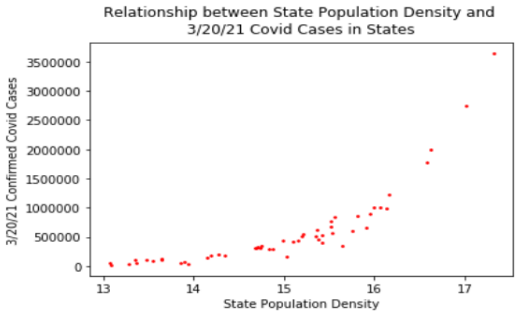
Project Deliverable 3 (v1 Final Report)

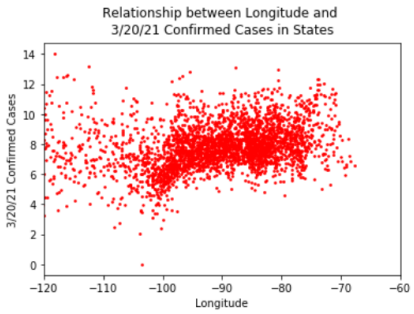
All data should have been collected. All project questions should have been reviewed, answered, and submitted in a written document outlining findings as a PR. You will also be asked to submit the associated data and a README explaining what each label/feature in your dataset represents. Your team should meet with the client before this deliverable.

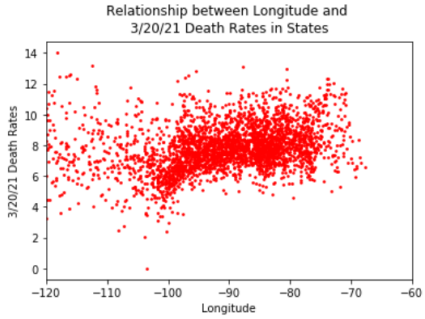
Checklist

1. All data is collected
2. Refine the preliminary analysis of the data performed in PD1&2
3. Answer another key question
    - What are additional social determinants of health (other than Health Insurance coverage) that affect the COVID-19 confirmed cases and deaths?
        - Population density?
            - According to our data visualization, there appears to be an exponential relationship between population density and coronavirus confirmed cases. The number of confirmed cases increases as state population density increases. People who live in densely populated states are more likely to contract coronavirus.



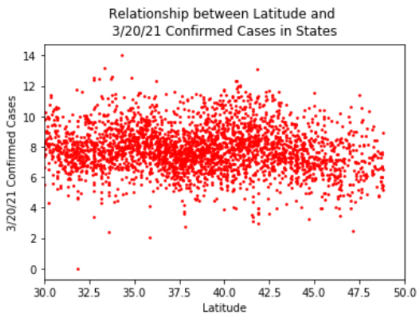
        - Physical location?
            - US counties located between 100W and 80W longitude have the highest *frequencies* of confirmed cases and death rates. These are counties that are located in the Eastern half of the US.

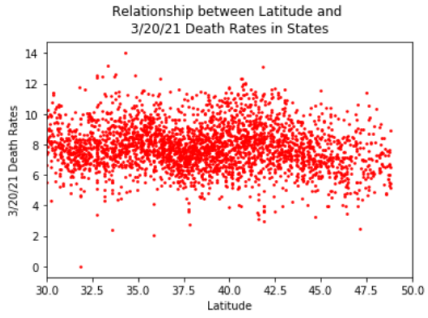




        - Climate?
            - Climate is determined by latitude. The US is approximately between 30N and 50N latitude. The frequencies of confirmed cases and death rates are

both distributed evenly between 30N and 50N. There is no apparent correlation between climate and confirmed cases.





4. Attempt to answer overarching project question
    - Is there a correlation between Health Insurance and COVID confirmed cases and deaths?
        - There seems to be a correlation, but it may be influenced by social determinants of health. People with health insurance coverage may have more opportunities for COVID testing, and if they test positive they would count towards the number of confirmed COVID cases.
5. Create a draft of your final report
6. Refine project scope and list of limitations with data and potential risks of achieving project goal
    - Project scope: What is the relationship between coronavirus recovery versus death rates and access to health insurances? How do other social determinants of health affect the spread and severity of coronavirus in the US?
    - Limitations:
        - Not including vaccine data
        - Health insurance data is from 2018
7. Submit a PR with the above report and modifications to original proposal

<u>All Project Questions:</u>

Key Focus Questions:

1. What is the relationship between coronavirus confirmed cases versus death rates and access to health insurance? How does this relationship change across various states and counties in the United States?
    - Confirmed cases and death rates seem directly related, and both of them also seem directly related to health insurance access. Counties with more confirmed cases always have more deaths compared to counties with less confirmed cases.
2. What are other social determinants of health that affect the severity and spread of coronavirus? How do these relationships compare with health insurance?
    a. Population: According to our data visualization, there appears to be an exponential relationship between population density and coronavirus confirmed cases. The number of confirmed cases increases as state population density

increases. It can be hypothesized that people who live in densely populated states are more likely to contract coronavirus.

b. Latitude: Generally, the relationship is consistent throughout the country. Climate has little effect on covid rates because latitude affects the climate and the latitude of the US is generally consistent.
c. Longitude: The longitude may affect covid rate since there are more covid cases on the east coast.

Guiding Questions for Initial Analysis:

1. How has the coronavirus impacted the United States as a whole? When was the first wave, second wave?
    a. Covid was first detected in the United States around late February and by mid-March, all 50 states, the District of Columbia, New York City, and four U.S. territories had reported cases of COVID-19. From the first step of our data preprocessing, we saw an average of about 10 confirmed coronavirus cases in March, and about 1035 confirmed cases in July. Within that time period, the average number of deaths due to coronavirus increased from about 0 to 41.
2. Do the waves occur at various points in different areas? How are the different demographics affected?
    a. The waves may occur at different times, if they start in denser populated areas the wave coils spread faster. However it may take some time for that wave to get to less populated counties, which in turn would lead to waves occurring at various timeframes.
3. First glance, what is the spread of healthcare access like? What states/counties have more access versus less access?
    a. The average percentages of the insured population among the US states are all at least 80%. The states with the most insured population are the states in the East Coast and states with the highest population of people. For example, Massachusetts has the highest percentage of insured population (96.7%) and high ranking states include New York, Connecticut, and California.

README.txt: