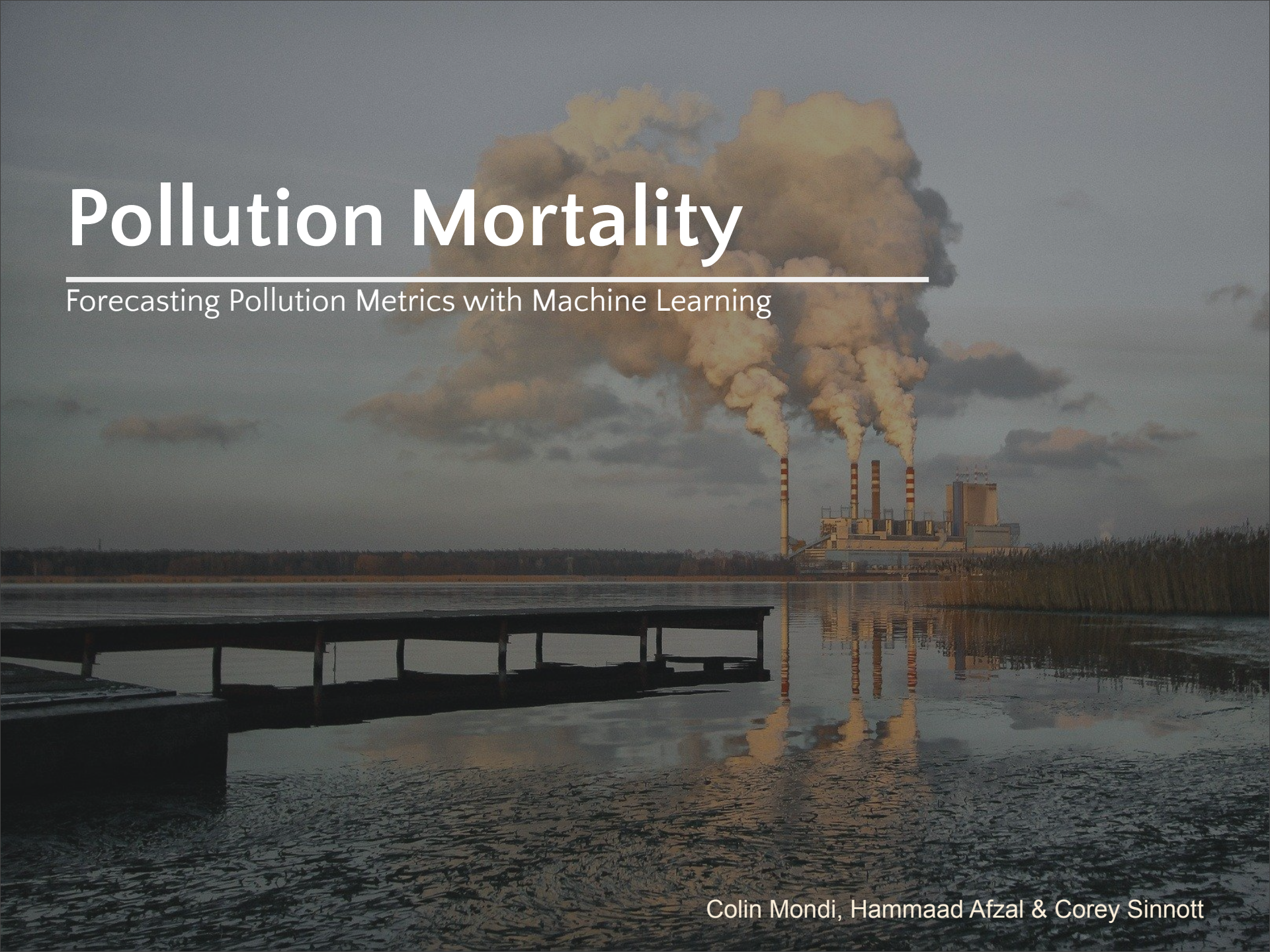


Pollution Mortality

Forecasting Pollution Metrics with Machine Learning

Colin Mondi, Hammaad Afzal & Corey Sinnott

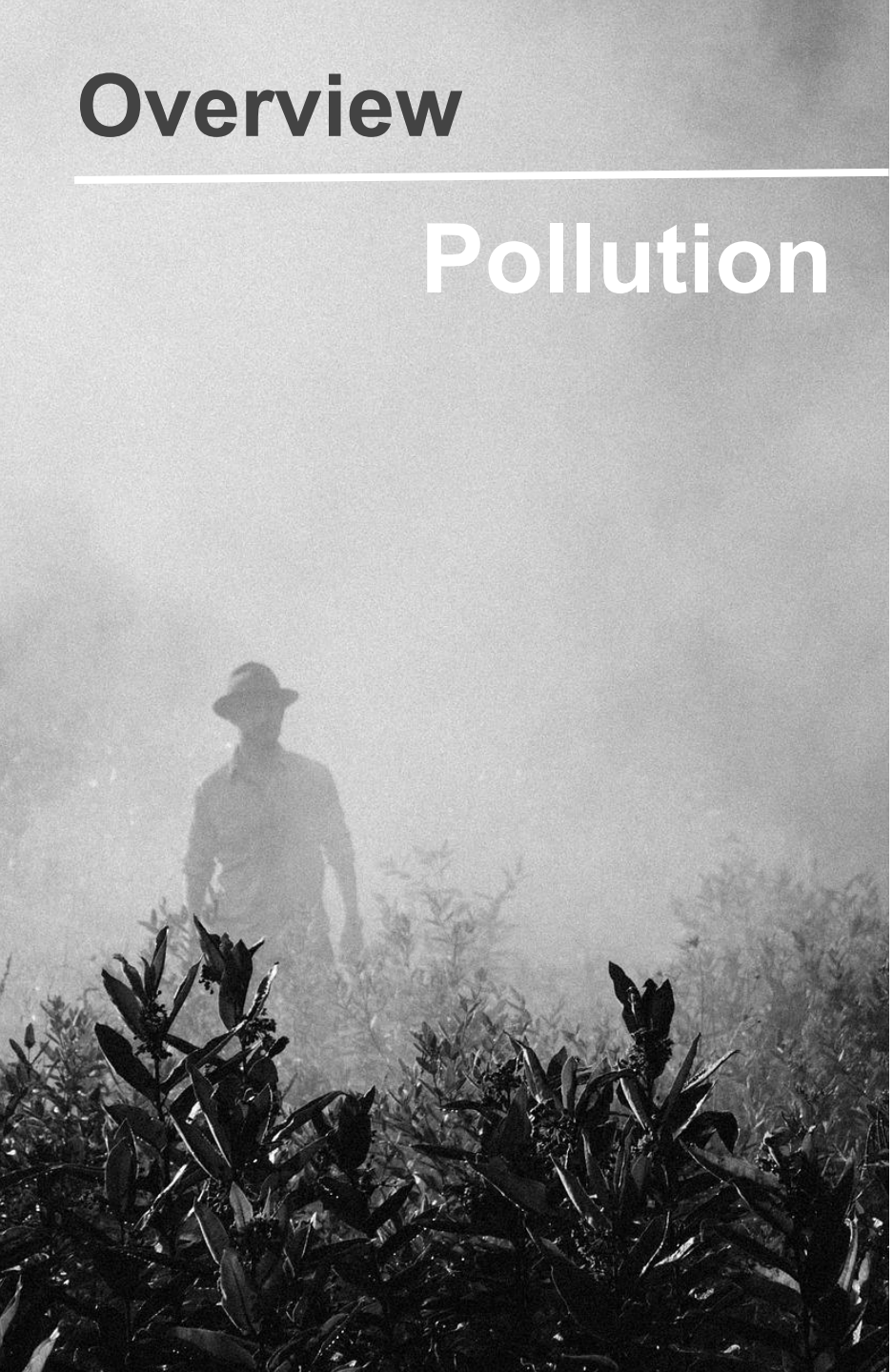
TOC

Problem Statement

Can we predict pollution mortality?

Overview

Pollution



- **Ambient /Outdoor Pollution**
  - Industrial processes
  - Motor vehicles
  - Wildfires
  - Biogenic emissions
    - Mold and spores

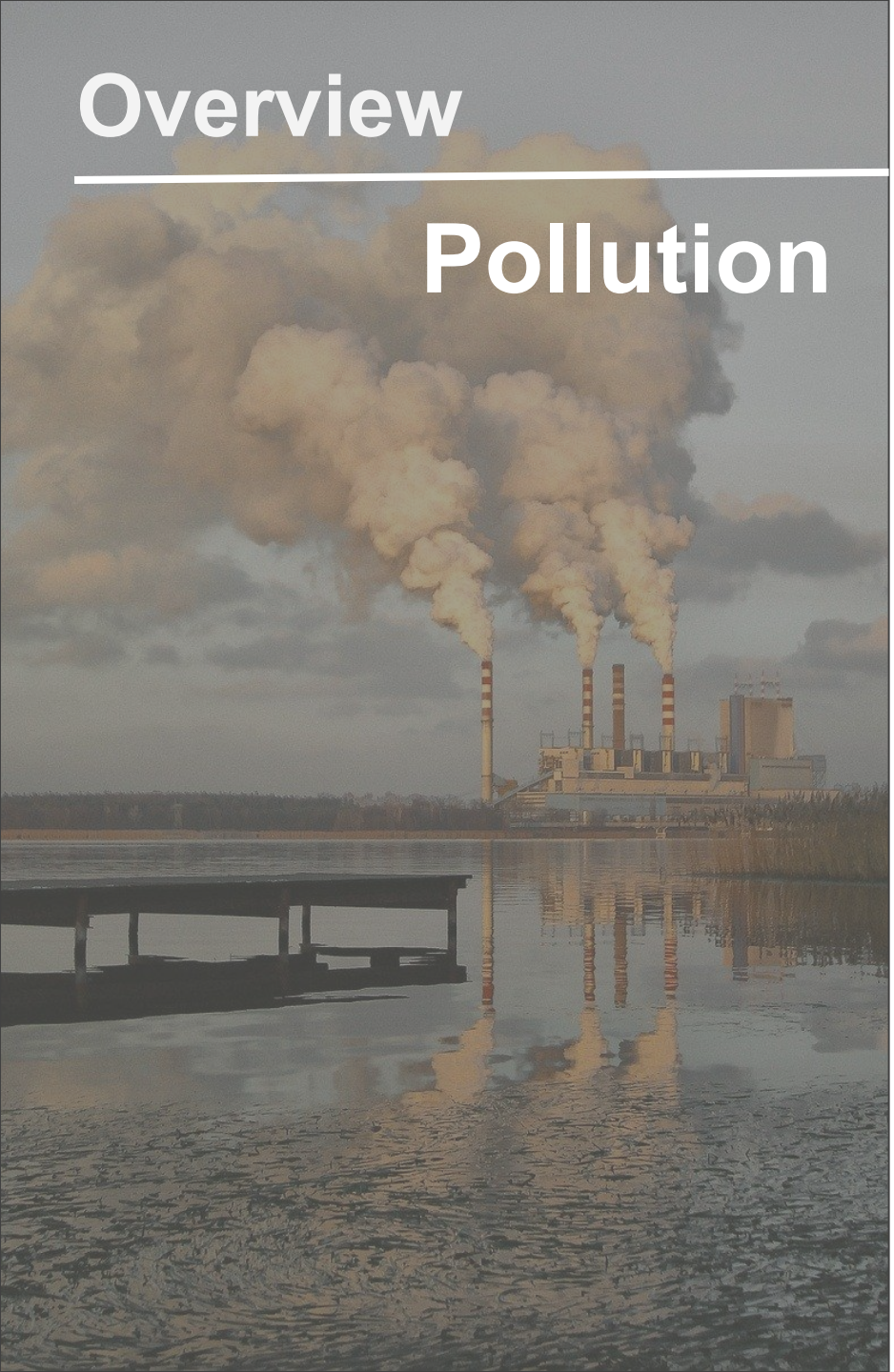
Overview

Pollution

- **Household / Indoor Pollution**
  - Asbestos
  - Toxic home cleaning products
  - Radon
  - Tobacco smoke
  - Wood fires and heating appliances
  - Biological pollutants
    - Mold, bacteria, etc.
  - Varnishes and paints
  - Synthetic fragrances and candles
  - Pesticides

