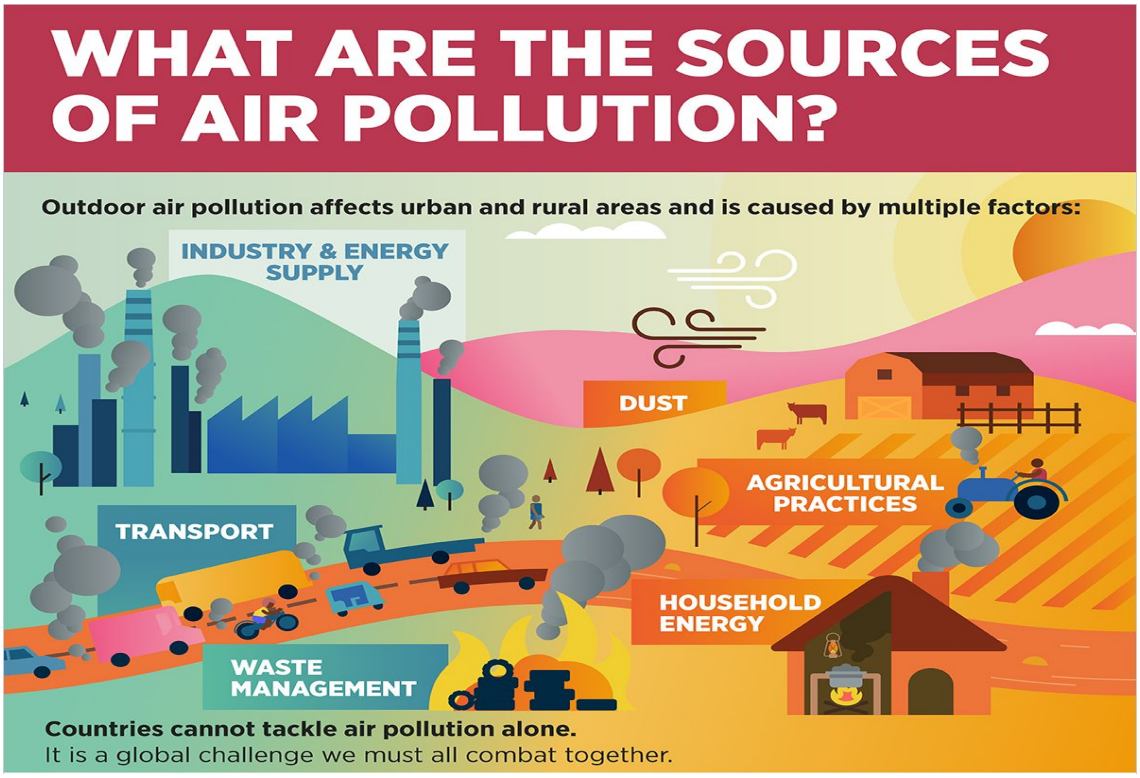
Air Pollution Analytics



Source : World Health Organization

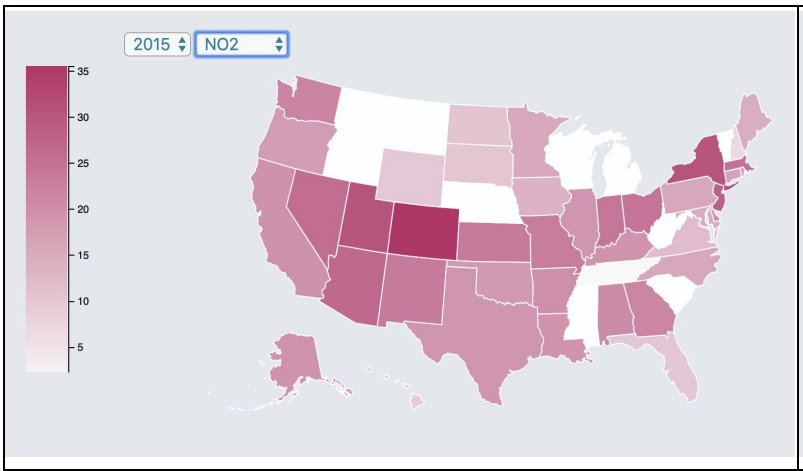
Air Quality Index Levels of Health Concern	Numerical Value	Meaning
Good	0-50	Air quality is considered satisfactory, and air pollution poses little or no risk.
Moderate	51-100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101-150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151 to 200	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health warnings of emergency conditions. The entire population is more likely to be affected.
Hazardous	301 to 500	Health alert: everyone may experience more serious health effects.

