

# *Modeling the Air Quality Index Across the Bay Area*



Berkeley  
UNIVERSITY OF CALIFORNIA

L. Drew Hill  
Environmental Health Sciences  
CE263N

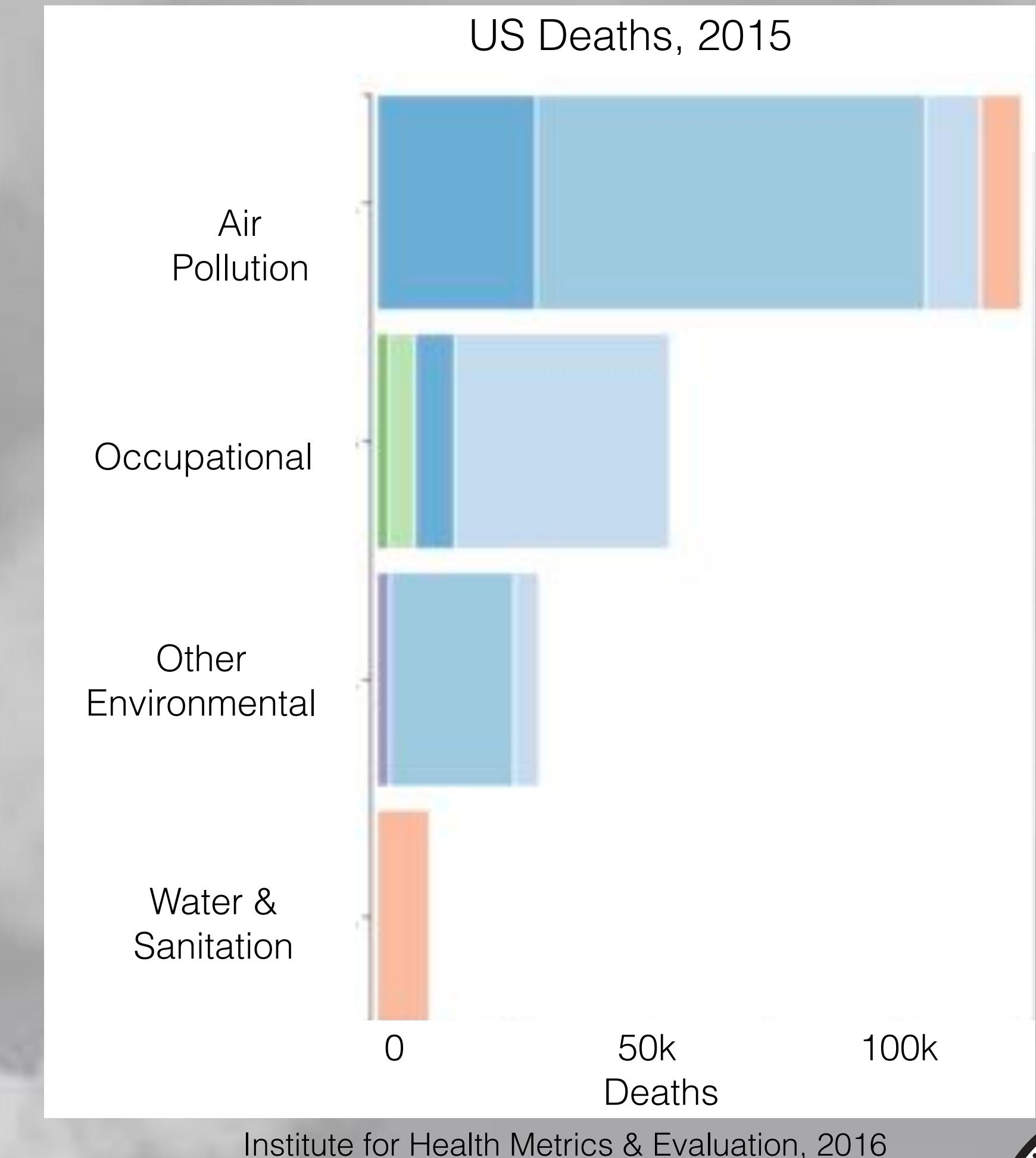
# Motivation

Air pollution is the #1 environmental cause of death

Responsible for ~ 100,000 US deaths annually

Familiar with individual pollutant-prediction models; aggregate is new

Aggregate air quality is a function of:



# Research Goals

Develop a Python class that can model an air quality map for a given hour of a day, under given weather conditions



+



=



# (US) Air Quality Index

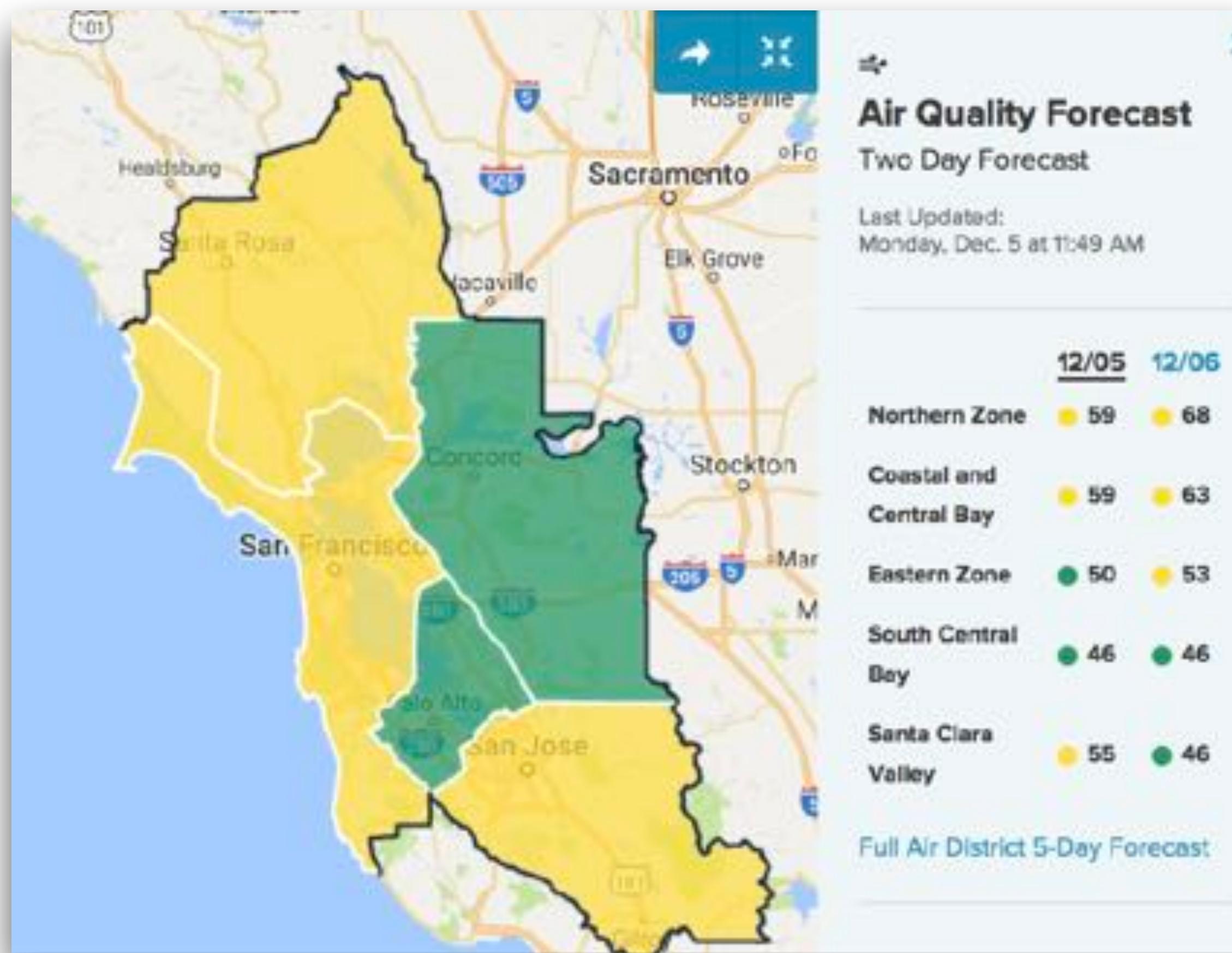
Aggregate measure of air quality

Incorporates data on 5 EPA Criteria Pollutants

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon



# Data



BAY AREA  
AIR QUALITY  
MANAGEMENT  
DISTRICT

EPA's local administrative district

17 air quality monitoring stations  
throughout the Bay Area

Calculate and report hourly  
average AQI





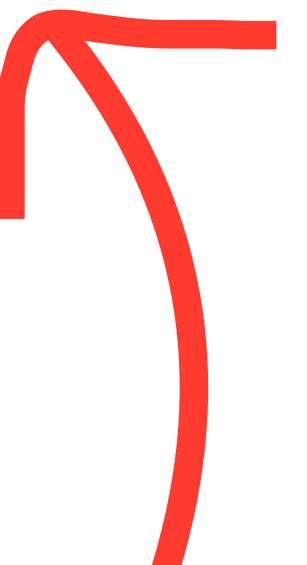
# Data

HTML tables updated every hour

Archived every day, indexed via URL that is pattern-dependent

```
'www.baaqmd.com/aqi_table/' + '11/25/2016'
```

Writing a program to scrape these tables should be straight-forward



Just an example; not the real format!



