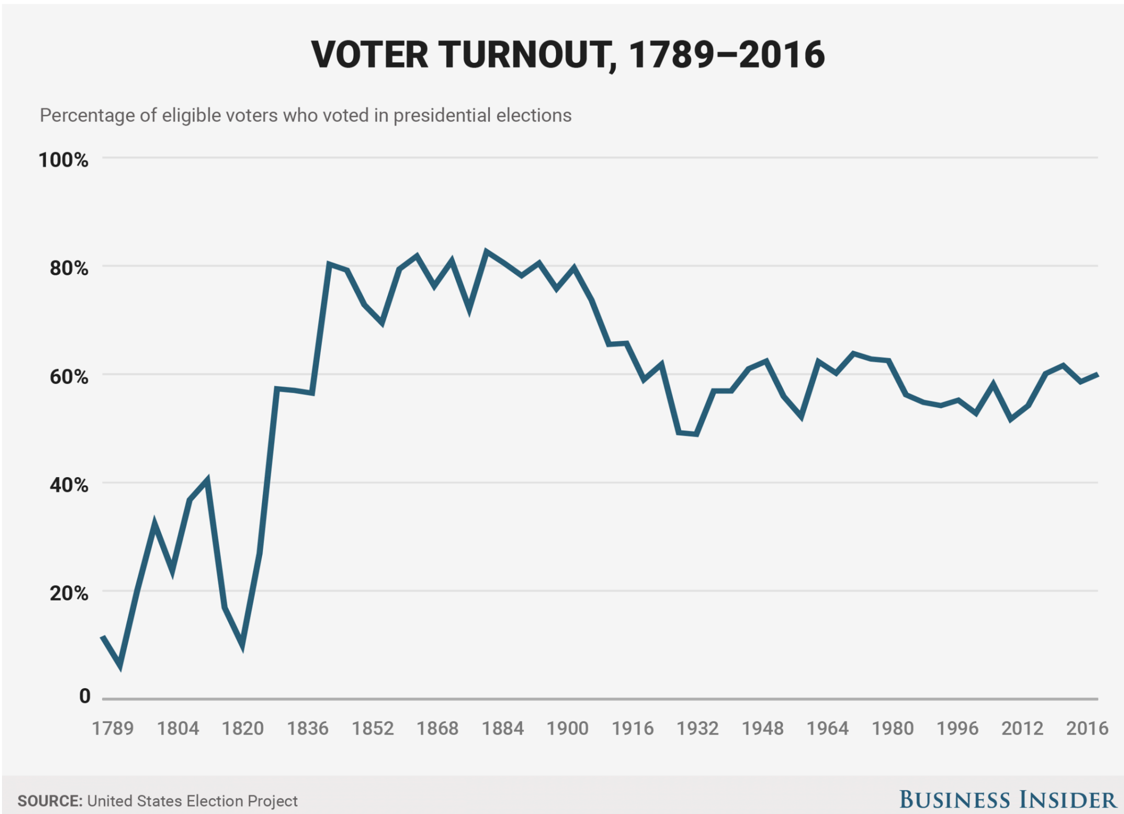
Improving Polling Accuracy in US Presidential Elections

A machine learning approach to developing likely voter models

Why do we care?

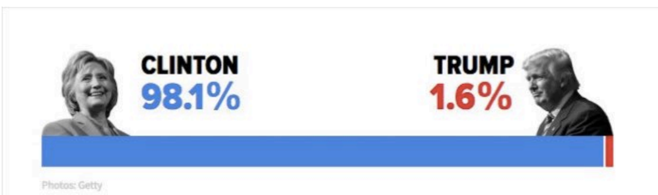


- For the past two decades, less than 60% of eligible voters participated in presidential elections
- Difficulty of determining who will vote is a primary factor in polling inaccuracy

