Executive Summary

Quinn Bankson, Paige Hathaway, Eric Leinweber, and Betsy Norwood

Policy Question

How does proximity to coal-fired power plants affect health outcomes in the United States?

Background

Historically, coal has been a significant source of energy production in the US. In 2021, the burning of coal produced 22% of the US's total electricity (U.S. Energy Information Administration). Coal has been known to have negative health consequences, as easily seen from coal miners and those who burn coal in their homes as a heating source. However, the health impacts of the large-scale burning of coal as an energy source are largely unknown. A review conducted by Duke University of 113 peer-reviewed studies that looked at the health effects of coal-burning power plants found that pollutants related to the burning of coal were linked to negative health outcomes such as lung cancer, premature deaths, and low birth-weights (Duke Health News). However, there are still major gaps in the knowledge of this subject, and more research is required to know the full extent of the negative health consequences caused by coal fired power plants. The lack of knowledge on this subject provided inspiration for this research project.

The Data

To form the data set that we used for our analysis, we collected air pollution data from the Environmental Protection Agency (EPA), respiratory health outcome data from the Institute for Health Metrics and Evaluation, and smoking data from the Center for Disease Control (CDC). All of the data collected was at the county-level to allow for the joining of the data, along with our feature engineering that relied on country centroids. Our data focuses on the continental US.

The data collected from the EPA focuses on Air Quality Index (AQI) data that the EPA collected in 2014. The AQI is the EPA's index for reporting air quality. Values range from 0-301+, with the groups of levels of concern as follows: good (0-50), moderate(51-100), unhealthy for sensitive groups (101-150), unhealthy (151-200), very unhealthy (201-300), and hazardous (301 and higher). The AQI data provided by the EPA also includes the number of days that carbon monoxide, nitrogen dioxide, ozone, particulate matter smaller than 10 micrometers, and particulate matter smaller than 2.5 micrometers were detected in the air in the counties with an AQI monitor in the year 2014. In our exploratory data analysis (EDA), one part of our analysis was to look at the missingness of our data set. Our exploration of missingness showed that the AQI variables have fairly high rates of missingness, while the rest of our data set has very low or no missing data. This is because of the amount of data the EPA collects on AQI- it is not feasible for the EPA to have AQI monitors in all of the counties in the US, so they place these monitors in approximately a third of the US counties.

The data collected from the Institute for Health Metrics and Evaluation looks at mortality from asthma, Pneumoconiosis, lung disease, and respiratory disease per 100,000 in the year 2014. It also includes percent change in mortality for each of these illnesses from the year 1984 to 2014.

Feature Engineering

Using the R package Geosphere, we made a matrix of the distances between all coordinates of centroids of counties in the US and all coordinates of coal power plants that are operating in the US. This distance was calculated by the Haversine formula, which is a formula that is used to calculate the distance between two points on a sphere, such as the Earth. The minimum distance was then pulled from the large matrix, one minimum distance was assigned to each county centroid representing its distance to the nearest plant, thus creating our distance_to_nearest_plant variable.

Tableau Maps

To provide a visual aid to our analysis of our data set, we created five maps on Tableau (shown in the Appendix). Figure 1 shows the locations of each coal fired power plant in the US. Figure 2 uses the distance_to_nearest_plant variable to show how far each county centroid is to a power plant (darker red denotes closest to a coal fired power plant, and darker blue denotes farthest to a coal fired power plant). Figures 3, 4, and 5 show percent change in lung, asthma, and respiratory mortality, respectively. For these figures, Tableau used state averages for the distance of each county from a power plant in that state, and the average county distance from a coal plant for each state is the shade of the state. For example, the southeast has a darker shade than the west, which means that people in the southeast, on average, live in a county that is closer to a coal plant than those in the west.

Regressions and Interpretations

Pr1model: This regression looked at the relationship between mortality from respiratory disease and some environmental pollutants and factors (air quality, ozone, NO2), AQI, and distance from plants. At a 90% confidence interval, the significant variables were good days*, moderate days*, unhealthy days for sensitive groups*, and unhealthy days*. At a 95% confidence interval, the only significant variable was the distance to plants. The distance to plant variable has a coefficient of -0.0208, suggesting that as the distance between the county centroid and the operating coal power plant increases by one kilometer, the expected morality by respiration decreases by 0.0208. This indicates that county centroids farther away from coal fired power plants are associated with lower respiratory mortality. At a 99% confidence interval, the overall regression was found significant as well as the nitrogen dioxide (NO2) variable.

