Estimating the Impact of Public Policies and Social Measures against COVID-19

Seungjae Ryan Lee
Princeton University

Abstract

Coronavirus Disease 2019 (COVID-19) is a disease caused by the SARS-CoV-2 virus that has been impacting the world heavily over the past few months. To combat these diseases, governments have enforced public policies, and communities have taken social distancing measures. These efforts have been recognized as crucial factors to the difference in the COVID-19 growth rates of countries around the world. We deploy neural network models to states in the United States to forecast the trend on daily deaths using public policy and social distancing data and analyze the significance and limitations of this approach.

1 Introduction

Coronavirus Disease 2019 (COVID-19) is a disease caused by the SARS-CoV-2 virus that has been affecting the world over the past few months. According to the World Health Organization (2020b), as of May 1, 2020, there are 3,175,207 confirmed cases of COVID-19 and 224,172 deaths, and among them, 84,771 confirmed cases and 6,403 deaths come from the last 24 hours.

The impact of COVID-19 varies widely between countries. As of May 1st, the United States leads the number of confirmed cases with 1,035,353 confirmed cases, contributing to nearly one-third of the total confirmed cases. The United States also has the greatest number of deaths, contributing to nearly one-fourth of the total deaths with 55,337 deaths (World Health Organization, 2020b). In contrast, the Republic of Korea (South Korea) only has 10,774 confirmed cases and 248 deaths, with 9 new confirmed cases and 1 new death on May 1st. Curiously, Korea had its first confirmed case on January 20th, one day earlier than the first confirmed case of the United States on the 21st. Furthermore, Korea had 3736 confirmed cases on March 1st, more than 60 times the confirmed cases of the United States then (62) (World Health Organization, 2020a). Yet, now the United States has nearly 100 times the confirmed cases and more than 220 times the deaths.

The United States has a higher population than South Korea, but this does not explain the huge discrepancy. The population of the United States is less than 7 times the population of South Korea, whereas the number of cases for the United States is more than 60 times that of South Korea. Furthermore, South Korea has a much higher population density with 517 people/km^2 compared to the United States with 34 people/km^2.

Scientists have attributed this discrepancy to multiple causes that can be broadly viewed in three categories: government interventions, communal efforts, and region-specific characteristics.

Government interventions include policies such as contact tracing, travel bans, and closure of facilities made effective by local governments. South Korea is an exemplary country that successfully deployed contact tracing (Zastrow, 2020). Communal efforts are voluntary social measures taken by individuals in the community. Social measures such as wearing masks and practicing social distancing have also been advocated by many researchers (Greenhalgh and Howard, 2020). Finally, region-specific static characteristics such as population density, average temperature, and climate have also been raised as possible causes for discrepancies.

To understand the significance of these features, we train various neural networks to predict the growth of COVID-19. We discover that recurrent neural networks utilize public policy and social distancing data better and achieve higher accuracy than fully connected neural networks.

2 Related Work

COVID-19 has garnered great interest from the scientific community and machine learning community over the past few months. Computer vision techniques have been used to analyze medical imaging such as X-rays or computed tomography (CT) images to diagnose COVID-19. Cohen, Morrison, and Dao (2020) compiled a database of medical imaging for COVID-19 cases for such studies. Xu et al. (2020) used the ResNet-18 architecture for feature extraction from CT images and achieved 86.7% accuracy on three-way classification among COVID-19, Influenza-A-viral-pneumonia, and irrelevant-to-infection. Song et al. (2020) also used ResNet-50 architecture to extract features from CT images and achieved an 86% accuracy and an F1-score of 0.87.

Researchers have also attempted to use natural language processing methods to understand COVID-19 using open-source pre-trained language models such as SciBERT (Beltagy, Lo, and Cohan, 2019) and ClinicalBERT (Alsentzer et al., 2019). With these models, they created search engines to browse information from over 47000 related scholarly articles in the CORD-19 (2020) dataset to assist the medical community to answer important questions.

Due to such quantity, it is nearly impossible to give a complete overview of prior work regarding COVID-19, so we focus on a few works especially relevant to our work. Our work focuses on predicting the number of deaths with deep neural networks to understand how social measures and government policies affect the spread of COVID-19. There have been multiple prior works predicting the growth of COVID-19 using epidemiological models or machine learning models.

Koczkodaj et al. (2020) approximated the growth of confirmed cases using nonlinear least squares to fit a simple exponential curve with two parameters. The authors predicted that by March 31st, there will be one million confirmed cases outside of China. This work is noteworthy in that the total number of confirmed cases outside of China did reach one million a couple of days after the predicted date. However, the authors assumed that during the early stages of COVID-19 infections, the growth will be exponential. This simple extrapolation and exponential growth assumption make the model inadequate. The number cannot grow exponentially forever, as there are only a finite number of people in the world. Furthermore, various intervention methods have been shown to decrease the rate of growth.

Cobb and Seale (2020) used a random forest model to investigate the relationship between the growth rate of COVID-19 and the shelter-in-place (SIP) orders among US counties. They first selected counties that had confirmed cases before March 16th and divided the counties into two categories: those that experienced a SIP order from March 19th to March 31st, and those that did not experience it. They found that the compound growth rate decreased by $14.4\% \pm 1.6\%$ for counties with SIP order, which was 7.8% more than $6.6\% + 1.4\%$ by those without SIP order. Then, a random forest model was trained to predict the compound growth rate after a SIP order given the following features: latitude, longitude, population, median age, number of physicians, median income, population/sq. mile, and water use per capita. They found that the most important features were population, longitude, and population density. Counties with high population and population density benefited greater with the SIP order than those with lower population or population density, and counties with a longitude between -79.7102° and -972363° benefited more than counties outside this zone.

Dandekar and Barbastathis (2020) used a neural network to quantify the effect of quarantine control on the spread of COVID-19, focusing on four locales: Wuhan, Italy, South Korea, and the United States. They use both classical epidemiological models and machine learning models to predict the spread of COVID-19. They discovered that classical SIR and SEIR models did not predict the stagnation in the number of infected people, whereas the neural network model trained with quarantine data predicted the stagnation correctly. The neural network was a two-layered fully connected neural network with 10 hidden units and the ReLU activation function, with the susceptible, infected, recovered, and quarantined population as the input and the "quarantine strength term" as the output.

Soures et al. (2020) created SIRNet, a combination of the SEIR model and neural networks to predict the effects of social distancing. First, the number of confirmed cases, recovered cases, population density, and mobility data was given as input to a linear cell (fully-connected layer) or an LSTM cell (recurrent layer). The outputs of this layer were then given to the SEIR cell, a recurrent neural network that follows the SEIR epidemiological model. The SIRNet was tested in 6 countries: Germany, Spain, France, Italy, South Korea, and the United States, and the model showed that "the epidemic is highly sensitive to changes in mobility rates."

Pal et al. (2020) used artificial neural networks to predict the long-term risk of countries using COVID-19 trend data and weather data. The authors used recurrent neural networks to extract features from these temporal data and predict the trend using these features. They then created a rule-based algorithm to group countries into various risk categories. This work is significant in that it incorporated weather data to predict the trend of COVID-19. However, the authors did not analyze the importance of various features, so it is uncertain how significant the weather data was for the prediction of the trend.

Overall, various machine learning methods have been used to predict the trend of COVID-19, but there has not yet been a comprehensive study that compares the importance of diverse features on public policy, community efforts, and the environment. Our study incorporates these features to train deep neural networks that predict the number of deaths by COVID-19. Then, by comparing the accuracies of neural networks with and without each feature, we estimate the predictive power of each feature. The results of this project could aid public policy decisions or serve as a verification of the effectiveness of social distancing.

3 Dataset

For this work, we selected four data sources directly related to COVID-19. The New York Times COVID-19 Data in the United States dataset (New York Times, 2020) provides the number of confirmed cases and deaths for each state in the United States. The COVID Tracking Project (2020) provides the number of tests administered per state and the data-quality grade of each state. The Google COVID-19 Community Mobility Reports (Google LLC, 2020) provide the change in mobility for six different categories of locations for each state. Finally, articles by Mervosh, Lu, and Swales (2020) and Mervosh et al. (2020) provide when each state went into shutdown and reopened.

We also include additional features unrelated to COVID-19 to characterize the states. These features are the latitude, longitude, population, and state area of each state. These features not only characterize the states but also allow us to generate more useful features through transformations.

3.1 The New York Times COVID-19 Data in the United States

The New York Times COVID-19 Data in the United States dataset (New York Times, 2020) offers a cumulative count of the number of confirmed cases and deaths in the United States. Confirmed cases are patients who test positive for COVID-19, and the number of confirmed cases is determined by reports from a federal, state, territorial, or local government agency. For deaths, only those confirmed in labs are counted to the database. The data is organized in country-level, state-level, and county-level, and starts from January 21, 2020, when the first reported case appeared in Washington. It is collected and updated daily by journalists across the United States monitoring news conferences, analyzing data releases, and directly seeking clarification from officials.

The data sometimes differ from the information reported by states and counties due to various special cases. For example, in this dataset, the patients are counted in the location where they receive treatment, which may not necessarily be where they live. However, some local health departments do not count these patients. Similarly, "when a resident of Florida died in Los Angeles, [the dataset] recorded her death as having occurred in California rather than Florida, though officials in Florida counted her case in their own records." The New York Times team also reports that sometimes officials transport patients identified in one state or county to another, often without explanation, resulting in inaccurate data.

There are many more reasons that this dataset may contain inaccuracies. First, there can be a temporal lag between true data and reported data. This may come from the time needed to test, verify, and report the case, but also from the innate properties of the SARS-CoV-2 virus. Since those who are infected by the virus remain asymptomatic for 2-14 days, the true number of confirmed cases may be drastically undercounted.

Another reason to suspect a difference between the reported number and the true number is the inaccuracies of the COVID-19 tests. There are various testing kits for COVID-19 with varying accuracies. False positives could inflate the number of confirmed cases and deaths, and false negatives could deflate the number of confirmed cases and deaths.

Furthermore, the number of confirmed cases can be especially inaccurate, as it is strictly dependent on the number of tests performed: the number of confirmed cases cannot exceed the number of tests performed. As the United States suffered from a lack of testing kits, the discrepancy between true confirmed cases and reported confirmed cases may be large. In fact, one study suggests that this true confirmed cases could be 28 to 55 times higher than the reported confirmed cases in Los Angeles County (Hopper, 2020).

Although better than the number of confirmed cases, the number of deaths is also inaccurate. The deaths may include people who may soon have died without COVID-19 (Wu et al., 2020) or exclude those who have died before being tested positive for COVID-19 (Chen, 2020). Despite such shortcomings, we use the number of deaths as our trend variable and our variable to predict, as it is more accurate than the number of confirmed cases.

Figure 1 shows the total confirmed cases over time, and Figure 2 shows the total deaths over time. New York has the highest number of confirmed cases and deaths, followed by New Jersey. Both plots show a flattening of the slopes of two states around April 20th, the second dotted line. The first dotted line is March 20th, when states started announcing stay-at-home orders. The flattening of the curve aligns with the studies that claim that social distancing efforts become apparent 30 days after they start.

