Analysis of Air Pollution and Lung Cancer Rates

Bailey Bjornstad
Ahmed Safdar

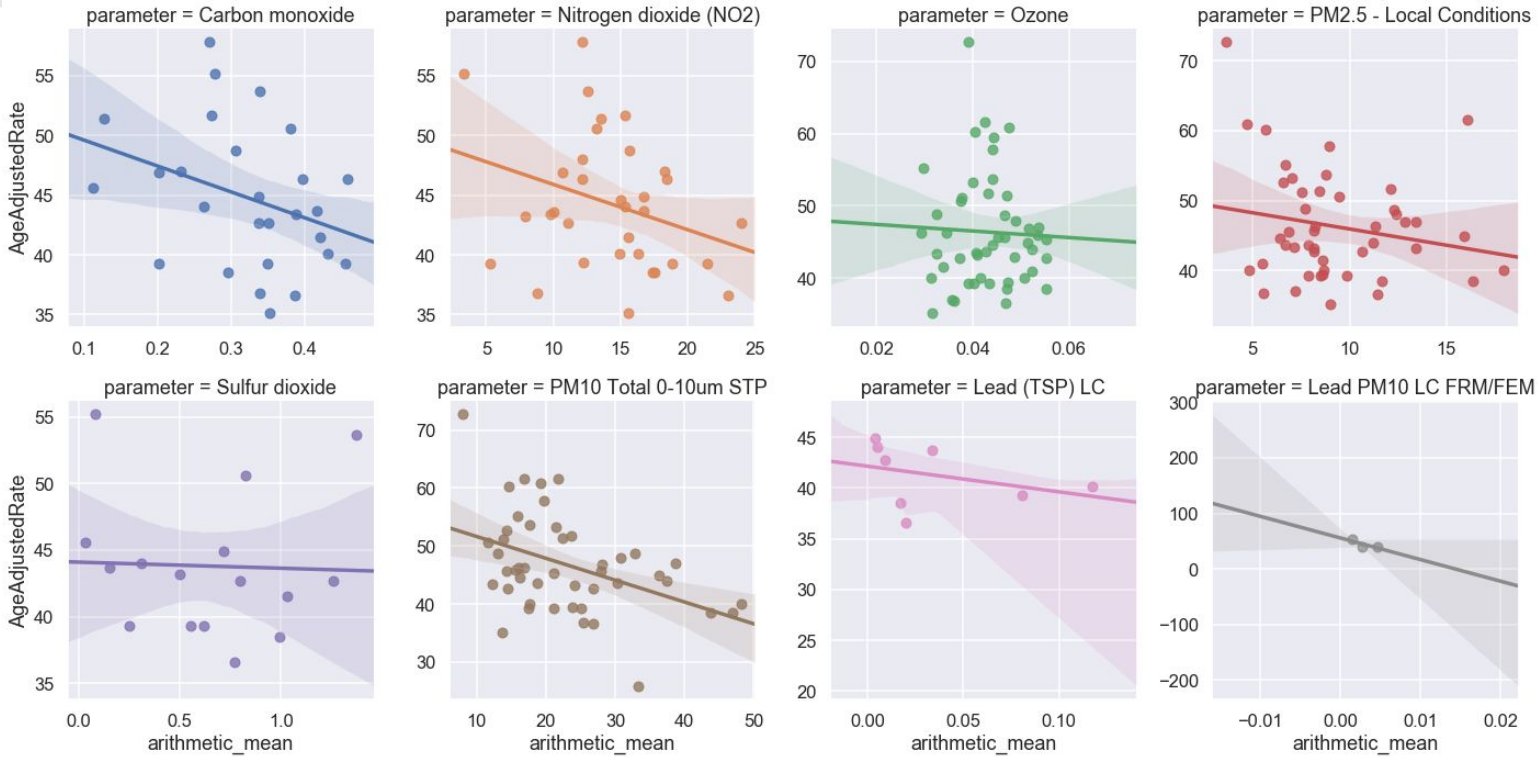
Purpose

- Determine if there is a correlation between levels of air pollution and rates of lung cancer
- Hypothesis:
  - Higher levels of air pollution result in increased incidence of lung cancer.
- Scenario: EPA or other regulatory agency is trying to identify highly problematic pollutants in order to curb rates of lung cancer.

