Exploring Variables Impacting Asthma and Low Birth Weights in California

Ashton Cloer, Elizabeth Good, Emily Kuhlman, Sakina Shahid/

https://github.com/emkuhlmann/KuhlmannGoodShahidCloer_ENV872_EDE_FinalProject

Contents

List of Tables

List of Figures

Rationale and Research Questions

Wildfires are a prevalent risk in the state of California and have been shown to burn nearly half a million acres of land annually. Peak fire events can cause an increase in air pollutants–PM2.5 specifically has been found to increase tenfold after burning (Gupta et al, 2017). PM2.5 concentrations are also affected by socioeconomic status (SES), whereas premature deaths attributable to PM2.5 have been found to increase as household income decreases (Sohrabi et al, 2020). This suggests that as household income decreases, families are unable to afford to reside in areas with cleaner air (Neidell, 2004). PM2.5 is attributed with causing higher incidence rates in asthma. A California based study found that decreases in PM2.5 concentrations significantly contributed towards reductions in asthma rates (Garcia et al, 2019). While recent policies have been passed with the intentions of reducing air pollutants in the state, trends in asthma rates have increased between 1993 and 2014 (Garcia et al, 2019). PM2.5 has also been correlated with the prevalence of term low birth weights (tLBW). Increased exposure to PM2.5 during pregnancy has been shown to increase the chance of tLBW (Ha et al, 2017). In addition to PM2.5, ozone is another criteria air pollutant that can still consistently reach unhealthy levels across the country despite numerous efforts and successes in decreasing its concentration over the past few decades (U.S. EPA, 2023). While ozone has not been shown to significantly contribute to asthma occurrences or other burdens of disease, it is a widely monitored air pollutant and highly reactive compound with health effects including damaging tissue, reducing lung function, and sensitizing the lungs to other irritants (Garcia et al, 2019, Sohrabi et al, 2020, Neidell, 2004).

Our objective is to analyze asthma rates and term low birth weights in California, over the years 2005 to 2019, to determine if there are any trends in the association between our two study objectives and PM2.5 and ozone concentrations. The analysis will investigate three research questions:

1. Have PM2.5 and ozone concentrations changed in California counties from 2005 to 2019?

2. Do PM2.5 and ozone concentrations impact asthma rates in California counties?

3. Do PM2.5 and ozone concentrations impact the incidence of low birth weights in California counties?

A timeseries analysis will be used to answer the first research question. Then, we will create two models to represent how PM2.5 concentrations and ozone concentrations impact the prevalence of both health effects in California counties with wildfire rates and socioeconomic status included as additional explanatory variables.

Dataset Information

Table 1: Information for Low Birthweight Percent, Asthma and Socioeconomic status

Detail	Description
Data Source	CDC National Environmental Public Health Tracking Network
Retrieved From	https://ephtracking.cdc.gov/DataExplorer/
Variables Used	Annual Number of Emergency Department Visits for Asthma,Percent of Low Birthweight (<2500g) Live Singleton Births by county, Percent of the population living in poverty by county
Data Range	2005-2019

Table 2: Information for Wildfire Rates

Detail	Description
Data Source	California Department of Forestry and Fire Protection (CalFire)
Retrieved From	CalFire Wildfire Activity Statistics publications (Redbooks)
Variables Used	Acres burned during large fire events (300 acres and greater) in areas that fall within CalFire direct protection areas
Data Range	2008-2019

Table 3: Information for Ozone and PM 2.5

Detail	Description
Data Source	U.S. EPA Air Data Pre-generated Data Files
Retrieved From	U.S. EPA Air Data Pre-generated Data Files; Annual Summary Data
Variables Used	Ozone 8-hour 2015 Arithmetic Mean and PM25 Annual 2012 Arithmetic Mean
Data Range	2005-2019

Data Wrangling Process

Wildfire Rates - Csv files were downloaded from the US EPA Air Data website for each year from 2015 through 2019. Each file was loaded into R and filtered for the most recent NAAQS standard, the state of California, and the 35 counties of interest. Then the variables of interest were selected, including location information for the air quality monitors, the pollutant standard, unit, and arithmetic mean, and the year. The files were then compiled into one complete data frame. To analyze ozone and PM2.5 data, the compiled data was filtered for the pollutant of interest, grouped by county and year, and a mean pollutant concentration was calculated.

PM 2.5 and Ozone Concentrations - Pdf redbooks of wildfire statistics for each year from 2008 through 2019 were downloaded from CalFire's website. The table of interest (Large Fires 300 Acres and Greater - State and Contract Counties Direct Protection Area) was loaded into Excel using the get data from pdf tool. Once each data table was loaded into its own excel sheet, only the columns for unit (the county in which the fire occurred), year, and acres burned and only the entries for fires that were managed by CalFire were kept. The choice to filter for only fires managed by CalFire was made due to the disclaimer included in the pdf data files, which stated that data was compiled from FEMA incident reports, and other federal or local agencies are not obligated to provide reports to CalFire for annual statistics, making the data an incomplete list of other agency large fires. Files were saved as csv files and were then read into R. For each file, data was grouped by county and year and the annual mean acreage burned was calculated. The files were compiled into one data frame and filtered for the 35 counties of interest.