Garbage In, Garbage Out

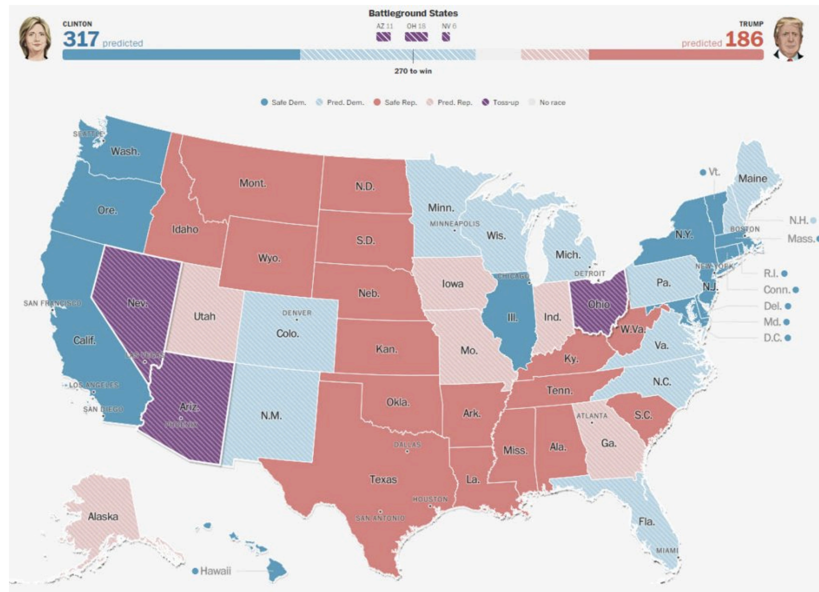
Huffington Post @HuffingtonPost

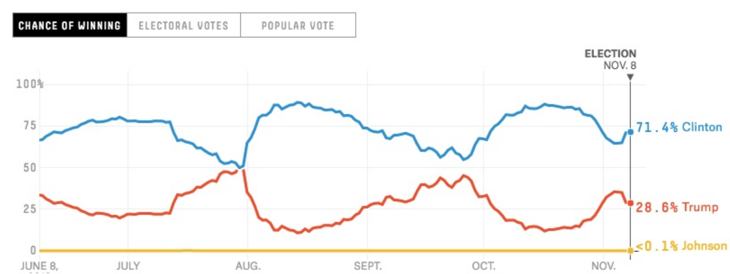
Our @pollsterpolls model gives @HillaryClinton a 98.1% chance of winning the presidency elections.huffingtonpost.com/2016/forecast/...



