Predicting Asian American Party Affiliation

Akash Palani

Data: National Asian American Voter Survey

- Telephone interview data
- National sample of Asian American Voters
- Demographics
- Social identities, attitudes
- Political attitudes, behavior, and policy views

Question: How well can we predict party affiliation using policy positions?

Exploratory Data Analysis

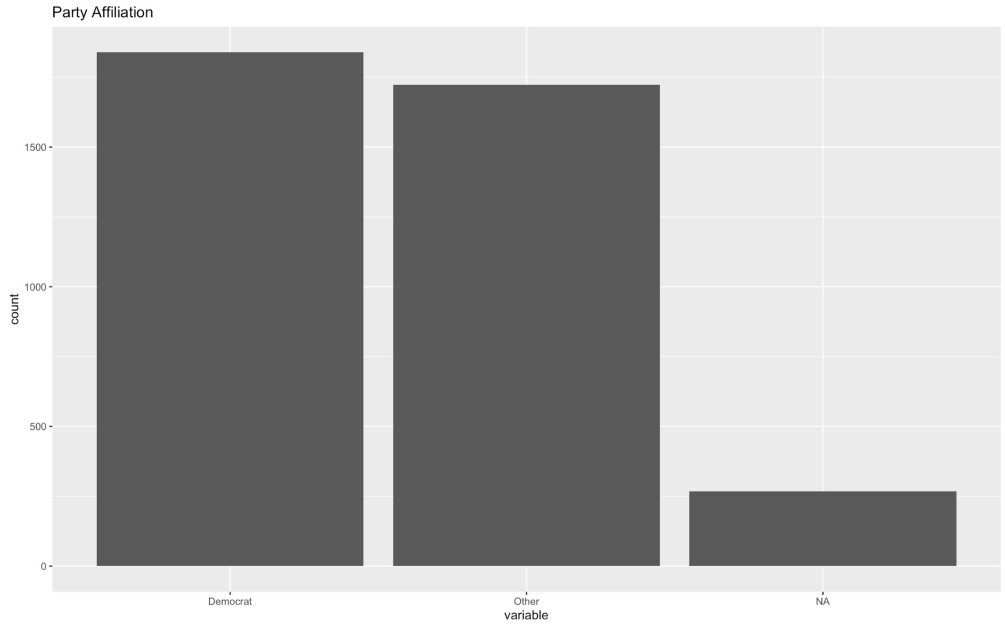
Variables of Interest

Response: Party Affiliation

Predictor Candidates:

- Most important issue facing US
- Obamacare (Support/Oppose/Don't Know)
- Free public college (S/O/DK)
- Syrian refugee entry (S/O/DK)
- Marijuana legalization (S/O/DK)
- Muslim ban (S/O/DK)
- Emission limits on power plans (S/O/DK)
- Government action to promote equal rights (S/O/DK)

