Project 3

By: Edgar Alvarado, Brad Anderson, Alexandria Orvis, Corbin Cox

Project Proposal

Overview:

We are utilizing data from the US Department of Health, Chronic Disease Indicators (CDI) webpage and the US Environmental Protection Agency, Air Quality Systems API webpage to create our visualizations.

Purpose:

The purpose of our project is to track air quality across the United States and compare the prevalence of 3 chronic diseases with air quality metrics using data from 2020. The goal is to analyze any possible correlation between local air quality per state and the prevalence of Asthma, COPD, and Cardiovascular Disease.

Table 3. Pollutant-Specific Sensitive Groups

When this pollutant has an AQI above 100…	Report these Sensitive Groups
Ozone	People with lung disease, children, older adults, people who are active outdoors (including outdoor workers), people with certain genetic variants, and people with diets limited in certain nutrients are the groups most at risk
PM2.5	People with heart or lung disease, older adults, children, and people of lower socioeconomic status are the groups most at risk
PM10	People with heart or lung disease, older adults, children, and people of lower socioeconomic status are the groups most at risk
CO	People with heart disease is the group most at risk
NO2	People with asthma, children, and older adults are the groups most at risk
SO2	People with asthma, children, and older adults are the groups most at risk

Notes: Statements may be combined so that each group is mentioned only once.

Datasets to be analyzed

US Department of Health: Chronic Disease Indicators (CDI) - https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi

This dataset collects studies done in each state to track the measures of various chronic diseases.

US Environmental Protection Agency: Air Quality System (AQS) API - https://aqs.epa.gov/aqsweb/documents/data_api.html

This dataset collects air quality metrics across the United States from various site locations. Different stations collect different reporting parameters.

Visualization 1 - Choropleth Map

Analysis

LINK: https://handro4.github.io/Project-3---Group-6/Dashboard/index.html
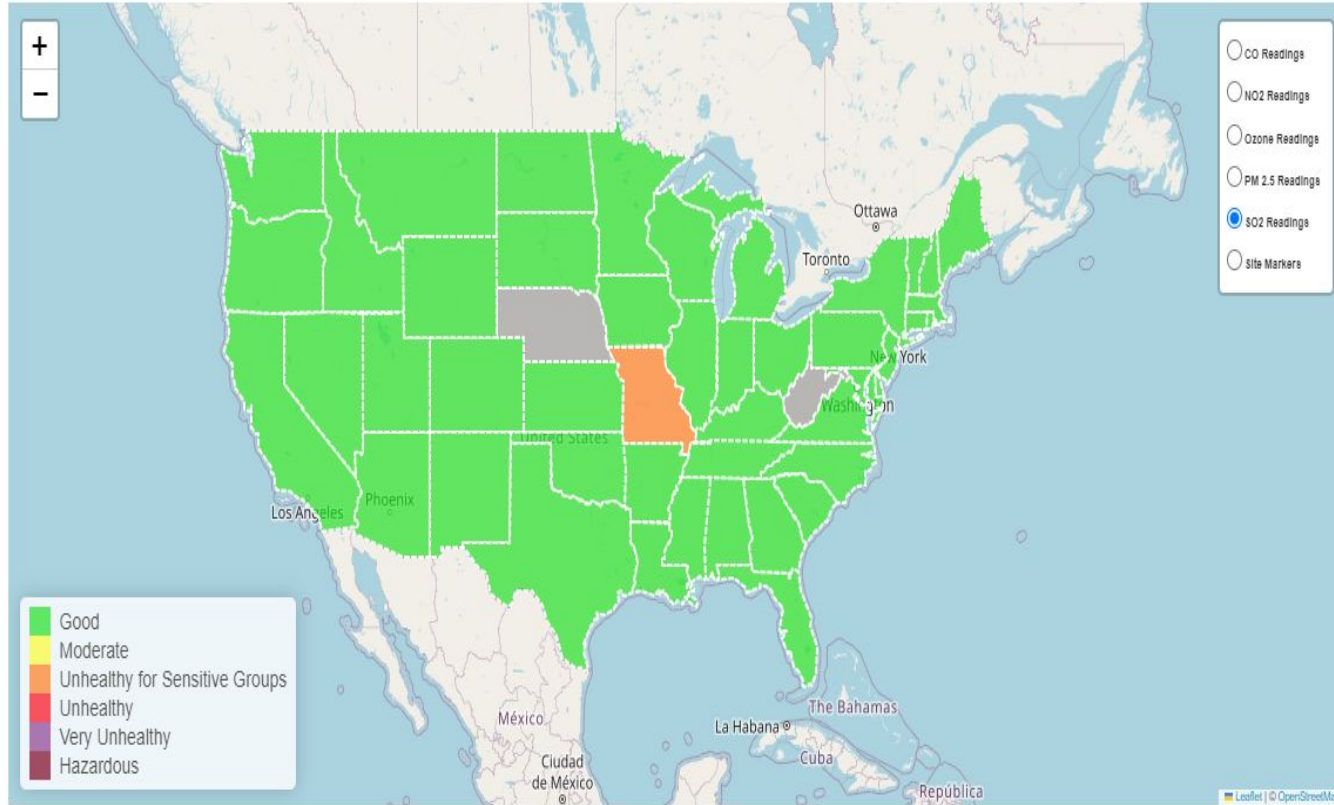
❖ The Choropleth map shows the impact of AQI pollutants on the states. Each state will be highlighted a color corresponding to the pollutant concentration in that state. You can toggle the different parameters on to see the effect each parameter has on each state.

