Effectiveness of Non-pharmaceutical Interventions and Sentiment Regarding COVID-19

A closer look at 'stay at home' policies in the U.S. and Twitter sentiment of topics relating to COVID-19 by Aashish Ghimire & Chase Mortensen. Report by Chase Mortensen.

INTRODUCTION

COVID-19, first recognized at the end of 2019, has spread throughout the world and to every state in the US. The purpose of this report is to address several subtasks listed in the Kaggle COVID-19 Open Research Dataset Challenge (CORD-19). In our proposal we hoped to address the following subtasks: "Rapid assessment of the likely efficacy of school closures, travel bans, bans on mass gatherings of various sizes, and other social distancing approaches," "Research on why people fail to comply with public health advice, even if they want to do so (e.g., social or financial costs may be too high)," and "Efforts to identify the underlying drivers of fear, anxiety, and stigma that fuel misinformation and rumor, particularly through social media."

In order to address the first subtask regarding the effectiveness of policies, we analyzed cases and deaths caused by COVID-19 in numerous states. We compared states with 'stay at home' orders with those without such policies. We were unable to complete all of the analyses which we hoped to accomplish for the second task involving failures to comply with health advice but were able to investigate right-wing and left-wing media around COVID-19. The third task, which involved social media, was addressed by a sentiment analysis from a novel Twitter dataset assembled using the tweepy package.

TASK 1: EFFECTIVENESS OF POLICIES

Dataset

Data for task 1 ultimately was acquired from the New York Times. We attempted to use several other datasets, but the NYTimes dataset was the only one we found that had current data. However, there was a tradeoff with the data, since it didn't include information on hospitalization - only cases and deaths in the U.S.

Analysis Techniques

Initially, we hoped to plot case counts and deaths in linear and log scale to identify major changes in individual states' responses to COVID-19 and then identify policies adopted by the state which could have led to the change. For example, Idaho issued a statewide 'stay at home' order on 3/25, which seemed to have an effect in the following weeks of slowing the growth of cases and deaths in the state (Fig. 1).

However, states like Idaho were fairly uncommon; in most cases, the curve didn't take such a dramatic turn and instead slowly shifted over longer periods of time. Additionally, analyzing policies was complicated by the sheer number of policies - sometimes in the 100s -

enacted by the states in addition to more local measures. Ultimately, we decided to compare states with no ‘stay at home’ order with those with ‘stay at home’ orders. Additionally, there was a third group of states which either had regional orders or advisories. We compared multiple states in each category with the national trend (Fig. 2) by adjusting the national trend to be proportional to each state’s population. This allowed us to analyze whether the cases and deaths in each state were above or below the expected numbers and adjusted for actions taken on a national level, such as the national emergency declaration on 3/13. This technique has drawbacks, such as more rural states having their curves delayed, but is still useful in comparing states.

Selected Results

There were four states with no ‘stay at home’ order: Arkansas, Iowa, Nebraska, and North Dakota. These rural states were generally well under the predicted curve (Fig.3-6). Nebraska was the most concerning state. The number of cases seen in logarithmic scale is nearly linear (Fig. 7), which indicates exponential growth. If this trend continues, it seems that a ‘stay at home’ order is inevitable.

States with regional orders or advisories include Kentucky, Massachusetts, Oklahoma, South Dakota, Utah, and Wyoming (Fig. 8-13). For the sake of brevity, I will only discuss Utah and Massachusetts. Utah (Fig. 12) is near the project number of cases, but significantly below the projected number of deaths (Wyoming also follows this pattern). There are several possible explanations for this. Utah has fairly high testing compared to the rest of the country, so with more testing and more cases, the death rate decreases. Additionally, Utah is a relatively young, healthy state, which could lower the death rate. Massachusetts, on the other hand, is older than Utah and has a higher population. It is surprising that Massachusetts doesn’t have a ‘stay at home’ order since the number of cases and the number of deaths have passed the national trend.

All other states not mentioned in the other two categories have statewide ‘stay at home’ orders in place. Some states, like Colorado (Fig. 14) and Nevada (Fig. 15), follow the national trendline fairly closely. Some, like New York (Fig. 16), have been hit particularly hard and are over the national trend. These states tend to be the more populous hubs with major airports and dense urban areas. However, Washington (Fig. 17) and California (Fig. 18) which both started above the national trend have managed to ‘flatten their curves’ and are now below the US trendlines, which indicates that the ‘stay at home’ policies seem to be having a positive effect, although there are many factors, including under-testing, which could be affecting the results.

TASK 2: Challenges in Policy Implementation

Dataset

We had hoped to investigate hospitalization rates in the U.S. versus hospitalization rates in countries with universal healthcare. However, we ended up focusing our time and energy on other parts of this report. However, we were able to investigate left-wing and right-wing media sentiments on the recent COVID-19 lockdown protests. Our dataset consisted of 40 news articles; 10 each from CNN, The New York Times, The Wall Street Journal, and Fox News.

Analysis Techniques

We used natural language processing using NLTK and TextBlob libraries to analyze sentiment and compared frequent words to see if left-leaning news sources had different sentiments from right-leaning news sources.

Selected Results

Our analysis revealed that in the selected articles, The New York Times had a higher positive sentiment, followed by CNN, then The Wall Street Journal, and then Fox News (Fig. 19). Additionally, by analyzing the intersection of word frequencies, we found that left-leaning (CNN and The New York Times) and right-leaning (The Wall Street Journal and Fox News) were more similar intra-group than any right-left combination. Right/Left assignment was taken from allsides.com.

TASK 3: Twitter Sentiment

Dataset

For the third task, we assembled our own novel dataset using the tweepy library to collect around 4.3 million tweets relating to a number of topics around COVID-19. These topics included Trump, Biden, CNN, FOX, Fauci, the economy, Republicans, and Democrats.

Analysis Techniques

We used common data science analysis techniques to acquaint ourselves with the data, such as histograms and line plots. We also used the natural language processing libraries NLTK and TextBlob to analyze the sentiments of the tweets.

Selected Results

Relating to COVID-19, we found that Trump was mentioned most often, followed distantly by the economy and CNN (Fig. 20). This is also seen in the line plots of tweets by categories (Fig. 21-22) which show a spike in tweets relating to COVID-19 and Trump around 8:00 PM MST, which could be a result of a white house press conference which occurred that evening. Interestingly, there was a spike for tweets relating to Republicans around 5:00 AM MST. Trump was the dominant (hourly) topic associated with COVID-19 throughout the entire day in which tweets were collected.

We found that sentiment was most positive for Republicans followed by Democrats using both NLTK and TextBlob (Fig. 23-24). However, TextBlob found that Trump had the most negative sentiment while NLTK found that Dr. Fauci had the most negative sentiment (in tweets concerning COVID-19).

Finally, for tweets covering both COVID-19 and the POTUS, we found that polarity increased in variation as subjectivity increased (Fig. 25). The same was true for other sub-topics as well.

GitHub Repository: https://github.com/GitAashishG/DataScienceFinalProject

FIGURES

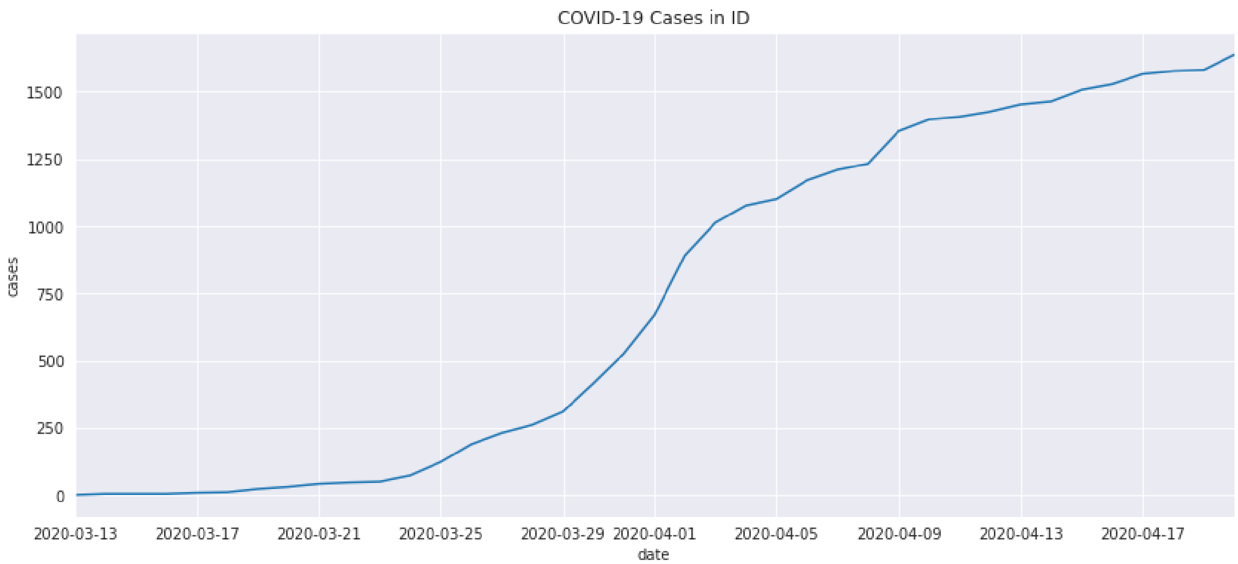

Figure 1. Linear scale of COVID-19 cases in Idaho.