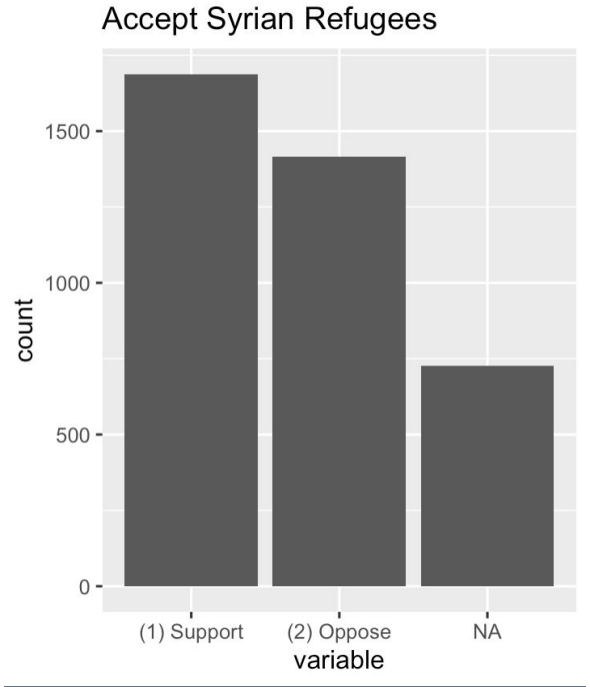
Distribution of Response

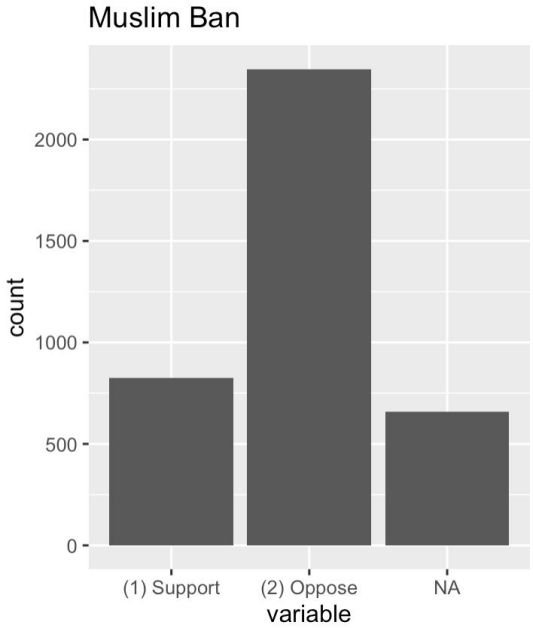


Missingness

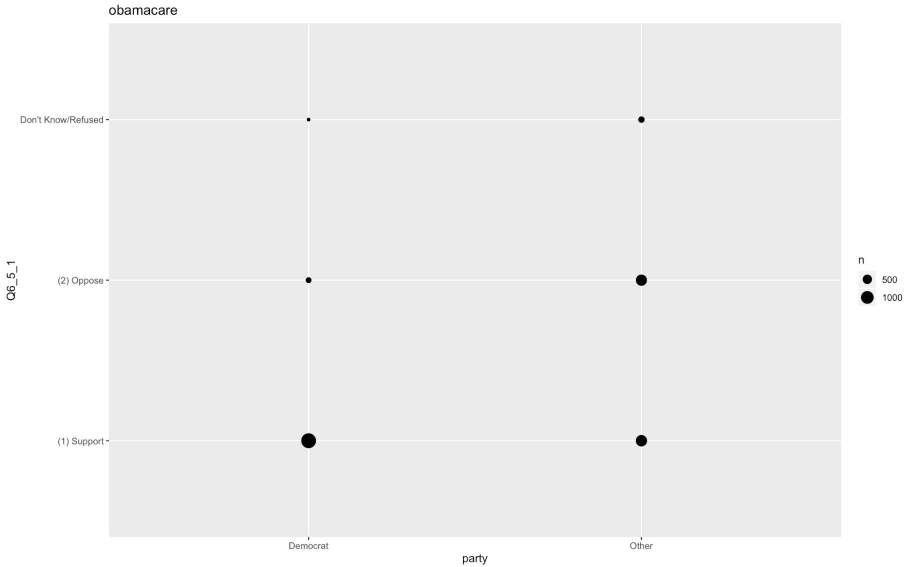
```
── Variable type: factor ──────────────────────────────────────────────────────────────────────────
  skim_variable n_missing complete_rate ordered n_unique top_counts
1 Q6_1A                56         0.870 FALSE         15 (01: 79, (14: 72, (13: 51, (09: 44
2 Q6_5_1               85         0.803 FALSE          2 (1): 209, (2): 137
3 Q6_5_2               61         0.858 FALSE          2 (1): 242, (2): 128
4 Q6_5_3              136         0.684 FALSE          2 (2): 174, (1): 121
5 Q6_5_4               75         0.826 FALSE          2 (2): 225, (1): 131
6 Q6_5_5              124         0.712 FALSE          2 (2): 193, (1): 114
7 Q6_5_6              107         0.752 FALSE          2 (1): 236, (2): 88
8 Q6_5_7              100         0.768 FALSE          2 (1): 248, (2): 83
9 party                50         0.884 FALSE          2 Oth: 226, Dem: 155

── Variable type: list ────────────────────────────────────────────────────────────────────────────
  skim_variable n_missing complete_rate n_unique min_length max_length
1 train                 0             1        1          9          9
```