2016/11/07, 11:25

4 728 RETWEETS **3 812** LIKES





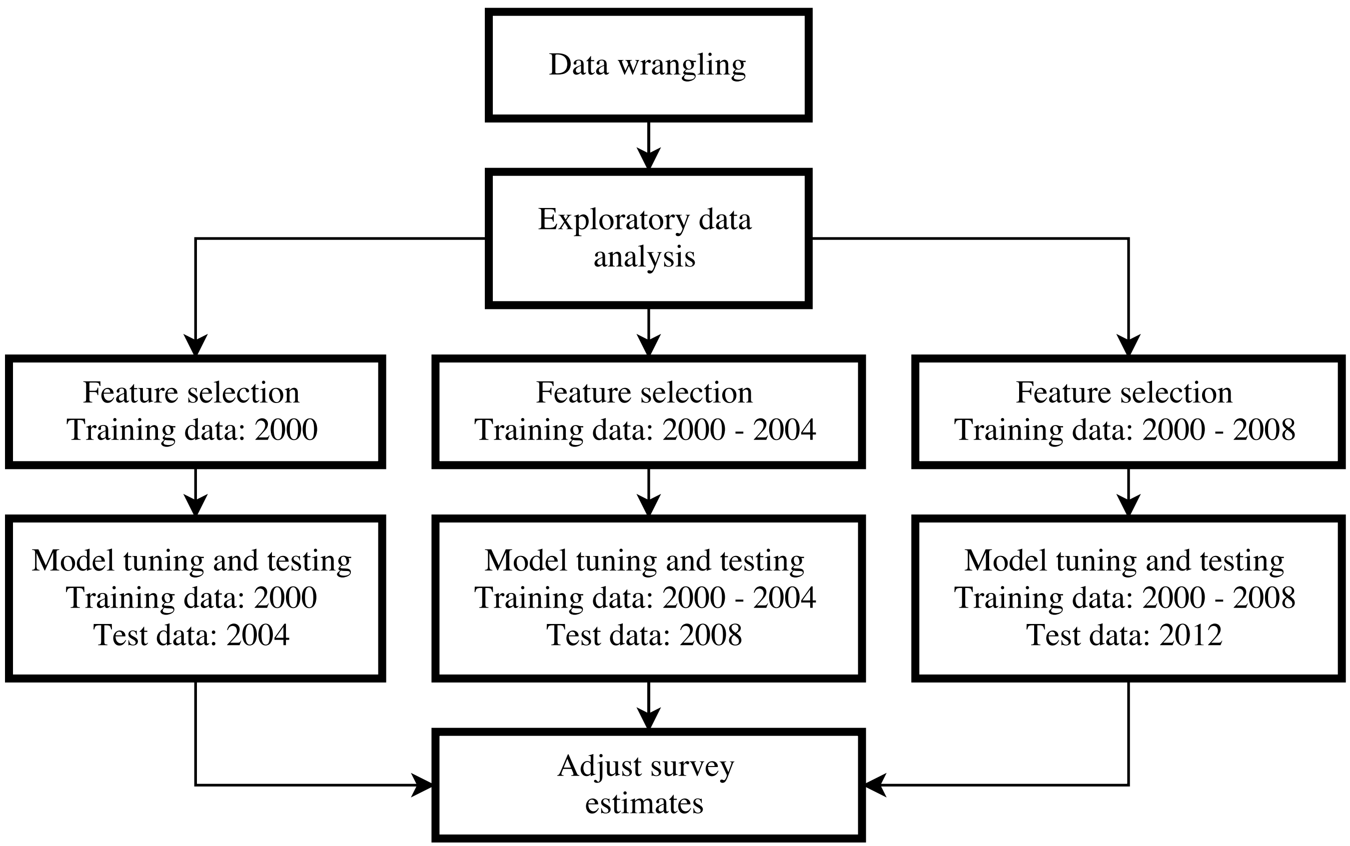
American National Election Studies

- ANES is a research organization specializing in time series studies of voting, public opinion, and political participation
- Cumulative time series data
  - 50000+ respondents, 900+ features
  - Each respondent interviewed prior to election as well as after
