Data Science Project

I will be studying air quality/pollution and its link to health and wellbeing and the cost of public health.

Long-term exposure to ozone and particle pollution has been found to injure the lungs and has been shown to cause or worsen asthma and other illnesses. I will be using data to study this and other relations of air pollutants to health. For example, the association between increased hospital admissions for asthma and particulate matter.

Additionally, I am interested in looking at other datasets to look for factors that might add context to the study. For example, rat inspection findings, restaurants in the area (type, e.g. fast food), population, socioeconomics, traffic, parks, and anything that might be an indicator or influencer of lifestyle and public health.

Currently: Data collection and exploratory data analysis

1. Basic visualizations in Tableau
2. Python, matplotlib, scipy, pandas, and seaborn

Air Quality

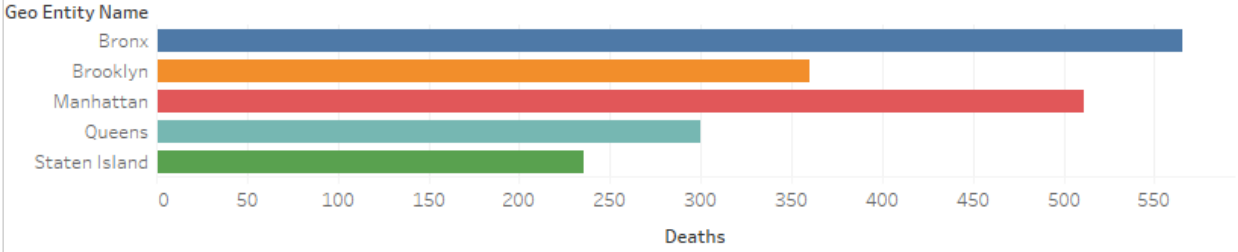
The AQI is an index for reporting daily air quality. It tells you how clean or polluted your air is, and what associated health effects might be a concern for you.
EPA calculates the AQI for five major air pollutants regulated by the Clean Air Act: ground-level ozone, particle pollution (also known as particulate matter), carbon monoxide, sulfur dioxide, and nitrogen dioxide. For each of these pollutants, EPA has established national air quality standards to protect public health .Ground-level ozone and airborne particles are the two pollutants that pose the greatest threat to human health in this country.

Next

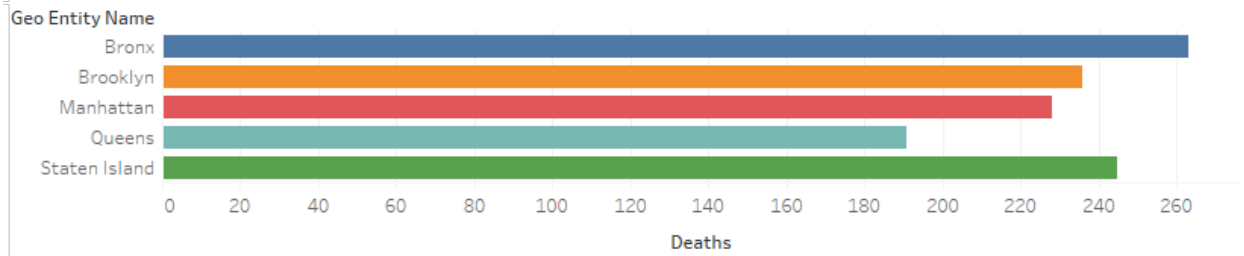
1. Analysis and correlations
2. See where the data exploration takes me, decide on course of action
3. BigQueryHelper
4. Clustering, ML, Inferences
5. Interactive graphs

NYC O3 and PM2.5 Attributed Deaths by Borough

O3 Attributed Deaths by Borough

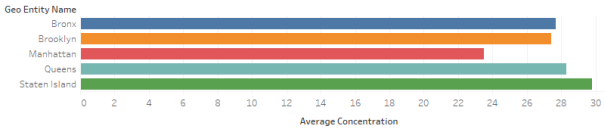


PM2.5 Attributed Deaths by Borough

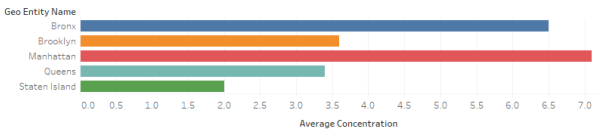


NYC Average Concentration of Air Pollutants (O3, PM2.5, NO, NO2, SO2, and EC) by Borough