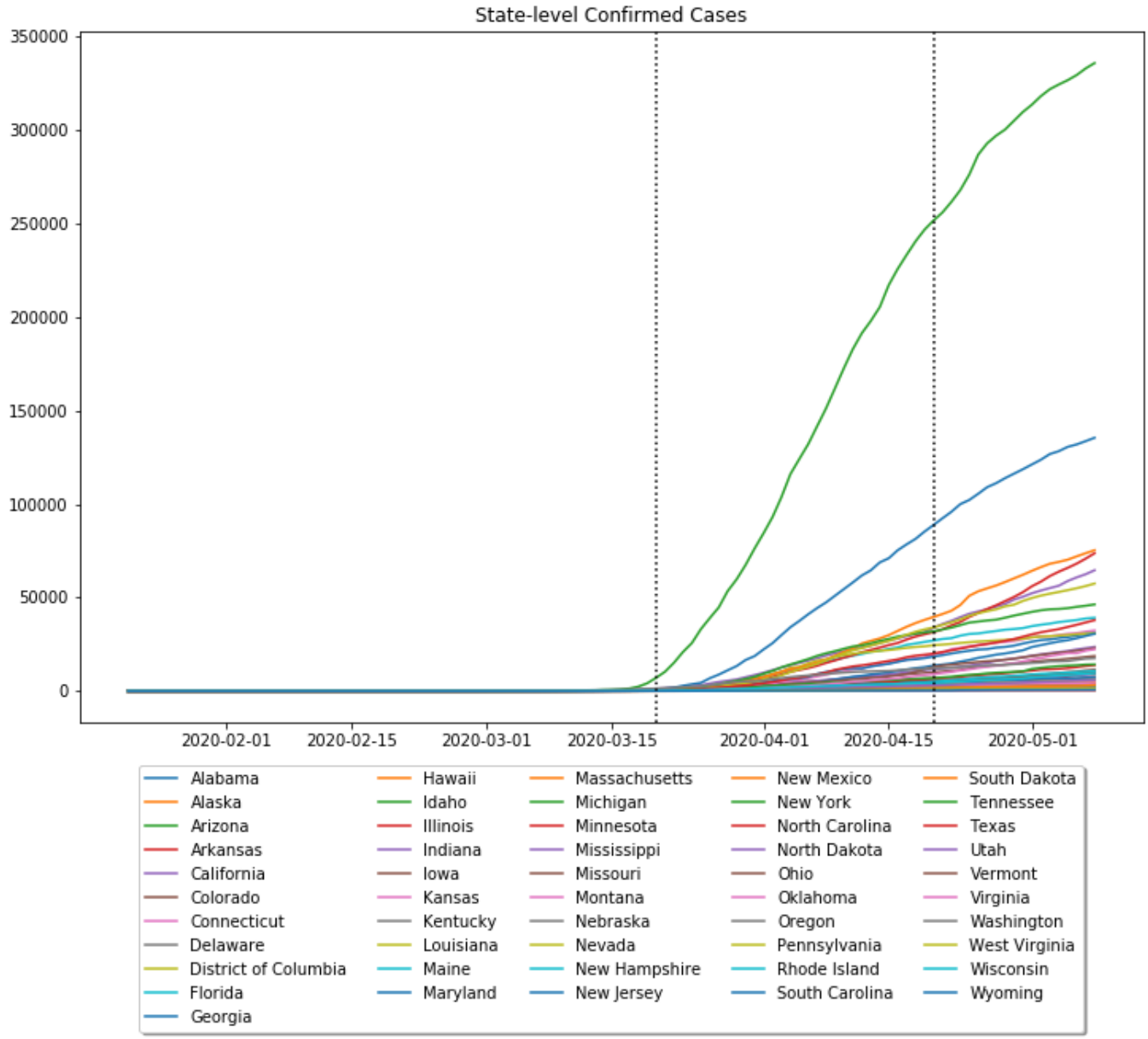

Figure 1. State-level confirmed cases from January 21st to May 8th. The left dotted line marks March 20th, and the right dotted line marks April 20th.

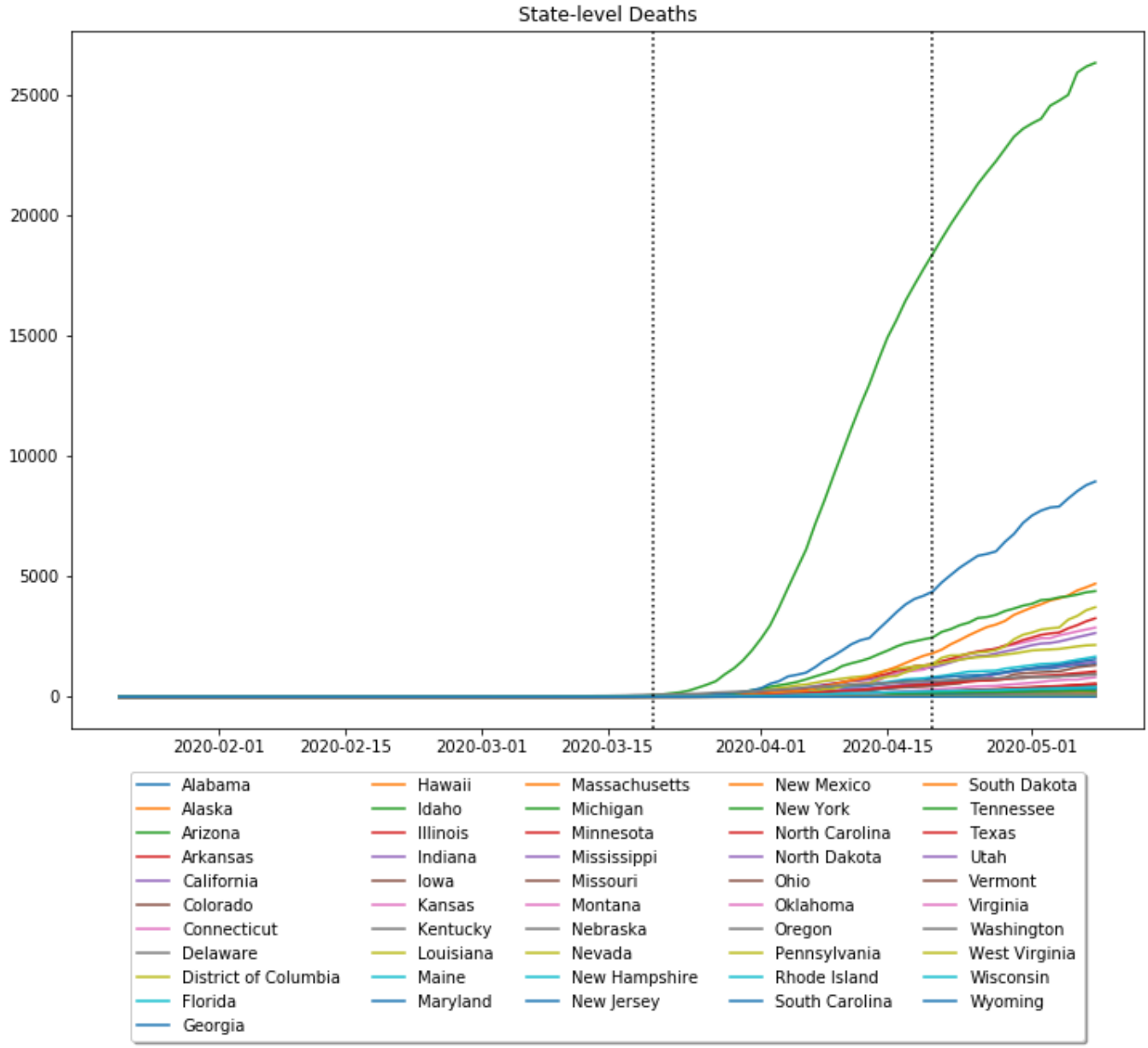

Figure 2. State-level deaths from January 21st to May 8th. The left dotted line marks March 20th, and the right dotted line marks April 20th.

3.2 The COVID Tracking Project

The COVID Tracking Project (2020) is a volunteer organization that collects COVID-19 data in the United States from all 50 states, 5 territories, and the District of Columbia. The data comes from state public health authorities, through reports, news sources, or other means. The project also assigns each state a "data-quality grade based on [their] assessment of the completeness of their reporting."

The COVID Tracking Project database provides many useful numbers. This ranges from the number of positive cases and deaths that can be found in other databases, but also more sophisticated data such as the number of tests administered, patients hospitalized, in ICU, or on a ventilator. However, most of this data is available only for the entire nation and not available in many states. Furthermore, they note that patient outcomes and demographics data are reported inconsistently. To use these more sophisticated variables, we have attempted to remove the early days of COVID-19 when there were few cases but found that there were still too many missing values. Therefore, from this dataset, we only use the total number of tests administered and the data-quality grade, which have no missing values and are relatively well-reported.

Figure 4 shows how the total number of tests per state increased over time. We see that New York, the state with the most confirmed cases and deaths, also has the most number of tests as of May 1st with 927438 tests, followed by California, Florida, and Texas. Except for these four states, the number of tests is similar between states, as shown by the histogram in Figure 3. Most states have less than 100,000 tests administered.

Each state is also given a letter grade: A+, A, B, C, D, and F. These grades are computed through 16 categories that measure how thorough and accessible the data is. These categories include whether the state reports the statistics on tests (total, positive, negative) and outcomes (hospitalized, ICU, ventilator, recovered). Furthermore, some categories check if the states give a demographic breakdown of confirmed cases and deaths. No state has data quality grade F. Also, nearly all states have a grade of C or above, and around two-thirds of the states have grade B or above, with grade B being the most common. The histogram in Figure 5 shows how many states have each data quality grade.

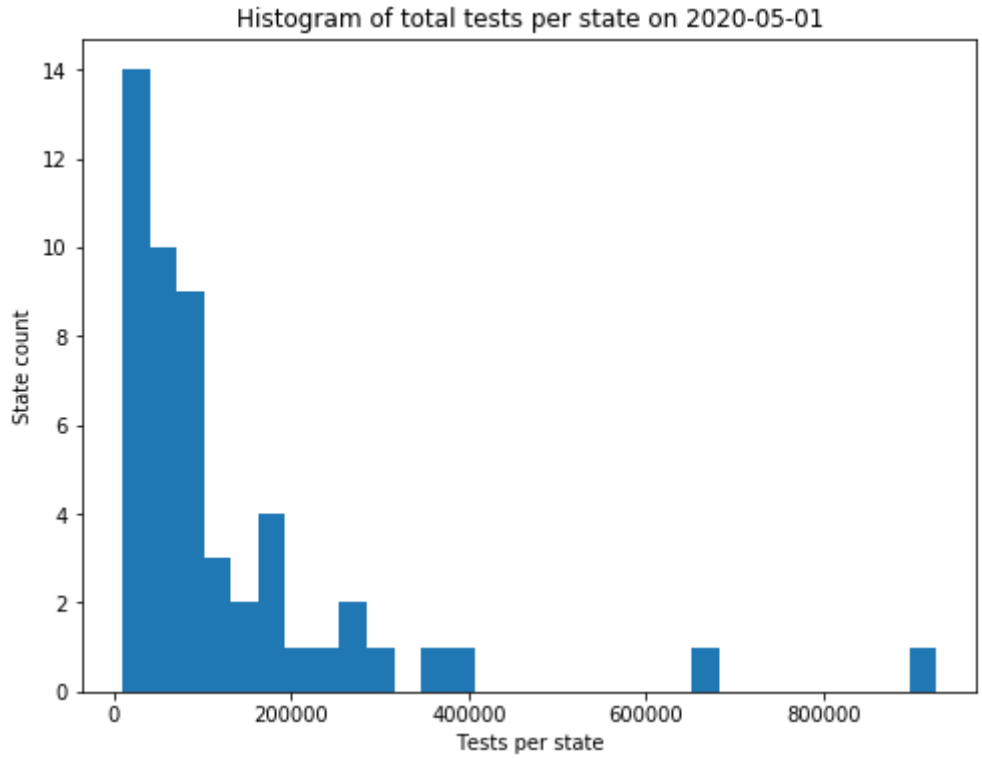

Figure 3. Histogram of total tests per state on May 1st, 2020. There are a few states that have tested more than 300,000 people, but most states have tested less.