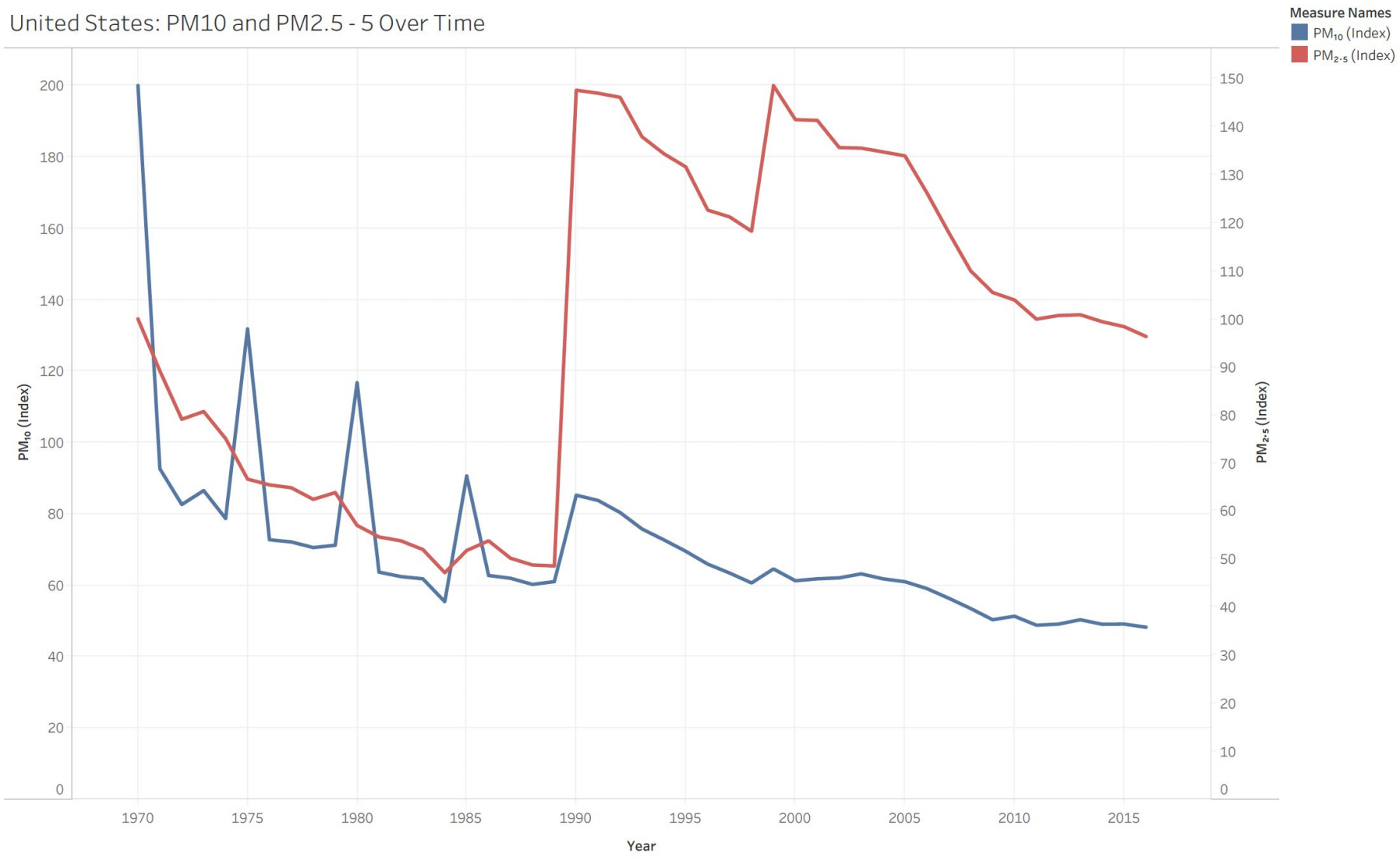
Overview

Pollution



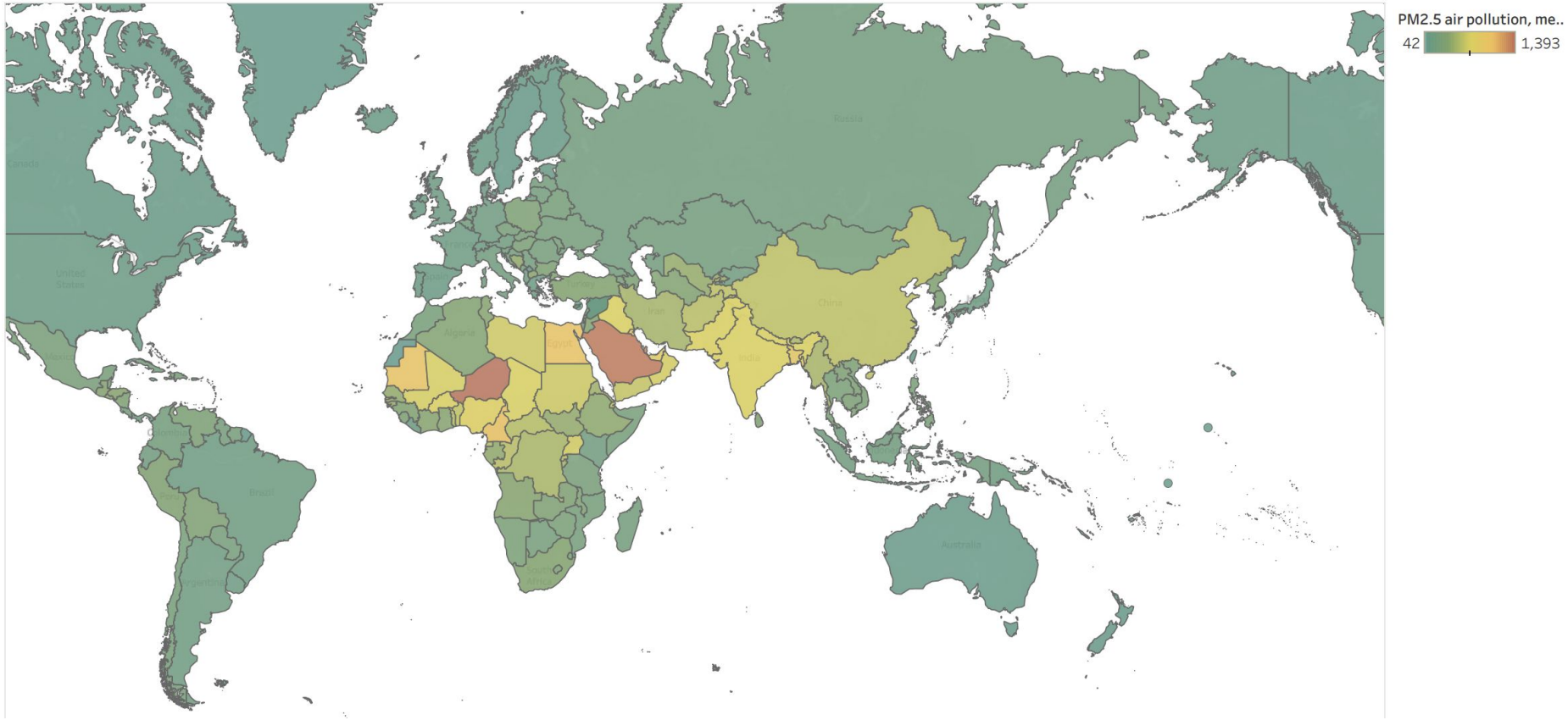
- **Particulate matter (PM)**
  - Ambient air pollutant
- **PM2.5**
  - Particles 2.5μm or smaller
  - Bi-product of combustion
- **PM10**
  - Particles 10μm or smaller
  - Mechanically generated dusts

Overview


United States: PM10 and PM2.5 - 5 Over Time
Measure Names
PM10 (Index)
PM2.5 (Index)
PM10 (Index)
PM2.5 (Index)
Year
0
20
40
60
80
100
120
140
160
180
200
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

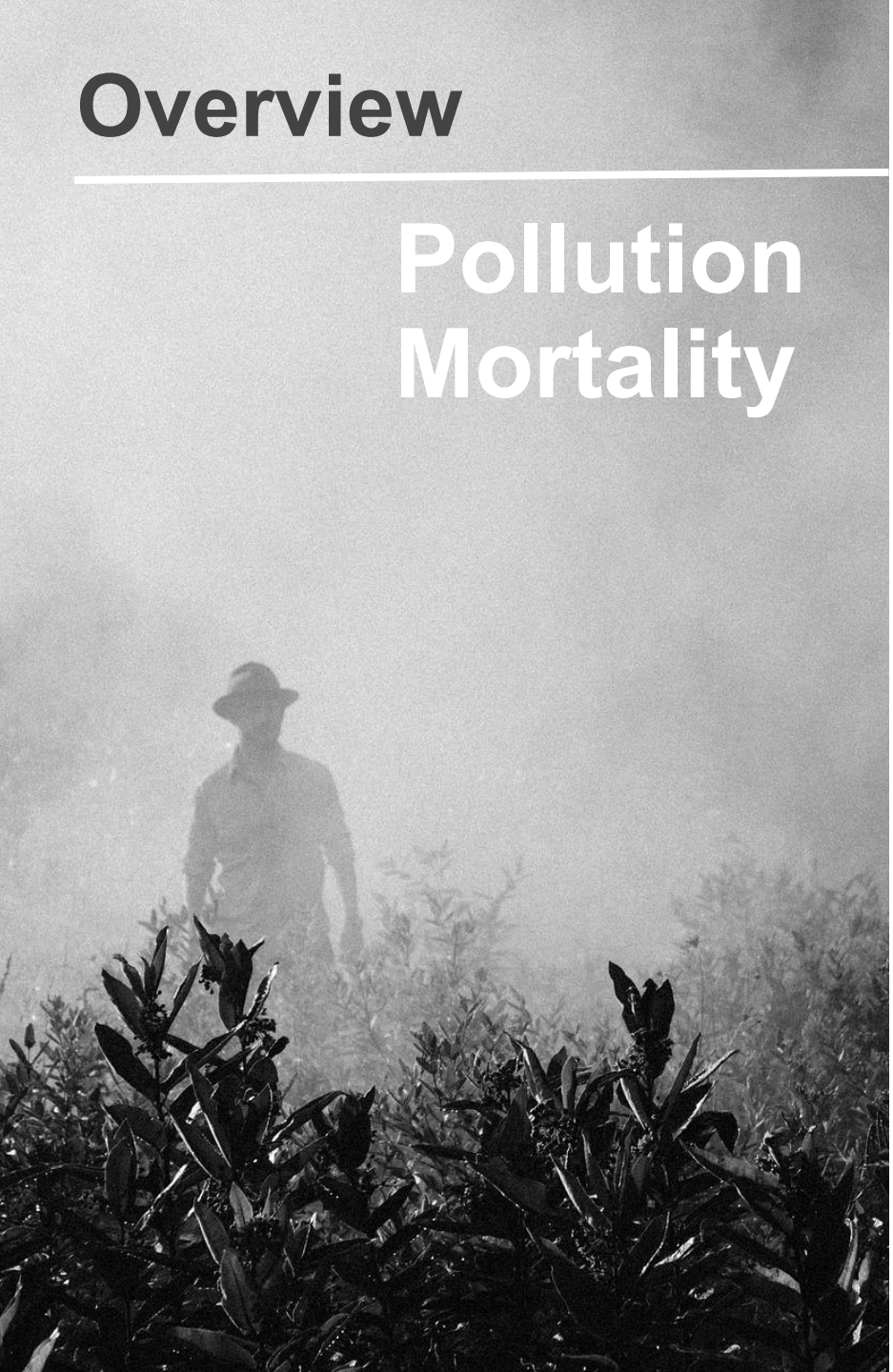
Overview

Worldwide PM2.5 Exposure



Overview

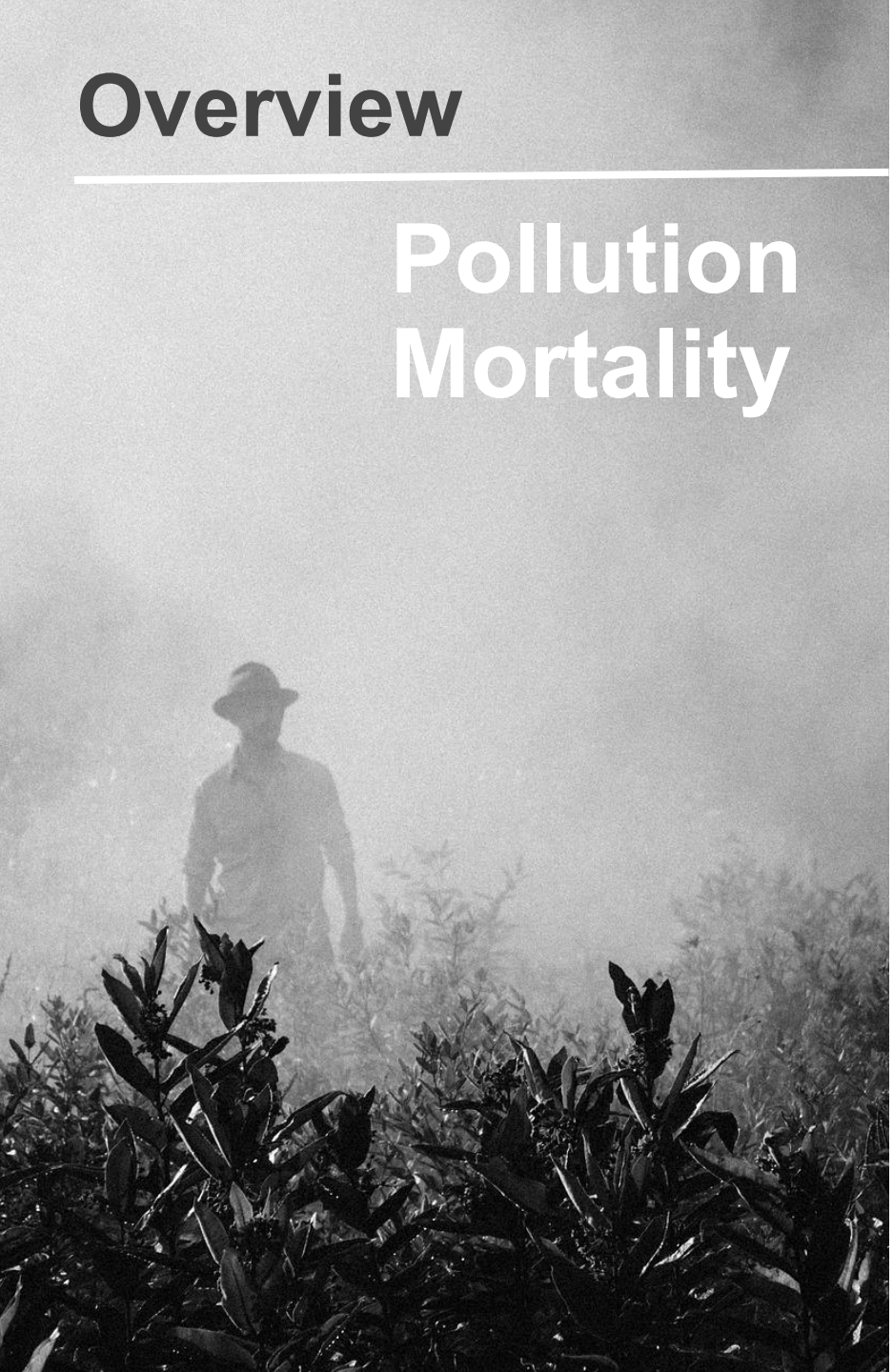
Pollution Mortality



- Not as simple as acute death by pollution.
  - Only 1.4% of pollution mortality is acute.
    - Most occur on days with extreme temperature flux.