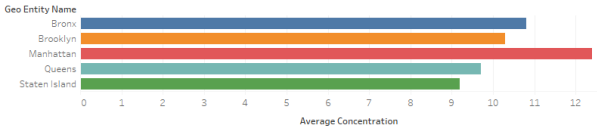
O3 Air Quality by Borough



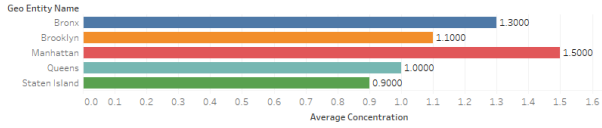
SO2 Air Quality by Borough



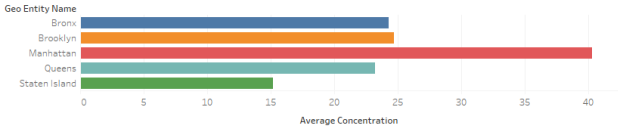
PM2.5 Air Quality by Borough



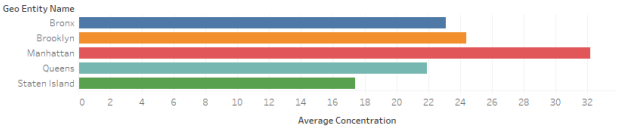
EC Air Quality by Boruough

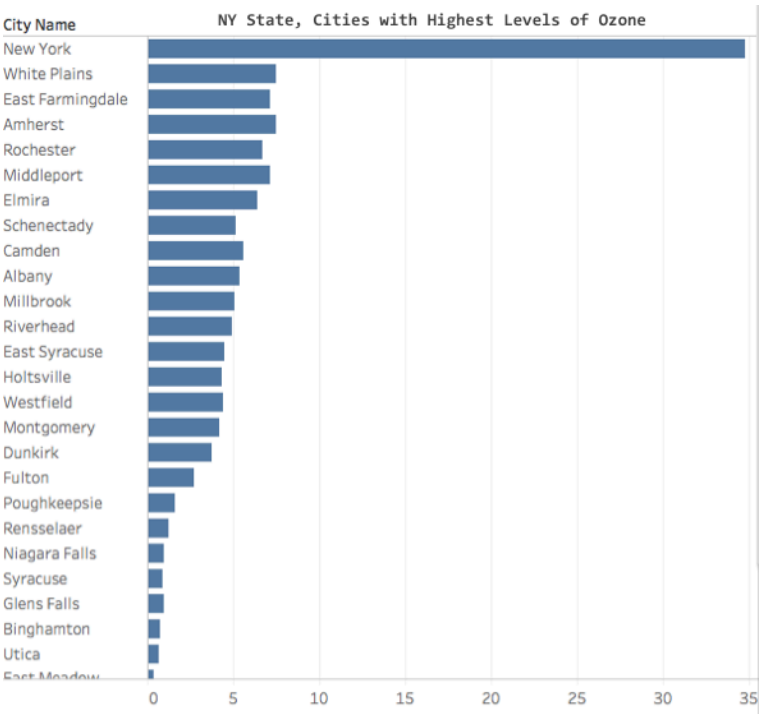


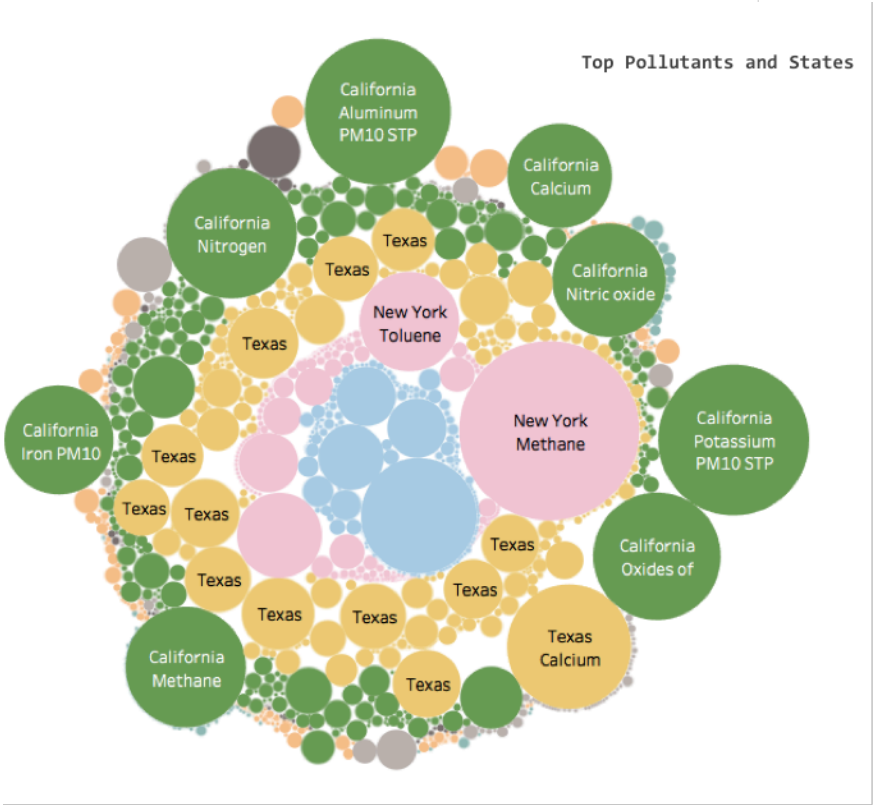
NO Air Quality by Borough

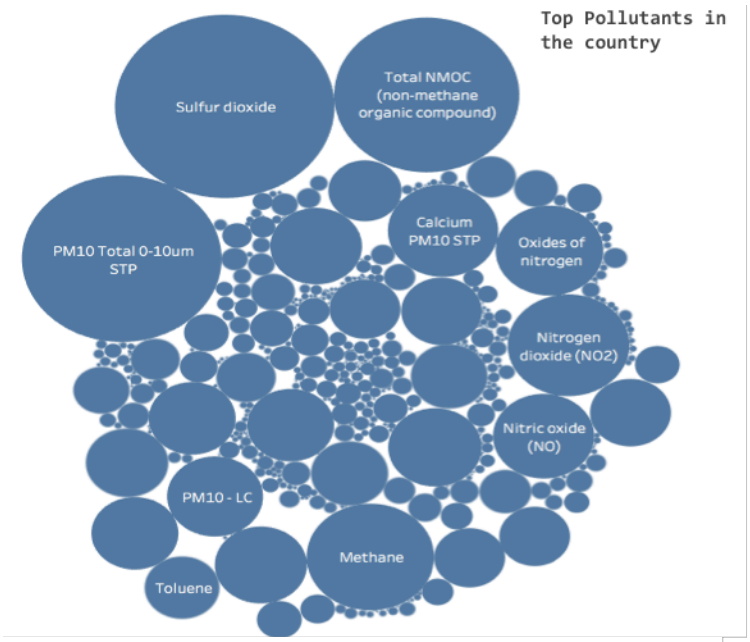


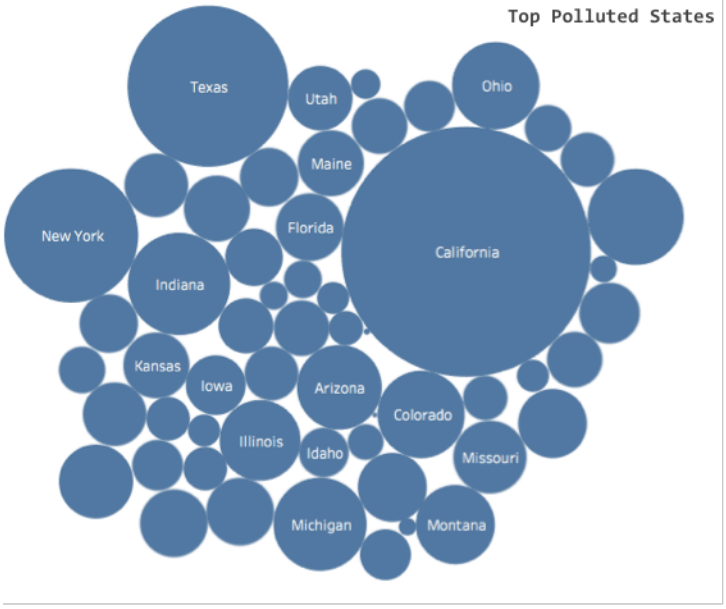
NO2 Air Quality by Borough



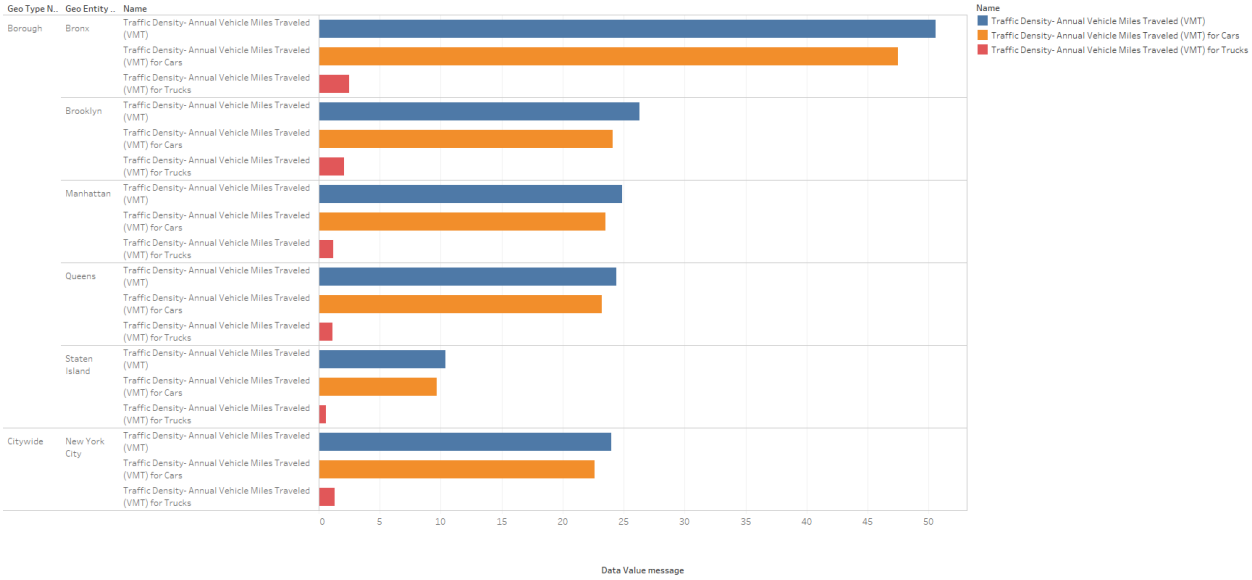

NY State, Cities with Highest Levels of Ozone
City Name
New York
White Plains
East Farmingdale
Amherst
Rochester
Middleport
Elmira
Schenectady
Camden
Albany
Millbrook
Riverhead
East Syracuse
Holtsville
