HSPH Capstone Project

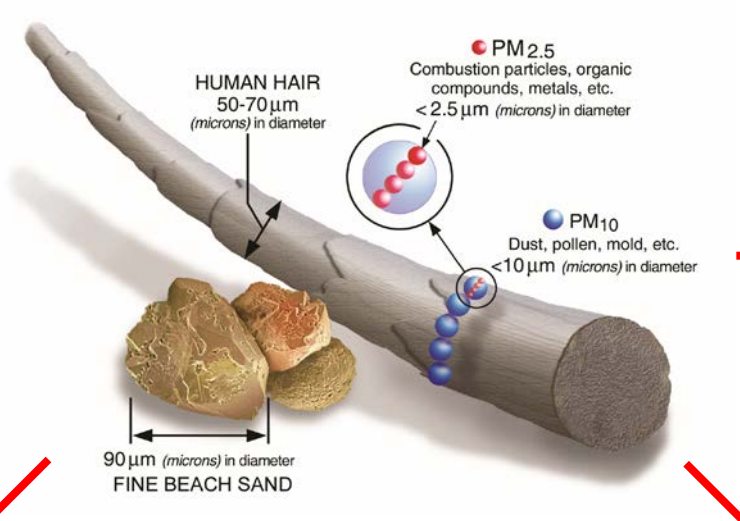
Air Pollutant Models

Keyan Halperin
Justin Lee
Chris Hase
Casey Meehan

Pollution Smollution…

Asthma

Heart Attack



Lung Cancer

Stroke

Partner Mission:

What's hurting us? How? What do we do?

