Residential Properties and Presidential Election Results: King County, WA

Elliott Day

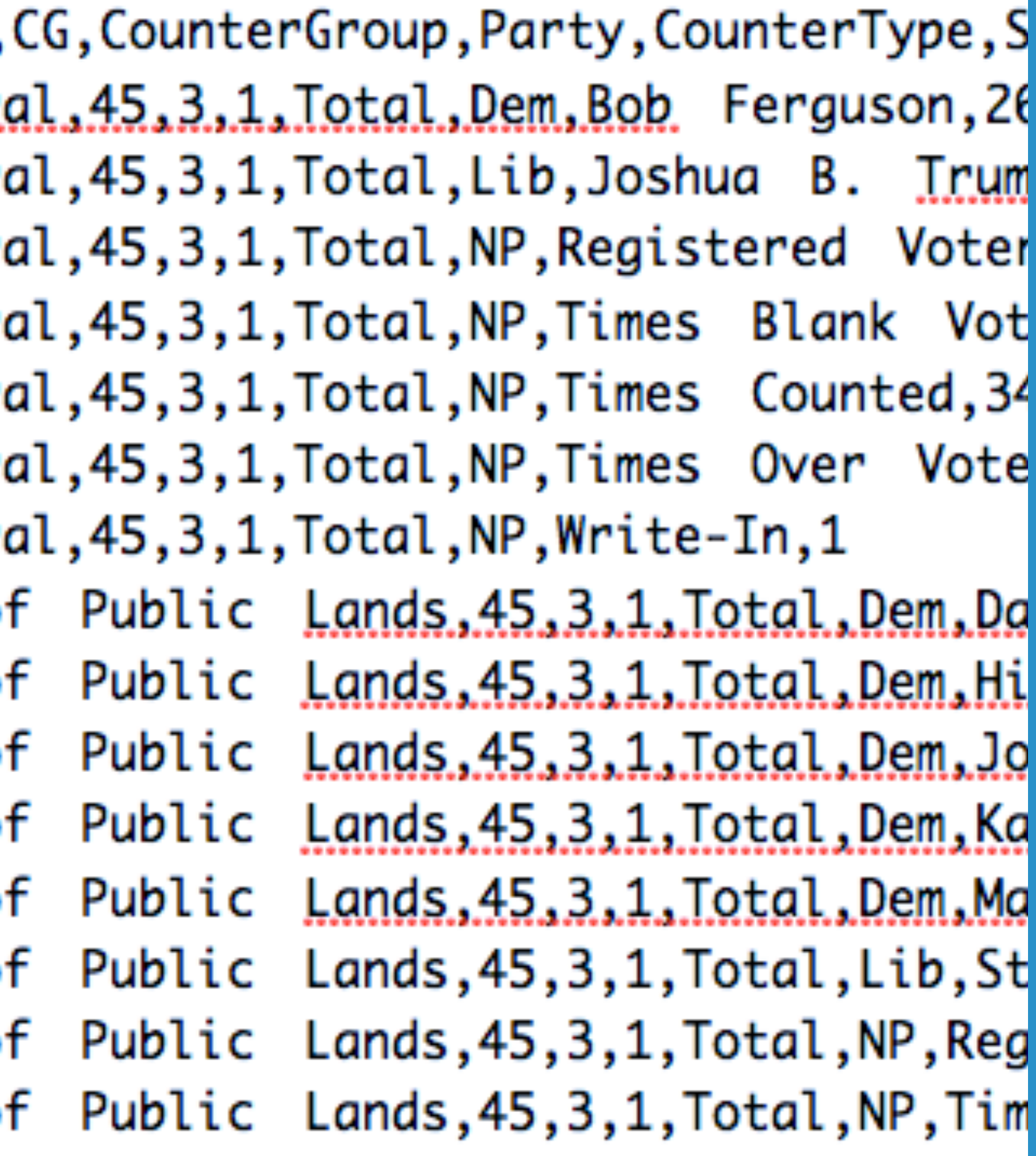
June 30, 2019

Objective

Determine the capacity of a county's residential property data to predict the outcome of a specific election. (Test a hunch)

My background with this topic

- Degree in geography, GIS user
- Background in local politics and government
- Nonprofit research role focused on elections
- Mapped 15 years of worth of the county's elections at the Atlas of King County Politics



Election Data

2016 Presidential election results for King County, Washington.

The County's Department of Elections counts the vote and makes a report of the results available online.

Precincts

Results are tabulated by voting precincts, polygons drawn to contain no more than 1,000 voters.

King County, Washington *(population 2.1 million, county seat: Seattle)* has between 2,500 and 2,800 precincts depending on the year. In this dataset, we'll have 2,532 target values.

Precinct size is inversely proportional to population density.

Why Housing Data?

Elections are typically predicted, analyzed, and explained in terms of the datapoints available through polling and voter registration.


GROCERIES
VEGE
OPEN


WA-906

Why Housing Data?

Hypothesis: enough information related to voting preference (in this one case at least) is contained in data on how and where people live that we can predict a precinct's vote based on its residential housing characteristics.

