

IOWA STATE UNIVERSITY

DEPARTMENT OF DATA SCIENCE

THE EFFECTIVENESS OF STAY HOME ORDERS TO REDUCE COVID-19 PREVALENCE

Anthony Song
Michael Egle
John Chandara
Jaydon Cobb

May 5, 2020

Abstract

The purpose of this investigation is to provide insight into how the citizens of the United States and the world responded to social distancing regulations, and effective how these rules have been. We also contrast the United States' situation with other countries and elucidate trends that are produced by countries successfully suppressing the spread and deadliness of COVID-19. Using the mobility data provided by Apple and various global immunology statistics we were able to develop answers for the questions above.

1 Background

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Common symptoms include fever, cough, fatigue, shortness of breath and loss of smell. The disease was first identified in December 2019 in Wuhan, the capital of China’s Hubei province, and has since spread globally, resulting in the ongoing 2019–20 coronavirus pandemic. The first case reported outside of China occurred in Thailand on January 13th, 2020, which spread from a woman traveling from the epicenter. As of April 29th, 2020, more than 3.17 million cases have been reported across 185 countries and territories, resulting in more than 224,000 deaths. In mid-March of 2020, the COVID-19 disease began to spread like wildfire throughout the United States. As for the federal response to this outbreak, the cabinet of the president delegated all response and protection duties to the states’ governors.

1.1 Obtaining Our Data

We obtained our data from three primary sources, The Apple Mobility Dataset[1], New York Times COVID-19 Dataset[3], and the Worldometer Dataset[4]. The steps below highlight how we obtained and cleaned each dataset.

1.2 Apple Mobility Data

1. Obtain the dataset.
2. Pivot dataset to increase rows by creating a column containing the dates.
3. Change the date to year-month-date format using `lubridate`.

1.3 New York Times COVID-19 Dataset

1. Read in the data directly from the GitHub URL.
2. Use the `lubridate` package to change the date column to type `Date` from character.
3. Break the data into four subdivisions for the four cities of interest. The names of these four new data frames are `nyc`, `mia`, `chi`, and `sf`.

1.4 Worldometer Dataset

1. Obtain the dataset.
2. The data is composed of death per million people Worldometer data for various countries on April 24th, 2020. Values were chosen after a preliminary comparison using Microsoft Excel.
3. Rate of decrease in mobility column data was calculated using a `stat_smooth()` regression line. The rate calculated from the local max to the local minimum.
4. Apply first row as column names, then convert characters to numeric values. Take absolute values for plotting purposes.

2 Our Exploration

2.1 Hypothesis

We wanted to look at the effects of staying inside on the spread of the virus and as a result, the deaths caused by the virus. In theory, if people stay home more, there will be fewer interactions and less chances to spread the virus. Our group made the decision to use deaths as our statistic to show the spread of the virus rather than cases. We did this for a variety of reasons. First, this virus is much worse than a common flu and businesses would not be shutting down entirely if not for the tens of thousands of deaths across the country. Second, some regions are testing their citizens much more regularly than others. Third, it's not uncommon for people to have the virus and never show symptoms at all. We attempted to simultaneously visualize the change in mobility in our data along with the number of new deaths reported each day.

