Team 1: J. Polk, V. Nguyen, M. Heningburg, B. Lober

B. Bushong

Final Project Memo

29 April 2020

Chasing the Coronavirus

The novel SARS-CoV-2 coronavirus emerged from the city of Wuhan China within 2019, giving it the nickname “COVID-19.” This virus can be readily spread via human-to-human contact, leading to a large spread in the disease and the outbreak of a global pandemic. Our analysis will focus on the initial spread of COVID-19 over time from the first case confirmations in January up to the first week of April. We will examine and compare the number of case confirmations, deaths, and recoveries in the countries with the most case confirmations as of April: The United States, Spain, Italy, France, and Germany. China will be included in our analysis as well due to it being the location of the initial outbreak and current doubts about their recent case confirmation numbers.

Our data visualizations will track confirmed COVID-19 cases, deaths, and recoveries over time and create predictions for the level of virus confirmations and recoveries in the upcoming future. To do this, our data was halted on Tuesday, April 14th, 2020 to compare the accuracy of our prediction after a week (ending on Tuesday, April 21st, 2020). Our complete data was obtained from an online data science community (Kaggle), which relays the case number, observation date, province/state confirmed, country/region confirmed, number confirmed, number of deaths, and number of recoveries. After we make our predictions, we will compare them to the actual outcome spread in the U.S. over the week. Additionally, we will look at specific U.S. foreign policy established by the U.S. Government as a whole, to see if these policies have any effect on the probability of spread of COVID-19. With these prediction tools, we anticipate to better understand the transmission of COVID-19 and have the tools to predict the time course of future epidemics.

To start, we wanted to create some visualizations to show not only the location of the highest case confirmations, but how each country compared in terms of the number of confirmations. We used the coronavirus tracker to focus on the top countries that had reported the most confirmations of the virus. We also wanted to include China since it was the origin of this virus. Once we collected this data, we wanted to import ggplot libraries to plot this data and also import the maps, mapprojec, and maptools libraries so we are able to display each country and its corresponding confirmations. After this, world map data was imported so we are able to display and look at information like different locations around the world and a plot of the world map. Since we were only focusing on the top countries, we had to subset the data, only including the countries used in the project. After this, we had to create a dataset that showed the number of confirmations by country so it can show on the plot. Because we only needed six countries to visualize, we created a subset within the map data to only show the countries we were looking at, which was the United States, Spain, Germany, France, Italy, and China, Once this was done, we used ggplot to plot the project data and used different functions within ggplot to show the number of confirmations by lightness or darkness of color, so we are able to display how other countries look in comparison to another. The darker the color, the more confirmations the country had. We also changed the legend scale to represent the number of confirmations in thousands, so that the legend is spaced out and is easier to read (figure 1). Looking at the data(figure2), the United States had the highest amount of case confirmations, and had about 415,000 more confirmed cases than the second leading country in terms of confirmations, which was Spain. China had the lowest amount of confirmations, but reports have questioned China's honesty in reporting the right amount of cases, so this could have affected the data in some way. It is interesting to see how when the virus was first being reported on, the United States was not even a factor, but now has the largest amount of confirmations.

After mapping the intensity of the pandemic up to April 14th, we began our prediction model for the United States. In doing so, we developed a compartmental model (a mathematical model used to describe systems and interactions) based on the SIR model with three compartments (susceptible, infected, and recovered). Susceptibles are healthy individuals that could become infected. The infected group are those that ideally have received the virus whether symptomatic or not. The recovered group represents those formerly infected but now not infected (aka recovered). In this model we used a system of equations (figure 3). As initial conditions, we calculated the recovery, confirmation, and death rates for the United States (figure 4). As of April 14th, the recovery rate was 5.13%, the mortality rate was 3.18%, and about 2.1% of the U.S. had coronavirus confirmed. Further initial conditions included the U.S. population on April 14th (329515369). Overall, in our SIR model, we predicted a small, yet steady increase in the recovered population and small drop in the infected population if there was no social distancing. However, the United States enacted full lockdown and social distancing precautions to prevent the spread of COVID-19. Thus, with social distancing precautions we predicted a larger decline in the infected population and a jump in the recovered population, with less people susceptible (figure 6). Clearly, the usage of social distancing has justified reasons, as it prevents any further confirmations and would allow the United States to continue with normal life.