Overview

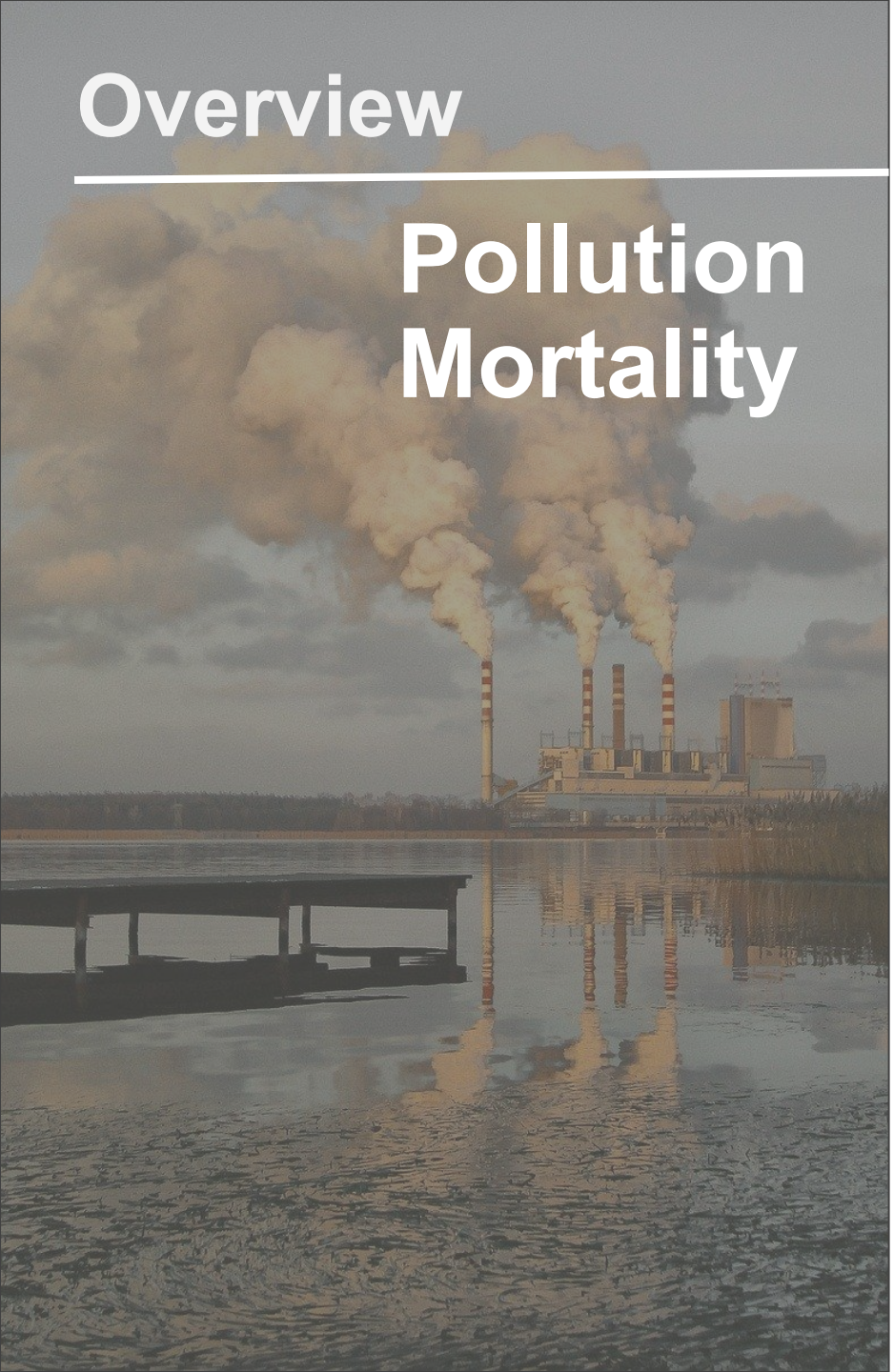
Pollution Mortality



- Defined by the World Health Organization (WHO) as a death that is premature due to pollution [1].
  - Stroke
  - Heart Disease
  - Pulmonary Disease
  - Lung Cancer
  - Acute Respiratory Infection

Overview

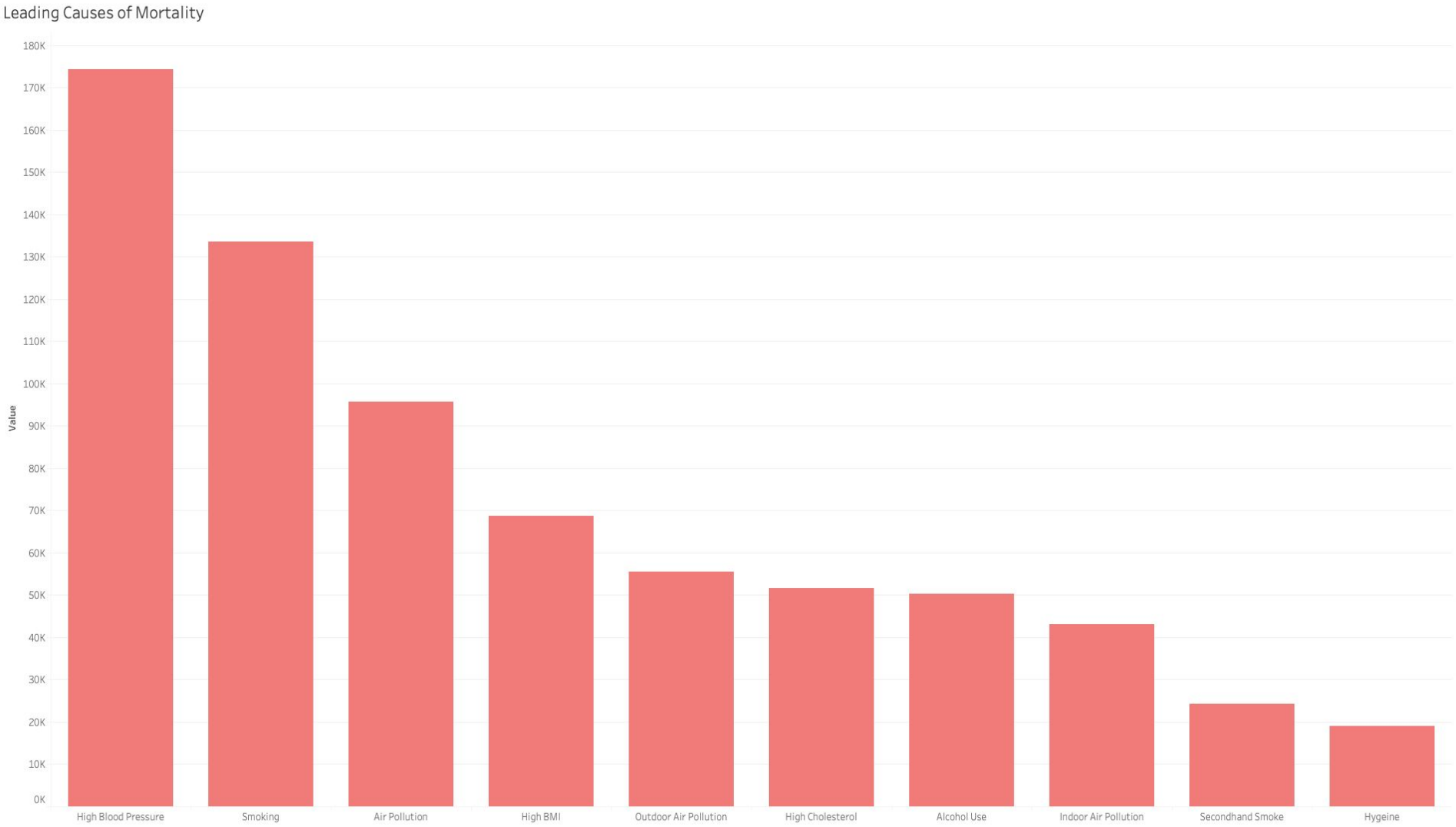
Pollution Mortality



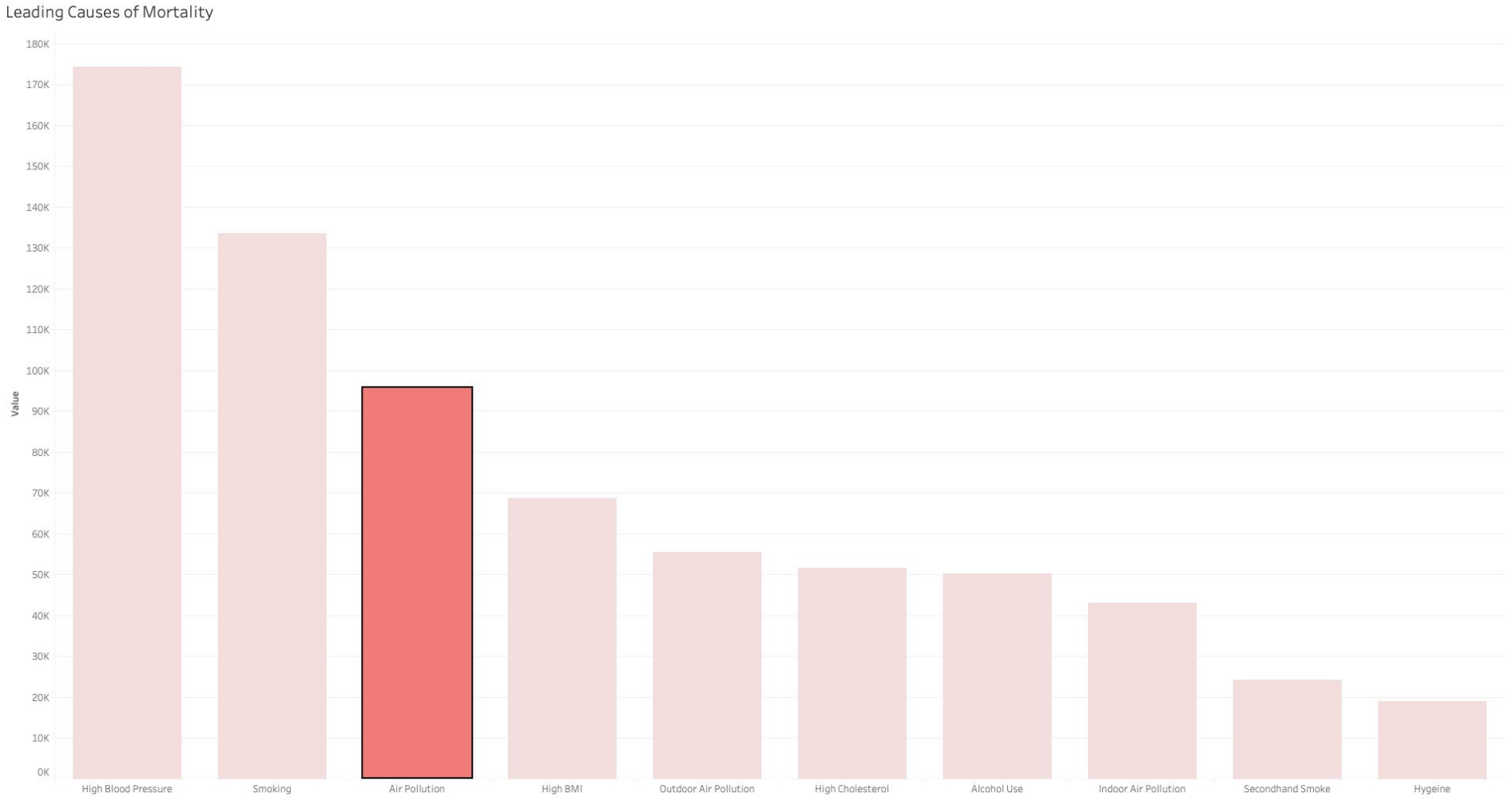
- 8 million annual deaths [1]
  - Ambient: 4.2 mil
  - Household: 3.8 mil
  - 1 in 8 deaths worldwide est.
  - Some researchers estimate as high as 40% of deaths caused by some form of pollution [4]

World Health Organization

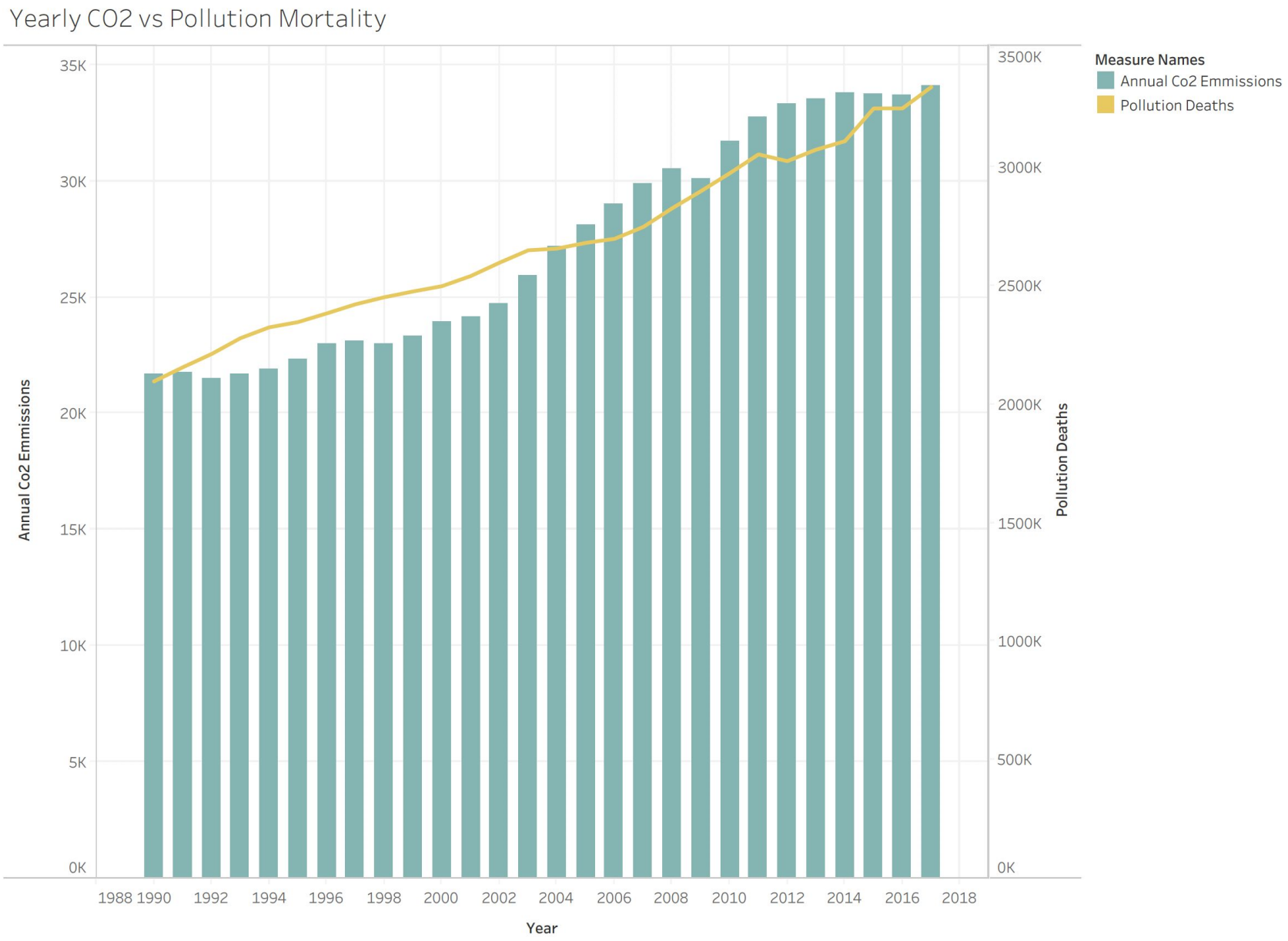
Overview


Leading Causes of Mortality
Value
180K
170K
160K
150K
140K
130K
120K
110K
100K
90K
80K
70K
60K
50K
40K
30K
20K
10K
0K
High Blood Pressure
Smoking
Air Pollution
High BMI
Outdoor Air Pollution
High Cholesterol
Alcohol Use
Indoor Air Pollution
Secondhand Smoke
Hygeine