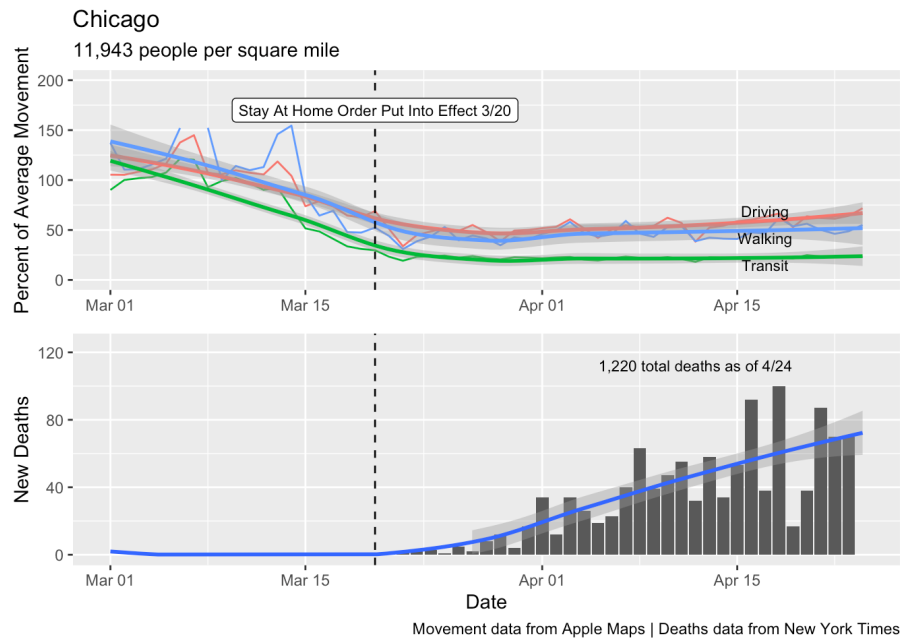


Figure 1: Change in mobility plotted with number of daily new deaths for Chicago, IL

2.2 Case Study of Chicago, IL

As an example, Chicago's data is shown above in *Figure 1*. The y-axis on the top half is the percentage of movement compared to an average day in the given region. As shown, there was a significant decrease in movement even before the stay at home order was officially put in place. However the numbers have slowly crept back towards the average, which may help to explain why the curve in Chicago still has not flattened.

2.3 Case Study of Other Regions

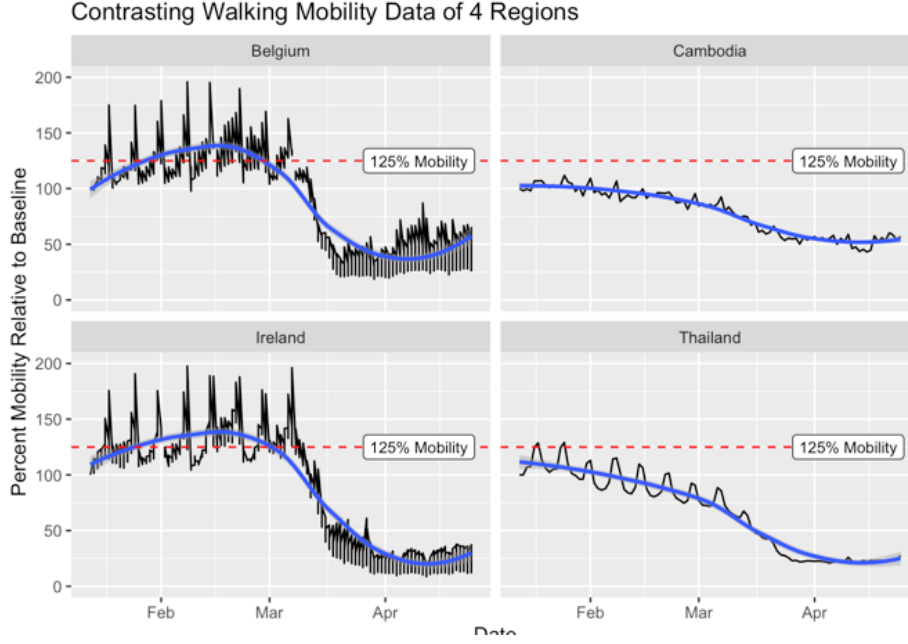


Figure 2: Comparison of mobility data for low death rate and high death rate regions

