What Happened?

Election 2016

Waylin Wang and Lukas Froehlich

Data Acquisition

- County level election results from townhall.com github repo
  - Contains results from 2008, 2012, and 2016
  - 3112 counties

- 130 demographic features from 2011 - 2015 through US Census API.
  - Includes information about education, income, citizenship, population, race, means of transportation, and relocation.
  - Feature transformation by population

Questions

Which demographics were most significant for voter turnout in 2016 compared to 2012?

Which demographics contributed most to change in voter behavior from 2012 to 2016?

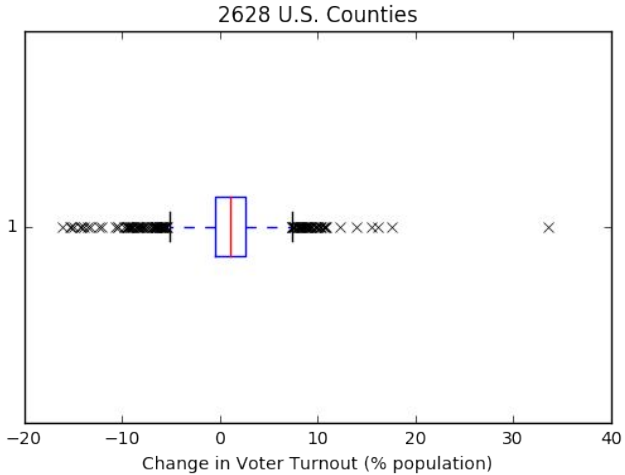
How well can county majority be predicted given demographics?

Exploring Voter Turnout

Voter Turnout

1. Create response variable for voter turnout

```
voter_turnout = (total_votes_2016 - total_votes_2012) / population
```



mean	1.045871
std	3.064287
min	-16.037837
25%	-0.522142
50%	1.143138
75%	2.655468
max	33.625456

2. Break counties into two groups
   a. Very positive change in voter turnout (< Q1)
   b. Very negative change in voter turnout (> Q3)

Voter Turnout

3. Fit random forest to demographics predicting voter turnout class

4. Select most significant variables

5. Compare variables between high and low turnout to see direction of effect

77% prediction accuracy

Demographic	Mean value in low turnout counties	Mean value in high turnout counties
% Black	14.74	5.65
In State-Some College	13.79	10.8
% Black Female	7.51	2.76
Median Age Male	0.28	0.15
Total Reporting 0 Income	10.68	10.6
# of Housing Units	6.5	4.5
Total Born In State education	44.11	36.86

Change in Voter Behaviour

Change in Voting Ratio

Goal: Identify which demographics had a change in voting ratio

1. Create new response variable, voting_ratio_change

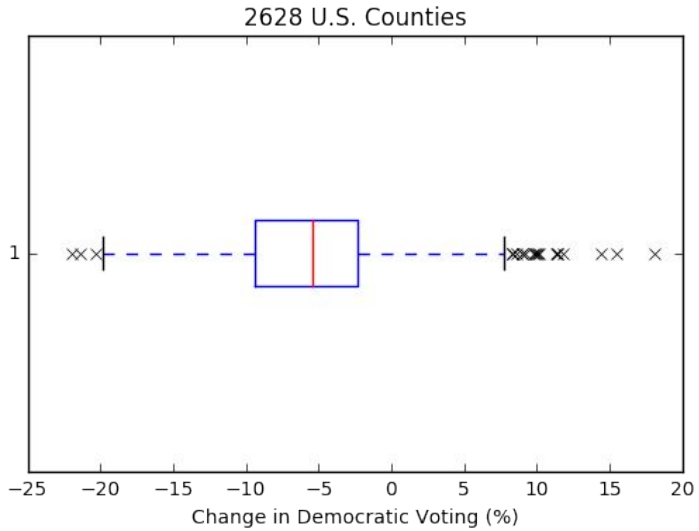
```
voting_ratio_2012 = dem_2012 / total_2012

voting_ratio_2016 = dem_2016 / total_2016

voting_ratio_change = ratio_voting_2016-ratio_voting_2012
```