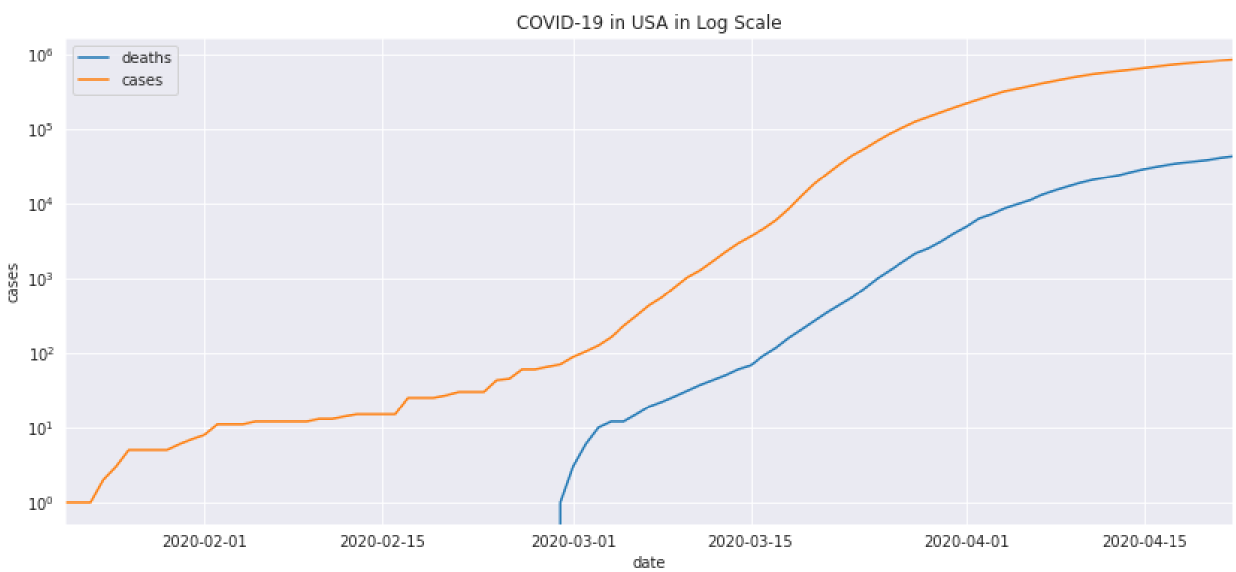

Figure 2. US COVID-19 cases and deaths in log scale.

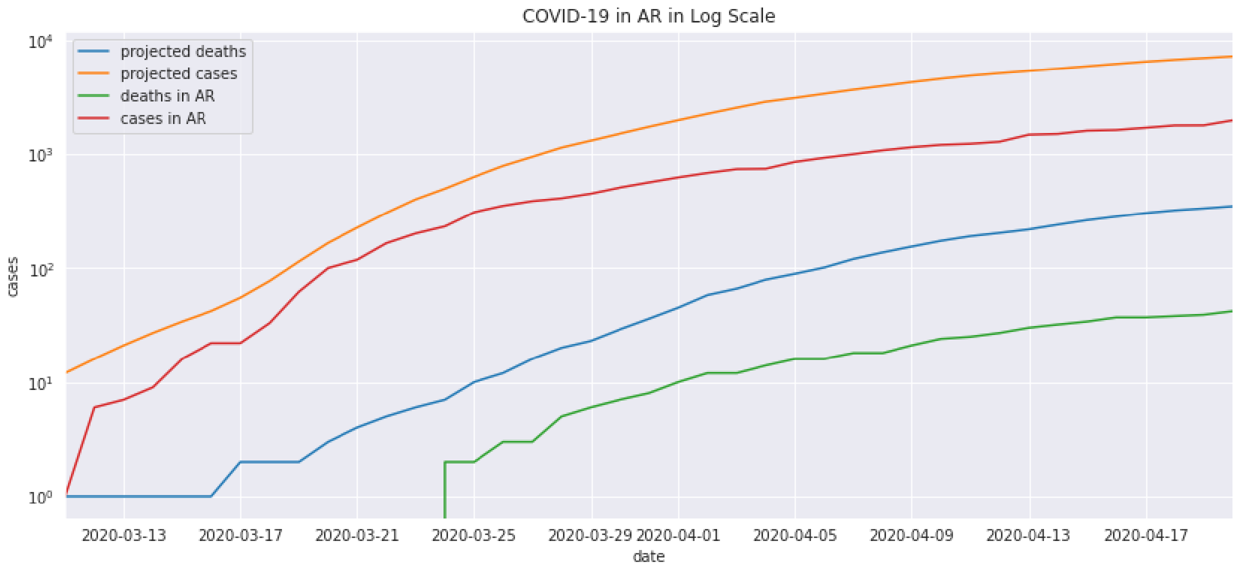

Figure 3. COVID-19 cases and deaths in Arkansas compared with US trends.

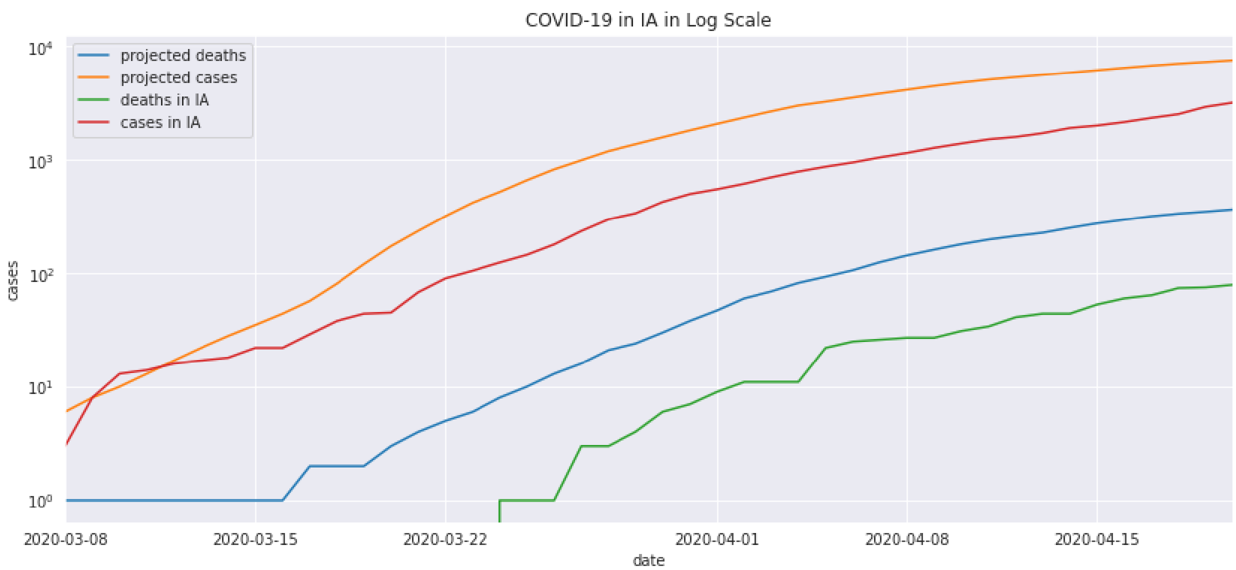

Figure 4. COVID-19 cases and deaths in Iowa compared with US trends.