After looking at the US analysis, We wanted to further explore the mobility data and look for trends amongst other countries and regions. Similar to the first analysis, we started exploring the death data in order to reduce confounding variables, like individuals that were hypochondriacs (people that misdiagnose minor symptoms), that might be associated with the reported number of COVID-19 cases. After viewing many countries in the upper and lower median of the Worldometer data set for deaths per million people, we extracted five extreme values for low death rate (LDR) and high death rate (HDR) countries. Due to the mysterious nature of this virus, we thought it would be best to start looking for trends in the upper (Belgium/Ireland, *Figure 2*) and lower (Cambodia/Thailand, *Figure 2*) extremities, this would illuminate any potential correlations that may be unclear in countries with death rates around the median. We wanted to specifically target only the walking data, as this would provide a better representation of social interactions and potential interactions that occur within a six foot radius of

another human. After plotting the relative walking mobility as a function of the date for each extracted country, we stumbled upon a trend that grouped the countries by their respective death rates as shown below.

Using `geom_smooth` to statistically create a smooth trendline, we found that all of our LDR and HDR countries could be represented by either a shallow or steep sine curve for LDR and HDR respectively. We found the best distinguishing factor to quantify these inherently different functions would be to measure the decreasing rate of change found at the mid-February to April time frame. The rate of change was determined by the data points produced by the `geom_smooth` regression line. We then measured the rate of change from the local maximum to the local minimum. We decided to look at the maximum and minimum of the regression line instead of the raw data because this would overlook any extreme values produced from highly mobile dates. Doing this for every extracted country, we obtained a data set that looked like *Table 1*. (ddp)

Country	Deaths per Million	Change in Mobility	Below 125%
Indonesia	3	-1.4512530	1
Cambodia	0	-0.7241492	1
Korea	5	-1.5005610	1
Taiwan	0.3	-0.5122063	1
Thailand	0.71	-1.2754960	1
Italy	426	-2.6000790	0
United Kingdom	299	-2.1510730	0
Spain	490	-3.5016930	0
Belgium	597	-2.7629390	0
Ireland	215	-2.6594180	0

Table 1: Data Set with the country, death rate, decreasing mobility rate and below 125% mobility, which is referred to as ddp.

2.4 Statistical Implications

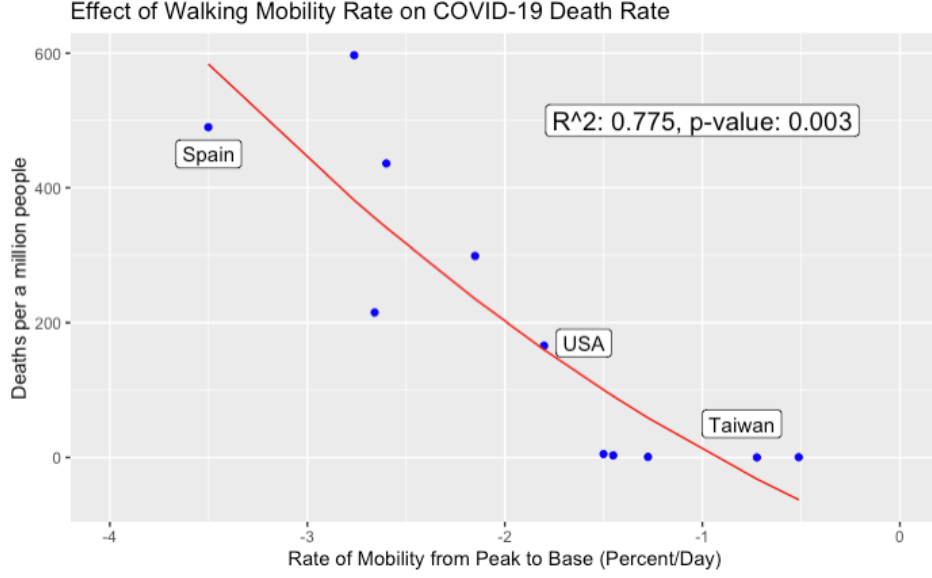


Figure 3: Regression model of death million people as a function of decreasing mobility