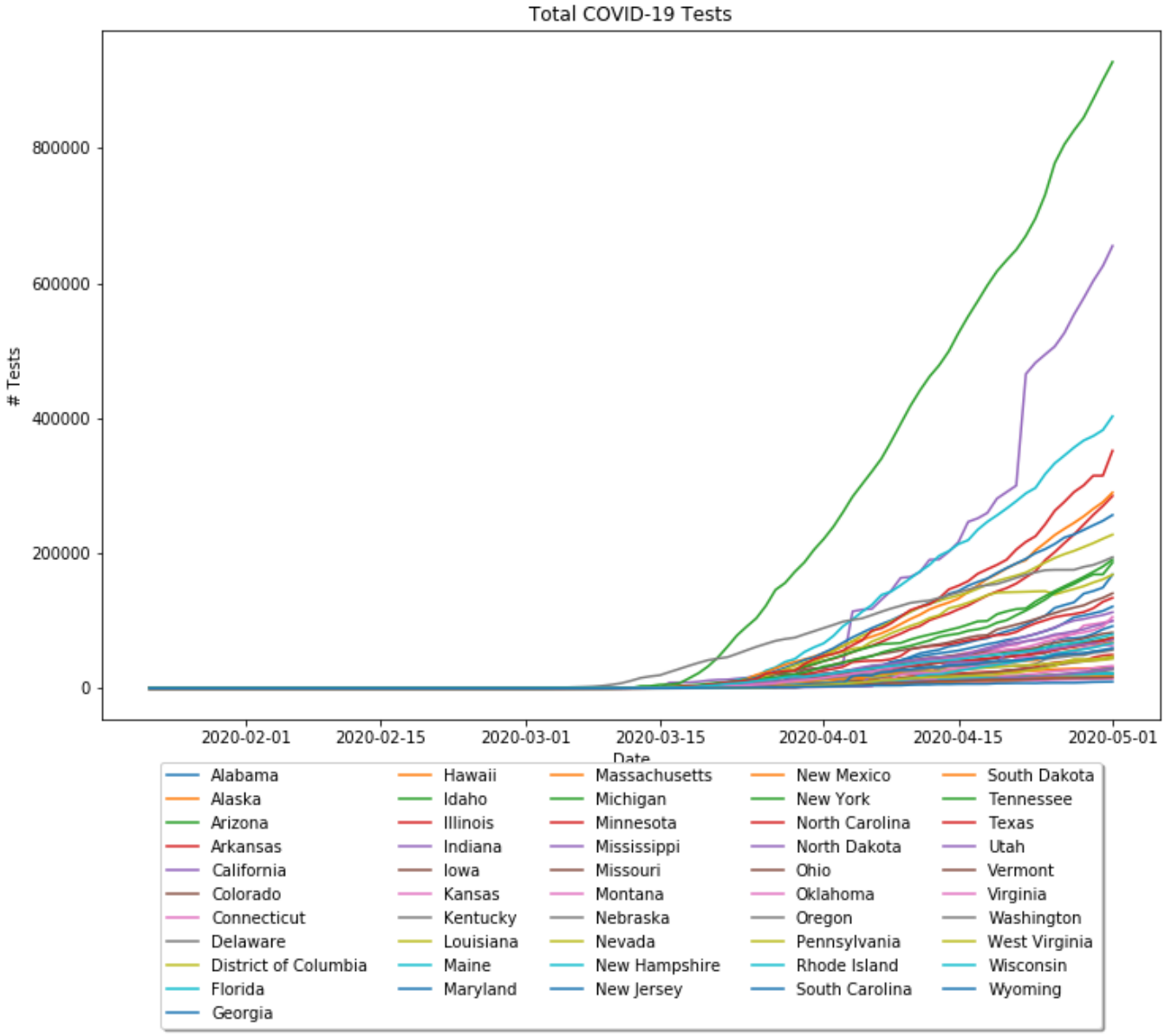

Figure 4. Growth in the number of tests per state. New York is the uppermost line on May 1st, followed by California, Florida, and Texas.

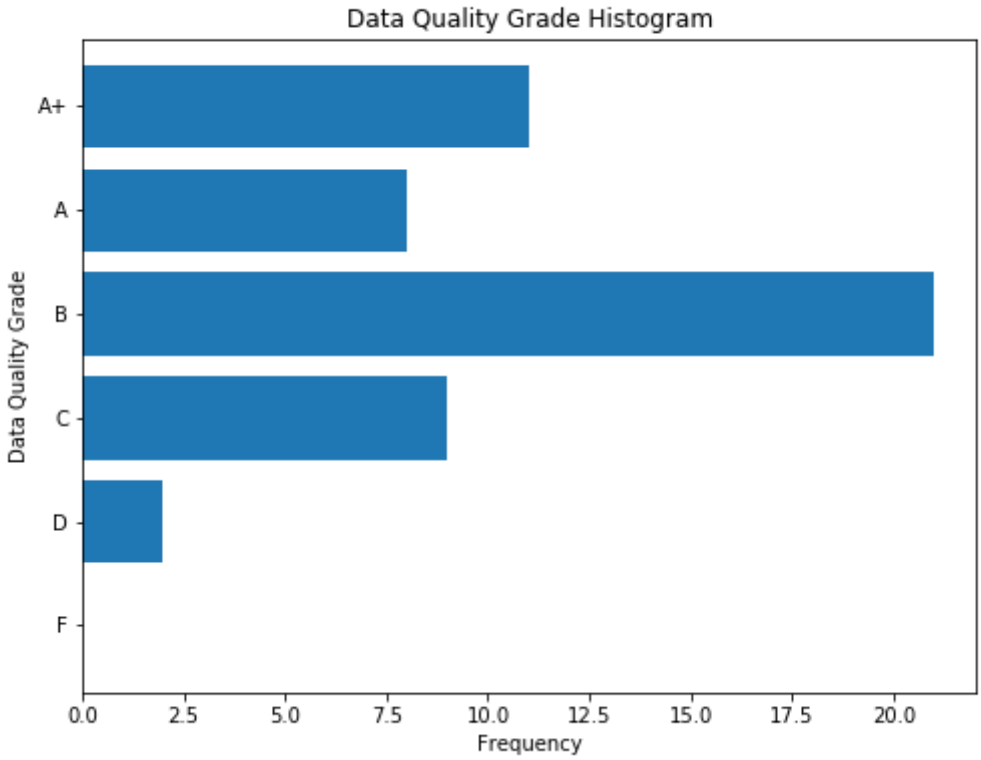

Figure 5. Histogram of data quality grades per state. The 5 territories are not shown in this histogram.

3.3 Google COVID-19 Community Mobility Report Dataset

Google LLC (2020) published a set of community mobility reports that displays the changes in visits to places such as grocery stores, parks, or transit stations. These reports were created by aggregating anonymized data from users that have voluntarily used Google's Location History setting. The dataset contains data for multiple countries, and for the United States, it also contains state-level and county-level data. The mobility data does not give absolute values on the number of visits, but only the changes in visits from the "baseline." The baseline is calculated by taking the median of the 4 weeks from January 3rd to February 6th, 2020. Because the number of visits can differ based on the day of the week, the baseline is calculated separately for each day of the week.

The mobility report displays changes in visits to six place categories: grocery and pharmacy, parks, transit stations, retail and recreation, residential, and workplaces. Note that parks include not only local and national parks but also public beaches, marinas, dog parks, plazas, and public gardens. Likewise, retail and recreation include various places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters.

Figure 6 shows the mean and standard deviations of each feature in the dataset. Workplace mobility and transit station mobility decreased overall while residential mobility increased, which could be a consequence of more people working from home and the unemployment rate rising. Grocery and pharmacy mobility did not change much as these were necessary trips. Interestingly, park mobility increased, showing that people still spent time outdoors. Figure 7 plots a Gaussian curve using these mean and standard deviation values against a histogram for each feature. Curiously, the mobility change in parks and grocery and pharmacy followed a Gaussian fit, whereas other features were more bimodal.

On the state level, the mobility report contains no missing data. However, on the county level, there is a lot of missing data. Figure 8 plots the ratio of missing county-level data for each of the six mobility features. For features such as workplace data, only 6% of the data is missing. However, nearly 75% of the data is missing for park mobility data. To use most of the data, we decided not to focus on individual counties and instead look at them at the state level.

Feature	μ	σ
Retail & Recreation	-15.067767	23.666614
Grocery & Pharmacy	-0.531132	12.827270
Parks	15.600275	33.646821
Transit Stations	-16.303406	24.612359
Workplaces	-18.796698	21.609043
Residential	7.678707	8.837952

Figure 6. Mean and standard deviations of mobility features

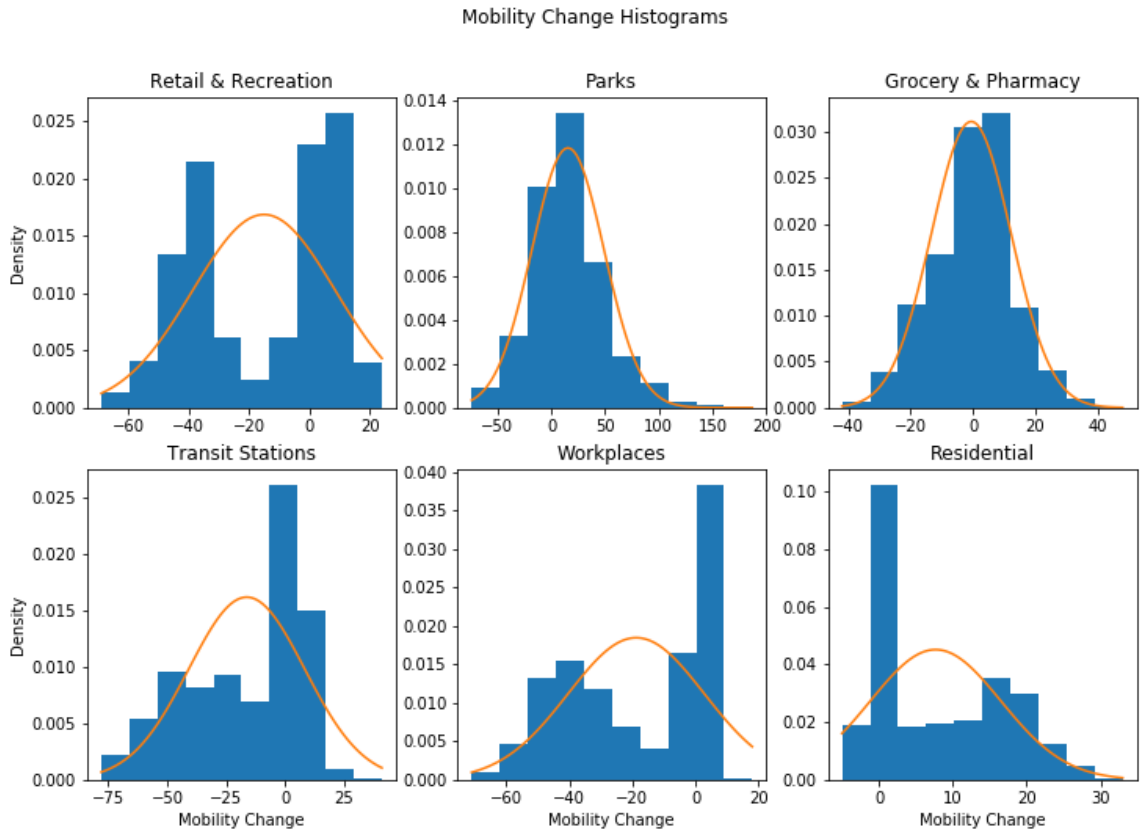

Figure 7. Histogram of state-level mobility change from February 15 to April 11.

