Kunal Lalwani, Emma Dennis-Knieriem,
Jeremy Banuelos, Trisha Ballakur

Predicting Stability after COVID-19

How long will it take for the number of travelers from China to the USA to rebound back to pre-outbreak baseline travelers during the COVID-19 outbreak based on SARS outbreak analysis?

The Problem

The threat of a megavirus has been an impending source of fear for the global society. From the Bubonic Plague of the 1350s to diseases like SARS, the idea of mass sickness and death has loomed. Although our medical technology has come along way and is rapidly increasing, epidemics cause global anxiety, which is only exacerbated by the media.

As we find ourselves in the midst of COVID-19, the coronavirus, we believe it is integral to contain fear with the inference of past results through science. By carefully studying a previous epidemic stemming from China, we have produced predictions to aid in comforting people during this time. Large scale fear also has economic implications. Many people hold onto their money instead of spending it when there is uncertainty about the future, which can lead to recessions in the economy. From the perspective of investors, it may keep a lot of business to be able to give clients an estimation of when the market will reach its original levels before the outbreak.

SARS

We began by examining the most recent epidemics that have caused global anxiety, which have been ebola and H1N1. However, ebola originated near the Ebola River in Africa. For a variety of historical and structurally perpetuated reasons such as the wealth gap, the US has received far fewer travelers from African countries affected by the ebola virus than from China. H1N1, or swine flu, broke out in the US, rendering travelling data from the US to the US meaningless. We instead chose to compare the modern coronavirus to the SARS outbreak in China in 2003. Although this happened close to two decades ago, making comparisons about travelers embarking from the same country helped us create a more accurate model than if we had used data from a country other than China, no matter how recent the outbreak.

Our goal was to create a model to predict how long it will take after the coronavirus' outbreak for travel numbers from China to the US to reach the same levels again, based on comparable data from during the SARS epidemic in China.

Applying Insight from SARS to COVID-19

Because the SARS outbreak was so similar in nature to the current outbreak of the coronavirus, we are using tourism data from around the time of the SARS outbreak to help with our predictability model of the effects of the coronavirus on tourism from China to the United States. We gleaned tourism data from the NTTO datasets from 2001 until 2005, specifically focusing on the number of Chinese citizens entering the US during each month in this period. This data, the example of an epidemic that rebounded, will be used to train our predictive model.

To forecast the potential effects of the coronavirus on tourism, we utilized data from “Final COR Volume.xlsx” provided by Fidelity Investments, and fed this into our predictive model to find a prediction for when the coronavirus will rebound.

To get our data into the most easily read forms, we compiled select data into Google spreadsheets from “Final COR Volume.xlsx” as well as data we found from during the SARS outbreak. We then downloaded these sheets into Excel files that could easily be read into R.

These showed us that there was a rebound in the number of travelers from China to the US to pre-outbreak levels after the SARS outbreak hit.

