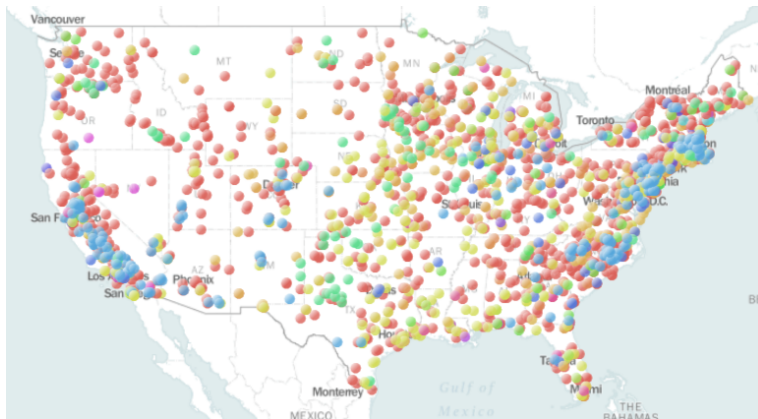


US Power Plant Analysis Pipeline:

Where in the US do I install a plant?



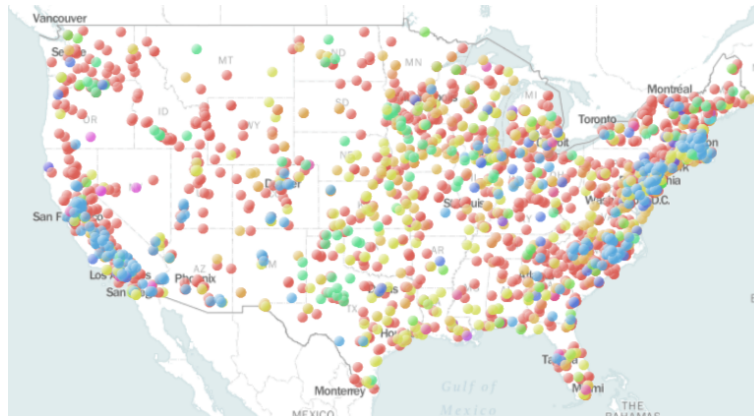
Deciding Factors

**Fuel/Cost
Availability**

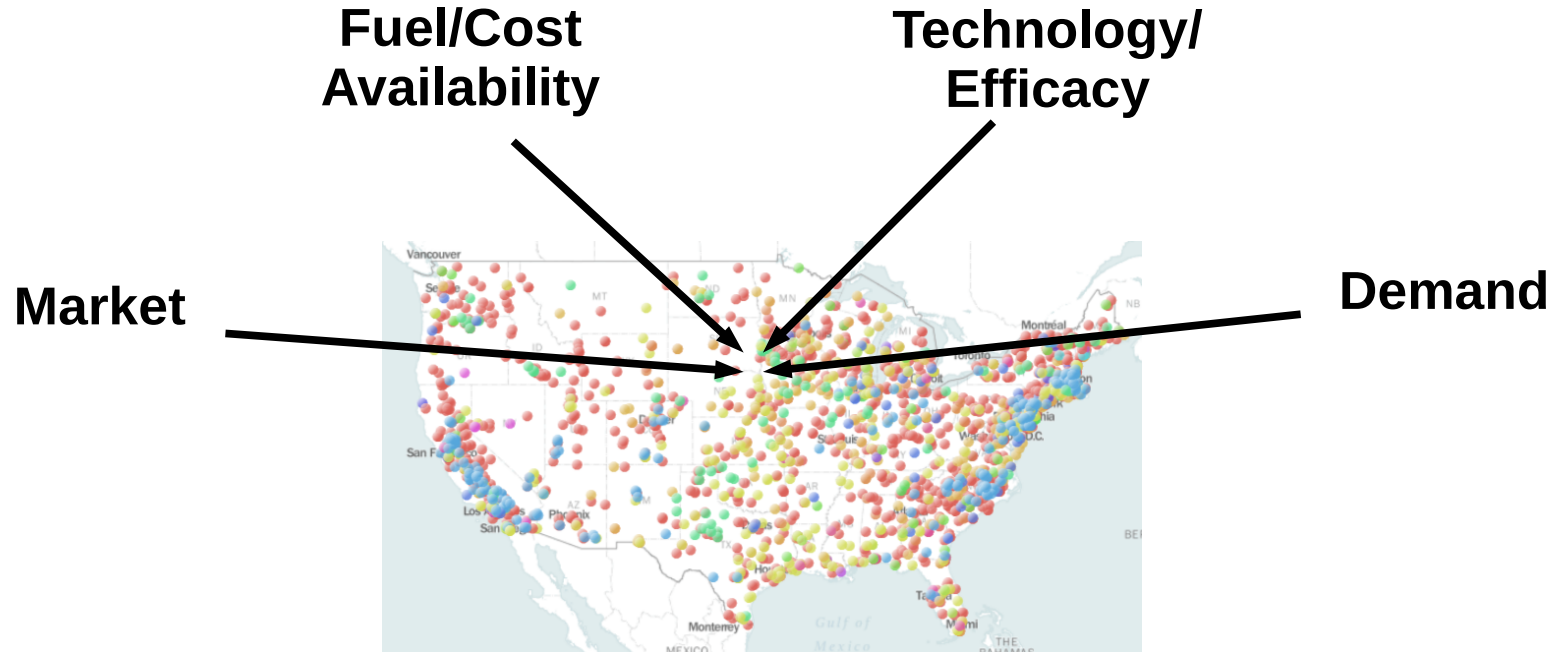
**Technology/
Efficacy**

Market

Demand



Deciding Factors



Data Set

	EIA-923 Generation and Costs	EIA-860 Plant Information	County Census Data
2001	Plant ID	Plant ID	County
.	State	Lat/Long (some entries)	State
.	Fuel Type/Consumption	Sector	Population
.	Net Generation	County	
