

The Effects of Air Quality on U.S. Life Expectancy

Michael T. Moen

2024-11-22

Contents

1	Abstract	2
2	Background	3
2.1	Data Sources	3
2.2	Ground-Level Ozone	3
2.3	Fine Particulate Matter	4
3	Methodology	5
3.1	Data Preparation	5
3.2	Predicting Mortality Due to Cardiovascular Factors from Air Quality	5
4	Results	5
5	Conclusion	5
6	Appendices	6
6.1	Apendix A: Merging Datasets with Python	6

1 Abstract

Write last once background, methodology, and results are all established.

2 Background

Ground-level ozone and fine particulate matter are two of the most significant air pollutants that impact public health. This analysis explores this issue by looking at the relationship of these two pollutants with the life expectancy of U.S. counties due to various mortality factors.

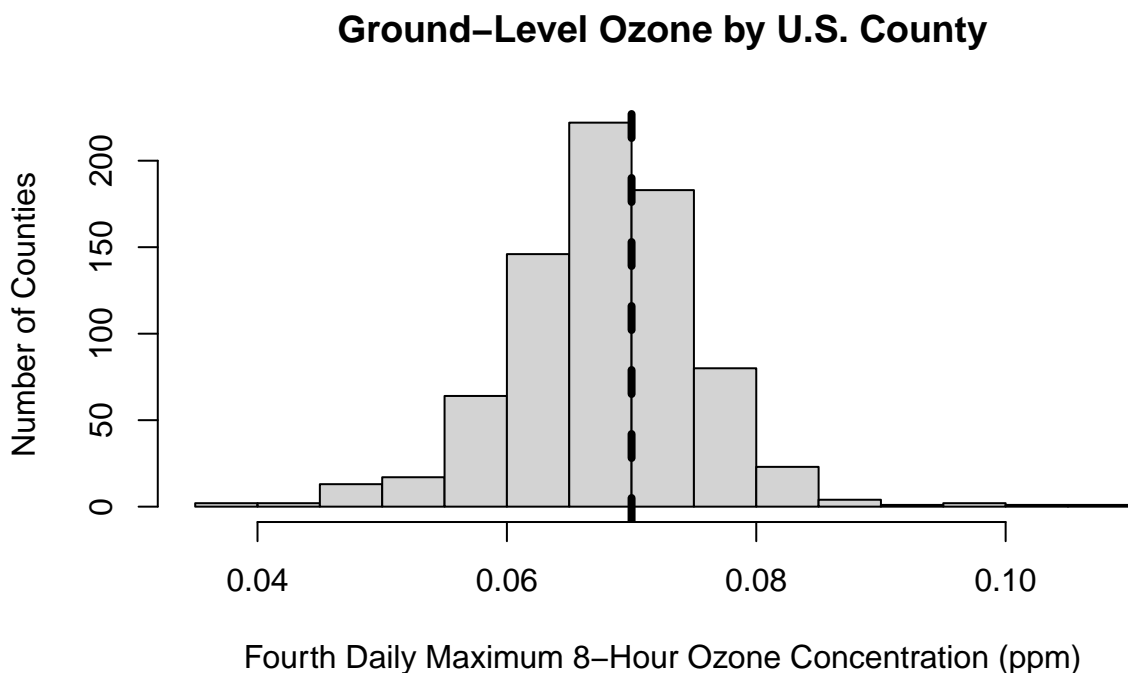
2.1 Data Sources

The data used in this analysis was collected from two main sources. The mortality data comes from the Institute for Health Metrics and Evaluation (IHME) and contains data for various mortality factors, with the most recent data being from 2014. Since this data is derived from death registration data, it is not missing data for any counties.

The air quality data in this analysis was published by the Environment Protection Agency (EPA) and covers the presence of various pollutants recorded in the United States by county in 2023. Since this methodology only allows the EPA to monitor the pollutants in counties with the appropriate facilities to record such data, it is much more sparse. Only about one-third of counties have any recorded data, and the counties that do have recorded data typically only monitor select pollutants. However, the two pollutants that have the most significant impact on public health, ground-level ozone and fine particulate matter, are the most widely recorded pollutants. Because of this, these pollutants are the focus of this analysis.