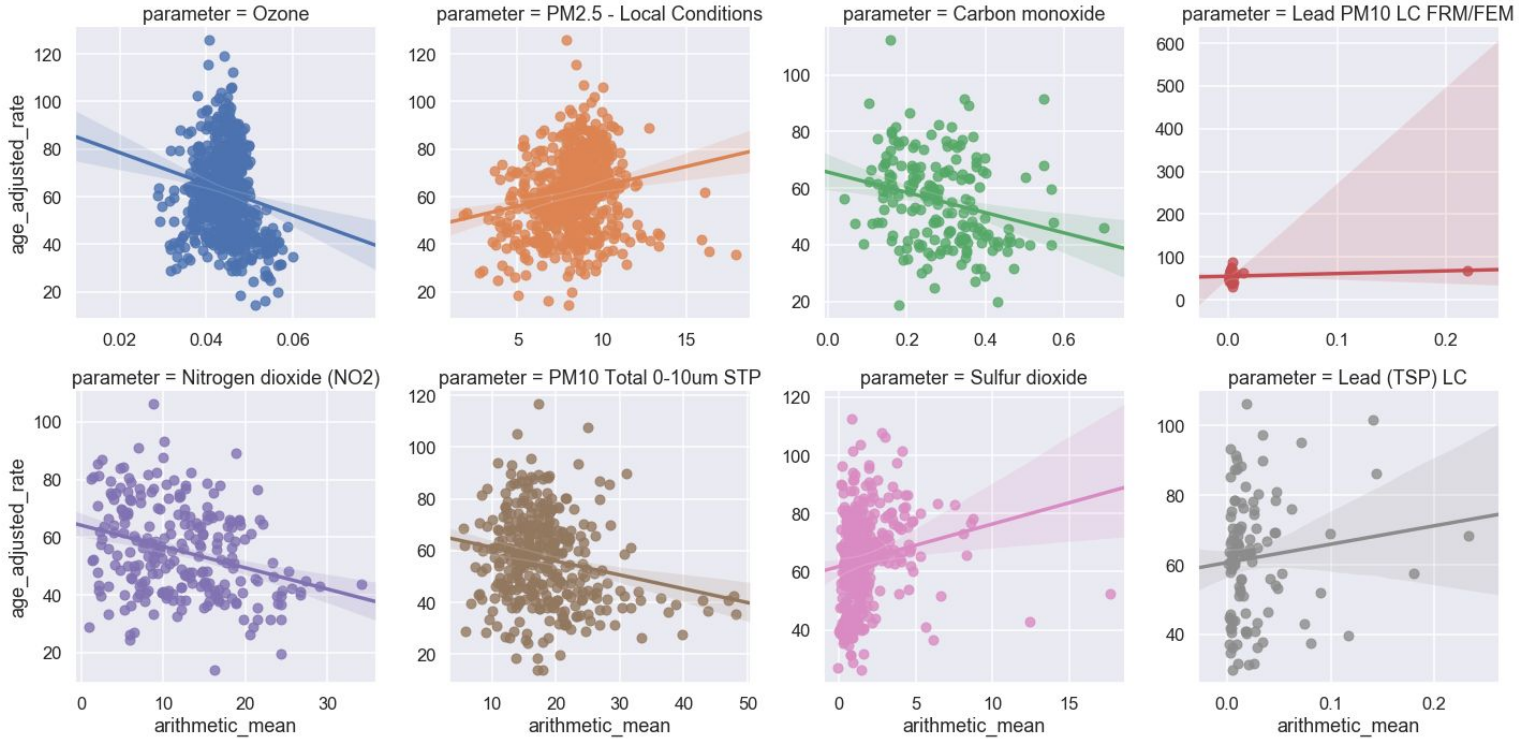
Sources

- EPA Air Quality System API
  - pyaqs: custom module to wrap API requests to pandas dataframes
- CDC Geographic Information Systems database