❖ A few things to note about this visualization
➢ The readings are taken from site locations that span across the country. Site locations are concentrated more heavily in certain states which could skew the state level results
➢ Most common readings were PM 2.5 and Ozone which could be why those have the most noticeable effect on the map
➢ Each station reported a yearly average for these readings. We are displaying the max average reading.
➢ Clearly most of these readings are in the green, which was not exactly the expected result.

Visualization 2 - Scatterplot

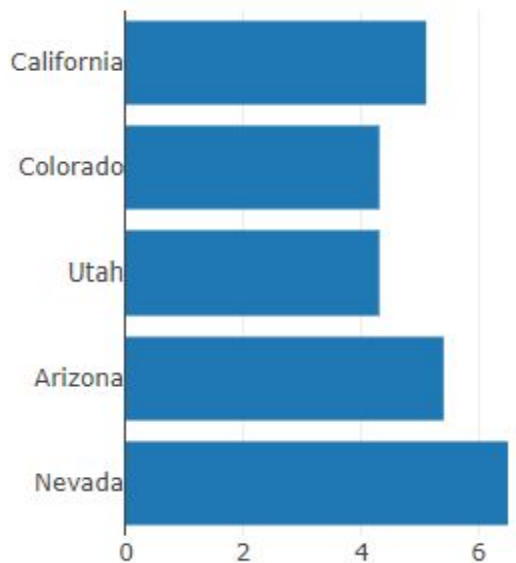Effect of PM 2.5 on Cardiovascular Disease Prevalence
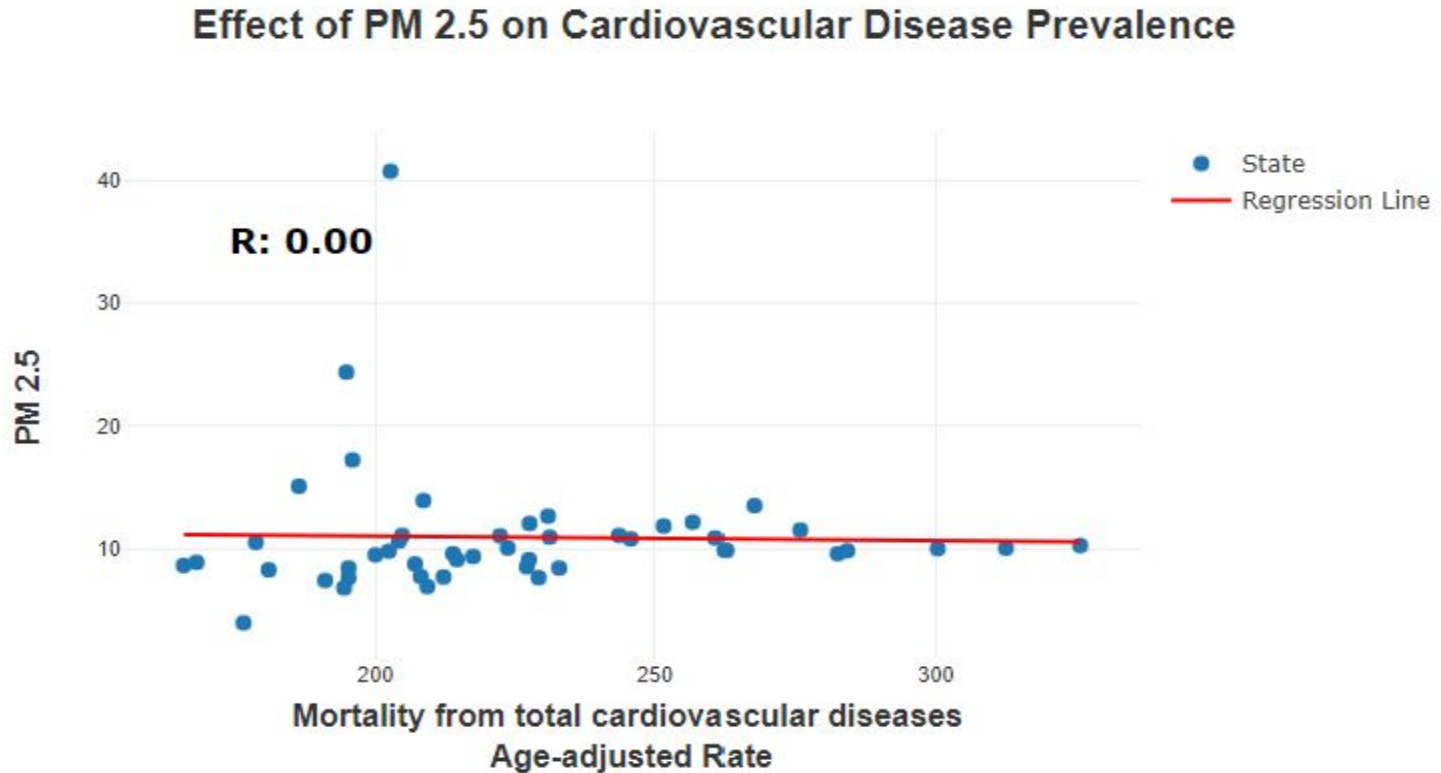
Analysis

