Project Proposal

Before you start writing any code, your group should outline the scope and purpose of your project. This helps provide direction and prevent [scope creep] (https://en.wikipedia.org/wiki/Scope_creep).

Project Title: Covid-19 Spread based on Human Factors and Social Distancing

Team Members: Jimmy Walsh, Sequoyah Cooper, Mathew Lack, Shweta, Shukla-Goyal, Liz Stevens

Project Description/Outline:

We are interested in using our new data skills to look at possible relationships between the Covid-19 pandemic, human factors (socio-economic status, unemployment), and state's decisions about social distancing.

Sources & Datasets to Be Used:
(the kind of data you'd like to work with/field you're interested in (e.g., geodata, weather data, etc.)

- Liz Ideas:
  - Geodata (# of hospitals & location of hospitals by county)
    - https://www.kaggle.com/roche-data-science-coalition/uncover#definitive-healthcare-usa-hospital-beds.csv
  - Socio-Economic/Financial Distress and relationship to Covid cases deaths
    - County Level Datasets (for socioeconomic indicators)
    - https://research.stlouisfed.org/resources/covid-19/preliminary/covid-19-households-financial-distress-part-2
    - https://www.ers.usda.gov/data-products/county-level-data-sets/
  - Unemployment and Restaurant/Industry Sales related to Covid-19
    - https://fred.stlouisfed.org/release/tables?rid=50&eid=3077
    - https://fred.stlouisfed.org/series/RSXFS?utm_source=series_page&utm_medium=related_content&utm_term=related_resources&utm_campaign=categories
  - Social Distancing Scoreboard
    - https://www.unacast.com/covid19/social-distancing-scoreboard
- Sequoyah:
  - https://wallethub.com/edu/most-aggressive-states-against-coronavirus/72307/
- Jimmy
  - https://raw.githubusercontent.com/nytimes/covid-19-data/183f76a2d78e7ab556339ed0534802e1b3e22689/us-counties.csv

- Shweta (API for live information about COVID-19)
    - https://coronavirus-19-api.herokuapp.com/all
    - https://coronavirus-19-api.herokuapp.com/countries

Research Questions to Answer:
(the kinds of questions you'll be asking of that data)

Is there a relationship between the spread of the pandemic and socio-economic factors of people infected?

- Does spread correlate with unemployment rates and industry?
- Where are hospitals located in relation to highly affected areas?
    - Is lack of hospital beds correlated with greater # of fatalities?
- Is there a relationship between states successfully practicing social distancing and reducing the amount of confirmed cases?
    - How does population density affect this?
    - How does when a state decides to social distance (# of cases before they implement policy) impact the effectiveness of flattening the curve (in # of cases and # of deaths)?

Rough Breakdown of Tasks:

Data Collection and API use: Shweta
Cleaning Data for Data Analysis: Sequoyah
Creating loops to analyze the data and figure out how to manipulate it for answering questions: Everyone (Headed by Jimmy/Liz)
Creating Visuals/Graphs: Liz/Matt
Summary Analysis of findings: Matt