2. Follow same process as “Voter Turnout”

Change in Voting Ratio



mean	-5.676965
std	5.351930
min	-21.940109
25%	-9.387943
50%	-5.361564
75%	-2.358579
max	18.064662

94.2% prediction accuracy

Demographic	Mean value for county that voted more GOP	Mean value for county that voted more DEM
% White Not Hispanic Male	45.33	32.55
Foreign Born below 100% poverty	0.4	1.9
Foreign Born at/above 1.5X poverty line	1.2	6.23
% White Not Hispanic Female	45.62	33.38
Citizen by Naturalization	0.82	3.88
HS or equal	27.76	17.68
Total Born In State education	49.82	32.99

Change in Voter Behaviour w/ Logistic Regression

- Which demographics contributed most to change in voter behavior from 2012 to 2016?

Approach

- Use Lasso/Elastic Net to select top features from 2012 and 2014 and compare Logistic Regression coefficients

Demographic Effects in 2012

Logistic Regression Coefficients (Accuracy: 87.93%)

Currently Married	-0.317283
Moved Within Same State	-0.280607
Other State-HS or equal	-0.252629
Less Than HS	-0.234204
Graduate	-0.164781
Total Reporing 0 Income	-0.137280
# White Not Hispanic	-0.097915
2 vehicles	-0.087223
Same house 1 year ago	-0.070133
Median Income	-0.004497
Median Income Born Outside US: Native	-0.001337
Median Income Born Other State	0.000057
Median Income Foreign Born	0.000827
Median Income Born In State	0.001208
Same house 1 year ago White	0.037653
100% below povery line	0.069505
Work outside County	0.078911
Foreign born naturalized	0.104808
Citizen by Naturalization	0.104808
Never Married	0.112771
Work outside State	0.134887
# of Housing Units	0.154294
Income 35-50K	0.213100
Education Count	0.296345
Income 50-65K	0.335249
Other State-Graduate	0.363238
Walked	0.590713

Demographic Effects in 2016

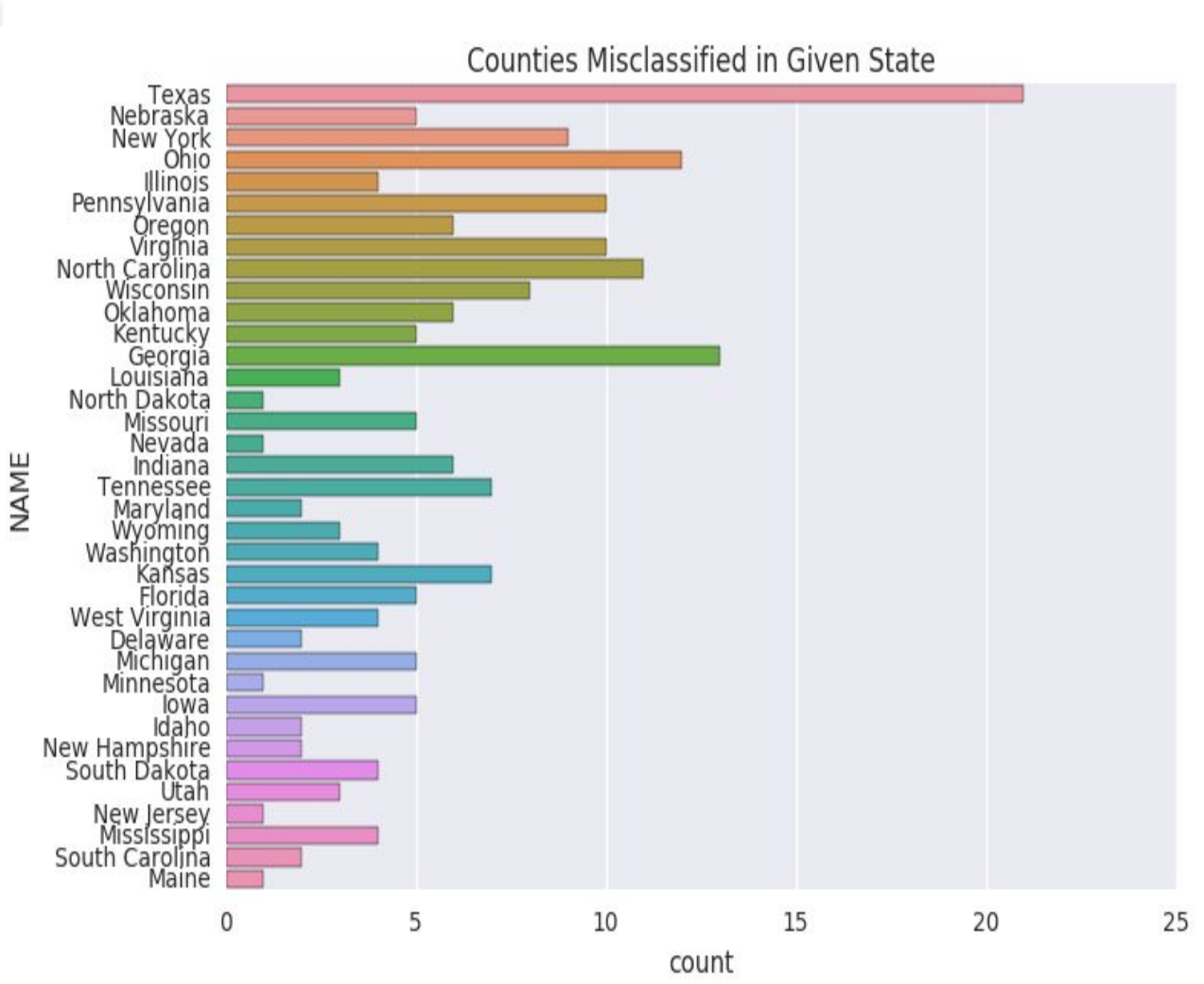
Logistic Regression Coefficients (Accuracy: 93.1%)