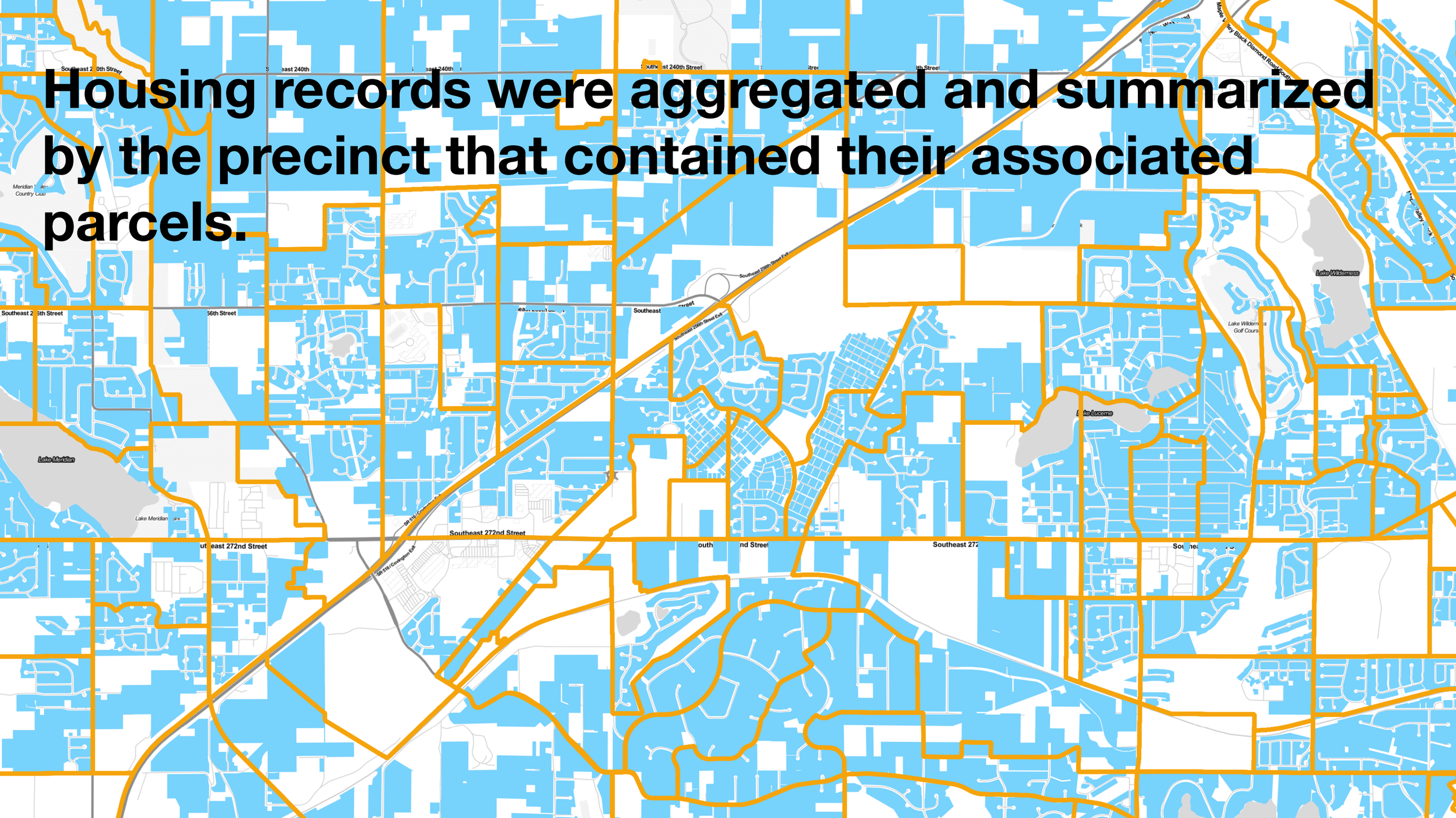
What's in it?

The County Assessor collects a lot of detailed and thorough information on properties, from road and sewer types to views and bathroom counts. These features are generally available for every parcel in the county.