In comparing our prediction to the results by date April 21st, 2020, we found that the number of case confirmations actually increased dramatically (figure 9) as opposed to our initial assumption of a steady decline in infected people with social distancing. However, our prediction for an increase in recovery rate was accurate. The difference is between the rate at which the U.S. initial infected population actually recovered was fairly large (figure 7), when our prediction was small on the assumption that U.S. citizens couldn’t actually social distance properly and that the U.S. healthcare system wasn’t where it needed to be at the time. And although we didn’t predict the number of deaths, there was a large increase in the number of U.S. deaths within the week, yet not nearly as many recoveries. Overall, within the week, the percent of U.S. citizens who had coronavirus increased to 3.6%. The recovery rate also increased to 6.7% and the mortality rate increased to 4.0%. In consideration of other top countries at the start of this prediction analysis, Spain had a 31.4% recovery rate and 9.6% mortality rate, Italy had a 19.0% recovery rate and 11.9% mortality rate, France had a 20.2% recovery rate and 10.3% mortality rate, and Germany had a 40.1% recovery rate and 2.1% mortality rate. Finally, for China, where it all began, they had a 67.9% recover rate and mortality rate of 3.8%. When considering these other top countries, the United States clearly is on the bottom in terms of the amount of people who are contracting and recovering. In terms of those who are dying in the United States, we actually seem to be on the lower end (after Germany and China), despite what reports may say.

In clear error analysis, we acknowledge our rate of infection and rate of recovery could be perfected to better estimate the future of coronavirus in the United States. In the future, we'd also like to take into more consideration the consistent variations of policy and the individual actions of U.S. Governors within each state. Additionally, we believe that the consideration of healthcare faults weren't considered enough in the increase in the number of deaths. Lack of staff, health equipment, and general care from those within and outside the health system all play a clear factor in properly monitoring and predicting the spread of COVID-19. It is clear, however, that the practice of social distancing is, in fact, beneficial and needed, despite the economic setbacks. Other preventative measures and policies explored in former top countries that are now doing much better should also be a consideration, despite apparent cultural differences that may make those in the U.S. reluctant to adopt into everyday life.

After careful analysis of the spread of COVID-19, it can be assumed that this pandemic is far from being over. Our predictions on the spread of the virus in the U.S. were partially incorrect: we predicted a decline in case confirmations and increase in recoveries, and in reality there were increases in both case confirmations and recoveries. As the amount of confirmed global cases starts to exceed 3 million, this demonstrates that the virus will perhaps continue to be transmitted beyond our expectations.

As the United States surpasses 1 million cases and becomes the top country with the most COVID-19 cases, future policies and guidelines are up for debate. The initial 30-day travel ban on most of Europe (excluding Ireland and the UK) may have reduced spread, but it is also notable that this ban did not apply to U.S. citizens or permanent residents, allowing them to possibly return to the U.S. as carriers of the virus. Furthermore, the U.S. has been slightly more unique in that stay-at-home orders implemented by states have occurred intermittently at different times rather than on a national scale. With the reopening of businesses in states such as Georgia that have not flattened their curve in cases, this could bear the risk of creating another wave of outbreaks. While some states have implemented fines and punitive measures for social gatherings, it is questionable if these measures will effectively prevent transmission. Ultimately, the future of COVID-19 in the U.S. will rely heavily on voluntary compliance during this public health emergency.

Data Figures

Figure 1

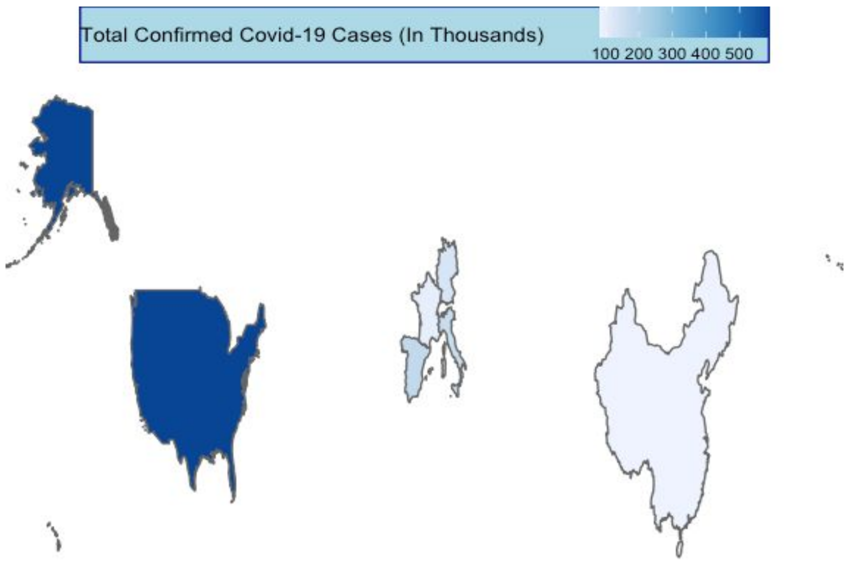


Figure 2

Country	Total Confirmed Covid-19 Cases (In thousands; As of April 14th)
United States	587.357
Spain	172.655
Germany	130.400
Italy	159.516
France	98.076
China	82.249

Figure 3

Basic SIR Model:

$$\frac{dS}{dt} = -bSI,$$

$$\frac{dI}{dt} = +bSI - cI,$$

$$\frac{dR}{dt} = +cI$$

Figure 4

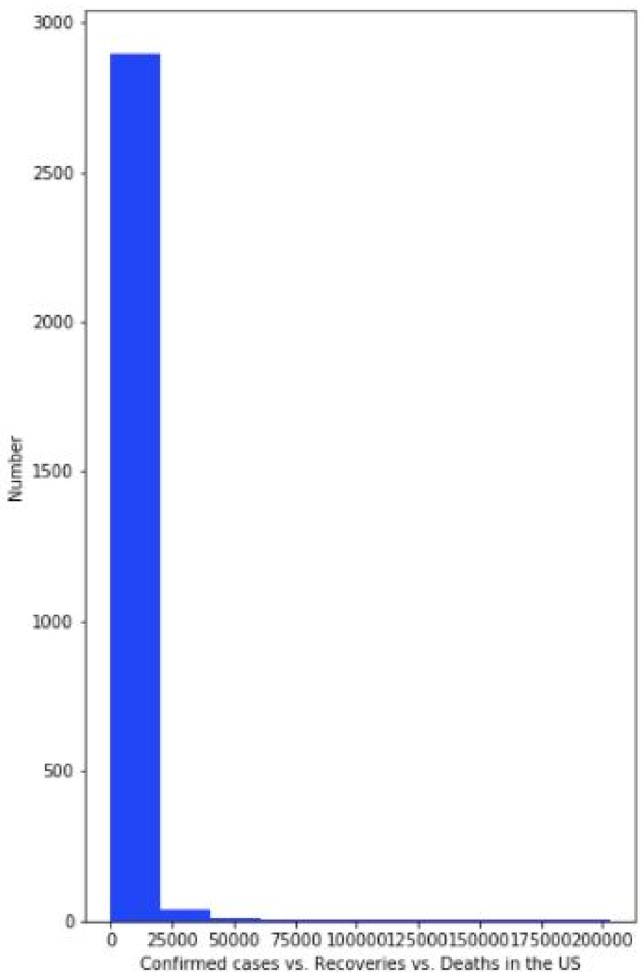


Figure 5

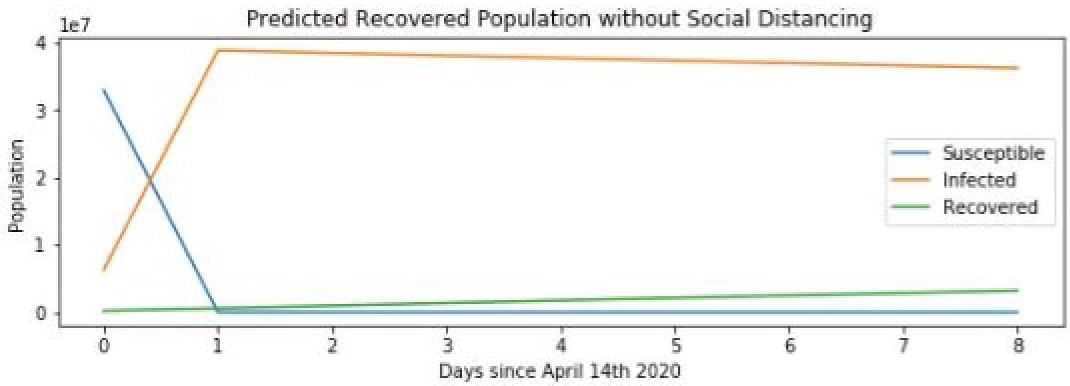


Figure 6

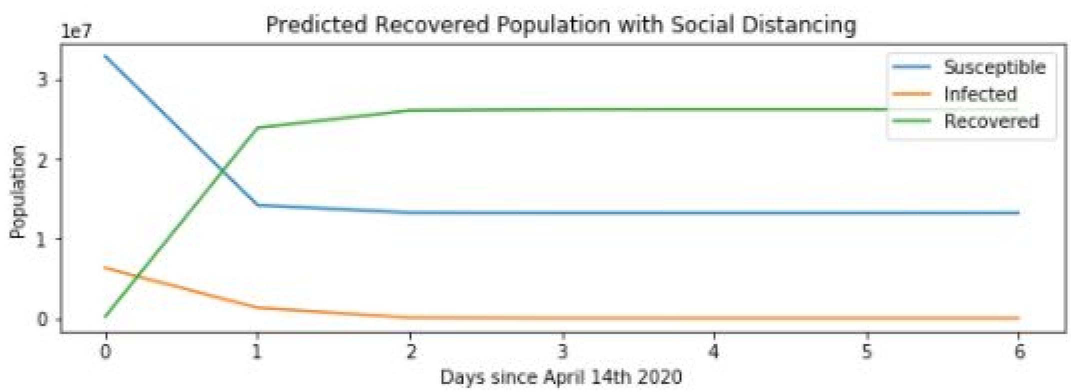