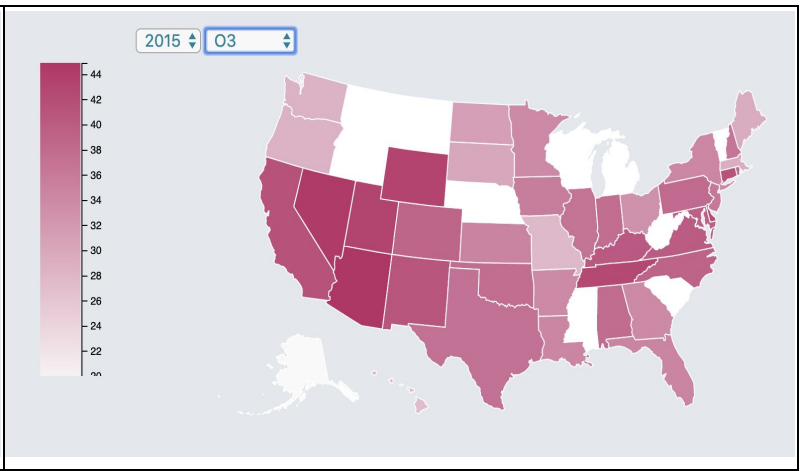
Source: airnow.gov

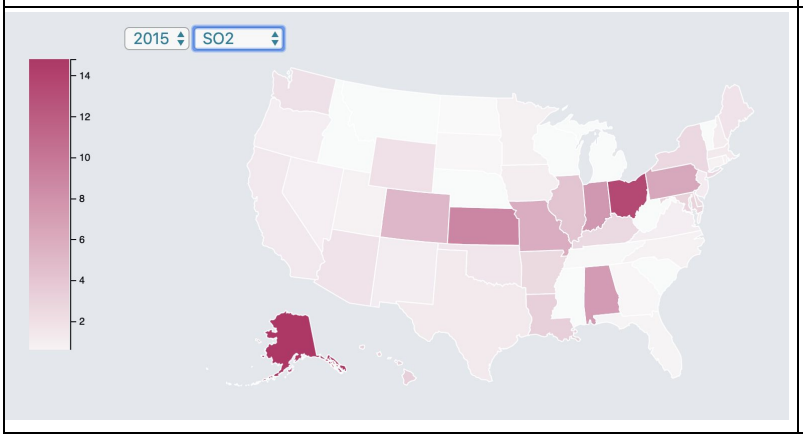
Dataset

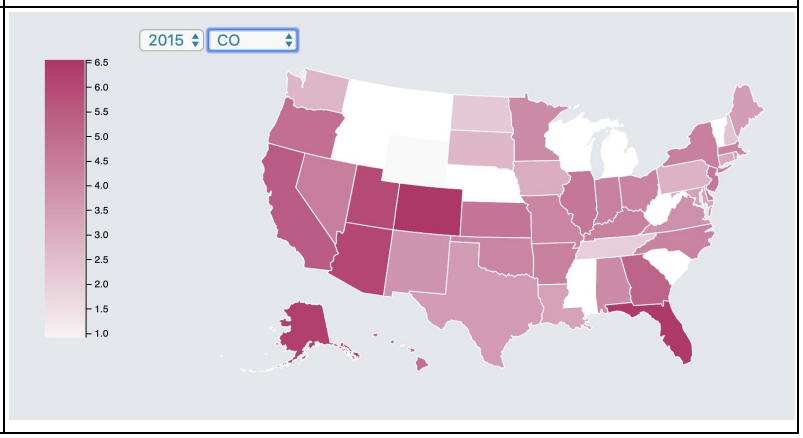
- Source:Environmental Protection Agency(EPA)
- Concentrations of pollutants in states and counties from the year 2000-2016
- EPA has established standards for six common air pollutants, which are referred to as “criteria” pollutants.
  - Carbon monoxide (CO)
  - Lead (Pb)
  - Nitrogen dioxide (NO2)
  - Ozone (O3)
  - Particulate matter (PM), and
  - Sulfur dioxide (SO2)
- Our dataset contains data about four of the above six pollutants: NO2, O3, SO2, and CO. Quantities for each pollutant: Mean, AQI, 1st Max Value and 1st Max Hour.

