
NEW YORK UNIVERSITY
TANDON SCHOOL OF ENGINEERING
CS 6513: BIG DATA

PROJECT PART 1

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: April 19 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-covid-19-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 numbers 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.