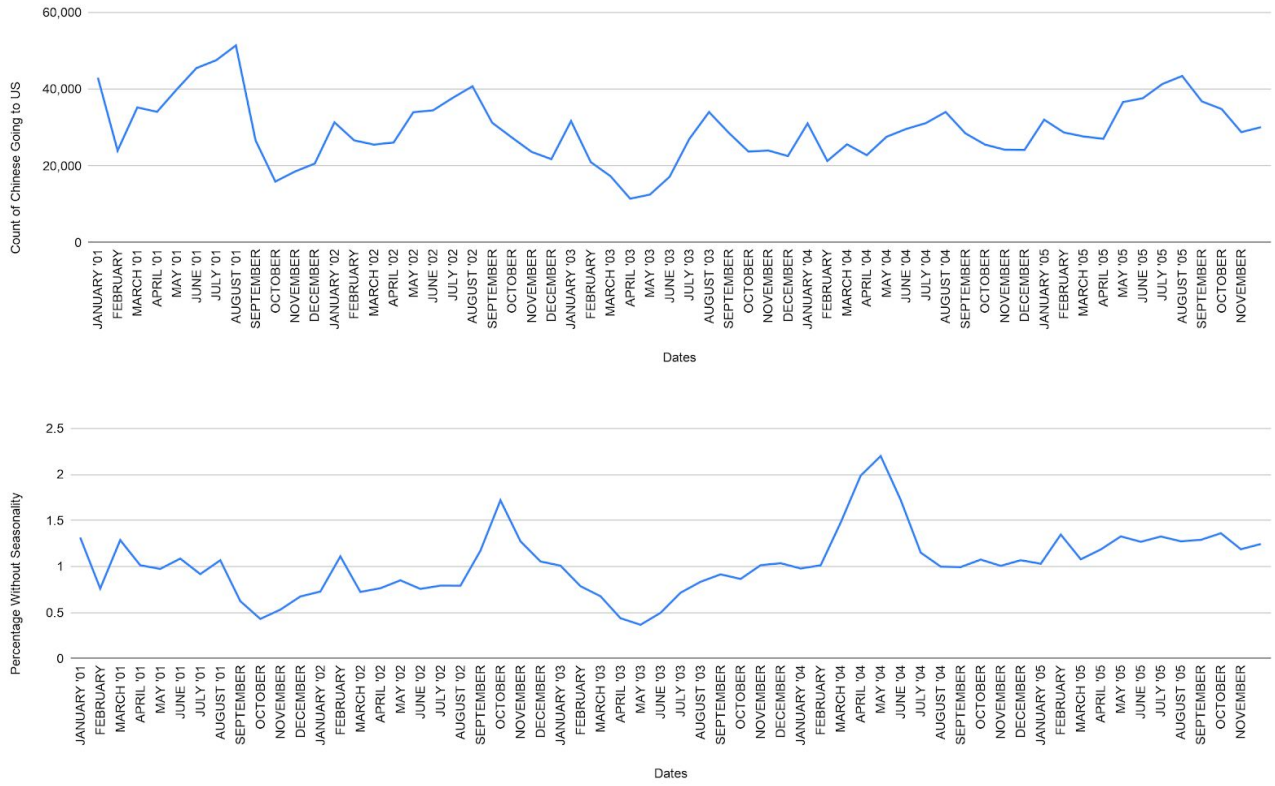

Fig 1.0. One major aspect we had to control for in our data was seasonality. There are trends within each year, such as lower levels of travel from China to the US during winter months. In order to find trends across years, we had to find a baseline of what was to be expected throughout the seasons without condensing all data within a year into one aggregated number.

Below is the practices we followed:

Method

We built a forecasting model to predict the trend in the current number of Chinese visitors coming to the USA. The dataset on the change of chinese visitors entering into the USA before, during and after the SARS outbreak were fed to the ARIMA (Auto Regressive Integrated Moving Average) model for a prediction. We removed the trend and seasonality using the inbuilt acf(), pacf() and also box plots to build a model. The model was validated and the efficiency was tested using the serial correlation test- Ljung-Box test upto lag=5. The outcome was p~0.2, showing no serial correlation.

We used this model as a baseline to fit our new time series data related to the coronavirus and predicted the time it would take to for the ratio of Chinese travelers to the USA to return to pre-outbreak levels. From our model, we observed that the number of travelers will return to pre-outbreak levels from July 2020 onwards. Below are visualizations that helped lead us to this conclusion.

Conclusion

We created this prediction model with the intention that Fidelity investors feel more secure or certain of the timeline of flux during the coronavirus. Further, our model can be used to aid in increasing the security of investments in certain sectors. With more time, we would like to explore how a change in tourism affects different market sectors in the U.S. We would also have liked to generalize this model to be inclusive of economies affected not only by epidemics in China, but worldwide. We would need sufficient data to control for factors such as inflation, major world events (such as 9/11 or the Olympics), and historically or environmentally devastated regions.