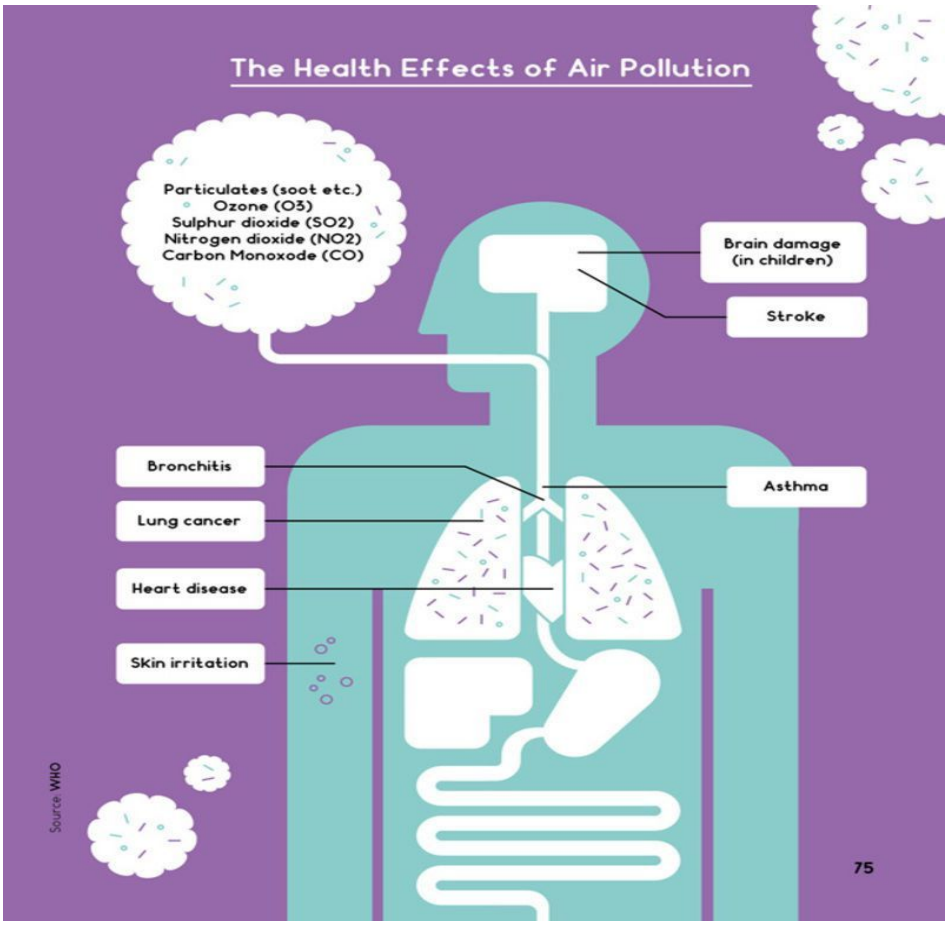
Distribution of Pollutants in terms of AQI levels