2.2 Ground-Level Ozone

Ground-level ozone is a pollutant that has been shown to have negative effects on the health of individuals, particularly those with respiratory issues such as asthma.

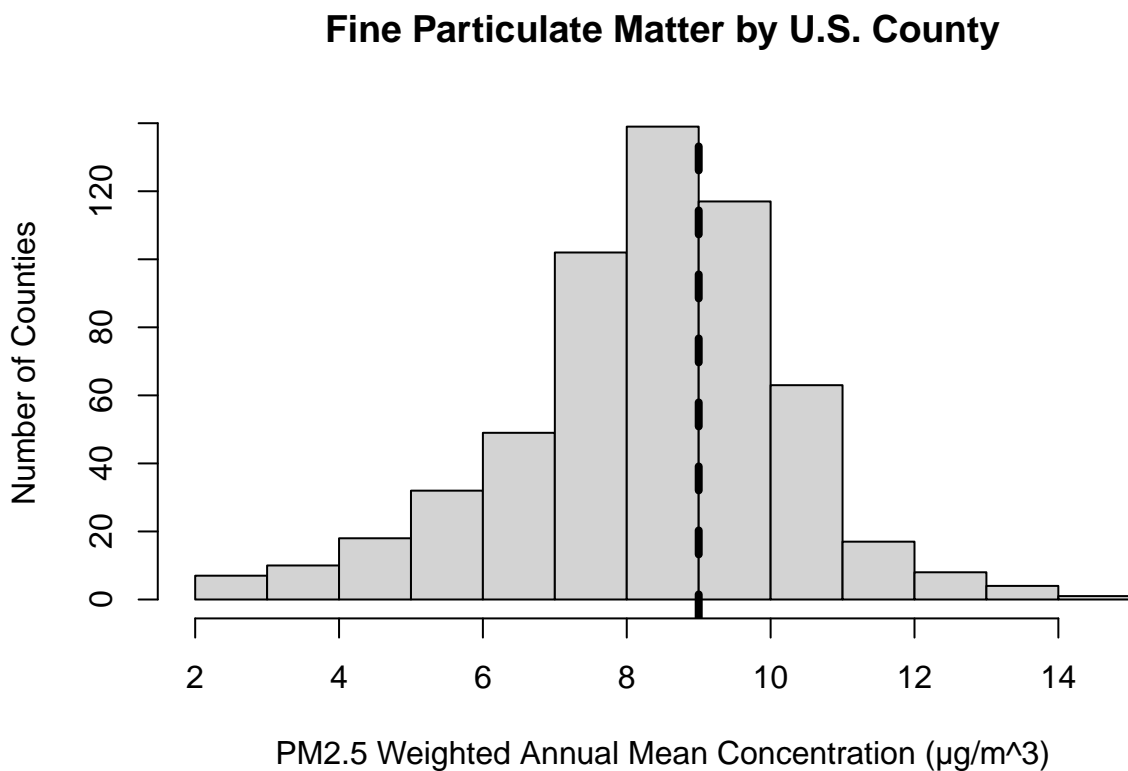


The histogram shows the severity of the ground-level ozone recorded by the EPA in 2023 for each county. Note that the EPA only recorded this data in 761 of the 3143 counties and county-equivalents (parishes in Louisiana, boroughs in Alaska, etc.) in the United States.

The EPA's National Ambient Air Quality Standards (NAAQS) sets the threshold for the ground-level ozone at 0.07 parts per million (ppm), which is represented by the vertical line in the histogram. The high number of counties above this threshold show that this pollutant has an unhealthy presence in many places in the United States.

2.3 Fine Particulate Matter

Fine particulate matter (PM_{2.5}) refers to particles with a diameter of 2.5 micrometers or smaller. These fine particles can be made up of anything, with some of the notable sources of PM_{2.5} including vehicle emissions, industrial activity, and wildfires. Like ground-level ozone, fine particulate matter has been shown to have negative impact on health, particularly among those with pre-existing respiratory issues.



The histogram shows the severity of the fine particulate matter (PM_{2.5}) recorded by the EPA in 2023 for each county. Note that the EPA only recorded this data in 562 of the 3143 counties and county-equivalents (parishes in Louisiana, boroughs in Alaska, etc.) in the United States.

