**Research Proposal**

1. Does the standard of living index matter in the spread of Corona?

2.Does weather play a role in confirmed Corona cases?

3.Does social distancing impact the number of confirmed cases of Corona? Also, evaluate the impact of lockdown strategies on confirmed cases.

4.Is the number of new confirmed cases directly driven by the number of tests performed?

Intro:

Given the recent epidemic, our team focus is to evaluate the relationship between social distancing measures and income per capita impacts of confirmed cases in the United States by county using time series data.

Null Hypothesis 1: Income per capita does not impact the number of confirmed cases by US county.

Null Hypothesis 2: The rate of cases spreading around the US is not impacted by the social distancing measures and lockdown status enforced by US states/counties.

Null Hypothesis 3: The number of new confirmed coronavirus cases is not directly tied to the number of tests performed. (There is not a direct % relationship apparent.)

Description:

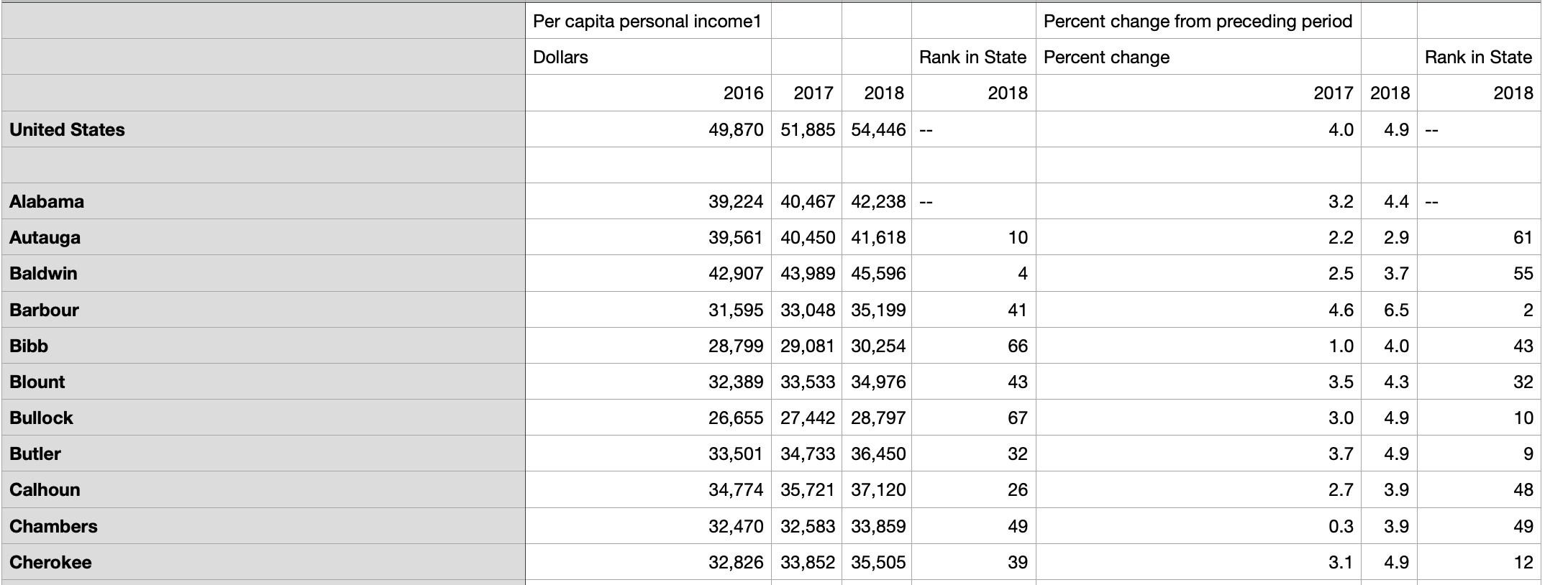
Our project team’s objective is to demonstrate the various factors that contribute to high rates of confirmed cases in the US. We hope to show how strong enforced social distancing measures can reduce the rate of confirmed cases. Understanding the impact of income and strong social measures can help lay a foundation in the response plan for future healthcare crises.

Motivation (“why do you want to do this?“):

The motivation behind this project is to use the skills we learned in class and apply it to real-world data to help gain further insight into Coronavirus. There are multiple facts and figures published about the virus every day, and we thought that this project would be the ideal opportunity for us to do a deep dive into this data ourselves and try to answer some of these pressing questions, which directly impact all of us.

We also want a description of your dataset itself:

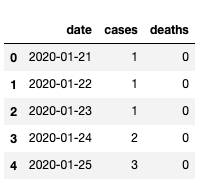
Dataset one for income by county: County level data on new confirmed cases since February, this data file is merged with income per capita from 2018



(“We’ve looked at it, it has xyz rows, these columns are categorical, these are the numeric ones, we plan to use each column in this way, the categorical ones break the data out in xyz way, etc.“)

Data Dictionary for Social Distancing:

|  |  |  |
| --- | --- | --- |
| Variable Name | Description | Type |
| Date | Date | Date field |
| Cases | Number of reported cases | Numerical |
| deaths | Number of reported deaths | Numerical |
| per\_decrease\_ret\_rec | Percentage decrease in traffic for retail stores and recreation. | Numerical |
| per\_decrease\_parks | Percentage decrease in traffic for parks. | Numerical |
| per\_decrease\_grocery | Percentage decrease in traffic for grocery & pharm | Numerical |
| Per\_decrease transit | Percentage decrease in traffic for transit stations. | Numerical |
| per\_decrease\_workplace | Percentage decrease in traffic for workplaces | Numerical |
| per\_decrease\_residential | Percentage decrease in traffic for residentials | Numerical |
| Num\_lockdown | Number of days to lockdown | Numerical |
| loc\_status | Lockdown status | Categorical |



US Cases

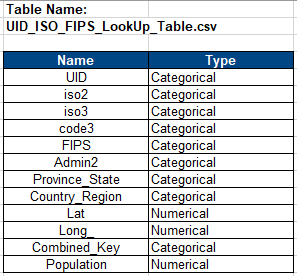
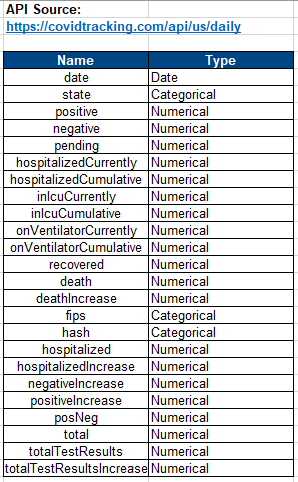
Data sources Websites:

· <https://www.kff.org/health-costs/issue-brief/state-data-and-policy-actions-to-address-coronavirus/>

· <https://s3.eu-central-1.amazonaws.com/hdx-ckan-filestore-prod/resources/8403502a-5c61-4a17-9567-dd1c50829f0f/data_ncov2019.csv?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Expires=180&X-Amz-Credential=AKIARZNKTAO7XQY7ED6N%2F20200416%2Feu-central-1%2Fs3%2Faws4_request&X-Amz-SignedHeaders=host&X-Amz-Date=20200416T204226Z&X-Amz-Signature=628393af08f35a57d77d8a80214d13edd2fc7ed331730648d6d84047afa655eb>

<https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data>

Latitude, Longitude, and Population for Counties. Merge on County/State or FIPS.



<https://covidtracking.com/api/us/daily>

Used to calculate number of tests and positive results. (positive + negative + pending = total)