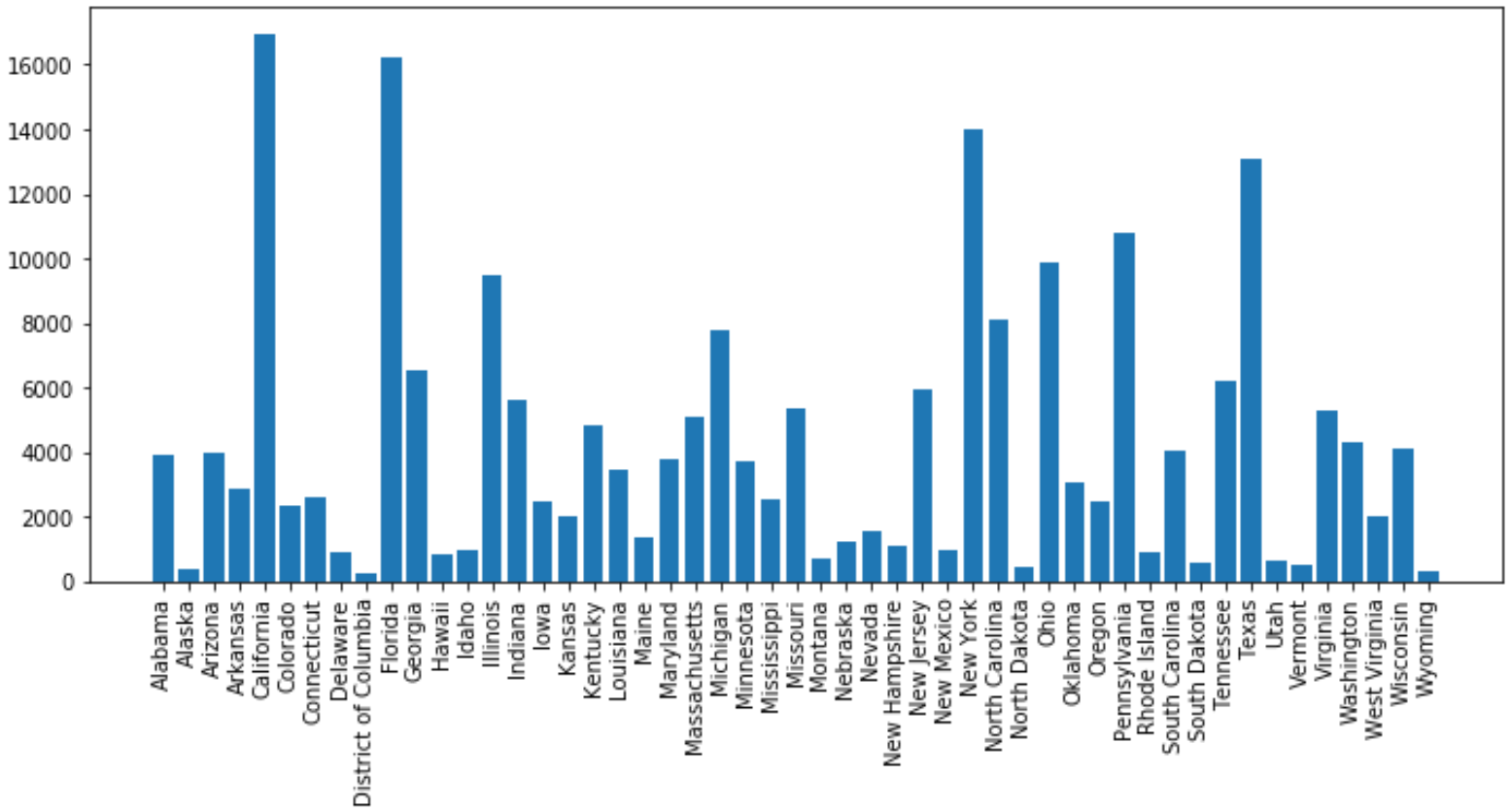




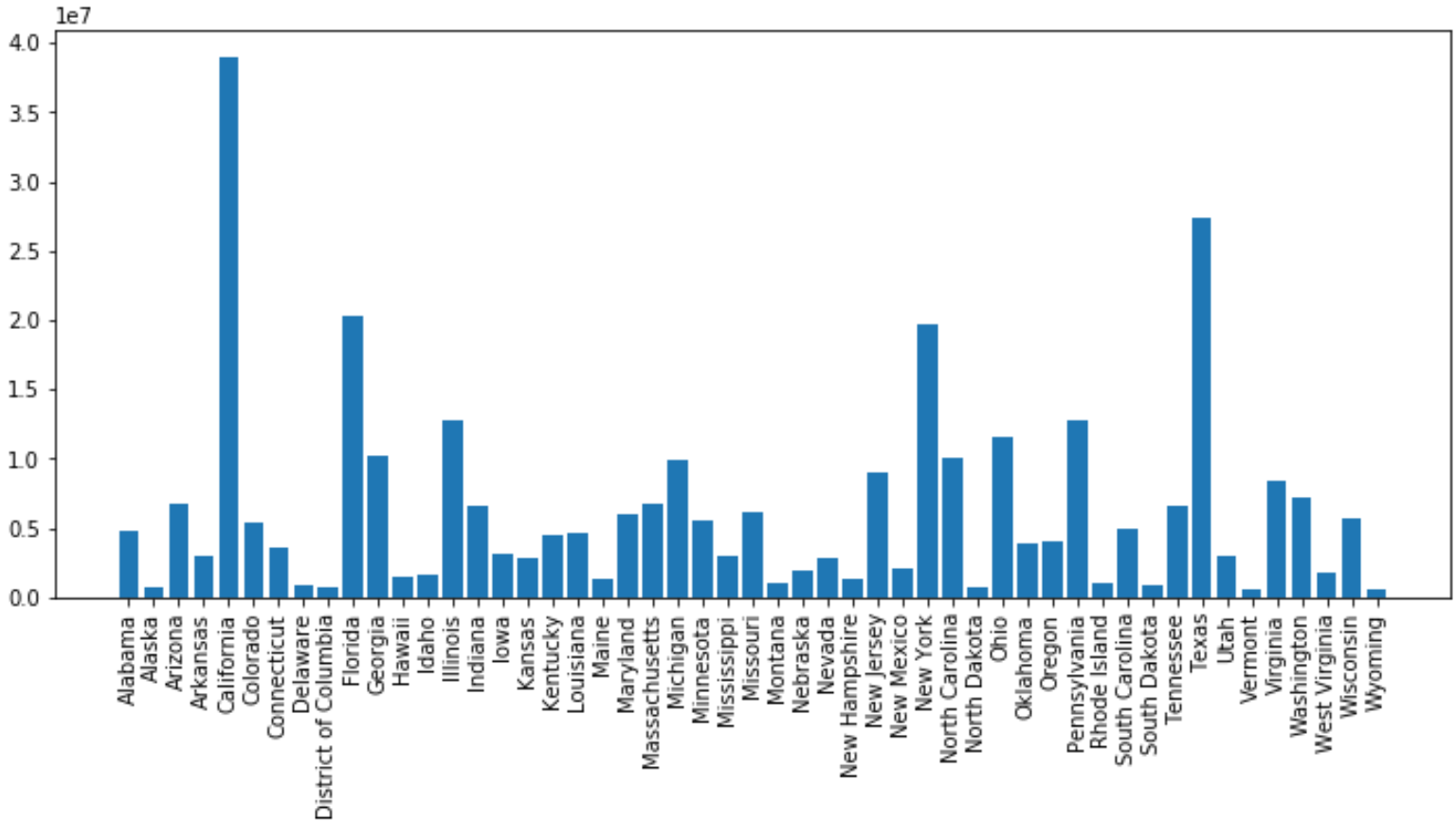




Source: World Health Organization



