New York University

Tandon School of Engineering

CS 6513: Big Data

Project Final Report

The Effect of COVID-19 Government Policies on Social Distancing and Infection Numbers

Group Name: OneTeammateRanAway

Name: Haifeng Zhang netID: hz2149
Name: Tianyi Zhao netID: tz1330

Date: May 10 2020

I Introduction

In the COVID-19 pandemic, social distancing has been a major mechanism to prevent outbreaks and to reduce the risk of infection. U.S state governments have raised policies on social contacts such as school closure, restaurant limits and other stay-at-home policies. However, the public may take measures not immediately or before the governments publish orders. This project will study on the community mobility change to the safety policies.

The project will focus on the effects of governments' stay-home policies on social distancing in the United States, especially in the New York State. We will use the public mobility data by Google and an open source government policy data. Furthermore, we will combine the infection numbers to our results and evaluate how effective the policies are in reducing COVID-19 spread.

II Objectives and Overview

This project has two main problems:

- How community mobility changes in response to the government stay-at-home orders in the United States.

- How effective is social distancing in reducing COVID-19 spread.

For the first problem, we will use community mobility data by Google to find the mobility percentage changes per day in March and April of each state in the US. And we can extract policy data of each state governments so we can match the time point to the mobility curve. We will compare the curve before and after the publish time of policies. Also, we may compare it to the same time period in 2019 or the previous weeks.

The results will be extended to the COVID-19 infection data. We will first apply a simple exponential model to the state infection curve to figure out whether the government measures reduce the increase rate. There might not be a direct conclusion since the infection number can be influenced by multiple factors. We will try to use some prediction models raised by researchers to compare the patterns of infection curve before and after governments' stay-at-home orders.

III Data

The project will use 3 open source data sets:

Data set name	Link	Description
Johns Hopkins dashboard data	https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data	Daily reports for new cases by Johns Hopkins CSSE, including daily case increase in each county in the United States
COVID-19 Community Mobility Reports	https://www.gstatic.com/covid19/mobility/Global_Mobility_Report.csv	These Community Mobility Reports aim to provide insights into what has changed in response to policies aimed at combating COVID-19. The reports chart movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.
Government imposed safety measures	https://data.humdata.org/dataset/acaps-covid19-government-measures-dataset	4000-row dataset that compiles official public safety measures such as social distancing, school closures, etc. imposed by countries around the world. Includes the dates when these measures were imposed and the official or news source. Compiled by Assessment Capacities Project (ACAPS).

IV Data Cleaning & Integration

The details are included in datasets-used.csv file in our project part2 github repository.

IV.1 Mobility Data

This data set is nicely organized so we did not need to do much cleaning. We extracted the mobility percentage change since March 1st in New York.

IV.2 Infection Data

We used time_series_covid19_confirmed_US.csv in from Johns Hopkins dashboard data. This data set is also well organized and we extracted the confirmed cases number in each counties as well as the total number in New York State.

IV.3 Government Policy Data

We extracted information about New York government policies from the government safety measures data set. We used keywords related to New York and the United States. We extracted data related to the United States to one file and the data related only to New York to another file. This data set is much more challenging because the data are all keywords and descriptions.

For the New York state policy data, we manually check each extracted row (around 10 items). However, this method could not be applied to large data set such as the total US data (around 800 items). This challenge remains to solve in the future.

V Data Analysis

This is part 3 of the project. In this part we mainly use data in New York state from Feb.15 to March.7.

V.1 Visualization

We first do visualization on mobility data and confirmed case data. We take New York State data as example. Figure 1. shows the visualized graph of mobility data from Feb.15 to March.7. Since the data are percentage changes comparing to a baseline which is generated from the average of mobility amount in each weekdays, the data of weekends show as a peaks because the average mobility in weekend before the pandemic broke out is lower.