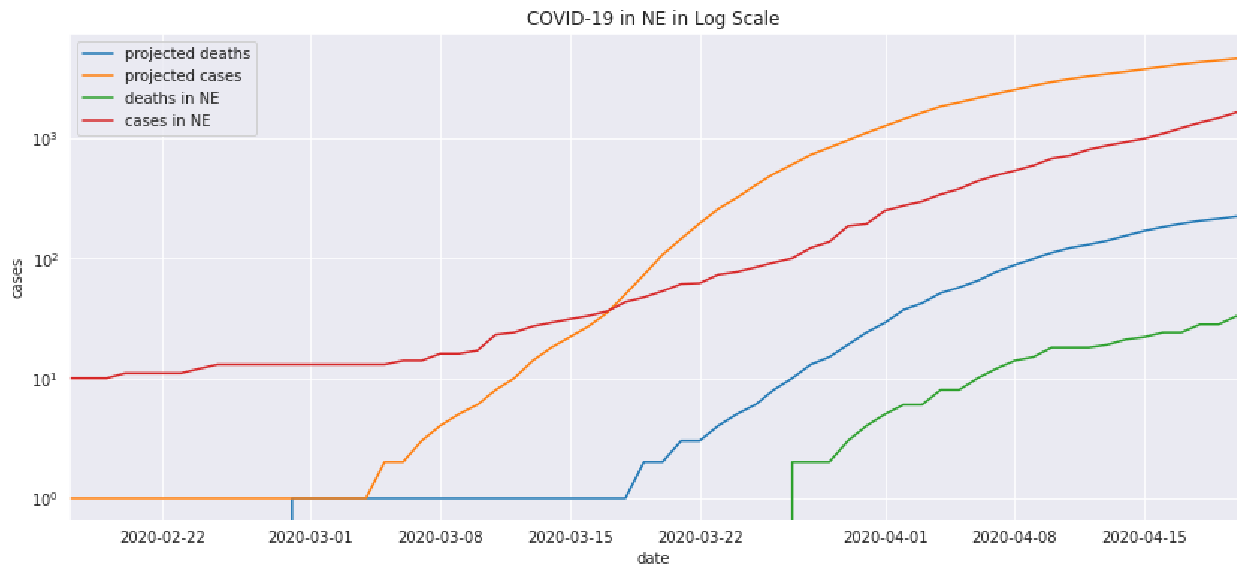

Figure 5. COVID-19 cases and deaths in Nebraska compared with US trends.

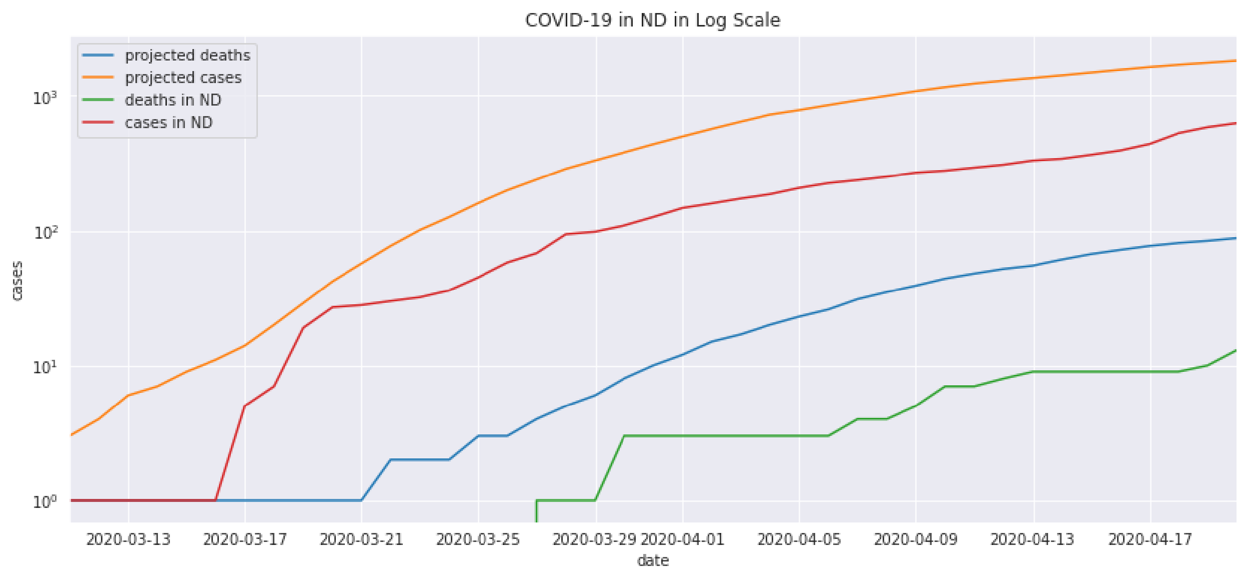

Figure 6. COVID-19 cases and deaths in North Dakota compared with US trends.

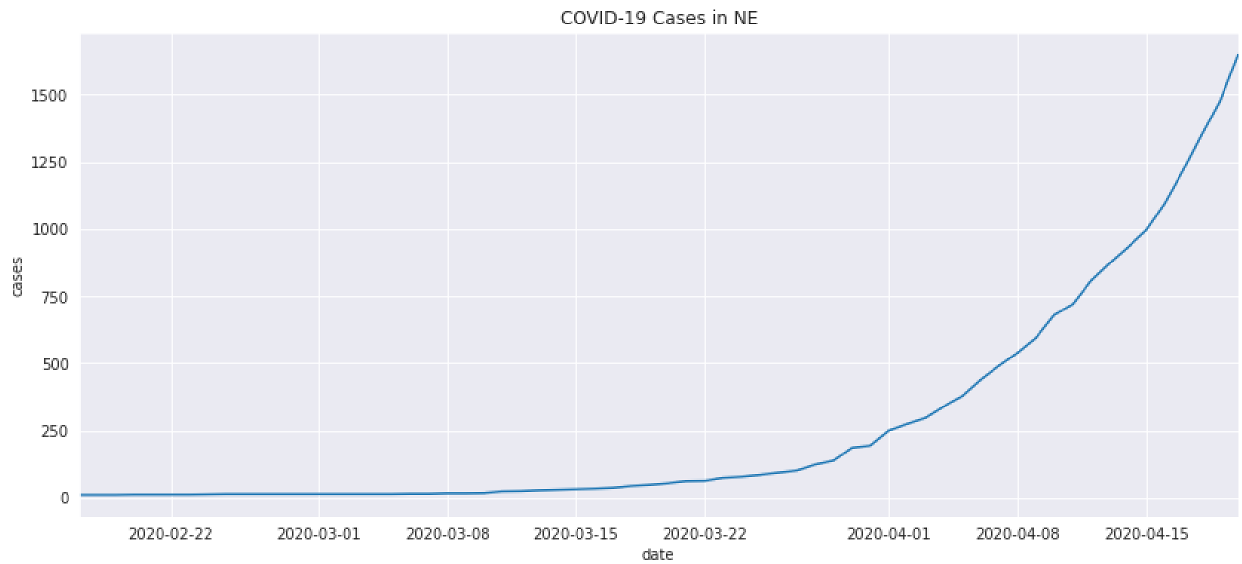

Figure 7. COVID-19 cases in Nebraska in linear scale.

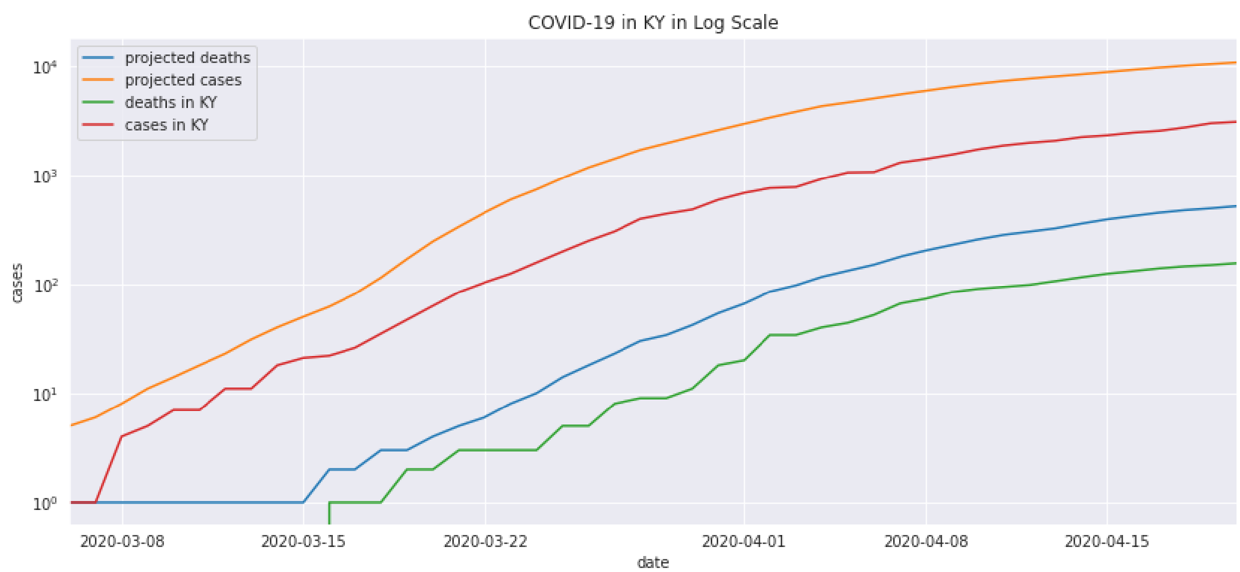

Figure 8. COVID-19 cases and deaths in Kentucky compared with US trends.

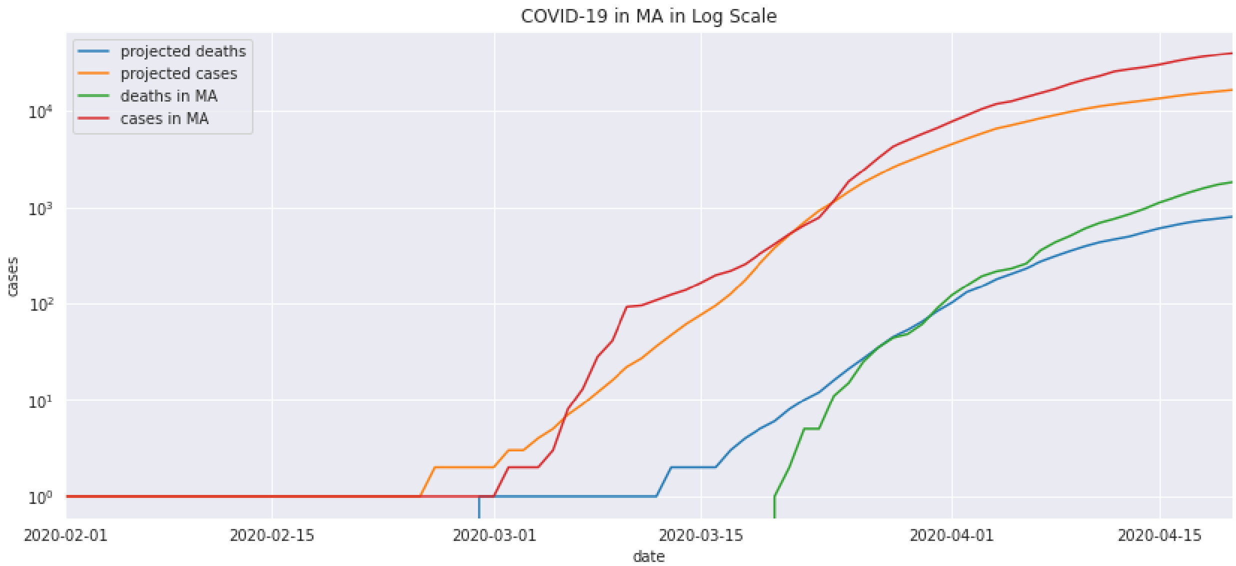

Figure 9. COVID-19 cases and deaths in Massachusetts compared with US trends.

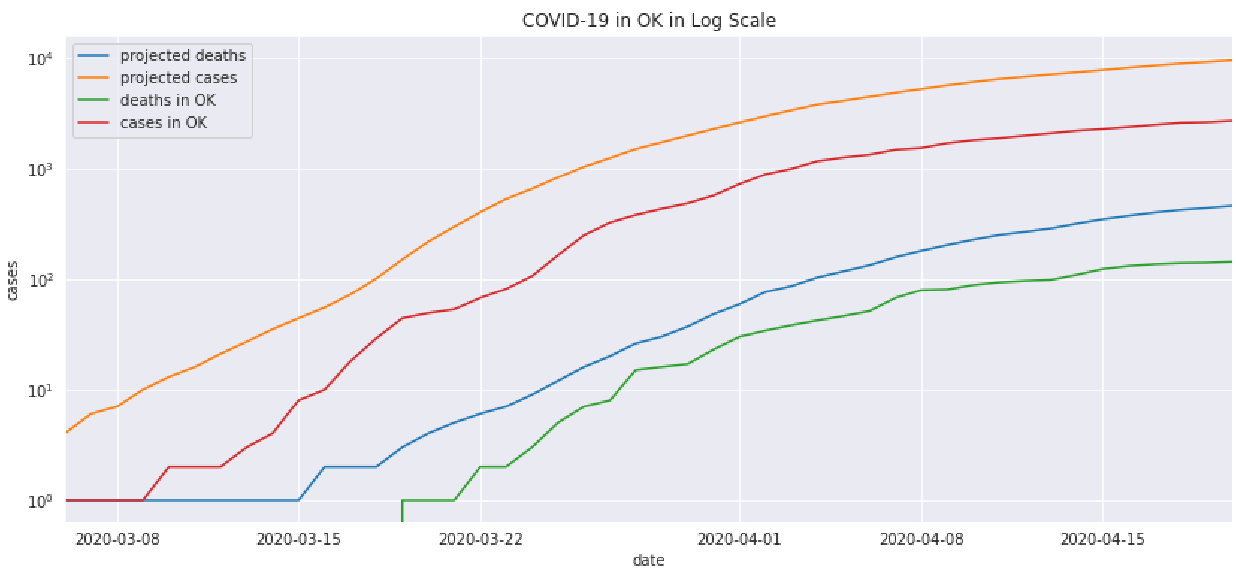

Figure 10. COVID-19 cases and deaths in Oklahoma compared with US trends.

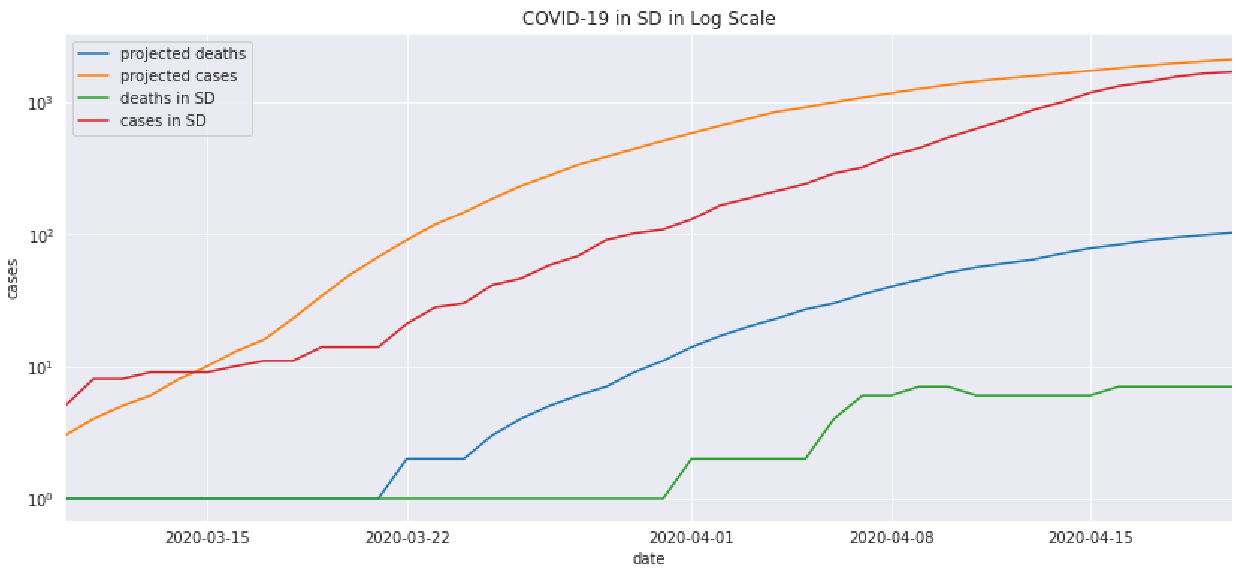

Figure 11. COVID-19 cases and deaths in South Dakota compared with US trends.