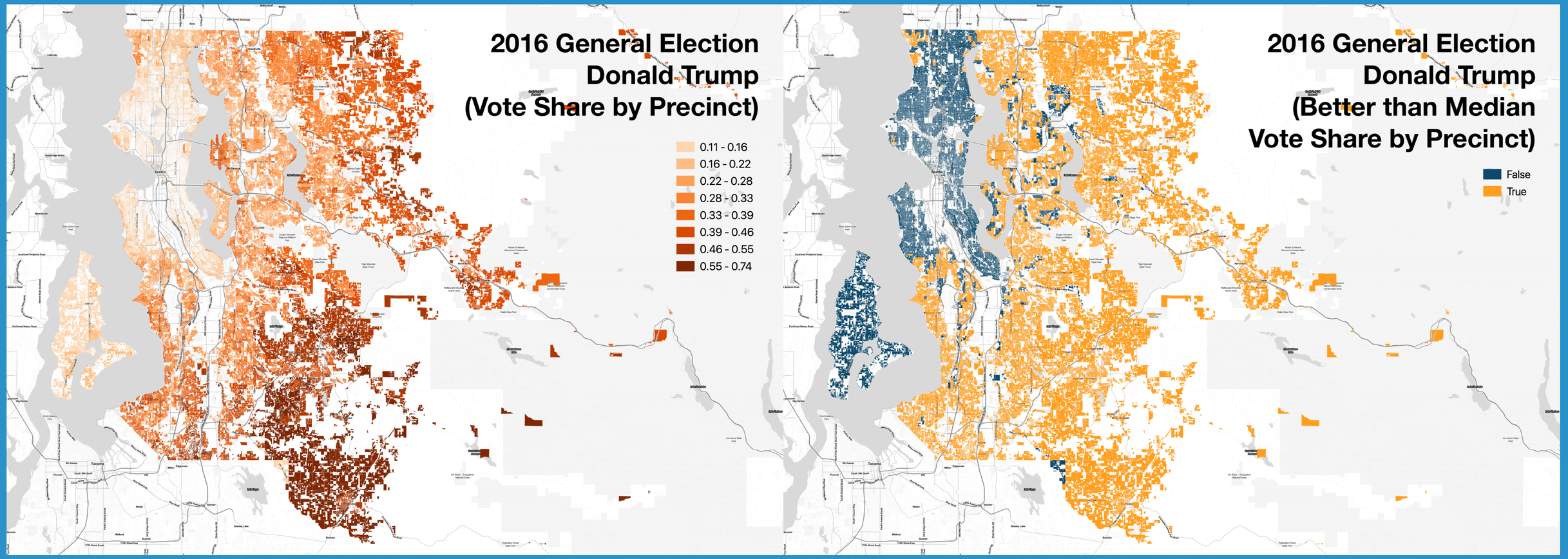
Example Features

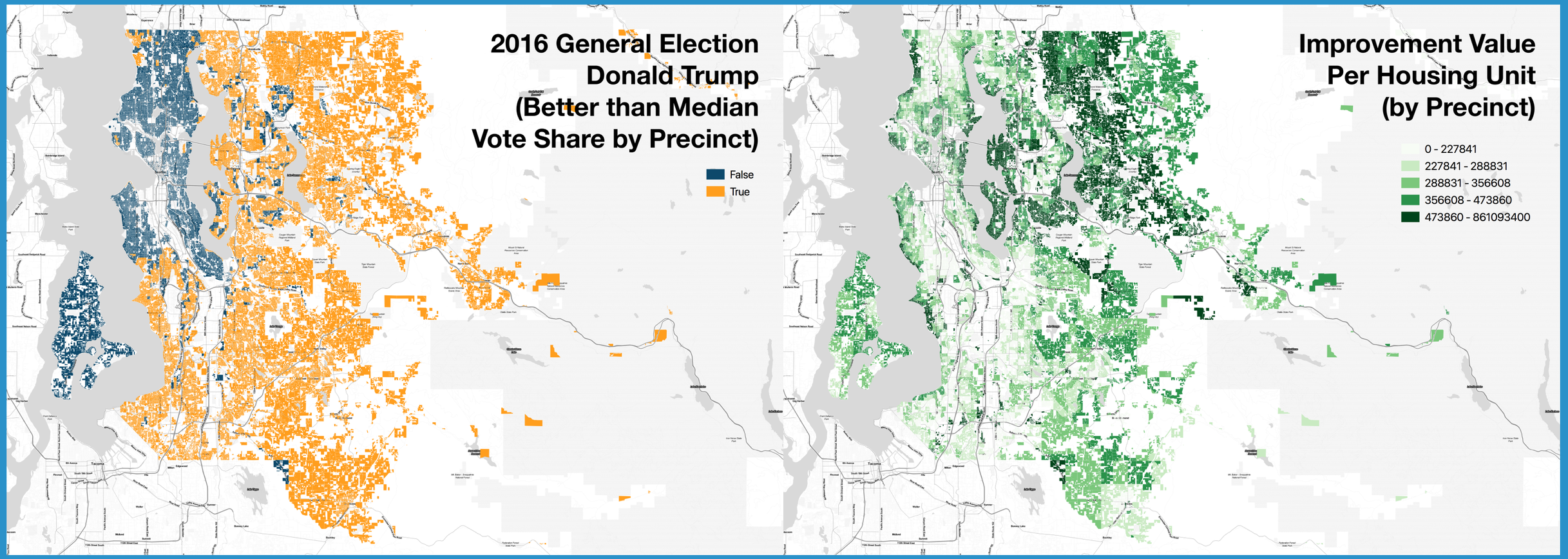
I combined several tables to create the features in the example dataset.

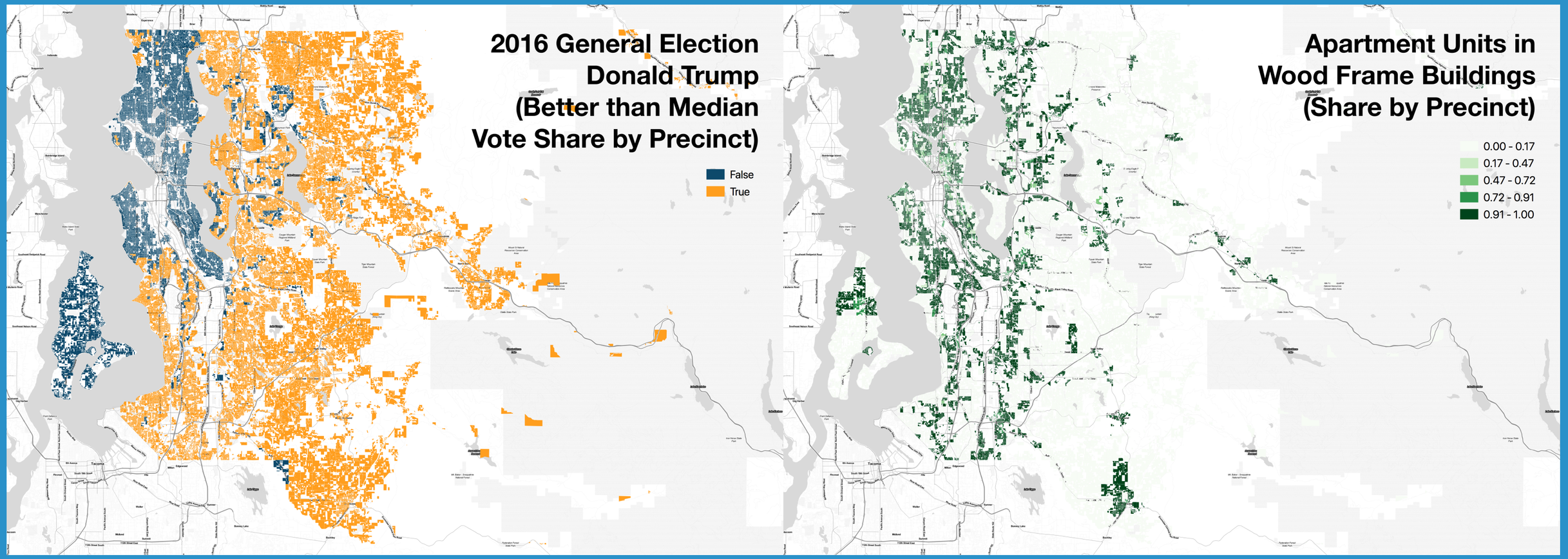
- Parcels
- Residential Buildings (up to 3 units)
- Apartment Buildings
- Condominium Units
- Condominium Complexes

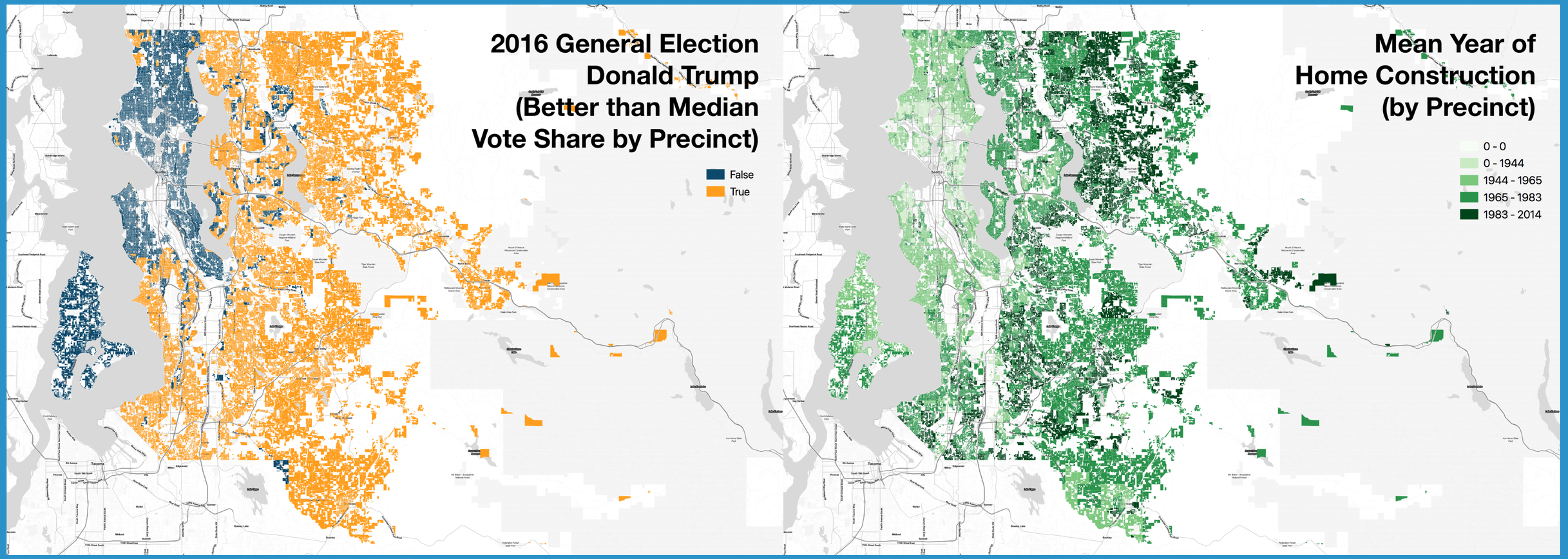

Housing records were aggregated and summarized by the precinct that contained their associated parcels.
Southeast 240th Street
Southeast 256th Street Exit
Southeast 256th Street
56th Street
Lake Wilderness
Lake Meridian
Southeast 272nd Street
SR 516/Covington Exit