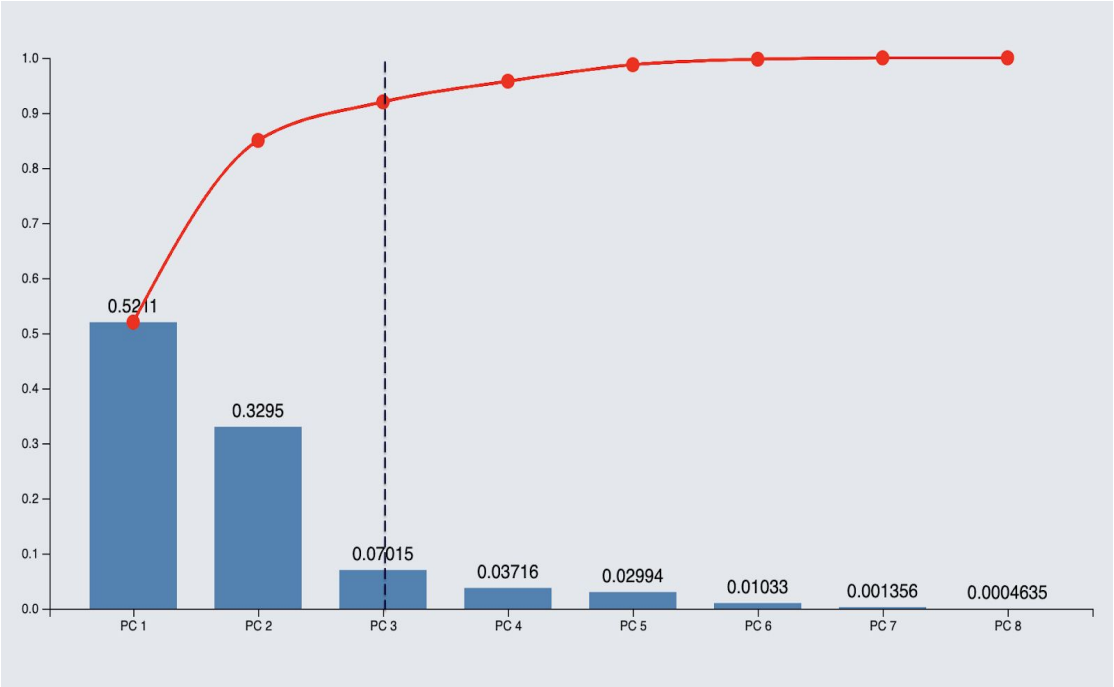
The death count from lung and bronchus cancer in 2015