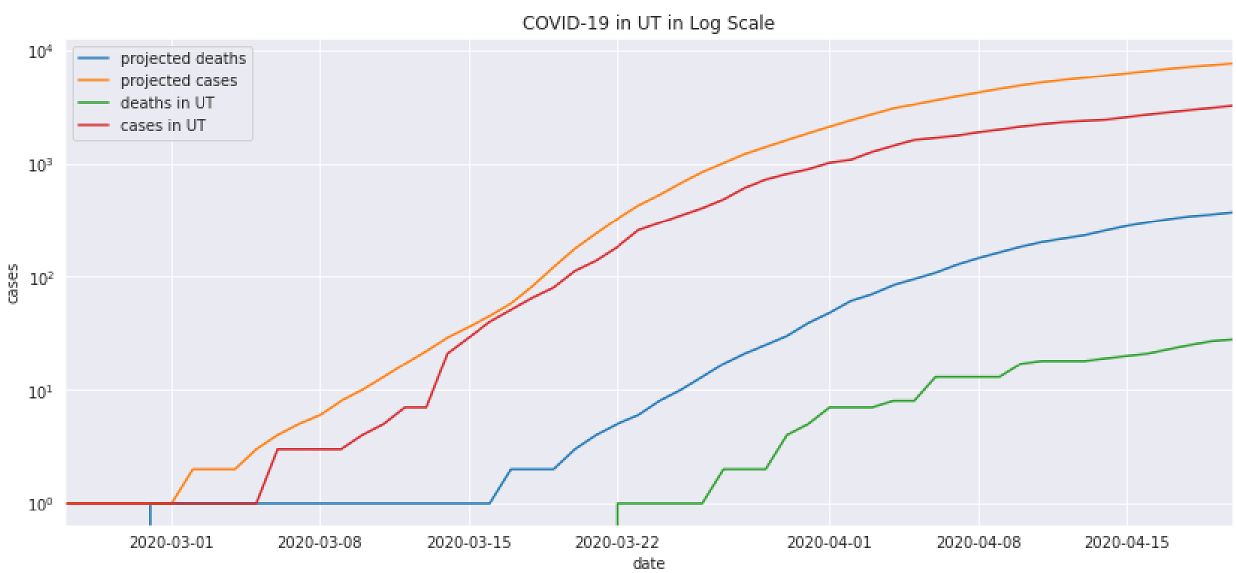

Figure 12. COVID-19 cases and deaths in Utah compared with US trends.

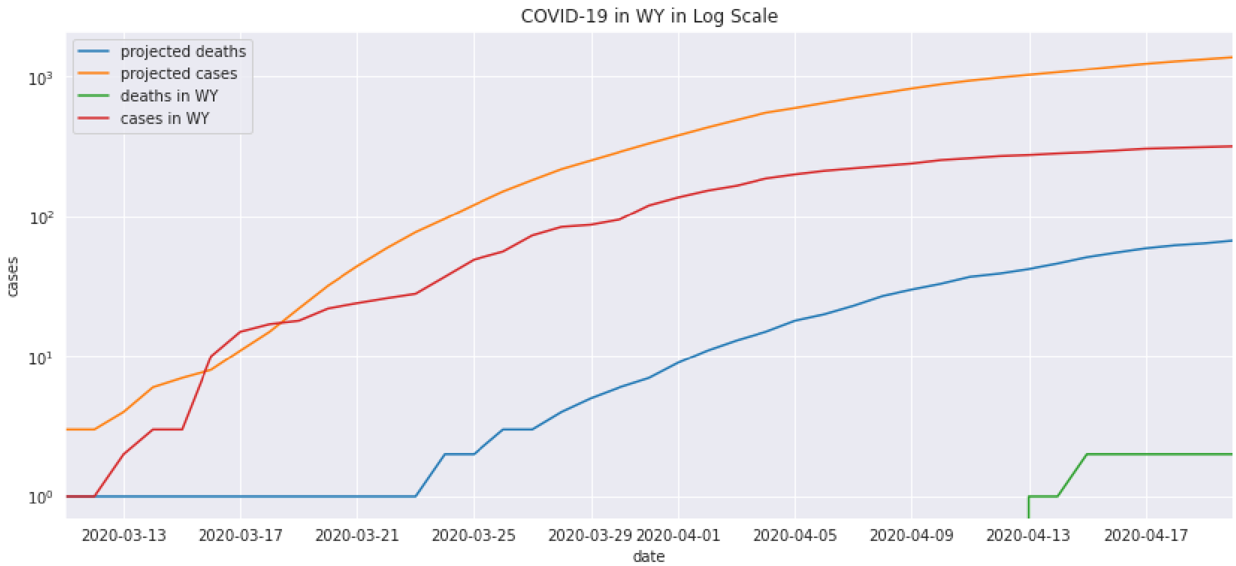

Figure 13. COVID-19 cases and deaths in Wyoming compared with US trends.

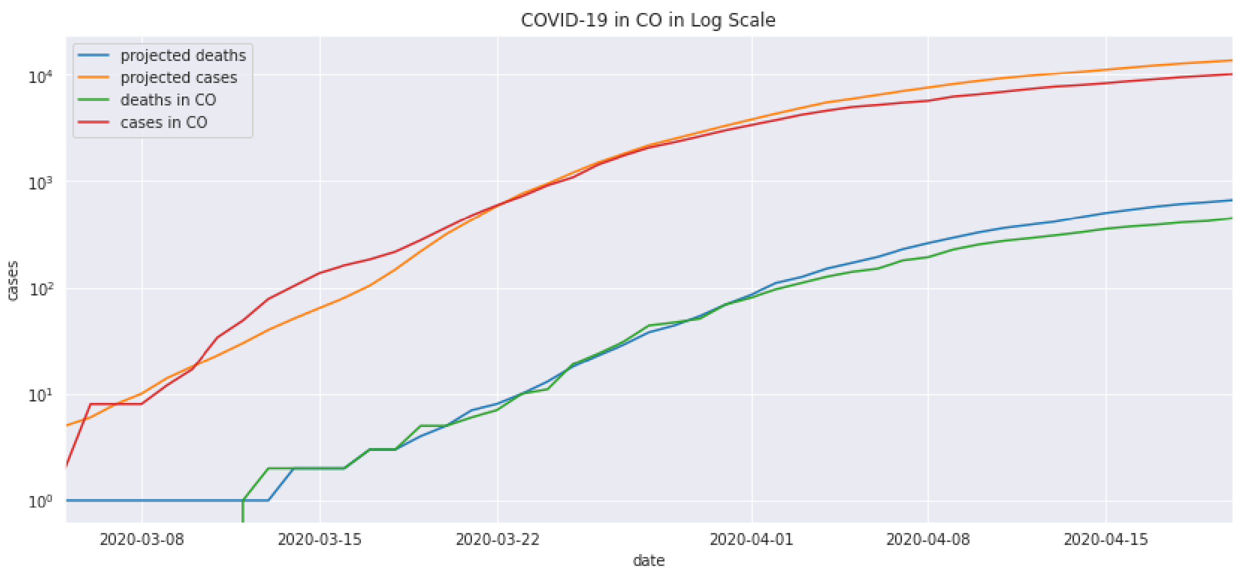

Figure 14. COVID-19 cases and deaths in Colorado compared with US trends.