1

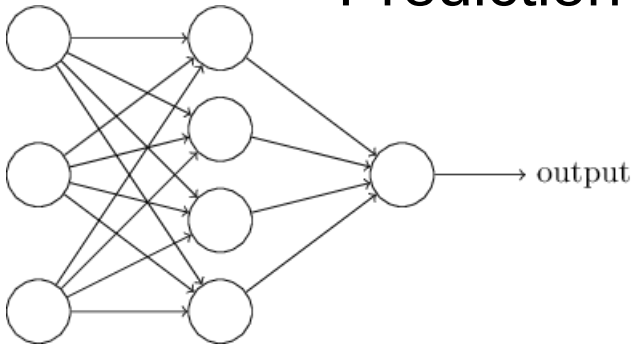
2

3

Confidence with Causation

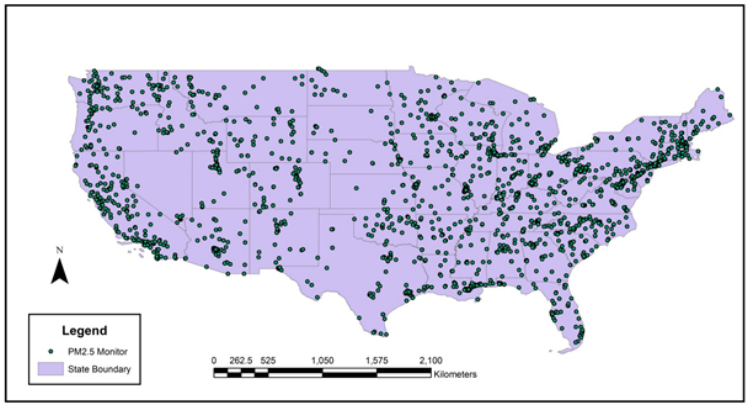
Geographic Data

Pollution Prediction



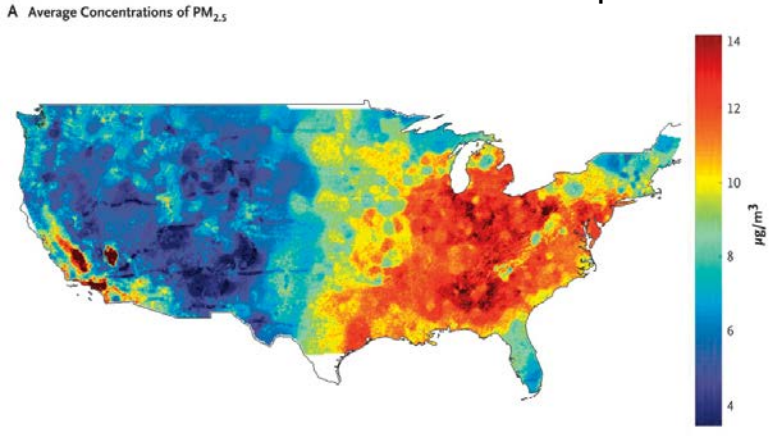
Causal Health Outcomes

$PM_{2.5}$ Predictions



Sensor Network

Prediction Map

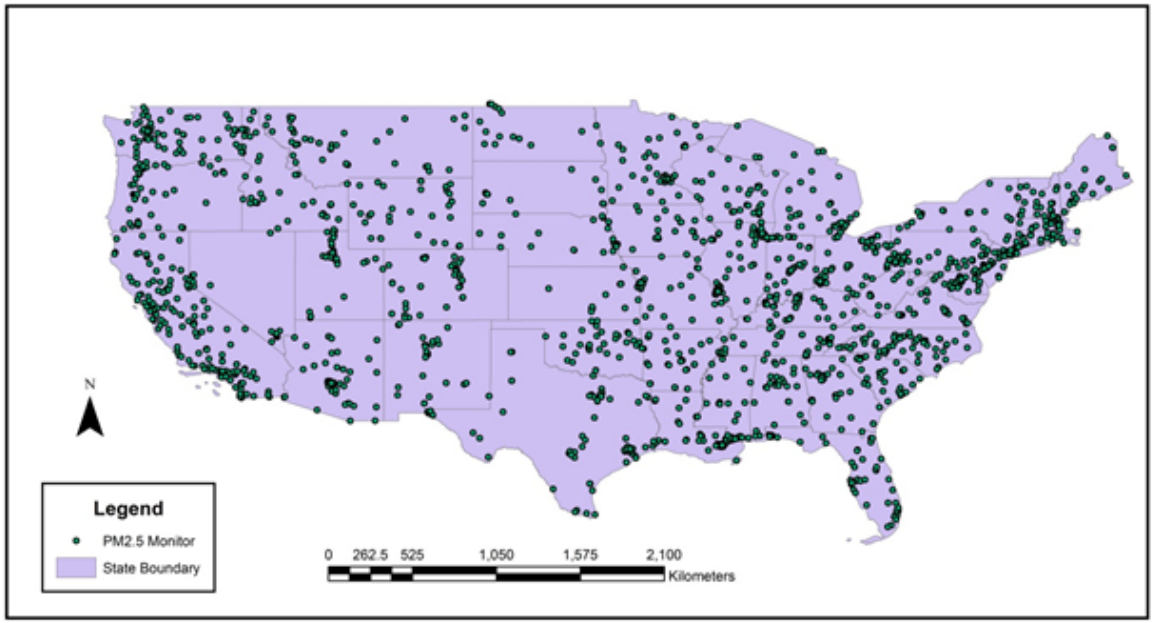


- We have $PM_{2.5}$ sensor outputs for locations with monitors
- Goal is to produce accurate $PM_{2.5}$ values for the entire US

$PM_{2.5}$ Sensors

- Daily $PM_{2.5}$ outputs from 2,156 monitors over the course of 16 years (2000 - 2016)