Westfield
Montgomery
Dunkirk
Fulton
Poughkeepsie
Rensselaer
Niagara Falls
Syracuse
Glens Falls
Binghamton
Utica
0
5
10
15
20
25
30
35

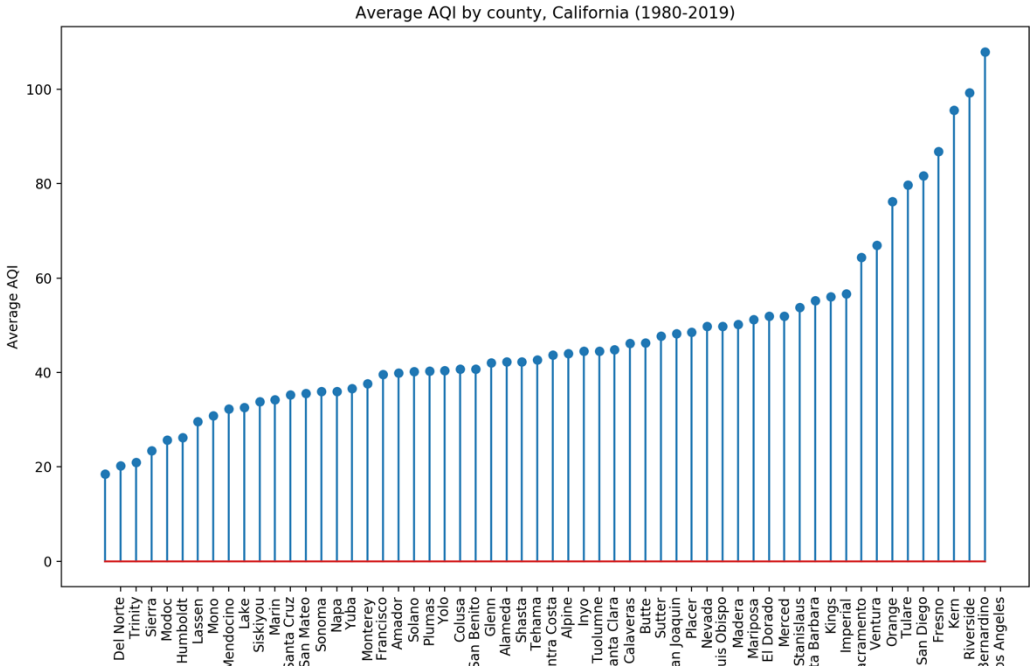

Top Pollutants and States
California Aluminum PM10 STP
California Calcium
California Nitrogen
Texas
Texas
California Nitric oxide
New York Toluene
Texas
New York Methane
California Potassium PM10 STP
California Iron PM10
Texas
Texas
Texas
Texas
California Oxides of
Texas
Texas
Texas
Texas
Texas Calcium
California Methane
Texas

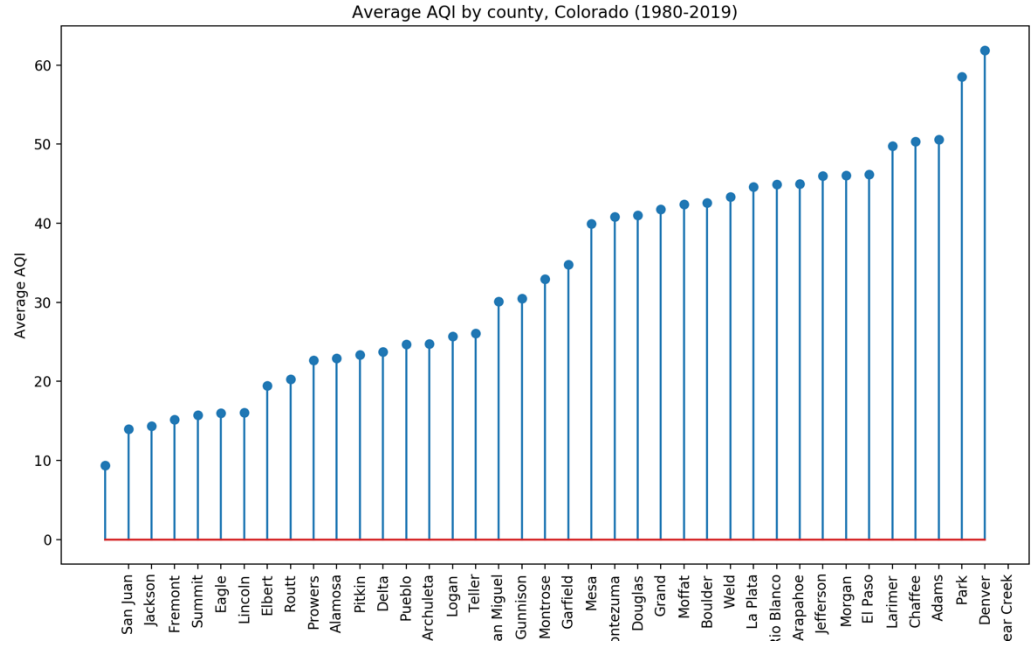

Top Pollutants in the country
Sulfur dioxide
Total NMOC (non-methane organic compound)
PM10 Total 0-10um STP
Calcium PM10 STP
Oxides of nitrogen
Nitrogen dioxide (NO2)
Nitric oxide (NO)
PM10 - LC
Methane
Toluene