Less Than HS	-0.265917
# White Not Hispanis Female	-0.262218
Moved Within Same State	-0.246613
Moved From Different State	-0.219069
Born Other State at/above 1.5X poverty line	-0.208518
Total Reporing 0 Income	-0.191759
Income > 75K	-0.181936
Currently Married	-0.168618
No Vehicle Owned	-0.156645
Same house 1 year ago	-0.090056
Graduate	-0.063102
Work in County	-0.043940
Bachelor	-0.039522
In State-HS or equal	-0.029615
# Hispanic	-0.010952
# White Not Hispanic	-0.008495
Median Income	-0.005299
1-1.5X of poverty line	-0.005140
Median Income Born Other State	-0.001758
Median Income Born Outside US: Native	0.000871
Median Income Foreign Born	0.003692
100% below povery line	0.023298
Same house 1 year ago White	0.026474
# Black Female	0.029098
Work outside County	0.036297
Foreign born naturalized	0.174013
Citizen by Naturalization	0.174013
Work outside State	0.196733
Never Married	0.202886
2 vehicles	0.220181
Education Count	0.289691
Other State-Bachelor	0.427380
Walked	0.574626
Other State-Graduate	0.836953

Difference

	Field	2016_coef	2012_coef	diff
1	# White Not Hispanis Female	-0.262218	0.000000	-0.262218
3	Moved From Different State	-0.219069	0.000000	-0.219069
4	Born Other State at/above 1.5X poverty line	-0.208518	0.000000	-0.208518
6	Income > 75K	-0.181936	0.000000	-0.181936
8	No Vehicle Owned	-0.156645	0.000000	-0.156645
11	Work in County	-0.043940	0.000000	-0.043940
12	Bachelor	-0.039522	0.000000	-0.039522
13	In State-HS or equal	-0.029615	0.000000	-0.029615
14	# Hispanic	-0.010952	0.000000	-0.010952
17	1-1.5X of poverty line	-0.005140	0.000000	-0.005140
18	Median Income Born Other State	-0.001758	0.000057	-0.001814
19	Median Income Born Outside US: Native	0.000871	-0.001337	0.002207
23	# Black Female	0.029098	0.000000	0.029098
29	2 vehicles	0.220181	-0.087223	0.307405
31	Other State-Bachelor	0.427380	0.000000	0.427380
34	Other State-HS or equal	0.000000	-0.252629	0.252629
35	Median Income Born In State	0.000000	0.001208	-0.001208
36	# of Housing Units	0.000000	0.154294	-0.154294
37	Income 35-50K	0.000000	0.213100	-0.213100
38	Income 50-65K	0.000000	0.335249	-0.335249