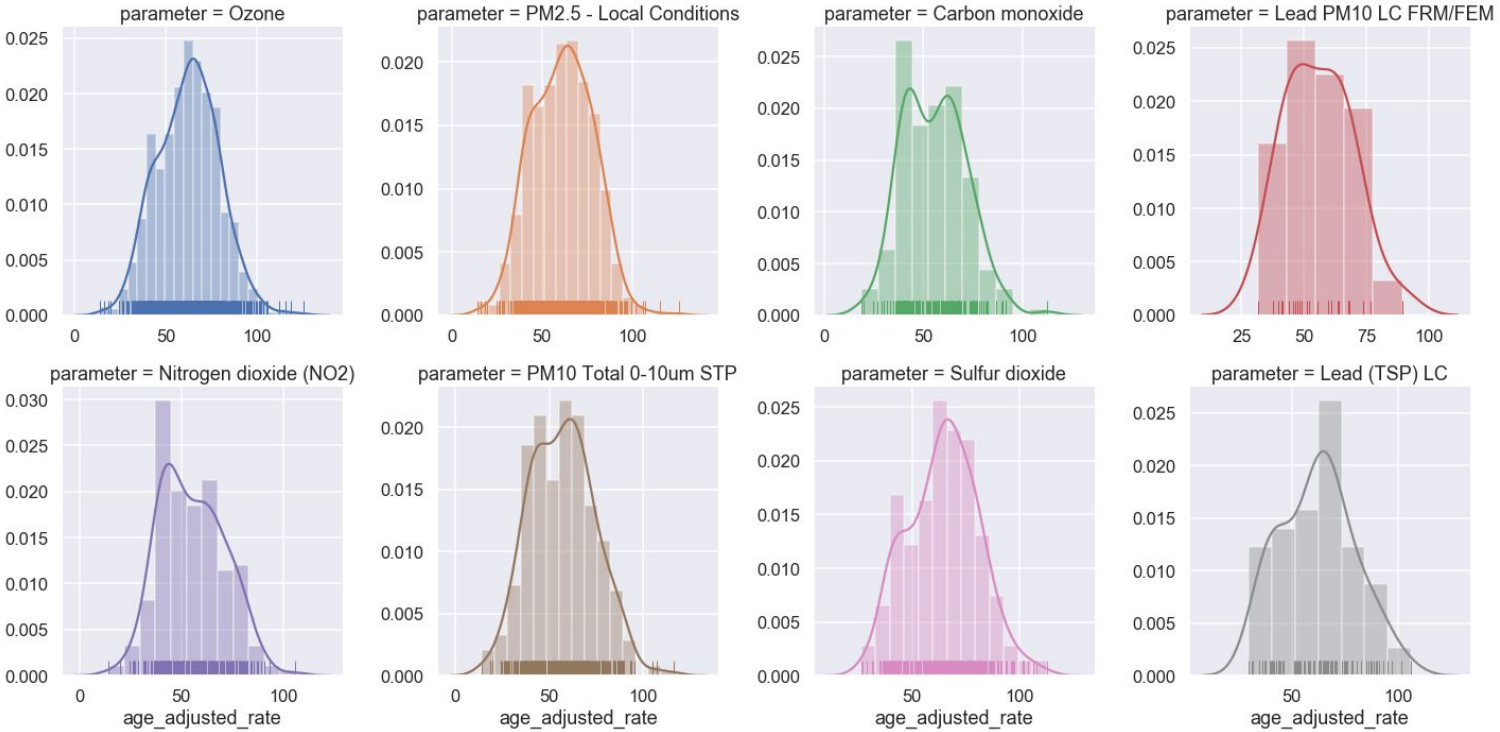
Results - CA


parameter = Carbon monoxide
parameter = Nitrogen dioxide (NO2)
parameter = Ozone
parameter = PM2.5 - Local Conditions
parameter = Sulfur dioxide
parameter = PM10 Total 0-10um STP
parameter = Lead (TSP) LC
parameter = Lead PM10 LC FRM/FEM
AgeAdjustedRate
arithmetic_mean