Overview

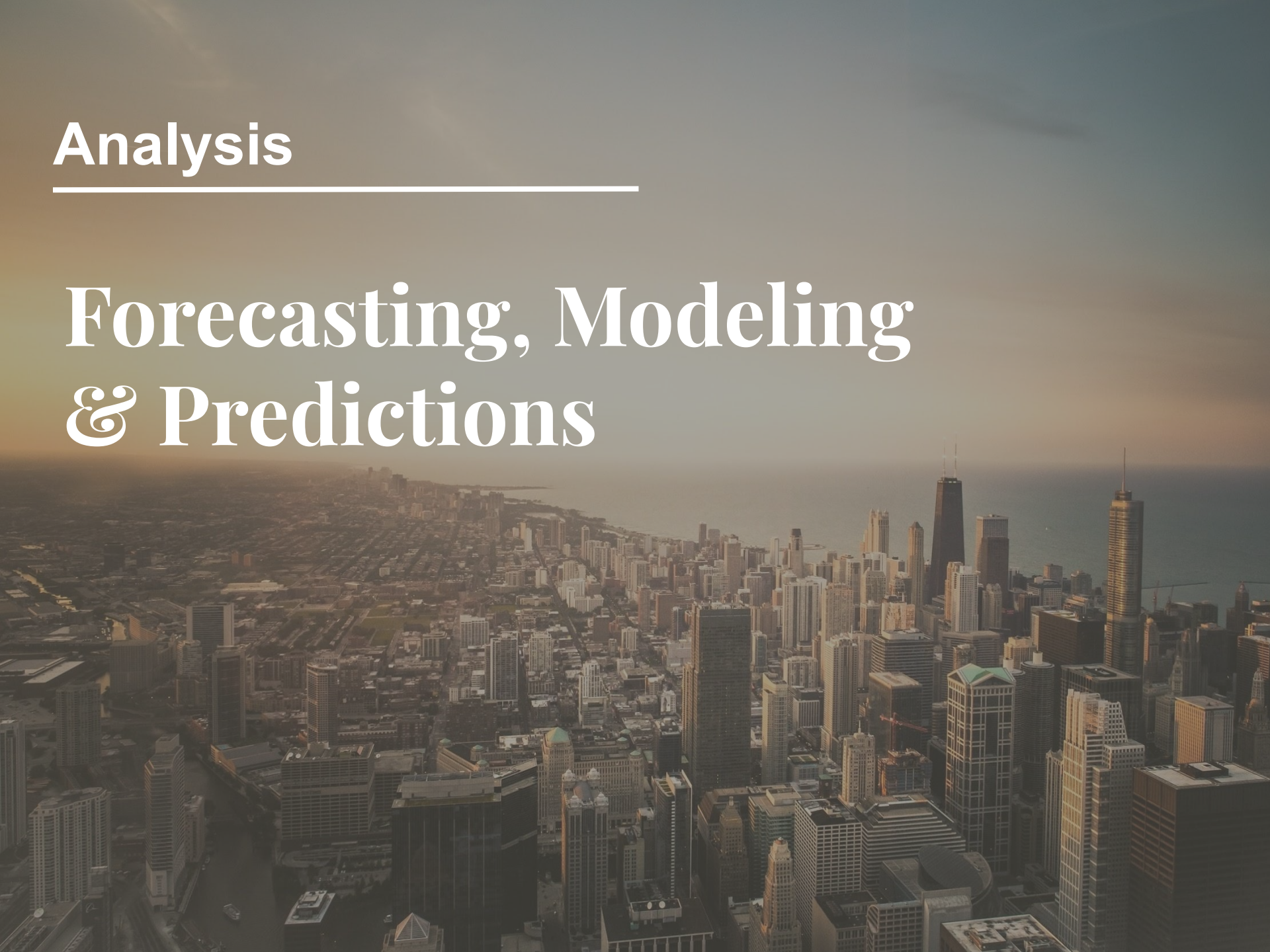

Leading Causes of Mortality
Value
180K
170K
160K
150K
140K
130K
120K
110K
100K
90K
80K
70K
60K
50K
40K
30K
20K
10K
0K
High Blood Pressure
Smoking
Air Pollution
High BMI
Outdoor Air Pollution
High Cholesterol
Alcohol Use
Indoor Air Pollution
Secondhand Smoke
Hygeine

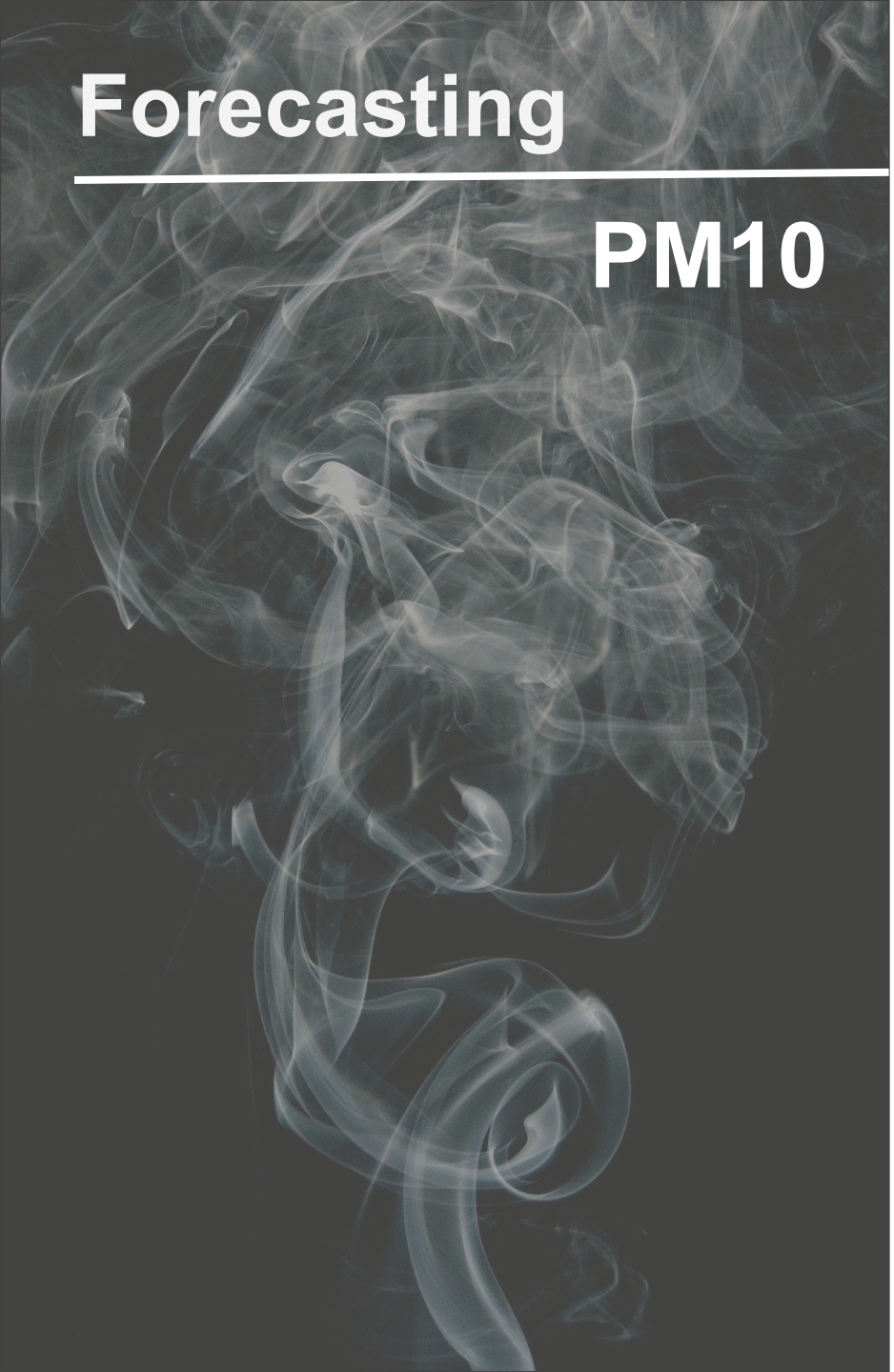
Overview


Yearly CO2 vs Pollution Mortality
Annual Co2 Emmissions
35K
30K
25K
20K
15K
10K
5K
0K
Pollution Deaths
3500K
3000K
2500K
2000K
1500K
1000K
500K
0K
1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018
Year
Measure Names
Annual Co2 Emmissions
Pollution Deaths

Analysis

Forecasting, Modeling & Predictions





Forecasting

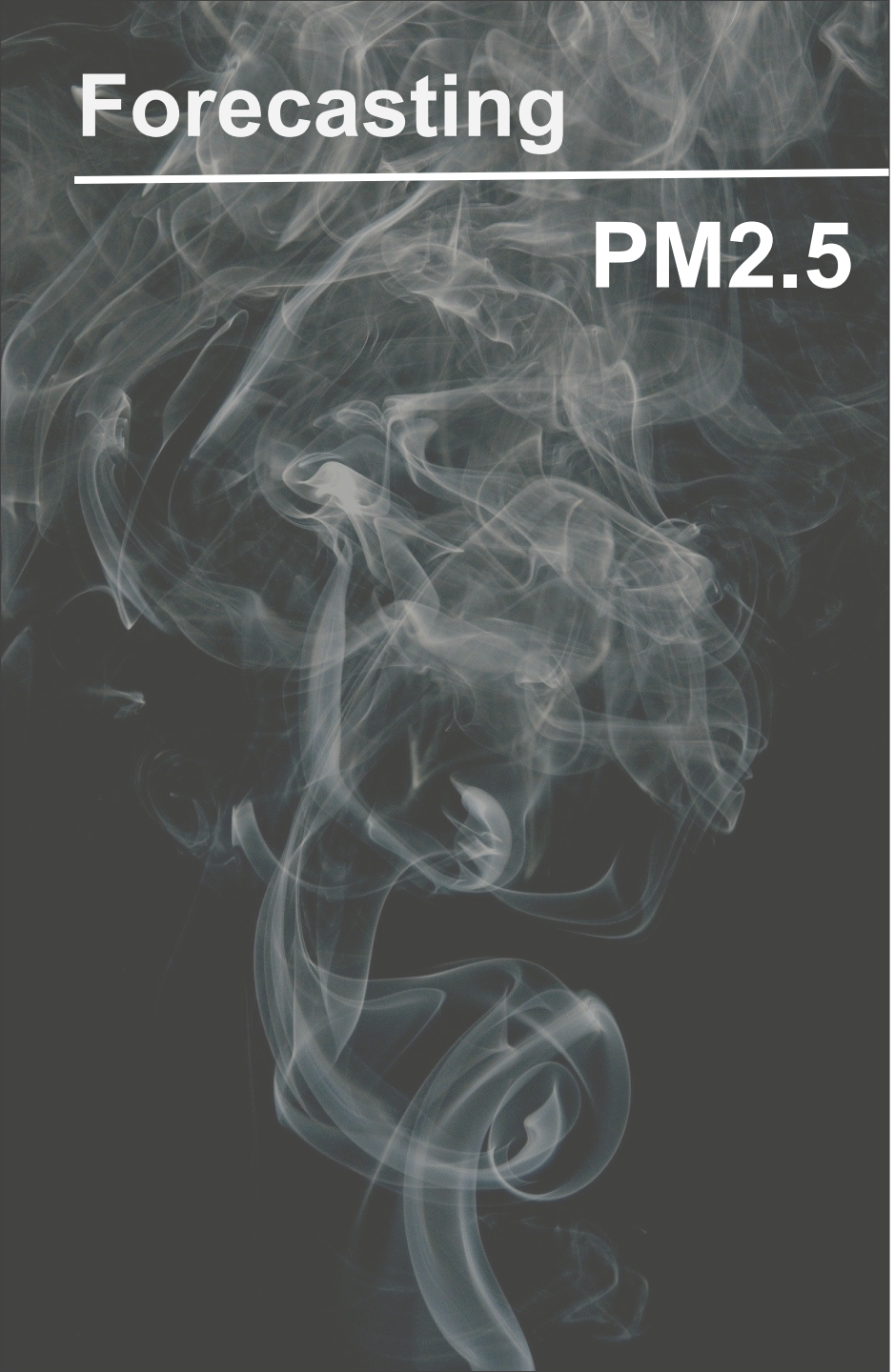
PM10

- **Summary of PM10 forecast**
  - one line explain
  - maybe two

Forecasting

PM10?