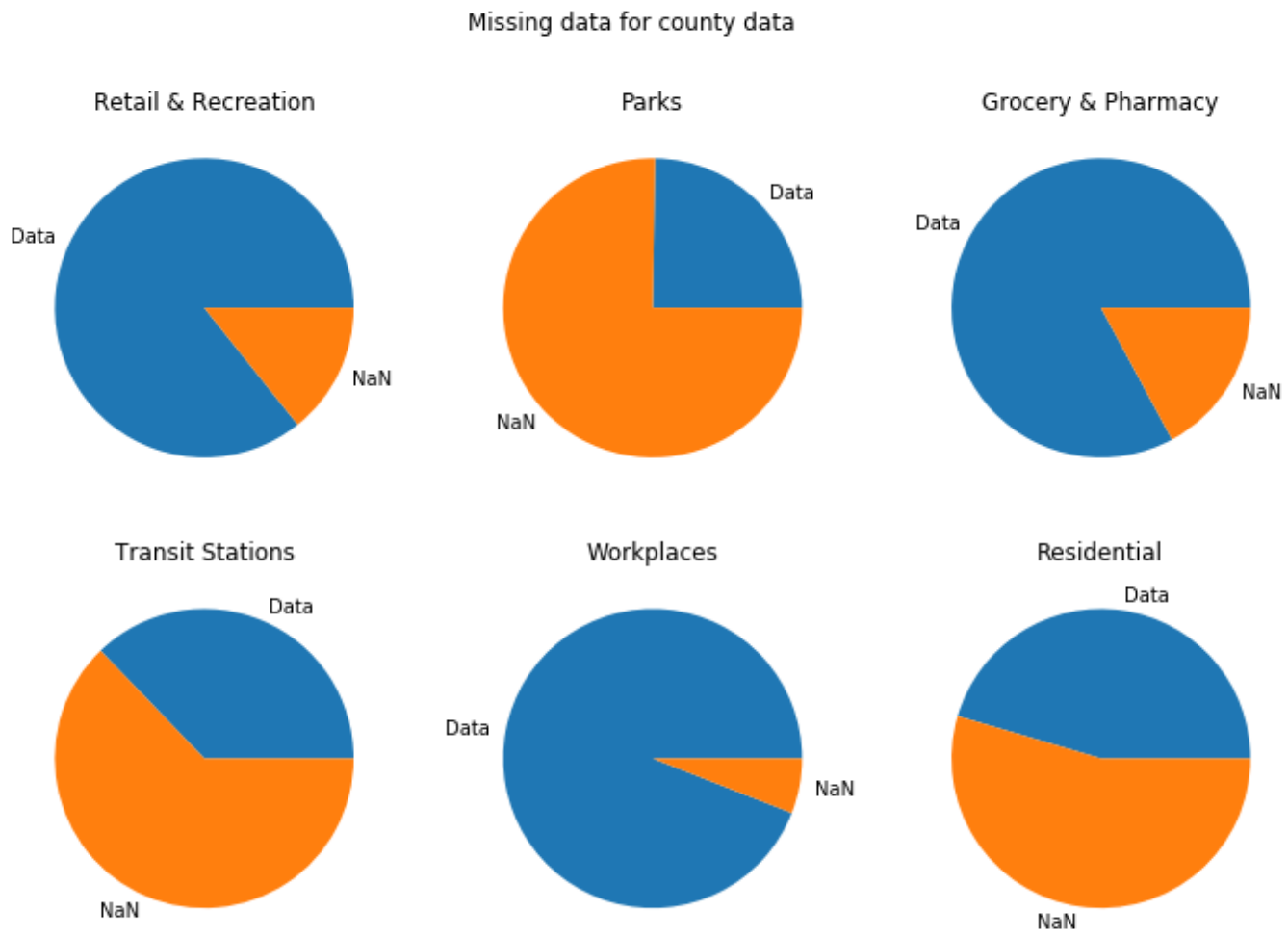

Figure 8. Pie charts of missing data in county-level mobility data.

3.4 New York Times Shutdown Dataset

Mervosh, Lu, and Swales (2020) and Mervosh et al. (2020) compiled the shutdown orders of each state with the date of effect and date of expiration. The data was provided as news articles, so we converted it to structured data. The articles only check whether the shutdown order is in place or has expired, so it does not provide to what extent the state went under shutdown and is reopening. For example, Michigan's shutdown order expires on May 15, but recreation and retail services have already reopened. In contrast, Arkansas has no shutdown order but placed restrictions closing gyms, hair salons, and other retail and beauty services. Therefore, the data cannot serve as an exact measure of the state policy, as it differs in stringency and scope. Rather, it provides a general idea of when the state decided to take action.

Figure 9 displays how many states were under stay-at-home order on each date. 43 states have been shut down for at least some period, and they were all shut down from April 7th to April 24th. As of May 9th, 27 states are still under stay-at-home orders.

Figure 10 shows when each state was under a stay-at-home order. We only plot states that announced stay-at-home orders before May 9th, and the states are ordered by their start dates. California was the earliest state to announce the order on March 19th, and South Carolina was the latest with April 7th. Washington has the latest reopening date, planned on July 12th, but some states (California, Kentucky, Maryland, Oregon) do not have plans to reopen.

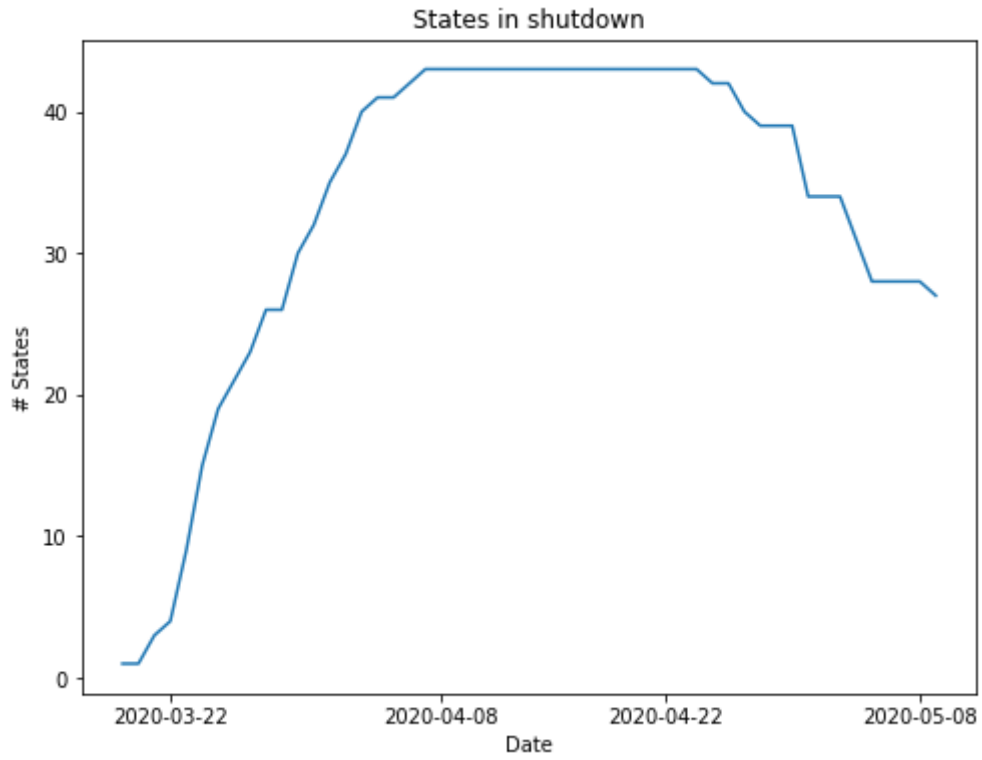

Figure 9. Number of states in shutdown for each date from March 19th to May 9th..

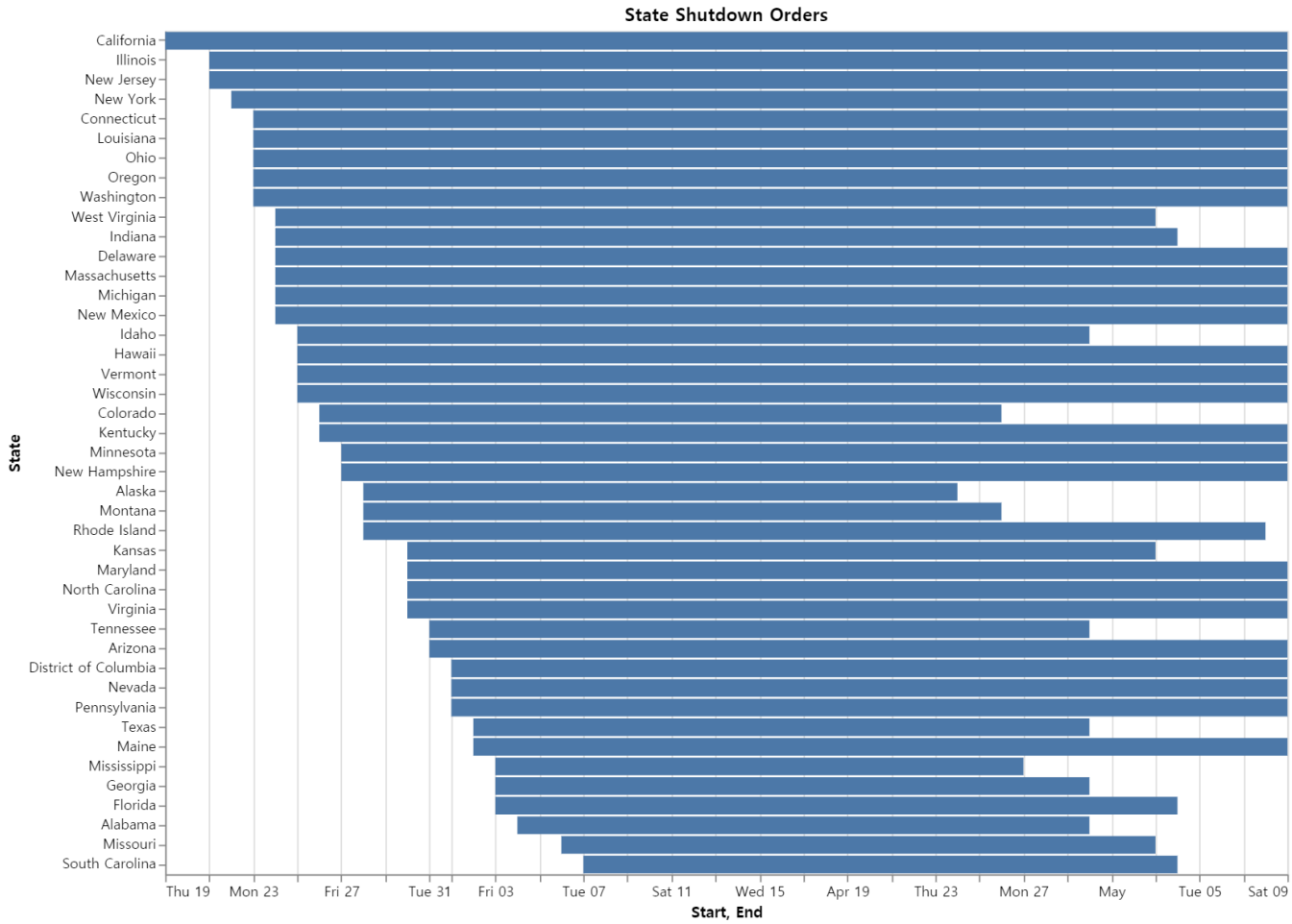

Figure 10. Gantt chart of state shutdown orders. The earliest date is March 19th, and the latest date is May 9th, the date the data was collected. States that did not reopen and states with reopening dates after May 9th was cut off at May 9th.