Population in 2015

Principal Component Analysis

- We found out the intrinsic dimensionality of the data using PCA.
- We passed the Air Quality Indices and mean values (normalized according to units) of the four pollutants to PCA. The results are as follows:

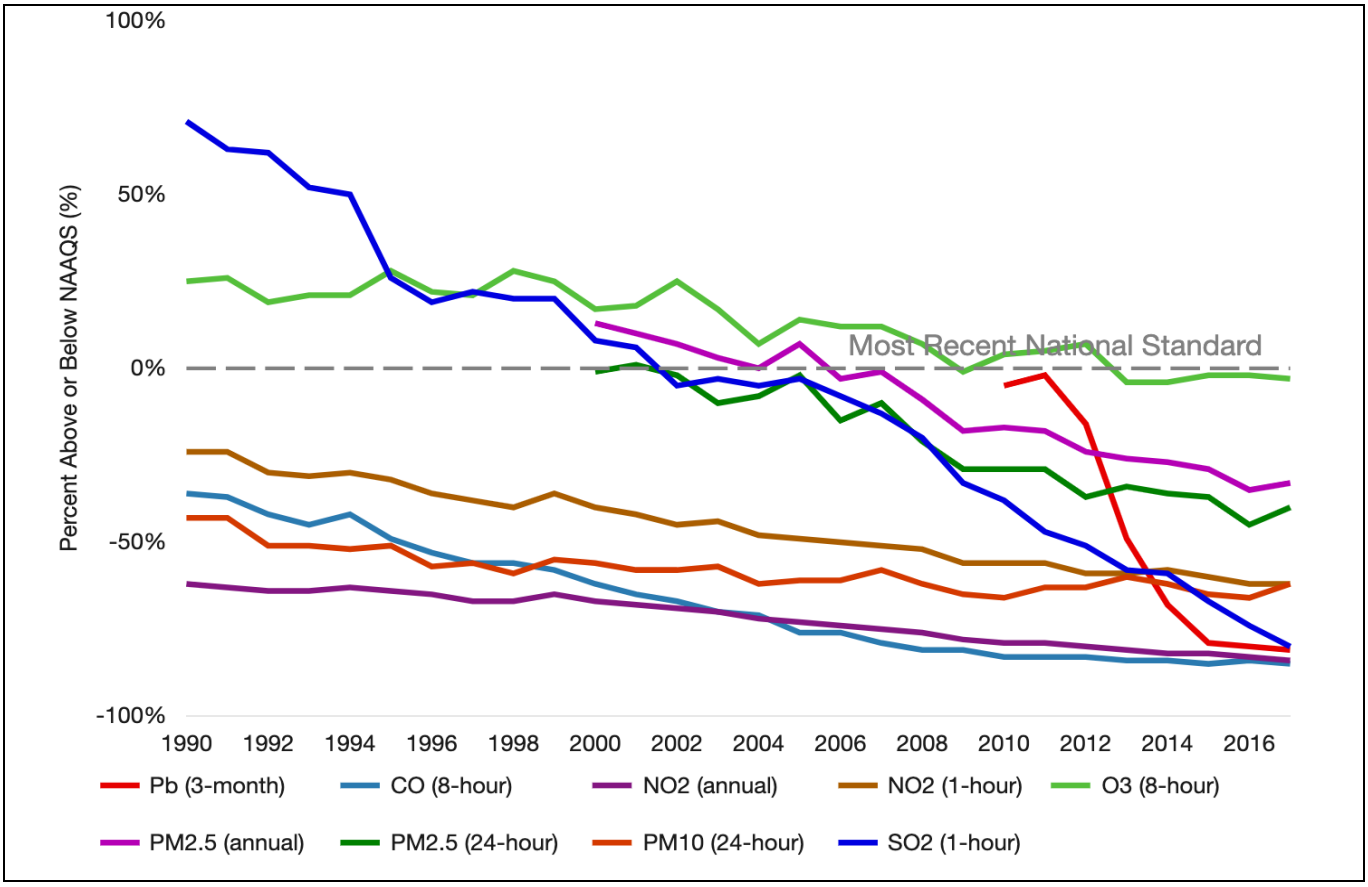