- Sensors tend to be located along the coasts and in more heavily populated areas



Model Inputs - Geographic Data

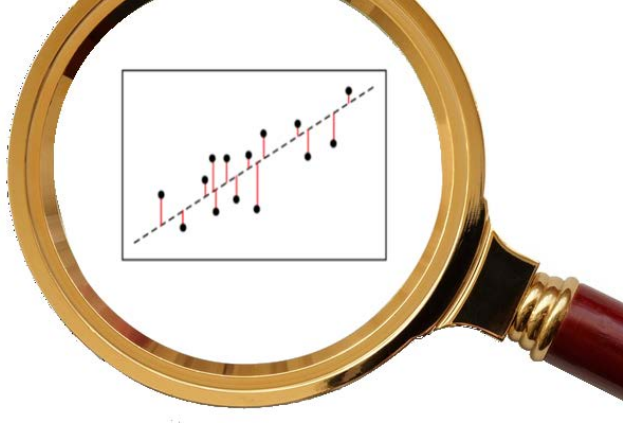
- Weather, topography, and satellite information for areas across the US
  - ~18 GB
- Examples include precipitation rate, air temperature, soil moisture content, and UV index
- Significant proportion of missing data - all variables have at least some missing values

Things to consider

- The data has spatial and temporal components that we need to consider when determining the structure of our model
  - RNNs, CNNs?