3.5 Additional Data

Following Cobb and Seale (2020), we include additional features that characterize how vulnerable the state is against COVID-19. We selected latitude, longitude, population, and state area as our features. The latitude and longitude of each state were found in the Dataset Publishing Language (Google, 2012). The population data and the state area measurements come from the US Census Bureau (2018, 2019). These features not only characterize the states but also allow us to generate more useful features through transformations. Two such transformations are the infected rate and the fatality rate computed by dividing the number of confirmed cases and deaths by the state population.

Figure 11 shows the distribution of latitudes and longitudes. The two noticeable outliers are Hawaii (latitude 19.898682, longitude -155.665857) and Alaska (latitude 63.588753, longitude -154.493062). Other mainland states have latitudes around 30 to 50 and longitudes around -130 to -65.

Figure 12 shows the total population of each state according to the US Census Bureau (2019). Four states have a significantly higher population than other states: California (39512223), Texas (28995881), Florida (21477737), and New York (19453561). The US Census Bureau (2018) also provides data on the land area, water area, and total area for each state. Most states have similar total areas and land areas except Alaska and Texas, which form the two outliers seen in the first two histograms of Figure 13. In terms of water area, Alaska again has the highest area, with the next outlier being Michigan. All other states have similar water area. With these additional data, we also computed population density. Figure 14 shows the histogram of population density, computed by dividing the state population by the land area. The outlier is the District of Columbia, which has density one order of magnitude greater than other states.

With these additional data, we also computed the population density. Figure 14 shows a histogram of the population density, computed by dividing the state population by the land area. One outlier is the District of Columbia, which has density one order of magnitude greater than other states.

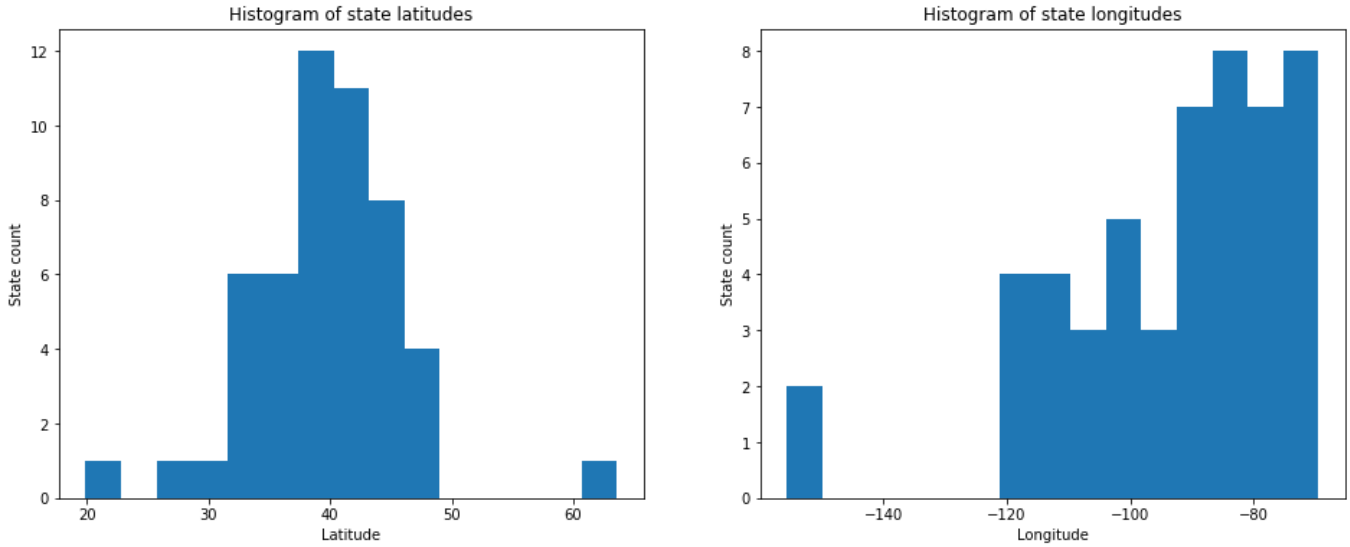

Figure 11. Histogram of state latitudes and longitudes.

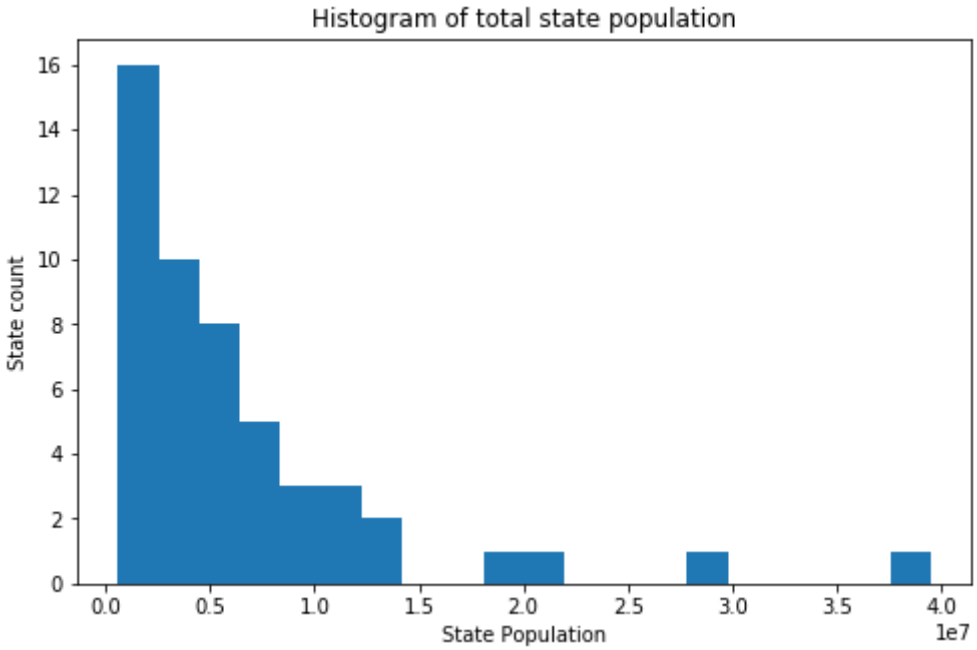

Figure 12. Histogram of state population.

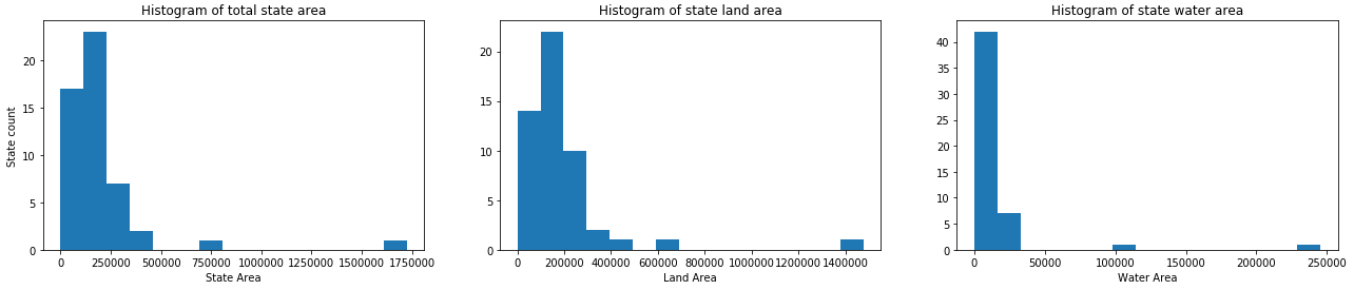

Figure 13. Histograms of state areas.

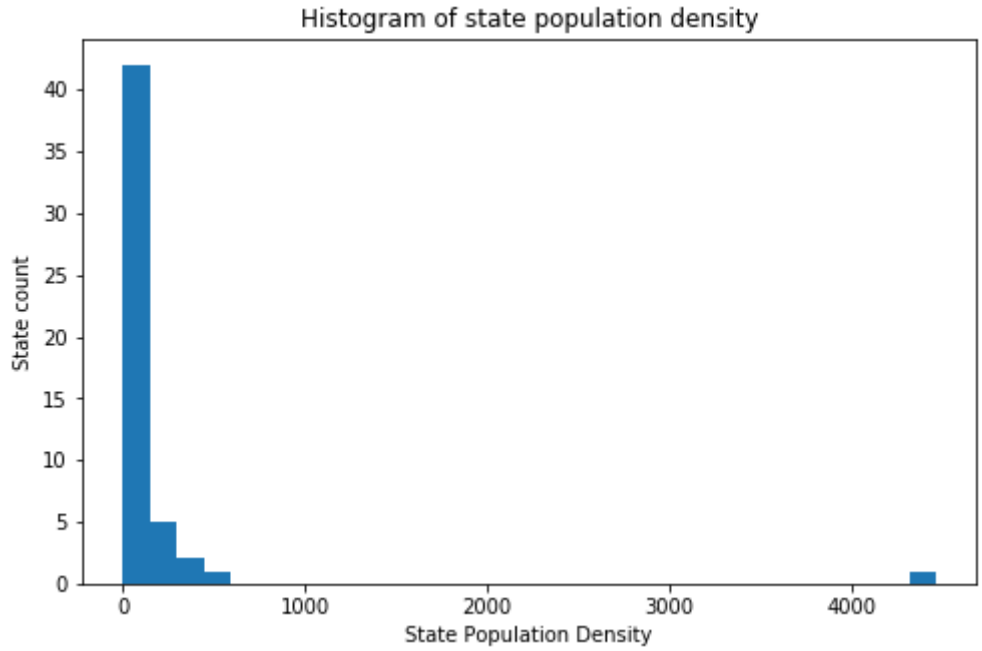

Figure 14. Histograms of state population density.

4 Experiments

We designed two experiments to understand the trend of COVID-19 and the impact of social distancing and government policies. Since the COVID-19 data is time-series data, we compare the predictions of fully connected neural networks and recurrent neural networks through the mean squared error loss function and permutation feature importance. Also, we design two different tasks - 7-day prediction task and 14-day prediction task - and analyze how the predictions differ for networks trained for these different tasks by visualizing the predicted trend against the true trend.

4.1 Data Preprocessing

Before training, we preprocess the datasets for neural network training. This includes reshaping the data to input and target variables, splitting the dataset into train, validation, and test sets, and standardizing the data.

Each data point consists of a 28-day "window." COVID-19 is known to have up to 14 days of asymptomatic period, so the first 14 days of this window are given to the model as the input. For the 14-day prediction model, the daily number of deaths in the second 14 days of the window becomes the target to predict. For the 7-day prediction model, the daily number of deaths in the last 7 days of the window becomes the target to predict.

Each data point contains 189 features: 182 features from 14 days of data with 13 features each day, and 7 features describing the state. Note that we include both cumulative and daily confirmed cases, deaths, and tests. The daily numbers are included since they are more directly relevant to predicting the daily number of deaths. However, since the data point has a 28-day window, so the daily numbers cannot substitute cumulative numbers, so we also include cumulative numbers.

Features	Type
Cumulative Cases	Time-series
Cumulative Deaths	Time-series
Cumulative Tests	Time-series
Daily Cases	Time-series
Daily Deaths	Time-series
Daily Tests	Time-series
Retail & Recreation Mobility	Time-series
Parks Mobility	Time-series
Grocery & Pharmacy Mobility	Time-series
Transit Stations Mobility	Time-series
Workplaces Mobility	Time-series
Residential Mobility	Time-series
Stay-at-home Order	Time-series
Total State Area	Non-time-series
Total Land Area	Non-time-series
Total Water Area	Non-time-series
Latitude	Non-time-series
Longitude	Non-time-series
State Population	Non-time-series
State Population Density	Non-time-series

Figure 15. Features given to neural network as input.