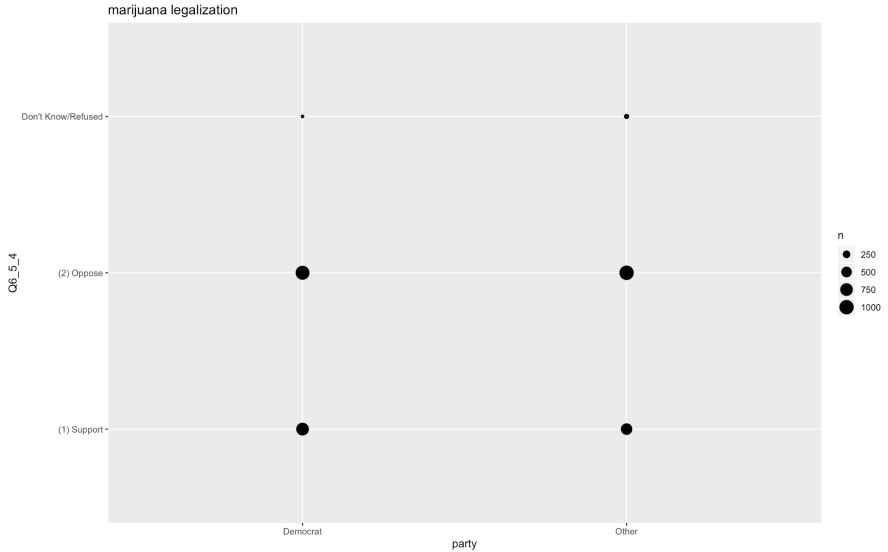
Distribution of Predictors





Covariation





Model Building

Candidate Models

Logistic Regression

- mod_1 = party ~ .
- mod_2 = party ~ Q6_1A + Q6_5_1 + Q6_5_2 + Q6_5_3 + Q6_5_7
- mod_3 = party ~ Q6_1A

LDA

- mod_1 = party ~ .
- mod_2 = party ~ Q6_1A + Q6_5_1 + Q6_5_2 + Q6_5_3 + Q6_5_7
- mod_3 = party ~ Q6_1A

Logistic Regression Performance

Error rates, confusion matrices

	model_name	train_error	test_error
	<chr>	<dbl>	<dbl>
1	mod_01	0.323	0.365
2	mod_02	0.325	0.373
3	mod_03	0.459	0.439

	model_name	party	pred_party	prop
	<chr>	<fct>	<chr>	<dbl>
1	mod_01	Democrat	Democrat	0.257
2	mod_01	Democrat	Other	0.190
3	mod_01	Other	Democrat	0.175
4	mod_01	Other	Other	0.378
5	mod_02	Democrat	Democrat	0.251
6	mod_02	Democrat	Other	0.196
7	mod_02	Other	Democrat	0.177
8	mod_02	Other	Other	0.376
9	mod_03	Democrat	Democrat	0.169
10	mod_03	Democrat	Other	0.278
11	mod_03	Other	Democrat	0.161
12	mod_03	Other	Other	0.392

LDA Performance

Error rates, confusion matrices

	model_name	train_error	test_error
	<chr>	<dbl>	<dbl>
1	mod_01	0.322	0.368
2	mod_02	0.323	0.376
3	mod_03	0.459	0.439

	model_name	party	pred_party	prop
	<chr>	<fct>	<fct>	<dbl>
1	mod_01	Democrat	Democrat	0.259
2	mod_01	Democrat	Other	0.188
3	mod_01	Other	Democrat	0.180
4	mod_01	Other	Other	0.373
5	mod_02	Democrat	Democrat	0.262
6	mod_02	Democrat	Other	0.185
7	mod_02	Other	Democrat	0.190
8	mod_02	Other	Other	0.362
9	mod_03	Democrat	Democrat	0.169
10	mod_03	Democrat	Other	0.278
11	mod_03	Other	Democrat	0.161
12	mod_03	Other	Other	0.392

Results: LDA Model 1

Coefficients of linear discriminants:	
	LD1
Q6_1A(02) Education	0.223095838
Q6_1A(03) Environment/Climate change	0.163689689
Q6_1A(04) Foreign Policy	0.403162236
Q6_1A(05) Government/Broken political system	0.120604991
Q6_1A(06) Health care	0.642738313
Q6_1A(07) Incomes/Gap between rich and poor	0.116736550
Q6_1A(08) Immigration	-0.215027309
Q6_1A(09) Racism or Racial discrimination	-0.095785289
Q6_1A(10) Sexism or Gender discrimination	0.464815022
Q6_1A(11) Social security	-0.348517162
Q6_1A(12) Taxes	-0.311295151
Q6_1A(13) Terrorism/ISIS/National security	0.242865586
Q6_5_1(2) Oppose	1.803018976
Q6_5_1Don't Know/Refused	1.088262040
Q6_5_2(2) Oppose	0.647295039
Q6_5_2Don't Know/Refused	0.130873765
Q6_5_3(2) Oppose	0.297840103
Q6_5_3Don't Know/Refused	0.141913572
Q6_5_4(2) Oppose	-0.067325116
Q6_5_4Don't Know/Refused	0.042646232
Q6_5_5(2) Oppose	-0.058014017
Q6_5_5Don't Know/Refused	0.295000687
Q6_5_6(2) Oppose	0.003167372
Q6_5_6Don't Know/Refused	0.084922101
Q6_5_7(2) Oppose	0.538065252
Q6_5_7Don't Know/Refused	0.195852245

Best train error: 0.322.

Next steps

Lasso, subset selection, more predictors (values, demographics, all?)