  - All pre and post-election interviews from 1948 to 2012

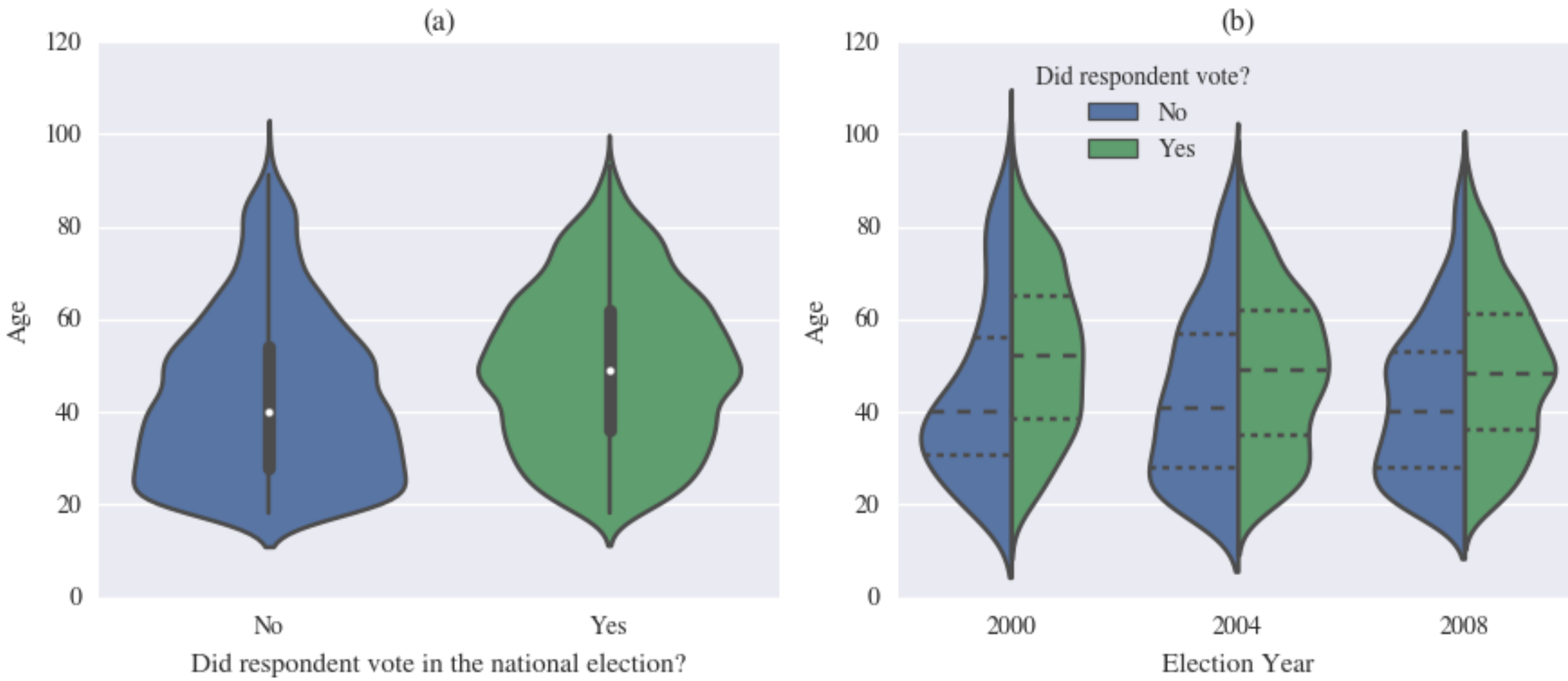
For This Project

- Two goals:
  - Find best classification model for separating voters from non-voters
  - Determine change in polling error
- Interested in recent elections
  - Data from 2000 to 2012
  - 9,374 respondents and 263 features

