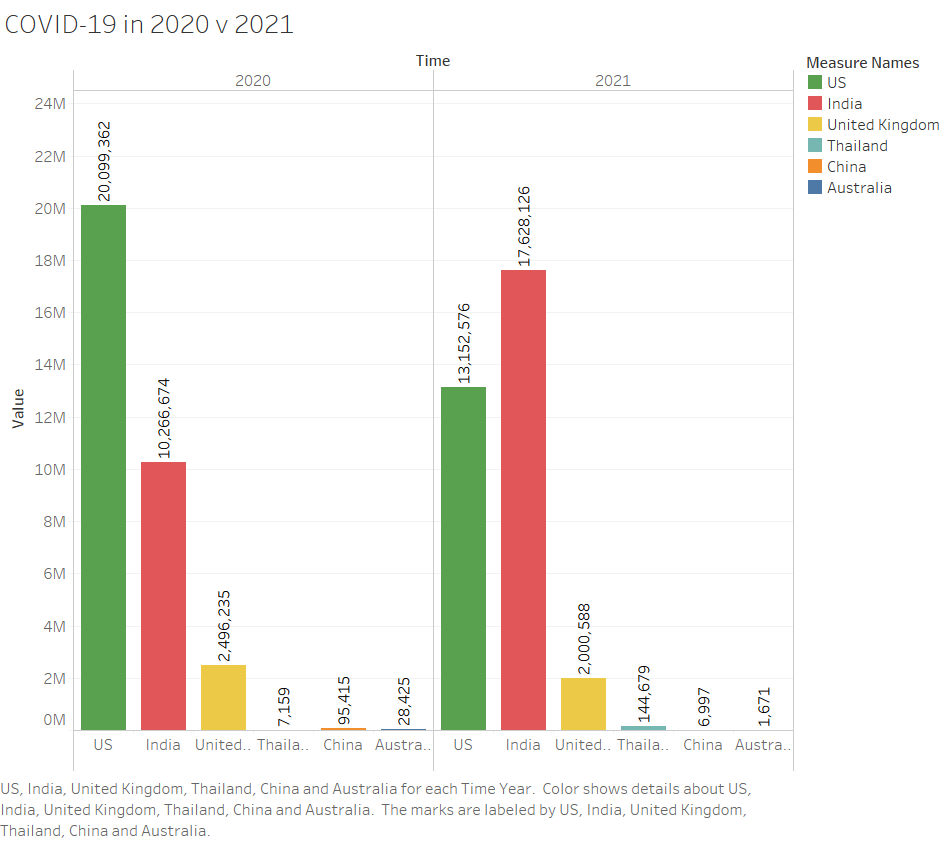
**Exhibit #1: COVID-19 Cases Times Series Trend Line (India, United States, United Kingdom)**



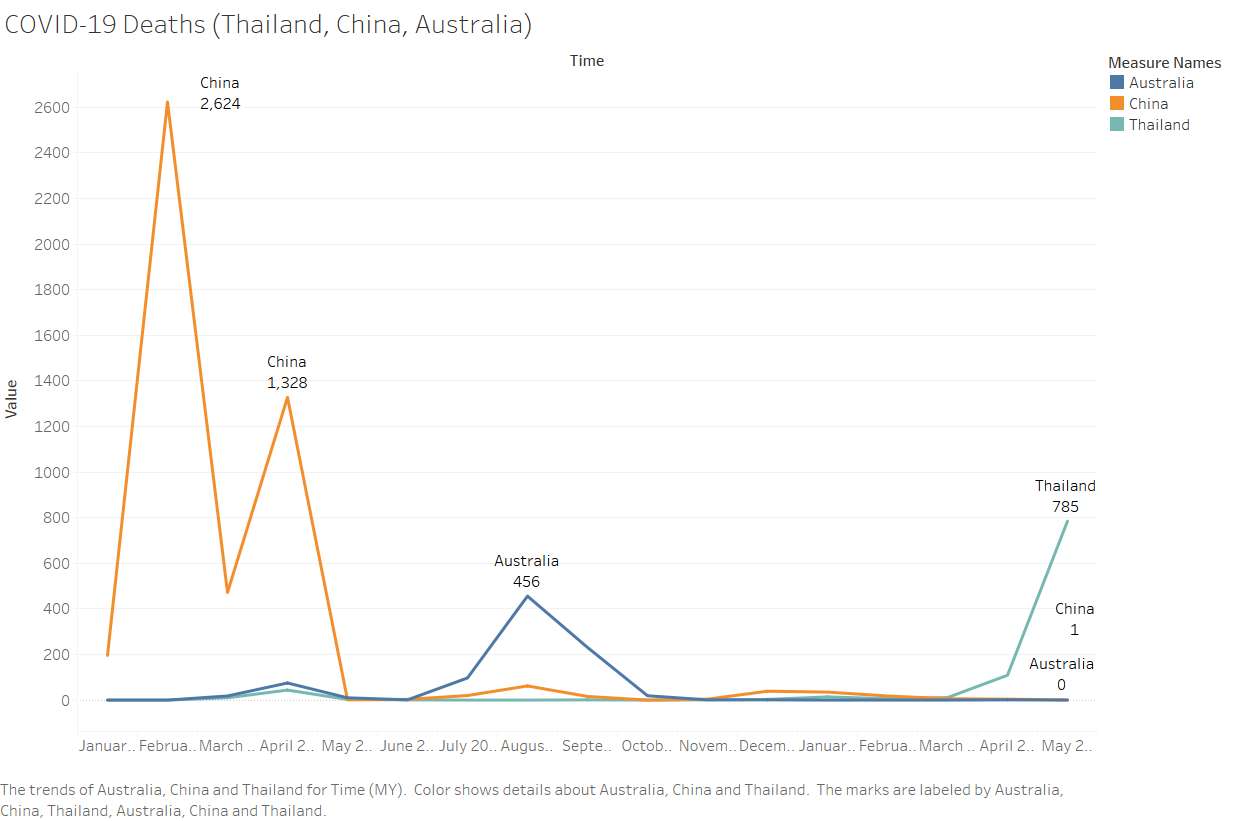
**Exhibit #2: COVID-19 Cases Times Series Trend Line (Australia, China, Thailand)**

****

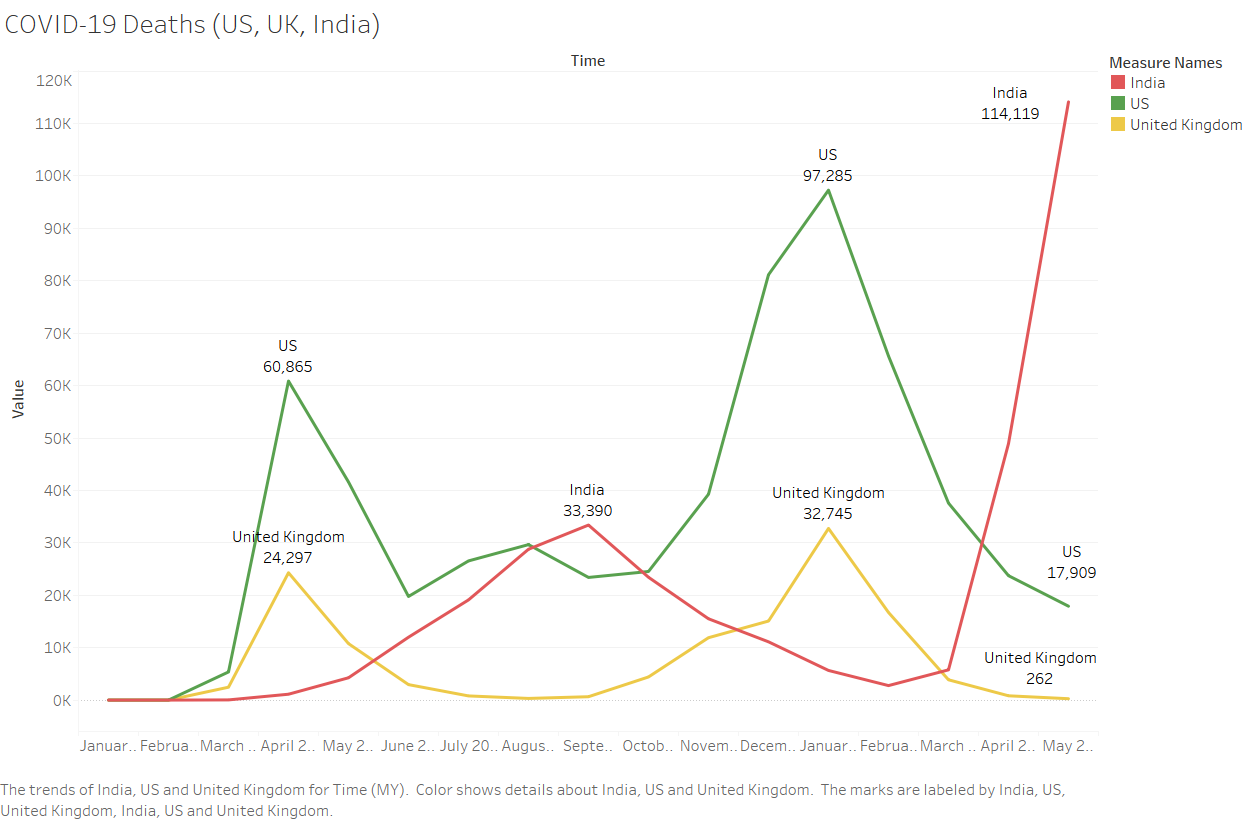
**Exhibit #3: COVID-19 Cases 2020 v. 2021**