- The areas without sensors may be quite different than the areas with sensors, so need to concern ourselves with degree of extrapolation of $PM_{2.5}$ predictions

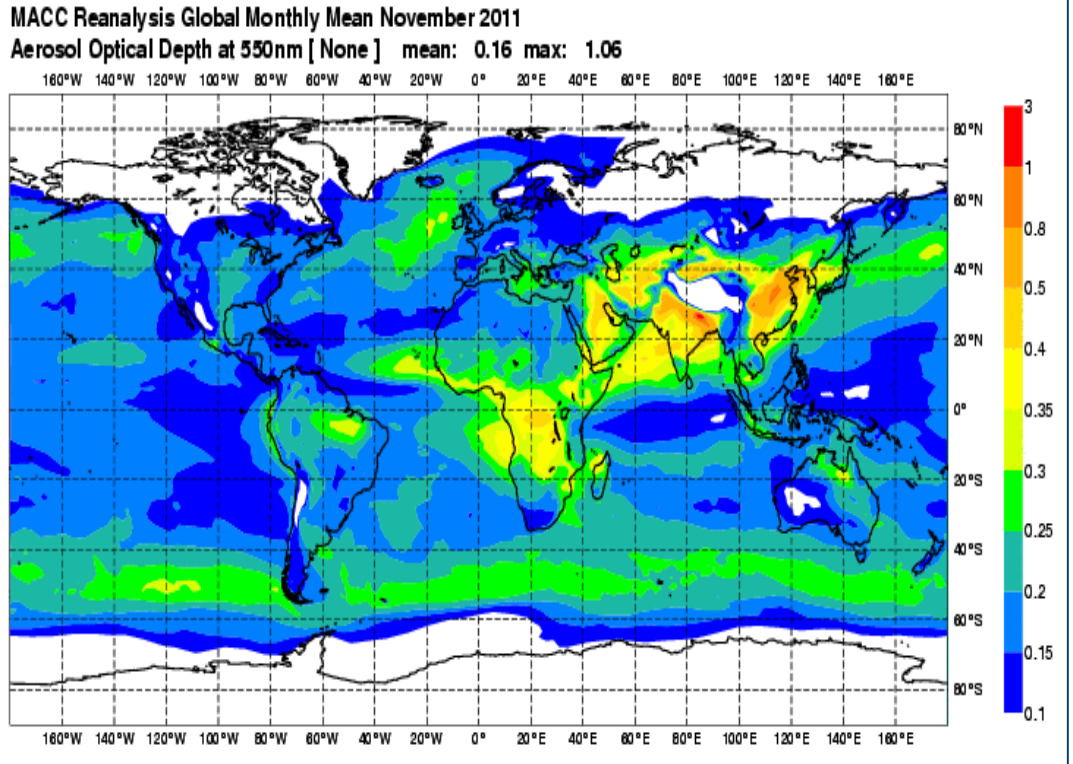
Promising Leads

