Maria Camila Coronado
EDLD 652 – Winter 2023

Final project proposal

With this project I would like to explore the relationships between school absenteeism and outdoor air pollution in the state of California, which is one of the potential locations for case studies in my dissertation.

1. **Research questions:**
   1.1. How is absenteeism distributed by county in California?
   1.2. How is outdoor air pollution distributed by county in California?
   1.3. Is there a relationship between outdoor PM 2.5 air pollution in California counties and school chronic absenteeism?

2. **Data description and sources:**

2.1. California schools' absenteeism data:

I will use data of chronic absenteeism from California Department of Education of the year 2021-2022, that is publicly available here: https://www.cde.ca.gov/ds/ad/filesabd.asp. This file contains 264938 rows and 13 columns with information on Academic Year, Aggregate Level, County Code, District Code, School Code, County Name, District Name, School Name, Charter (ALL/Y/N), Reporting Category, Chronic Absenteeism Eligible, Cumulative Enrollment, Chronic Absenteeism Count, Chronic Absenteeism Rate. According to the website chronic absenteeism rate is "The unduplicated count of students determined to be chronically absent (Chronic Absenteeism Count) divided by the Chronic Absenteeism Enrollment at the selected entity for the selected population using the available filters.". I will use the chronic absenteeism rate of the total students of each academic institution and average it by county to create the visualizations.

2.2. Outdoor air pollution by county in California:

Particulate matter of is deemed as a cause for respiratory illnesses in both children and adults. Particles of less than 2.5 µg/ m3 (respirable) can infiltrate the gas-exchange region of the lungs. PM 2.5 has been related to health effects like reduced lung function, asthma and other pulmonary diseases in children and adults who were exposed to high levels of endotoxins present in PM (Morakinyo et al., 2016) . Average daily outdoor PM 2.5 air pollution by county in California is available as one of the county health rankings of the state here: https://www.countyhealthrankings.org/explore-health-rankings/california/data-and-resources

2.3. California counties:

I found a shapefile of California counties that I could potentially use to map the absenteeism data from the California Department of Education. The data is publicly available here: https://data.ca.gov/dataset/ca-geographic-boundaries.

3. Visualization ideas:

3.1. Geographic distributions:

For questions 1 and 2 I imagine being able to create 2 visualizations that explain the answers. The first visualization could be a map with the county boundaries of the state, with information on average school absenteeism by country similar to the one I am showing in Figure (1A). Next to it, I imagine a bar graph showing the top 15 counties with more chronic absenteeism in the state. I show a similar example in figure (1B).

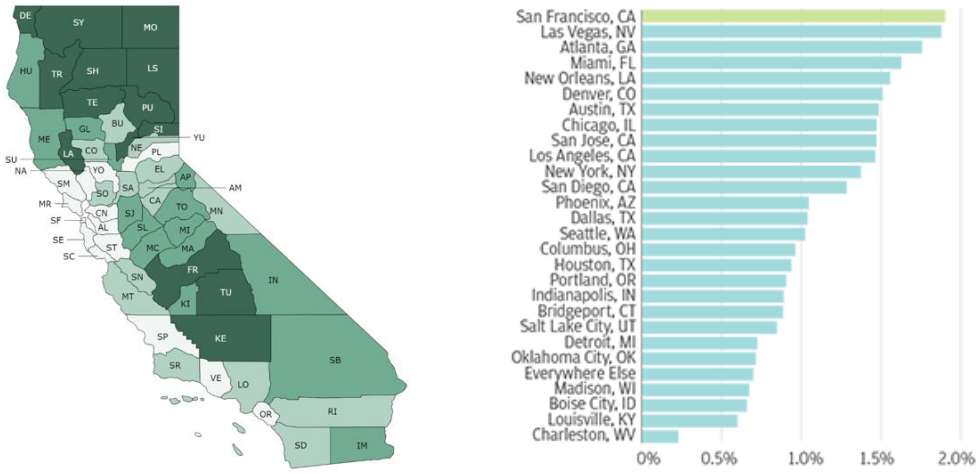

Figure 1. (A) California counties with gradients of green according to data. Source : https://www.countyhealthrankings.org/explore-health-rankings/california/data-and-resources. (B) Bar graph with percentages. Source: https://www.jpmorganchase.com/institute/research/cities-local-communities/institute-san-francisco-economy.

Each map should easily convey which counties have more or less chronic absenteeism in schools, and more or less daily PM 2.5 average concentrations.

3.2. Correlations between absenteeism and PM 2.5 by county:

For the research question 3, I imagine a scatterplot or a bubble chart where each plot represents a county in California. One axis would represent the absenteeism percentages by county and the other would represent the average PM 2.5 values. Since both are continuous variables, I think this would be a good way to show their associations.

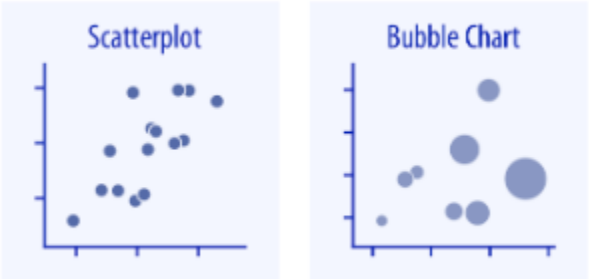

Figure 2. Visualizations of x-y relationships. Wilke Chapter 5. https://clauswilke.com/dataviz/directory-of-visualizations.html

4. Intended audience:

All visualizations would be intended for laypersons. Every visualization would be accompanied with definitions and explanations of the information that each graphic is showing, as well as a short paragraph of analysis.

References:

data.ca.gov. (2019). *CA Geographic Boundaries - Datasets - California Open Data*. https://data.ca.gov/dataset/ca-geographic-boundaries

Morakinyo, O. M., Mokgobu, M. I., Mukhola, M. S., & Hunter, R. P. (2016). Health Outcomes of Exposure to Biological and Chemical Components of Inhalable and Respirable Particulate Matter. *International Journal of Environmental Research and Public Health*, *13*(6). https://doi.org/10.3390/IJERPH13060592

Rankings, C. H. (2023). *Data and Resources | County Health Rankings & Roadmaps*. https://www.countyhealthrankings.org/explore-health-rankings/california/data-and-resources

Wilke, C. (2019). Fundamentals of Data Visualization. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). https://clauswilke.com/dataviz/directory-of-visualizations.html