- ❖ This scatter plot compares each AQI parameter to asthma *prevalence* (% of the population with the disease), chronic obstructive pulmonary disease *prevalence*, and cardiovascular mortality *rates* (deaths per 100,000 population).
- ❖ We plotted a linear regression the r-squared value to show how predictive the line is of the data points
  - ➢ None of the regression lines show predictive values for the relationship between air quality measures and health outcomes
- ❖ There are several reasons the data might not be related:
  - ➢ Inconsistent data collection between states - states have a widely varying number of stations
  - ➢ PM 2.5 and Ozone are measured more frequently than the other parameters

Visualization 3 - Bar Chart

Top 5 Polluted States for Ozone

(Bar chart showing Age-adjusted Prevalence by state)

- California: ~5.0
- Colorado: ~4.3
- Utah: ~4.3
- Arizona: ~5.4
- Nevada: ~6.5

Age-adjusted Prevalence

Analysis

- ❖ The bar chart data will show how close the data is per the top 5 states
  - ➢ Due to inconsistent data logging per state
- ❖ All charts show the top 5 states with the reported pollution
  - ➢ Correlating the health outcomes to them show minimal correlation
  - ➢ We know that there is some data that we didn't include due to research constraints

Conclusions

Our data did not show any significant relationship between any AQI parameter (**OZONE, NO2, SO2, CO, PM 2.5**)and the 3 chronic diseases we observed (***Asthma, Cardiovascular Disease, COPD***).

Limitations:

❖ We did not have full data for 4 different states which were omitted from the dataset: **Alaska** (ALL), **Nebraska** (ALL), **West Virginia** (ALL), **South Carolina** (CO)
❖ The air quality and disease prevalence data is only from the year 2020
❖ It's possible that there was a more accurate or useful way way to interpret the AQI data

Sources

US Department of Health: Chronic Disease Indicators (CDI) - https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi

US Environmental Protection Agency: Air Quality System (AQS) API - https://aqs.epa.gov/aqsweb/documents/data_api.html

Choropleth documentation - https://leafletjs.com/examples/choropleth/

Linear Regression Package - https://unpkg.com/d3-regression@1.3.10/dist/d3-regression.min.js

Airnow Air Quality Index - https://www.airnow.gov/education/students/what-is-the-aqi/

THE END