Figure 7

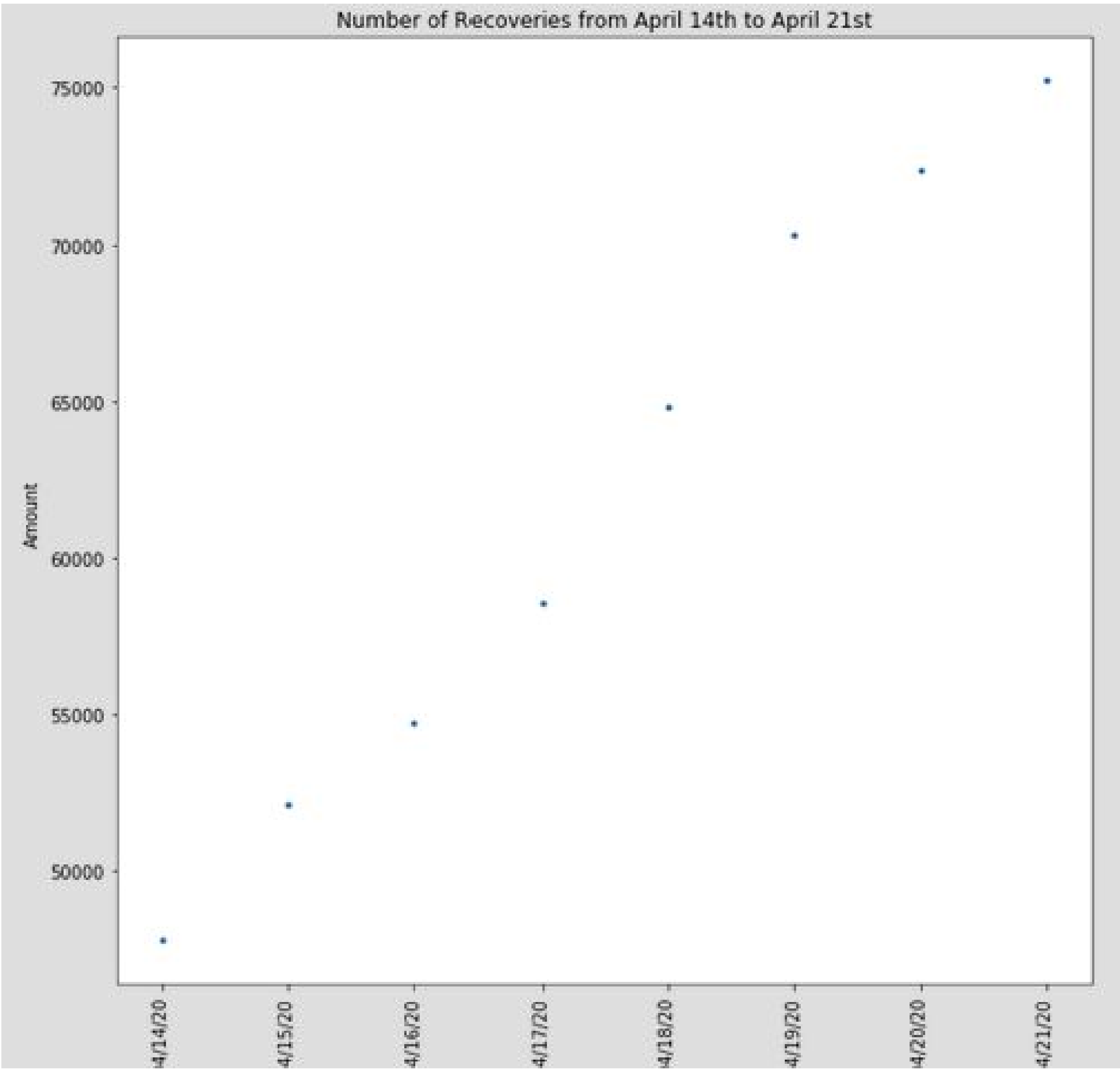

Number of Recoveries from April 14th to April 21st
Amount
75000
70000
65000
60000
55000
50000
4/14/20
4/15/20
4/16/20
4/17/20
4/18/20
4/19/20
4/20/20
4/21/20