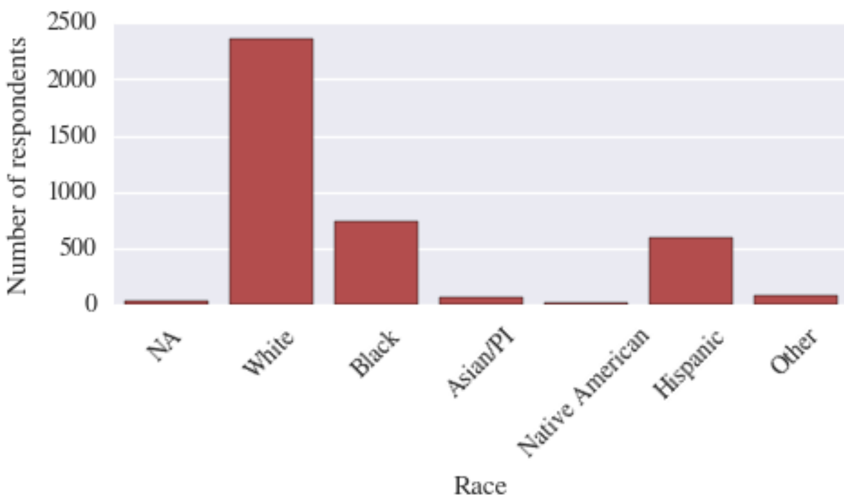
Project Flow

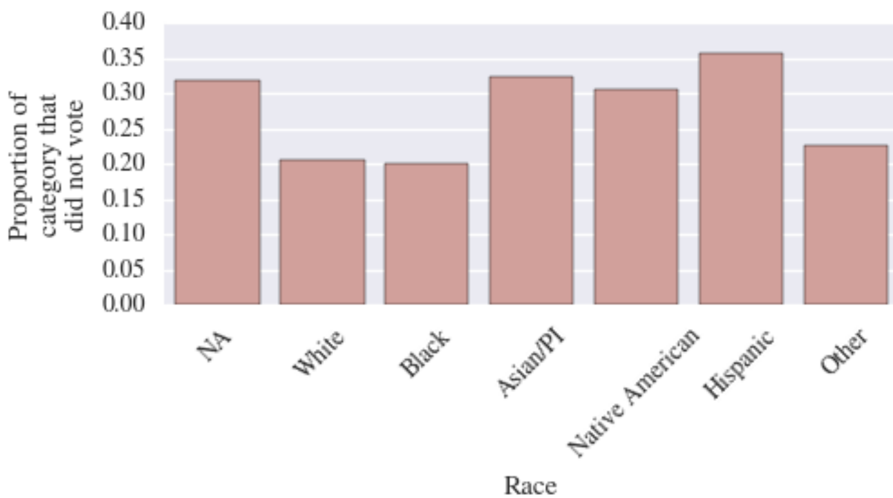


Non-Voters Skew Younger



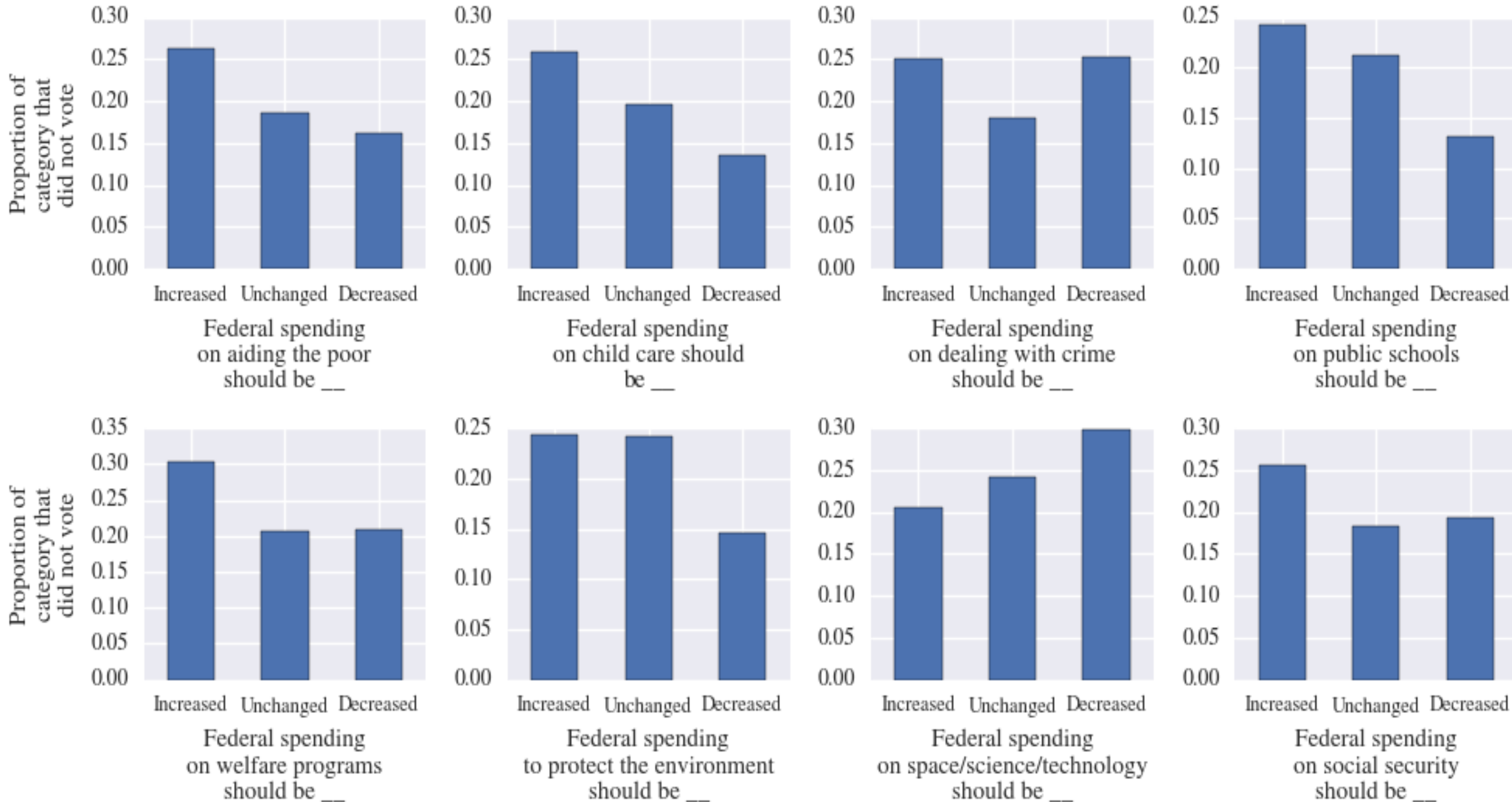
Non-Voter Proportions by Ethnicity



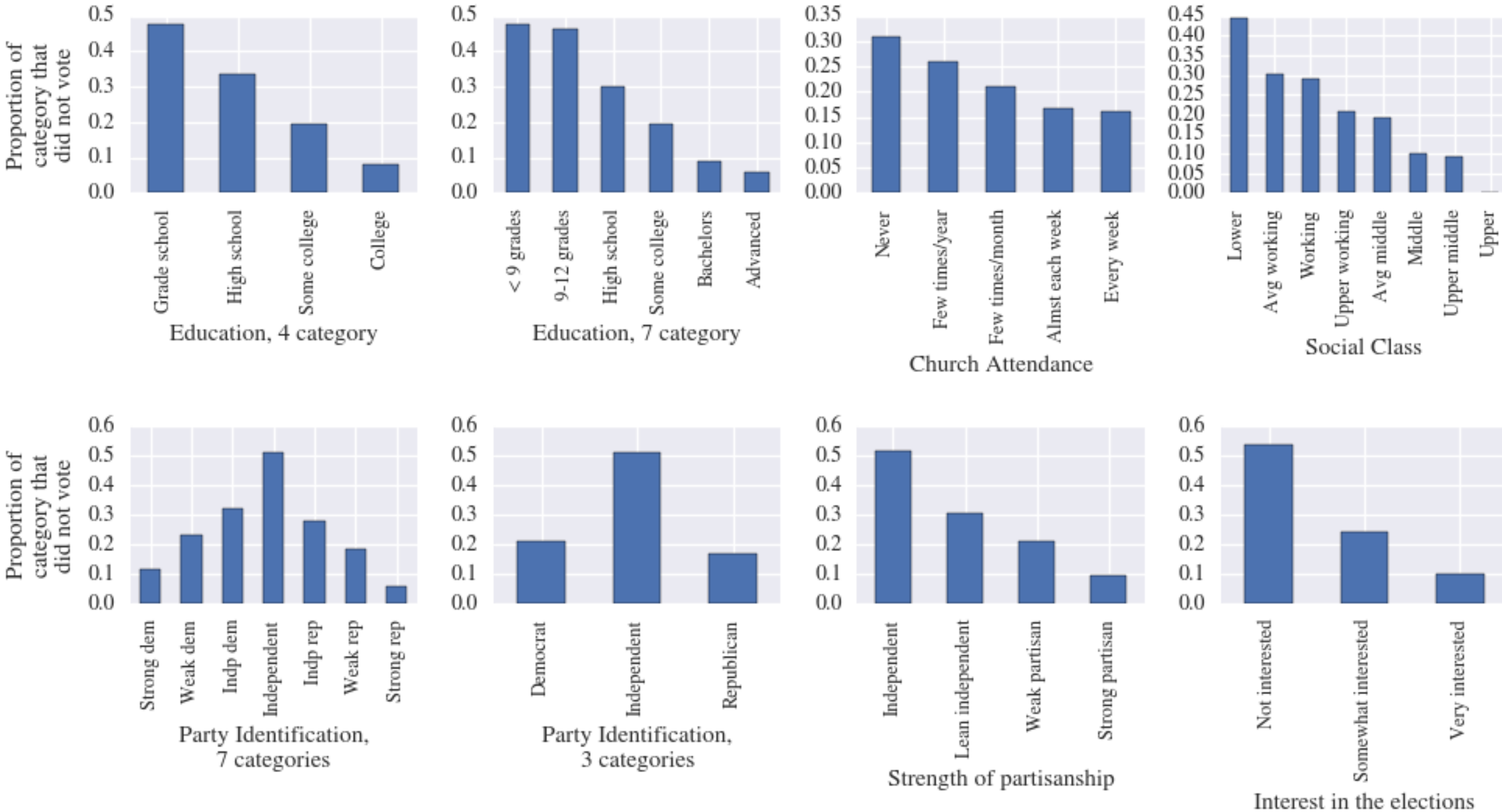


- Whites and blacks have highest vote participation rates
- Hispanics have the lowest despite being third largest racial group