SES (Socioeconomic status) - The data was downloaded as csv files from the CDC National Public Health Tracking Network Data Explorer and were read into R studio. There were no NAs present in the dataset. The column titled value was renamed to Percent_Poverty and was changed to a numeric column. Column Year and CountyFIPS were changed to factors. Column County and State were changed to characters. California population data was downloaded from our project's Git Repository. Columns for CountyFIPS and Year were changed to factors to match the socioeconomic dataframe. Next, the population and socioeconomic data frame were joined. In the combined data frame, a column was added that used Percent_Poverty to normalize the number of individual's living in poverty per 100k residents. Finally, the counties in the dataframe were reduced to the 35 included across all datasets in the project. Two additional datasets were created for analysis, one that averaged the Percent_Poverty by County and one that averaged the Percent_Poverty by year. All datasets were saved and uploaded to the Git Repository.

Asthma Rates - The asthma dataset was retrieved from the CDC National Environmental Public Health Tracking Network Data Explorer and the CSV file was read into R Studio. There were NAs present in the dataset from Sierra and Alpine counties. These counties were removed from the dataframe. The visits column was changed from factor to numeric and year was changed from integer to factor. Next, the tidycensus R package was used to retrieve county population data for California from 2005-2019. ChatGPT was used to help write the code to create a for loop to fetch the data for each year and create a dataframe including all years. Annual population data was only available for California counties with populations greater than 65,000 residents. This dataframe was wrangled by changing column names to match the asthma dataframe and selecting the CountyFIPS, County, year and Population columns. This processed dataset was saved as a CSV and joined to the asthma dataframe. The number of annual asthma visits was then normalized by the county population for each year and multiplied by 100,000 to calculate the number of emergency department visits for asthma per year per 100,000 residents of each county. Finally, the counties in this dataframe were reduced to only include the 35 counties present across all datasets being used in this analysis.

Low Birthweight - Low Birthweights for California were retrieved from the CDC National Environmental Public Health Tracking Network and the CSV file was read into R Studio. The data was read in to show missing values as NA, and these counties and data points were omitted. County FIPS, State FIPS and Year were read in as factors. Data for years 2000 to 2005 and 2020 was filtered out to match the other datasets. The data was read in as a percent of the population in each California county so it was not normalized for 100k per population that give birth. This file was then filtered to remove counties to match the other datasets.

```
##    |                                                            |
```