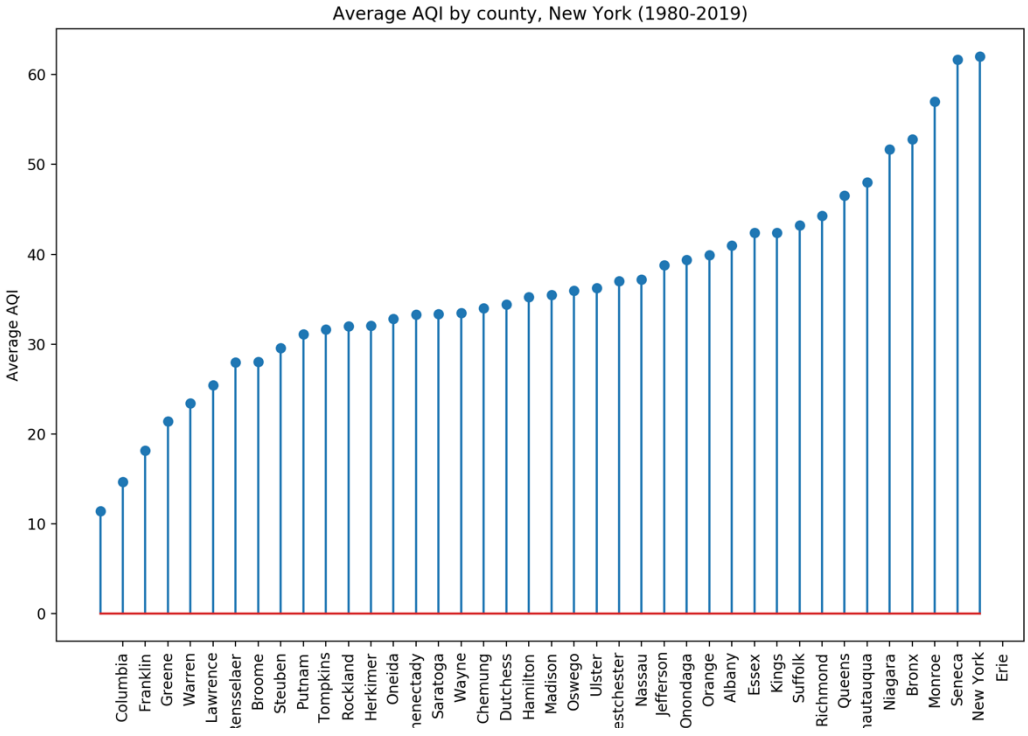

Top Polluted States
Texas
Utah
Ohio
Maine
New York
Florida
California
Indiana
Kansas
Iowa
Arizona
Colorado
Illinois
Idaho
Missouri
Michigan
Montana