Non-voting hurts Democratic policies



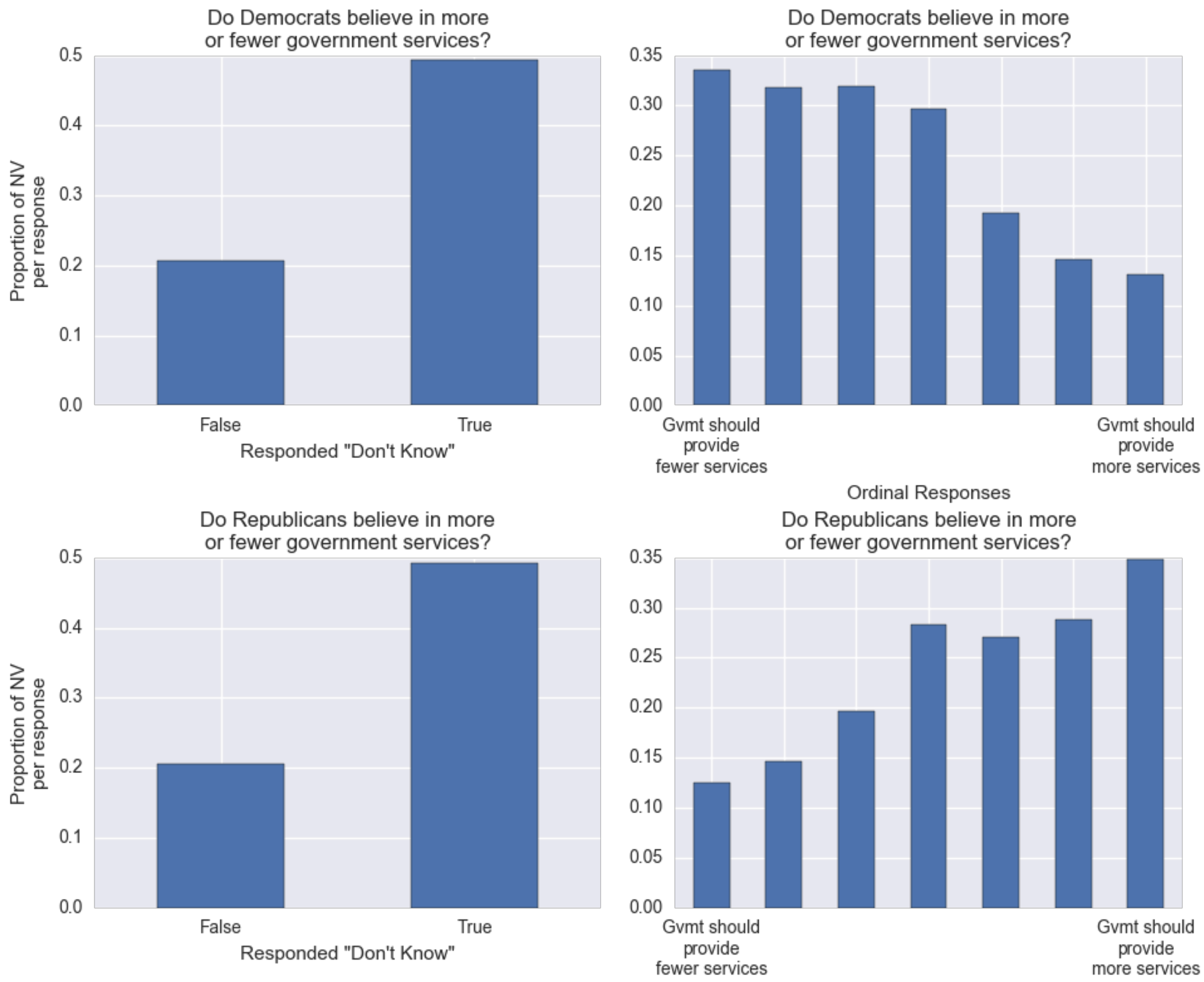
More Patterns



Non-Voters and Political Knowledge

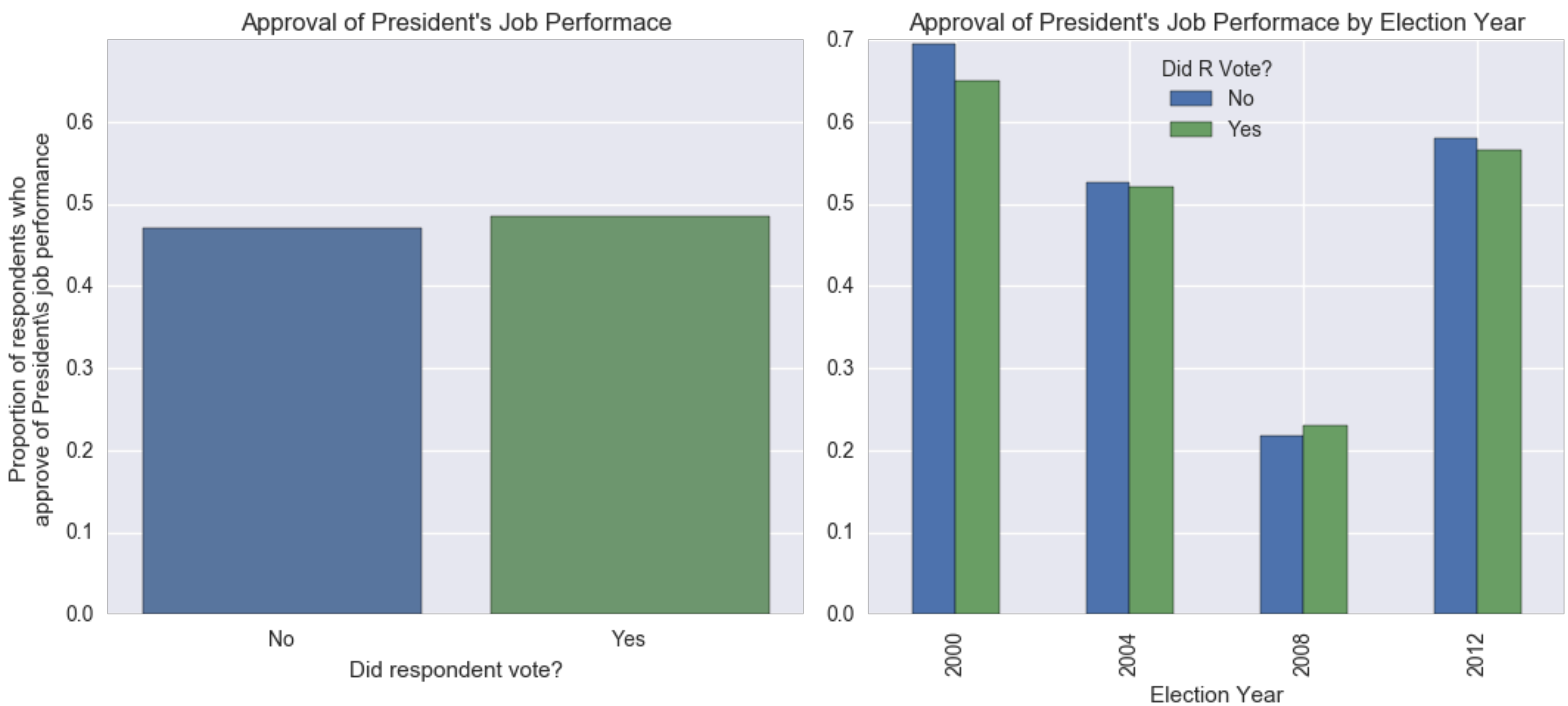
- ANES time series survey includes a battery of questions testing political knowledge
- Do wrong or "don't know" responses indicate a greater likelihood of being a non-voter?