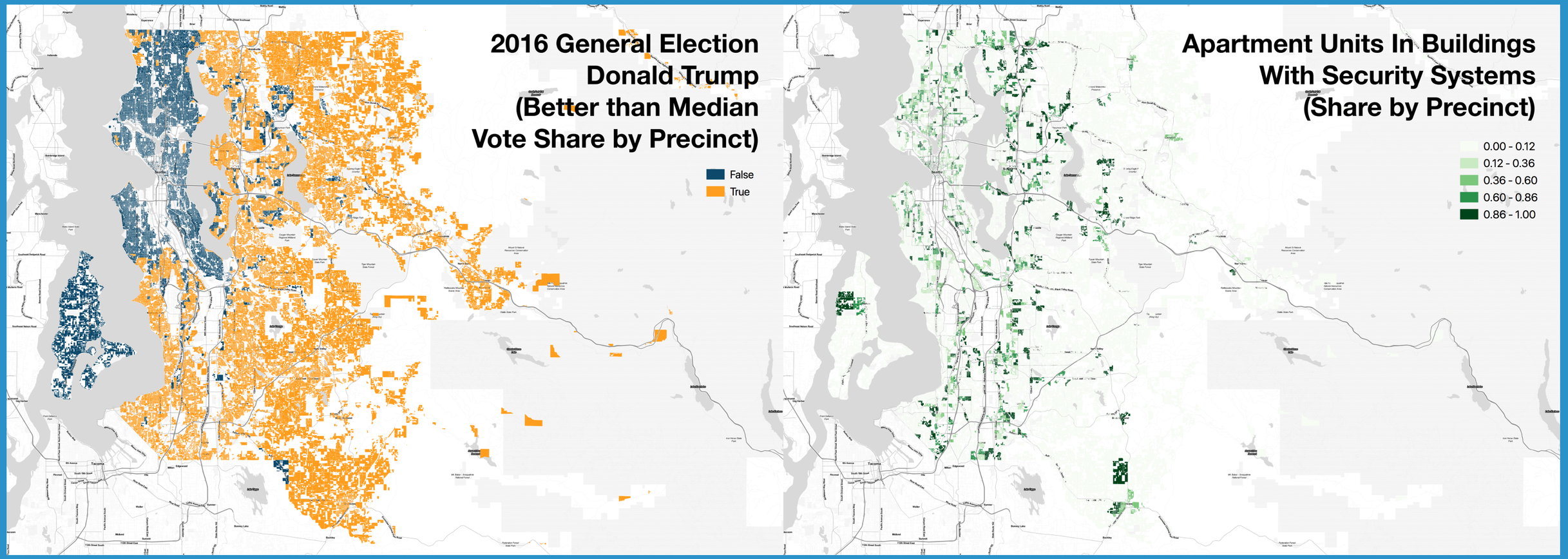
Target Variable

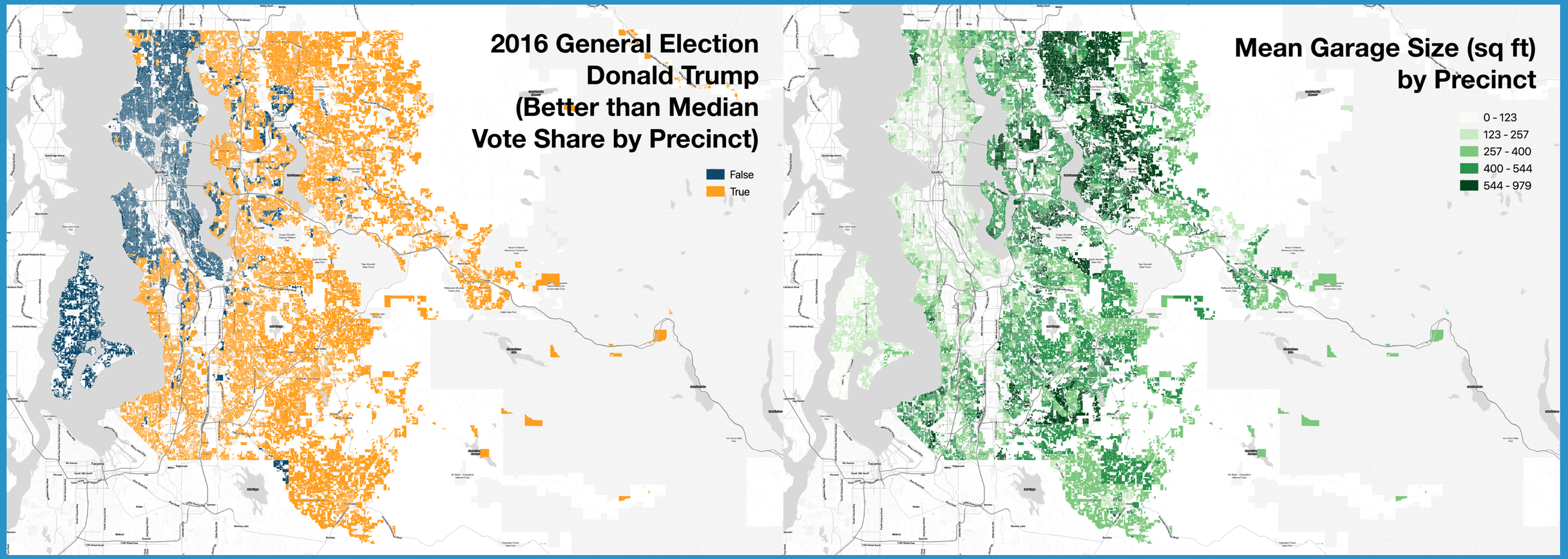



2016 General Election
Donald Trump
(Better than Median
Vote Share by Precinct)
False
True
Improvement Value
Per Housing Unit
(by Precinct)
0 - 227841
227841 - 288831
288831 - 356608
356608 - 473860
473860 - 861093400