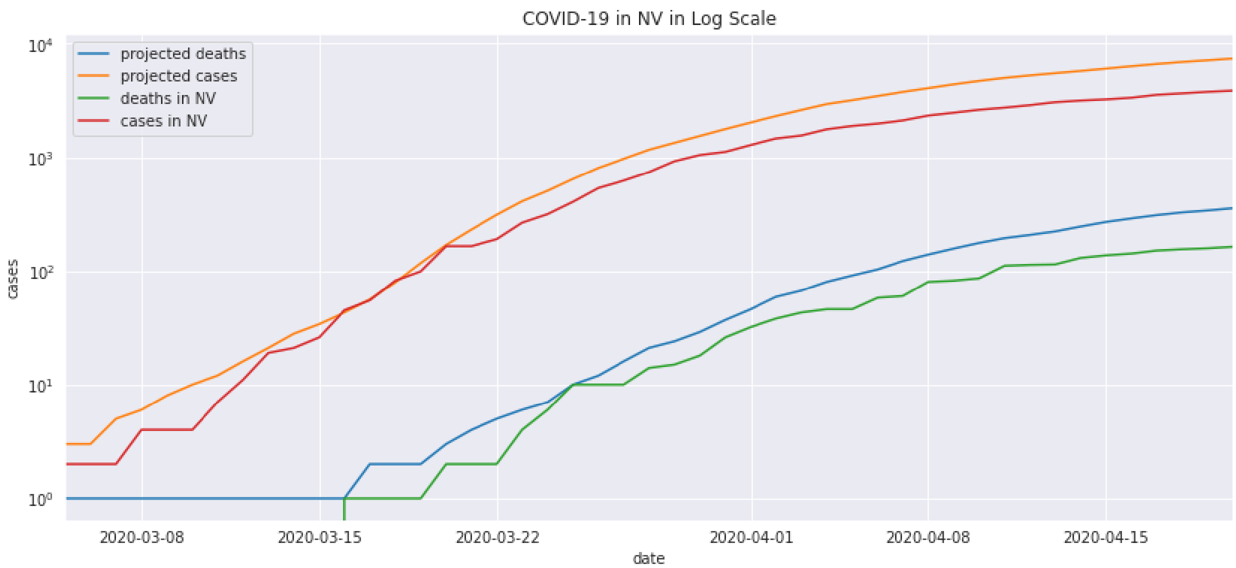

Figure 15. COVID-19 cases and deaths in Nevada compared with US trends.

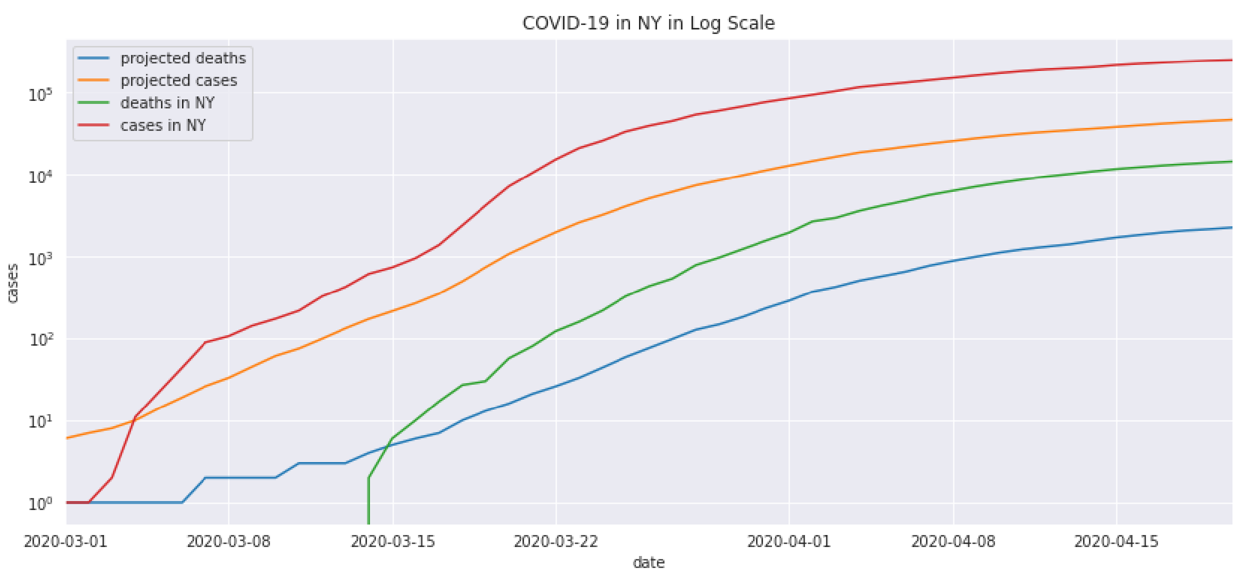

Figure 16. COVID-19 cases and deaths in New York compared with US trends.

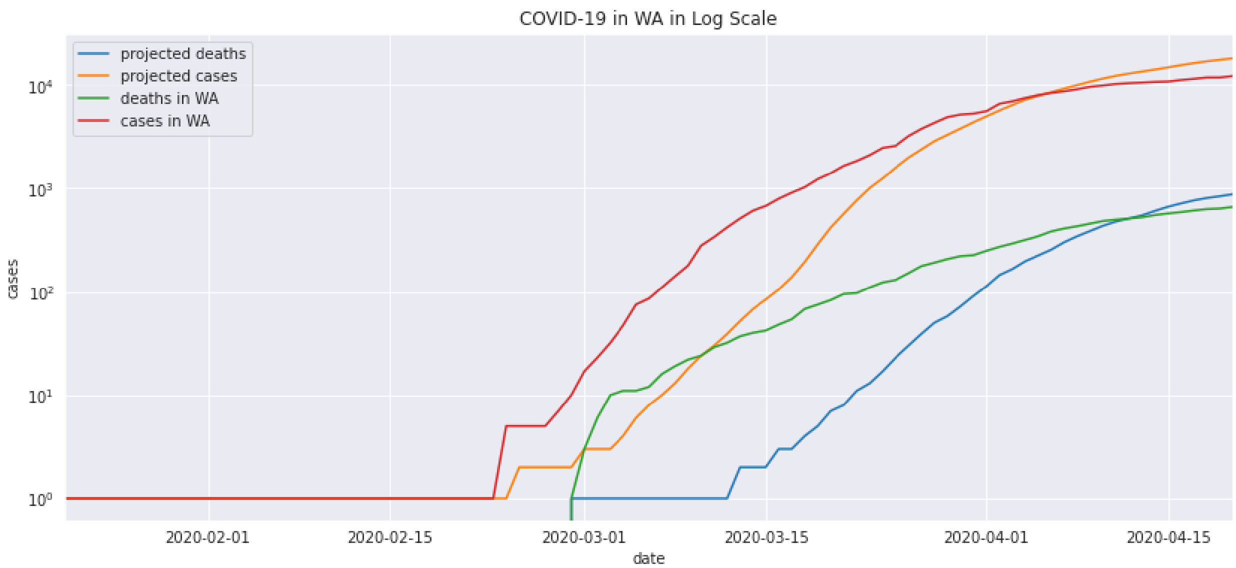

Figure 17. COVID-19 cases and deaths in Washington compared with US trends.