The NAAQS sets the threshold for the fine particulate matter at 9.0 µg/m³, which is represented by the vertical line in the histogram. The high number of counties above this threshold show that this pollutant has an unhealthy presence in many places in the United States.

3 Methodology

3.1 Data Preparation

The two datasets used in this analysis were found online as CSV and XLSX files. These file were downloaded and joined using Python's pandas library using the the county FIPS codes (see Appendix A for details). This merged dataset was then imported into R for analysis.

Since the air quality data does not cover every U.S. county, only the cases that contain both

3.2 Predicting Mortality Due to Cardiovascular Factors from Air Quality

4 Results

5 Conclusion

Summarize findings and explain importance (or lack thereof) of findings. Write once finished with other sections

6 Appendices

6.1 Appendix A: Merging Datasets with Python

```
import pandas as pd

# Read and clean mortality data
mortality_df = pd.read_csv('mort.csv')
mortality_df = mortality_df.dropna(subset=['FIPS'])
mortality_df['FIPS'] = mortality_df['FIPS'].astype(int)

# Read and clean air quality data
aqi_df = pd.read_excel('ctyfactbook2023.xlsx',
                       sheet_name='County Factbook 2023',
                       skiprows=2)
aqi_df = aqi_df.dropna(subset=['County FIPS Code'])
aqi_df['County FIPS Code'] = aqi_df['County FIPS Code'].astype(int)

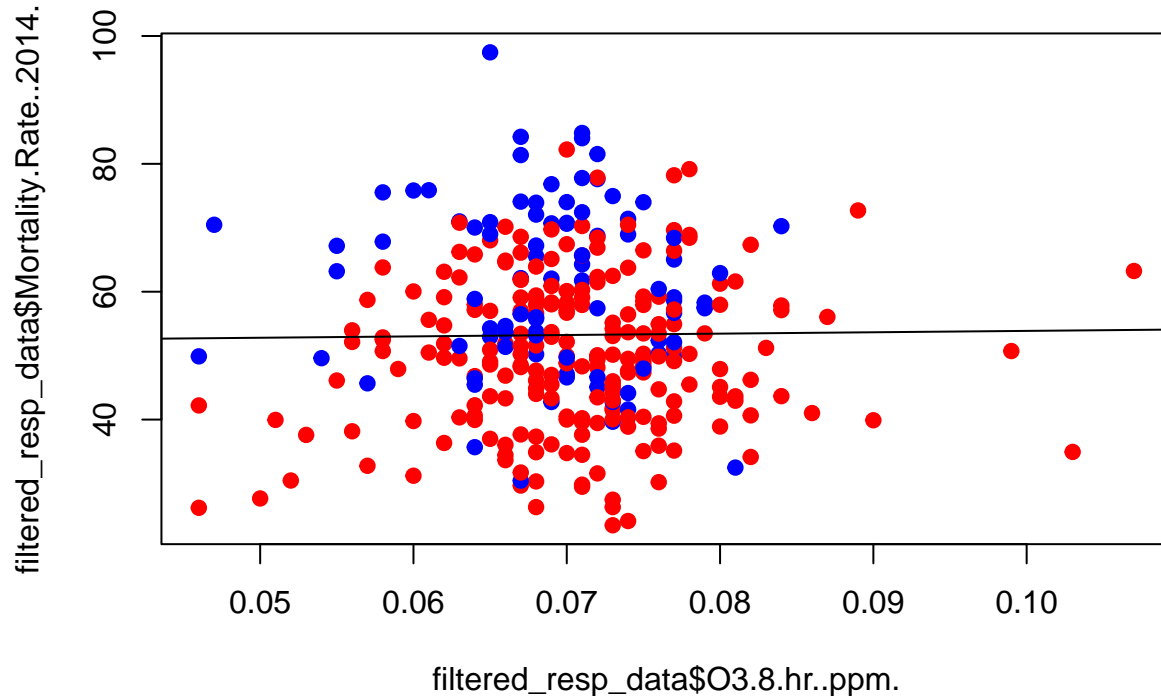
# Join datasets on FIPS codes
merged_df = pd.merge(mortality_df, aqi_df,
                     left_on='FIPS', right_on='County FIPS Code', how='inner')

# Export merged dataframe to CSV
merged_df.to_csv('air_quality_mortality.csv', index=False)
```