****

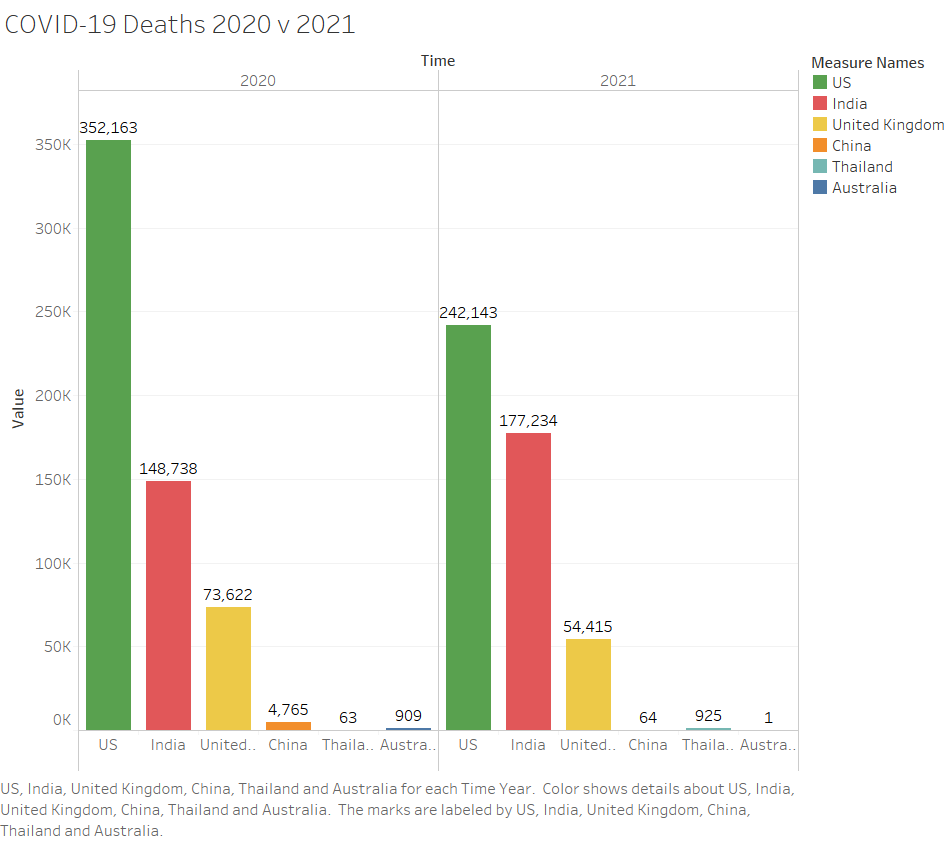
**Exhibit #4: COVID-19 Deaths (Australia, China, Thailand)**

****

**Exhibit #5: COVID-19 Deaths (India, United States, United Kingdom)**

****

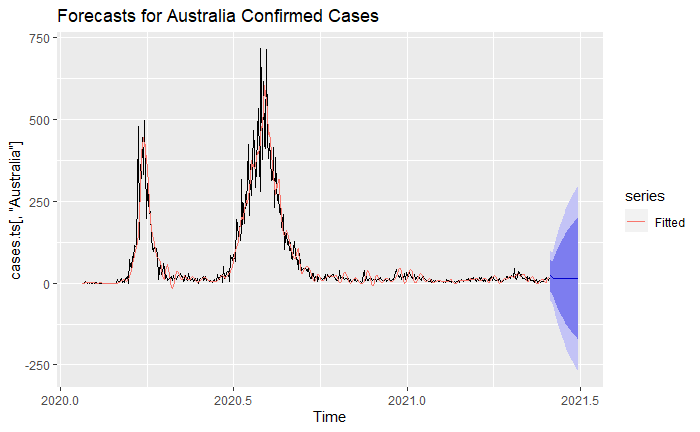
**Exhibit #6: COVID-19 Deaths 2020 v. 2021**

****

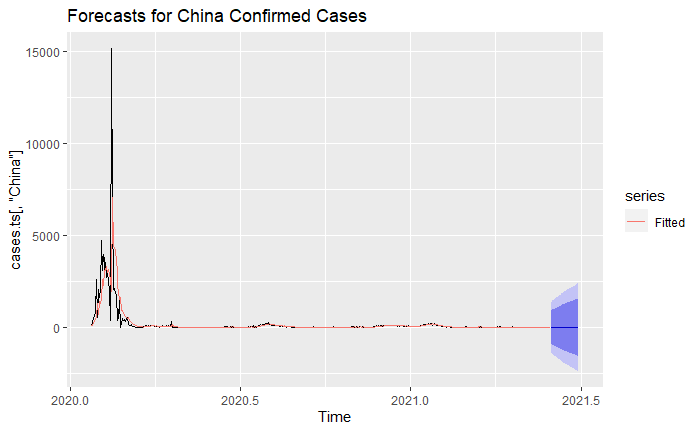
**Exhibit #7: Map of Nations**

****

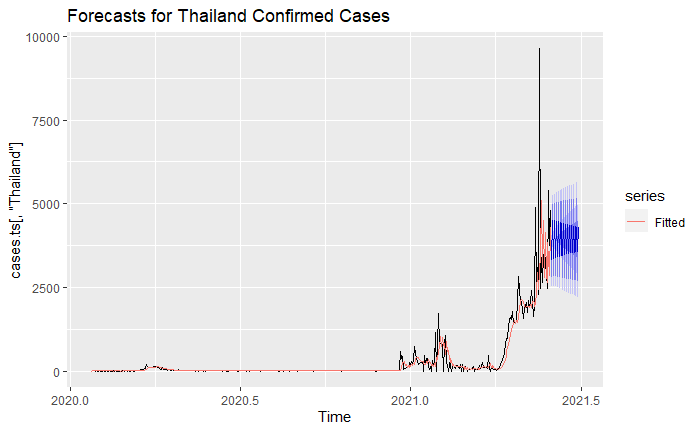
**Exhibit #8: Forecasts for Australia Confirmed Cases**

****

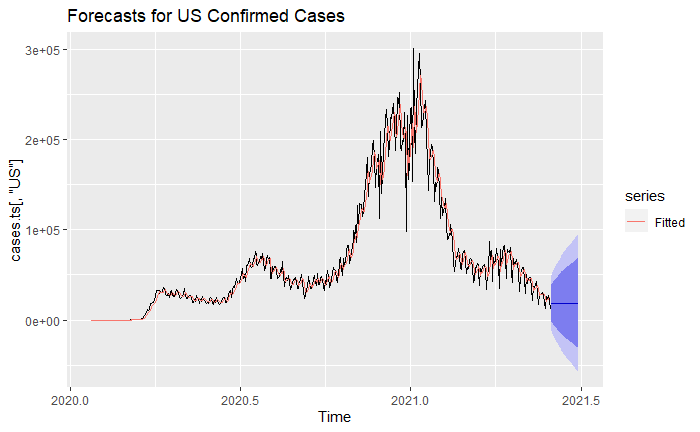
**Exhibit #9: Forecasts for China Confirmed Cases**

****

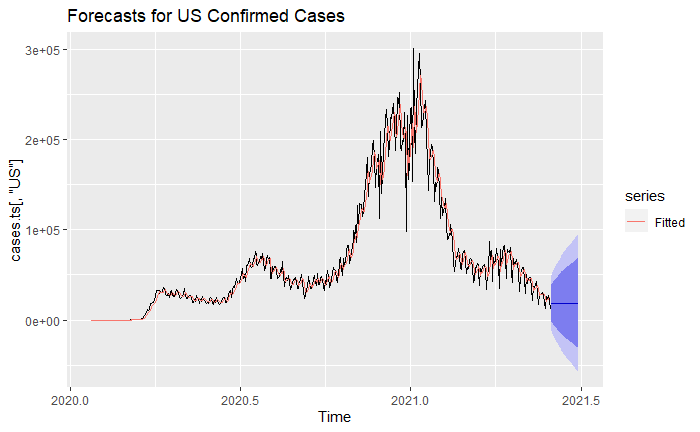
**Exhibit #10: Forecasts for Thailand Confirmed Cases**