Figure 8

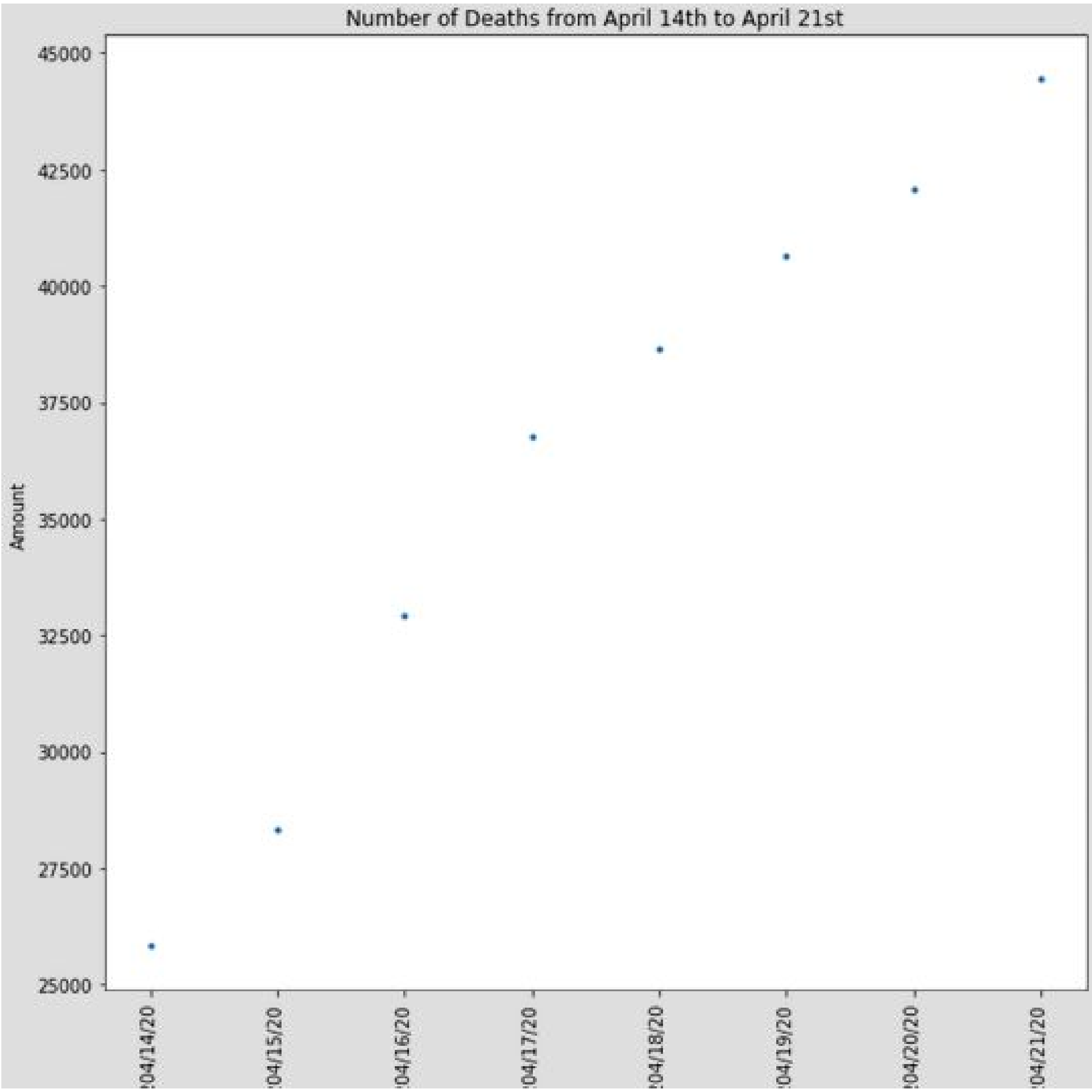
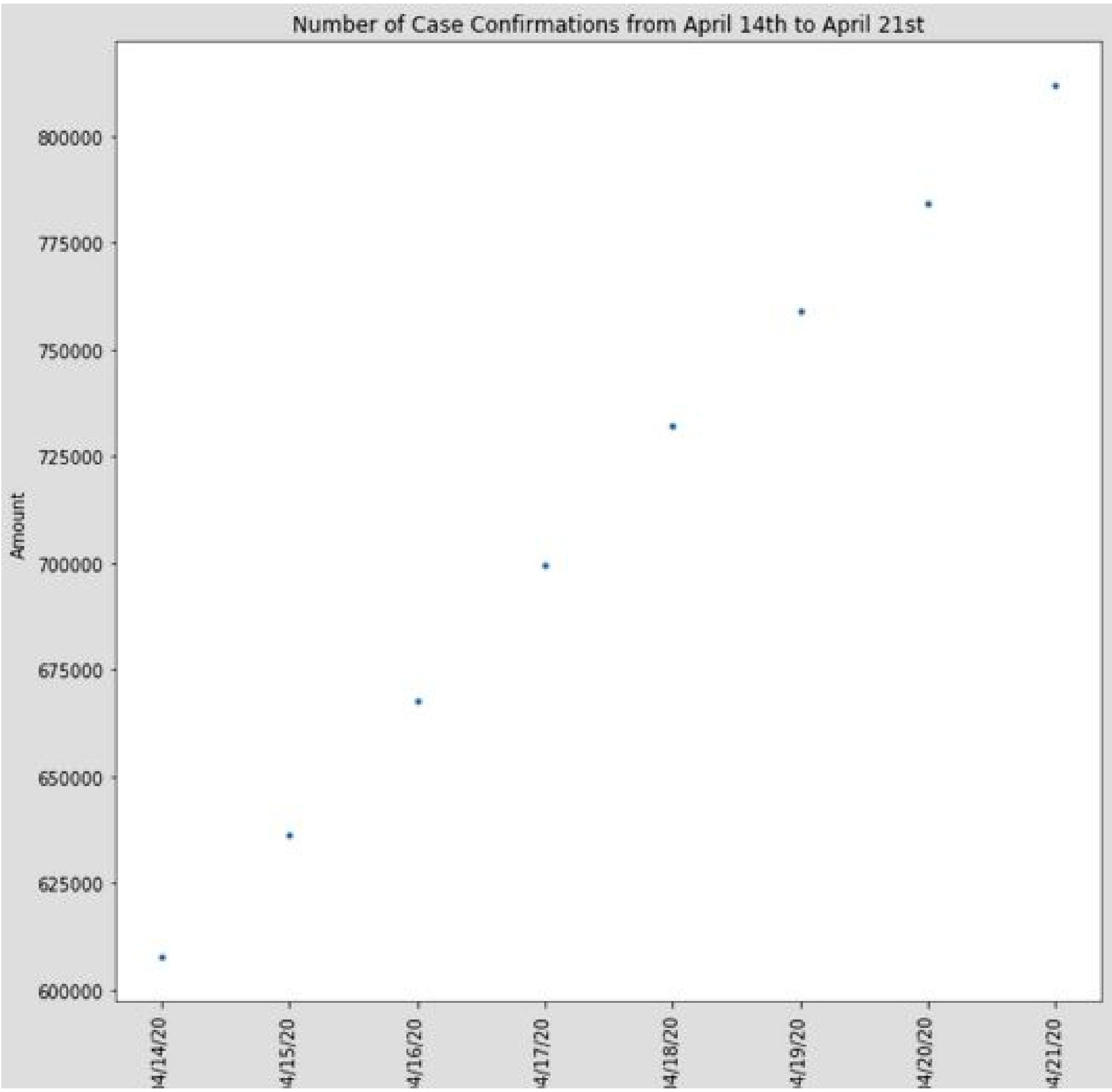

Number of Deaths from April 14th to April 21st
Amount
45000
42500
40000
37500
35000
32500
30000
27500
25000
2004/14/20
2004/15/20
2004/16/20
2004/17/20
2004/18/20
2004/19/20
2004/20/20
2004/21/20

Figure 9


Number of Case Confirmations from April 14th to April 21st
Amount
800000
775000
750000
725000
700000
675000
650000
625000
600000
4/14/20
4/15/20
4/16/20
4/17/20
4/18/20
4/19/20
4/20/20
4/21/20

References

Chen, Yuxuan. "A Novel SEIR Model Based on the Data of Quarantined, Suspected, Hospitalized, Identified and Recovered Cases to Forecast the Trend of 2019-nCov Epidemics After February 12, 2020," March 12, 2020, 1–26.

"Coronavirus (COVID-19)." Google News, Google, 2020, news.google.com/covid19/map?hl=en-US&gl=US&ceid=US%3Aen.

"Novel Corona Virus 2019 Dataset." *Kaggle*, Kaggle, 2020, www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset.

"U.S. and World Population Clock." *Population Clock*, U.S. Government Census, 2020, www.census.gov/popclock/.