# Predicting Pothole Repair Times for the City of Seattle

Richard Anderson


# Data Pipeline

**Obtain Data**

Socrata
Making Data Social™

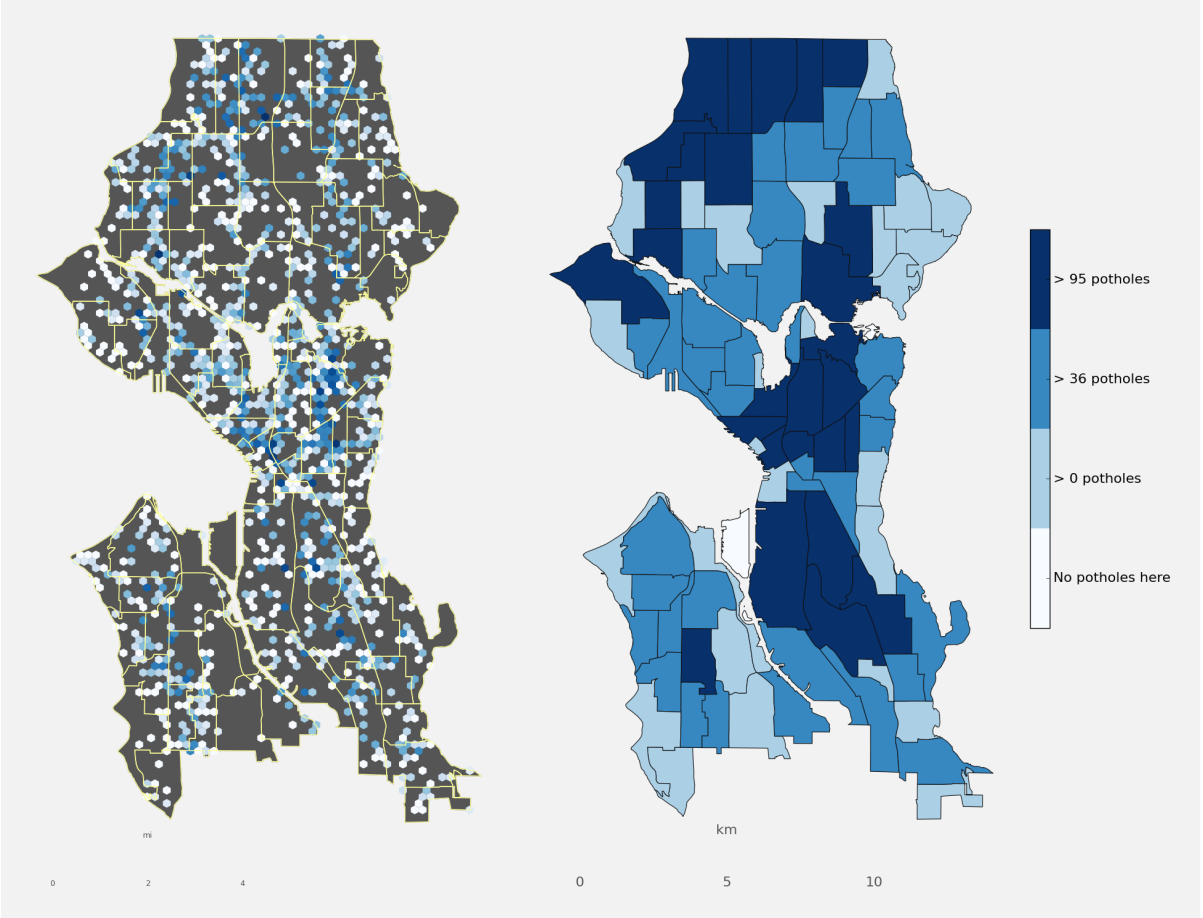

seattle.gov
Department of Transportation



NOAA NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
UNITED STATES DEPARTMENT OF COMMERCE

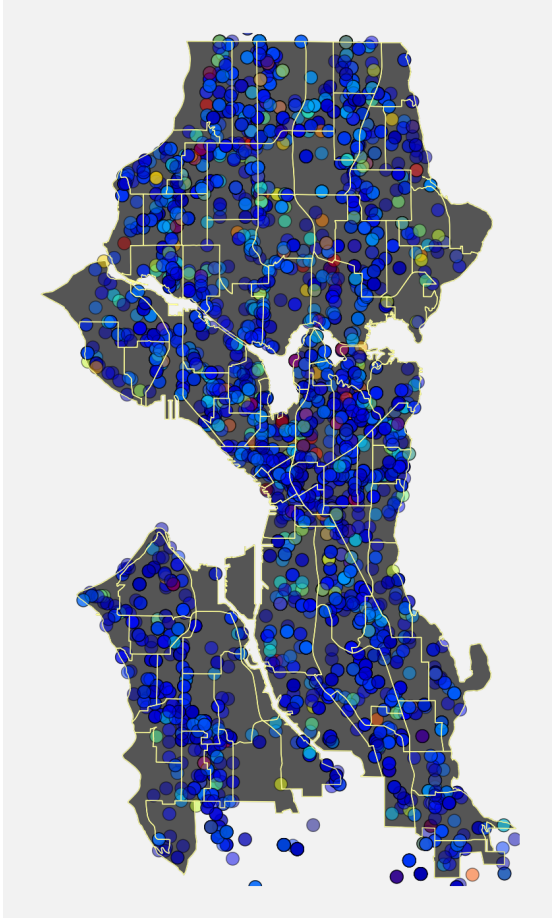
# Data Pipeline

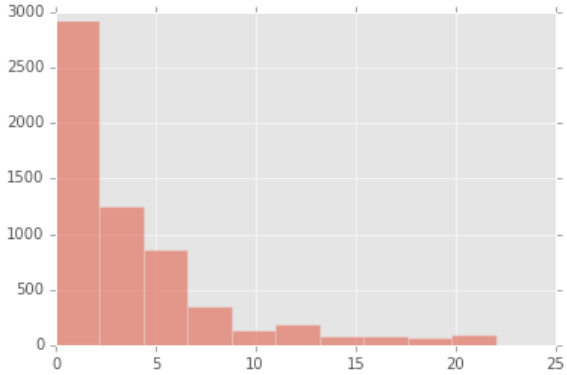
**Obtain Data** → **Geocode Locations**

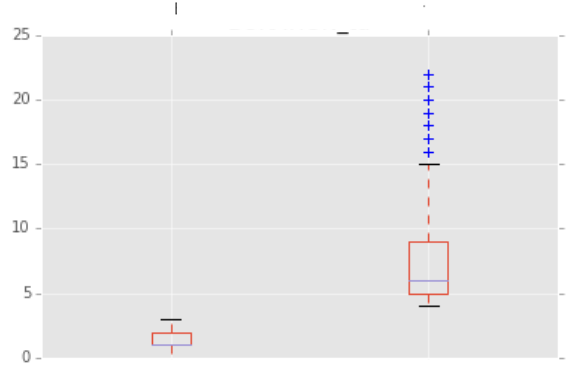


# Data Pipeline

**Obtain Data** → **Geocode Locations**

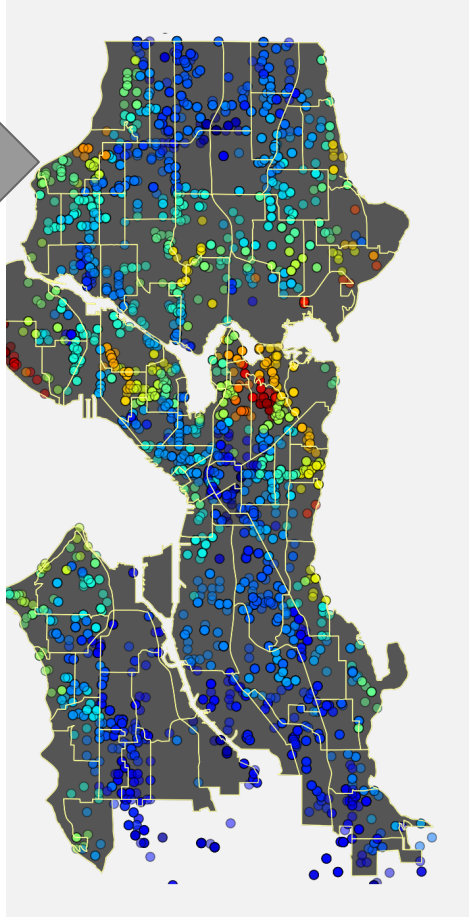