2016	Tech Type		

Data Set

	EIA-923 Generation and Costs	EIA-860 Plant Information	County Census Data
2001	Plant ID	Plant ID	County
.	State	Lat/Long (some entries)	State
.	Fuel Type/Consumption	Sector	Population
.	Net Generation	County	
2016	Tech Type		

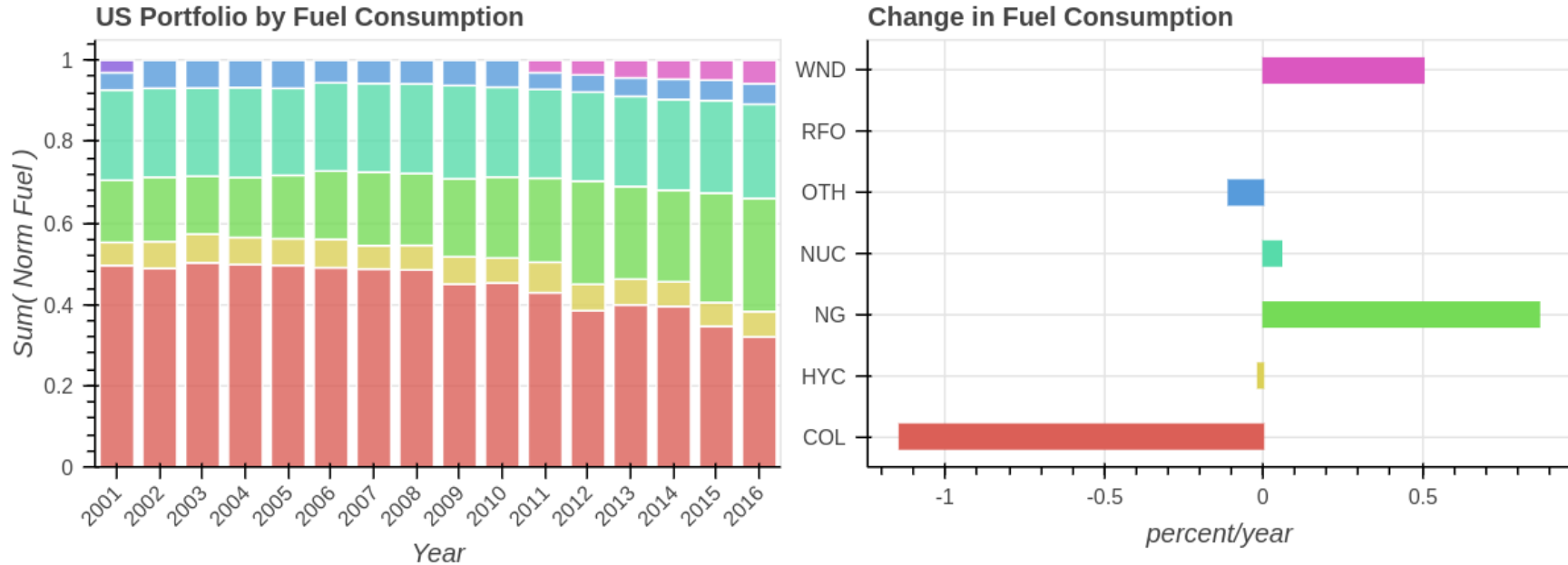
Data Set

	EIA-923 Generation and Costs	EIA-860 Plant Information	County Census Data
2001	Plant ID	Plant ID	County
.	State	Lat/Long (some entries)	State
.	Fuel Type/Consumption	Sector	Population
.	Net Generation	County	
2016	Tech Type		