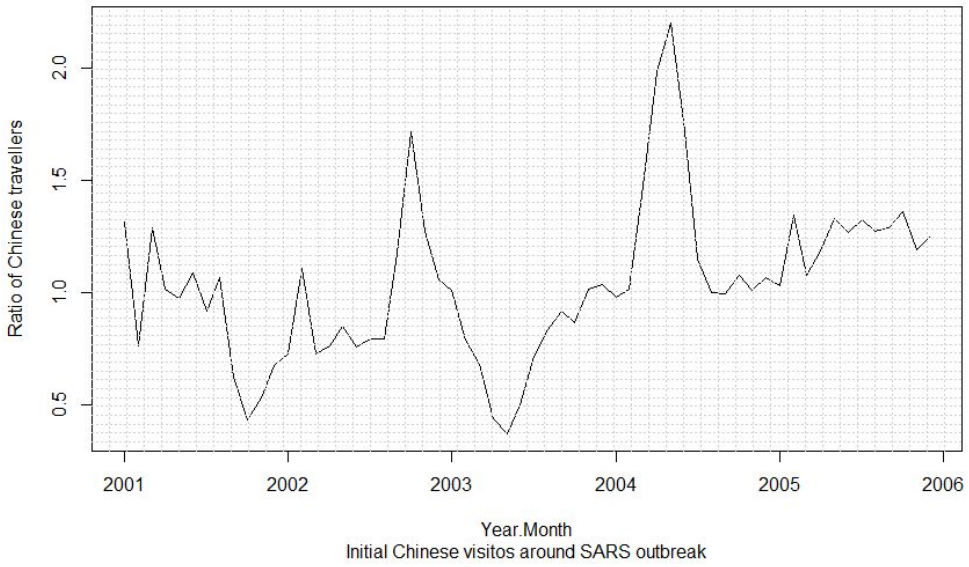

Fig 1.1 The continuation of Fig 1.0. On the y-axis is the ratio of travelers from China to the US over all travelers to the US per month each year. Before March 2003, the drastic drop in the middle of the plot, the average ratio of Chinese travelers was slightly less than 1.0. After the SARS outbreak (March 2003), the ratio stabilized back to an average of above 1.0.

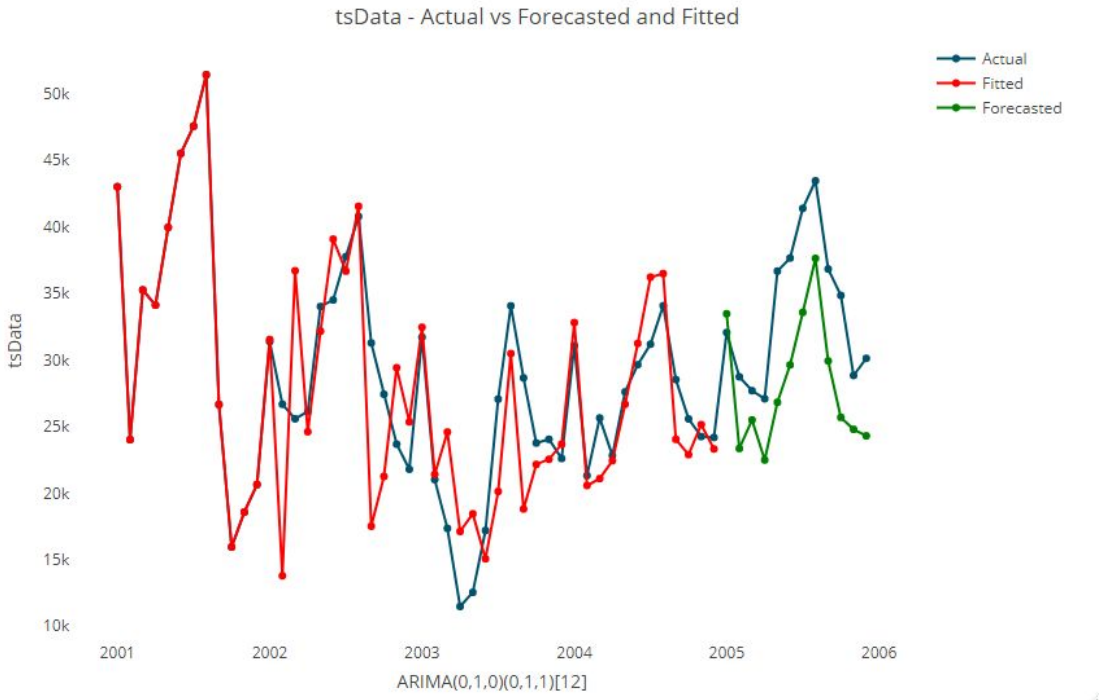

Fig 2. The model curve fit over the original data in the red shows- an approximate measure and forecasted for the next 24 months. The model will be used to predict the situation for nConV outbreak.

