**Effectiveness of State Policy Actions to Address COVID-19**

**Team Members:**

Kacper Malysa · Mingxi Xia · Roman Sydorchuk · Spencer Sumner · Yurun Liu

**1. Project Proposal**

**A formal description of the project.**

* The impact of COVID-19 has dramatically altered the daily lives of U.S. citizens. Public policy related to the pandemic can be described as haphazardly implemented and contentious in nature. It seems clear that, given the socioeconomic cost of the pandemic, it is vital to identify the public health policies that produce the best outcomes. This project seeks to examine public health policies implemented in the past six months by various American states in order to identify those that work, and those that do not.
* A number of data repositories have gathered publicly available information on the pandemic’s outcomes, and the public responses to the disease since the beginning of the crisis. In this project, we plan to use this information to evaluate the effects of different types of policies over time on Covid metrics such as infection, hospitalization, and deaths. We also are examining the differences in Covid metrics across different states based on what policies they chose to implement and when they did so. Our ultimate goal is to identify the most effective public health policies for fighting Covid-19, and to establish the optimal timing of policy implementation.

**A specific question or set of questions that the project seeks to address.**

* Which states have most effective policies
* Which policies are most effective
  + Stay at Home Order
  + Quarantine for Travelers
  + Non-Essential Business Closures
  + Large Gatherings Ban
  + Restaurant Limits
  + Bar Closures
  + Face Covering Requirement
* What variables constitute a policy to be implemented (e.g. population, etc.)
  + Population size
  + Population density
  + Infection rate
  + Death rate
* What variables constitute a policy to be effective
  + Same as above for now

**A proposed methodology/approach to the analysis that will be performed (tentative).**

* Linear regression analysis
  + Simple regression
  + Multiple regression
* Logic regression analysis
  + For which state policy is very successful or weak successful
* Z/T/F distribution and test
  + Test each variable is related with the Y value.
* Possibly KNN
  + Build a training and testing model
  + Find the predictor and see if it is same as the true statement

**A metric or set of metrics which will measure analysis results.**

* 95% Confidence or α equal to 0.05
* p-value to decided the hypothesis

**2. Project Outline**

**Literature review and related work.**

The COVID-19 pandemic has provoked sweeping responses in public health policies across the globe. Given that the pandemic continues to this day, research into these policies and their effects on health outcomes is still ongoing, however, several prior publications do exist that attempt to quantify these consequences. Some of the earliest research on this subject was conducted by Hsiang et al. (2020), who focused on the effects of anti-contagion policies across the countries of China, South Korea, Italy, Iran, France, and the United States. This research concluded that anti-contagion policies generally resulted in significant reductions to COVID growth rates, although the precise effects of these policies differed across populations (Hsiang et al., 2020). In the case of the U.S., the combined effect of implemented COVID policies were associated with a reduction in daily COVID growth rates of 31.61% between March 3rd and April 6th of 2020 (Hsiang et al., 2020). Much of this reduction in COVID growth rate can be attributed to social distancing measures, business closures, and policies focused on home isolation (Courtemanche et al., 2020; Dave et al., 2020; Hsiang et al., 2020). Within the U.S. itself the combined effects of social distancing and shelter-in-place orders were predicted to have prevented as many as 90% of COVID cases in the first two months of the pandemic (Courtemanche et al., 2020). This result must be taken with some caution, however, for a number of reasons. First, these results were obtained via an extrapolation of predicted COVID spread models compared to actual observed growth rates (Courtemanche et al., 2020). Second, the precise effect associated with shelter-in-place orders has been observed to vary across different states and across areas with differing population densities (Dave et al., 2020). States that had high population densities, and who adopted shelter-in-place orders earlier on in the pandemic appear to have seen a more dramatic reduction in COVID infection rates than less densely populated states who implemented their orders later (Dave et al., 2020). While these results are encouraging for supporters of public policy initiatives, we must caution that research in this area is still ongoing, and that confounding factors do exist. One such confounding factor is the large-scale adoption of social distancing measures by the public on their own initiative. As noted by Ziedan et al., (2020), much of the decline in mobility, economic activity, and social interaction between individuals can be attributed to private decisions made by individuals either before or during the adoption of COVID public health policies. These confounding factors make it all the more important for us to closely examine the available data. The identification of the most effective policies for combating COVID-19 is crucial, both for avoiding wasteful restrictions and for saving lives.