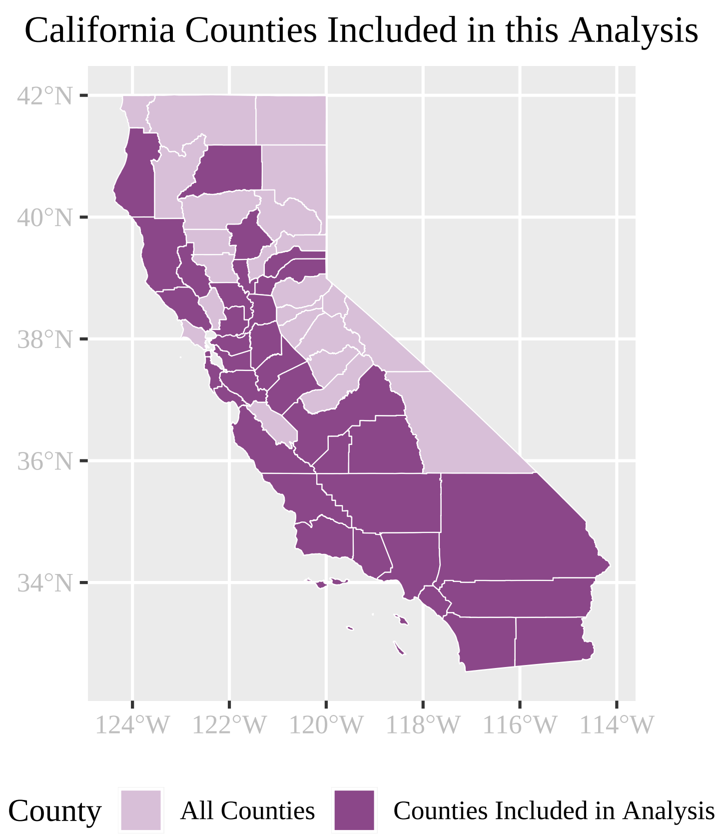


Figure 1: Map of Counties

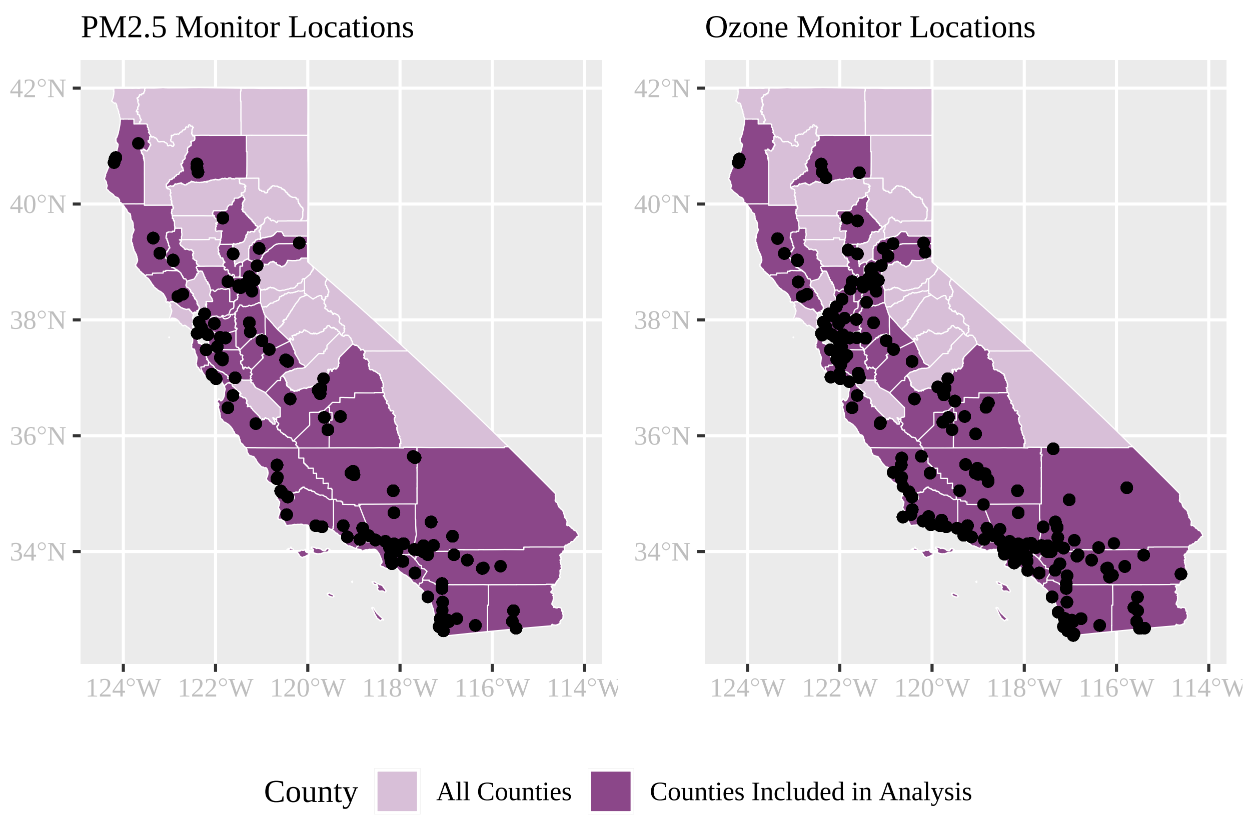


Figure 2: Map of Counties by Monitoring Sites

Exploratory Analysis

To understand the spread of the data, exploratory analysis were conducted on each of the variables utilising various tools

Wildfire Rates

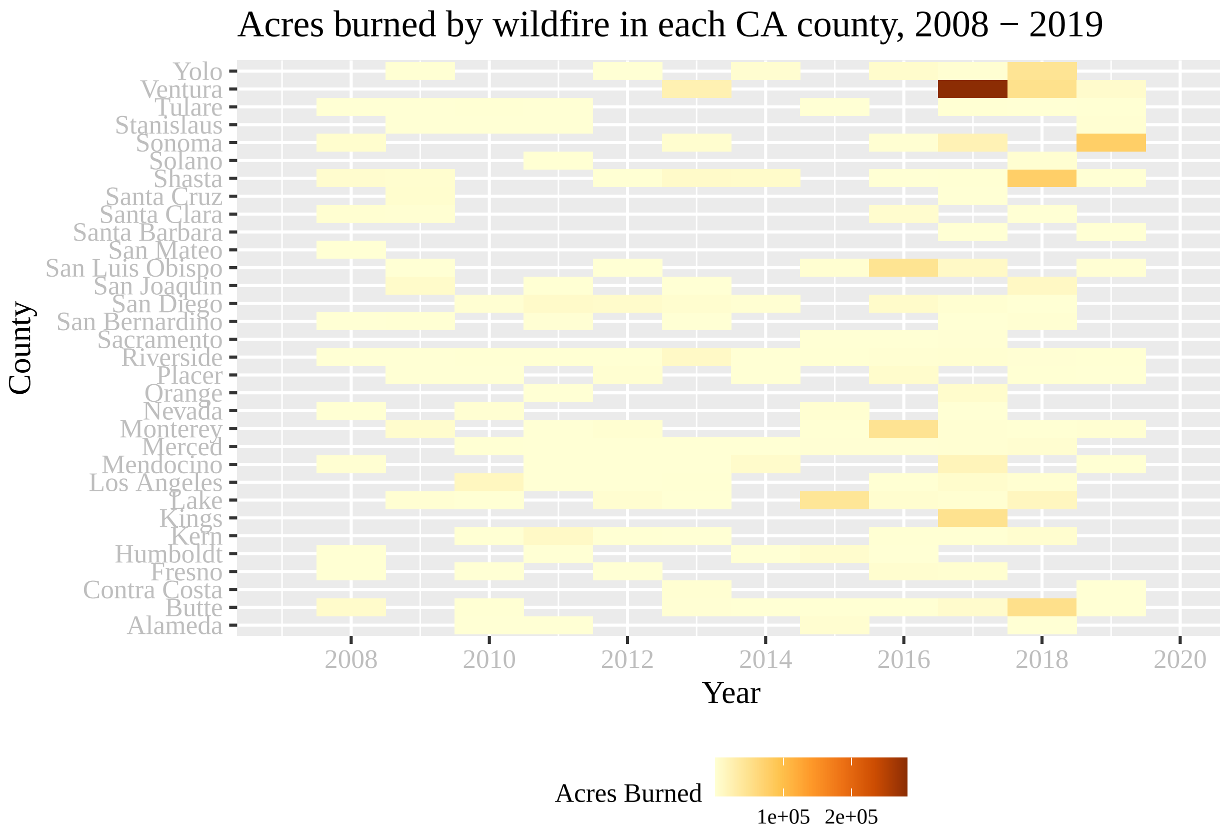


Figure 3: Wildfire Rates Heatmap

The heat map visualizes the average acreage burned in wildfires managed by CalFire for each California county of interest from 2008 through 2019. Blocks missing data had no significant wildfires managed by CalFire in that year, indicating zero acres burned. Ventura county in 2017 has by far the highest acreage burned represented on the map at 281893 acres. Otherwise, the median acreage burned in a county in one year is around 1000 acres.