Pr2model: This regression looked at the relationship between mortality from respiratory disease and AQI levels, as well as types of days as categorized by AQI levels. Of the variables, 4

were found to be significant at the 95% confidence level. These were "good days", "moderate days", "unhealthy for sensitive group days" and "unhealthy for sensitive group days". The good days and moderate days variables had negative correlation with mortality from respiratory disease, which could be anticipated. Unhealthy days had a positive effect, meaning an anticipated increase in mortality by respiration. Unhealthy for sensitive groups days had an unexpected negative correlation with mortality_resp. The distance_km variable was significant at a 99% confidence level, indicating that holding all other variables constant, the negative effect of increasing distance_km is significant in the model.

Final_mod1: This regression looks only at the effect of distance_km on mortality_resp_2014. The regression found distance_km to be significant above a 99.99% confidence level in the model. However, the R squared value is 0.02148, meaning that 2% of the data fit the model. There are clearly other important indicators of respiratory mortality that are not included in this model.

Smoking_mod2: We used the same regression from final_mod1, but included states and current_smokers. The states with significant outputs were Arizona, Arkansas, Connecticut, DC, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Virginia, Washington, and Wisconsin. Overall, when the state and which one resides has a statistically significant impact on your health outcomes. Current_smokers also had a statistically significant coefficient of 2.677, meaning that as mortality from respiratory disease in 2014 increased, the amount of current smokers increased by 2.677 persons. This shows that many other factors, such as smoking, can have an impact on mortality from respiratory illness. Moderating variables are important to consider when analyzing the relationship between coal fired power plants and health outcomes.

Conclusion

Our analysis of our data set suggests that coal fired power plants may have some significant impacts on one's health depending on proximity to a coal fired power plant. However, we do note that we could not account for variables such as genetics or systemic effects on health outcomes when conducting our research- we recognize that these factors also likely have an effect on health outcomes. Although the mechanism for measuring coal plants' distances from each county centroid works well in R code, it is not a nuanced enough tool to truly investigate health outcomes in different counties. The feature does not take into account where people live in each county. Populations could be more or less dense towards the centroid of the county, and there is no measure to account for this. Measuring the centroids' distance from operating coal plants is not a strong substitute for measuring the counties' densely populated areas distances from coal plants. The feature also does not account for winds or other

geographies. Although mathematically effective, the feature does not necessarily reveal the proximity effect of coal fired power plants in enough detail to make accurate predictions about health outcomes.

Works Cited

Duke Health News. *Despite Studies, Health Effects of Coal-Burning Power Plants
    Remain Unknown*. Duke University and Duke University Health System. 12 Oct. 2018.
    https://surgery.duke.edu/news/despite-studies-health-effects-coal-burning-power-plants-re
    main-unknown.

U.S. Energy Information Administration. *Electricity explained: Electricity in the United States*.
   U.S. Energy Information Administration. 15 July 2022.
    https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php.

Appendix

Coal Fired Power Plants (Operating)

Is Plant

1
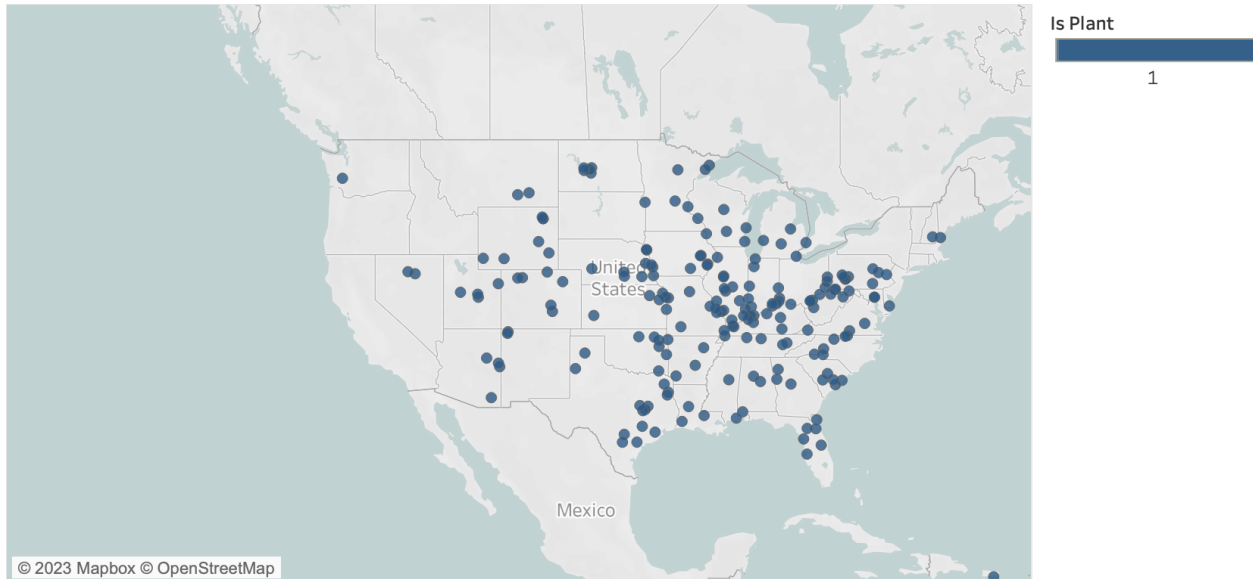
Map based on Long and Lat. Color shows Is Plant as an attribute. The view is filtered on Is Plant as an attribute, which includes values greater than or equal to 1.