Data Cleaning/Validation, Find Missing Values, Label Matching

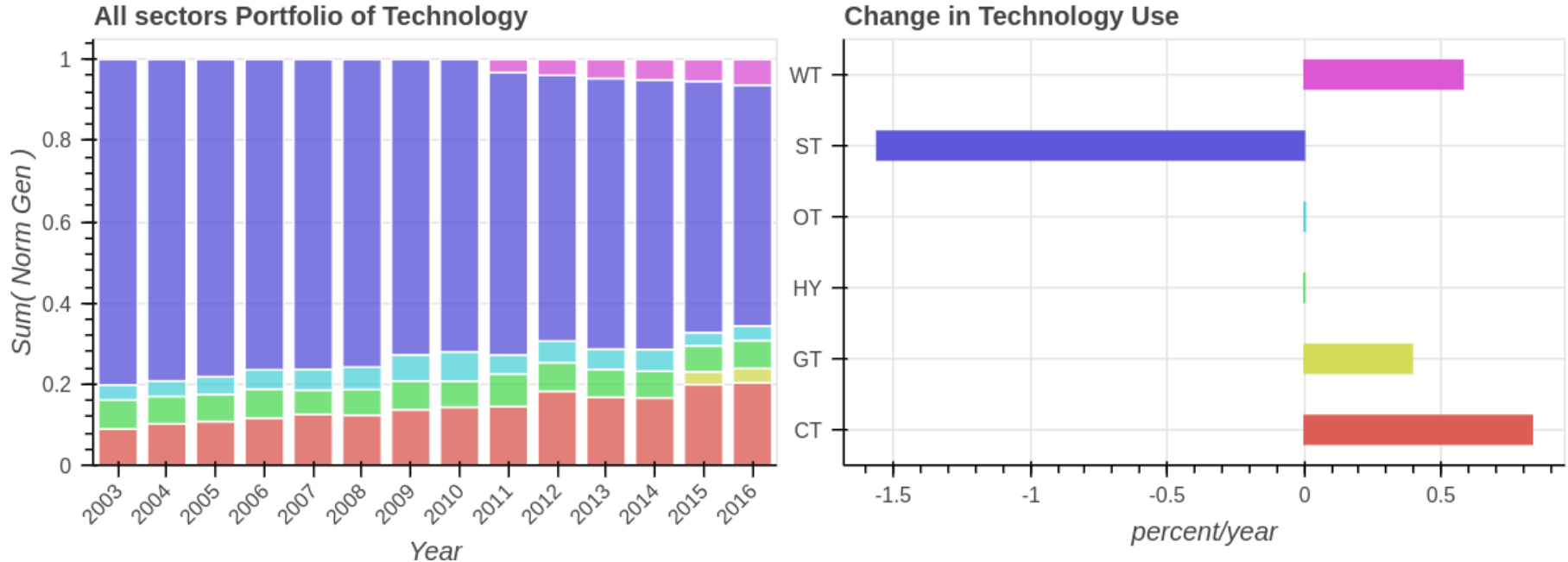
Compiled Data
11,000 plants, 16 years
180,000 samples, 13 Features

Fuel Trends



Natural Gas and Wind replacing Coal

Technology Trends



Combined Cycle and Wind Turbine Replacing Steam Technology

Predictive Analytics

Potential Indicators

Geography

**Population/Power
Density**

**Fuel Used, Sector,
Tech Trends**

Data Set Features

**Lat/Long, County,
State**

**Population,
Neighboring Gen.**

**Fuel consumption,
Efficiency, Sector**

Prediction of Interest

**Fuel Type Used
18 classes**

**Capacity
Continuous**

**Technology
13 classes**

Back Up Slides

Data Pipeline

**EIA-923
Generation and Costs**

**EIA 860
Plant Information**

County Census Data

Predictive Features

Prediction of Interest

Data Cleaning/Preparation

U.S. Department of Energy, The Energy Information Administration (EIA)
EIA-923 Monthly Generation and Fuel Consumption Time Series File, 2012 File
Sources: EIA-923 and EIA-860 Reports

Plant Id	Combined Heat And Power Plant	Nuclear Unit Id	Plant Name
2N			Bankhead Dam

Data validation

Net Generation (Megawatthours)	
	-147,503
	-206,679
	-773,280

Feature Engineering

- $\text{Gen/Fuel} = \text{Efficiency}$
- $f(\text{Lat}, \text{Long}, \text{Gen}) = \text{power density}$
- $f(\text{Lat}, \text{Long}, \text{Pop}) = \text{Pop density}$

Feature Matching

Tot_MMBtu January	TOT_MMBTU _JAN
----------------------	-------------------

'PLNTCODE,N,5,0','PLNTCODE',
'PLANT_CODE','Plant Code'

Geography

**Population
Density**

**Fuel Used, Sector,
Tech Trends**

**Fuel Type Used
18 classes**

**Capacity
Continuous**

**Technology
13 classes**

Proposed Models

- **Fuel Type**
 - Unsupervised cluster of Lat/Long to give regions
 - Naive Bayes within each cluster to give
- **Capacity**
 - GLM with population and power density
- **Technology**
 - Random Forest with Fuel Type, Capacity, Sector