Misclassified Counties



Classifying County Voting

County Voting

Goal: Try and uncover interesting characteristics while classifying which way a county voted given the demographics

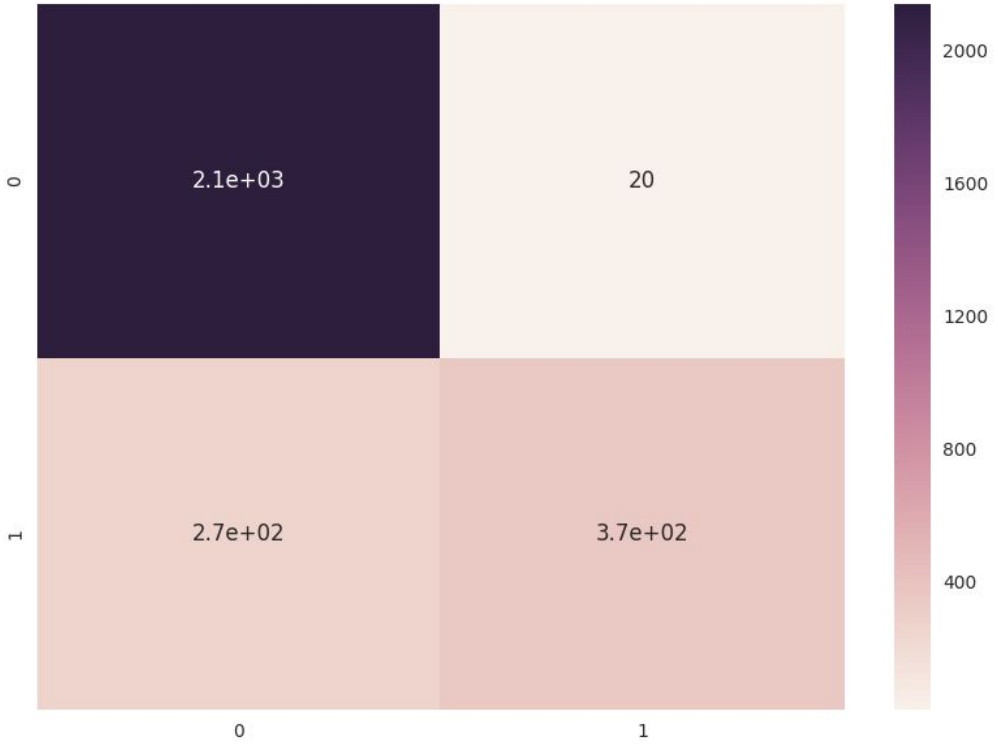
1. Create new classification variable: “dem” “gop”
2. Predict!