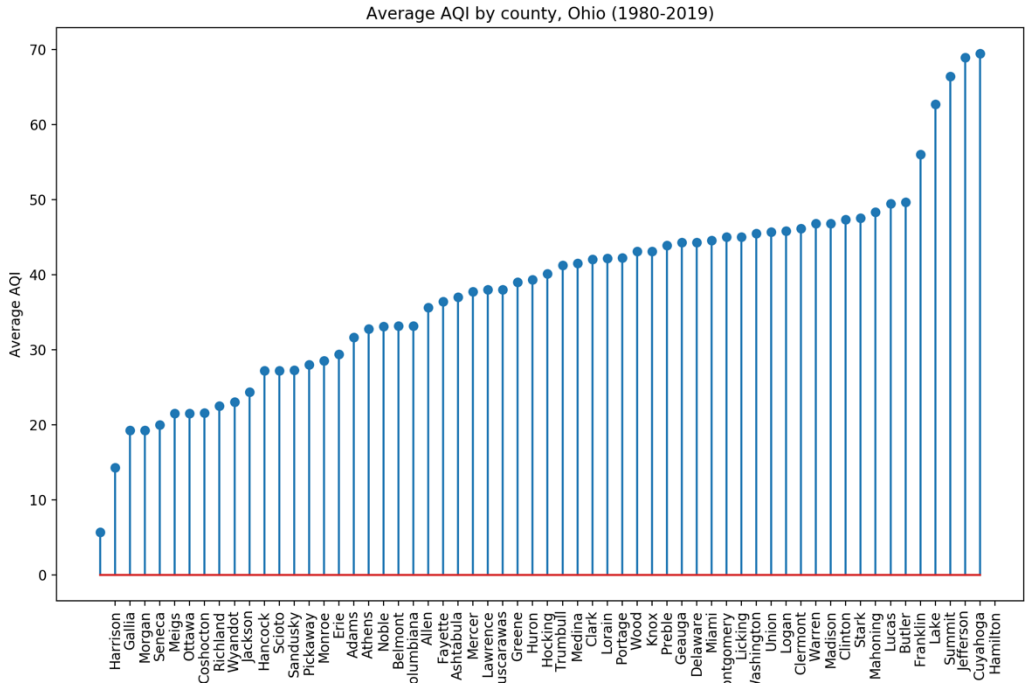
Traffic Density

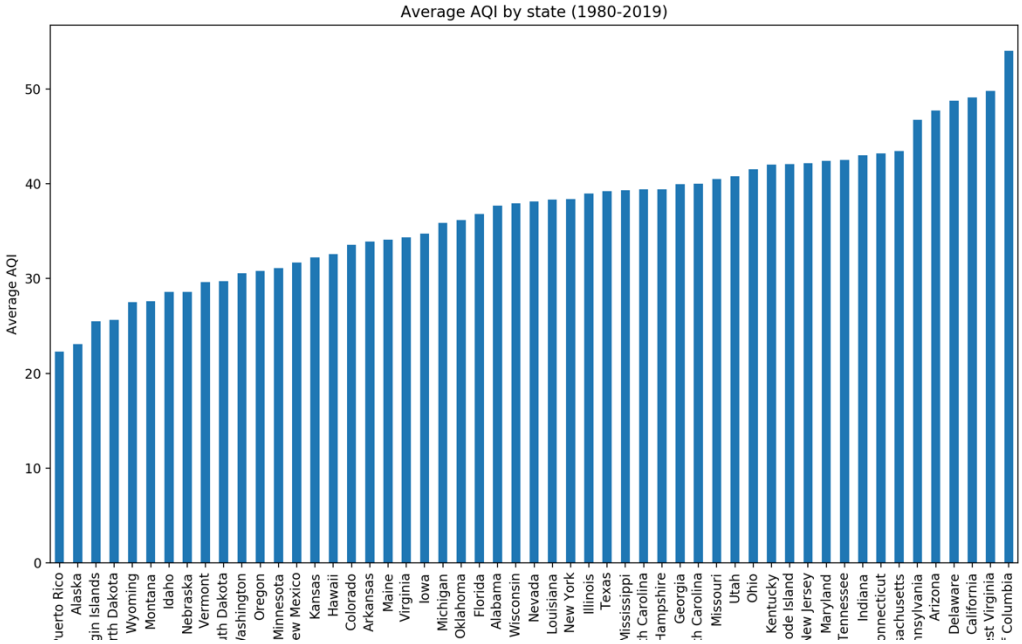


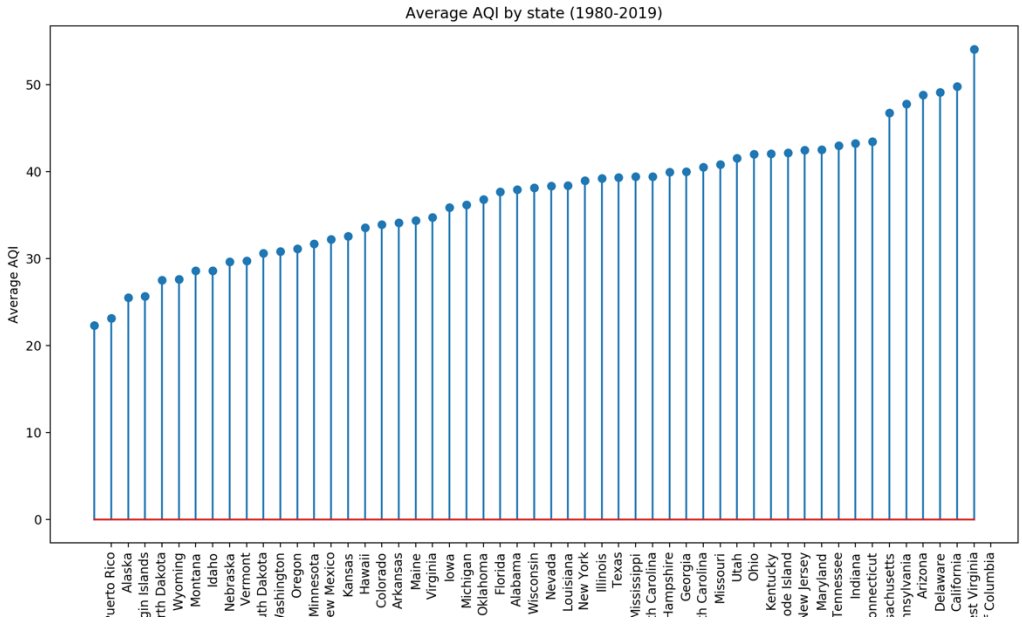

Average AQI by county, California (1980-2019)
Average AQI
0
20
40
60
80
100
Del Norte
Trinity
Sierra
Modoc
Humboldt
Lassen
Mono
Mendocino
Lake
Siskiyou
Marin
Santa Cruz
San Mateo
Sonoma
Napa
Yuba
Monterey
Francisco
Amador
Solano
Plumas
Yolo
Colusa
San Benito
Glenn
Alameda
Shasta
Tehama
ntra Costa
Alpine
Inyo
Tuolumne
anta Clara
Calaveras
Butte
Sutter
an Joaquin
Placer
Nevada
uis Obispo
Madera
Mariposa
El Dorado
Merced
Stanislaus
ta Barbara
Kings
Imperial
cramento
Ventura
Orange
Tulare
San Diego
Fresno
Kern
Riverside
ernardino
s Angeles

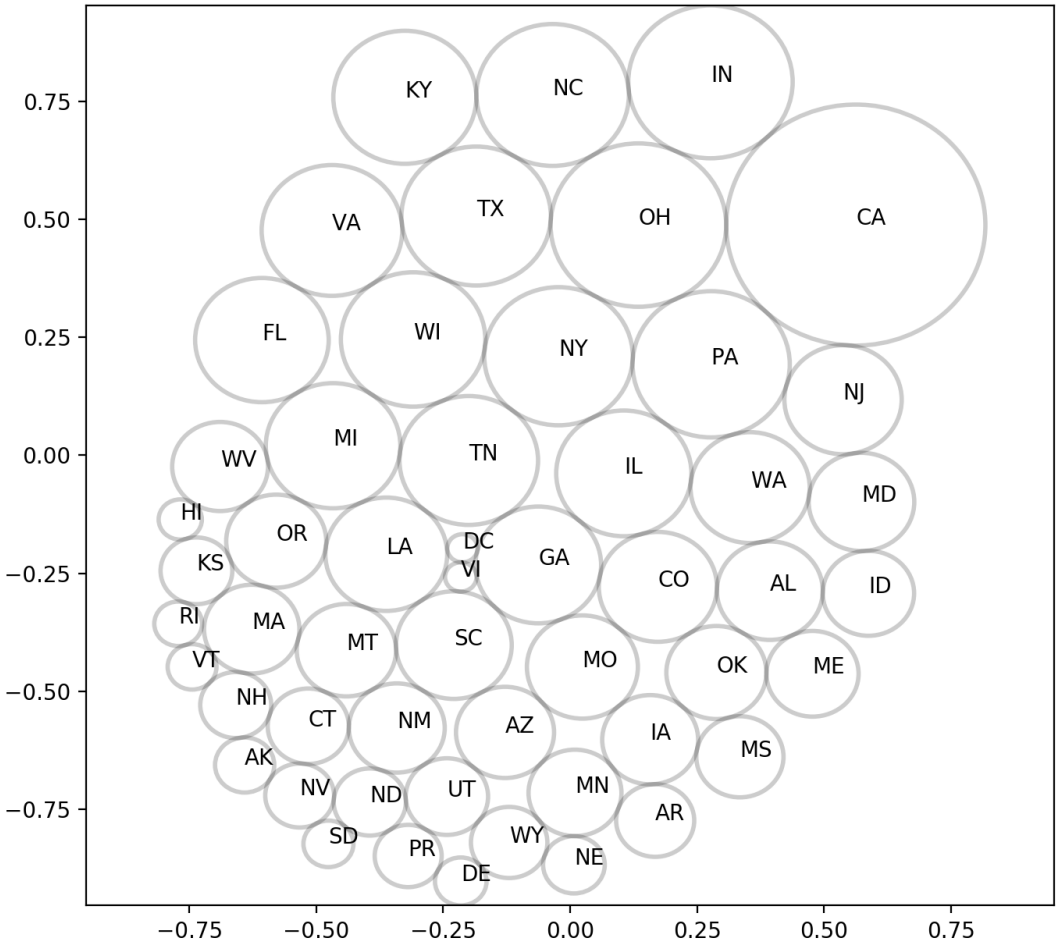

Average AQI by county, Colorado (1980-2019)
Average AQI
0
10
20
30
40
50
60
San Juan
Jackson
Fremont
Summit
Eagle
Lincoln
Elbert
Routt
Prowers
Alamosa
Pitkin
Delta
Pueblo
Archuleta
Logan
Teller
an Miguel
Gunnison
Montrose
Garfield
Mesa
ontezuma
Douglas
Grand
Moffat
Boulder
Weld
La Plata
io Blanco
Arapahoe
Jefferson
Morgan
El Paso
Larimer
Chaffee
Adams
Park
Denver
ear Creek

