Project name:
Relationship between health insurance access and coverage on COVID rates (recovery & death) and vaccination access

Team:
Linsy Wang linsyw@bu.edu
Sooyoun Lee selena98@bu.edu
Dana Zheng daynah@bu.edu

Project Description:

Collecting information on COVID-19, specifically focusing on total number of recovered patients, total number of deaths and total number of current active cases. GitHub and Kaggle contain worldwide databases on the coronavirus, with frequent updates. The data gathered from the worldwide databases will be compared with health insurance access, collected from the CDC, in the United States across different demographics, gender, race, and location. Upon determining a relationship between coronavirus rates and health insurance access across different demographics, information will also be obtained on the total number of vaccinations across the same demographics. The ultimate goal is to determine the relationship among health insurance access and coronavirus recovery vs. death rates. With increased access to vaccinations, the goal is also to uncover how the previously determined relationship translates to vaccination access and completion. As there is data on coronavirus rates across the world, but only health insurance access within the United States, the primary focus of the relationship will be on the citizens of the USA.

Datasets:

https://www.kaggle.com/imdevskp/corona-virus-report?select=country_wise_latest.csv
https://www.kaggle.com/gpreda/covid-world-vaccination-progress
https://covid19.census.gov/datasets/health-insurance-coverage-states/data
https://www.cdc.gov/nchs/covid19/pulse/health-insurance-coverage.htm
https://www.cdc.gov/nchs/covid19/pulse/reduced-access-to-care.htm
https://ourworldindata.org/coronavirus/country/united-states?country=~USA

Updated Potential Risks and Limitations:

- Refining to just cases in the US
- won't be able to see the relationships between various countries, as we will be focusing on the US
- There could be inaccuracies in the data of cases

General Questions:

- Is the usa_county_wise dataset sufficient?

- At the moment, we have processed data from March - July, which is before the vaccination process has started
- Do the coronavirus waves occur at various points in different areas? How are the different demographics affected?

Key Question(s) answered:

1. How has the coronavirus impacted the United States as a whole?

The virus was first detected in the United States around late February. 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.

Next Steps:

- Pre-process and clean the population insurance dataset using similar methods
- Finish pre-processing and cleaning the covid cases and deaths datasets (to most recent dates)
    - https://ourworldindata.org/coronavirus/country/united-states?country=~USA
        - Daily confirmed cases (per million), death rates (per million), reproduction rates, ICU rates, hospital patients, testing rates, total vaccination rates
        - It also includes data on population density, gdp per capita, cardiovascular/diabetes prevalence, life expectancy, and female/male smoker
- Brainstorm methods on how to portray/determine correlation between our variables