While exploring the rate of change we noticed another trend that was intrinsic to LDR and HDR countries, the frequency of mobility percentages above 125%. Intuitively, high frequency of dates with high mobility is bound to spread bacteria quickly, but we did not continue this exploration because after preliminary tests, we found that these values did not show much promise in correlation. From *Table 1*, we plotted the deaths per million as a function of decreasing rate in mobility for all countries. The graph seemed to show a polynomial trend and after some regression fitting using the `lm()` function, we were able to obtain a model, as shown in *Figure 3*, with a coefficient of determination (r^2) of 0.775 and p-value of 0.003. This showed a promising trend that would help us answer our questions.

3 Data Wrangling and Visualization

Important Notes

1. Most data is prepared and plotted using dplyr pipeline (% > %).
2. Annotations on visualizations are prepared using annotate() function.

3.1 Preparation

1. Change the dates to rows so we can plot the data efficiently.
2. Change the date values to year/month/date to increase accessibility.
3. Use the filter function to extract specific walking and country data, then bind the specific sets into a new data set.

3.2 Data Visualization Preparation

In order to show how the mobility data and new deaths per day data increased or decreased together over time, we put the plots on top of one another as seen above in *Figure 1*.

1. Preparation of the top half of *Figure 1*: The line plot represents the three types of mobility data that was provided by Apple: walking, driving, and transit. We separated these due to the fact that low levels of walking and transit use are more important than use of driving.

```
chi_travel_viz <- chi_travel %>%  
  ggplot(  
    aes(  
      x = date,  
      y = percent_change,  
      group = transportation_type,  
      color = transportation_type  
    )  
  ) +  
  geom_line() +  
  geom_smooth()
```

2. Preparation of the bottom half of Figure 1: Histogram of each day's newly reported deaths with a trend line to account for possible corrections in the data.

```
chi_deaths_viz <- chi %>%  
  ggplot(  
    aes(  
      x = date,  
      y = new_deaths  
    )  
  ) +  
  geom_histogram(  
    stat = "identity"  
  ) +  
  geom_smooth()
```

3. Combination of plots: We utilized the ggpubr package for its **ggarrange** function, giving the user the ability to combine two plots on top of one another. In doing this, we now have a plot with corresponding x-axes for the top and bottom.

```
ggarrange(  
  chi_travel_viz,  
  chi_deaths_viz,  
  nrow = 2  
)
```

3.3 Data Visualization (Figure 2 & 3)

1. Preparation of *Figure 2*: Line graph of multiple countries mobility data with a regression line using `facet_wrap` (*figure 2*)

```
qplot(date, percent_change, date = indo)
  + stat_smooth(aes(
    outfit = indo_tha<<-..y..
  ))
```

2. Rate of change graph and calculation for each country.

```
indo_tha <- data.frame(indo_tha)

fit_indo$day = seq.int(nrow(fit_indo))

vc_indo <- c(max(
  fit_indo$fit_indo),
  min(fit_indo$fit_indo)
)

Max_min_indo <-
  fit_indo[fit_indo$fit_indo %in% vc_indo, ]

Rate_of_Change <-
  (max_min_indo[2,1] - max_min_indo[1,1]) /
  (max_min_indo[2,2] - max_min_indo[1,2])
```

3. Preparation of *Figure 3*: Death rate of country as a function of mobility rate with a polynomial regression, r^2 and p-value statistics.

```
regression = lm(
  dpp$'Deaths per a million people' ~
  poly(
    dpp$'Rate of decrease in Mobility',
    2,
    raw = TRUE
  ),
  data = dpp
)

pred_dpp <- data.frame(
  death = predict(regressor, dpp),
  rate = dpp$'Rate of decrease in Mobility'
)

summary(regression) # r^2 and p-value

ggplot(
  data = dpp,
  aes(
    y = 'Death per a million people',
    x = 'Rate of decrease in Mobility'
  )) +
  geom_line(
    color = 'red',
    data = pred_dpp,
    aes(x = rate, y = death)
  )
)
```

4 Discussion

Using the Apple Mobility data, New York Times COVID-19 GitHub data, and Worldometer data we strove to answer four important questions: Have citizens obeyed social distancing regulations? Has it been effective? How are we (United States) handling this compared to other regions? What can we learn from regions that are successfully keeping COVID-19 prevalence low?

4.1 Remarks on America's Response

Starting off, we wanted to take a look into how the citizens of the United States are coping with the new regulations on social interaction. In the scope of this project, we answered this by looking at four large cities: New York City, Chicago, Miami and San Francisco. For these four cities, we can confirm that a vast majority of these citizens are obeying the government implemented social distancing regulations. This can be qualitatively and quantitatively observed by an intense decrease in walking, driving and transit mobility for the four cities. Furthermore, by incorporating the death toll, population density and the “Stay at Home Order” implementation date, we can draw conclusions on the regulation’s effectiveness. Based on these four cities, we can see that San Francisco had the most effective implementation of social distancing regulations. With the second highest population density, San Francisco, as of April 24th, 2020, kept their death toll to 22, which is extremely low compared to New York City (11,157 deaths) and Chicago (1,220 deaths). Outside of the mobility data, a factor that may explain San Francisco’s effective regulations is the well-timed implementation date. Compared to the other cities, San Francisco mandated their regulations at least 4 days before the other cities. This shines light on the effectiveness of early action in a global pandemic.

4.2 Remarks on Other Responses

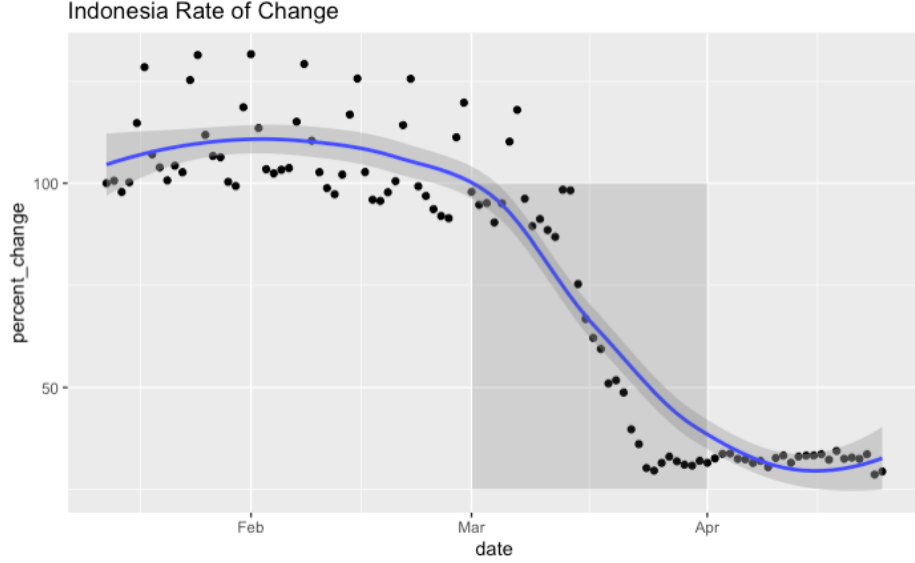


Figure 4: Graph of Indonesia, regression line points were extracted for calculations of rate of change (ROC). The shaded region represents the region in which ROC was extracted.