Figure 1

Distance To Nearest Plant

Distance To Nearest Pl..

4,243   733,768
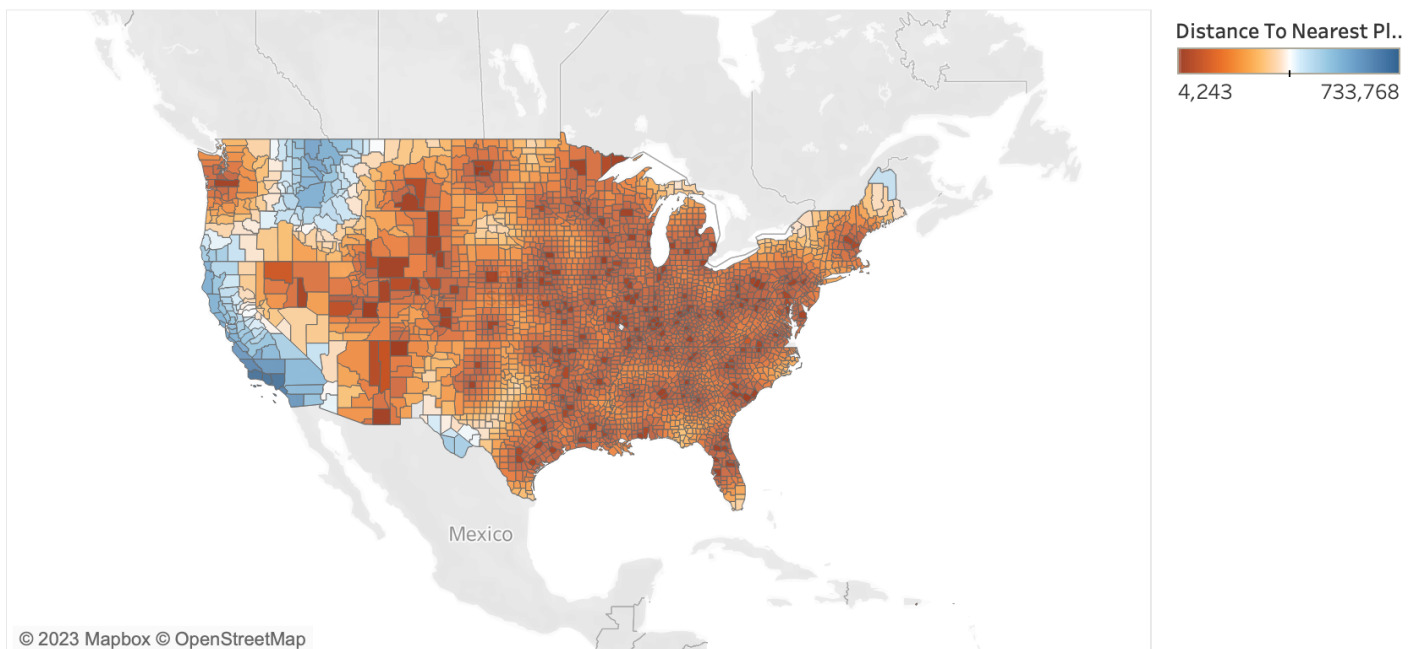
Map based on Long and