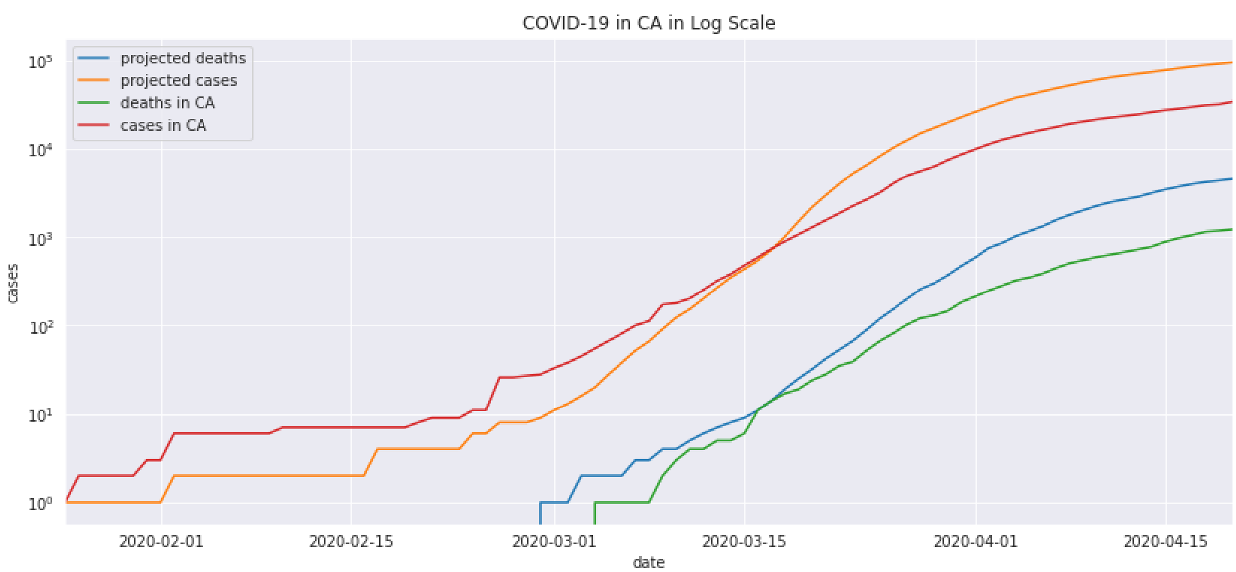

Figure 18. COVID-19 cases and deaths in California compared with US trends.

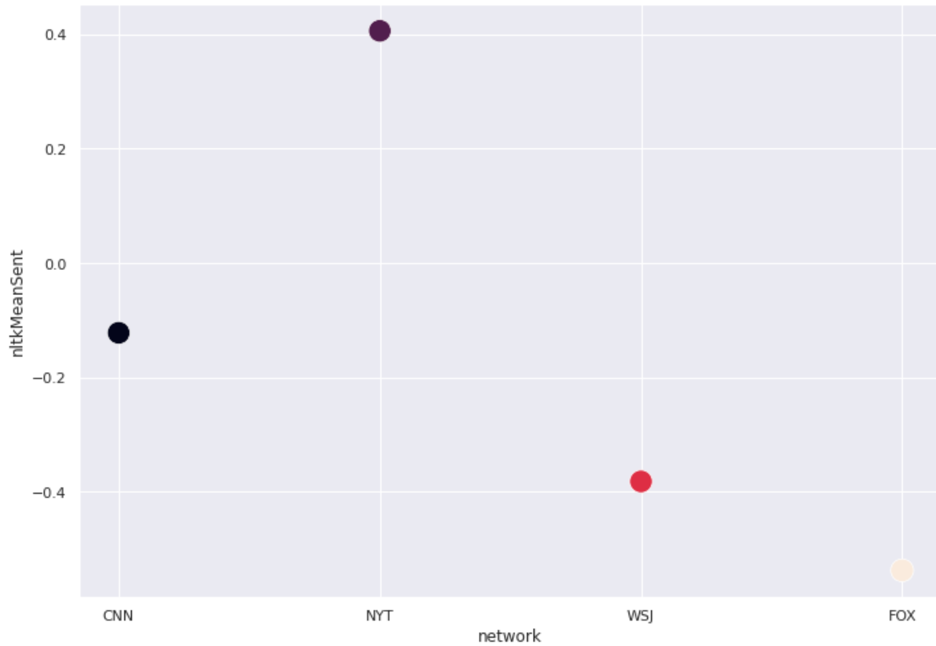

Figure 19. Mean sentiment in articles addressing COVID-19 shut down protests among popular left-leaning and right-wing news sources.

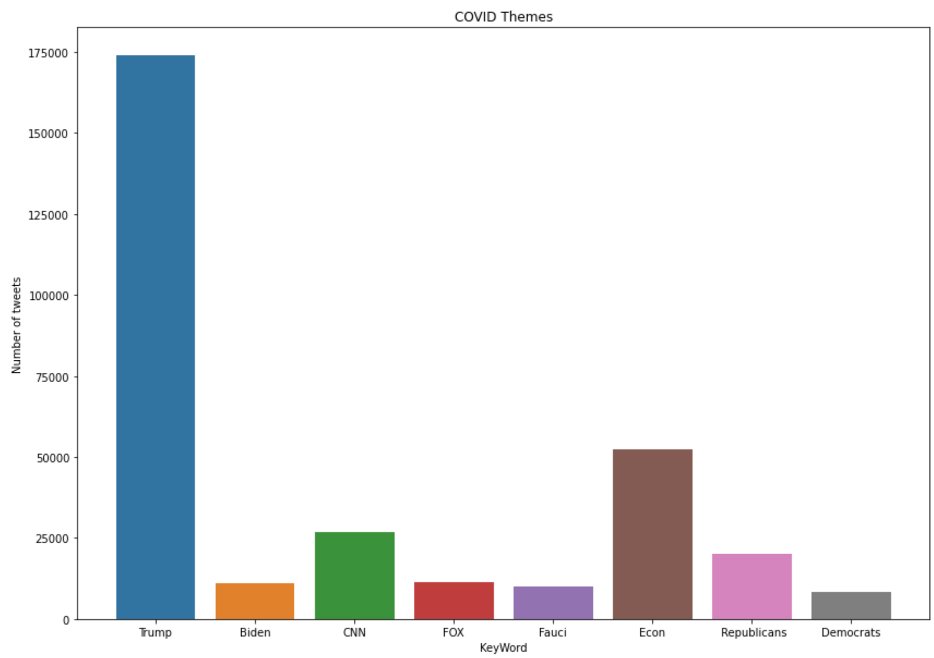

Figure 20. Counts of tweets in the Twitter sentiment dataset relating to each topic.

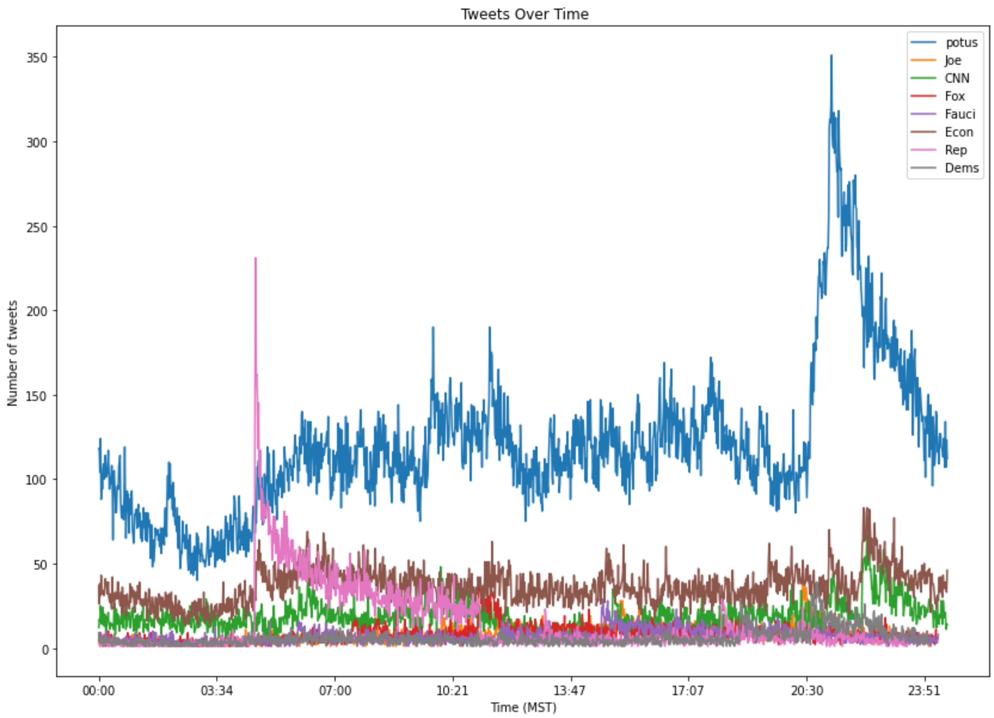

Figure 21. Line plot of tweets relating to sub-topics in minutes.