Using SKLearn's MLP

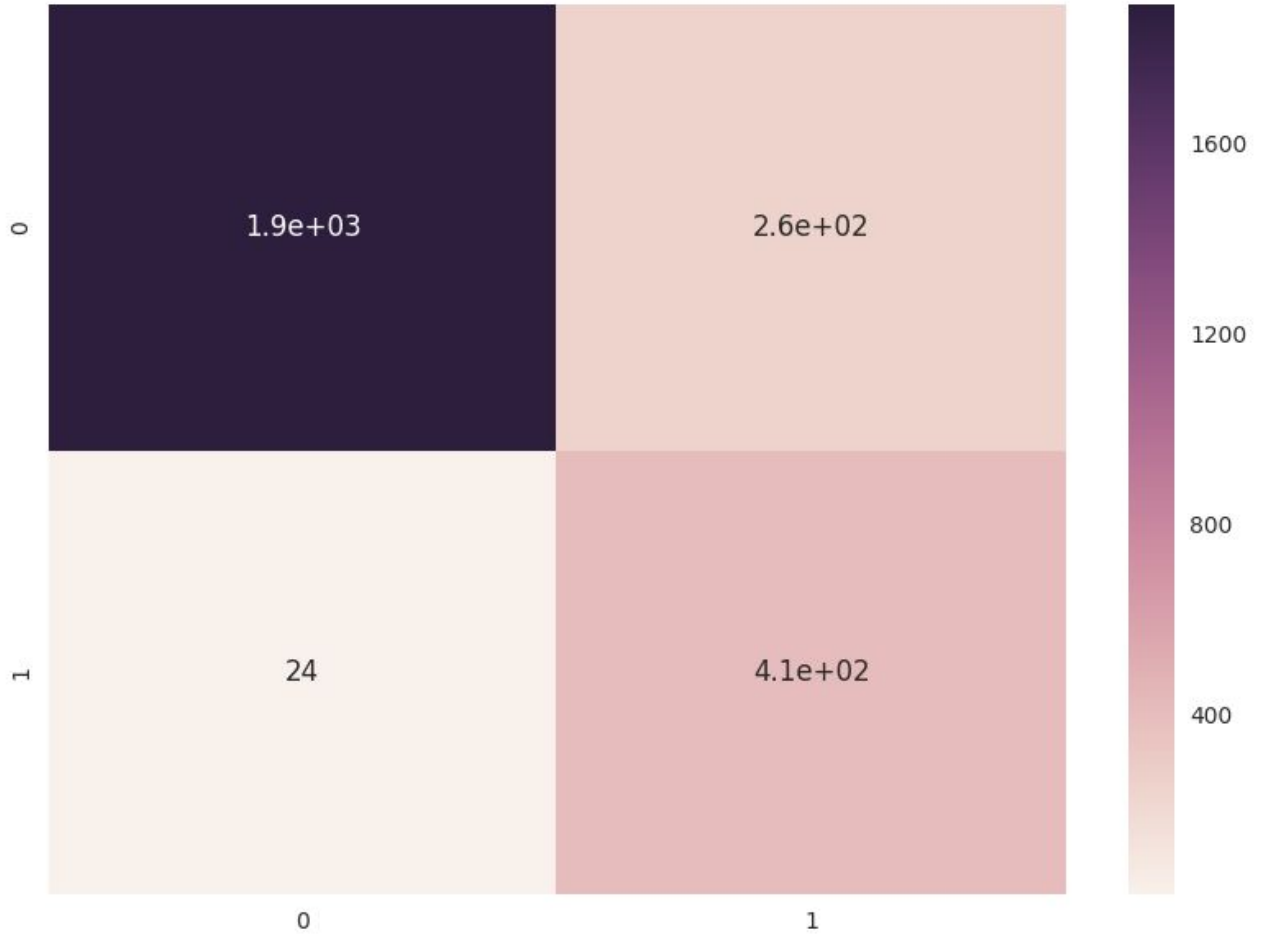
- Loss function: Log-Sigmoid
- Adaptive Learning Rate
- LBFGS solver
- 2 Hidden Layers with 100 nodes/layer
- Learning Rate: 0.001
- L1 Penalty: 1