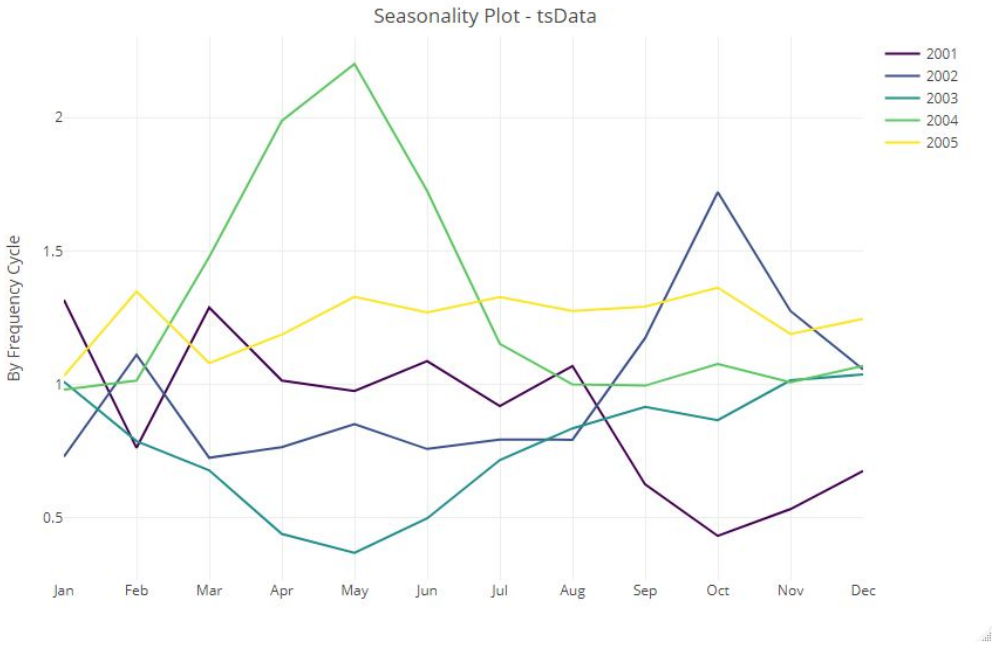

Fig 3. This plot overlays the year for 2001 through 2005 inclusively. The seasonality trend of year wise was analysed using the graph to remove the moving average.

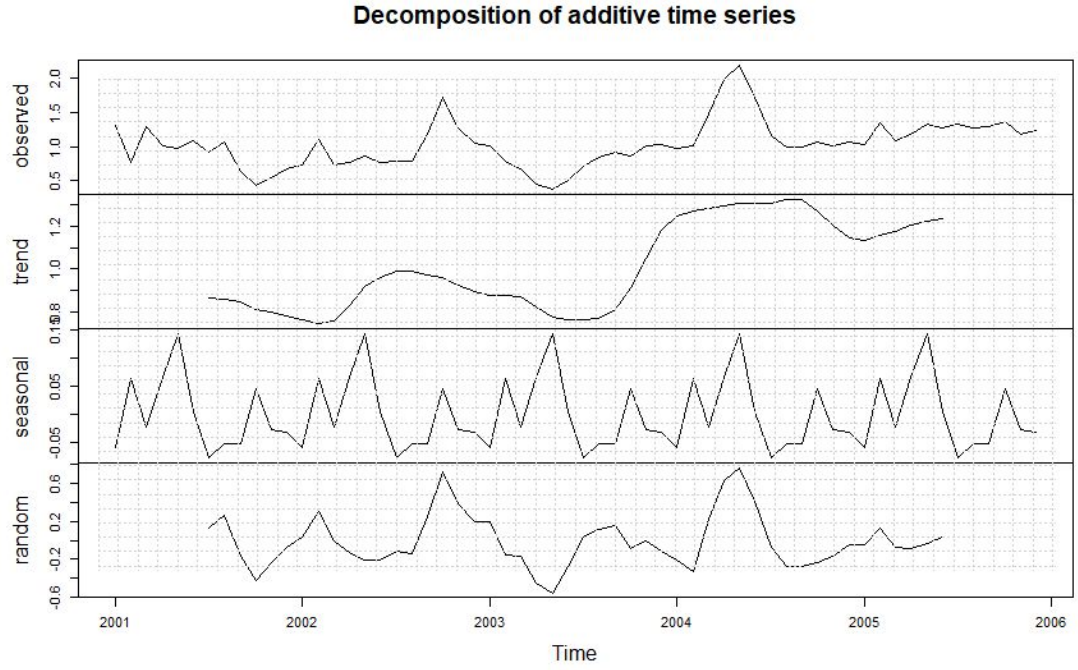

Fig 4. The graph shows a linear trend and seasonality component along with a polynomial function in the end.

