RELATIONSHIP BETWEEN HEALTH INSURANCE ACCESS AND COVERAGE ON COVID RATES (RECOVERY & DEATH) AND VACCINATION ACCESS

Contact	Name: Linsy Wang, Sooyoun Lee, Natalia Matuk, Dana Zheng Email: linsyw@bu.edu, selena98@bu.edu, nataliam@bu.edu, daynah@bu.edu
Organization	N/A
Organization Description	N/A
Project Type	Data Science
Project Description	**This project involves:** 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.

Data Sets	Links to data sets e.g. APIs where data will be sourced from, folders/ files in the google drive that we were given, links to sites to be scraped, etc.
	
	COVID - list of links below
	
	Country daily/biweekly cases/deaths for countries and vaccination data:
	https://github.com/owid/covid-19-data/tree/master/public/data
	
	(New) Cases/deaths/recoveries:
	https://www.kaggle.com/imdevskp/corona-virus-report?select=country_wise_latest.csv
	    ● Fields:
	        ○ Country/Region

- # Confirmed
- # Deaths
- # Recovered
- # Active
- # New cases
- # New deaths
- # New recovered
- # Deaths/100 cases
- # Recovered/100 cases

Vaccinations: https://www.kaggle.com/gpreda/covid-world-vaccination-progress
- Fields:
  - Country
  - Date
  - # Total vaccinations
  - # People vaccinated
  - # People fully vaccinated
  - # Daily vaccinations
  - % Total vaccinations
  - % People vaccinated

US states insurance based on income:
https://covid19.census.gov/datasets/health-insurance-coverage-states/data

Uninsured population during COVID (April 23, 2020 - January 18, 2021)
https://www.cdc.gov/nchs/covid19/pulse/health-insurance-coverage.htm
- Fields
  - % Uninsured at the time of interview
  - % Private health insurance coverage
  - % Public health insurance coverage
  - Categories/groups
    - Age
    - Gender
    - Race/Hispanic ethnicity
    - Education
    - State

Reduced access to healthcare caused by COVID 19: (April 23, 2020 - January 18, 2021)
https://www.cdc.gov/nchs/covid19/pulse/reduced-access-to-care.htm
- Fields
  - % Delayed medical care
  - % Did not get needed care
  - % Delayed or did not get care
  - Categories/groups
    - Age
    - Gender
    - Race/Hispanic ethnicity
    - Education

	■ State
Suggested Steps	Steps to complete the project including data collection, data cleaning/ processing steps, and analysis 1. Gather and consolidate all data from GitHub, Kaggle, CDC and NIAD (potentially). Separate data into coronavirus rates, health insurances access across chosen demographics and vaccination rates. 2. For data cleaning, utilize Pandas in Python in order to organize and clean gathered data for faster visualization and analysis 3. Data visualization: a. Matplotlib: to illustrate the correlation (if it exists) between coronavirus data and health insurance status (subgroups may include demographics) 4. Data Analysis: Utilize visualizations from Matplotlib to support claims and answers to Key Focus Questions
Questions to be answered in Analysis	Key Focus Questions: 1. What is the relationship between coronavirus recovery versus death rates and access to health insurances? How does this relationship change across various demographics within the United States? 2. With access to the coronavirus vaccine, is there a connection between the coronavirus recovery versus death rates and increased or decreased vaccination? Does health insurance access impact vaccinations at all? 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. Do the waves occur at various points in different areas? How are the different demographics affected? 2. First glance, what is the spread of healthcare access like? What demographics have more access versus less access? 3. Difference in coronavirus cases with regards to just private or public health insurance?
Additional Information	- Team meeting schedule: weekly to biweekly

List of Limitations/ Potential Risks	- Vaccination data may not be extensive as the distribution process is still undergoing - The distribution of vaccine does not appear to have any impact on insurance as the phases are based on occupation, age, and preexisting health conditions - Note: Current vaccination data may be biased to reflect people who already have health insurance since the current distribution phase is for those who have preexisting health conditions - Focusing on the United States, other countries may have different information for healthcare and covid data, this will help keep our data more consistent - If just focusing on the United States and the relationship of healthcare and covid data there won't be an opportunity to compare the relationship in the United States to other countries in the world - Possibly no strong correlation between coronavirus cases and healthcare - Accuracy of coronavirus cases, people may not be regularly getting tested or be asymptomatic - There is a big discrepancy between testing requirements (i.e. college students are required to get weekly tested v.s. people who may have not gotten tested at all since COVID-19)

Links:
https://catalog.data.gov/dataset?_groups_limit=0
https://www.pewresearch.org/download-datasets/
https://ucsd.libguides.com/data-statistics/gender
https://www.kaggle.com/datasets

COVID -
https://www.cdc.gov/library/researchguides/2019novelcoronavirus/databasesjournals.html
https://github.com/owid/covid-19-data/tree/master/public/data
https://www.niaid.nih.gov/research/accessing-clinical-data
https://www.kaggle.com/imdevskp/corona-virus-report?select=country_wise_latest.csv
https://www.kaggle.com/gpreda/covid-world-vaccination-progress
https://data.cdc.gov/NCHS/Indicators-of-Health-Insurance-Coverage-at-the-Tim/jb9g-gnvr
https://www.cdc.gov/nchs/covid19/pulse/health-insurance-coverage.htm
https://www.cdc.gov/nchs/covid19/pulse/reduced-access-to-care.htm

insurance
https://covid19.census.gov/datasets/health-insurance-coverage-states/data