****

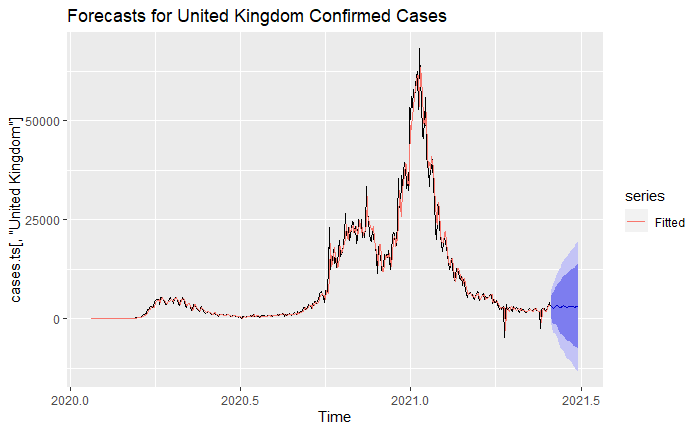
**Exhibit #11: Forecasts for Thailand Confirmed Cases**

****

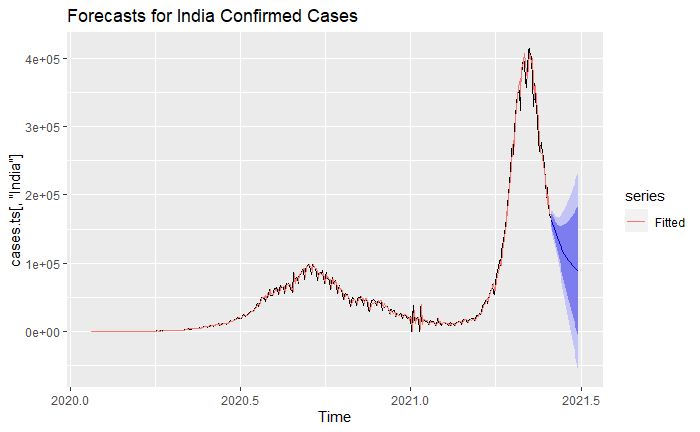
**Exhibit #12: Forecasts for US Confirmed Cases**

****

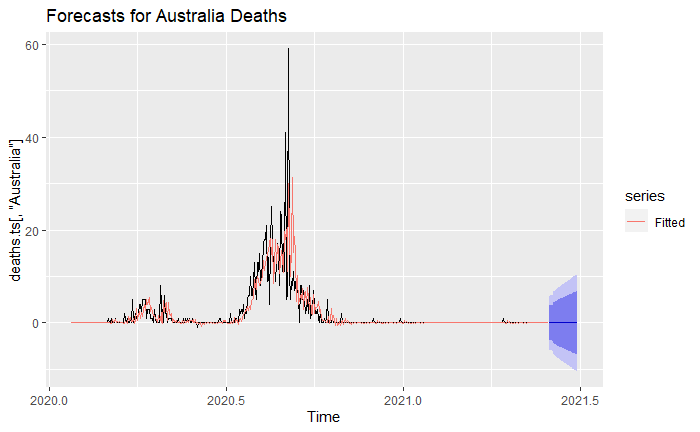
**Exhibit #13: Forecasts for UK Confirmed Cases**

****

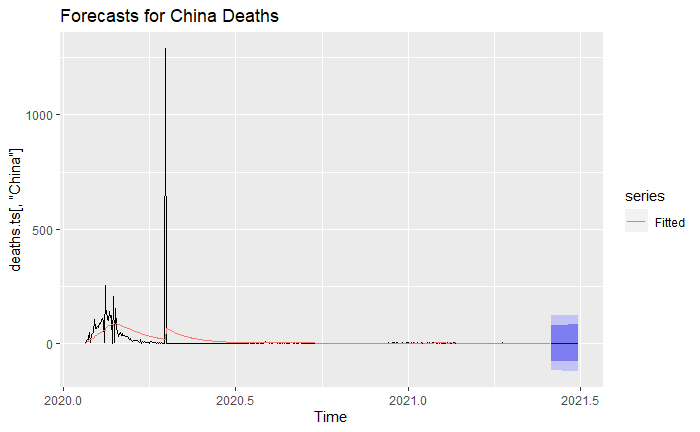
**Exhibit #14: Forecasts for India Confirmed Cases**

****

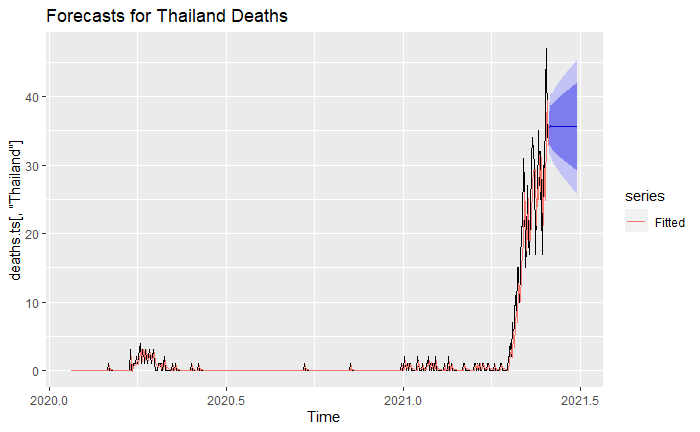
**Exhibit #15: Forecasts for Australia Deaths**

****

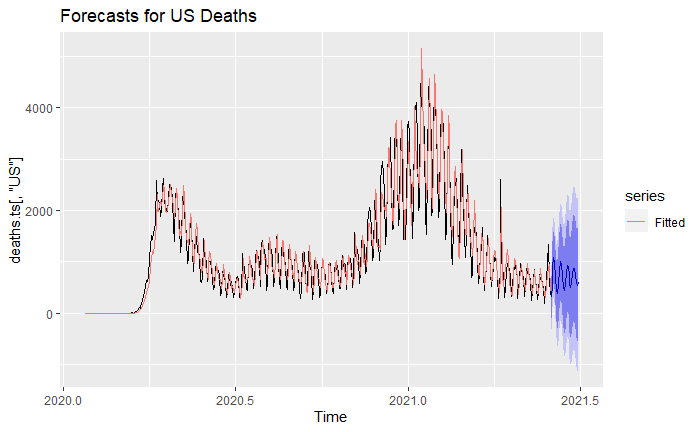
**Exhibit #16: Forecasts for China Deaths**

****

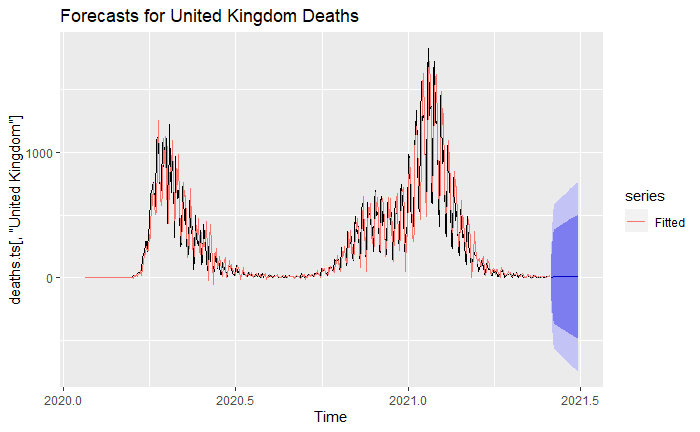
**Exhibit #17: Forecasts for Thailand Deaths**

****

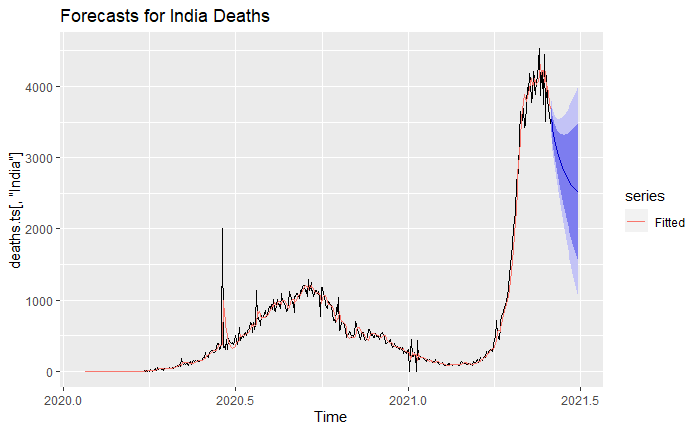
**Exhibit #18: Forecasts for US Deaths**

****

**Exhibit #19: Forecasts for UK Deaths**

****

**Exhibit #20: Forecasts for India Deaths**

****

R Script:

|  |
| --- |
| ---  title: "covid project 11/16/21 TS"  output: html\_document  ---  ```{r setup, include=FALSE}  knitr::opts\_chunk$set(echo = TRUE)  #prework  rm(list=ls())  gc()  cat('\f')  library('fpp2')  library('dplyr')  library('janitor')  ```  ```{r}  setwd('C:/Users/nkheth1/Desktop/Time Series and Forecasting/Final Project')  #daily.covid.worldwide=read.csv('covid\_19\_data.csv')  global.cases <- read.csv('time\_series\_covid\_19\_confirmed.csv')  #us.cases=read.csv('time\_series\_covid\_19\_confirmed\_US.csv')  global.deaths <- read.csv('time\_series\_covid\_19\_deaths.csv')  #us.deaths=read.csv('time\_series\_covid\_19\_deaths\_US.csv')  global.recovered <- read.csv('time\_series\_covid\_19\_recovered.csv')  ```  ```{r, echo=FALSE}  #df.list <- list(daily.covid.worldwide,global.cases,us.cases,global.deaths,us.deaths,global.recovered)  #lapply(df.list,anyNA  #daily.covid.worldwid=daily.covid.worldwide$SNo=NULL # just an ID variable, not useful for us  #confirmed cases data cleaning  global.cases <- global.cases[,-c(1,3,4)] #removing columns for region, latitude, and longitude  global.cases <- global.cases %>% #sum different regional data under single country data  group\_by(Country.Region) %>%  summarise\_all(sum)  global.cases.m <- as.matrix(global.cases)  global.cases.m.t <- t(global.cases.m) #convert data frame into matrix and transpose into columns  global.cases=as.data.frame(global.cases.m.t) #convert back into data frame  global.cases <- global.cases %>% #use country (row 1) as column and variable names  row\_to\_names(row\_number(1))  global.cases <- sapply(global.cases, as.numeric) #convert data frame from string to numeric  anyNA(global.cases) #FALSE  cases.ts <- ts(global.cases[,c("Australia","China","Thailand", "US", "United Kingdom", "India")],  start = c(2020,23),  frequency = 365)  cases.ts <- diff(cases.ts, 1)  global.cases <- as.data.frame(cases.ts)  write.csv(global.cases, "C://Users/nkheth1/Desktop/Time Series and Forecasting/Final Project/globalcases.csv",  row.names = FALSE)  #confirmed deaths data cleaning  global.deaths <- global.deaths[,-c(1,3,4)]  global.deaths <- global.deaths %>%  group\_by(Country.Region) %>%  summarise\_all(sum)  global.deaths.m <- as.matrix(global.deaths)  global.deaths.m.t <- t(global.deaths.m)  global.deaths=as.data.frame(global.deaths.m.t)  global.deaths <- global.deaths %>%  row\_to\_names(row\_number(1))  global.deaths <- sapply(global.deaths, as.numeric)  anyNA(global.deaths) #FALSE  deaths.ts <- ts(global.deaths[,c("Australia","China","Thailand", "US", "United Kingdom", "India")],  start = c(2020,23),  frequency = 365)  deaths.ts <- diff(deaths.ts, 1)  global.deaths <- as.data.frame(deaths.ts)  write.csv(global.deaths, "C://Users/nkheth1/Desktop/Time Series and Forecasting/Final Project/globaldeaths.csv",  row.names = FALSE)  #confirmed recovered data cleaning  global.recovered <- global.recovered[,-c(1,3,4)]  global.recovered <- global.recovered %>%  group\_by(Country.Region) %>%  summarise\_all(sum)  global.recovered.m <- as.matrix(global.recovered)  global.recovered.m.t <- t(global.recovered.m)  global.recovered <- as.data.frame(global.recovered.m.t)  global.recovered <- global.recovered %>%  row\_to\_names(row\_number(1))  global.recovered <- sapply(global.recovered, as.numeric)  anyNA(global.recovered) #FALSE  recovered.ts <- ts(global.recovered[,c("Australia","China","Thailand", "US", "United Kingdom", "India")],  start = c(2020,23),  frequency = 365)  recovered.ts <- diff(recovered.ts, 1)  global.recovered <- as.data.frame(recovered.ts)  write.csv(global.recovered, "C://Users/nkheth1/Desktop/Time Series and Forecasting/Final Project/globalrecovered.csv",  row.names = FALSE)  ```  ```{r, echo = FALSE}  ###Time Series Elements and Partitioning  autoplot(cases.ts)  ggAcf(global.cases[,"Australia"], lag.max = 60) #trend  ggAcf(global.cases[,"China"], lag.max = 60) #trend  ggAcf(global.cases[,"Thailand"], lag.max = 60) #trend  ggAcf(global.cases[,"US"], lag.max = 60) #trend  ggAcf(global.cases[,"United Kingdom"], lag.max = 60) #trend  ggAcf(global.cases[,"India"], lag.max = 60) #trend  autoplot(deaths.ts)  ggAcf(global.deaths[,"Australia"], lag.max = 60) #trend  ggAcf(global.deaths[,"China"], lag.max = 60) #trend  ggAcf(global.deaths[,"Thailand"], lag.max = 60) #trend  ggAcf(global.deaths[,"US"], lag.max = 60) #trend, seems to have seasonality frequency 7  ggAcf(global.deaths[,"United Kingdom"], lag.max = 60) #trend, also seem to have seasonality frequency 7  ggAcf(global.deaths[,"India"], lag.max = 60) #trend  autoplot(recovered.ts)  ggAcf(global.recovered[,"Australia"], lag.max = 60) #trend  ggAcf(global.recovered[,"China"], lag.max = 60) #trend  ggAcf(global.recovered[,"Thailand"], lag.max = 60) #trend  ggAcf(global.recovered[,"US"], lag.max = 60) #unreliable data  ggAcf(global.recovered[,"United Kingdom"], lag.max = 60) #trend  ggAcf(global.recovered[,"India"], lag.max = 60) #trend  cases.test <- tail(cases.ts, max(nrow(cases.ts)\*0.2, 30))  cases.train <- head(cases.ts, nrow(cases.ts) - nrow(cases.test))  deaths.test <- tail(deaths.ts, max(nrow(cases.ts)\*0.2, 30))  deaths.train <- head(deaths.ts, nrow(deaths.ts) - nrow(cases.test))  recovered.test <- tail(recovered.ts, max(nrow(recovered.ts)\*0.2, 30))  recovered.train <- head(recovered.ts, nrow(recovered.ts) - nrow(recovered.test))  ```  ```{r, echo = FALSE}  ### Model Building and Evaluation for Confirmed Cases  ##Australia  cases.drift1 <- rwf(cases.train[,"Australia"], nrow(cases.test), drift = TRUE)  checkresiduals(cases.drift1) #not adequate  accuracy(cases.drift1, cases.test[,"Australia"])[2,] #RMSE = 11.825, MAPE = 74.03  cases.ses1 <- ses(cases.train[,"Australia"], nrow(cases.test))  checkresiduals(cases.ses1) #not adequate  accuracy(cases.ses1, cases.test[,"Australia"])[2,] #RMSE=10.892, MAPE = 63.101  cases.holt1 <- holt(cases.train[,"Australia"], nrow(cases.test))  checkresiduals(cases.holt1) #not adequate  accuracy(cases.holt1, cases.test[,"Australia"])[2,] #RMSE = 52.426, MAPE = 505.355  cases.arima1 <- auto.arima(cases.train[,"Australia"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  cases.arima1 #(p,d,q) == (1,1,3) AIC = 4076.03  checkresiduals(cases.arima1) #not adequate  cases.pred1 <- forecast(cases.arima1, nrow(cases.test))  accuracy(cases.pred1, cases.test[,"Australia"])[2,] #RMSE = 10.462, MAPE = 59.365  cases.arima1.1 <- auto.arima(cases.train[,"Australia"], seasonal = FALSE)  cases.arima1.1 #(p,d,q) == (1,1,3) AIC = 4076.03 (extremely high)  checkresiduals(cases.arima1.1) #not adequate  cases.pred1.1 <- forecast(cases.arima1.1, nrow(cases.test))  