The dataset was split into the non-test set and test sets. The test set contains features from April 4th to April 17th to predict the number of daily new deaths from April 18th to May 1st (14-day prediction model) or from April 25 to May 1st (7-day prediction model). The non-test set contains features from February 15th to April 3rd. Each feature in the non-test data was standardized to have mean 0 and variance 1 with the following formula:

$$X = \frac{X - \bar{X}}{s}$$

where X is the data, $\bar{X}$ is the mean of the data, and s is the standard deviation of the data. For the test data, we use the mean and standard deviation of the non-test set to standardize the test data.

$$X_{test} = \frac{X_{test} - \bar{X}_{nontest}}{s_{nontest}}$$

20% of the non-test set was used for hold-out validation, and the remaining data were used for training. The resulting training set contained 1509 data points, and the validation set contained 378 data points.

4.2 Model Architecture

Different neural network architectures have been used widely for time series data. Fully connected neural networks are used in most architectures at the end of the network to interpret features extracted through recurrent neural networks or convolutional neural networks. Due to sharing weights across the data, convolutional neural networks (CNNs) have the benefit of having a smaller number of weights than fully connected layers and can learn transition-invariant characteristics. Therefore, CNNs have been used in most computer vision architectures for feature extraction. Recurrent neural networks (RNNs) sequentially interpret the data, using its internal memory to remember other parts of the sequence. This allows RNNs to extract fixed-length features from data with variable lengths, and thus they have been used as feature extractors in natural language processing tasks where the inputs are sentences or audio of variable sizes.

Our data is small enough that CNN's benefit of shared weights would be minimal. Therefore, in this paper, we test two neural network architectures: fully connected neural networks and neural networks with recurrent layers. For a fair comparison, we use models of similar size by setting the number of parameters to be roughly equal. Figure 16 shows the number of parameters for each neural network we trained.

Task	Network	Parameters
7-day prediction	FCNN	1119239
7-day prediction	RNN	1094663
14-day prediction	FCNN	1169471
14-day prediction	RNN	1152063

Figure 16. Number of parameters for each neural network.

The time-series features for 14 days, each with 13 features, are flattened to create a feature vector of size 182. Furthermore, there are 7 non-time-series features, resulting in a total of 189 features per data point. The fully connected neural network is given the concatenated feature vector with 189 features. This input passes through two hidden layers, each with 1024 and 896 hidden nodes, and goes to the output layer with 7 or 14 nodes, depending on the task. The recurrent neural network has two input locations. A many-to-one bidirectional LSTM (BiLSTM) layer receives daily features for 14 days and outputs a vector of size 512, 256 from each direction. Then, this vector concatenated with the non-time-series features is given to a fully connected hidden layer of size 1024 and finally goes to the output layer with 7 or 14 nodes, depending on the task. Every BiLSTM and linear layer is followed by a rectified linear unit (ReLU), with the exception of the output layer (Nair and Hinton, 2010). Both architectures are shown in Figure 17.

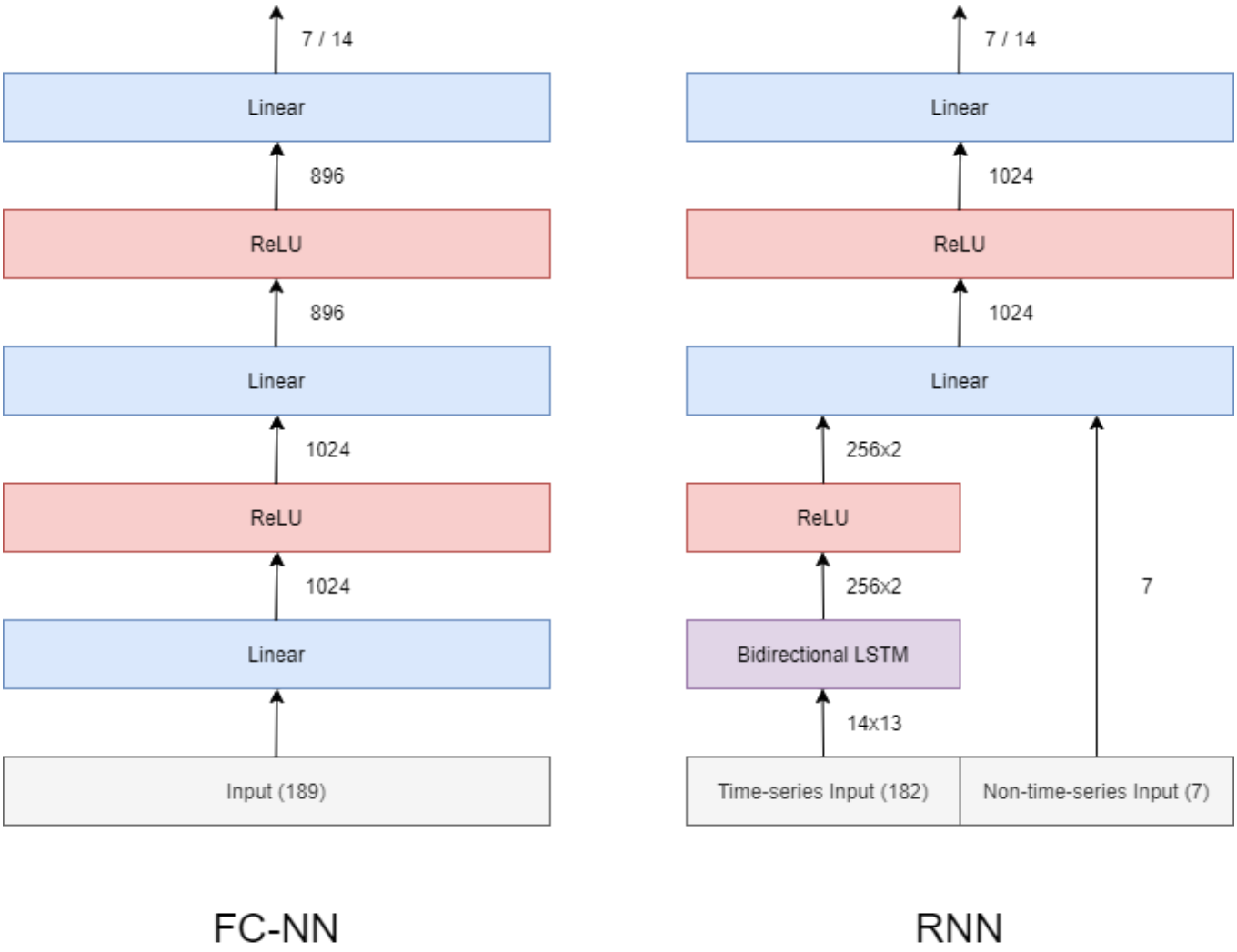

Figure 17. A visualization of the neural network architectures

4.3 Training

The neural networks were trained through gradient descent. We used batch gradient descent, where the entire training set was used for each gradient descent step. For both networks, we use the same hyperparameters for training. The training is done for 1000 epochs with the Adam optimizer (Kingma and Ba, 2014) with the mean squared error loss function. The learning rate was initialized as 0.0001 and was reduced by 70% if the validation loss does not decrease for 100 epochs.

Figure 18 shows the loss graph of FC-NN and RNN for the 7-day prediction. The validation loss of both FC-NN (green) and RNN (orange) stopped decreasing after approximately 300 epochs. For the 14-day prediction shown in Figure 19, the validation loss plateaus after 500 epochs.

All training was done with 4 2.60 GHz Intel Skylake CPU-cores and 1 NVIDIA V100 GPU. Each neural network took 2-3 seconds to train. For preliminary training done solely on 4 2.60GHz Kaby Lake CPU-cores, the training took 2.5 minutes for FC-NNs and 22 minutes for RNNs.

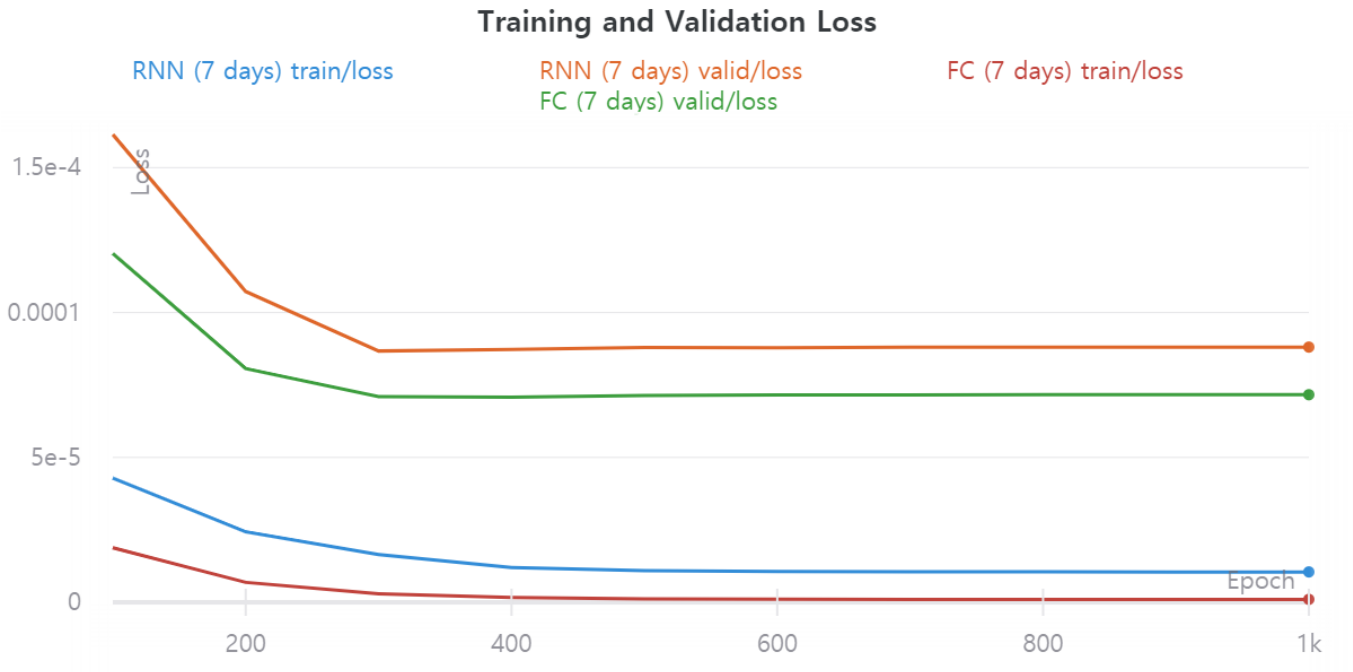

Figure 18. Training and validation loss graph for FC-NN and RNN for 7-day prediction