**All data sources and reference data with descriptions.**

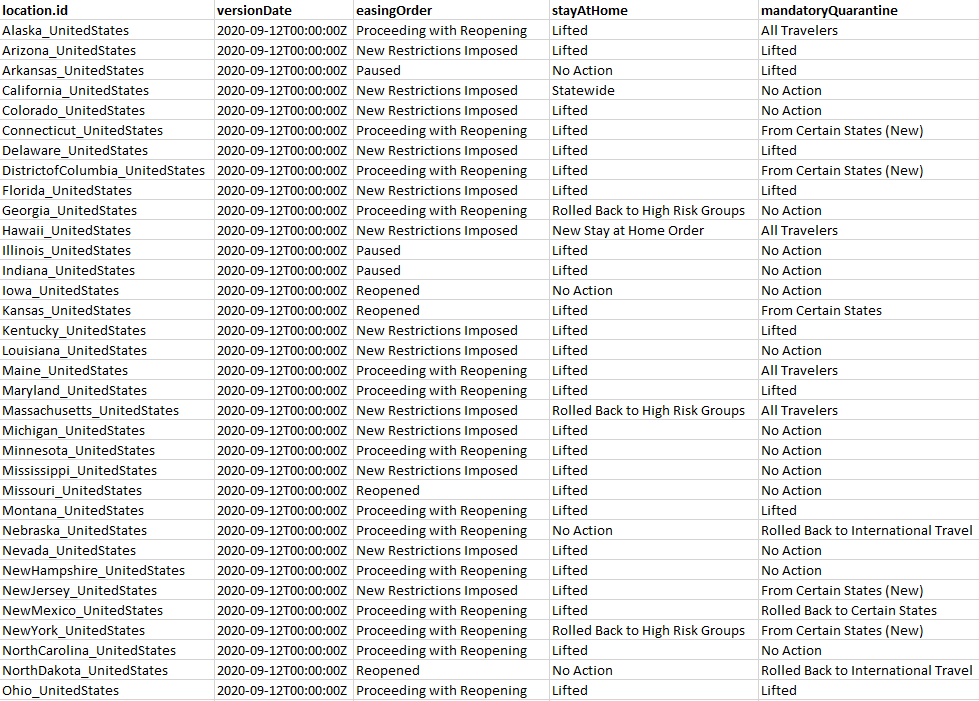
*Datasets:*

* Data Lake (Must request access):
  + <https://c3.ai/products/c3-ai-covid-19-data-lake/>
* Kaiser Family Foundation State Data (Data Lake uses the same data source):
  + <https://www.kff.org/coronavirus-covid-19/issue-brief/state-data-and-policy-actions-to-address-coronavirus/>

*Documentation:*

* C3.ai COVID\_19 API Documentation (5.1)

Trial Data Example (Part of CSV created from R using the data listed above):



* Screenshot above doesn’t show all of the data. The data has 49 observations and 26 dimensions.

*Feature Descriptions:*

| **Field** | **Data type** | **Description** |
| --- | --- | --- |
| **location** | OutbreakLocation | C3.ai Type OutbreakLocation where the policy was enacted. |
| **easingOrder** | string | "Yes" if the location is easing their social distancing measures, "No" otherwise. |
| **stayAtHome** | string | Description of the latest status of the stay-at-home order. |
| **mandatoryQuarantine** | string | Description of status of mandatory quarantine for travelers. |
| **nonEssentialBusiness** | string | Description of restrictions on non-essential businesses. |
| **largeGatherings** | string | Description of restrictions on large gatherings. |
| **schoolClosure** | string | Description of school closures or restrictions. |
| **restaurantLimit** | string | Description of restrictions on restaurants. |
| **PrimaryElectionPostponement** | string | Description of postponement or cancellation of primary elections. |
| **emergencyDeclaration** | string | "Yes" if a state of emergency was declared, "No" otherwise. |
| **waiveTreatmentCost** | string | Description of policies regarding cost sharing for COVID-19 treatment. |
| **freeVaccine** | string | Description of policies requiring free cost COVID-19 vaccines when available. |
| **waiverOfPriorAuthorizationRequirements** | string | Description of policies requiring a waiver of prior authorization requirements. May be superseded by the federal Families First Coronavirus Response Act. |
| **prescriptionRefill** | string | Description of policies regarding early prescriptions refills. |
| **premiumPaymentGracePeriod** | string | Description of policies regarding premium payment grace periods. |
| **marketplaceSpecialEnrollmentPeriod** | string | "Yes" if the special enrollment period for the state's insurance marketplace extended, "No" otherwise. |
| **section1135Waiver** | string | Description of approval status of the Section 1135 waiver. |
| **paidSickLeaves** | string | Description of status of paid sick leave policies adding to federal emergency leave. |
| **expandsAccesstoTelehealthServices** | string | "Yes" if expanded access to Tele-health services are issued, "No" otherwise. |
| **lastSavedTimestamp** | datetime | Datetime of last update for this version. |
| **version** | int | Incrementing version ID for all policies. |
| **versionDate** | datetime | Date of the policy version. |
| **numSavedVersions** | int | Total number of versions of this policy available with allversionsforpolicy. |
| **savedVersion** | int | Incrementing version ID for this policy. |