Northern Zone

Stations	Monday, December 5, 2016										
Hourly ~:	0	1	2	3	4	5	6	7	8	9	10
Napa	116	102	94	87	79	70	64	64	62	56	
Napa Valley College *	69	60	58	63	54	53	48	49	48	38	44
San Rafael	8	8	8	8	12	12	10	13	15	18	18
Sebastopol	40	32	27	22	19	22	23	32	33	27	18
Vallejo	20	18	18	23	20	12	6	7	16	60	74

Coast & Central Bay

Stations	Monday, December 5, 2016										
Hourly ~:	0	1	2	3	4	5	6	7	8	9	10
Berkeley Aquatic Park	55	48	43	38	36	32	31	30	21	4	8
Lancy College Fwy	54	81	60	49	39	34	30	25	25	19	16
Oakland East	44	37	29	25	21	23	28	27	26	21	31
Oakland West	56	51	55	55	57	55	53	44	41	31	42
San Francisco - Arkansas St.	37	52	56	52	38	25	35	52	37	41	42
San Pablo - Rumrill	20	23	24	31	32	30	32	28	35	32	55



# Data

Hourly Wx data available from NOAA  
as ASCII files

Available for several Bay Area locations

I will use OAK International Airport  
as a general Bay Wx estimate



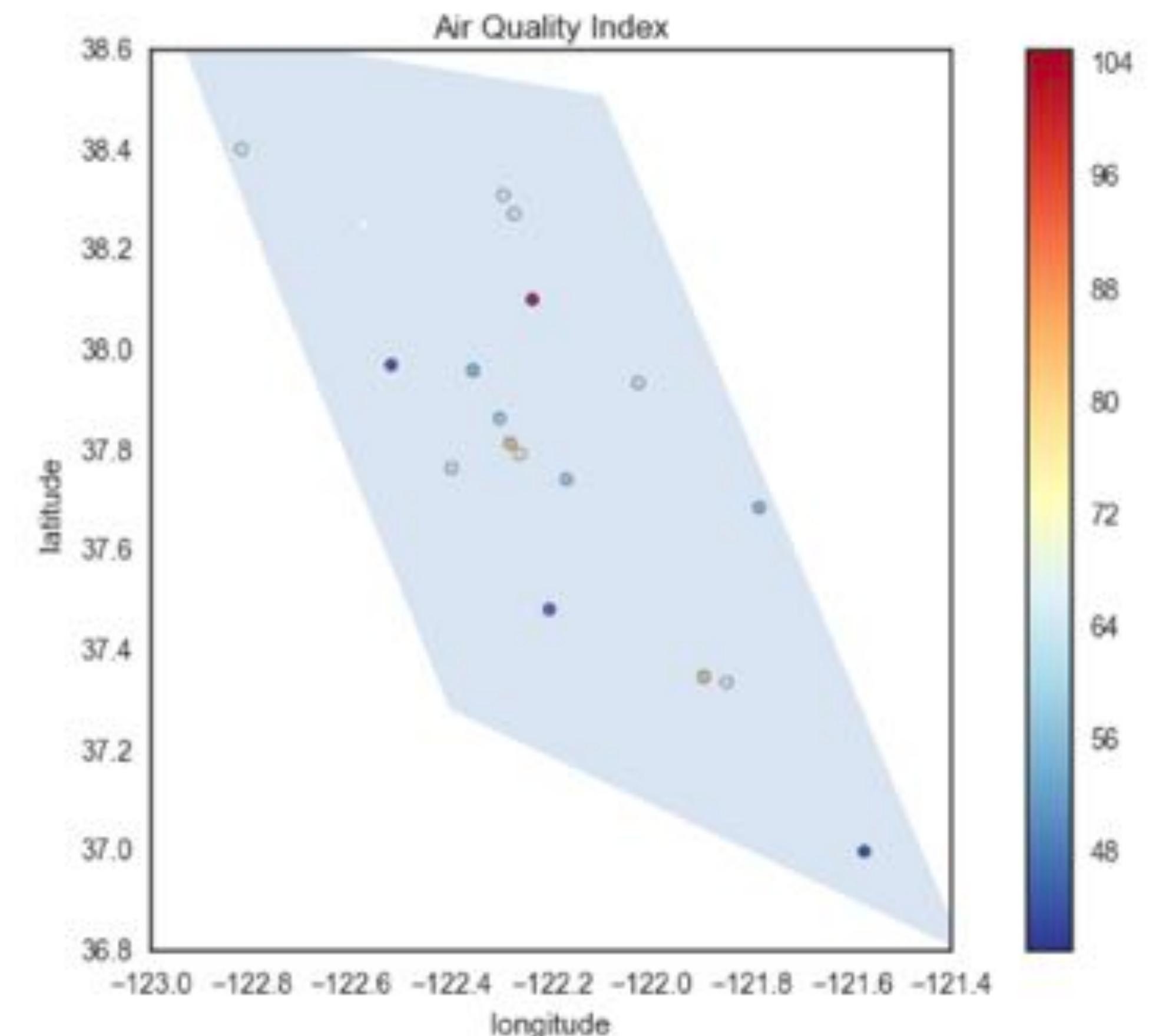
# Prediction

Plot each monitor

Use Shapely to produce buffers around each monitor

Aggregate buffers into a reasonably sized polygon

Produce a grid of points for prediction, cropped to fit this polygon

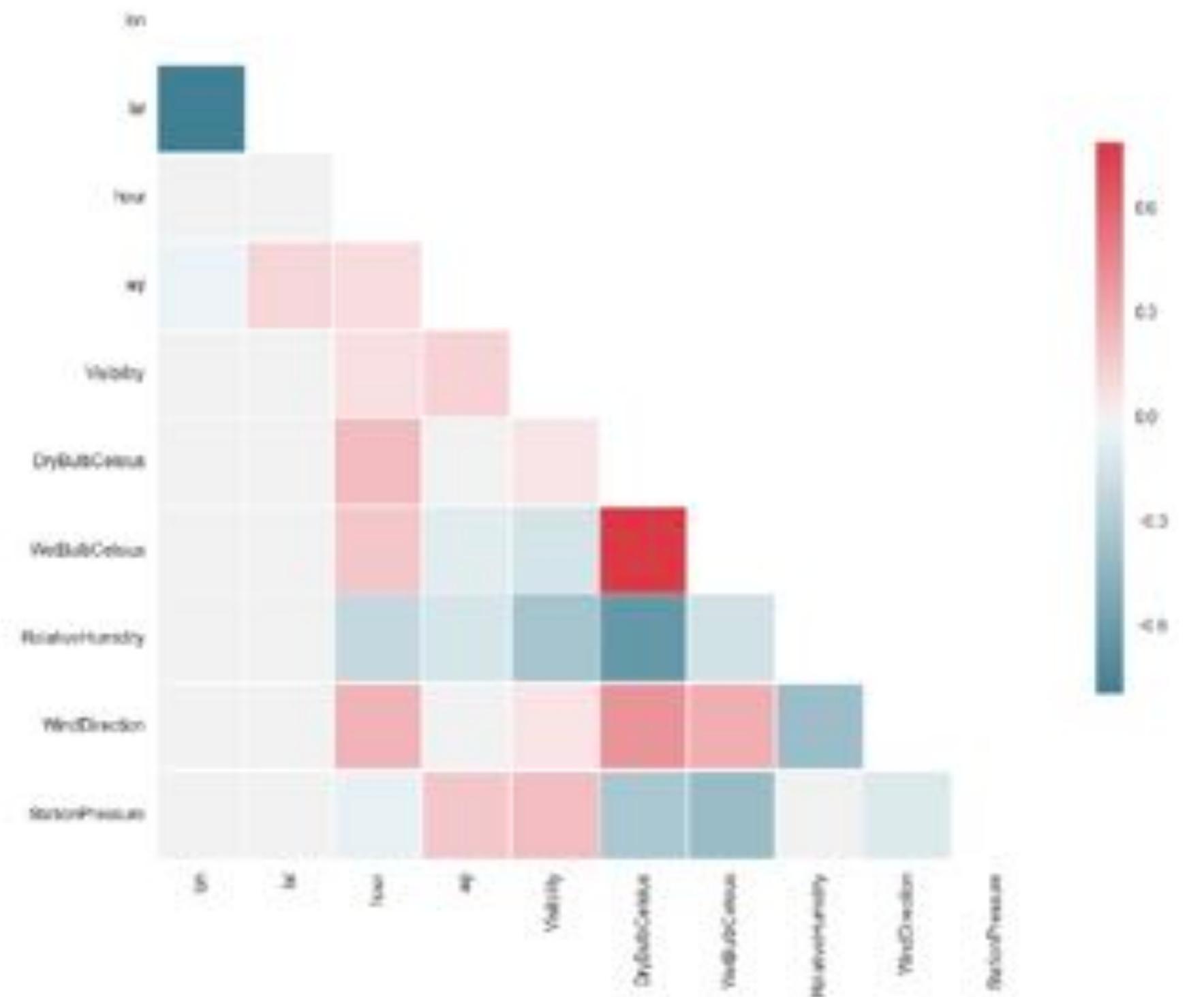


# Prediction

Graphically examine data

Choose a subset of variables with high bivariate correlation with AQI (reduce computational time)

\*\*Employ k-fold cross-validated Gaussian modeling\*\*



# Thanks!