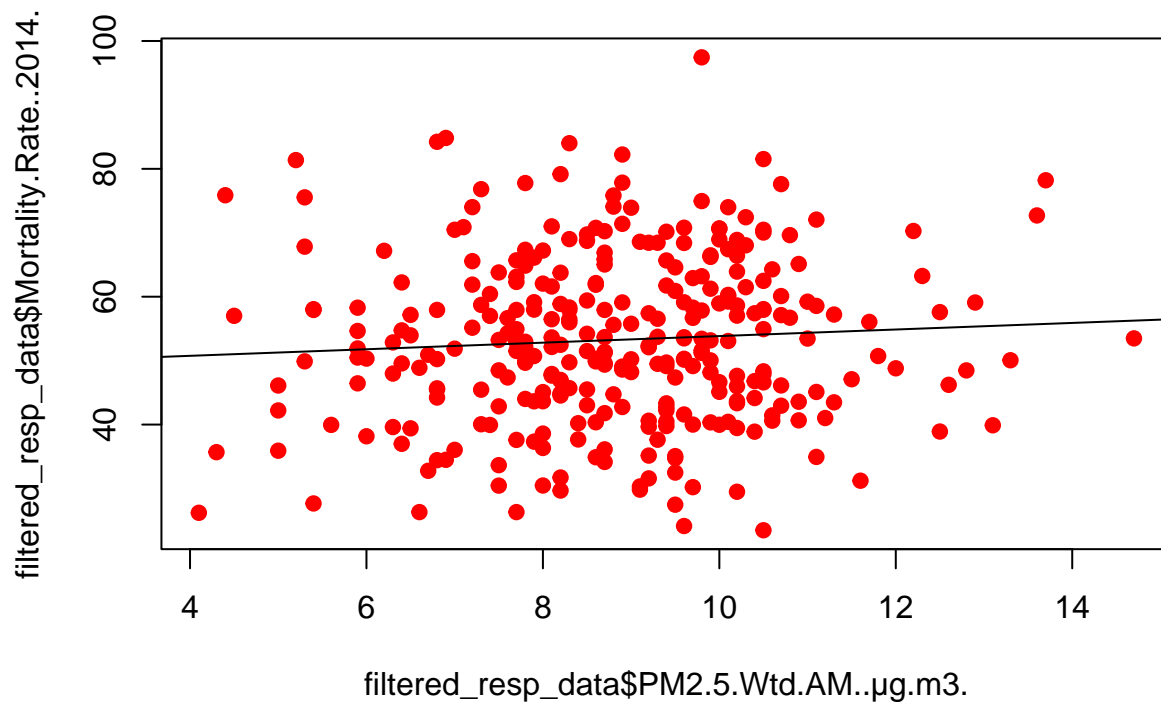
The two datasets used in this analysis were joined using Python's pandas library. Note that some columns were dropped and renamed for ease of use, but this was omitted from this code block.

```
##
## Call:
## lm(formula = Mortality.Rate..2014. ~ 03.8.hr..ppm., data = filtered_resp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.821  -9.470  -0.490   8.845  44.320
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    51.726     6.678   7.745 1.36e-13 ***
## 03.8.hr..ppm.    21.434    94.632   0.226   0.821
## ---
```