Ozone and PM 2.5 Concentrations

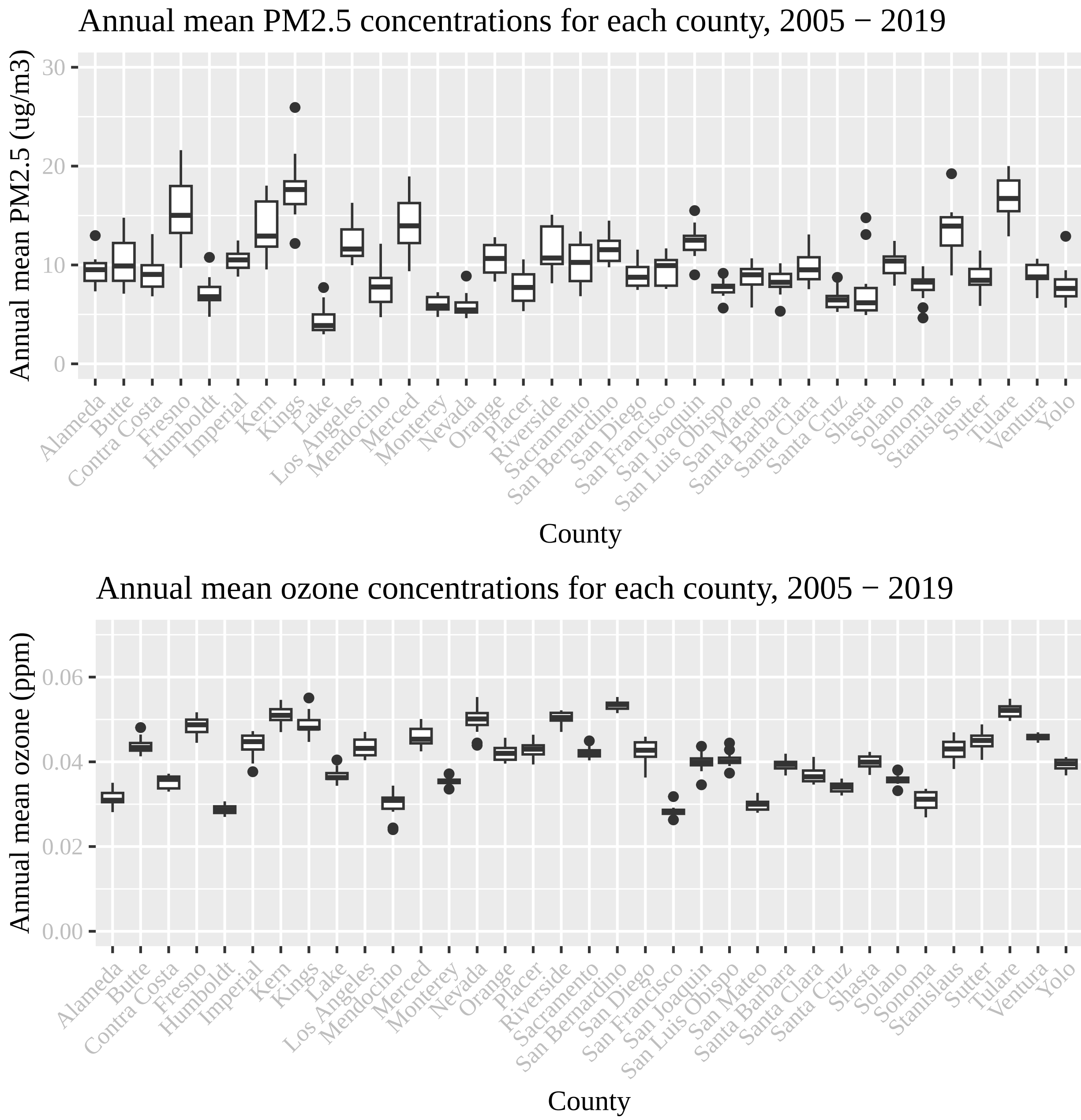
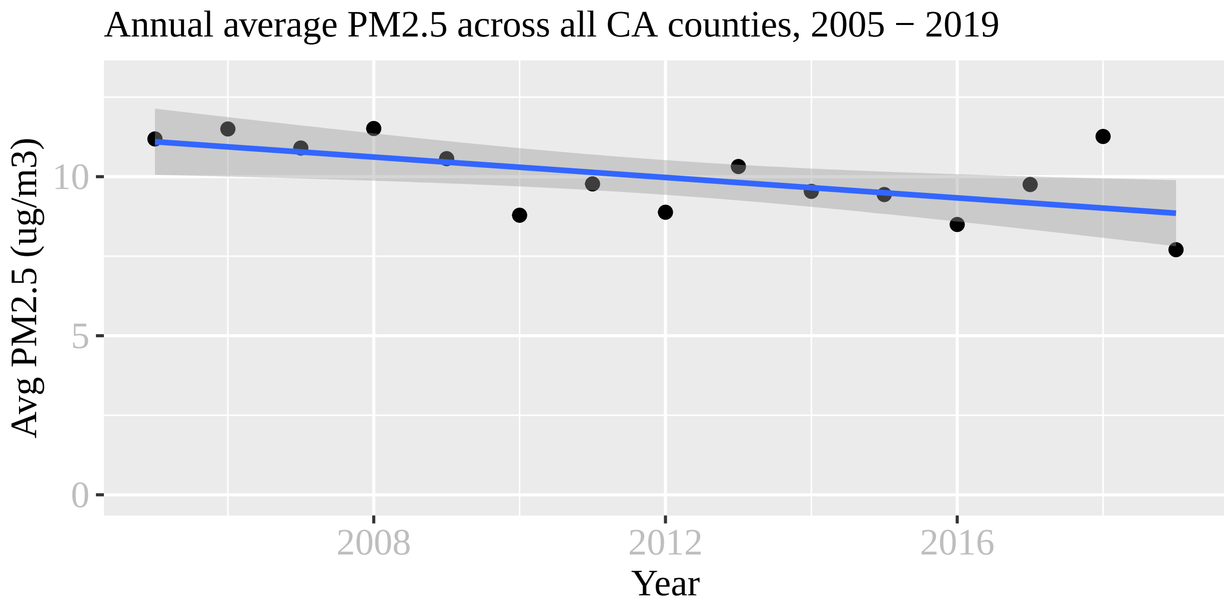