Lat. Color shows Distance To Nearest Plant as an attribute. Details are shown for State and County.
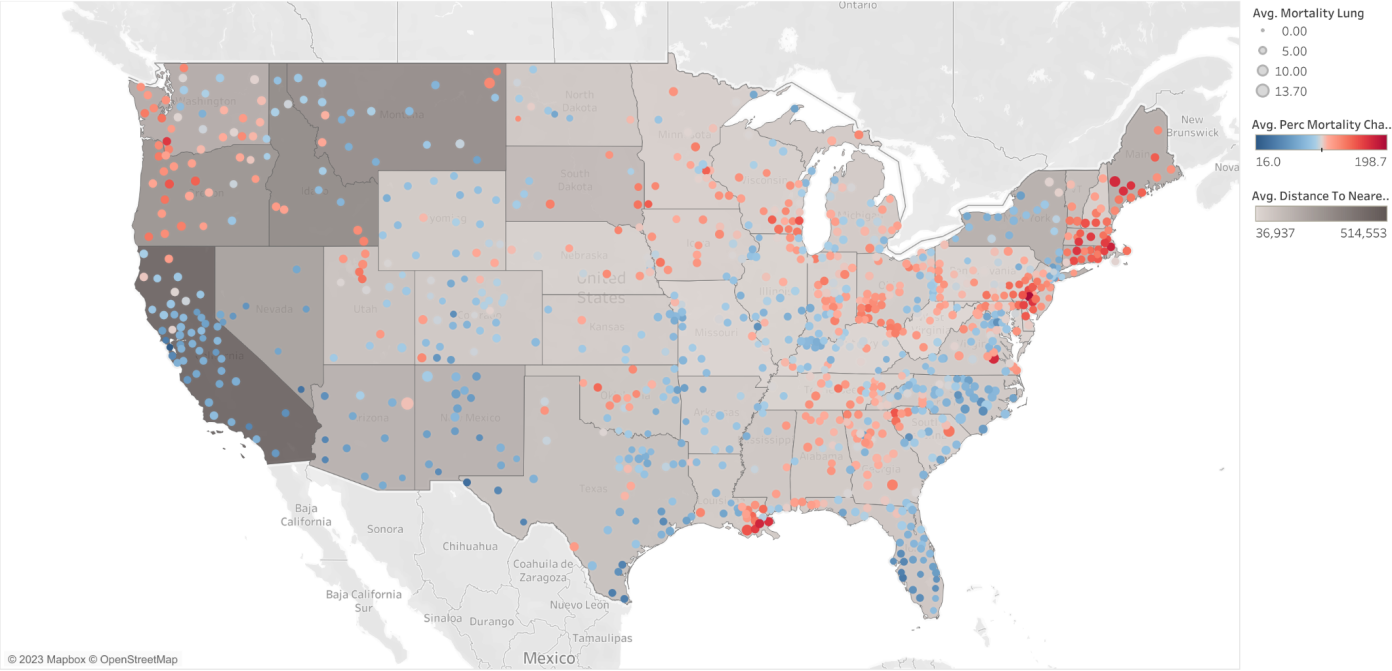
Figure 2