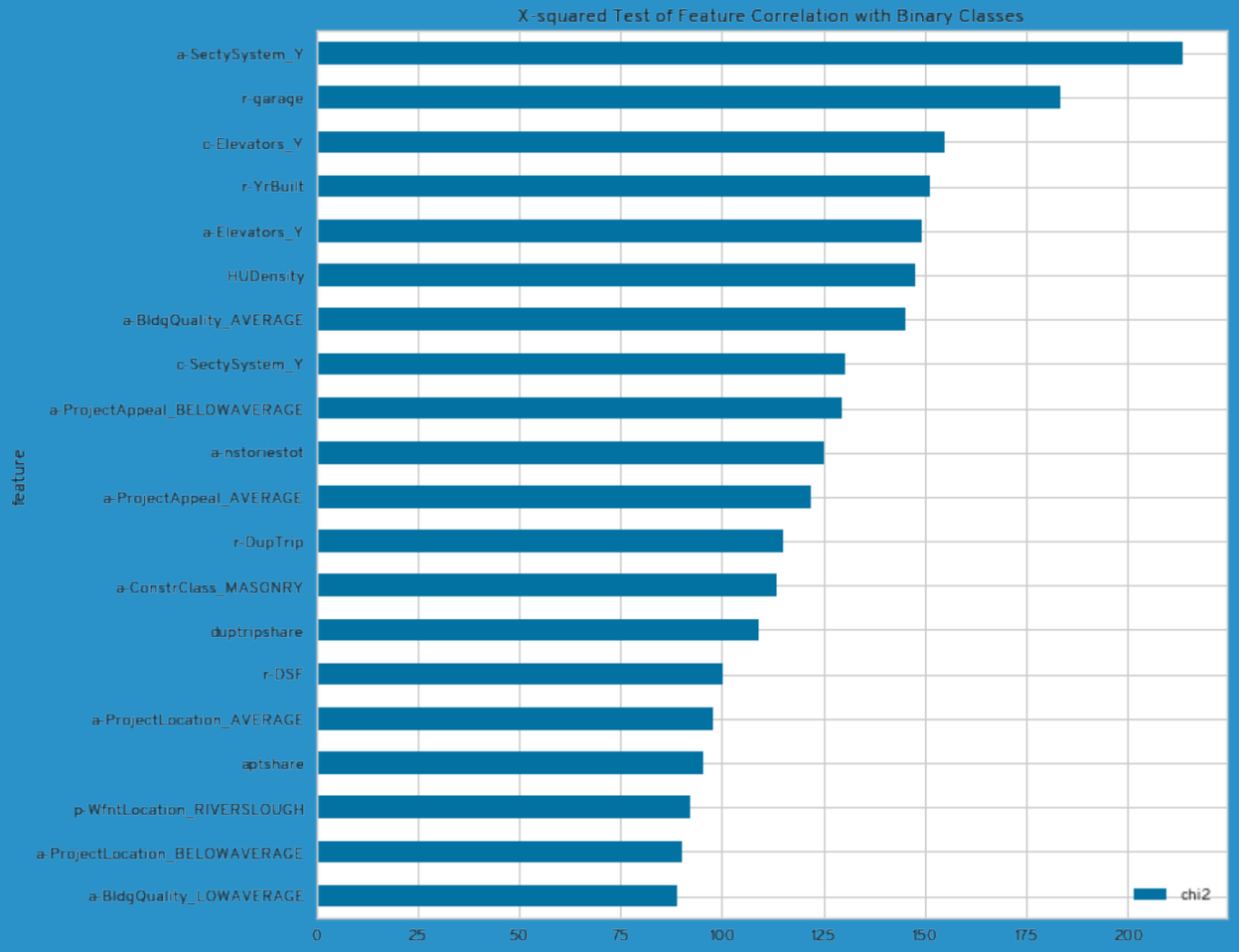

2016 General Election
Donald Trump
(Better than Median
Vote Share by Precinct)
False
True
Apartment Units in
Wood Frame Buildings
(Share by Precinct)
0.00 - 0.17
0.17 - 0.47
0.47 - 0.72
0.72 - 0.91
0.91 - 1.00


2016 General Election
Donald Trump
(Better than Median
Vote Share by Precinct)
False
True
Mean Year of
Home Construction
(by Precinct)
0 - 0
0 - 1944
1944 - 1965
1965 - 1983
1983 - 2014