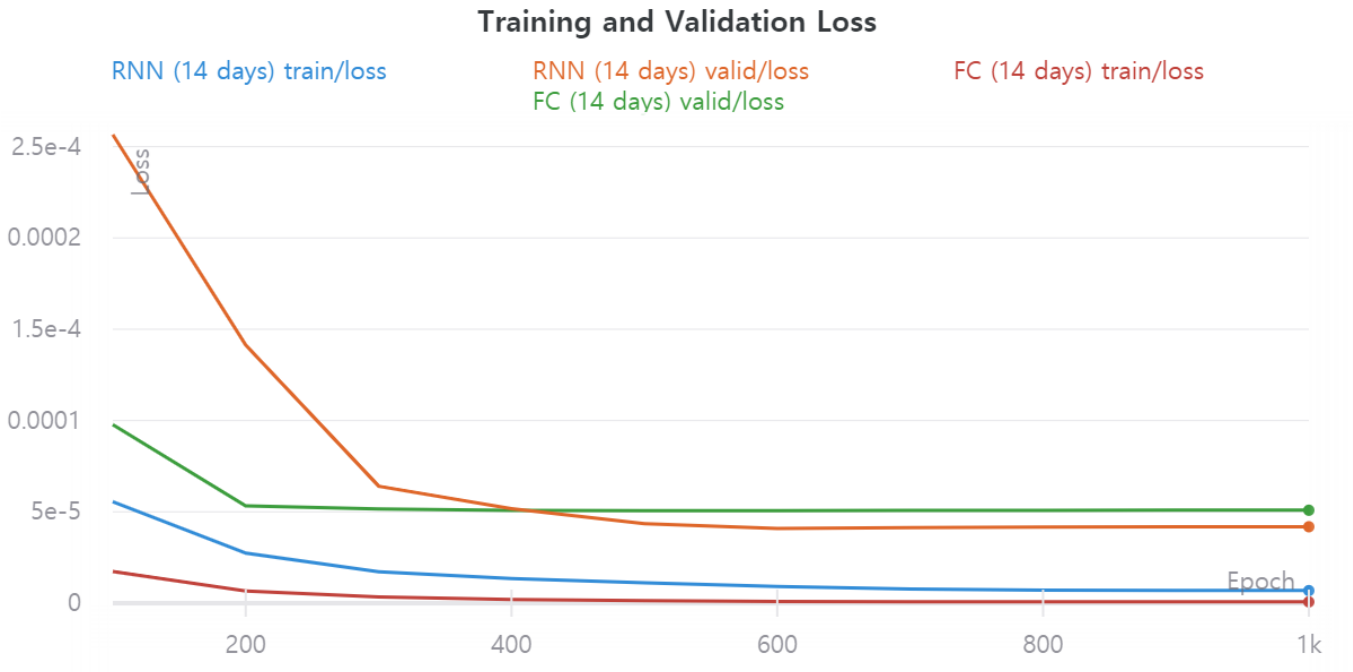

Figure 19. Training and validation loss graph for FC-NN and RNN for 14-day prediction

5 Results

5.1 Fully Connected Networks (FC-NN) vs. Recurrent Neural Networks (RNN)

The two neural network architectures - FC-NN and RNN - converged with a similar training loss and validation loss (Figure 18, 19). To analyze the accuracy of each model, we computed the mean squared error on the test set. We found that the RNN generated predictions more consistent with the data with lower loss. Furthermore, the RNN was able to consistently outperform FC-NN on each of the outputs and did not diminish in performance when predicting the number of deaths on later days (Figure 20).

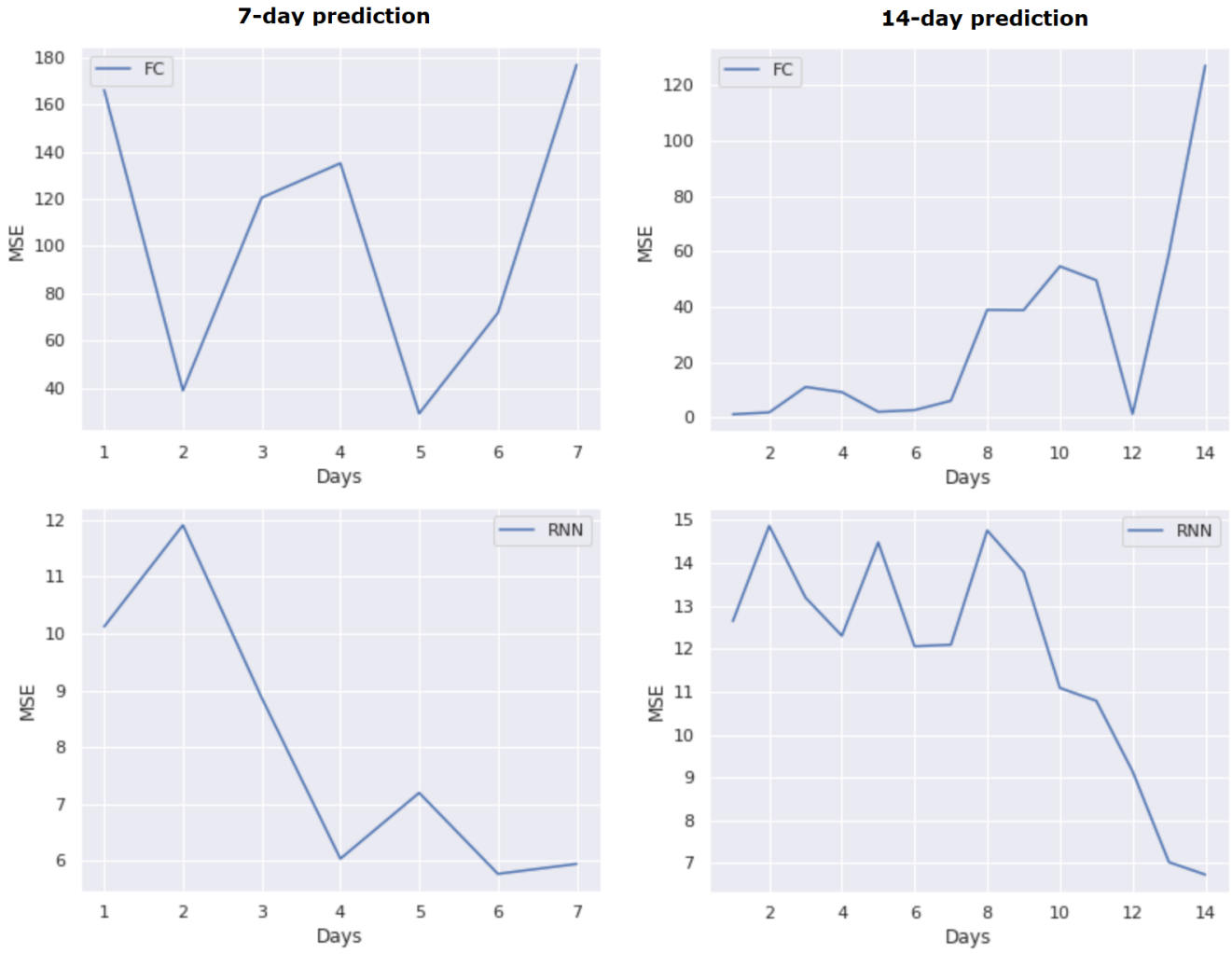

Figure 20. Mean squared error for each of the 7 or 14 days predicted by FC-NN (above) and RNN (below) for the 7-day prediction task (left) and the 14-day prediction task (right).

To justify the results, we also computed the permutation feature importance for each neural network. Permutation feature importance is a model inspection technique that measures the decrease in the model performance when a single feature value is randomly shuffled across the data (Breiman, 2001). We calculated the permutation feature importance for the 7-day prediction task.

Figure 21 displays the features with the top 10 feature importance. For fully connected neural networks, excluding the basic features on the number of cases, deaths, and tests, the only feature with significant positive feature importance is the mobility change in residential areas. This implies that the network did not use most of the social distancing mobility data or the stay-at-home government policy. In contrast, for the recurrent neural network, 5 mobility and government policy features display positive feature importance, implying that the neural network interpreted this data to improve its prediction.

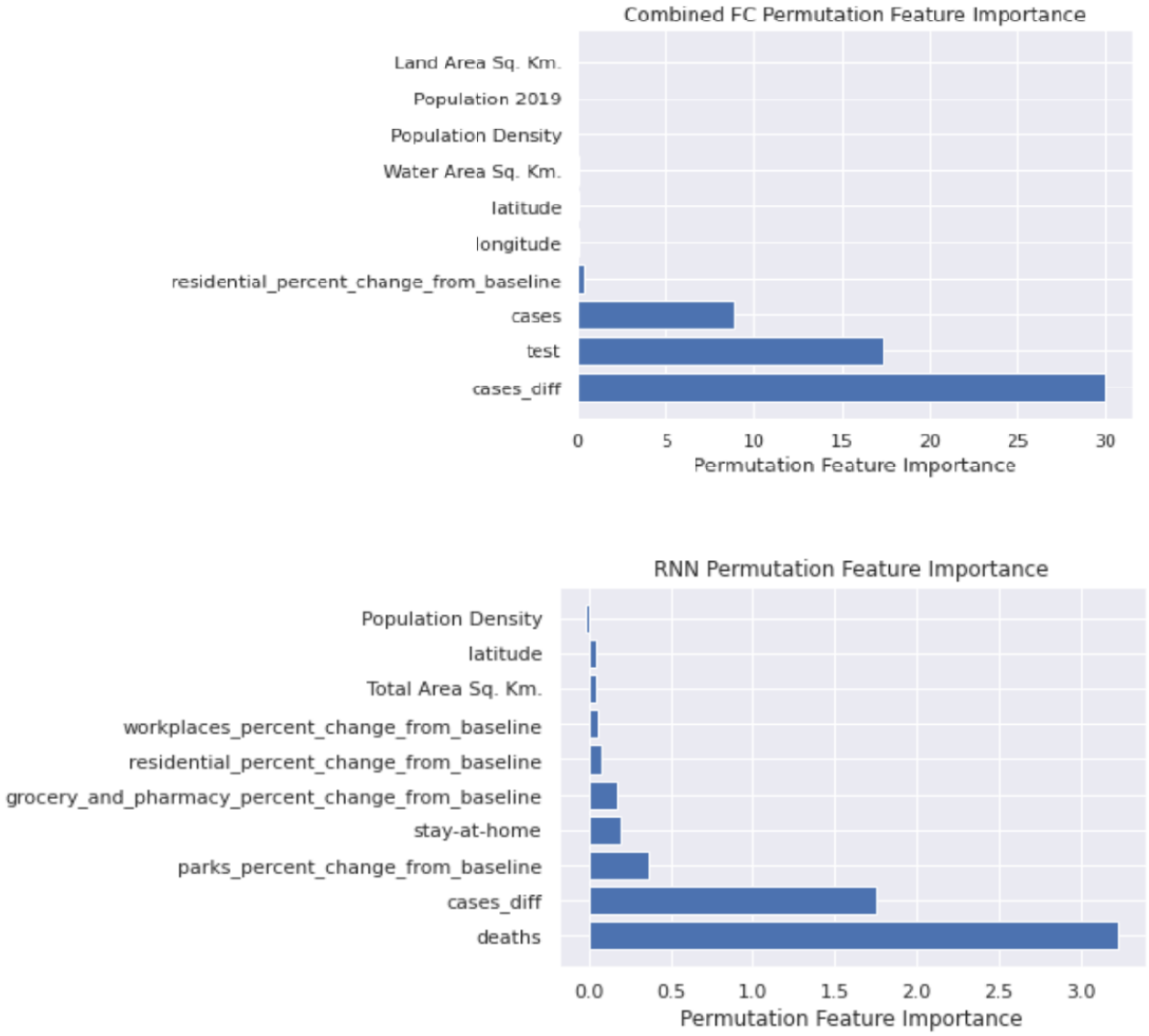

Figure 21. The permutation feature importance of the 10 features with the highest importance. The feature importance was computed both with FC-NN (above) and RNN (below) on the 7-day prediction task.

Because the recurrent neural networks outperformed fully connected neural networks on all days and utilized social distancing and government policy features more, we exclusively used recurrent neural networks for the remainder of our experiments.

5.2 7-day Prediction vs. 14-day Prediction

We compared the predictions of the recurrent neural network trained on the 7-day prediction task and the recurrent neural network trained on the 14-day prediction task. For each state, the networks were given the test-set data from April 4th to April 17th to predict the number of deaths from April 25th to May 1st. For the network trained on the 14-day prediction task, only the last 7 outputs were used.