*Other Notes:*

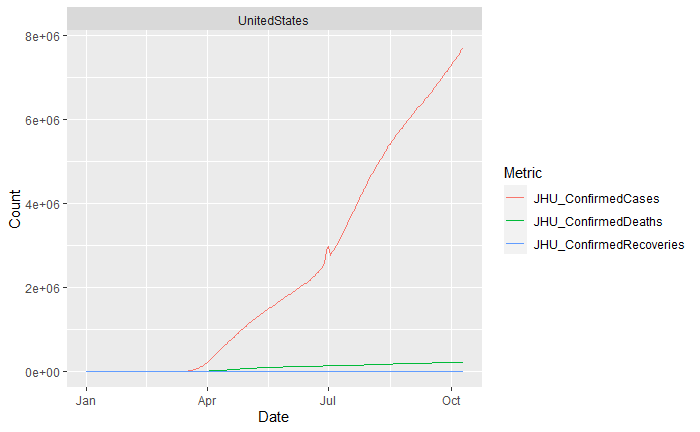
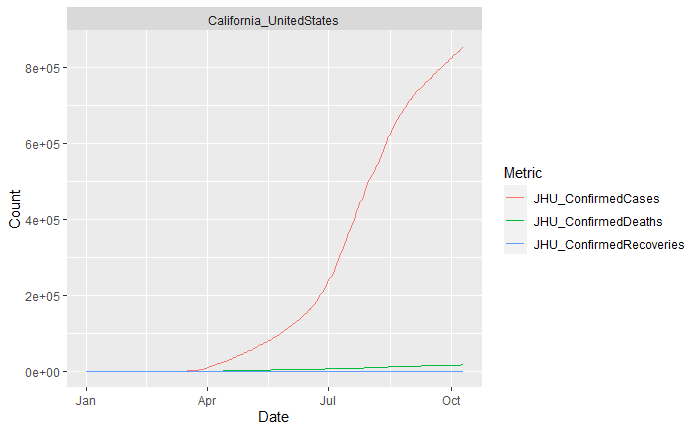
* This specific table only gives us current policies, but the previously mentioned “allVersionsForPolicy” has the historical data for all states.
* The dates for when different policies started and ended can be compared with different data from the c3.ai data lake such as death rate, hospitality rate, etc.

**Data processing and pipeline**

* Data input - Pulling the data from c3.ai Data Lake using their API and documentation into R
* Data cleaning - Removing null values and unnecessary or redundant data from dataset, as well as joining necessary tables when the time comes
* Transformation - When pulling the data from “allVersionsForPolicy”, we have to get each state separately, so data joins will be necessary and will require more further clean up. After that, comparison between different data from the Data Lake can be made.

**Data stylized facts**

* Example graphs that represent different rates over time. These types of data are the ones that we can use for comparisons against policies start and end dates.



**Model selection - feature selection requirements, classification/regression approaches, reference/baseline model, etc.**

* The goal of this project is to find outN and predict which policies established by different states helped prevent the spread and effects of COVID-19 in the united states. As we move on with the project, we will compare the previously mentioned methods such as Linear regression, Z/T/F distribution, KNN, and so on.

**Software packages, applications, libraries, and associated tools, etc**

* Software:
  + R
* Packages and Libraries:
  + tidyverse
  + httr
  + jsonlite
  + fitdistrplus
* Associated Tools:
  + C3.ai COVID-19 Data Lake

**References**

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Affairs*, *39*(7), 1237–1246. https://doi.org/10.1377/hlthaff.2020.00608

Dave, D., Friedson, A. I., Matsuzawa, K., & Sabia, J. J. (2020). When Do Shelter-in-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time. *Economic Inquiry*, *n/a*(n/a). https://doi.org/10.1111/ecin.12944

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L. Y., Hultgren, A., Krasovich, E., Lau, P., Lee, J., Rolf, E., Tseng, J., & Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, *584*(7820), 262–267. https://doi.org/10.1038/s41586-020-2404-8

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Ziedan, E., Simon, K., & Wing, C. (2020). *Effects of State COVID-19 Closure Policy on NON-COVID-19 Health Care Utilization* (No. w27621; p. w27621). National Bureau of Economic Research. https://doi.org/10.3386/w27621