accuracy(cases.pred1.1, cases.test[,"Australia"])[2,] #RMSE = 10.462, MAPE = 59.365  ##China  cases.drift2 <- rwf(cases.train[,"China"], nrow(cases.test), drift = TRUE)  checkresiduals(cases.drift2) #not adequate  accuracy(cases.drift2, cases.test[,"China"])[2,] #RMSE = 14.412, MAPE = 46.748  cases.ses2 <- ses(cases.train[,"China"], nrow(cases.test))  checkresiduals(cases.ses2) #not adequate  accuracy(cases.ses2, cases.test[,"China"])[2,] #RMSE= 9.94, MAPE = 36.582  cases.holt2 <- holt(cases.train[,"China"], nrow(cases.test))  checkresiduals(cases.holt2) #not adequate  accuracy(cases.holt2, cases.test[,"China"])[2,] #RMSE = 95.072, MAPE = 471.443  cases.arima2 <- auto.arima(cases.train[,"China"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  cases.arima2 #(p,d,q) == (0,1,2) AIC = 6357.25  checkresiduals(cases.arima2) #not adequate  cases.pred2 <- forecast(cases.arima2, nrow(cases.test))  accuracy(cases.pred2, cases.test[,"China"])[2,] #RMSE = 10.173, MAPE = 43.823  cases.arima2.1 <- auto.arima(cases.train[,"China"], seasonal = FALSE)  cases.arima2.1 #(p,d,q) == (2,1,1) AIC = 4076.03 (extremely high)  checkresiduals(cases.arima2.1) #not adequate  cases.pred2.1 <- forecast(cases.arima2.1, nrow(cases.test))  accuracy(cases.pred2.1, cases.test[,"China"])[2,] #RMSE = 9.936, MAPE = 37.203  ##Thailand  cases.drift3 <- rwf(cases.train[,"Thailand"], nrow(cases.test), drift = TRUE)  checkresiduals(cases.drift3) #not adequate  accuracy(cases.drift3, cases.test[,"Thailand"])[2,] #RMSE = 1898.542, MAPE = Infinite  cases.ses3 <- ses(cases.train[,"Thailand"], nrow(cases.test))  checkresiduals(cases.ses3) #not adequate  accuracy(cases.ses3, cases.test[,"Thailand"])[2,] #RMSE= 1919.253, MAPE = Infinite  cases.holt3 <- holt(cases.train[,"Thailand"], nrow(cases.test))  checkresiduals(cases.holt3) #not adequate  accuracy(cases.holt3, cases.test[,"Thailand"])[2,] #RMSE = 1891.844, MAPE = Infinite  cases.arima3 <- auto.arima(cases.train[,"Thailand"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  cases.arima3 #(p,d,q) == (2,1,3) AIC = 4883.82  checkresiduals(cases.arima3) #not adequate  cases.pred3 <- forecast(cases.arima3, nrow(cases.test))  accuracy(cases.pred3, cases.test[,"Thailand"])[2,] #RMSE = 1843.526, MAPE = Infinite  #Rough search returns the same model as exhaustive search  cases.arima3.1 <- auto.arima(cases.train[,"Thailand"], seasonal = FALSE)  cases.arima3.1 #(p,d,q) == (2,1,3) AIC = 4883.32 (extremely high)  checkresiduals(cases.arima3.1) #not adequate  cases.pred3.1 <- forecast(cases.arima3.1, nrow(cases.test))  accuracy(cases.pred3.1, cases.test[,"Thailand"])[2,] #RMSE = 1843.526, MAPE = Infinite  ##US  cases.drift4 <- rwf(cases.train[,"US"], nrow(cases.test), drift = TRUE)  checkresiduals(cases.drift4) #not adequate, does not eliminate trend from the data  accuracy(cases.drift4, cases.test[,"US"])[2,] #RMSE = 4.3x10^4, MAPE = 1.08x10^2  cases.ses4 <- ses(cases.train[,"US"], nrow(cases.test))  checkresiduals(cases.ses4) #not adequate, seasonality in the residuals  accuracy(cases.ses4, cases.test[,"US"])[2,] #RMSE= 2.75x10^4, MAPE = 68.8  cases.holt4 <- holt(cases.train[,"US"], nrow(cases.test))  checkresiduals(cases.holt4) #not adequate, seems to have some seasonality in residuals  accuracy(cases.holt4, cases.test[,"US"])[2,] #RMSE = 1.546x10^5, MAPE = 3.57x10^2  cases.arima4 <- auto.arima(cases.train[,"US"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  cases.arima4 #(p,d,q) == (3,1,2) AIC = 8700.13  checkresiduals(cases.arima4) #not adequate  cases.pred4 <- forecast(cases.arima4, nrow(cases.test))  accuracy(cases.pred4, cases.test[,"US"])[2,] #RMSE = 3.18x10^4, MAPE = 78.41  cases.arima4.1 <- auto.arima(cases.train[,"US"], seasonal = FALSE)  cases.arima4.1 #(p,d,q) == (0,1,1) AIC = 8756.75  checkresiduals(cases.arima4.1) #not adequate  cases.pred4.1 <- forecast(cases.arima4.1, nrow(cases.test))  accuracy(cases.pred4.1, cases.test[,"US"])[2,] #RMSE = 2.753x10^4, MAPE = 68.81  ##United Kingdom  cases.drift5 <- rwf(cases.train[,"United Kingdom"], nrow(cases.test), drift = TRUE)  checkresiduals(cases.drift5) #not adequate, does not eliminate trend from the data  accuracy(cases.drift5, cases.test[,"United Kingdom"])[2,] #RMSE = 1.022x10^4, MAPE = 3.415x10^2  cases.ses5 <- ses(cases.train[,"United Kingdom"], nrow(cases.test))  checkresiduals(cases.ses5) #not adequate, seasonality in the residuals  accuracy(cases.ses5, cases.test[,"United Kingdom"])[2,] #RMSE= 8.536x10^3, MAPE = 2.836x10^2  cases.holt5 <- holt(cases.train[,"United Kingdom"], nrow(cases.test))  checkresiduals(cases.holt5) #not adequate, seems to have some seasonality in residuals  accuracy(cases.holt5, cases.test[,"United Kingdom"])[2,] #RMSE = 8.589x10^3, MAPE = 2.855x10^2  cases.arima5 <- auto.arima(cases.train[,"United Kingdom"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  cases.arima5 #(p,d,q) == (3,1,2) AIC = 7252.98  checkresiduals(cases.arima5) #not adequate  cases.pred5 <- forecast(cases.arima5, nrow(cases.test))  accuracy(cases.pred5, cases.test[,"United Kingdom"])[2,] #RMSE = 8.305x10^3, MAPE = 2.76x10^2  cases.arima5.1 <- auto.arima(cases.train[,"United Kingdom"], seasonal = FALSE)  cases.arima5.1 #(p,d,q) == (0,1,1) AIC = 7289.06  checkresiduals(cases.arima5.1) #not adequate  cases.pred5.1 <- forecast(cases.arima5.1, nrow(cases.test))  accuracy(cases.pred5.1, cases.test[,"United Kingdom"])[2,] #RMSE = 8.536x10^3, MAPE = 2.836x10^2  ##India  cases.drift6 <- rwf(cases.train[,"India"], nrow(cases.test), drift = TRUE)  checkresiduals(cases.drift6) #not adequate, seems to have seasonality in the data  accuracy(cases.drift6, cases.test[,"India"])[2,] #RMSE = 2.072x10^5, MAPE = 71.08  cases.ses6 <- ses(cases.train[,"India"], nrow(cases.test))  checkresiduals(cases.ses6) #not adequate, seasonality in the residuals  accuracy(cases.ses6, cases.test[,"India"])[2,] #RMSE= 2.097x10^5, MAPE = 73.933  cases.holt6 <- holt(cases.train[,"India"], nrow(cases.test))  checkresiduals(cases.holt6) #not adequate, seems to have some seasonality in residuals  accuracy(cases.holt6, cases.test[,"India"])[2,] #RMSE = 2.023X10^5, MAPE = 6.886  cases.arima6 <- auto.arima(cases.train[,"India"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  cases.arima6 #(p,d,q) == (2,1,3) AIC = 7665.03  checkresiduals(cases.arima6) #not adequate, seasonality in the residuals  cases.pred6 <- forecast(cases.arima6, nrow(cases.test))  accuracy(cases.pred6, cases.test[,"India"])[2,] #RMSE = 2.076x10^5, MAPE = 71.73  #Rough search returns same model as exhaustive search  cases.arima6.1 <- auto.arima(cases.train[,"India"], seasonal = FALSE)  cases.arima6.1 #(p,d,q) == (2,1,3) AIC = 7665/03  checkresiduals(cases.arima6.1) #not adequate  