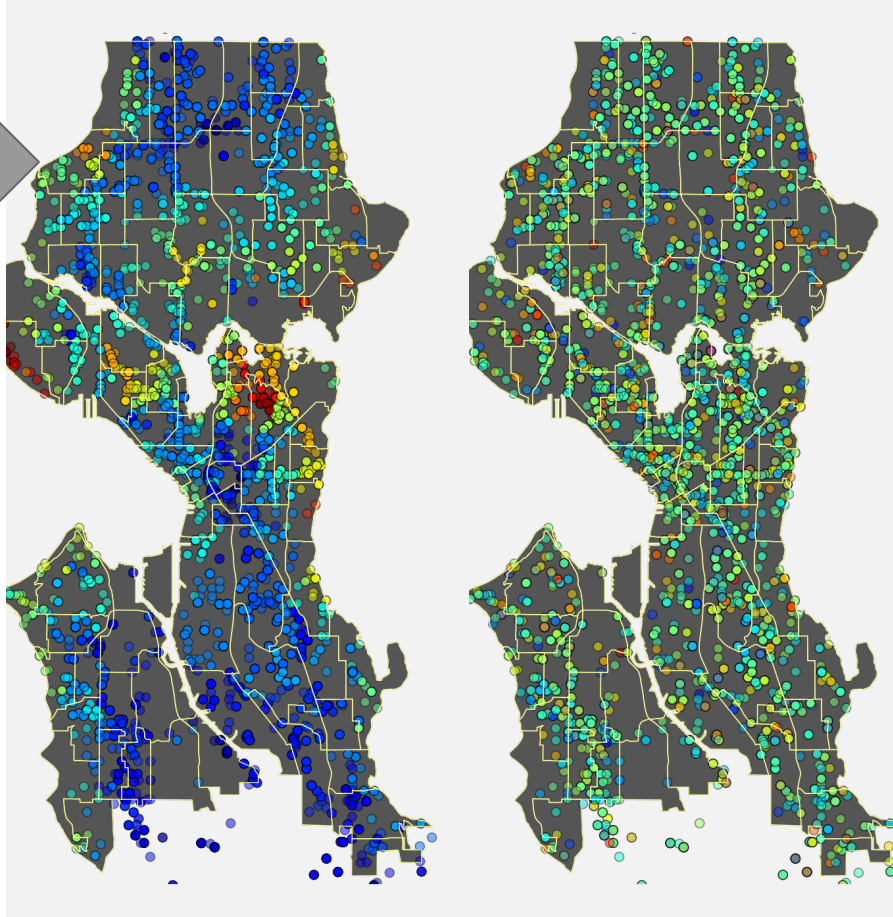
# Data Pipeline

**Obtain Data** → **Geocode Locations** → **Engineer Features**



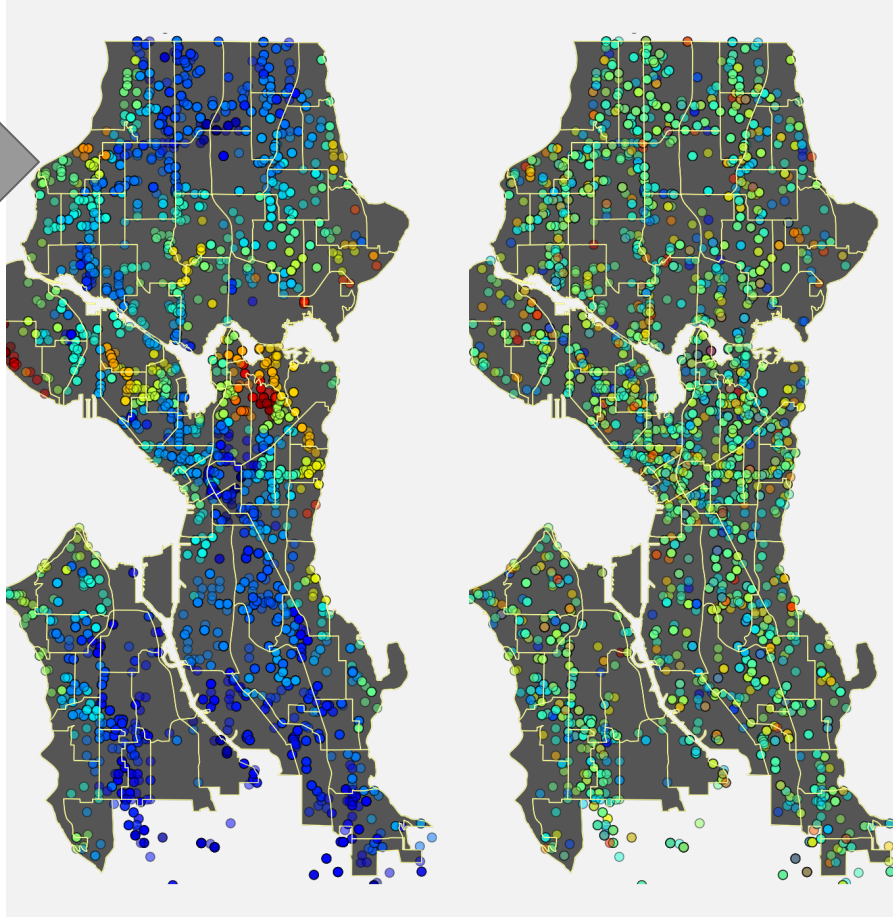
# Data Pipeline

Obtain Data → Geocode Locations → Engineer Features



# Data Pipeline

**Obtain Data** → **Geocode Locations** → **Engineer Features**



## Additional Features

- Cumulative Number of unrepaired potholes
- Minimum Distance to closest prominent city landmark
- Number of months until end of fiscal year
- Seattle Neighborhood
- Closest Street Features

# Data Pipeline


**Exploratory Data Analysis**

Scatterplots