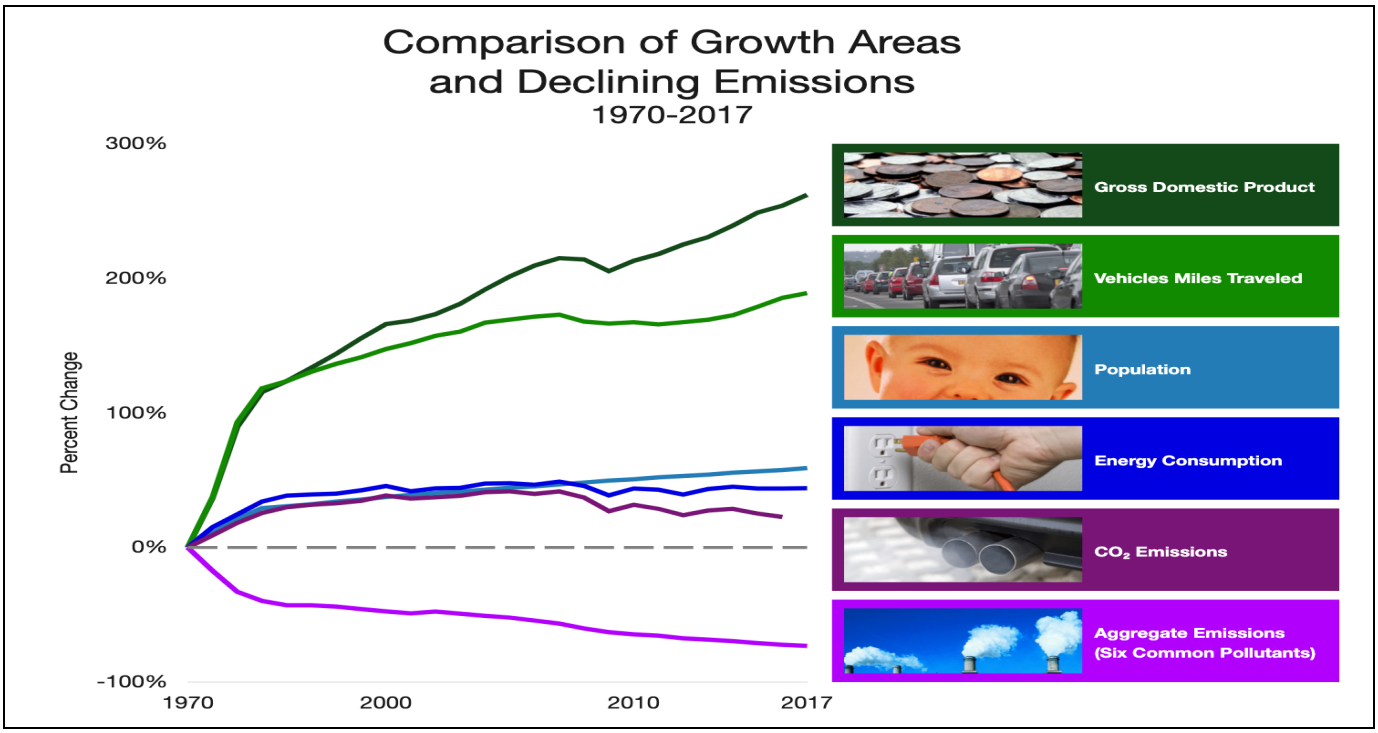
- More than 90% of the variance is explained by the first three components.

	Attribute names	Squared PCA loadings
1	NO2 AQI	0.828124
2	O3 Mean	0.79649
3	O3 AQI	0.631591
0	NO2 Mean	0.451281
6	CO Mean	0.17798
5	SO2 AQI	0.16062
7	CO AQI	0.138663
4	SO2 Mean	0.02343