cases.pred6.1 <- forecast(cases.arima6.1, nrow(cases.test))  accuracy(cases.pred6.1, cases.test[,"India"])[2,] #RMSE = 2.076x10^5, MAPE = 71.73  ```  ```{r, echo = FALSE}  ### Model Building and Evaluation for Deaths  ##Australia  deaths.drift1 <- rwf(deaths.train[,"Australia"], nrow(deaths.test), drift = TRUE)  checkresiduals(deaths.drift1) #not adequate  accuracy(deaths.drift1, deaths.test[,"Australia"])[2,] #RMSE = 0.101, MAPE = 100.00  deaths.ses1 <- ses(deaths.train[,"Australia"], nrow(deaths.test))  checkresiduals(deaths.ses1) #not adequate  accuracy(deaths.ses1, deaths.test[,"Australia"])[2,] #RMSE= 0.101, MAPE = Infinite  deaths.holt1 <- holt(deaths.train[,"Australia"], nrow(deaths.test))  checkresiduals(deaths.holt1) #not adequate  accuracy(deaths.holt1, deaths.test[,"Australia"])[2,] #RMSE = 0.10, MAPE = Infinite  deaths.arima1 <- auto.arima(deaths.train[,"Australia"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  deaths.arima1 #(p,d,q) == (0,1,5) AIC = 2059.79  checkresiduals(deaths.arima1) #not adequate  deaths.pred1 <- forecast(deaths.arima1, nrow(deaths.test))  accuracy(deaths.pred1, deaths.test[,"Australia"])[2,] #RMSE = 0.101, MAPE = Infinite  #Rough search returns the same model as exhaustive search  deaths.arima1.1 <- auto.arima(deaths.train[,"Australia"], seasonal = FALSE)  deaths.arima1.1 #(p,d,q) == (3,1,2) AIC = 2076.63 (extremely high)  checkresiduals(deaths.arima1.1) #not adequate  deaths.pred1.1 <- forecast(deaths.arima1.1, nrow(deaths.test))  accuracy(deaths.pred1.1, deaths.test[,"Australia"])[2,] #RMSE = 0.101, MAPE = Infinite  autoplot(deaths.ts[,"Australia"])  ##China  autoplot(deaths.ts[,"China"])  deaths.drift2 <- rwf(deaths.train[,"China"], nrow(deaths.test), drift = TRUE)  checkresiduals(deaths.drift2) #not adequate  accuracy(deaths.drift2, deaths.test[,"China"])[2,] #RMSE = 0.441, MAPE = Infinite  deaths.ses2 <- ses(deaths.train[,"China"], nrow(deaths.test))  checkresiduals(deaths.ses2) #not adequate  accuracy(deaths.ses2, deaths.test[,"China"])[2,] #RMSE= 0.932, MAPE = Infinite  deaths.holt2 <- holt(deaths.train[,"China"], nrow(deaths.test))  checkresiduals(deaths.holt2) #not adequate  accuracy(deaths.holt2, deaths.test[,"China"])[2,] #RMSE = 3.303, MAPE = Infinite  deaths.arima2 <- auto.arima(deaths.train[,"China"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  deaths.arima2 #(p,d,q) == (0,1,1) AIC = 4446.8  checkresiduals(deaths.arima2) #not adequate  deaths.pred2 <- forecast(deaths.arima2, nrow(deaths.test))  accuracy(deaths.pred2, deaths.test[,"China"])[2,] #RMSE = 0.39, MAPE = Infinite  #Rough search returns the same model as exhaustive search  deaths.arima2.1 <- auto.arima(deaths.train[,"China"], seasonal = FALSE)  deaths.arima2.1 #(p,d,q) == (1,1,2) AIC = 4449.21  checkresiduals(deaths.arima2.1) #not adequate  deaths.pred2.1 <- forecast(deaths.arima2.1, nrow(deaths.test))  accuracy(deaths.pred2.1, deaths.test[,"China"])[2,] #RMSE = 0.912, MAPE = Infinite  ##Thailand  autoplot(deaths.ts[,"Thailand"])  deaths.drift3 <- rwf(deaths.train[,"Thailand"], nrow(deaths.test), drift = TRUE)  checkresiduals(deaths.drift3) #not adequate  accuracy(deaths.drift3, deaths.test[,"Thailand"])[2,] #RMSE = 15.574, MAPE = 100  deaths.ses3 <- ses(deaths.train[,"Thailand"], nrow(deaths.test))  checkresiduals(deaths.ses3) #not adequate  accuracy(deaths.ses3, deaths.test[,"Thailand"])[2,] #RMSE= 15.361, MAPE = Infinite  deaths.holt3 <- holt(deaths.train[,"Thailand"], nrow(deaths.test))  checkresiduals(deaths.holt3) #not adequate  accuracy(deaths.holt3, deaths.test[,"Thailand"])[2,] #RMSE = 15.313, MAPE = Infinite  deaths.arima3 <- auto.arima(deaths.train[,"Thailand"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  deaths.arima3 #(p,d,q) == (2,1,1) AIC = 497.44  checkresiduals(deaths.arima3) #not adequate  deaths.pred3 <- forecast(deaths.arima3, nrow(deaths.test))  accuracy(deaths.pred3, deaths.test[,"Thailand"])[2,] #RMSE = 15.362, MAPE = Infinite  #Rough search returns the same model as exhaustive search  deaths.arima3.1 <- auto.arima(deaths.train[,"Thailand"], seasonal = FALSE)  deaths.arima3.1 #(p,d,q) == (2,1,1) AIC = 497.44  checkresiduals(deaths.arima3.1) #not adequate  deaths.pred3.1 <- forecast(deaths.arima3.1, nrow(deaths.test))  accuracy(deaths.pred3.1, deaths.test[,"Thailand"])[2,] #RMSE = 15.362, MAPE = Infinite  ##US  deaths.drift4 <- rwf(deaths.train[,"US"], nrow(deaths.test), drift = TRUE)  checkresiduals(deaths.drift4) #not adequate, there is some seasonality in the residuals  accuracy(deaths.drift4, deaths.test[,"US"])[2,] #RMSE = 2.116x10^3, MAPE = 3.276x10^2  deaths.ses4 <- ses(deaths.train[,"US"], nrow(deaths.test))  checkresiduals(deaths.ses4) #not adequate, there is some seasonality in the residuals  accuracy(deaths.ses4, deaths.test[,"US"])[2,] #RMSE= 1.754x10^3, MAPE = 2.714x10^2  deaths.holt4 <- holt(deaths.train[,"US"], nrow(deaths.test))  checkresiduals(deaths.holt4) #not adequate, still some seasonality left  accuracy(deaths.holt4, deaths.test[,"US"])[2,] #RMSE = 1.761x10^3, MAPE = 2.726x10^2  deaths.arima4 <- auto.arima(deaths.train[,"US"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  deaths.arima4 #(p,d,q) == (5,1,0) AIC = 5605.79  checkresiduals(deaths.arima4) #not adequate  deaths.pred4 <- forecast(deaths.arima4, nrow(deaths.test))  accuracy(deaths.pred4, deaths.test[,"US"])[2,] #RMSE = 1.144x10^3, MAPE = 1.79x10^2  #Rough search returns the same model as exhaustive search  deaths.arima4.1 <- auto.arima(deaths.train[,"US"], seasonal = FALSE)  deaths.arima4.1 #(p,d,q) == (3,1,2) AIC = 5664.33  checkresiduals(deaths.arima4.1) #not adequate  deaths.pred4.1 <- forecast(deaths.arima4.1, nrow(deaths.test))  accuracy(deaths.pred4.1, deaths.test[,"US"])[2,] #RMSE = 1.032x10^3, MAPE = 1.616x10^2  ##Unite Kingdom  deaths.drift5 <- rwf(deaths.train[,"United Kingdom"], nrow(deaths.test), drift = TRUE)  checkresiduals(deaths.drift5) #not adequate, there is some seasonality in the residuals  accuracy(deaths.drift5, deaths.test[,"United Kingdom"])[2,] #RMSE = 5.41x10^2, MAPE = 4.655x10^3  deaths.ses5 <- ses(deaths.train[,"United Kingdom"], nrow(deaths.test))  checkresiduals(deaths.ses5) #not adequate, there is some seasonality in the residuals  accuracy(deaths.ses5, deaths.test[,"United Kingdom"])[2,] #RMSE= 4.974x10^2, MAPE = 4.134x10^3  deaths.holt5 <- holt(deaths.train[,"United Kingdom"], nrow(deaths.test))  checkresiduals(deaths.holt5) #not adequate, still some seasonality left  accuracy(deaths.holt5, deaths.test[,"United Kingdom"])[2,] #RMSE = 1.689x10^2, MAPE = 1.402x10^3  deaths.arima5 <- auto.arima(deaths.train[,"United Kingdom"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  deaths.arima5 #(p,d,q) == (5,1,0) AIC = 4944.39  