PM10 Forecasting GRAPH HERE



Forecasting

PM2.5

- **Summary of PM2.5 forecast**
  - one line explain
  - maybe two

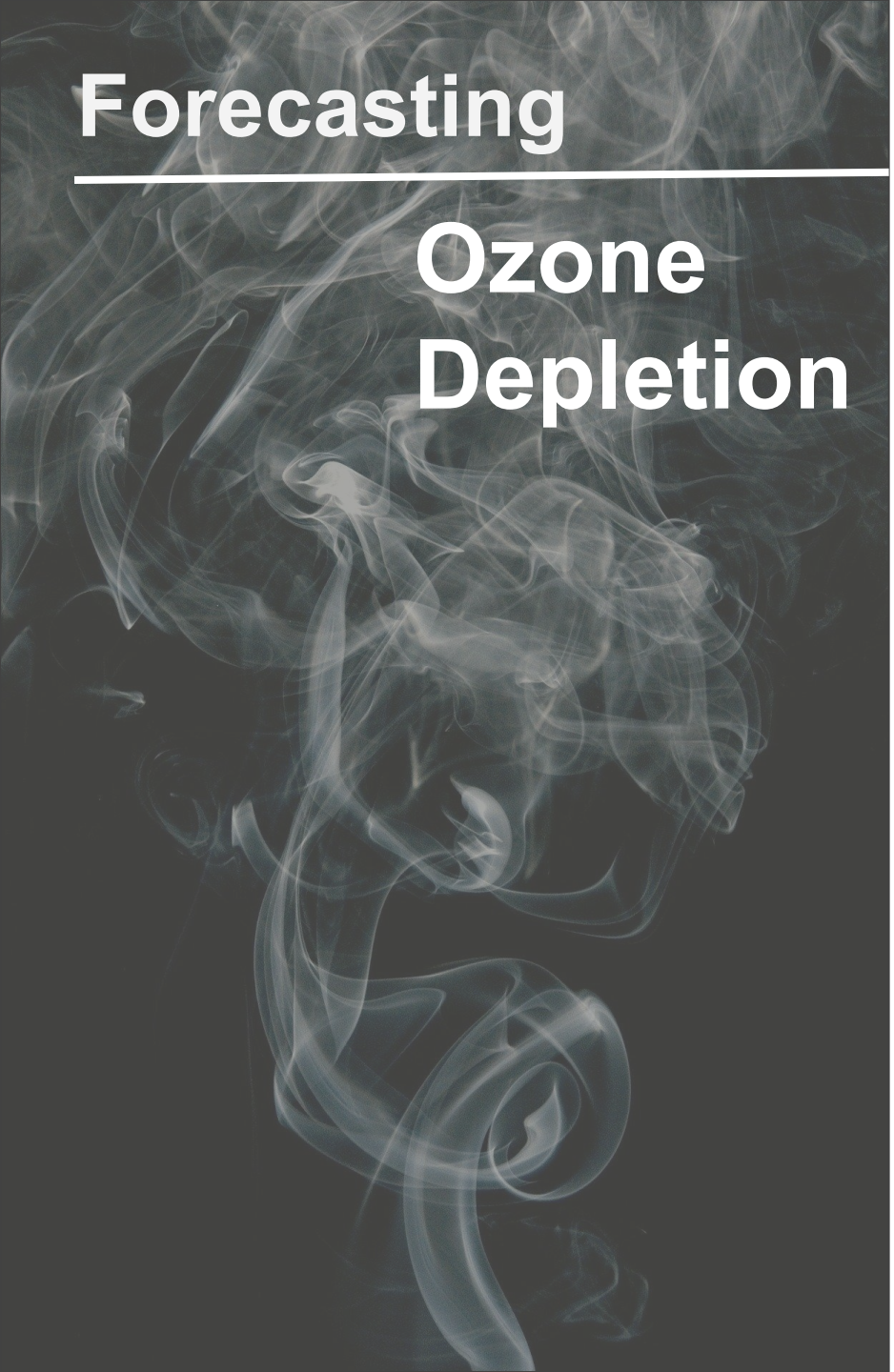
Forecasting

PM2.5

PM2.5 Forecasting GRAPH HERE

Forecasting

Ozone Depletion



- **Summary of Ozone forecast**
  - one line explain
  - maybe two

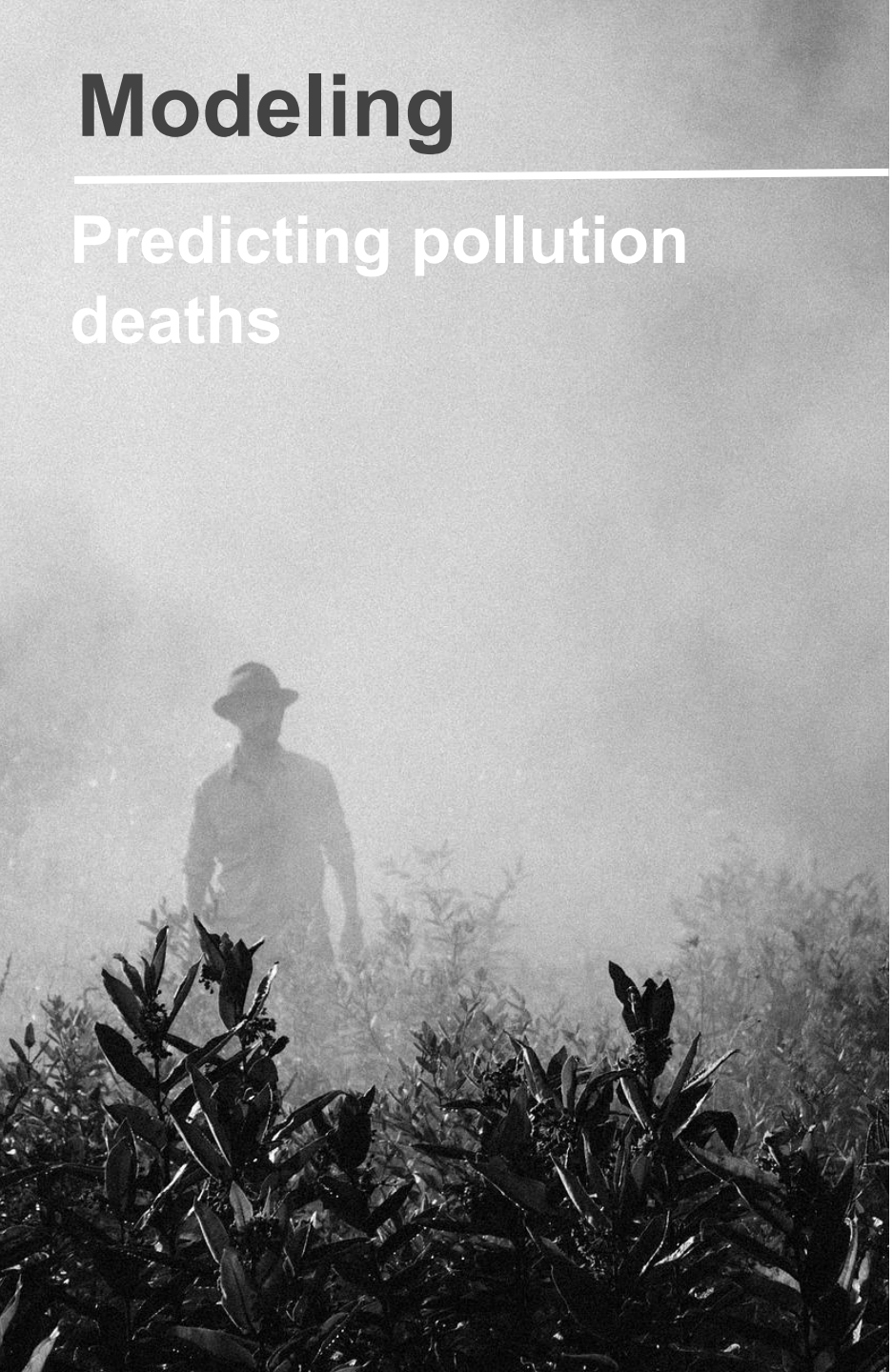
Forecasting

Ozone Depletion

Ozone Forecasting GRAPH HERE

Modeling

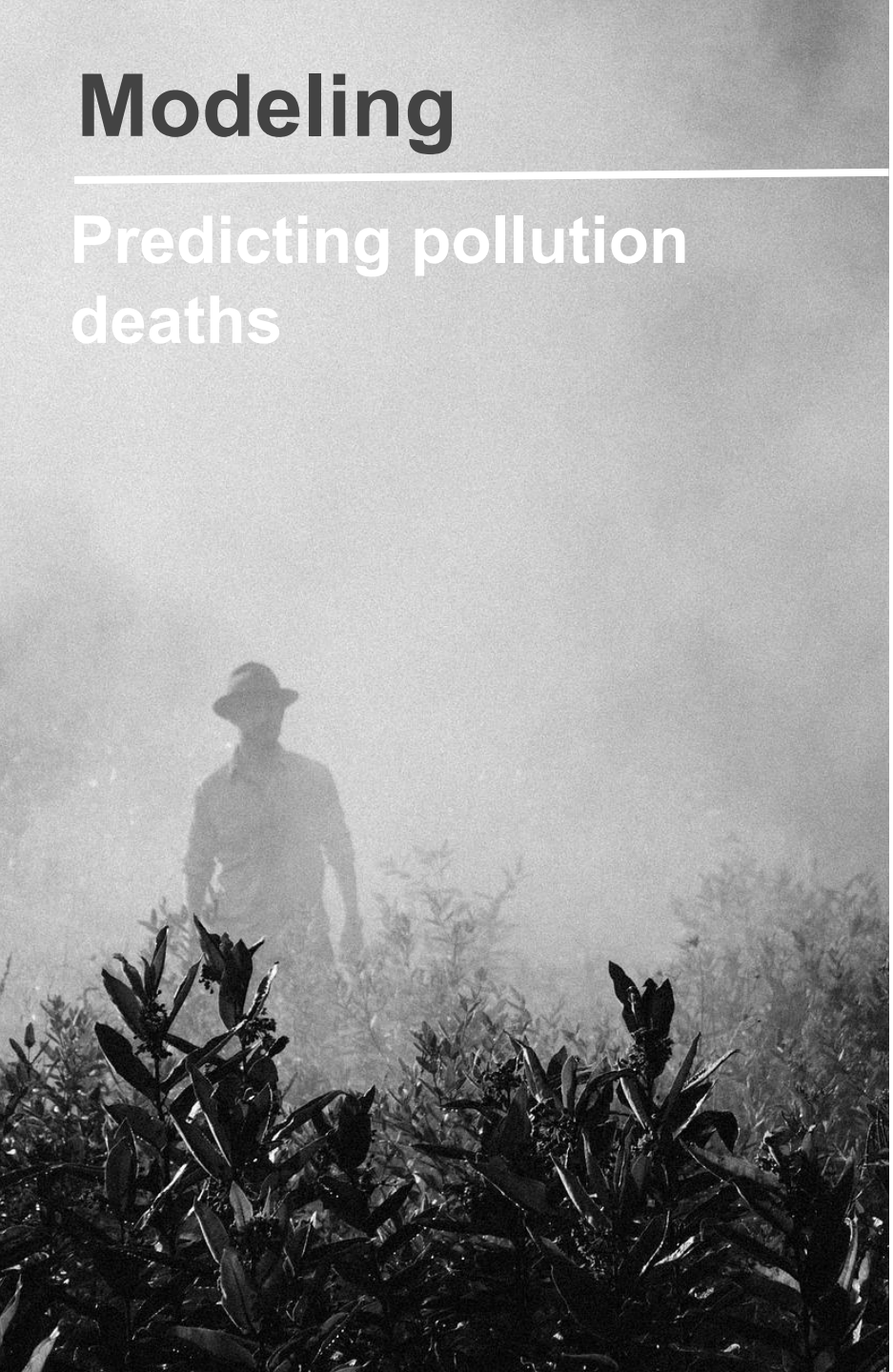
Predicting pollution deaths



- **Final Variables**
  - Annual CO_2 emissions
  - National health spending
    - to account for other factors of disease
  - Life expectancy
    - to balance other means of mortality
  - Ozone depleting emissions
  - Mean daily ozone
  - Population
    - Greatly correlated with mortality, therefore results will be reported with and without population.

Modeling

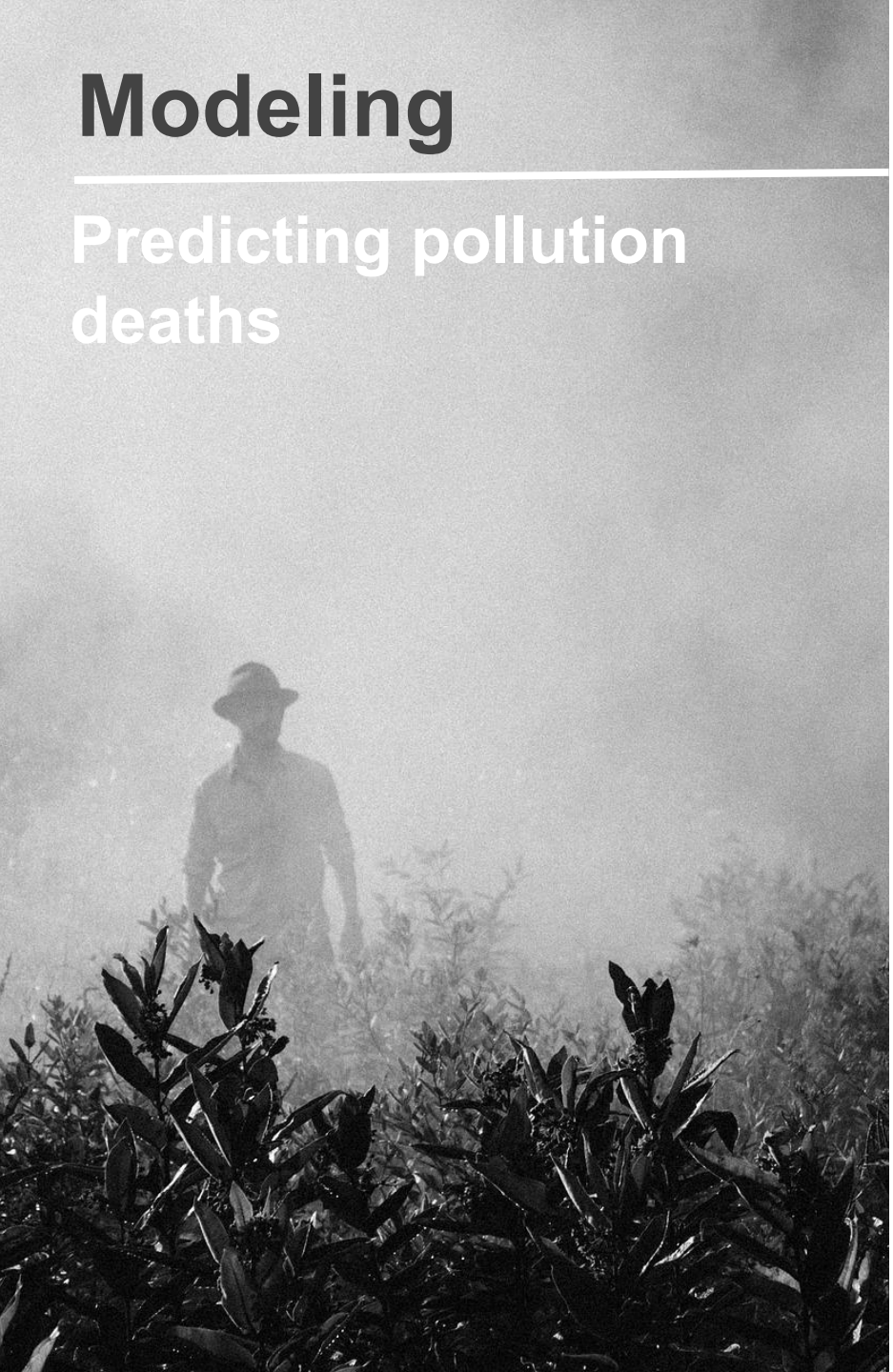
Predicting pollution deaths



- **Final Variables**
  - Total pollution mortality
    - Explored utilizing a crude rate.
    - Ultimately, it was more appropriate to predict based on total deaths, due to the other variables being based on summed values.
    - The option to convert to a rate is available post-analysis.