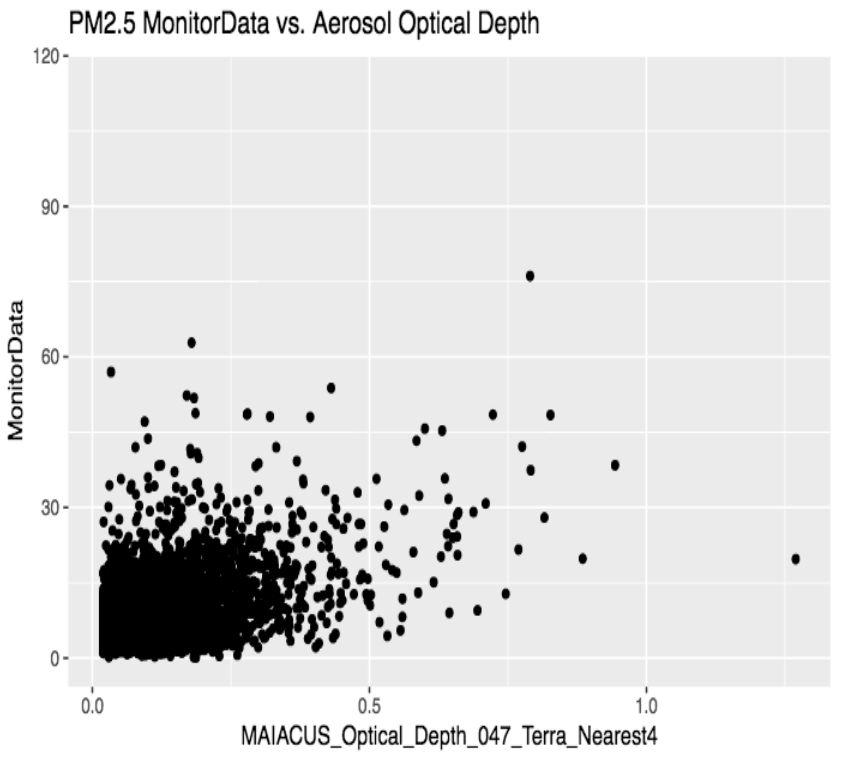


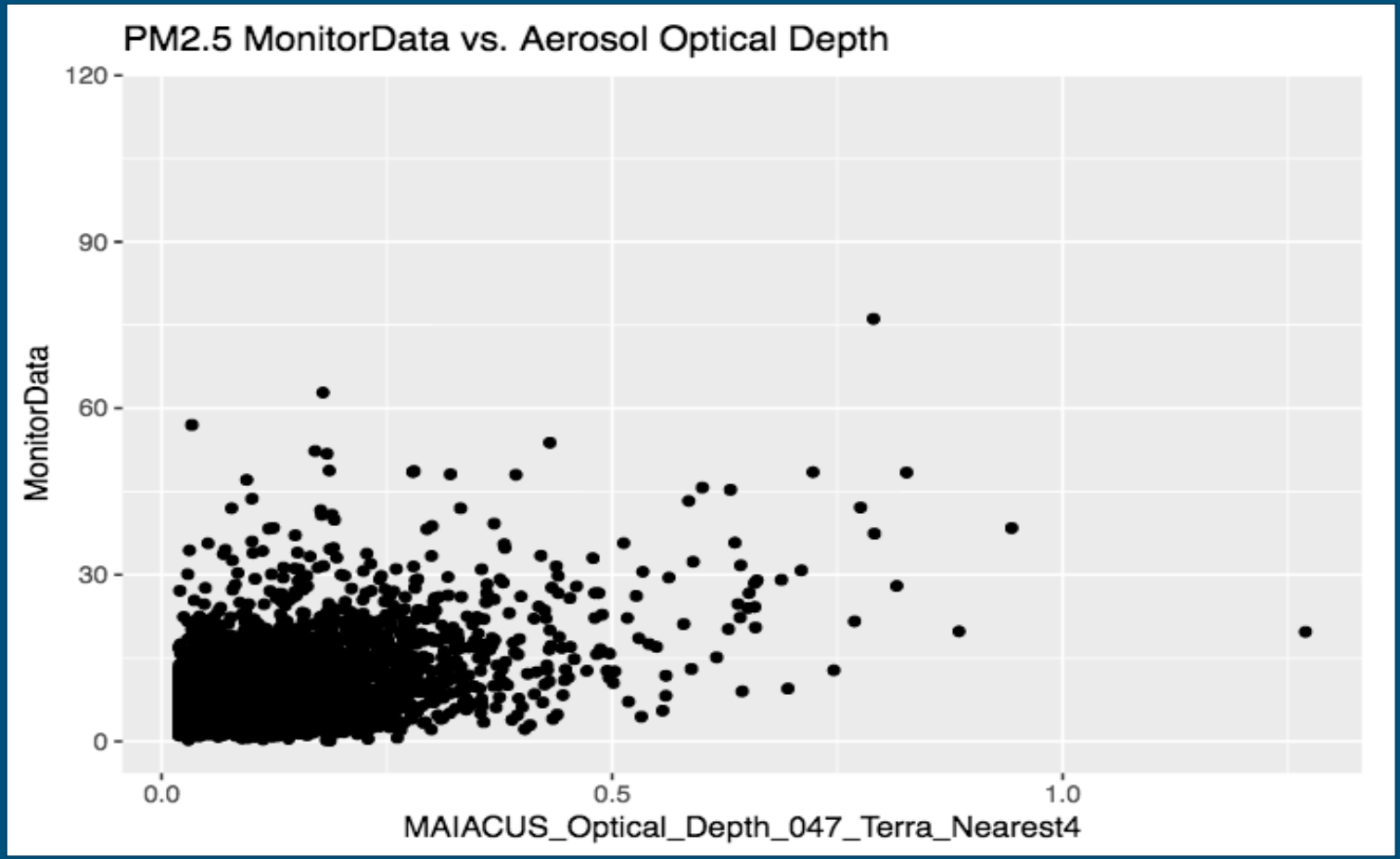
- Properties of locale of interest
  - Elevation, road density, vegetation, natural features
  - Satellite measurements of atmospheric properties