Broadening the scope of our regional analysis, we wanted to understand how the United States was battling this virus compared to the rest of the world. To answer this we looked at the relative death rate, percent mobility and decreasing rate of mobility. Qualitatively, we observed the trends of mobility spikes that occurred from mid-February to early-March. These spikes reached values greater than 150% the baseline mobility and follow similar trends of HDR regions like Ireland and Belgium shown in *Figure 2*. Next we looked at the deaths per million people, which range from 0 to 597 (*table1*), the United States is in the middle at 166. Finally, we looked at the calculated decreasing mobility rate, which ranges from -0.51 to -3.5 (*Table 1*) percent mobility per day, the United States is in the middle at -1.8 . These values indicate that the United States is battling COVID-19 moderately, meaning our practices of containment could be improved. Digging deeper, we wanted to understand how we could quarantine more effectively, by learning from countries successfully battling the virus. We looked specifically at the five countries located in the bottom right corner of

Figure 3. These countries showed low death rate, mobility under 125% of the baseline during mid-February to early-March and larger, and less-negative (shallow) rate of change. Our interpretation is that these countries were more inclined to react proactively at the beginning of this pandemic. These countries maintained their baseline mobility and slowly decreased it as the severity of the pandemic increased as seen by the mobility under 125% and decreasing mobility rate respectively. Furthermore, this correlated to a low death rate and a statistically significant regression line ($p - value = 0.003$). Although we have shown this to be significant, it is important to understand this model may fail when generalized for all countries that lie between the extreme values. Fortunately, in our investigation, countries that took early action by implementing laws and reducing mobility early, analogous to San Francisco, were able to effectively contain this virus by consistently keeping their mobility, and potentially their social interactions, below their baseline values.

5 Conclusion

Amidst the chaos caused by an unexpected global pandemic, researchers in all fields have been collaborating to understand the physiological, sociological and economic pathology of this unknown virus. To join this mysterious expedition, we wanted to look into new datasets that might uncover valuable information. Using the Apple mobility, New York Times COVID-19 and Worldometer data, we were able to unveil potential trends on how citizens have obeyed social distancing regulations, the effectiveness of these regulations and what we can learn from countries successfully battling this virus. Looking into four densely populated cities in the United States we were able to uncover the effectiveness of early action as displayed by San Francisco, a city that had a relatively high population density and an extremely low death toll. San Francisco shows, compared to Chicago, Miami and New York City, that the early implementation of the “Stay at Home” order may have decreased the relative mobility of the city’s residents, in turn decreasing the spread of the disease and ultimately, lowering the death toll. Zooming out to a global level we found there was a correlation between periodic spikes of high mobility percentage, large decreasing rates in mobility around the time of panic (March-April) and deaths per million people (dpp) for various regions. Cities with high death rates (Belgium: 597 dpp, Ireland: 215 dpp), seemed to have frequent spikes of high mobility (150 – 200%) at a crucial incubation period for the virus. These high mortalities ultimately lead to more severe decreases in mobility when the virus starts to exponentially infect the residents. From these features and trends, we can conclude that countries and cities that followed social distancing regulations and to early action, benefited by maintaining a low death rate. In future studies, it may be important to study confounding variables, like culture, diet and immunity, that could possibly strengthen or weaken our claims. In the scope of our investigation, we believe that the mobility data has uncovered novel trends that help us understand the epidemiology of this mysterious Coronavirus.

Addendum

This paper is dedicated to the hundreds of thousands of lives who have met an unfortunate demise to the horrible COVID-19 viral outbreak. May their souls rest in peace as we continue our efforts to fight this disease in an effort to save lives.

References

- [1] “COVID-19 - Mobility Trends Reports.” *Apple*. 3 Jan. 2020, www.apple.com/covid19/mobility.
- [2] Burkert, Andi. “Flattening the COVID-19 Curves.” *Scientific American Blog Network*, Scientific American, 17 Mar. 2020, blogs.scientificamerican.com/observations/flattening-the-covid-19-curves/.
- [3] The New York Times. “Coronavirus in the U.S.: Latest Map and Case Count.” *The New York Times*, The New York Times, 3 Mar. 2020, <https://github.com/nytimes/covid-19-data>.
- [4] “Coronavirus Cases:” *Worldometer*, www.worldometers.info/coronavirus/.