Squared loadings

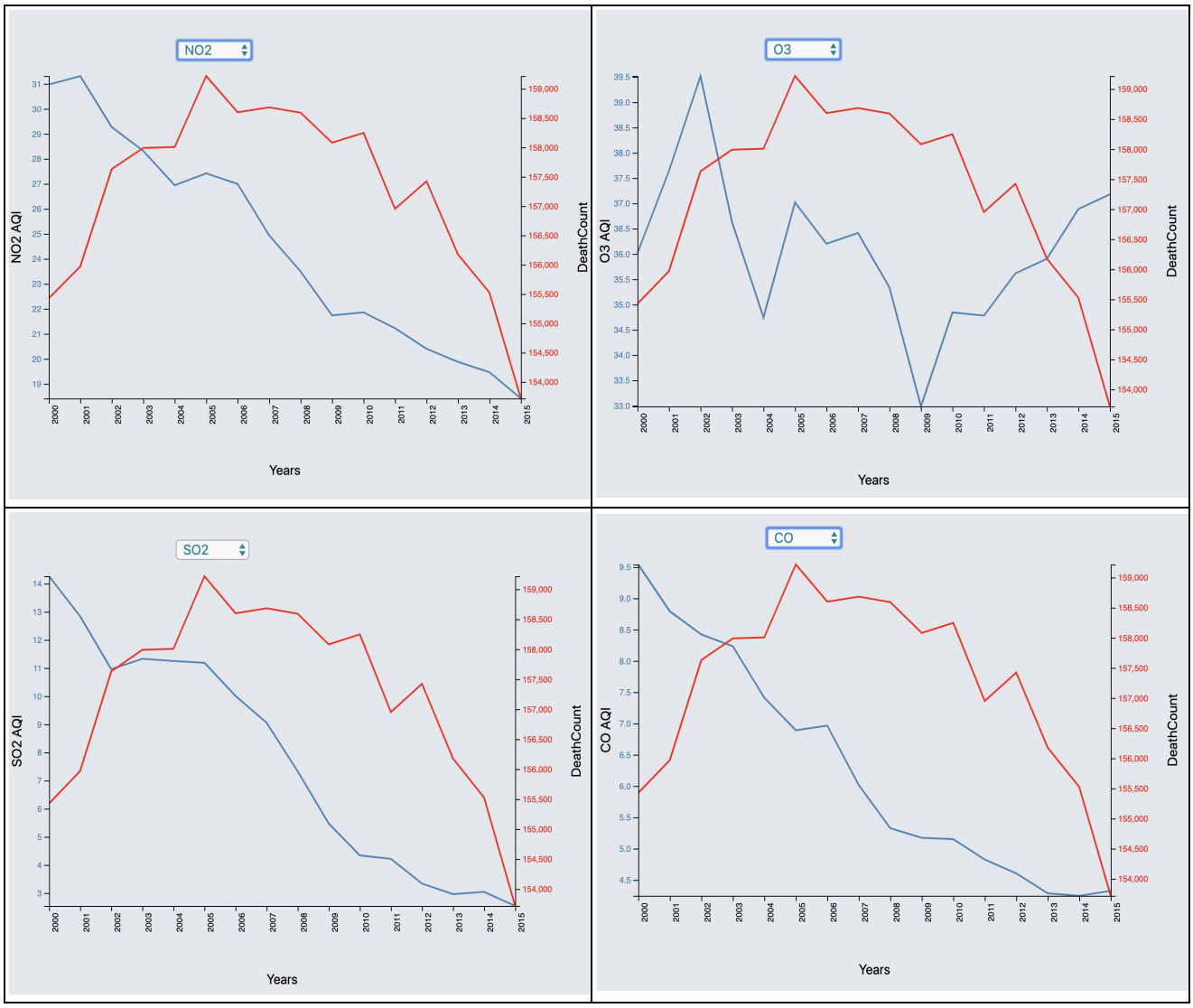
- Calculated the sum of squares of factor loadings on the first three principal components for each of the eight attributes.
- The table shows the importance of each attribute
- We see that NO2 and O3 have the highest PCA loadings. Thus, we come to the conclusion that NO2 and O3 are the most important pollutants in this dataset.





- For more than forty years, the Clean Air Act has been a key part of cutting pollution as the U.S. economy has grown.
- EPA works with state, local and tribal governments to reduce emissions of the 187 hazardous air pollutants.

Pollutants AQI and Death Count for the entire US from 2000 - 2015



- Today Americans breathe cleaner air and face lower risks of adverse health effects
- The U.S. leads the world in having clean air and a strong economy due to implementation of the Clean Air Act
- The counts of deaths due to respiratory diseases has significantly plummeted over the years.