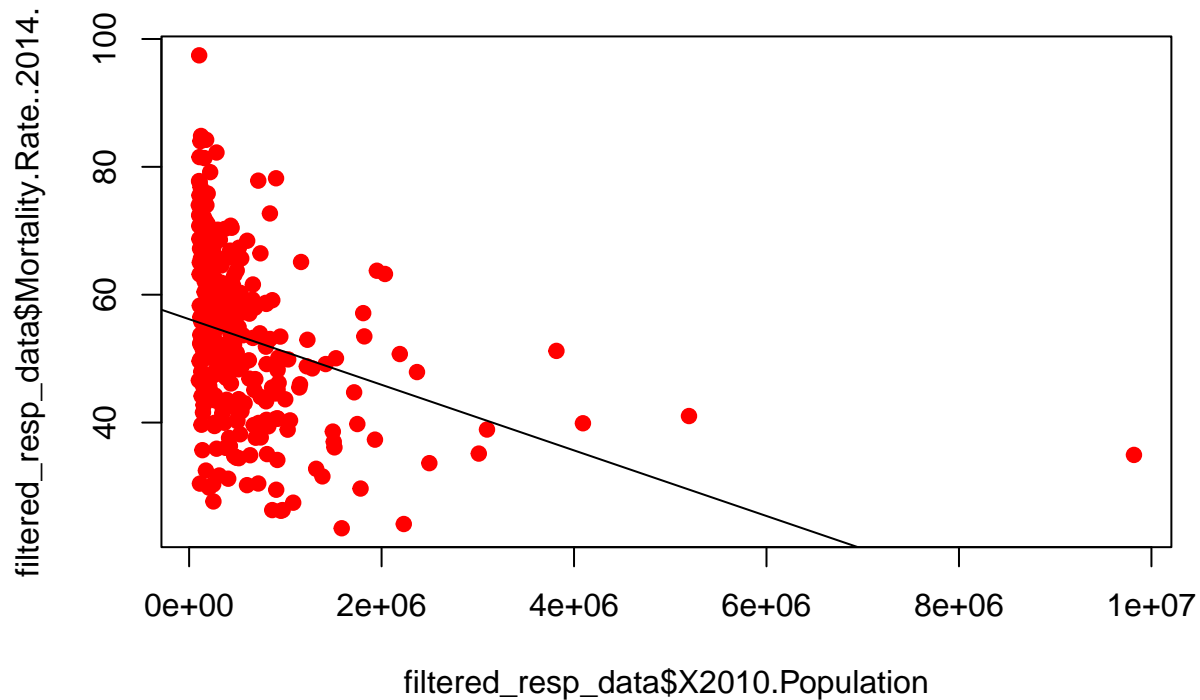
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.14 on 311 degrees of freedom
## Multiple R-squared:  0.0001649, Adjusted R-squared:  -0.00305
## F-statistic: 0.0513 on 1 and 311 DF,  p-value: 0.821
```



```
##
## Call:
## lm(formula = Mortality.Rate..2014. ~ PM2.5.Wtd.AM..µg.m3., data = filtered_resp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30.625  -9.159  -0.616   8.753  43.705
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    48.6927     3.7791  12.885  <2e-16 ***
## PM2.5.Wtd.AM..µg.m3.  0.5145     0.4202   1.224   0.222
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.11 on 311 degrees of freedom
## Multiple R-squared:  0.004797, Adjusted R-squared:  0.001597
## F-statistic: 1.499 on 1 and 311 DF,  p-value: 0.2218
```



```
##
## Call:
## lm(formula = Mortality.Rate..2014. ~ X2010.Population, data = filtered_resp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.218  -9.031  -0.680   7.677  41.803
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.617e+01  8.636e-01  65.045  < 2e-16 ***
## X2010.Population -5.131e-06  8.710e-07  -5.891  9.94e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.46 on 311 degrees of freedom
## Multiple R-squared:  0.1004, Adjusted R-squared:  0.09751
## F-statistic: 34.71 on 1 and 311 DF,  p-value: 9.941e-09
```

```
##
## Call:
## lm(formula = Mortality.Rate..2014. ~ 03.8.hr..ppm. + PM2.5.Wtd.AM..µg.m3. +
##     X2010.Population, data = filtered_resp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.046  -9.013  -0.504   7.884  41.474
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.761e+01  6.597e+00   5.702 2.77e-08 ***
## 03.8.hr..ppm.    1.782e+02  1.116e+02   1.596   0.111
## PM2.5.Wtd.AM..µg.m3.  7.569e-01  4.709e-01   1.607   0.109
## X2010.Population  -6.199e-06  9.341e-07  -6.636 1.44e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.3 on 309 degrees of freedom
## Multiple R-squared:  0.1297, Adjusted R-squared:  0.1213
## F-statistic: 15.35 on 3 and 309 DF,  p-value: 2.462e-09
```