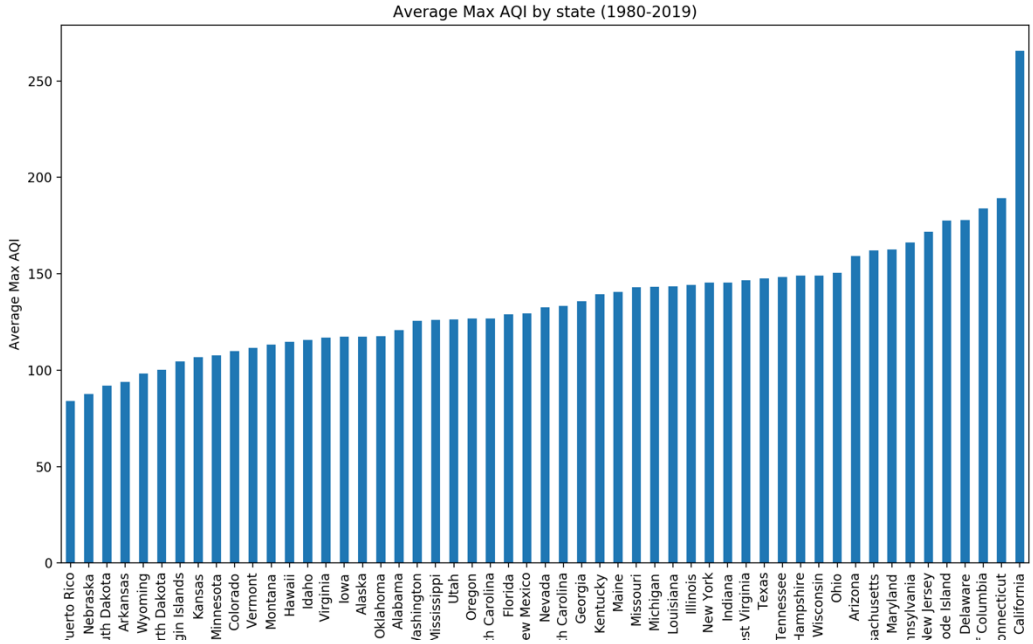

Average AQI by county, New York (1980-2019)
Average AQI
0
10
20
30
40
50
60
Columbia
Franklin
Greene
Warren
Lawrence
Broome
Steuben
Putnam
Tompkins
Rockland
Herkimer
Oneida
Saratoga
Wayne
Chemung
Dutchess
Hamilton
Madison
Oswego
Ulster
Nassau
Jefferson
Onondaga
Orange
Albany
Essex
Kings
Suffolk
Richmond
Queens
Niagara
Bronx
Monroe
Seneca
New York
Erie

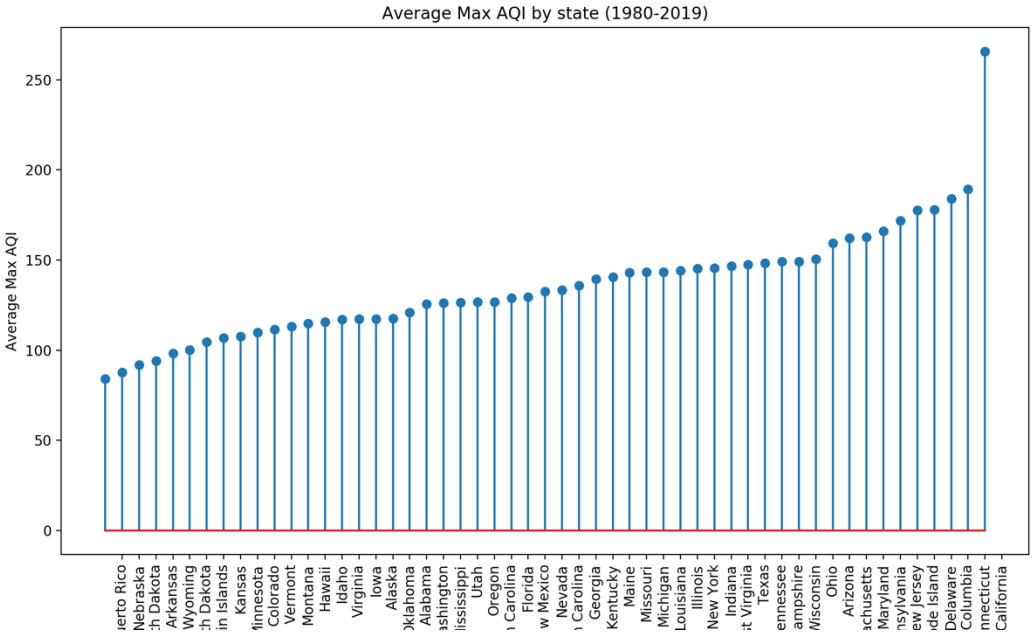

Average AQI by county, Ohio (1980-2019)
Average AQI
0
10
20
30
40
50
60
70
Harrison
Gallia
Morgan
Seneca
Meigs
Ottawa
Coshocton
Richland
Wyandot
Jackson
Hancock
Scioto
Sandusky
Pickaway
Monroe
Erie
Adams
Athens
Noble
Belmont
Allen
Fayette
Ashtabula
Mercer
Lawrence
Greene
Huron
Hocking
Trumbull
Medina
Clark
Lorain
Portage
Wood
Knox
Preble
Geauga
Delaware
Miami
Licking
Union
Logan
Clermont
Warren
Madison
Clinton
Stark
Mahoning
Lucas
Butler
Franklin
Lake
Summit
Jefferson
Cuyahoga
Hamilton