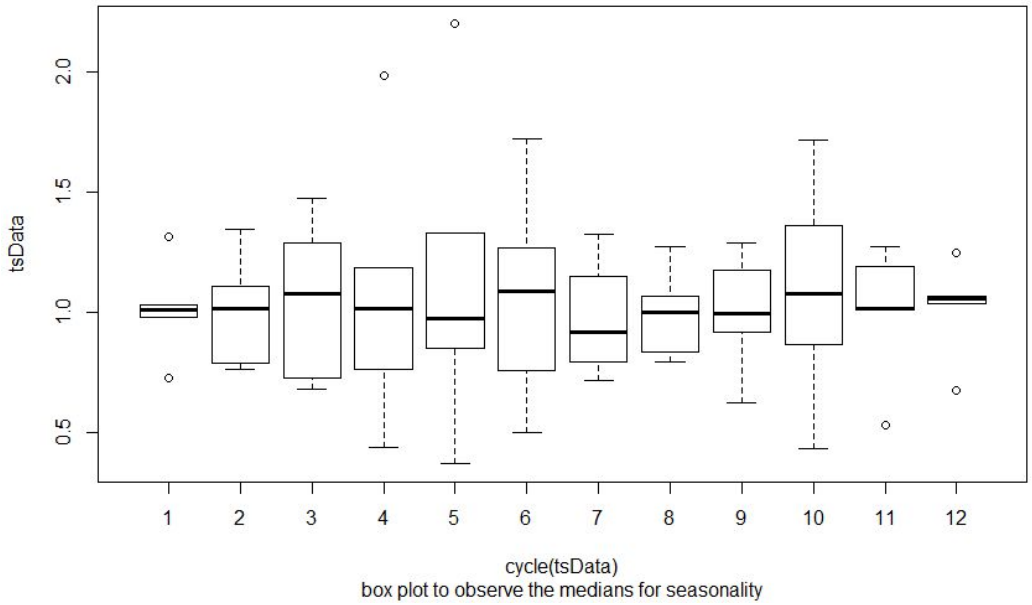

Fig 5. The box plot shows the moving median component to be removed for time series analysis.

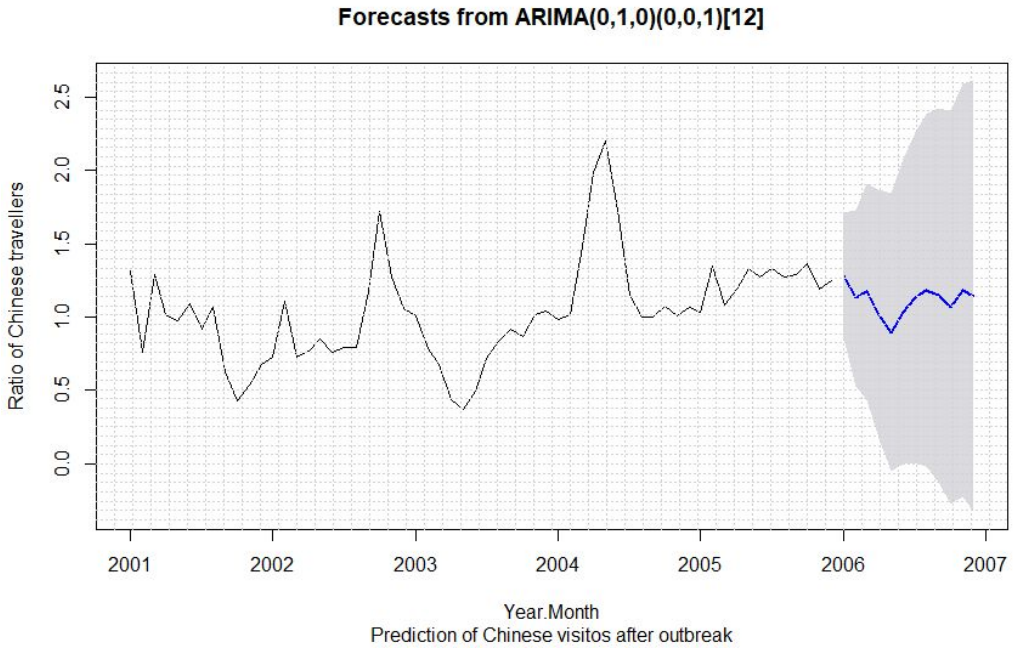

Fig 6. Applying our model to the SARS data, we predicted the ratio of Chinese travelers past the data we had. As expected, our model predicts a stable continuation of the post-outbreak level (above 1.0).