Modeling

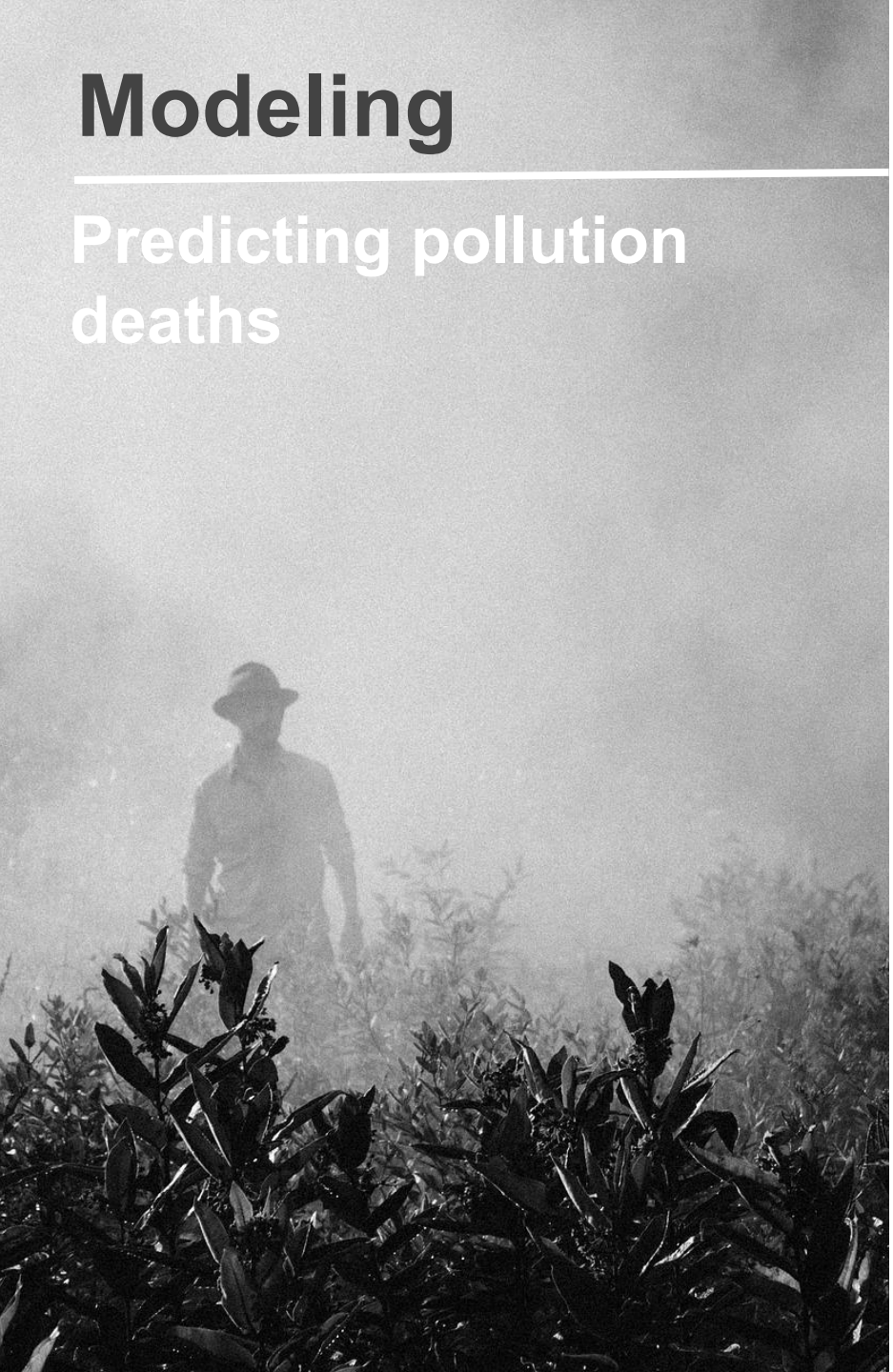
Predicting pollution deaths



- **LassoLARS**
  - Highest performing regression
    - Compared to many other regressions, including neural network.
  - Pre-processing:
    - Standard scaling
    - Imputing by mean
    - Polynomial features
    - PCA yielded ~ the same results
  - Best parameters:
    - Alpha val = 3
    - Normalization = True

Modeling

Predicting pollution deaths



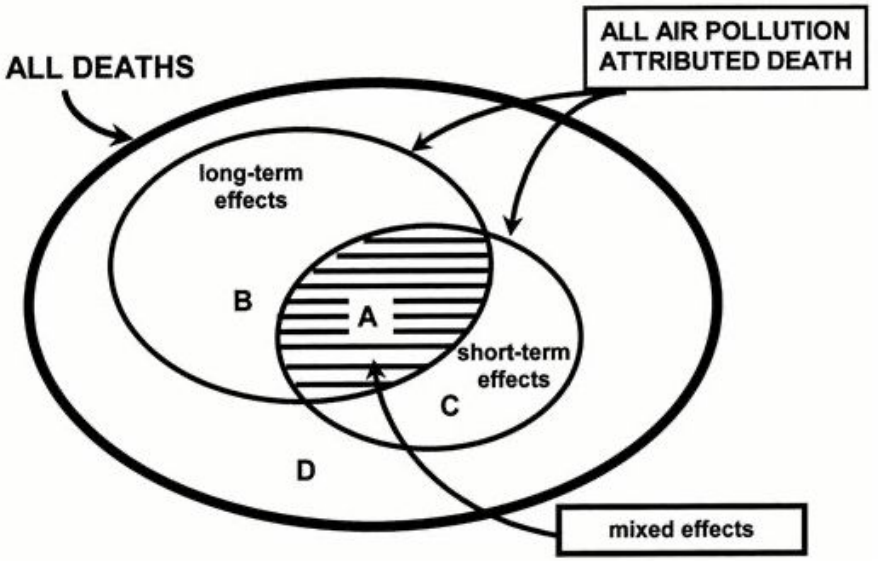
- **LassoLARS**
  - **Without population:**
    - r^2: **0.82**
    - RMSE: **43k**
  - **With population:**
    - r^2: **0.99**
    - RMSE: **7424**
  - **Null Hypothesis**
    - RMSE: **104k**

Analysis

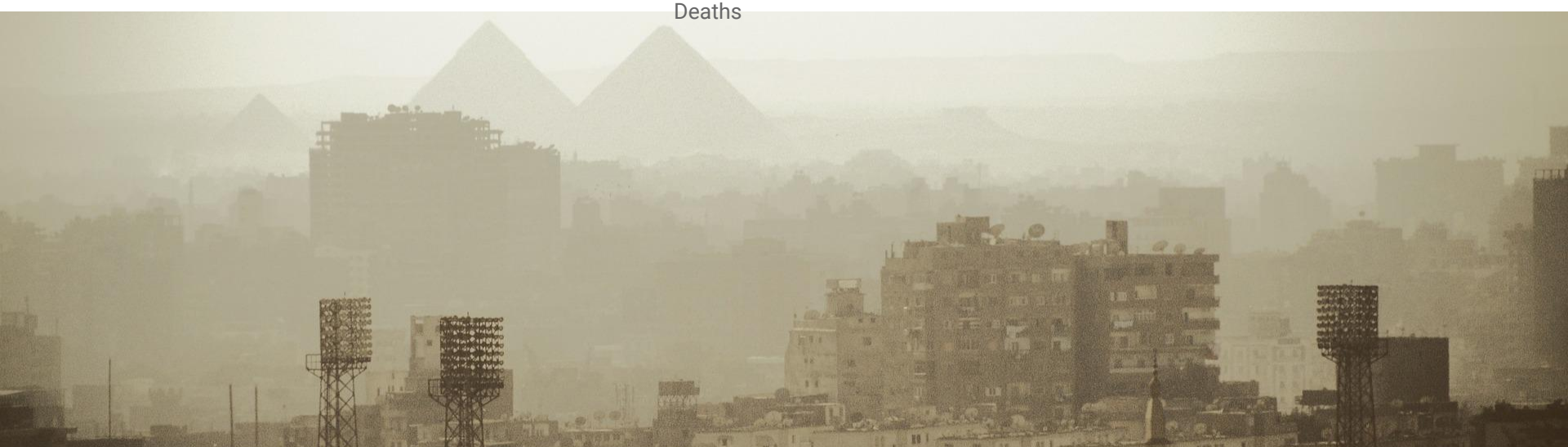
Challenges & Limitations

Challenges & Limitations

- **Lag**
  - Pollution causes acute death in extreme cases only

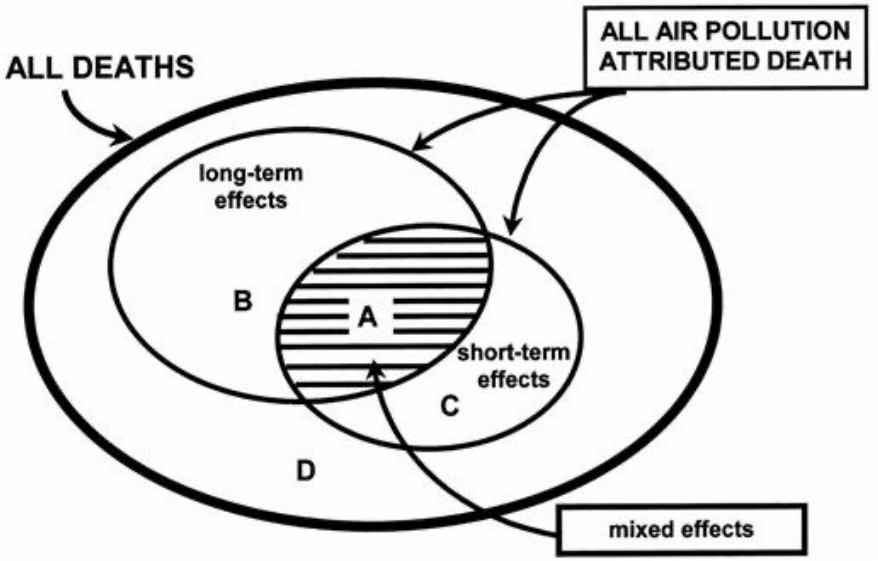


Source: The Distributed Lag between Air Pollution and Daily Deaths

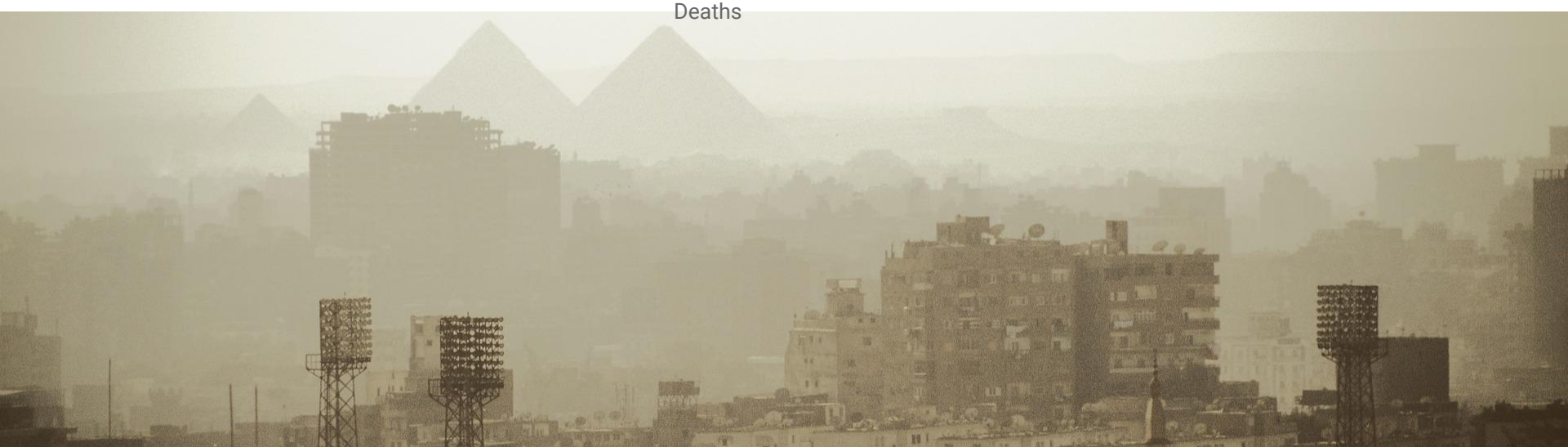


Challenges & Limitations

- **Lag**
  - Subject matter experts have yet to agree on a standard for calculating lag.
  - Most reports ignore lag.