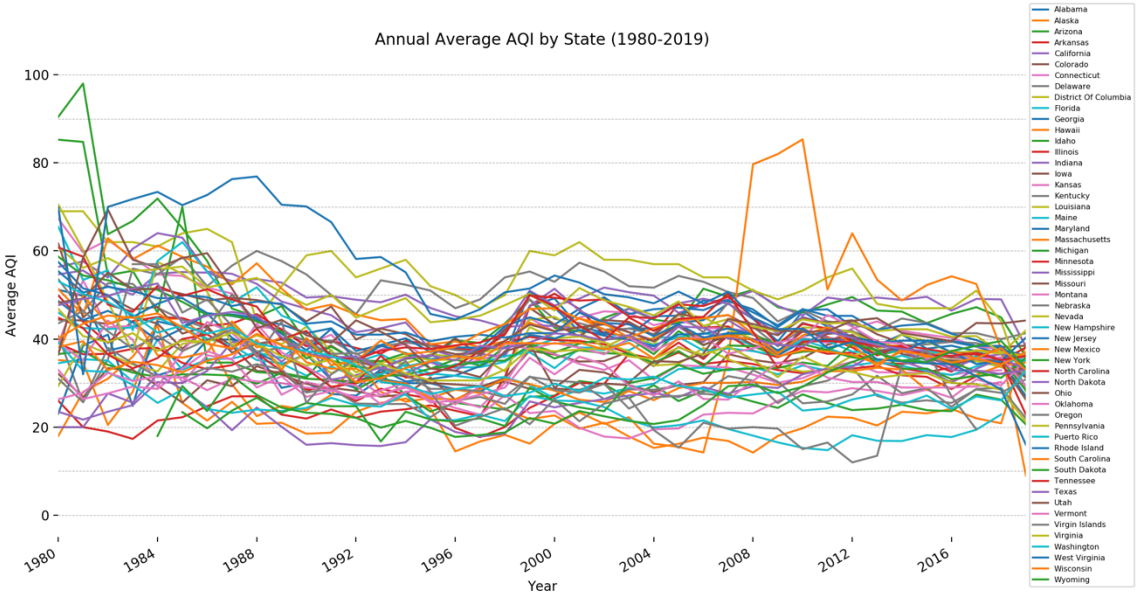

Average AQI by state (1980-2019)
Average AQI
0
10
20
30
40
50
uerto Rico
Alaska
gin Islands
rth Dakota
Wyoming
Montana
Idaho
Nebraska
Vermont
th Dakota
/ashington
Oregon
Minnesota
ew Mexico
Kansas
Hawaii
Colorado
Arkansas
Maine
Virginia
Iowa
Michigan
Oklahoma
Florida
Alabama
Wisconsin
Nevada
Louisiana
New York
Illinois
Texas
Mississippi
h Carolina
Hampshire
Georgia
h Carolina
Missouri
Utah
Ohio
Kentucky
ode Island
lew Jersey
Maryland
Tennessee
Indiana
onnecticut
achusetts
nsylvania
Arizona
Delaware
California
st Virginia
Columbia

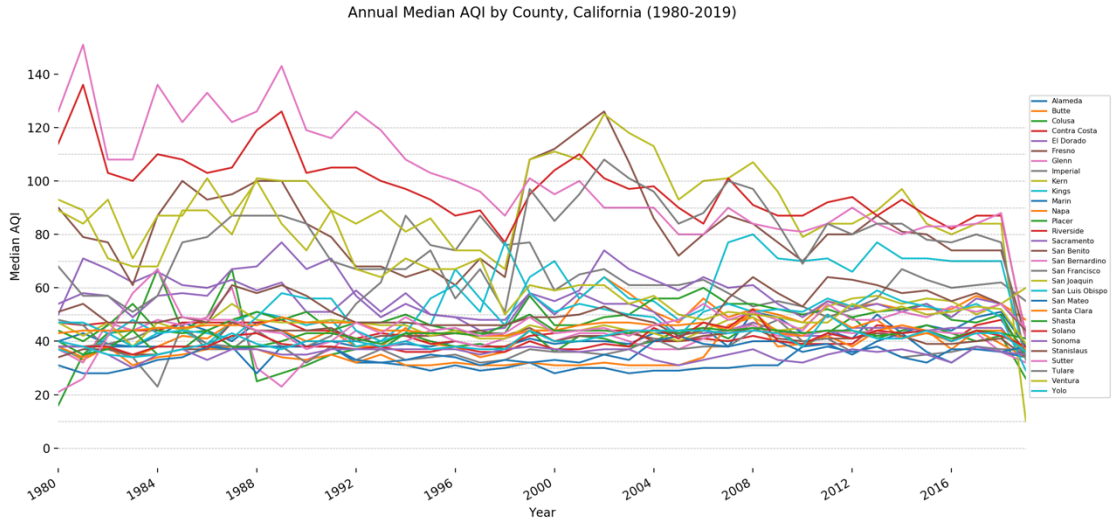

Average AQI by state (1980-2019)
Average AQI
0
10
20
30
40
50
uerto Rico
Alaska
gin Islands
rth Dakota
Wyoming
Montana
Idaho
Nebraska
Vermont
th Dakota
/ashington
Oregon
Minnesota
ew Mexico
Kansas
Hawaii
Colorado
Arkansas
Maine
Virginia
Iowa
Michigan
Oklahoma
Florida
Alabama
Wisconsin
Nevada
Louisiana
New York
Illinois
Texas
Mississippi
h Carolina
Hampshire
Georgia
h Carolina
Missouri
Utah
Ohio
Kentucky
ode Island
lew Jersey
Maryland
Tennessee
Indiana
onnecticut
achusetts
nsylvania
Arizona
Delaware
California
st Virginia
Columbia

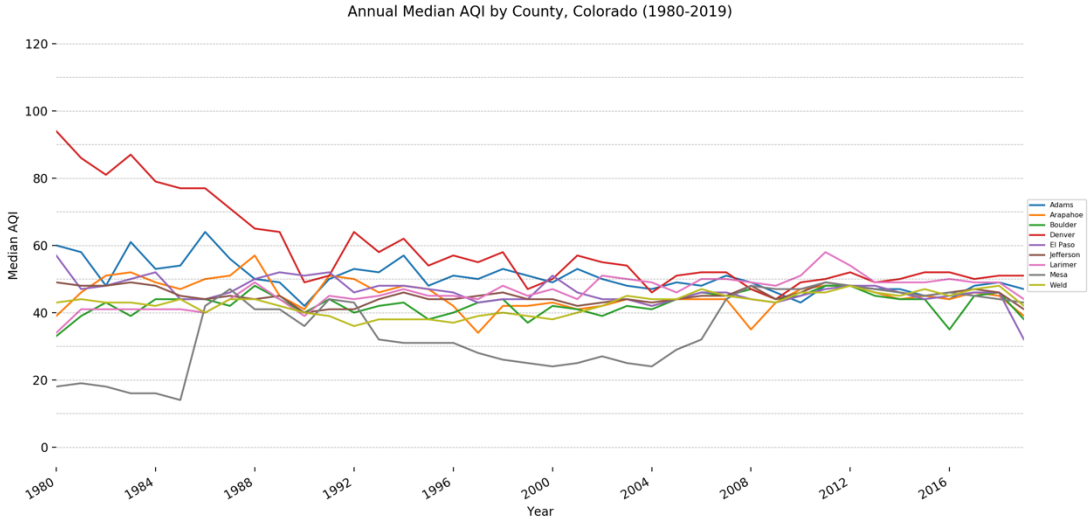

KY
NC
IN
VA
TX
OH
CA
FL
WI
NY
PA
NJ
MI
TN
IL
WA
MD
WV
HI
OR
LA
DC
VI
GA
CO
AL
ID
KS
RI
MA
MT
SC
MO
OK
ME
VT
NH
CT
NM
AZ
IA
MS
AK
NV
ND
UT
MN
AR
SD
PR
WY
NE
DE
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75