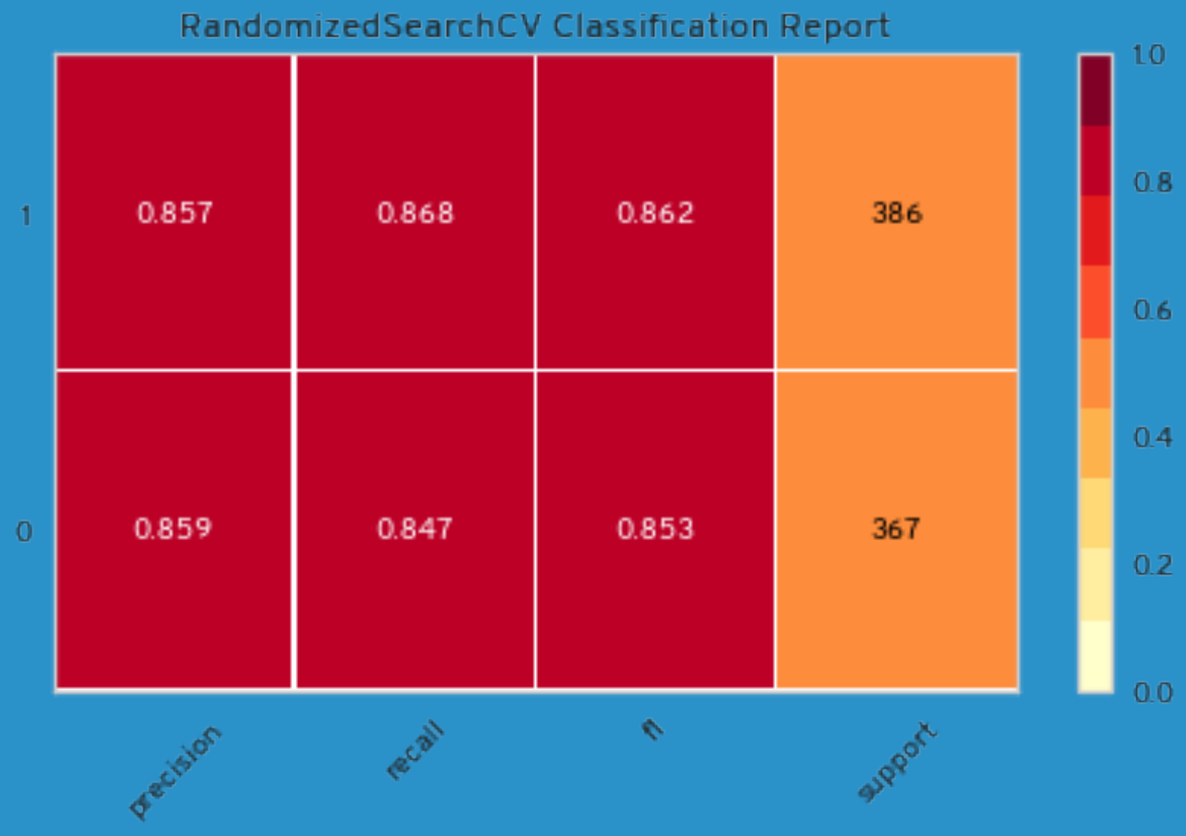

2016 General Election
Donald Trump
(Better than Median
Vote Share by Precinct)
False
True
Apartment Units In Buildings
With Security Systems
(Share by Precinct)
0.00 - 0.12
0.12 - 0.36
0.36 - 0.60
0.60 - 0.86
0.86 - 1.00

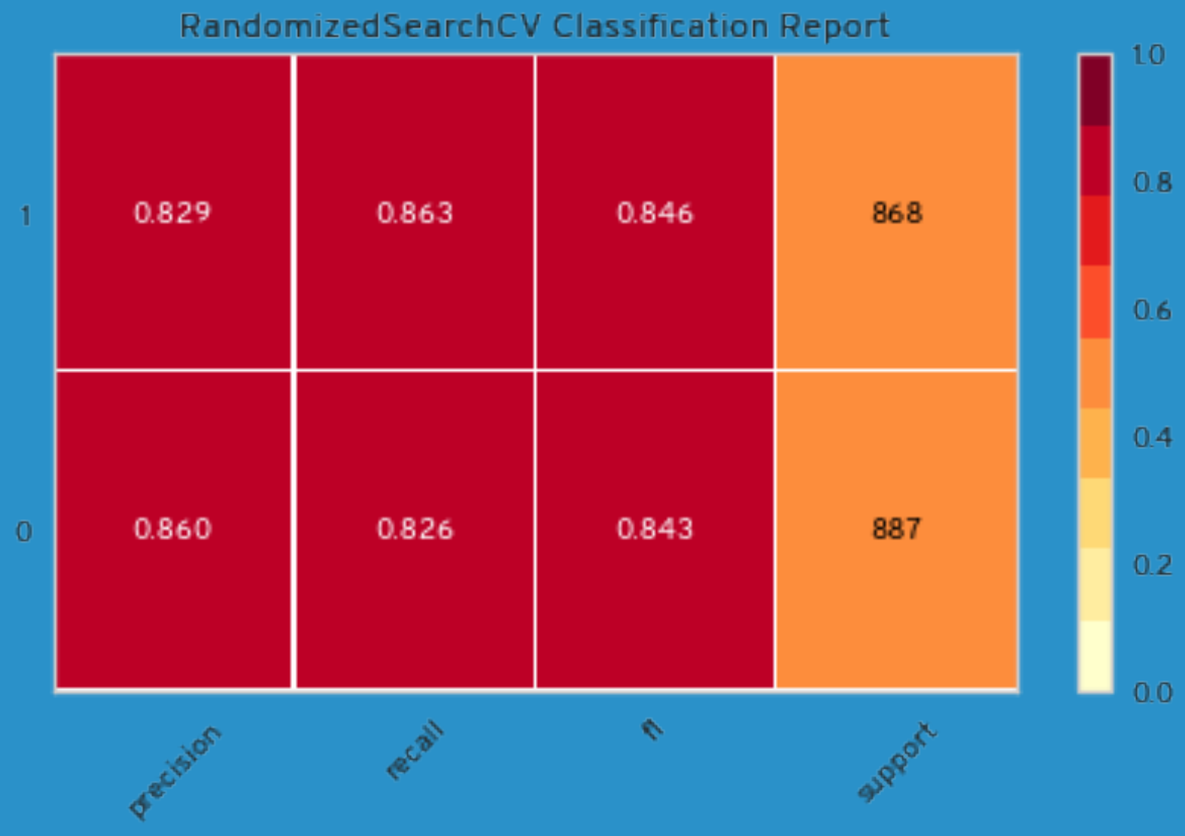

2016 General Election
Donald Trump
(Better than Median
Vote Share by Precinct)
False
True
Mean Garage Size (sq ft)
by Precinct
0 - 123
123 - 257
257 - 400
400 - 544
544 - 979

Feature Correlation (Chi-squared, SelectKBest)