Figure 22 shows the predictions of the RNN trained to predict the 7-day trend (red) and the RNN trained to predict the 14-day trend (blue). We find that one RNN does not strictly outperform another and that they achieve similar performance in most states. Also, the accuracy of the predictions varies widely depending on the state. For New York, both RNNs were able to predict the trend very well, but both RNNs fail to predict the trend in Montana. This indicates that the inaccuracies were not caused by the limitations of the model, but instead caused by the differences in states that the model could not characterize given the input.

Generally, the model did not predict states that had less than 10 deaths over the 7 days. These states include Alaska, Hawaii, Montana, and Wyoming. We conjecture that this is because small errors on data have amplified effect on these states as they have small numbers of deaths.

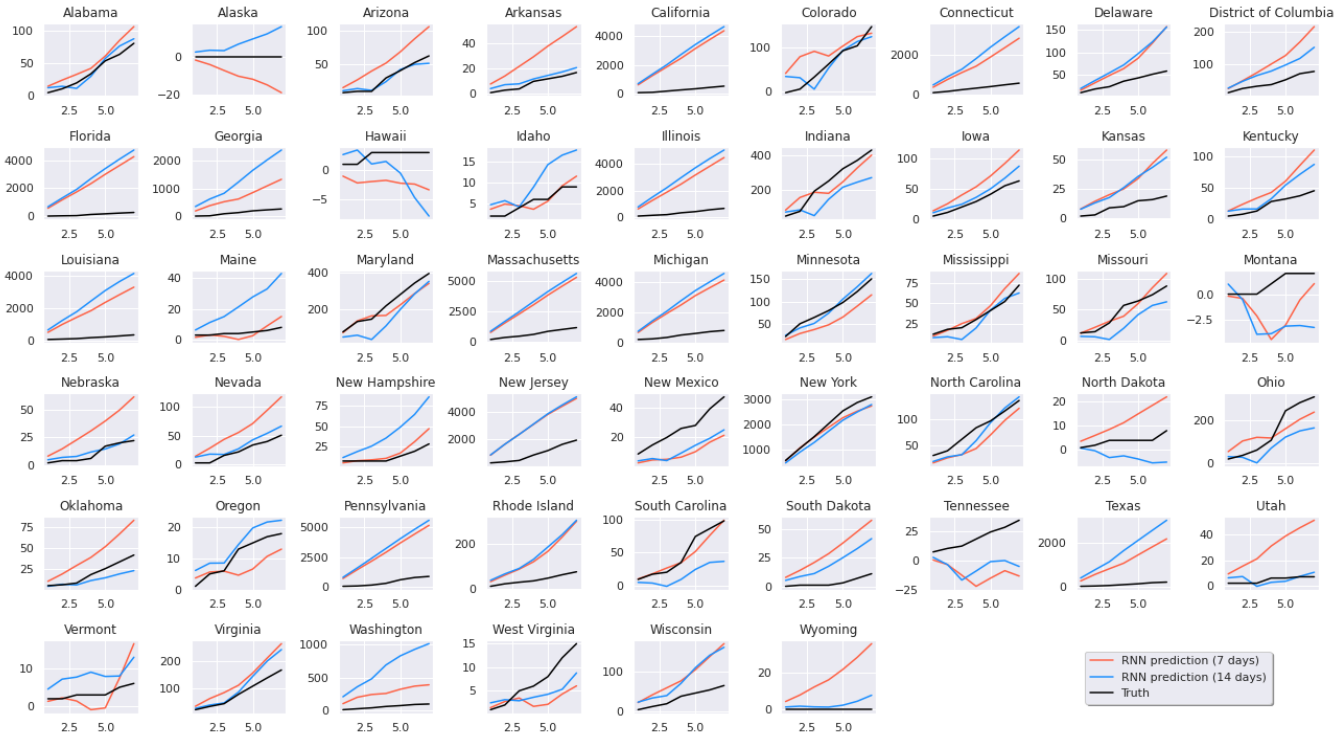

Figure 22. The predictions of the RNNs trained in 7-day prediction task (red) and 14-day prediction task (blue). The y-axis is the cumulative deaths since April 4th. The true trend is also displayed in black.

6 Conclusion and Future Work

In our experiments, we verified that recurrent neural networks perform better than fully connected neural networks given COVID-19 time-series data and found that it is likely because the recurrent neural networks can utilize social measures and government policies. We also found that both networks trained for 7-day prediction and 14-day prediction both had similar predictions, hinting that the prediction errors were a result of state characteristics not present in data.

One promising method to increase the accuracy of the neural networks would be to gather more data to include more features. These features could be time-series data that describe social measures and government policies, or they could be non-time-series data that describe the state's innate properties. Some good candidates for additional features are the daily mask consumption or the daily temperature.

Another avenue to explore is using different neural network architectures. There are many different ways to incorporate fully connected neural networks and recurrent networks. One promising method would be to use an ensemble of multiple neural networks to predict the daily deaths, as ensemble methods have proven to be successful in increasing prediction accuracy.

Finally, a deeper analysis of the differences in predicted and true trends for each state could result in surprising conclusions. The model's accuracy varies widely between states, and it is unclear what causes such discrepancies. Categorizing the states by accuracies and labeling the categories could help us discover new state characteristics that influence the trend of COVID-19.

References

Emily Alsentzer, John R. Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew B. A. McDermott. Publicly available clinical bert embeddings, 2019.

Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text. In EMNLP. Association for Computational Linguistics, 2019. URL https://www.aclweb.org/anthology/D19-1371 (https://www.aclweb.org/anthology/D19-1371).

Leo Breiman. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324 (https://doi.org/10.1023/A:1010933404324)

Lulu Yilun Chen. Heartbreak in the streets of wuhan, Mar 2020. URL https://www.bloomberg.com/news/articles/2020-03-12/heartbreak-in-the-streets-of-wuhan (https://www.bloomberg.com/news/articles/2020-03-12/heartbreak-in-the-streets-of-wuhan).

J.S. Cobb and M.A. Seale. Examining the effect of social distancing on the compound growth rate of sars-cov-2 at thecounty level (united states) using statistical analyses and a random forest machine learning model. Public Health, 2020. ISSN 0033-3506. doi: https://doi.org/10.1016/j.puhe.2020.04.016 (https://doi.org/10.1016/j.puhe.2020.04.016). URL http://www.sciencedirect.com/science/article/pii/S0033350620301219 (http://www.sciencedirect.com/science/article/pii/S0033350620301219)

Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. arXiv 2003.11597, 2020. URL https://github.com/ieee8023/covid-chestxray-dataset (https://github.com/ieee8023/covid-chestxray-dataset).

CORD-19. Covid-19 open research dataset (cord-19). https://pages.semanticscholar.org/coronavirus-research (https://pages.semanticscholar.org/coronavirus-research), 2020

The COVID Tracking Project "The COVID Tracking Project" https://covidtracking.com/ (https://covidtracking.com/) Accessed: 2 May 2020.

Raj Dandekar and George Barbastathis. Quantifying the effect of quarantine control in covid-19 infectious spread using machine learning. medRxiv, 2020. doi: 10.1101/2020.04.03.20052084. URL https://www.medrxiv.org/content/early/2020/04/06/2020.04.03.20052084 (https://www.medrxiv.org/content/early/2020/04/06/2020.04.03.20052084).

Trisha Greenhalgh and Jeremy Howard "Masks for all? The science says yes." https://www.fast.ai/2020/04/13/masks-summary/ (https://www.fast.ai/2020/04/13/masks-summary/) Accessed: 12 May 2020.

Google, "Dataset publishing language | states.csv." URL https://developers.google.com/public-data/docs/canonical/states_csv (https://developers.google.com/public-data/docs/canonical/states_csv), 2012.

Google LLC "Google COVID-19 Community Mobility Reports." https://www.google.com/covid19/mobility/ (https://www.google.com/covid19/mobility/) Accessed: 2 May, 2020.

Leigh Hopper. Early antibody testing suggests covid-19 infections in l.a. county greatly exceed documented cases, Apr 2020. URL https://news.usc.edu/168987/antibody-testing-results-covid-19-infections-los-angeles-county/ (https://news.usc.edu/168987/antibody-testing-results-covid-19-infections-los-angeles-county/).

Kaiser Family Foundation. 2020. State Data And Policy Actions To Address Coronavirus. https://www.kff.org/health-costs/issue-brief/state-data-and-policy-actions-to-address-coronavirus/ (https://www.kff.org/health-costs/issue-brief/state-data-and-policy-actions-to-address-coronavirus/) Accessed 2 May 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

W.W. Koczkodaj, M.A. Mansournia, W. Pedrycz, A. Wolny-Dominiak, P.F. Zabrodskii, D. Strzaška, T. Armstrong, A.H. Zolfaghari, M. Debski, and J. Mazurek. 1,000,000 cases of covid-19 outside of china: The date predicted by a simple heuristic. Global Epidemiology, page 100023, 2020. ISSN 2590-1133. doi: https://doi.org/10.1016/j.gloepi.2020.100023 (https://doi.org/10.1016/j.gloepi.2020.100023). URL http://www.sciencedirect.com/science/article/pii/S2590113320300079 (http://www.sciencedirect.com/science/article/pii/S2590113320300079).

Sarah Mervosh, Denise Lu and Vanessa Swales. See Which States and Cities Have Told Residents to Stay at Home, Apr 2020. URL https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html (https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html).

Sarah Mervosh, Jasmine C. Lee, Lazaro Gamio, and Nadja Popovich. See which states are reopening and which are still shut down, May 2020. URL https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html (https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html).

Vinod Nair, and Geoffrey Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of ICML. 27. 807-814, 2010.

The New York Times "Coronavirus (Covid-19) Data in the United States" https://github.com/nytimes/covid-19-data (https://github.com/nytimes/covid-19-data) Accessed: 2 May 2020.

Ratnabali Pal, Arif Ahmed Sekh, Samarjit Kar, and Dilip K. Prasad. Neural network based country wise risk prediction of covid-19, 2020.

Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Huiying Zhao, Yusheng Jie, Ruixuan Wang, Yutian Chong, Jun Shen, Yunfei Zha, and Yuedong Yang. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. medRxiv, 2020. doi: 10.1101/2020.02.23.20026930. URL https://www.medrxiv.org/content/early/2020/02/25/2020.02.23.20026930 (https://www.medrxiv.org/content/early/2020/02/25/2020.02.23.20026930).

Nicholas Soures, David Chambers, Zachariah Carmichael, Anurag Daram, Dimpy P. Shah, Kal Clark, Lloyd Potter, and Dhireesha Kudithipudi. Sirnet: Understanding social distancing measures with hybrid neural network model for covid-19 infectious spread, 2020.

US Census Bureau. State area measurements and internal point coordinates, Aug 2018. URL https://www.census.gov/geographies/reference-files/2010/geo/state-area.html (https://www.census.gov/geographies/reference-files/2010/geo/state-area.html).

US Census Bureau. State population totals: 2010-2019, December 2019. URL https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html (https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html).

World Health Organization. Coronavirus disease 2019 (covid-19): situation report, 41. Technical documents, 2020a

World Health Organization. Coronavirus disease 2019 (covid-19): situation report, 102. Technical documents, 2020b

Jin Wu, Allison Mccann, Josh Katz, and Elian Peltier. 63,000 missing deaths: Tracking the true toll of the coro-navirus outbreak, Apr 2020. URL https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html (https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html).

Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Yanfei Chen, Junwei Su, Guanjing Lang, Yongtao Li, Hong Zhao, Kaijin Xu, Lingxiang Ruan, and Wei Wu. Deep learning system to screen coronavirus disease 2019 pneumonia, 2020.

Mark Zastrow "South Korea is reporting intimate details of COVID-19 cases: has it helped?" https://www.nature.com/articles/d41586-020-00740-y (https://www.nature.com/articles/d41586-020-00740-y)