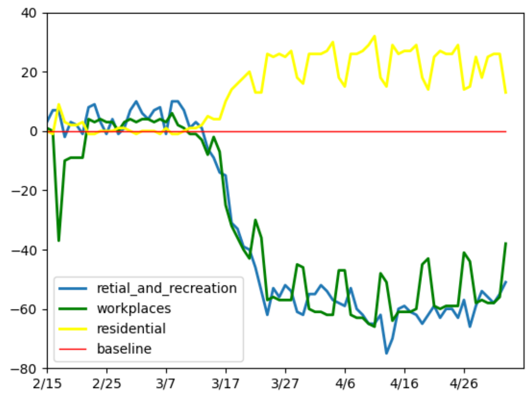

Figure 1: New York State mobility data visualization.

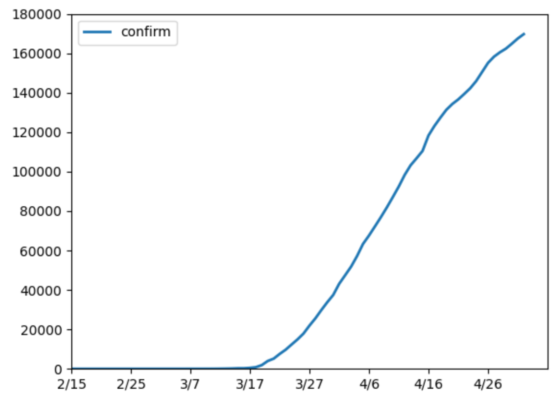

Figure 2: New York State COVID-19 confirmed cases data visualization.

Figure 2. is the visualized data of confirmed case numbers in New York State. We will use models to evaluate the changes in following parts.

V.2 Mobility vs. Policy

From the manually checked data, we have the summary of important government safety policies in New York States. Comparing the time point in the mobility curve, we can know that the policies lowered retial and recreation mobility as well as workplaces mobility and increased residential mobility. However, a few days before the orders released, the curve had started to decrease, which means there were already more people staying at home. Figure 4 added the state of emergency declaring time to the mobility graph.

Date	Policy
3/20	State of emergency
3/21	Ban gatherings and release staying at home order
3/26	CDC statement on COVID-19 self-quarantine guidance for New York
4/9	NYC health emergency

Table 1: New York State important policies related to COVID-19.