Figure 4: Ozone and PM 2.5 Box and Whiskers Plots Show Annual Summary Statistics by County

For Ozone, following a Tukey HSD test, San Bernardino, Tulare, and Kern counties were grouped as having the highest annual average ozone concentrations ranging from 0.051 to 0.053 ppm. Five counties (San Francisco, Humboldt, San Mateo, Mendocino, and Sonoma) were grouped as having the lowest annual average ozone concentrations ranging from 0.028 to 0.031 ppm.

For PM 2.5, following a Tukey HSD test, Kings, Tulare, and Fresno counties were grouped as having the highest annual average PM2.5 concentrations ranging from 15.7 to 17.8 ug/m3. Monterey, Nevada, and Lake counties were grouped as having the lowest annual average PM2.5 concentrations ranging from 4.4 to 6.0 ug/m3.



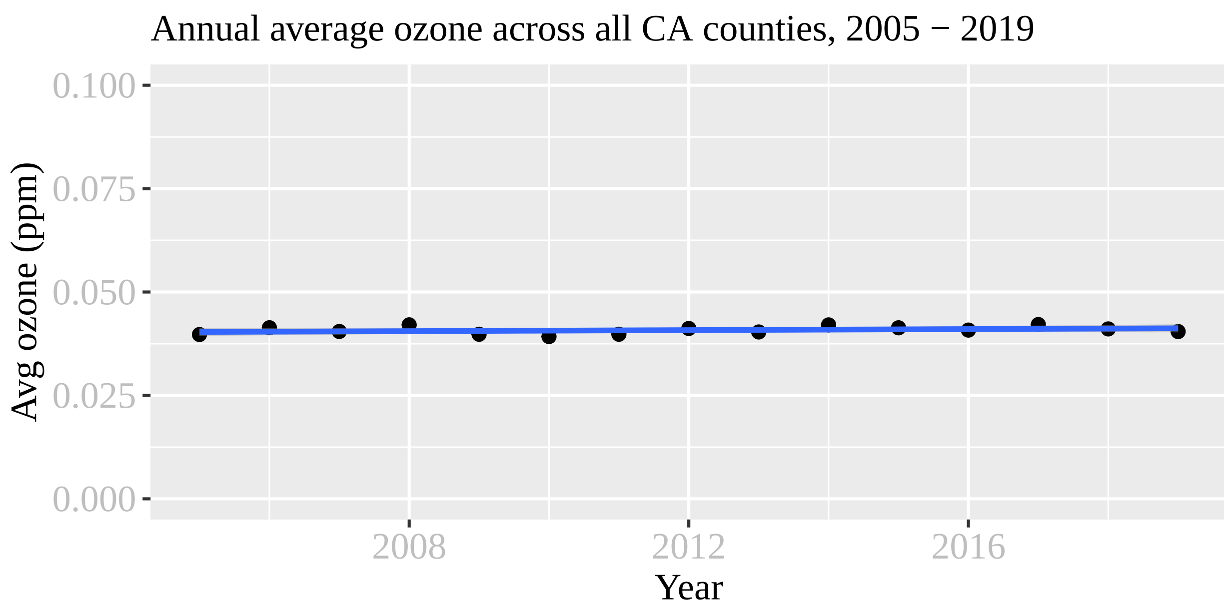


Figure 5: Ozone and PM 2.5 Time Series

Socioeconomic Status

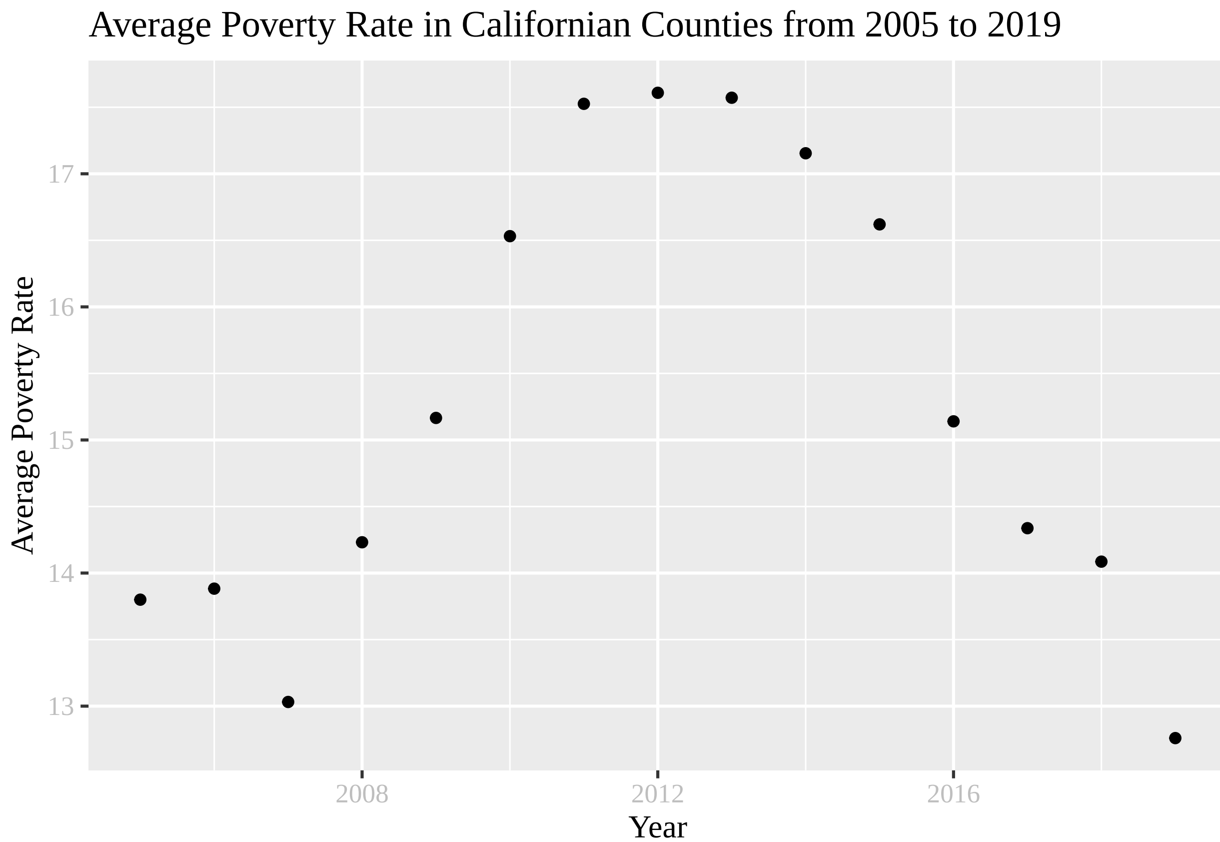


Figure 6: Socioeconomic Status Visualization

The average poverty rate in California counties, between the years 2005 to 2019, showed a quadratic pattern. The average poverty rate peaked from 2011 to 2013 at over 18% and reached its lowest in 2019 at 12.5%.

There was a significant difference between the mean poverty rate between the 35 test counties in California. Tulare county stood out at having the highest poverty rate at 24.47% while San Mateo had the lowest at 7.15%, both were statistically significant.