- Spatial and temporal nearby terms

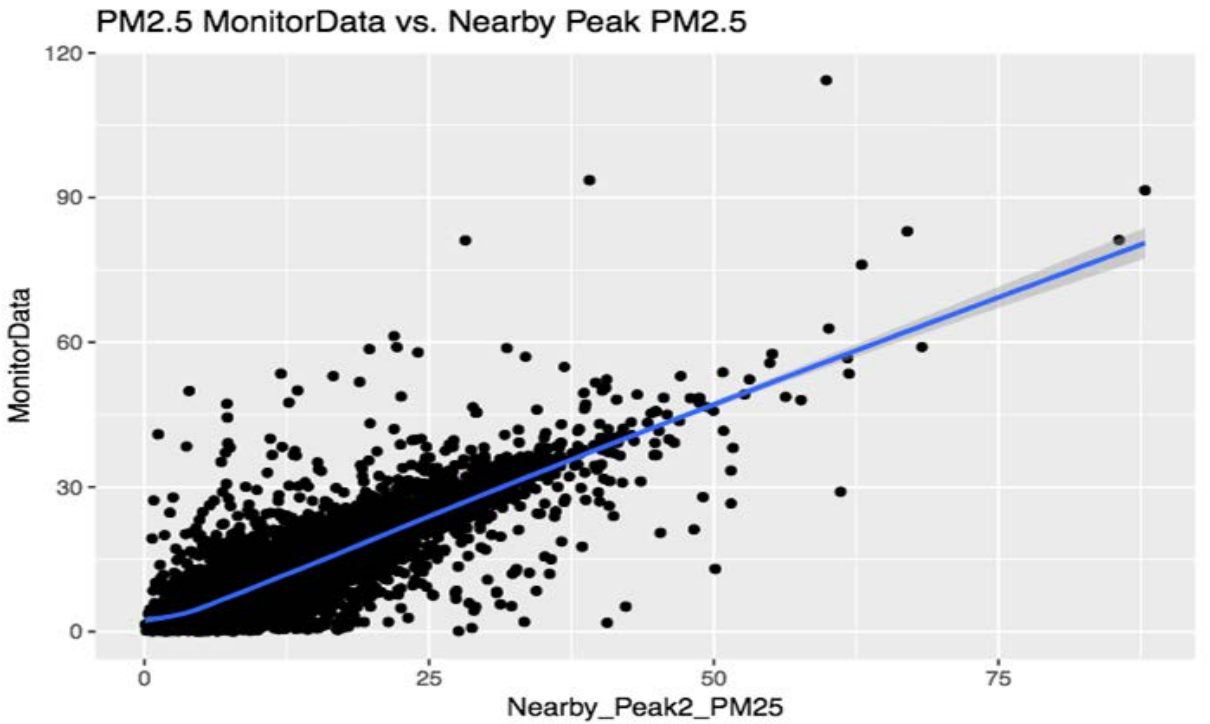
Promising Leads - Aerosol Depth



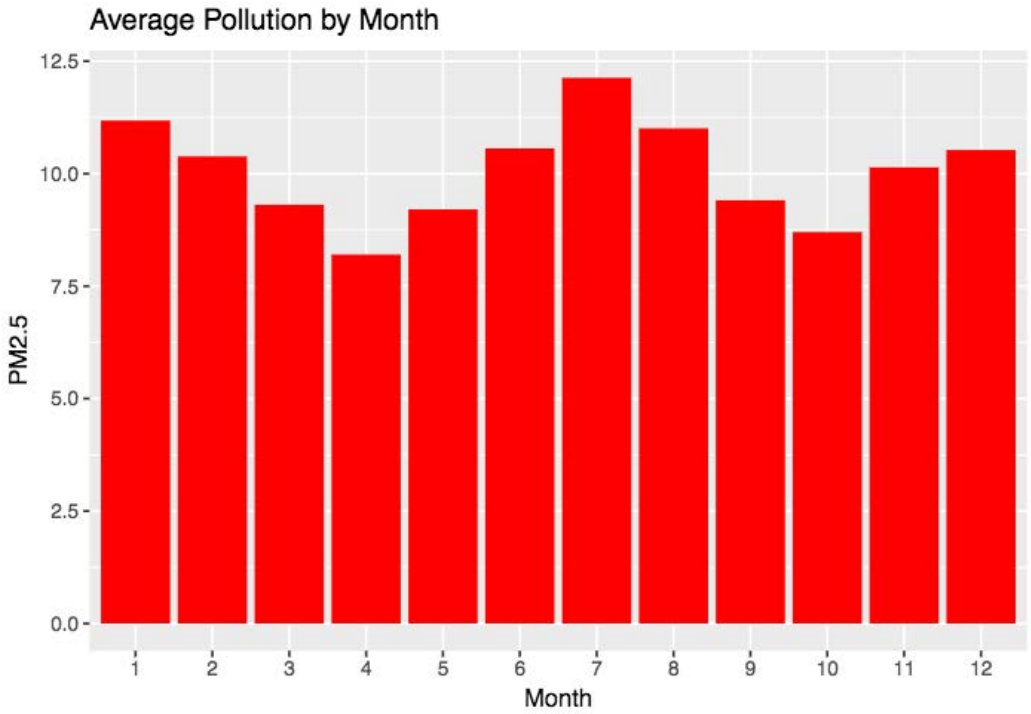



PM2.5 MonitorData vs. Aerosol Optical Depth
MonitorData
120
90
60
30
0
0.0
0.5
1.0
MAIACUS_Optical_Depth_047_Terra_Nearest4