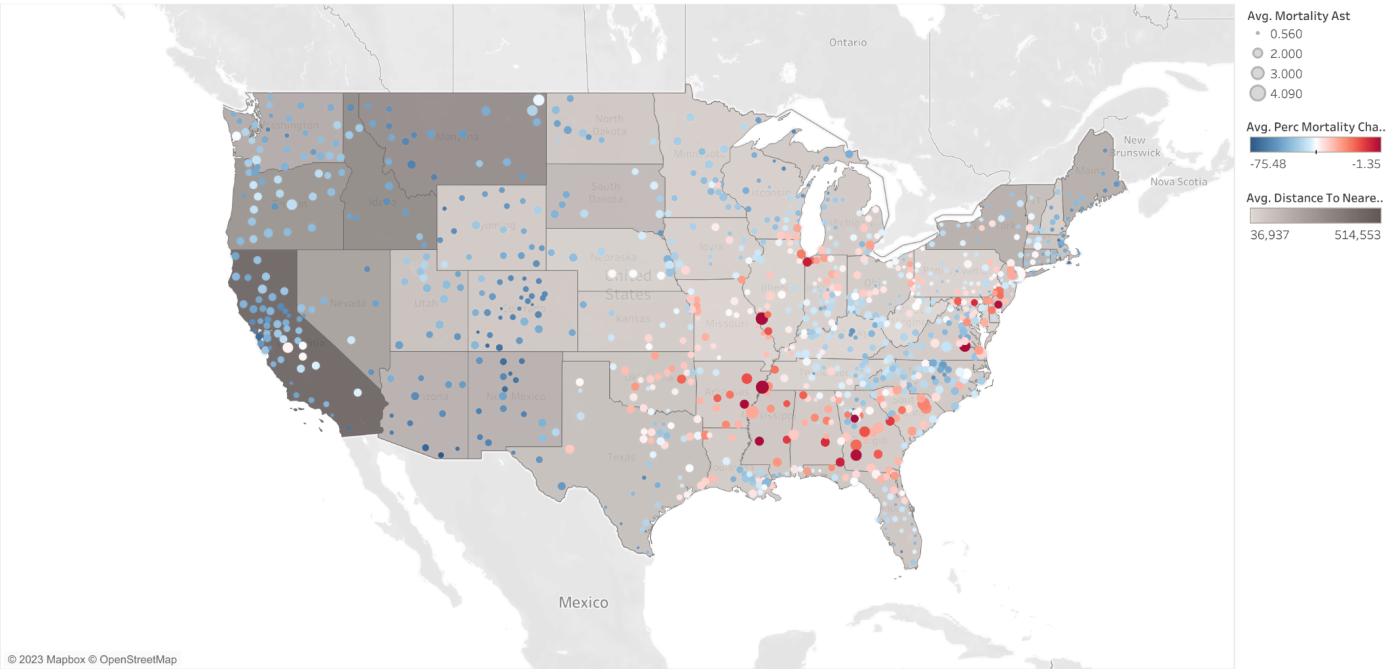Lung Mortality

Figure 3

Asthma Mortality

Figure 4

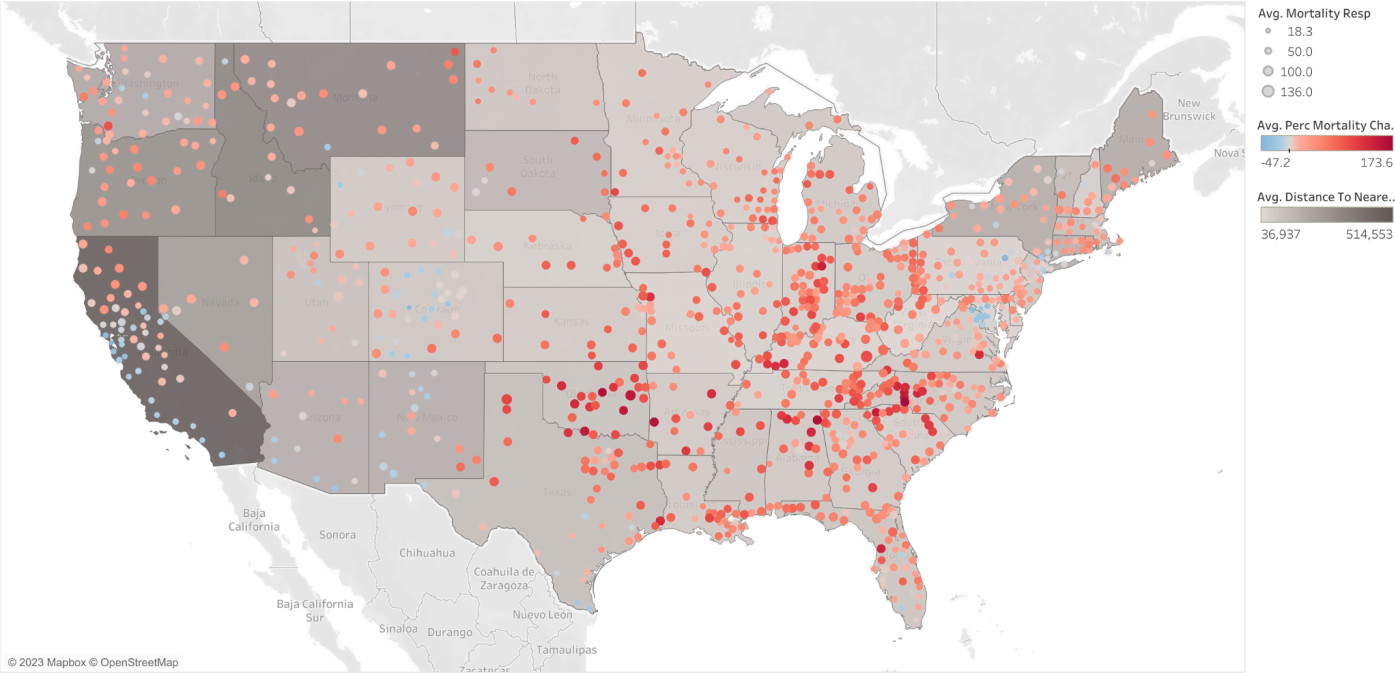**Figure 5**