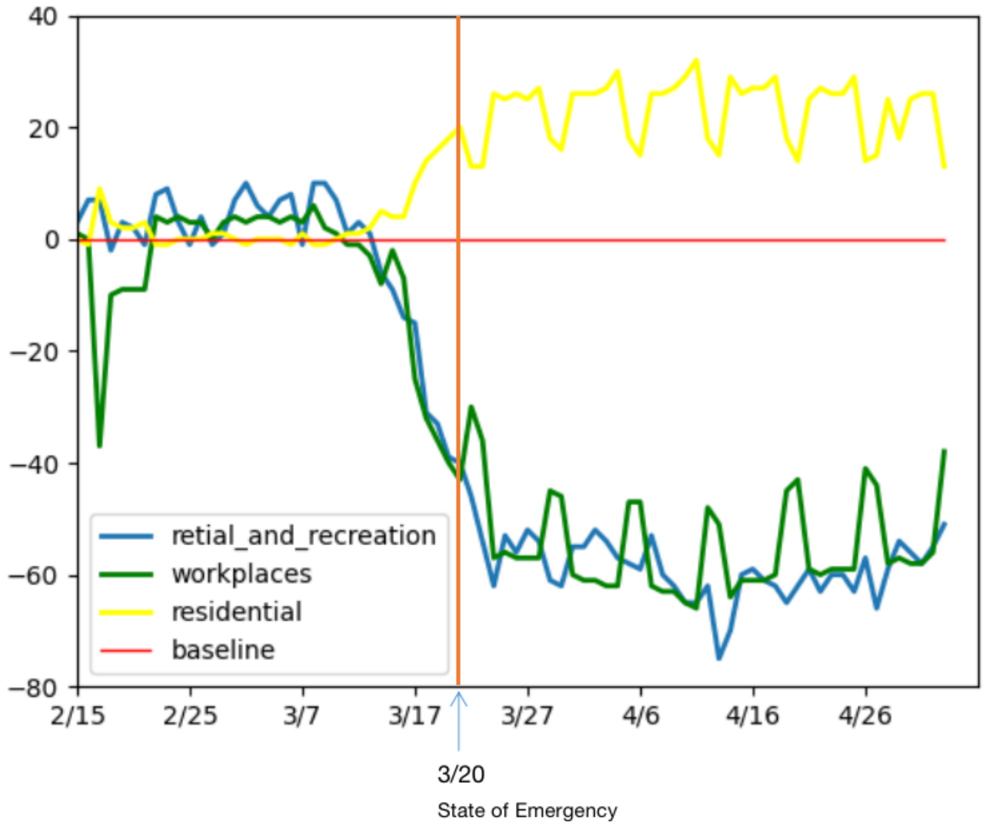

Figure 3: Mobility vs. Policy.

V.3 Mobility vs. Confirm Number

This part does not generate a good result as what we expected. Since the mobility percentage kept at a very low level after March 20th, comparing mobility curve to confirm cases curve makes less sense.

V.4 Model Fitting and Analysis

We use exponential curve to fit the beginning of increasing trends of confirmed cases in the New York State. The results are in Table 2. The parameter d indicates the starting day of the curve that is cut the be fitted. d is the days after Jan.15.

The first curve d=60 is March 16th.

The function of fitted curve will be $y = \exp(a) * \exp(b * x) = \exp(a + b * x)$

d	a	b
60	6.83351	0.12719027
70	9.7400657	0.06593247
80	10.88381126	0.04169873

Table 2: Parameters for fitted exponential curve.

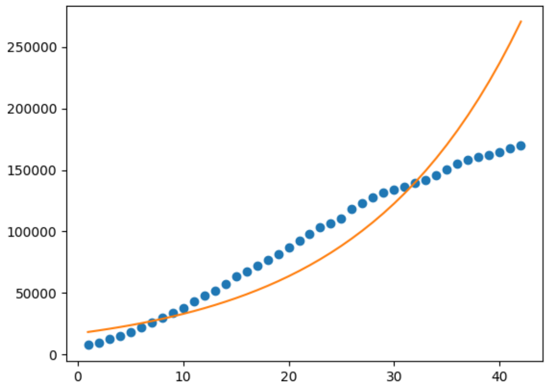

Figure 4: Fitted curve with $d = 70$.

We consider $d = 70$ the best fitted curve for the confirmed cases data. Among the fitted curves, parameter b decreases as we choose the starting day to be late, which means that the exponential rate keeps decreasing after the government released social distancing orders. However, this result is kind of tricky since there are multiple factors that influences the increasing rate of cases number.

VI Conclusion

By taking New York State as an example, we go through the whole process of analyzing how government policies affect social distancing and mobility. After extracting related data and data cleaning, each data set is visualized. We manually process the policy data, where better measures to integrate this data set are remained for discuss. Comparing the mobility data with important policy timeline gives good results, from which we know that governments are effective in social distancing but still many people took actions ahead of orders. However, comparing mobility data to confirmed cases curve is less meaningful. By fitting the confirmed cases curve, we find that the spreading trend of COVID-19 was affected by social distancing, which gives the answer for our questions.

This project is challenging. Many of our assumptions turned out to be meaningless and many results were not as we expected. But we still got some good results by working on multiple ways.

VII Github

Github repository for the project:
https://github.com/timming1437/CS6513-BigData-Project.

VIII Reference

1. Abouk, R., Heydari, B. The Immediate Effect of COVID-19 Policies on Social Distancing Behavior in the United States. SSRN: https://ssrn.com/abstract=3571421 or http://dx.doi.org/10.2139/ssrn.3571421.
2. Adolph, C., Amano, K., Bang-Jensen, B., Fullman, N. & Wilkerson, J. Pandemic politics: Timing state-level social distancing responses to covid-19. *medRxiv (2020).*