checkresiduals(deaths.arima5) #not adequate  deaths.pred5 <- forecast(deaths.arima5, nrow(deaths.test))  accuracy(deaths.pred5, deaths.test[,"United Kingdom"])[2,] #RMSE = 4.532x10^2, MAPE = 3.803x10^3  #Rough search returns the same model as exhaustive search  deaths.arima5.1 <- auto.arima(deaths.train[,"United Kingdom"], seasonal = FALSE)  deaths.arima5.1 #(p,d,q) == (1,1,5) AIC = 5003.89  checkresiduals(deaths.arima5.1) #not adequate  deaths.pred5.1 <- forecast(deaths.arima5.1, nrow(deaths.test))  accuracy(deaths.pred5.1, deaths.test[,"United Kingdom"])[2,] #RMSE = 4.215x10^2, MAPE = 3.548x10^3  ##India  deaths.drift6 <- rwf(deaths.train[,"India"], nrow(deaths.test), drift = TRUE)  checkresiduals(deaths.drift6) #not adequate  accuracy(deaths.drift6, deaths.test[,"India"])[2,] #RMSE = 2.303x10^3, MAPE = 67.7  deaths.ses6 <- ses(deaths.train[,"India"], nrow(deaths.test))  checkresiduals(deaths.ses6) #not adequate  accuracy(deaths.ses6, deaths.test[,"India"])[2,] #RMSE= 2.32x10^3, MAPE = 69.573  deaths.holt6 <- holt(deaths.train[,"India"], nrow(deaths.test))  checkresiduals(deaths.holt6) #not adequate  accuracy(deaths.holt6, deaths.test[,"India"])[2,] #RMSE = 2.452x10^3, MAPE = 87.9  deaths.arima6 <- auto.arima(deaths.train[,"India"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  deaths.arima6 #(p,d,q) == (1,1,3) AIC = 4830.06  checkresiduals(deaths.arima6) #not adequate  deaths.pred6 <- forecast(deaths.arima6, nrow(deaths.test))  accuracy(deaths.pred6, deaths.test[,"India"])[2,] #RMSE = 2.321x10^3, MAPE = 70.153  #Rough search returns the same model as exhaustive search  deaths.arima6.1 <- auto.arima(deaths.train[,"India"], seasonal = FALSE)  deaths.arima6.1 #(p,d,q) == (1,1,3) AIC = 5664.33  checkresiduals(deaths.arima6.1) #not adequate  deaths.pred6.1 <- forecast(deaths.arima6.1, nrow(deaths.test))  accuracy(deaths.pred6.1, deaths.test[,"India"])[2,] #RMSE = 2.321x10^3, MAPE = 70.153  ```  ```{r, echo = FALSE}  ### Model Building and Evaluation for Recovered  ##Australia  recovered.drift1 <- rwf(recovered.train[,"Australia"], nrow(recovered.test), drift = TRUE)  checkresiduals(recovered.drift1) #not adequate  accuracy(recovered.drift1, recovered.test[,"Australia"])[2,] #RMSE = 7.175, MAPE = Infinite  recovered.ses1 <- ses(recovered.train[,"Australia"], nrow(recovered.test))  checkresiduals(recovered.ses1) #not adequate  accuracy(recovered.ses1, recovered.test[,"Australia"])[2,] #RMSE= 7.767, MAPE = Infinite  recovered.holt1 <- holt(recovered.train[,"Australia"], nrow(recovered.test))  checkresiduals(recovered.holt1) #not adequate  accuracy(recovered.holt1, recovered.test[,"Australia"])[2,] #RMSE = 6.945, MAPE = Infinite  ###best MAPE  recovered.arima1 <- auto.arima(recovered.train[,"Australia"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  recovered.arima1 #(p,d,q) == (1,0,2) AIC = 5364.41  checkresiduals(recovered.arima1) #not adequate  recovered.pred1 <- forecast(recovered.arima1, nrow(recovered.test))  accuracy(recovered.pred1, recovered.test[,"Australia"])[2,] #RMSE = 44.965, MAPE = Infinite  #Rough search returns the same model as exhaustive search  recovered.arima1.1 <- auto.arima(recovered.train[,"Australia"], seasonal = FALSE)  recovered.arima1.1 #(p,d,q) == (1,0,2) AIC = 5364.41 (extremely high)  checkresiduals(recovered.arima1.1) #not adequate  recovered.pred1.1 <- forecast(recovered.arima1.1, nrow(recovered.test))  accuracy(recovered.pred1.1, recovered.test[,"Australia"])[2,] #RMSE = 44.965, MAPE = Infinite  ##China  autoplot(recovered.ts[,"China"])  recovered.drift2 <- rwf(recovered.train[,"China"], nrow(recovered.test), drift = TRUE)  checkresiduals(recovered.drift2) #not adequate  accuracy(recovered.drift2, recovered.test[,"China"])[2,] #RMSE = 61.791, MAPE = 307.04  recovered.ses2 <- ses(recovered.train[,"China"], nrow(recovered.test))  checkresiduals(recovered.ses2) #not adequate  accuracy(recovered.ses2, recovered.test[,"China"])[2,] #RMSE= 54.61, MAPE = 270.254  recovered.holt2 <- holt(recovered.train[,"China"], nrow(recovered.test))  checkresiduals(recovered.holt2) #not adequate  accuracy(recovered.holt2, recovered.test[,"China"])[2,] #RMSE = 351.246, MAPE = 1.61x10^3  recovered.arima2 <- auto.arima(recovered.train[,"China"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  recovered.arima2 #(p,d,q) == (0,1,2) AIC = 5296.13  checkresiduals(recovered.arima2) #not adequate  recovered.pred2 <- forecast(recovered.arima2, nrow(recovered.test))  accuracy(recovered.pred2, recovered.test[,"China"])[2,] #RMSE = 55.401, MAPE = 274.091  #Rough search returns the same model as exhaustive search  recovered.arima2.1 <- auto.arima(recovered.train[,"China"], seasonal = FALSE)  recovered.arima2.1 #(p,d,q) == (0,1,2) AIC = 5296.13  checkresiduals(recovered.arima2.1) #not adequate  recovered.pred2.1 <- forecast(recovered.arima2.1, nrow(recovered.test))  accuracy(recovered.pred2.1, recovered.test[,"China"])[2,] #RMSE = 55.401, MAPE = 274.091  ##Thailand  recovered.drift3 <- rwf(recovered.train[,"Thailand"], nrow(recovered.test), drift = TRUE)  checkresiduals(recovered.drift3) #not adequate  accuracy(recovered.drift3, recovered.test[,"Thailand"])[2,] #RMSE = 70.669, MAPE = Infinite  recovered.ses3 <- ses(recovered.train[,"Thailand"], nrow(recovered.test))  checkresiduals(recovered.ses3) #not adequate  accuracy(recovered.ses3, recovered.test[,"Thailand"])[2,] #RMSE= 435.89, MAPE = Infinite  recovered.holt3 <- holt(recovered.train[,"Thailand"], nrow(recovered.test))  checkresiduals(recovered.holt3) #not adequate  accuracy(recovered.holt3, recovered.test[,"Thailand"])[2,] #RMSE = 511.855, MAPE = Infinite  recovered.arima3 <- auto.arima(recovered.train[,"Thailand"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  recovered.arima3 #(p,d,q) == (0,1,5) AIC = 4956.9  checkresiduals(recovered.arima3) #not adequate  recovered.pred3 <- forecast(recovered.arima3, nrow(recovered.test))  accuracy(recovered.pred3, recovered.test[,"Thailand"])[2,] #RMSE = 312.458, MAPE = Infinite  #Rough search returns the same model as exhaustive search  recovered.arima3.1 <- auto.arima(recovered.train[,"Thailand"], seasonal = FALSE)  recovered.arima3.1 #(p,d,q) == (2,1,2) AIC = 4966.08  checkresiduals(recovered.arima3.1) #not adequate  recovered.pred3.1 <- forecast(recovered.arima3.1, nrow(recovered.test))  accuracy(recovered.pred3.1, recovered.test[,"Thailand"])[2,] #RMSE = 303.01, MAPE = Infinite  ##US  #recovered.drift4 <- rwf(recovered.train[,"US"], nrow(recovered.test), drift = TRUE)  #checkresiduals(recovered.drift4) #not adequate, there is some seasonality in the residuals  #accuracy(recovered.drift4, recovered.test[,"US"])[2,] #RMSE = 2.116x10^3, MAPE = 3.276x10^2  #recovered.ses4 <- ses(recovered.train[,"US"], nrow(recovered.test))  #checkresiduals(recovered.ses4) #not adequate, there is some seasonality in the residuals  #accuracy(recovered.ses4, recovered.test[,"US"])[2,] #RMSE= 1.754x10^3, MAPE = 2.714x10^2  #recovered.holt4 <- holt(recovered.train[,"US"], nrow(recovered.test))  #checkresiduals(recovered.holt4) #not adequate, still some seasonality left  #accuracy(recovered.holt4, recovered.test[,"US"])[2,] #RMSE = 1.761x10^3, MAPE = 2.726x10^2  #recovered.arima4 <- auto.arima(recovered.train[,"US"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  #recovered.arima4 #(p,d,q) == (5,1,0) AIC = 5605.79  #checkresiduals(recovered.arima4) #not adequate  #recovered.pred4 <- forecast(recovered.arima4, nrow(recovered.test))  #accuracy(recovered.pred4, recovered.test[,"US"])[2,] #RMSE = 1.144x10^3, MAPE = 1.79x10^2  #Rough search returns the same model as exhaustive search  #recovered.arima4.1 <- auto.arima(recovered.train[,"US"], seasonal = FALSE)  #recovered.arima4.1 #(p,d,q) == (3,1,2) AIC = 5664.33  #checkresiduals(recovered.arima4.1) #not adequate  #recovered.pred4.1 <- forecast(recovered.arima4.1, nrow(recovered.test))  #accuracy(recovered.pred4.1, recovered.test[,"US"])[2,] #RMSE = 1.032x10^3, MAPE = 1.616x10^2  ##Unite Kingdom  recovered.drift5 <- rwf(recovered.train[,"United Kingdom"], nrow(recovered.test), drift = TRUE)  checkresiduals(recovered.drift5) #not adequate, there is some seasonality in the residuals  accuracy(recovered.drift5, recovered.test[,"United Kingdom"])[2,] #RMSE = 72.28, MAPE = Infinite  recovered.ses5 <- ses(recovered.train[,"United Kingdom"], nrow(recovered.test))  checkresiduals(recovered.ses5) #not adequate, there is some seasonality in the residuals  accuracy(recovered.ses5, recovered.test[,"United Kingdom"])[2,] #RMSE= 58.287, MAPE = Infinite  recovered.holt5 <- holt(recovered.train[,"United Kingdom"], nrow(recovered.test))  checkresiduals(recovered.holt5) #not adequate, still some seasonality left  accuracy(recovered.holt5, recovered.test[,"United Kingdom"])[2,] #RMSE = 68.894, MAPE = Infinite  recovered.arima5 <- auto.arima(recovered.train[,"United Kingdom"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  recovered.arima5 #(p,d,q) == (1,1,4) AIC = 4019.25  checkresiduals(recovered.arima5) #not adequate  recovered.pred5 <- forecast(recovered.arima5, nrow(recovered.test))  accuracy(recovered.pred5, recovered.test[,"United Kingdom"])[2,] #RMSE = 57.819, MAPE = Infinite  #Rough search returns the same model as exhaustive search  recovered.arima5.1 <- auto.arima(recovered.train[,"United Kingdom"], seasonal = FALSE)  recovered.arima5.1 #(p,d,q) == (1,1,4) AIC = 4019.03  checkresiduals(recovered.arima5.1) #not adequate  recovered.pred5.1 <- forecast(recovered.arima5.1, nrow(recovered.test))  accuracy(recovered.pred5.1, recovered.test[,"United Kingdom"])[2,] #RMSE = 57.819, MAPE = Infinite  ##India  recovered.drift6 <- rwf(recovered.train[,"India"], nrow(recovered.test), drift = TRUE)  checkresiduals(recovered.drift6) #not adequate  accuracy(recovered.drift6, recovered.test[,"India"])[2,] #RMSE = 1.956x10^5, MAPE = 70.242  recovered.ses6 <- ses(recovered.train[,"India"], nrow(recovered.test))  checkresiduals(recovered.ses6) #not adequate  accuracy(recovered.ses6, recovered.test[,"India"])[2,] #RMSE= 1.965x10^5, MAPE = 70.237  recovered.holt6 <- holt(recovered.train[,"India"], nrow(recovered.test))  checkresiduals(recovered.holt6) #not adequate  accuracy(recovered.holt6, recovered.test[,"India"])[2,] #RMSE = 2.086x10^5, MAPE = 84.51  recovered.arima6 <- auto.arima(recovered.train[,"India"], seasonal = FALSE, stepwise = FALSE, approximation = FALSE)  recovered.arima6 #(p,d,q) == (2,2,3) AIC = 7632.19  checkresiduals(recovered.arima6) #not adequate  recovered.pred6 <- forecast(recovered.arima6, nrow(recovered.test))  accuracy(recovered.pred6, recovered.test[,"India"])[2,] #RMSE = 2.082x10^5, MAPE = 83.82  #Rough search returns the same model as exhaustive search  recovered.arima6.1 <- auto.arima(recovered.train[,"India"], seasonal = FALSE)  recovered.arima6.1 #(p,d,q) == (0,2,2) AIC = 7636.97  checkresiduals(recovered.arima6.1) #not adequate  recovered.pred6.1 <- forecast(recovered.arima6.1, nrow(recovered.test))  accuracy(recovered.pred6.1, recovered.test[,"India"])[2,] #RMSE = 2.086x10^5, MAPE = 84.547  ```  Forecasts for Confirmed Cases  ```{r, echo = FALSE}  #Australia  cases.arima1 <- Arima(cases.ts[,"Australia"], order = c(1,1,3))  cases.pred1 <- forecast(cases.arima1, 30)  autoplot(cases.pred1) + autolayer(fitted(cases.pred1), series = "Fitted")  #China  cases.arima2 <- Arima(cases.ts[,"China"], order = c(2,1,1))  cases.pred2 <- forecast(cases.arima2, 30)  autoplot(cases.pred2) + autolayer(fitted(cases.pred2), series = "Fitted")  #Thailand  cases.arima3 <- Arima(cases.ts[,"Thailand"], order = c(2,1,3))  cases.pred3 <- forecast(cases.arima3, 30)  autoplot(cases.pred3) + autolayer(fitted(cases.pred3), series = "Fitted")  #US  cases.arima4 <- Arima(cases.ts[,"US"], order = c(0,1,1))  cases.pred4 <- forecast(cases.arima4, 30)  autoplot(cases.pred4) + autolayer(fitted(cases.pred4), series = "Fitted")  #United Kingdom  cases.arima5 <- Arima(cases.ts[,"United Kingdom"], order = c(3,1,2))  cases.pred5 <- forecast(cases.arima5, 30)  autoplot(cases.pred5) + autolayer(fitted(cases.pred5), series = "Fitted")  #India  cases.arima6 <- Arima(cases.ts[,"India"], order = c(2,1,3))  cases.pred6 <- forecast(cases.arima6, 30)  autoplot(cases.pred6) + autolayer(fitted(cases.pred6), series = "Fitted")  ```  ```{r, echo = FALSE}  #Australia  deaths.arima1 <- Arima(deaths.ts[,"Australia"], order = c(0,1,5))  deaths.pred1 <- forecast(deaths.arima1, 30)  autoplot(deaths.pred1) + autolayer(fitted(deaths.pred1), series = "Fitted")  #China  deaths.arima2 <- Arima(deaths.ts[,"China"], order = c(0,1,1))  deaths.pred2 <- forecast(deaths.arima2, 30)  autoplot(deaths.pred2) + autolayer(fitted(deaths.pred2), series = "Fitted")  #Thailand  deaths.arima3 <- Arima(deaths.ts[,"Thailand"], order = c(2,1,1))  deaths.pred3 <- forecast(deaths.arima3, 30)  autoplot(deaths.pred3) + autolayer(fitted(deaths.pred3), series = "Fitted")  #US  deaths.arima4 <- Arima(deaths.ts[,"US"], order = c(3,1,2))  deaths.pred4 <- forecast(deaths.arima4, 30)  autoplot(deaths.pred4) + autolayer(fitted(deaths.pred4), series = "Fitted")  #United Kingdom  deaths.arima5 <- Arima(deaths.ts[,"United Kingdom"], order = c(1,1,5))  deaths.pred5 <- forecast(deaths.arima5, 30)  autoplot(deaths.pred5) + autolayer(fitted(deaths.pred5), series = "Fitted")  #India  deaths.arima6 <- Arima(deaths.ts[,"India"], order = c(1,1,3))  deaths.pred6 <- forecast(deaths.arima6, 30)  autoplot(deaths.pred6) + autolayer(fitted(deaths.pred6), series = "Fitted") |