Numerical Features

Categorical Features

**Models and Algorithms**

Logistic Regression

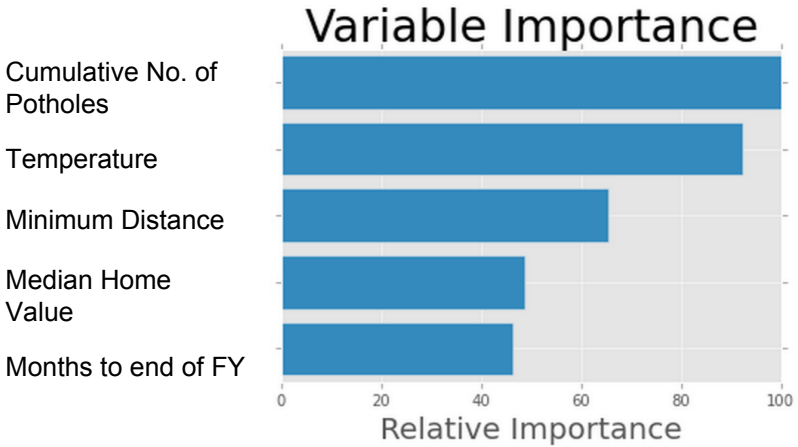
Random Forest Classifier

130 features

Parameter search

# Data Pipeline

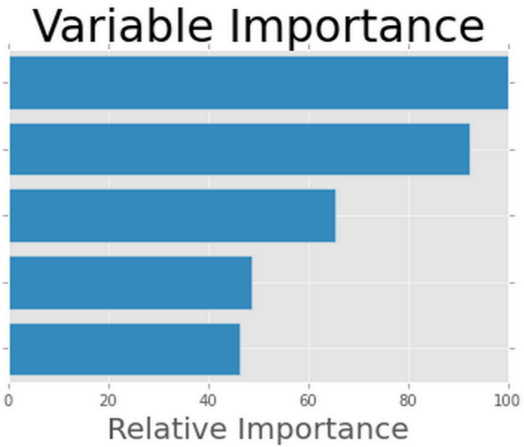
**Obtain Data** → **Geocode Locations** → **Engineer Features** → **Implement ML Models** → **Draw Conclusions**



# Data Pipeline

**Obtain Data** → **Geocode Locations** → **Engineer Features** → **Implement ML Models** → **Draw Conclusions**

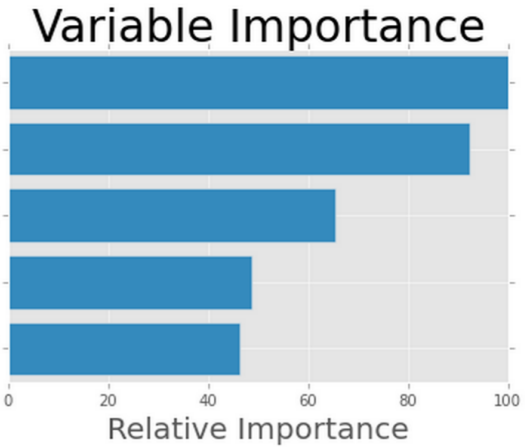
| | |
|---|---|
| +++ | Cumulative No. of Potholes |
| | Temperature |
| -- | Minimum Distance |
| + | Median Home Value |
| -- | Months to end of FY |



# Data Pipeline

**Obtain Data** → **Geocode Locations** → **Engineer Features** → **Implement ML Models** → **Draw Conclusions**

| | |
|---|---|
| +++ | Cumulative No. of Potholes |
| | Temperature |
| -- | Minimum Distance |
| + | Median Home Value |
| -- | Months to end of FY |




**Future Work:**

- Challenge is to find variables that are both independent and add the most information to the model
- Careful investigation of whether median home value is really driving repair times.