Promising Leads - Nearby Terms

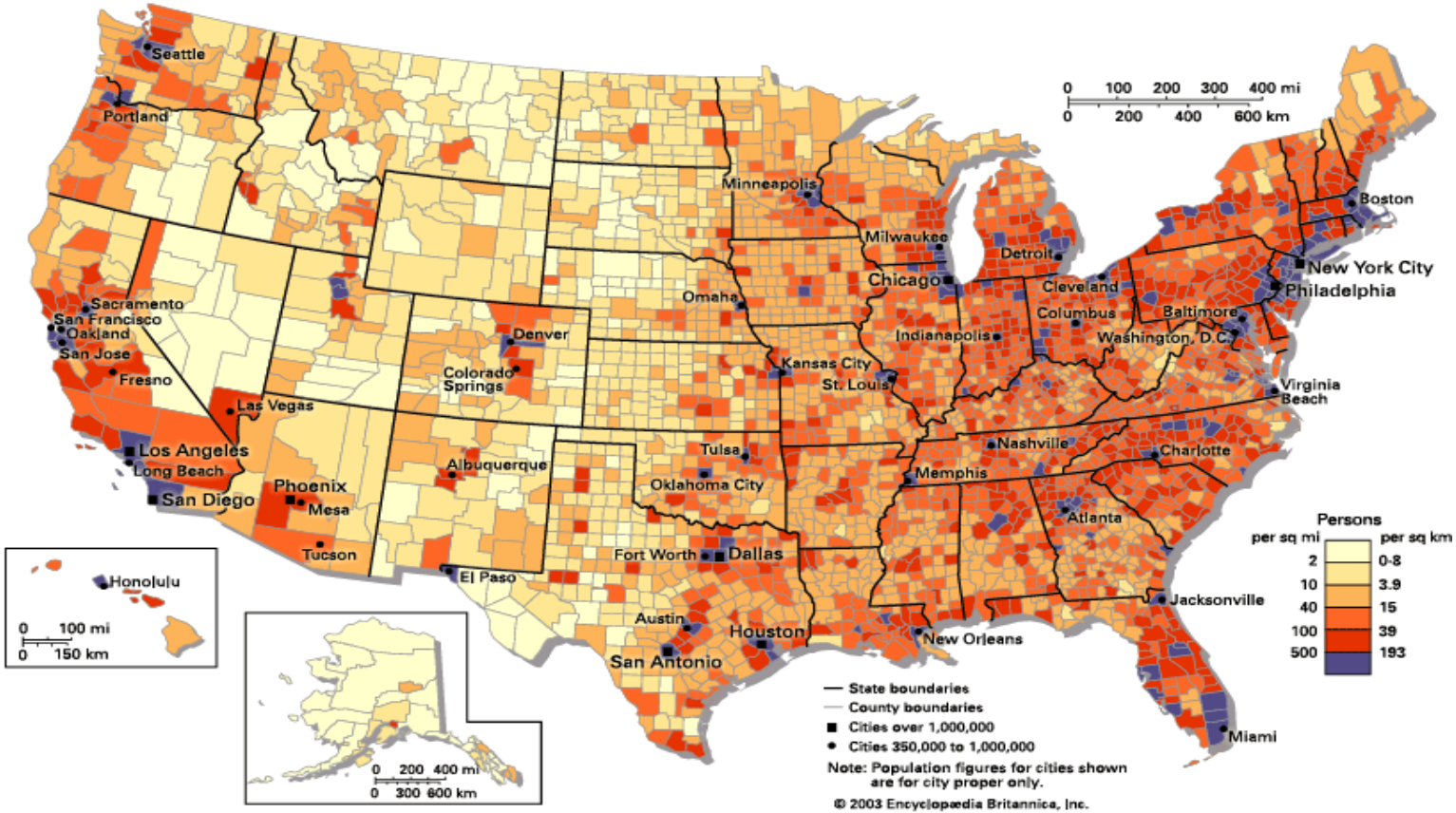


Promising Leads - Seasonal Variations

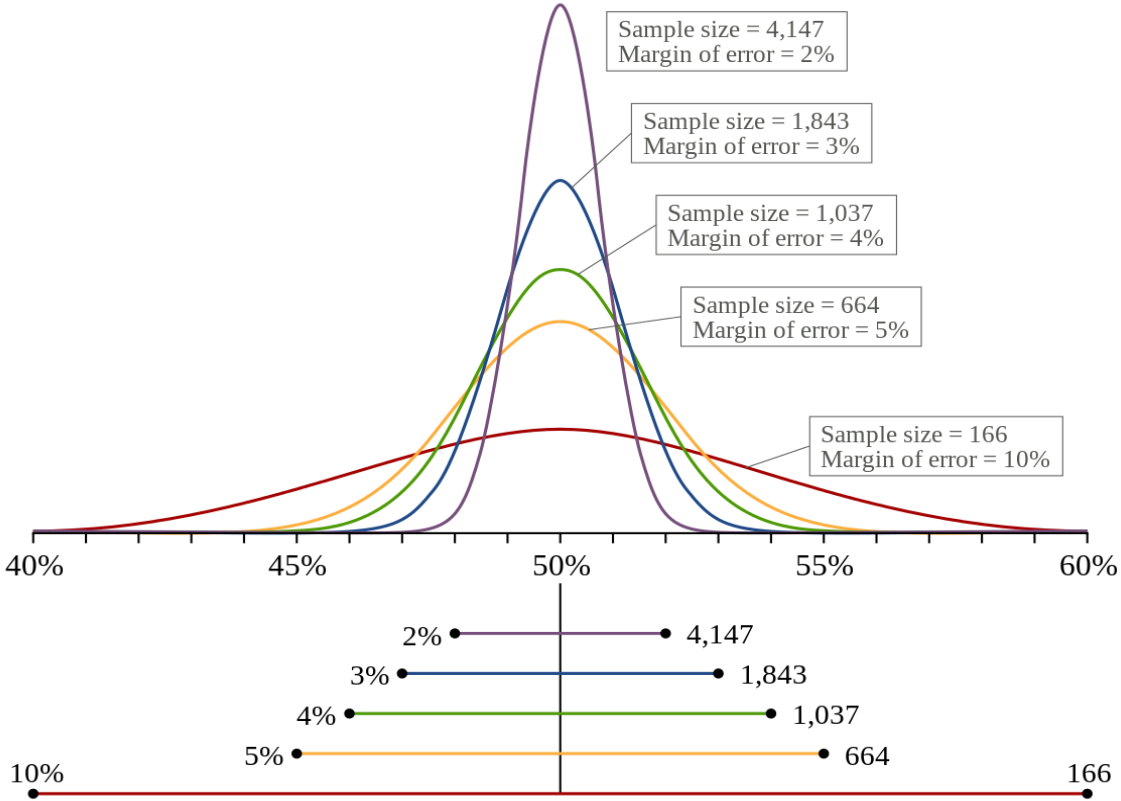


Extensions

Incorporating External Data



Prediction Margin of Error



Optimal Sensor Locations