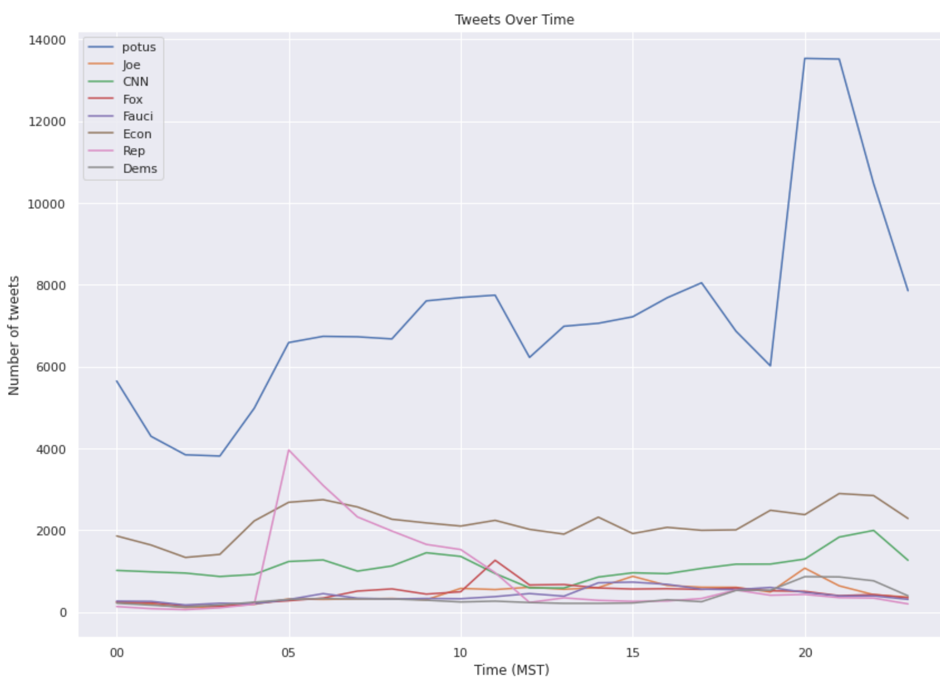

Figure 22. Hourly line plot of tweets relating to each COVID-19 sub-topic.

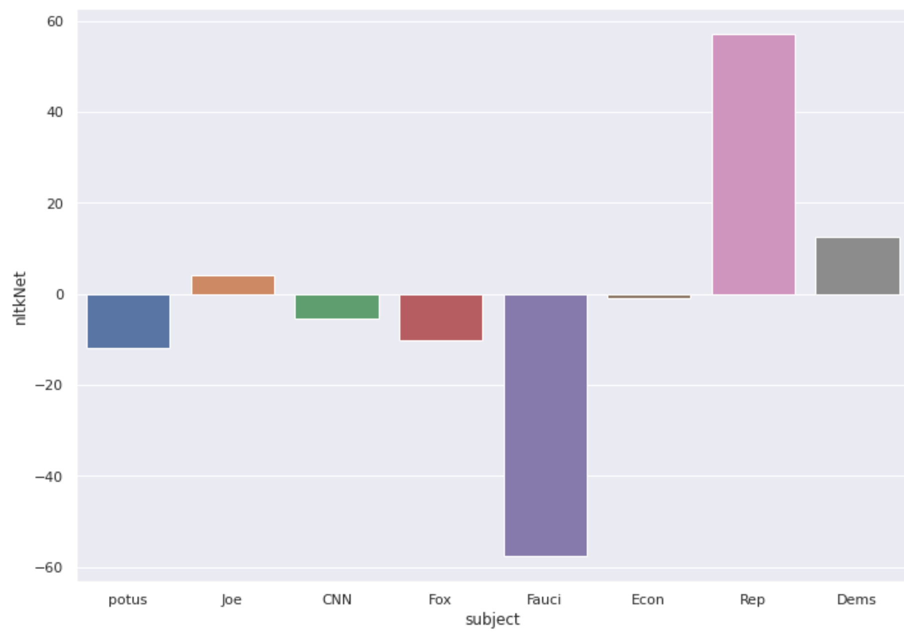

Figure 23. Sentiments of tweets concerning COVID-19 and each sub-topic according to the NLTK library.

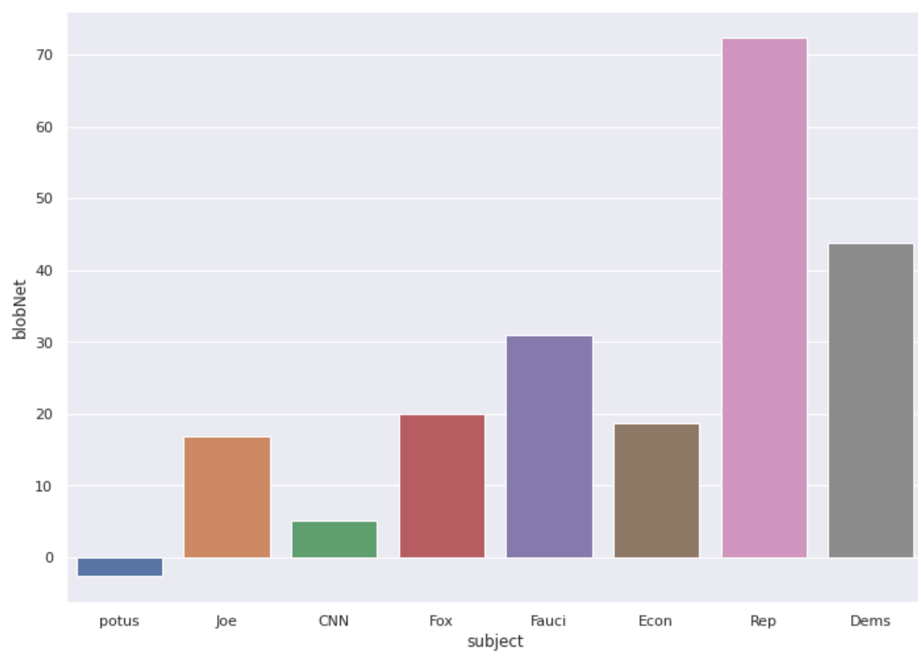

Figure 24. Sentiments of tweets concerning COVID-19 and each sub-topic according to the TextBlob library.

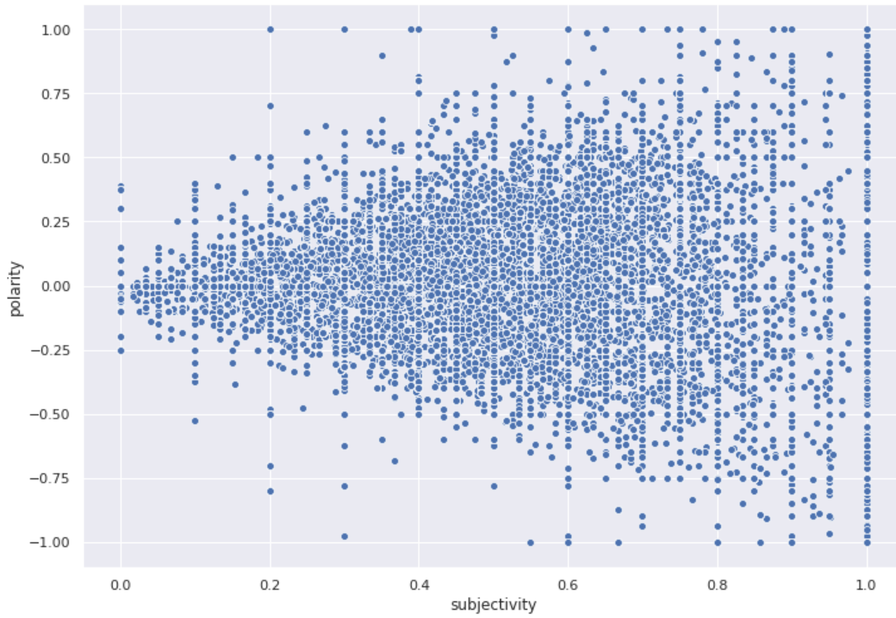

Figure 25. Tweets relating to POTUS and COVID-19 plotted according to polarity and subjectivity according to the NLTK library.