Asthma Rates

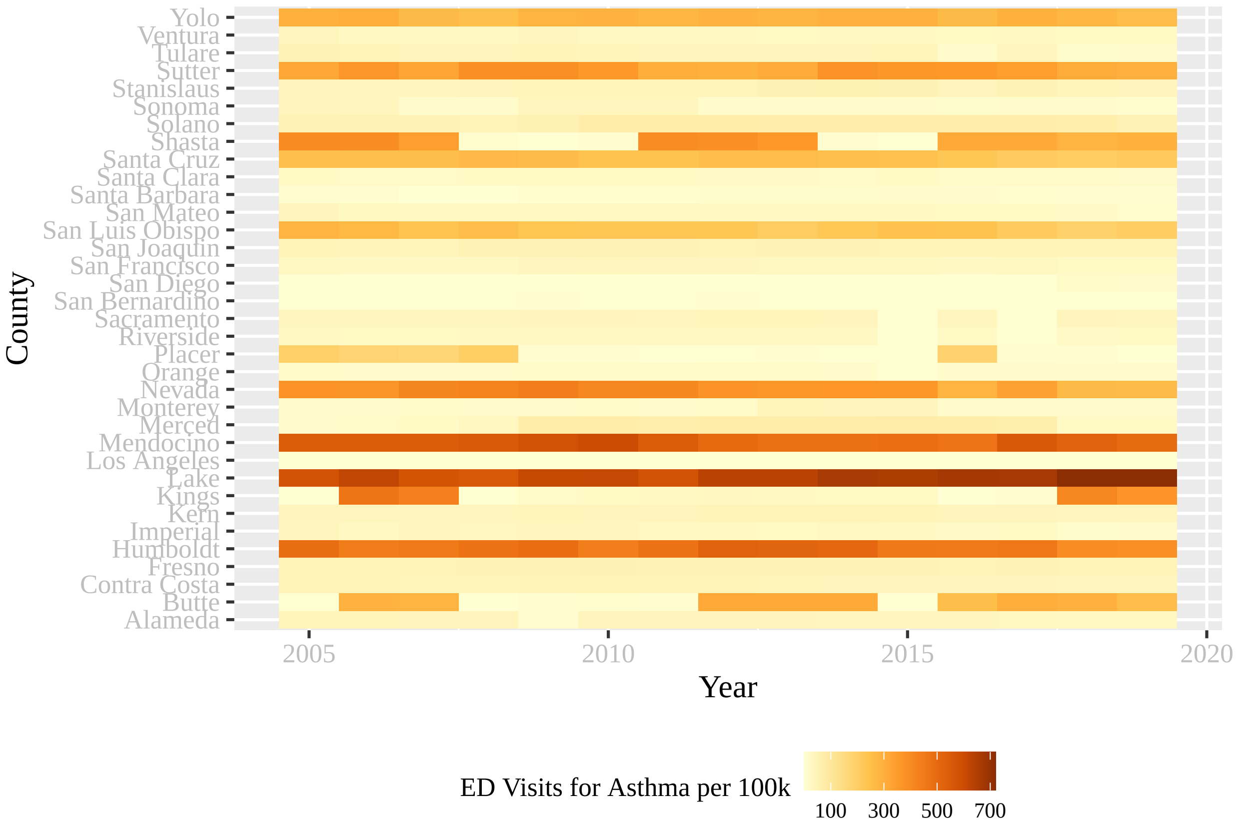


Figure 7: Asthma Rates Heatmap

From a Tukey HSD test analyzing emergency department visits for asthma by California county, Lake County stands out as the county with the highest average ED visit rate for asthma of 634 per 100k residents. 23 of the 35 counties included in the analysis were grouped together to have a statistically non-significant difference between their mean ED visit rates for asthma ranging from 4 to 67 ED visits per 100k residents.

The heatmap shows that most counties have annual emergency department visits for asthma below 300 per 100k residents. Three counties, Mendocino, Humboldt, and Lake, appear to have consistently higher numbers of ED visits for asthma for all years included in the analysis. Most counties have fairly consistent numbers of visits across the study period. However, Shasta, Kings, and Butte counties have periods where the annual visits jump up by several hundred, and there is some overlap in the periods of these jumps between the three counties.

Low Birthweight

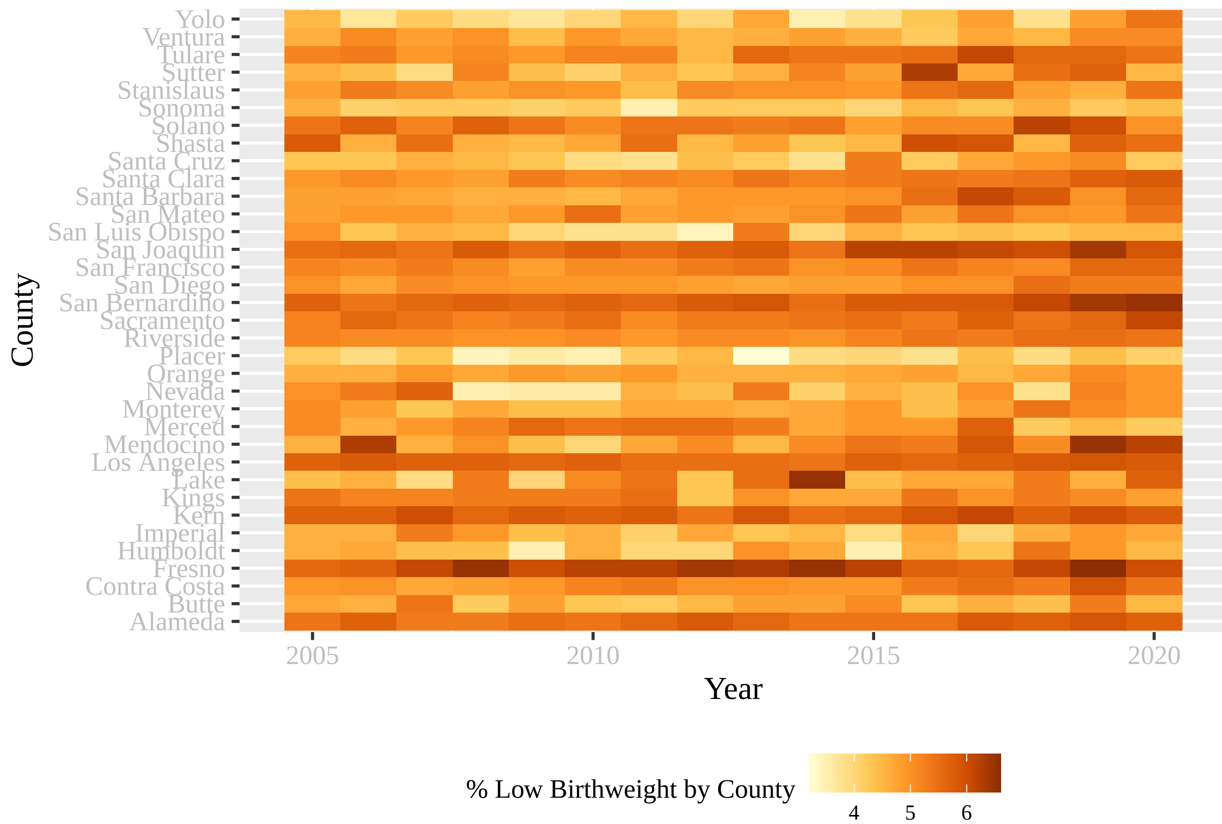


Figure 8: Low Birthweight Heatmap

There was minimal difference between percentages of low birth weight births among California counties according to a Tukey HSD test. However, Fresno County stood out as the county with the highest percent of births at a low birth weight (6.1%), and this was statistically significantly higher than the other 34 counties analyzed. Placer County stood out as the county with the lowest percent of births at a low birth weight (3.9%), which was statistically significantly lower than the other 34 counties.

The heatmap shows the change in % of Low Birthweight per county throughout the years 2005-2019. % Low Birthweight is reported based on a persons' residence address and not where they were hospitalized. Some counties like Fresno, San Joaquin, Bernardino and Solano have high % of Low Birthweights. Placer has a relatively low amount of % low birthweights. Overall, there is variability in low birthweight for all counties for all years but a general increase in % low birthweight rates in the time period 2016-2020.