Cross-Validated F1 Score: .81

Accuracy on predicting 2016: 89.04%

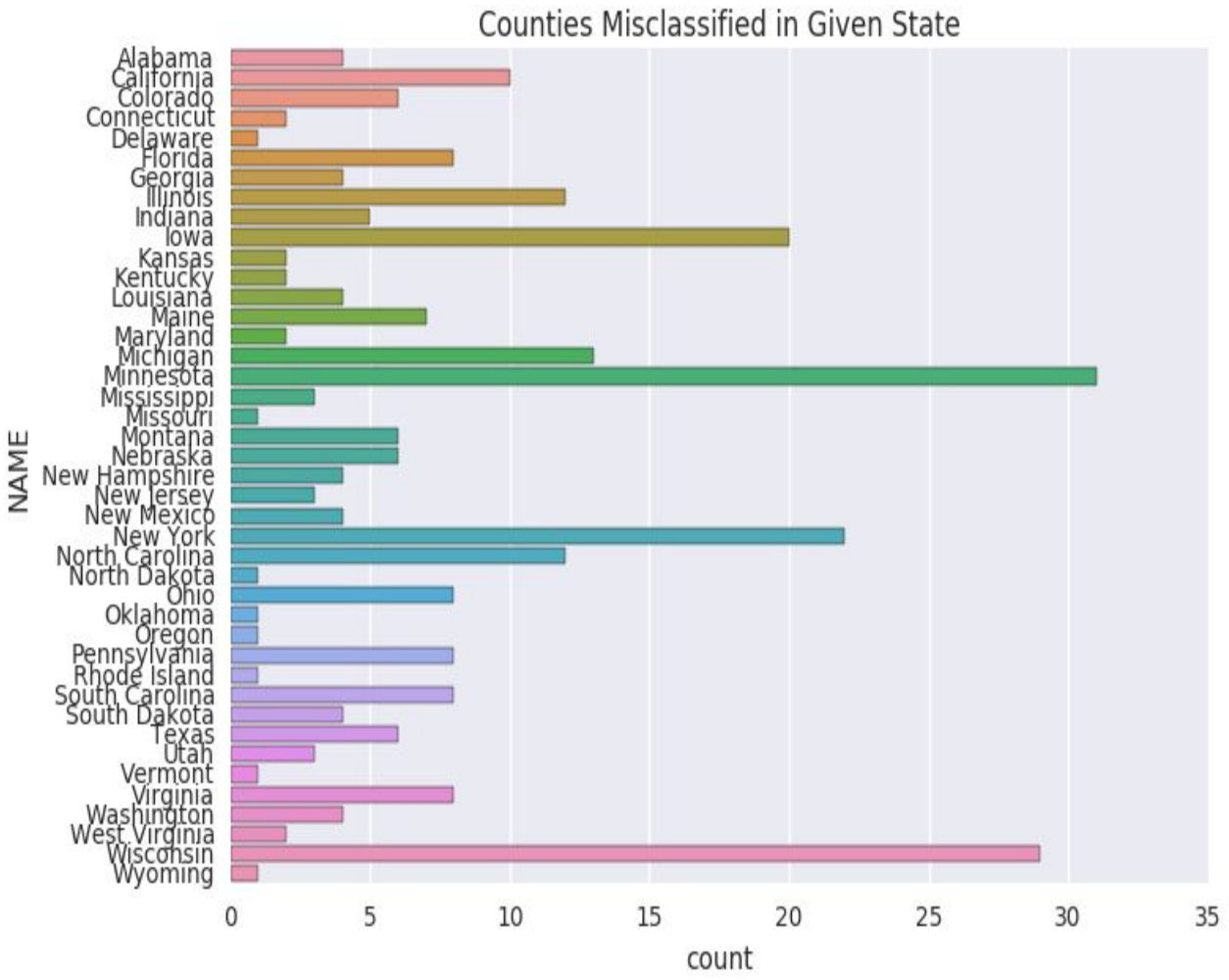


Classifying 2016 Demographics using 2012 Model



Misclassified Counties

Minnesota	31
Wisconsin	29
New York	22
Iowa	20
Michigan	13
Illinois	12
North Carolina	12
California	10
Virginia	8
Florida	8
South Carolina	8
Pennsylvania	8
Ohio	8
Maine	7
Nebraska	6
Montana	6
Texas	6
Colorado	6
Indiana	5
Louisiana	4
Georgia	4
Alabama	4
New Hampshire	4
New Mexico	4
Washington	4
South Dakota	4



Reference

Voter Turnout

Positive Change in Voter Turnout

Polk County 33.63
Jerome County 17.65
Jasper County 16.22
Sumter County 15.60
McKenzie County 14.05

Negative Change in Voter Turnout

Otero County -16.04
Terry County -15.33
Newton County -15.21
Iron County -14.92
Whitman County -14.24

Change in Voting Ratio

Positive Change in Voting Ratio

```
Salt Lake County 18.06
Cache County 15.51
Davis County 14.39
Arlington County 11.79
Falls Church City 11.41
```

Negative Change in Voting Ratio

```
Clark County -21.94
Henderson County -21.41
Monroe County -20.31
Iron County -19.84
Reynolds County -19.12
```