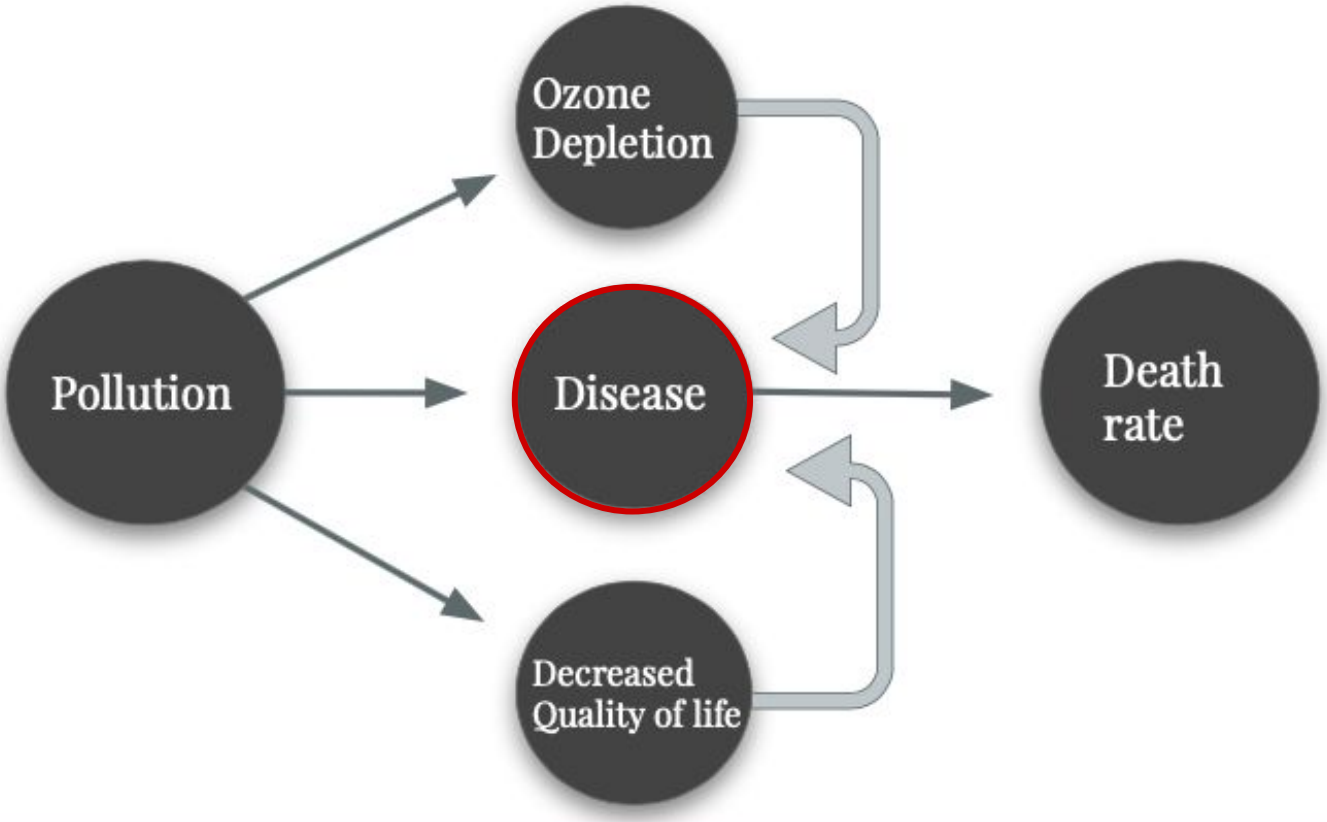


Source: The Distributed Lag between Air Pollution and Daily Deaths



Challenges & Limitations

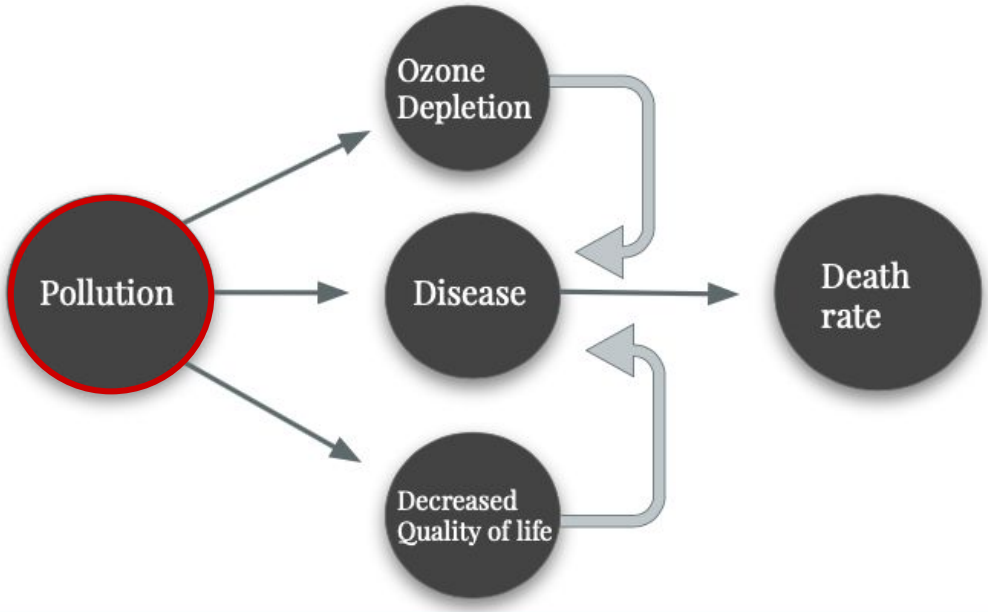
Complicated disease mechanism



Challenges & Limitations

Complicated disease mechanism [3]

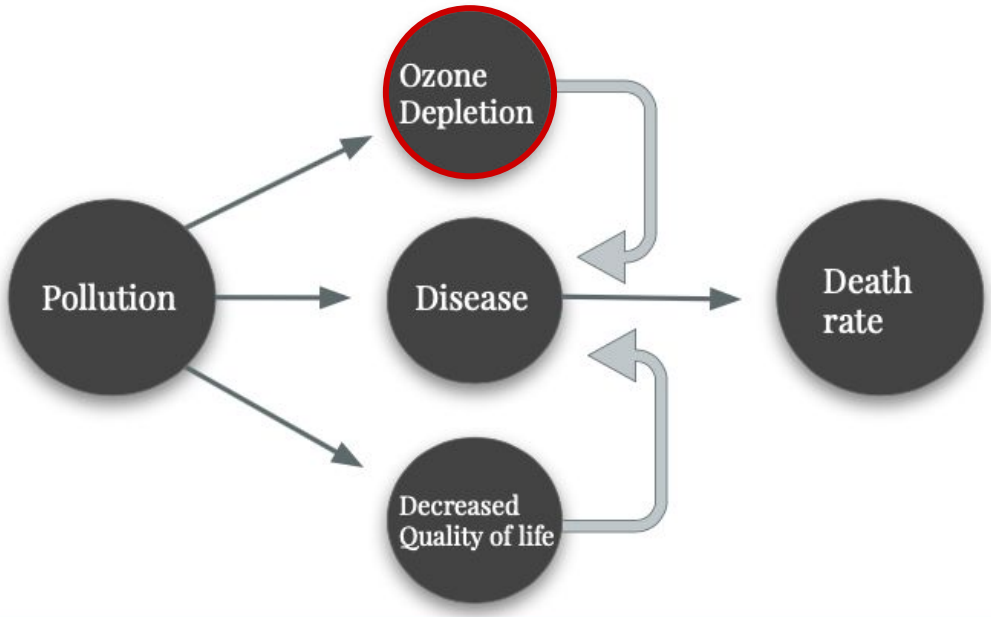
- PM10
  - Oxidative stress
  - Inflammation
- PM2.5
  - Oxidative stress
  - Inflammation
  - Autonomous nervous system activation
  - Changes lung microbiome
  - Specific intracellular pathways still being discovered



Challenges & Limitations

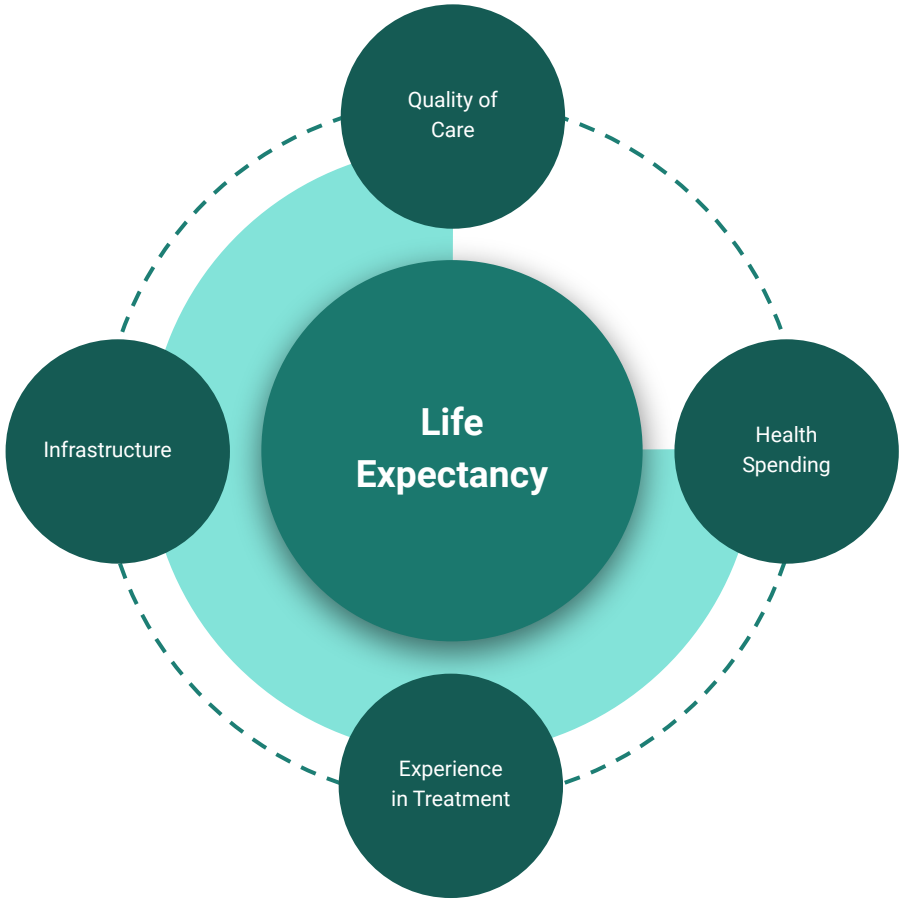
Complicated disease mechanism [3]

- Decreasing ozone
  - Higher rates of skin cancer



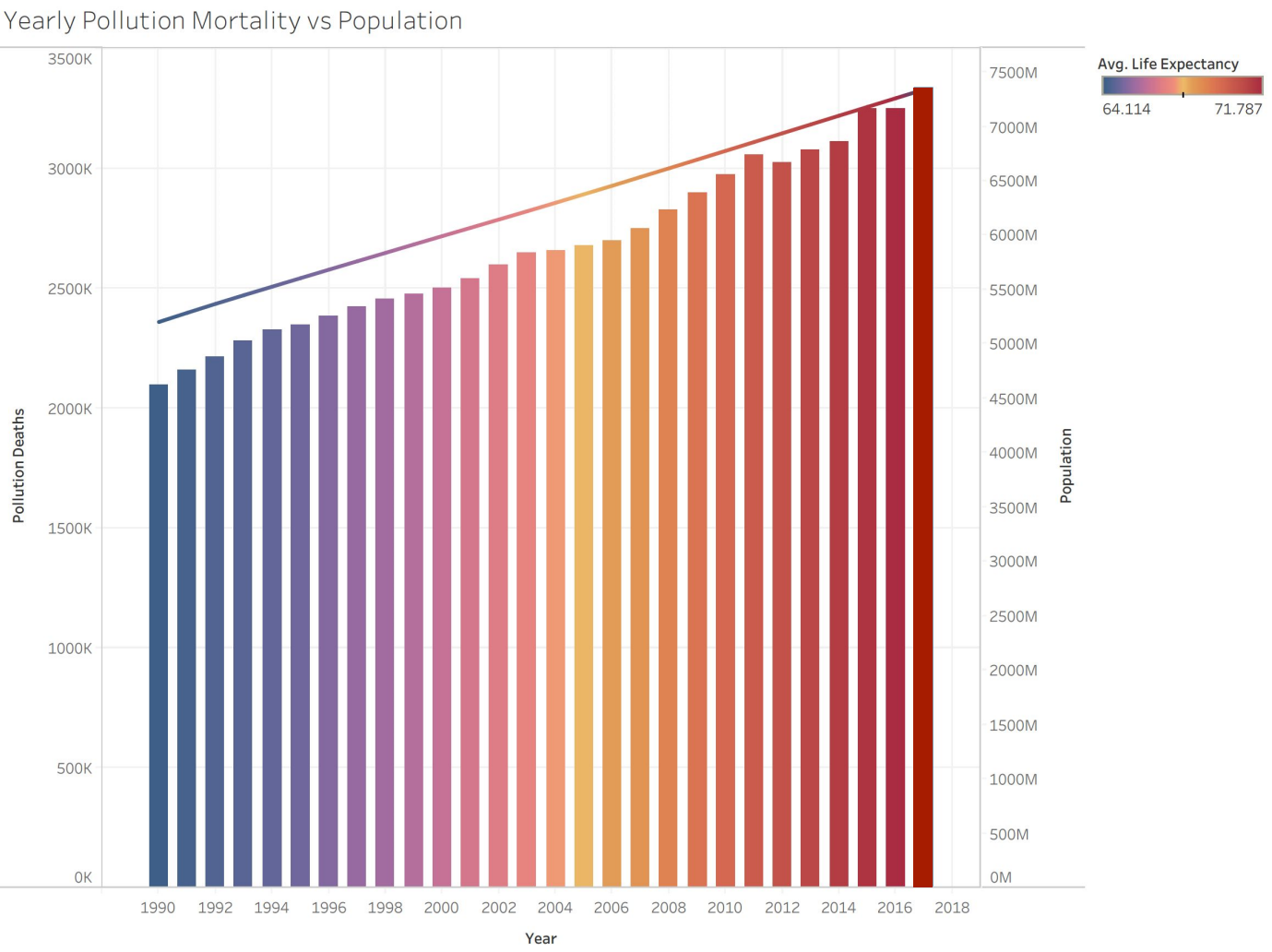
Challenges & Limitations

Healthcare and Quality of Life



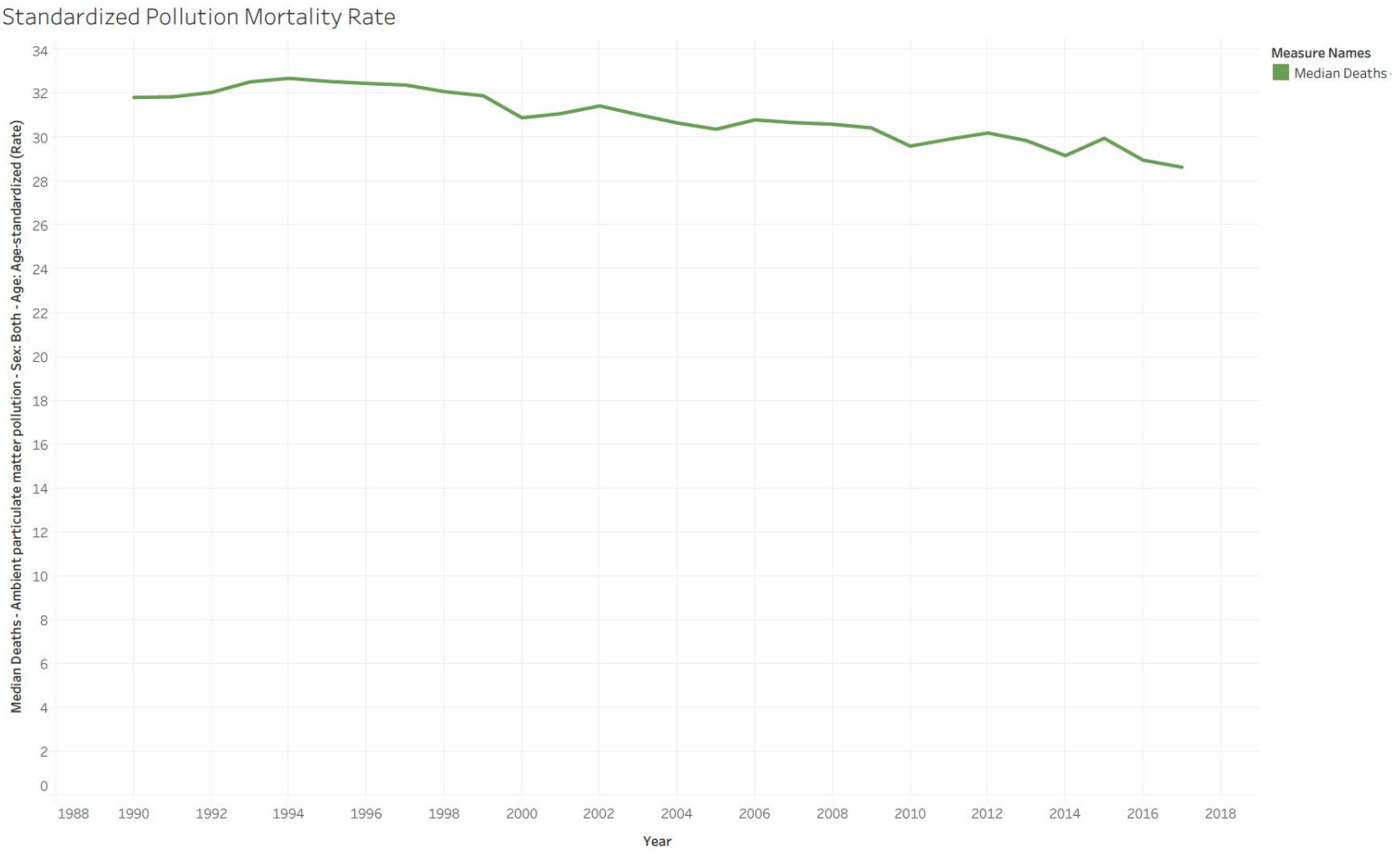
Challenges & Limitations

Pollution is highly correlated with population.