Analysis

Multiple Linear Regression Models

Table 4: Estimating Variation in ER Visits for Asthma

Predictor	Estimate	Std. Error	Test Statistic	p-value
Intercept	387.79	67.790	5.72	0.000
Mean Annual Ozone	-7367.43	1770.141	-4.16	0.000
Mean Annual PM2.5	-26.79	3.772	-7.10	0.000
Percent Poverty	18.92	2.302	8.22	0.000
Acres Burned	0.00	0.000	1.51	0.132

Table 5: Estimating Variation in Low Birth Weight Rates

Predictor	Estimate	Std. Error	Test Statistic	p-value
Intercept	3.80	0.294	12.92	0.000
Mean Annual Ozone	9.92	7.676	1.29	0.198
Mean Annual PM2.5	0.07	0.016	4.14	0.000
Percent Poverty	0.01	0.010	0.93	0.356
Acres Burned	0.00	0.000	-1.87	0.063

Question 1: Have ozone and PM2.5 concentrations changed in California counties from 2005 to 2019?

Null Hypothesis: Ozone and PM2.5 concentrations have not changed significantly in California counties from 2005 to 2019. Alternative Hypothesis: Ozone and PM2.5 concentrations have changed significantly in California counties from 2005 to 2019.

Based on statistical models, we conclude the following:

There has been no statistically significant trend in ozone concentrations across the California counties of interest from 2005 through 2019 (Mann-Kendall p value = 0.37).

There has been a statistically significant trend in PM2.5 concentrations across the California counties of interest from 2005 through 2019 (Mann-Kendall p value = 0.023). A one year increase is associated with a 0.16 ug/m3 decrease in PM2.5 concentrations (p value $<$ 0.05).

Question 2: Do PM2.5 and ozone concentrations impact asthma rates in California counties?

Null Hypothesis: Ozone and PM2.5 concentrations have not changed significantly in California counties from 2005 to 2019. Alternative Hypothesis: Ozone and PM2.5 concentrations have changed significantly in California counties from 2005 to 2019.

Based on statistical models, we conclude the following: A one ppm decrease in ozone concentration is associated with a 7367 unit increase in ER visits for asthma per 100k ($p < 0.001$).

A one ug/m3 decrease in PM2.5 concentration is associated with a 27 unit increase in ER visits for asthma per 100k ($p < 0.001$).

A one unit increase in percent poverty is associated with a 19 unit increase in ER visits for asthma per 100k ($p < 0.001$).

Question 3: Do PM2.5 and ozone concentrations impact the incidence of low birth weights in California counties?

Null Hypothesis: PM 2.5 and Ozone concentrations do not impact birth weights in California counties. Alternative Hypothesis: PM 2.5 and Ozone concentrations impact birth weights in California counties.

Based on statistical models, we conclude the following:

A one ug/m3 increase in PM2.5 concentration is associated with a 0.07 unit increase in percent underweight births ($p < 0.001$).

Summary and Conclusions

The outputs of the multiple linear regression models, particularly for emergency department visits for asthma, suggest the need for further evaluation in future analyses. Previous literature suggests that increases in air pollution concentrations, especially particulate matter, increase the prevalence of asthma related episodes and ED visits/hospitalizations. The model in this analysis gives the opposite trend, and a large standard error value for ozone as a predictor indicates additional exploration may be needed. There could be many more potential predictors to include in the model, such as residence (urban vs rural, proximity to pollution point sources), and medical history factors. Pollution levels can also fluctuate drastically on short time scales, so annual averages may not be the best method for this analysis. If higher time resolution data could be obtained for asthma-related ED visits, air pollution data down to hourly averages could be analyzed for potentially more accurate results.

References

Garcia, E., Berhane, K., Islam, T., McConnell, R., Urman, R., Zhanghue, C., Gilliland, F. 2019. Association of changes in air quality with incident asthma in children in California, 1993-2014. JAMA Network, 321(19). https://jamanetwork.com/journals/jama/article-abstract/2733972.

Gupta, P., Doraiswamy, P., Levy, R., Pikelnaya, O., Maibach, J., Feenstra, B., Polidori, A., Kiros, F., Mills, C. 2018. Impact of California fires on local and regional air quality: the role of a low-cost sensor network and satellite observations. GeoHealth, 2(6). https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018GH000136.

Ha, S., Zhu, Y., Liu, D., Sherman, S., and Mendola, P. 2017. Ambient temperature and air quality in relation to small for gestational age and term low birthweight. Environmental Research, 155. https://www.sciencedirect.com/science/article/pii/S0013935116307332?casa_token=s4b6BG9ZTAoAAAAA:V3seg2IniGknqc3bX9z4FukGwF0NexfAVv--VGpyzrdjQAYYRcRIs1YfHVzYuScyPLmue5xErcA.

Neidell, M. 2004. Air pollution, health, and socioeconomic status: the effect of outdoor air quality on childhood asthma. Journal of Health Economics, 23(6). https://www.sciencedirect.com/science/article/pii/S0167629604000864?casa_token=6is2hlKvma4AAAAA:iCgz317RhRJVyQTcX--eg0qeWlnxWgf4u8eWko1hJLIigxlrmwdsNXsb0k5k6OxvMH--NKy3KTM.

Sohrabi, S., Zietsman, J., and Khreis, H. 2020. Burden of disease assessment of ambient air pollution and premature mortality in urban areas: the role of socioeconomic status and transportation. International Journal of Environmental Research of Public Health, 17(4). https://www.mdpi.com/1660-4601/17/4/1166.

U.S. EPA. 2023. Air Pollution: Current and Future Challenges. https://www.epa.gov/clean-air-act-overview/air-pollution-current-and-future-challenges