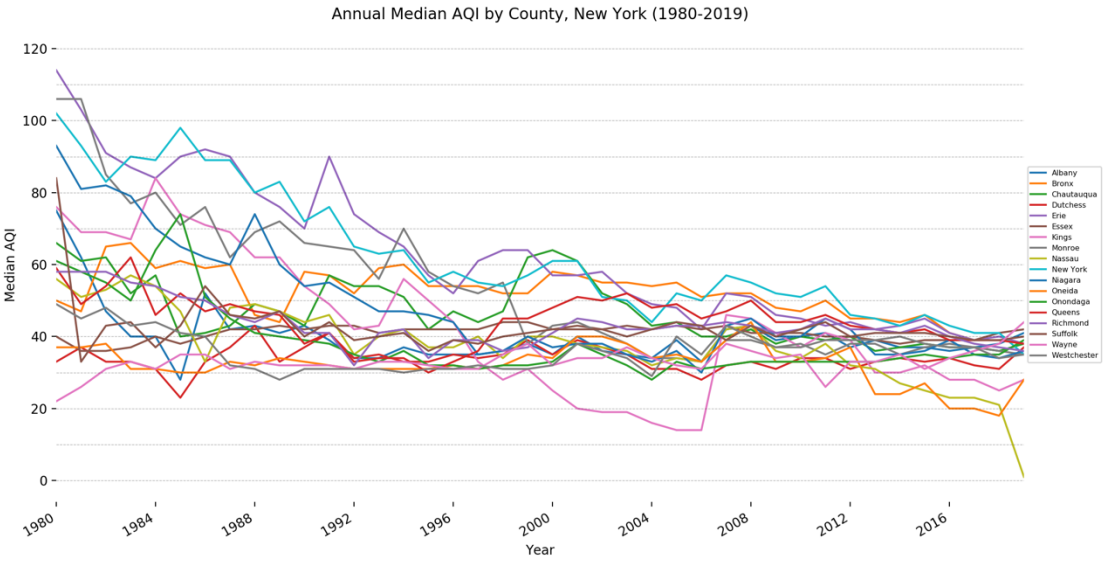

Average Max AQI by state (1980-2019)
Average Max AQI
0
50
100
150
200
250
uerto Rico
Nebraska
th Dakota
Arkansas
Wyoming
rth Dakota
gin Islands
Kansas
Minnesota
Colorado
Vermont
Montana
Hawaii
Idaho
Virginia
Iowa
Alaska
Oklahoma
Alabama
ashington
Mississippi
Utah
Oregon
h Carolina
Florida
ew Mexico
Nevada
h Carolina
Georgia
Kentucky
Maine
Missouri
Michigan
Louisiana
Illinois
New York
Indiana
st Virginia
Texas
Tennessee
Hampshire
Wisconsin
Ohio
Arizona
achusetts
Maryland
nsylvania
New Jersey
ode Island
Delaware
Columbia
onnecticut
California

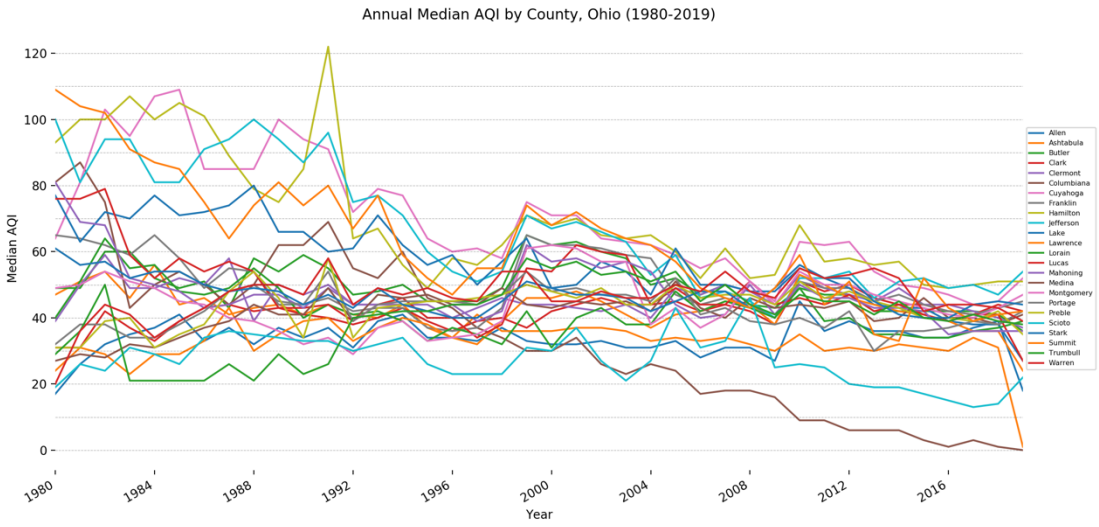

Average Max AQI by state (1980-2019)
Average Max AQI
0
50
100
150
200
250
uerto Rico
Nebraska
th Dakota
Arkansas
Wyoming
th Dakota
n Islands
Kansas
innesota
Colorado
Vermont
Montana
Hawaii
Idaho
Virginia
Iowa
Alaska
klahoma
Alabama
shington
ississippi
Utah
Oregon
Carolina
Florida
w Mexico
Nevada
Carolina
Georgia
Kentucky
Maine
Missouri
Michigan
Louisiana
Illinois
New York
Indiana
t Virginia
Texas
ennessee
ampshire
Visconsin
Ohio
Arizona
achusetts
Maryland
nsylvania
ew Jersey
de Island
Delaware
Columbia
nnecticut
California


Annual Average AQI by State (1980-2019)
Average AQI
Year
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District Of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virgin Islands
Virginia
Washington
West Virginia
Wisconsin
Wyoming


Annual Median AQI by County, California (1980-2019)
Median AQI
Year
Alameda
Butte
Colusa
Contra Costa
El Dorado
Fresno
Glenn
Imperial
Kern
Kings
Marin
Napa
Placer
Riverside
Sacramento
San Benito
San Bernardino
San Francisco
San Joaquin
San Luis Obispo
San Mateo
Santa Clara
Shasta
Solano
Sonoma
Stanislaus
Sutter
Tulare
Ventura
Yolo


Annual Median AQI by County, Colorado (1980-2019)
Median AQI
Year
Adams
Arapahoe
Boulder
Denver
El Paso
Jefferson
Larimer
Mesa
Weld


Annual Median AQI by County, New York (1980-2019)
Median AQI
Year
Albany
Bronx
Chautauqua
Dutchess
Erie
Essex
Kings
Monroe
Nassau
New York
Niagara
Oneida
Onondaga
Queens
Richmond
Suffolk
Wayne
Westchester



Possible inferences/Analysis

- Predict temperature from AQI
- Effect of air pollution on temperature and humidity
- Seasonal Air Quality Trends
- Correlation between air pollution and global warming
- Other factors are possible: rat inspection findings, restaurants in the area (type, e.g. fast food), population, socioeconomics, traffic, parks, and anything that might be an indicator or influencer of lifestyle and public health.
- Cluster states/cities/counties/geoloc according to air quality. Are there common features among them? (socioeconomics, traffic, parks)

Datasets
Air Quality:
https://www.epa.gov/outdoor-air-quality-data
https://data.Cityofnewyork.Us/environment/air-quality/c3uy-2p5r
https://www.kaggle.com/epa/air-quality
Asthma hospitalizations:
http://a816-dohbesp.nyc.gov/IndicatorPublic/PublicTracking.aspx?theme_code=2,3&subtopic_id=11
https://data.cccnewyork.org/data/download