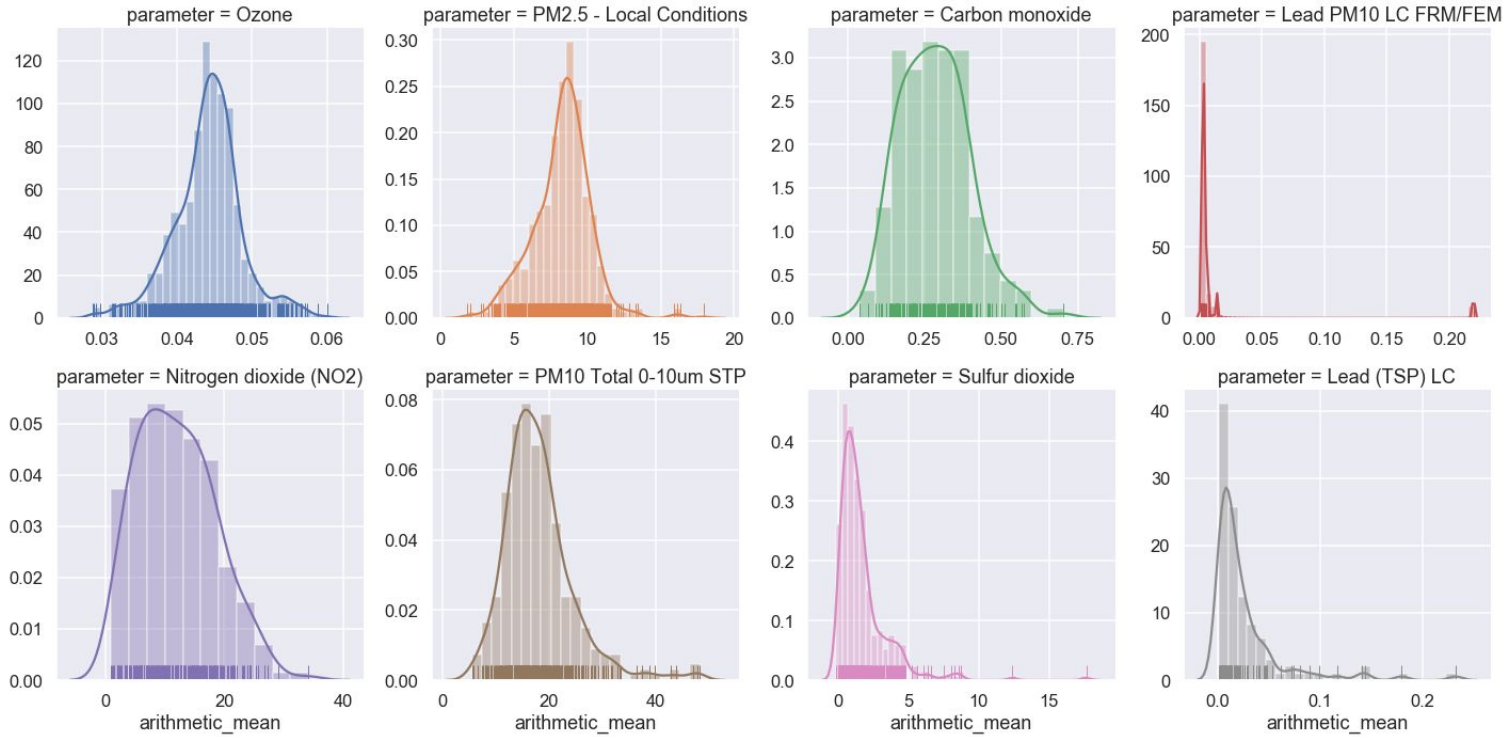
Results - United States



Distributions – Age Adjusted Cancer Rate



Distributions - Average Air Quality



Hypothesis Testing

Alternative Hypothesis: these is a correlation between levels of {parameter} in the air and rates of lung cancer in the United States.

Null Hypothesis: there is no correlation between levels of {parameter} in the air and rates of lung cancer in the United States.

Testing: automated testing for each parameter

- Split into groups based on mean air quality readings (into halves and thirds)
- No conclusive T-test results

Conclusion

- Results: Inconclusive
  - We are unable to make any recommendations due to the lack of data points. Further study is required.
  - In general, cancer is a very complicated disease with many risk factors and triggering events.
  - Moreover, air is very fluid.
- Perhaps looking at instances of chronic lung disease would prove more fruitful.