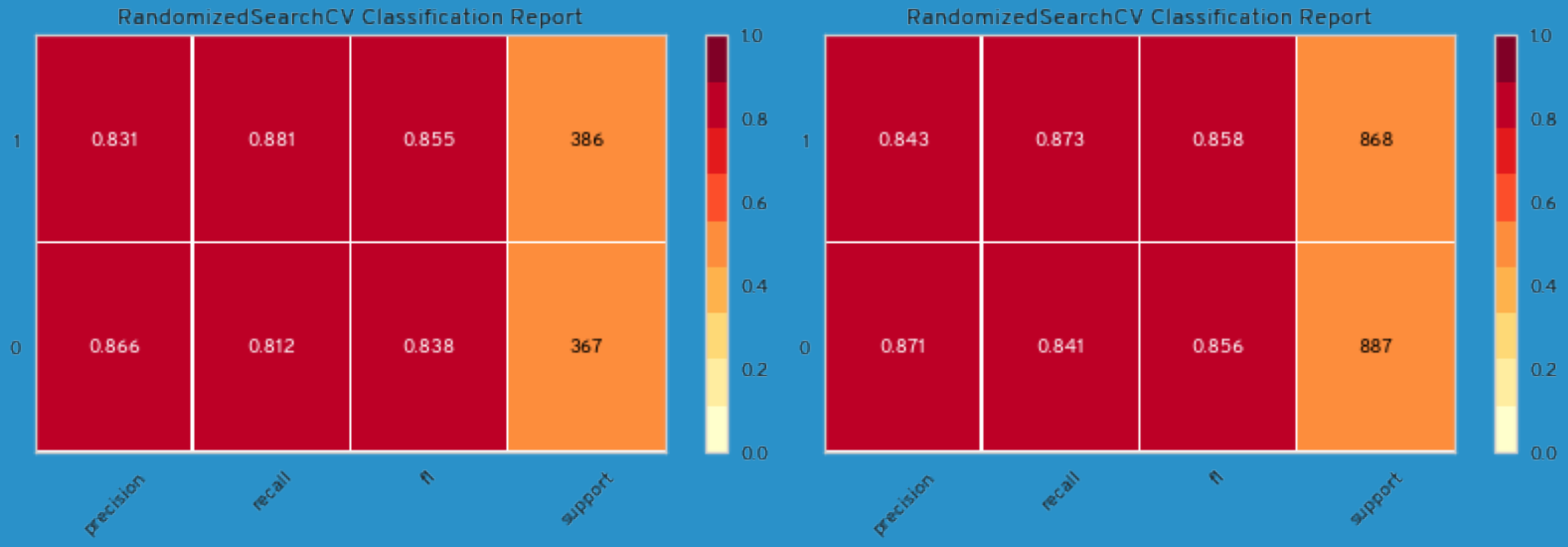


Logistic Regression

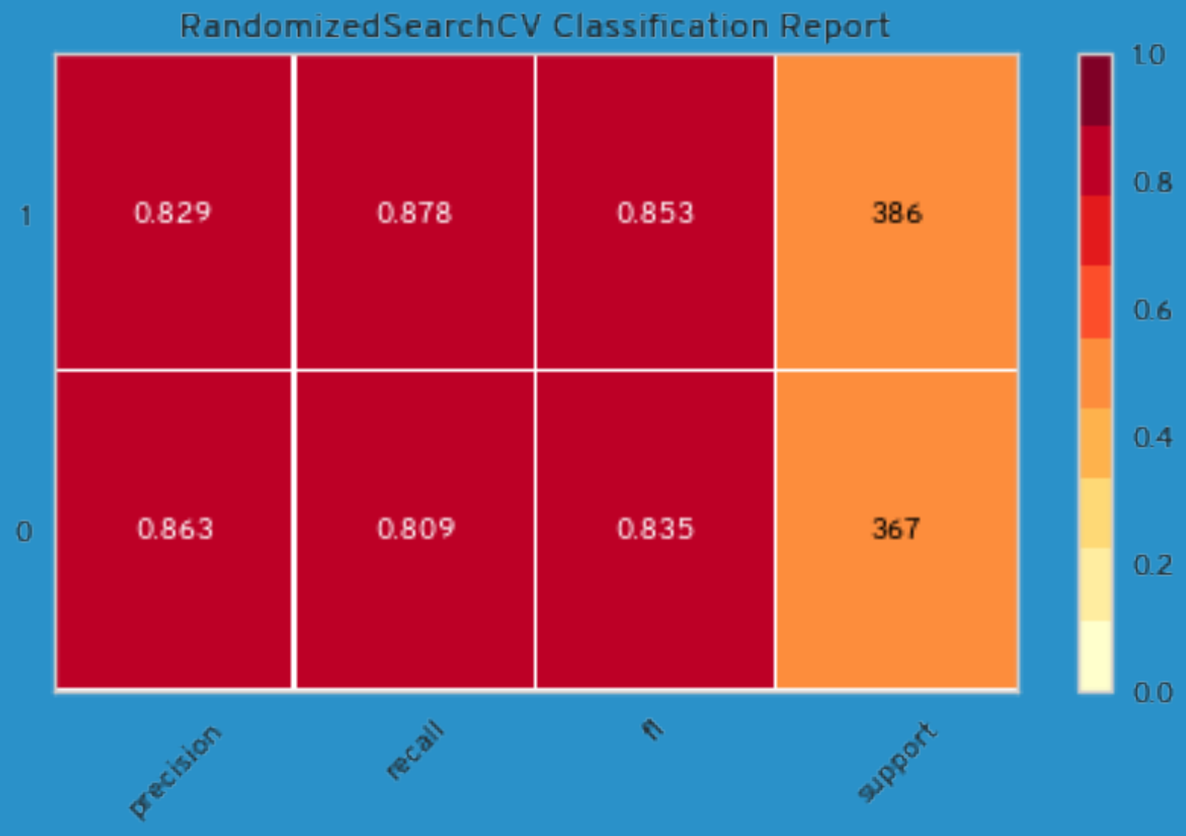


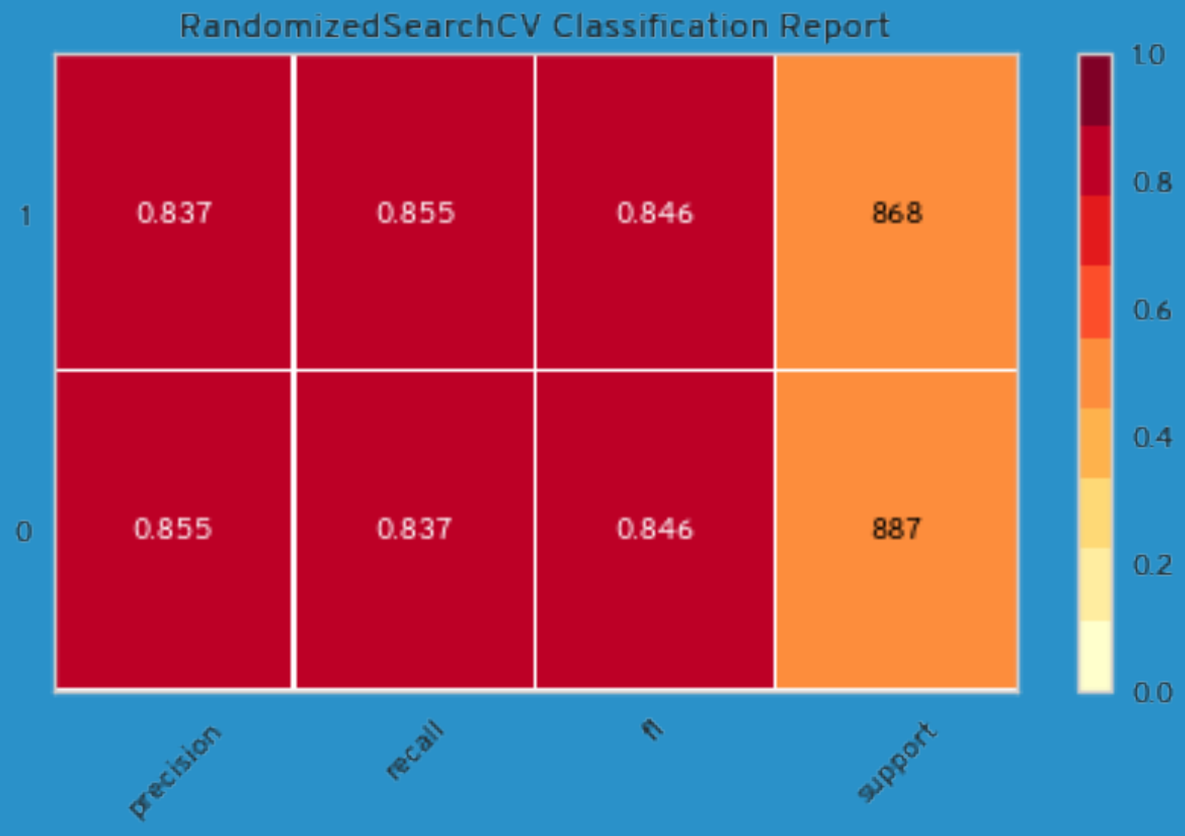


Random Forest Classification

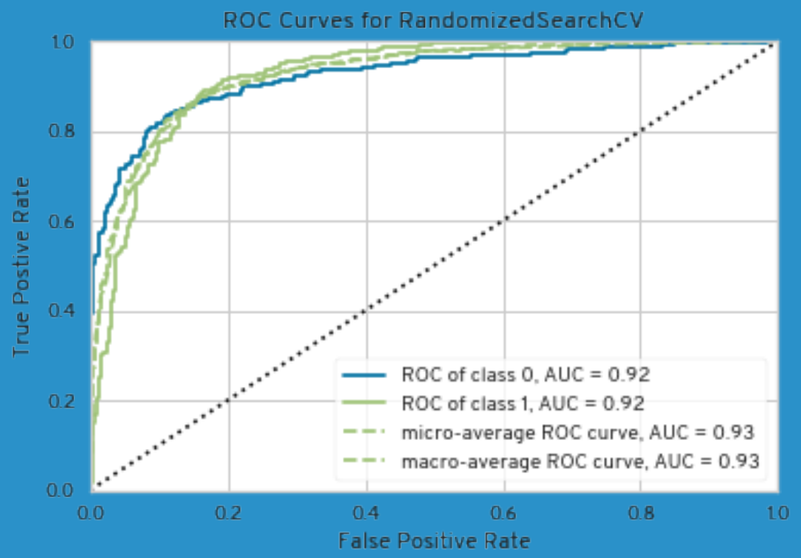


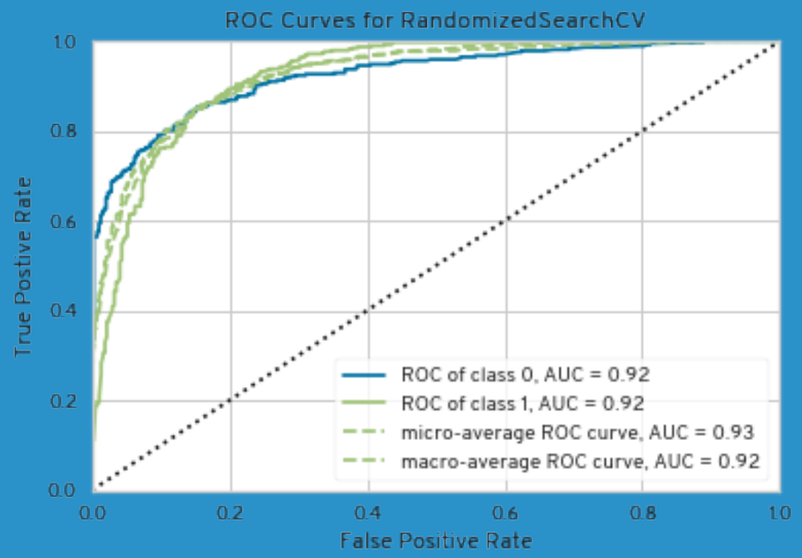
XGBoost Classification

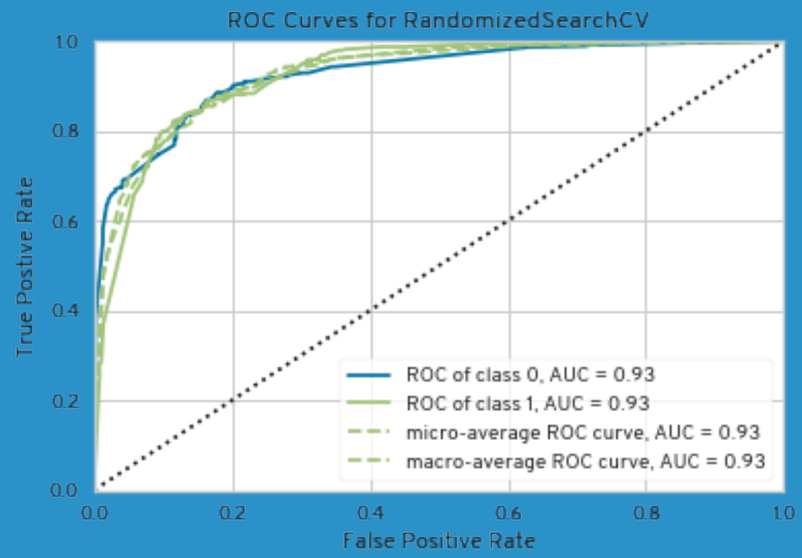




ROC Curves


ROC Curves for RandomizedSearchCV
True Postive Rate
False Positive Rate
ROC of class 0, AUC = 0.92
ROC of class 1, AUC = 0.92
micro-average ROC curve, AUC = 0.93
macro-average ROC curve, AUC = 0.93


ROC Curves for RandomizedSearchCV
True Postive Rate
False Positive Rate
ROC of class 0, AUC = 0.92
ROC of class 1, AUC = 0.92
micro-average ROC curve, AUC = 0.93
macro-average ROC curve, AUC = 0.92


ROC Curves for RandomizedSearchCV
True Postive Rate
False Positive Rate
ROC of class 0, AUC = 0.93
ROC of class 1, AUC = 0.93
micro-average ROC curve, AUC = 0.93
macro-average ROC curve, AUC = 0.93

Mapping Predictive Performance

- Show *where* predidctions were accurate
- Show *where* and *by how much* predictions were inaccurate

Accuracy Comparison

Error Comparison

Sparser features

How does manually selecting a small set of features compare to our K best approach?

Feature Selection 2: By Hand

Model Comparison

- Roughly equivalent performance
- Primary issue is the relative lack of data
- Especially given the richness of the feature set

Practical Uses

- Election narratives in media are not always empirical
- Campaigns and electeds want to understand ongoing trends

Shortcomings

- This single case, with ~2,500 examples, doesn't give us quite enough runway for tuning steps requiring extra validation
- The two tree based algorithms in particular require more data to realize their strengths

Next Steps

- Since I have several thousand example elections, including countywide and