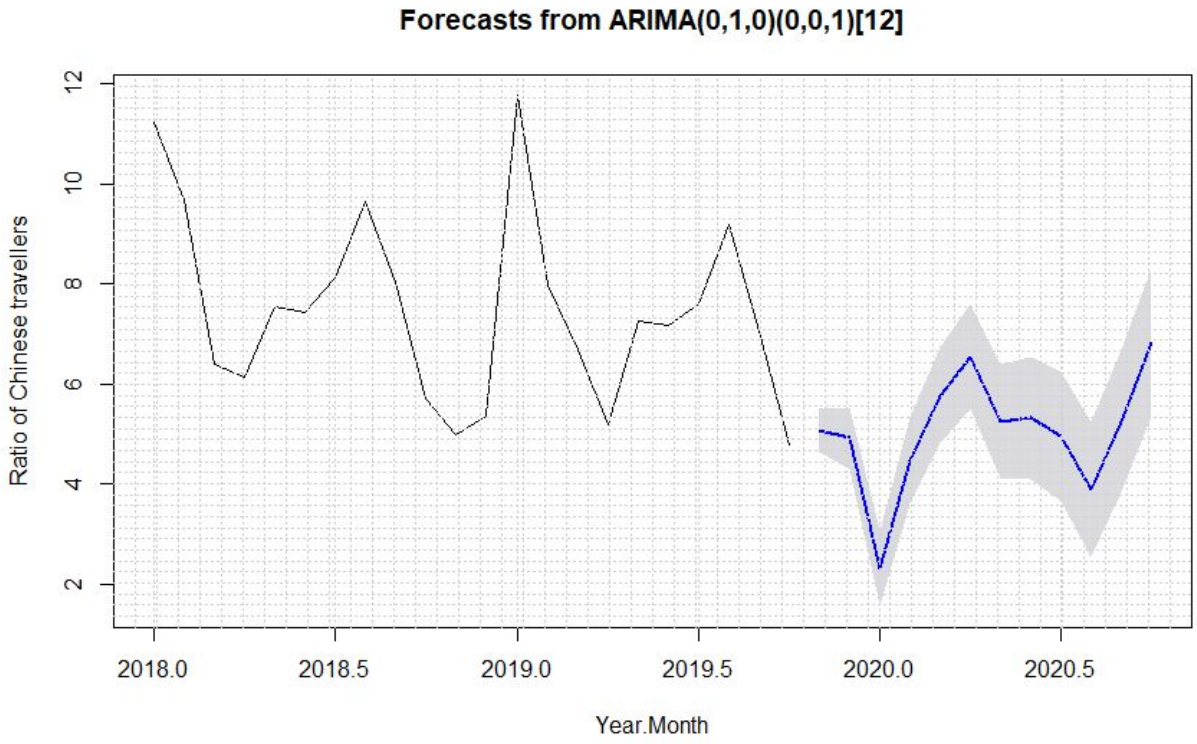

Fig 7. The grey ribbon around the blue forecast is narrower when looking close into the future and wider when looking farther into the future, because there is more uncertainty. They represent the upper and the lower bounding uncertainties. This predictability model uses data from the rebound time of tourism from China during the SARS outbreak. We expect tourism from China to drop drastically during the first couple of months of the coronavirus and for tourism counts to begin to rebound in the middle of 2020.

Assumptions

- No future travel bans from China given current political climate, which could greatly affect future tourism from China along with the coronavirus
- The coronavirus does not turn into a pandemic (controlled by about the same time SARS was)
- United States' tourism was still rebounding from the 9/11 attacks before the SARS outbreak where tourism in general decreased by 50%
- Uncertainties from other countries and the global political climate will not have an effect on tourism from China to the United States

The Team

Before Saturday morning, only 2 out of 4 had met before. Out of our four team members, two are sophomores at Brown, one is a junior at Brown, and our final member is a master's student

at UCSD. Moreover, all of us are studying different fields: data science, machine learning, applied math, graphics, cyber security, and political science. As such, we all applied a plethora of skills to come up with our research question and create the visualizations and model.

Acknowledgements

Eric[1], the Fidelity representative, was a fantastic resource for us to bounce our ideas off of and consult. Specifically, he encouraged us to simplify our project and gave us insight on complexities we might encounter. Further, he even stayed up around 10 pm via Slack to help us find a flaw in the way we were graphing our data. Because of his advice, we spent over five hours solidifying our problem and do research before starting to vet that it was something of use and feasible.

[1] Mystery Solved!