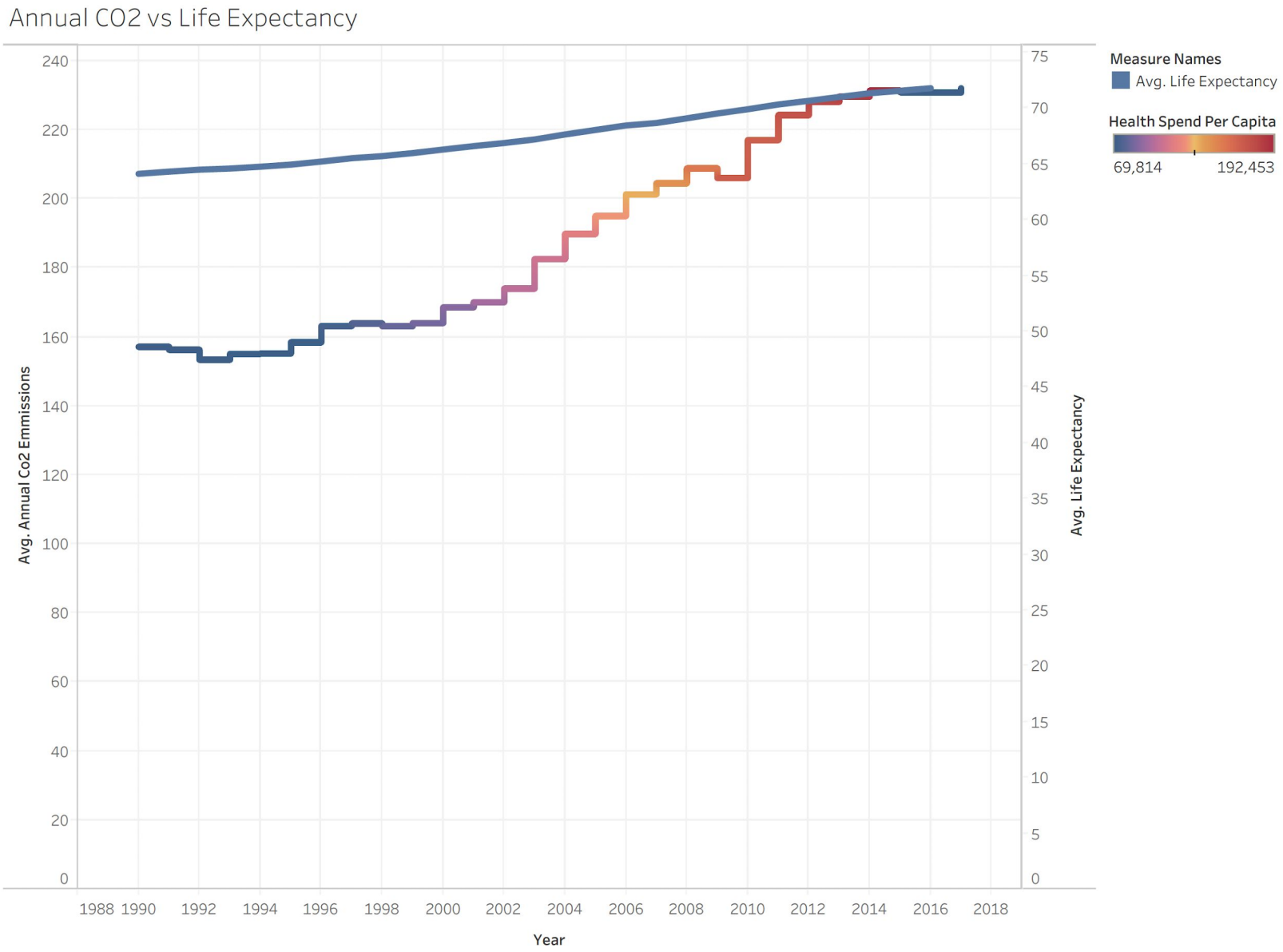


Challenges & Limitations

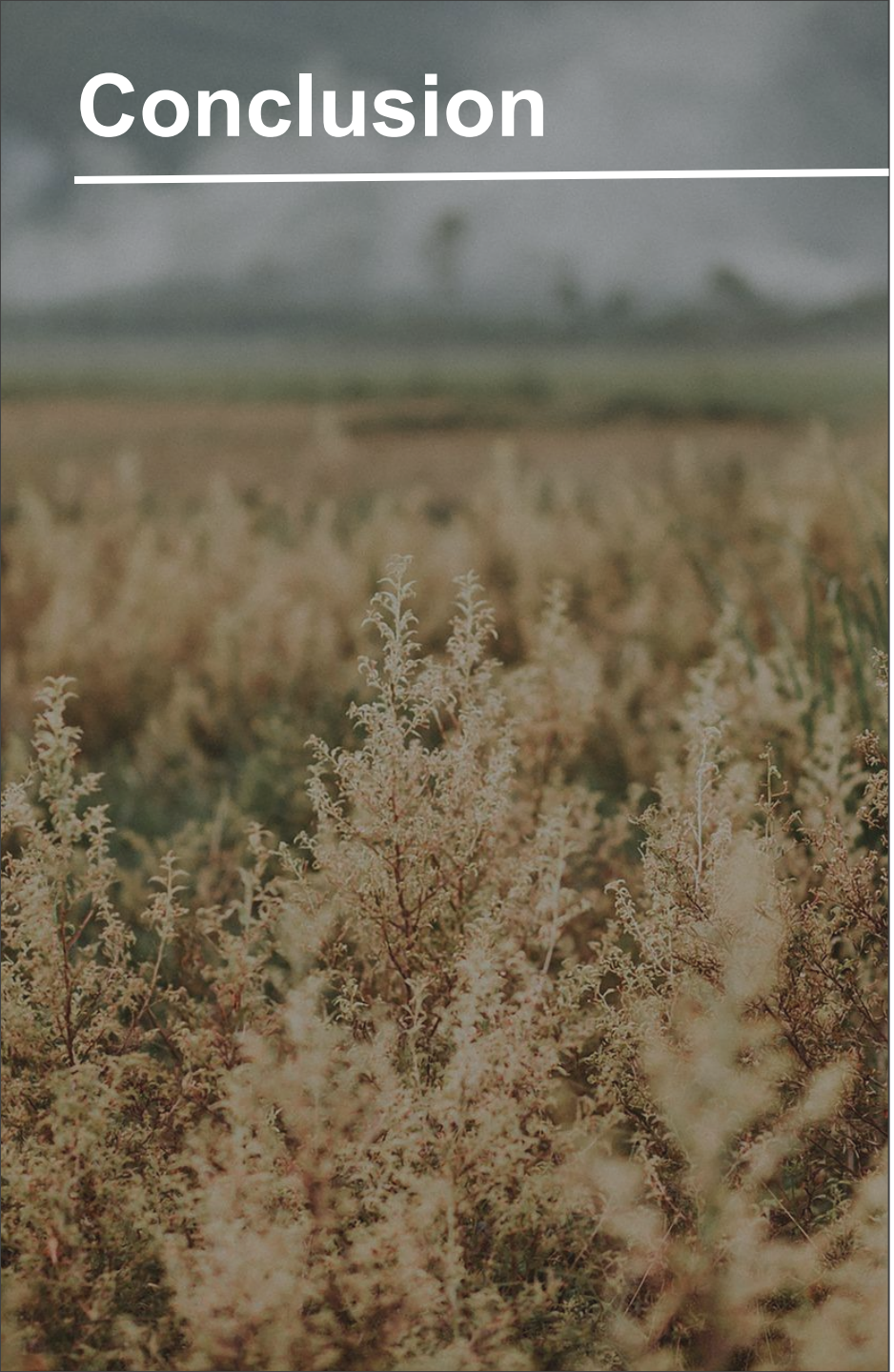
Population is directly correlated with mortality.



Challenges & Limitations

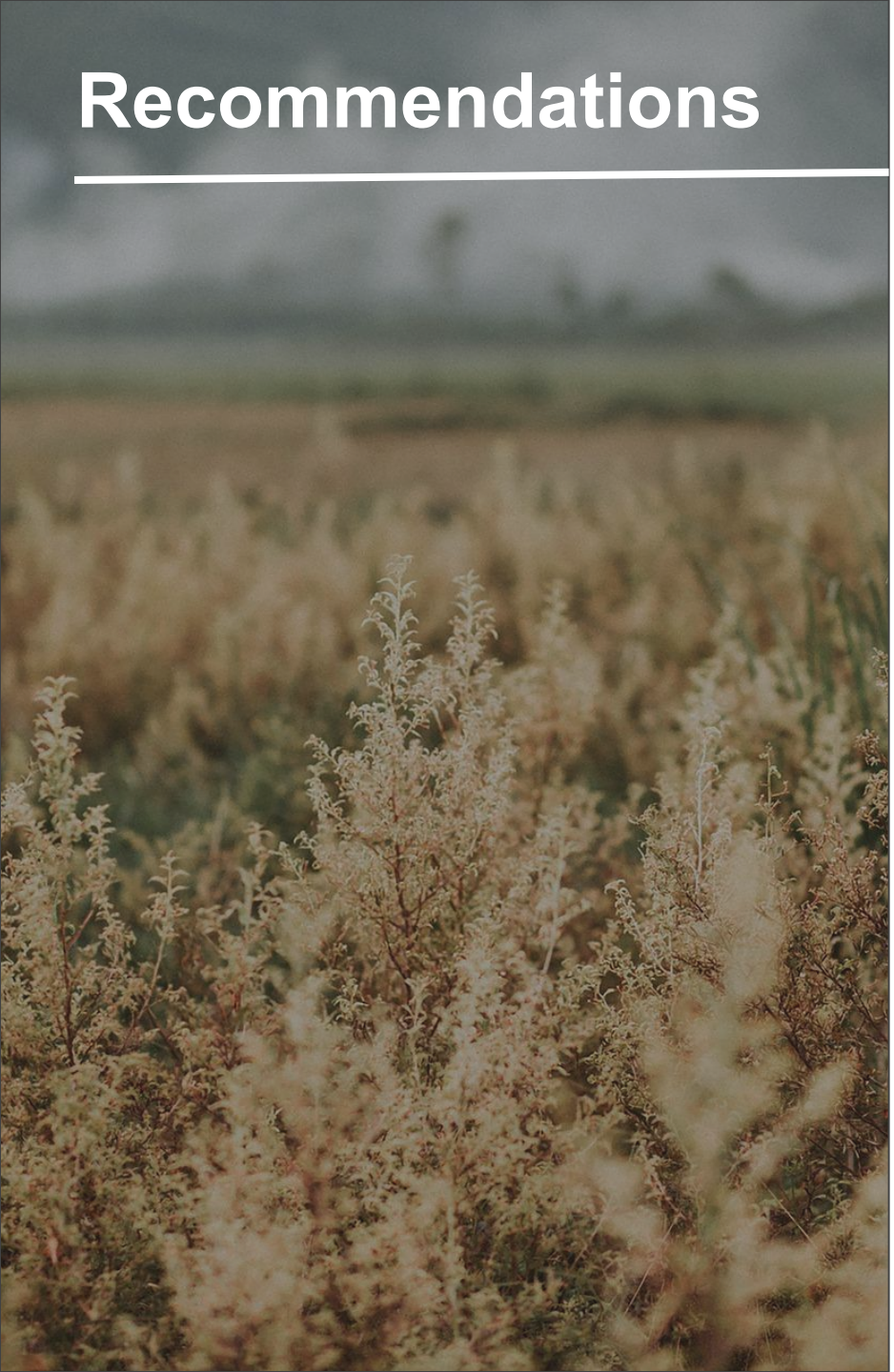


Conclusion



- Pollution is one of the leading influencers of premature mortality.
- Pollution mortality is rising with population, and can be predicted with high confidence.
- The per capita rate is offset by an increase in life-expectancy.
- The intertwined relationship between pollution metrics and population makes analysis complex and leads to indeterminate conclusions.

Recommendations



- Standardize a mathematical approach to represent lag.
- Separate analysis into groups by a metric such as GDP.
- Focus on disease mechanisms, and more targeted predictions.

Pollution Mortality

Forecasting Pollution Metrics with Machine Learning

Any Questions?

All images courtesy of Pixabay.com

Sources

1. Word Health Organization: https://www.who.int/health-topics/air-pollution#tab=tab_1
2. The Distributed Lag between Air Pollution and Deaths. https://www.researchgate.net/publication/12533027_The_Distributed_Lag_between_Air_Pollution_and_Daily_Deaths
3. The Mechanism of Air Pollution and Particulate Matter in Cardiovascular Diseases. https://pubmed.ncbi.nlm.nih.gov/28303426/
4. Pollution Causes 40 Percent of Deaths Worldwide: https://www.sciencedaily.com/releases/2007/08/070813162438.htm
5. National Particle Component Toxicity (NPACT) Initiative: integrated epidemiologic and toxicologic studies of health effects or particulate matter components: https://pubmed.ncbi.nlm.nih.gov/24377209/