Example

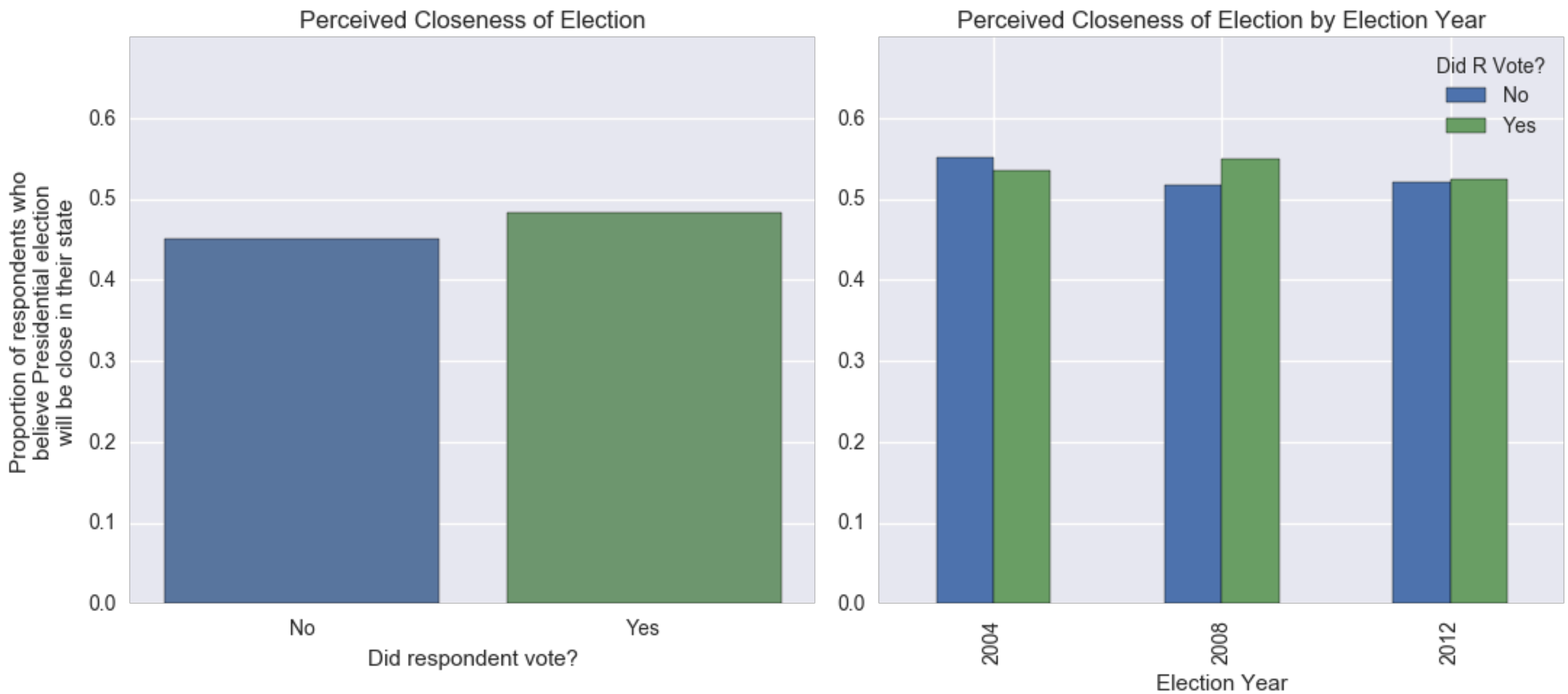

Do Democrats believe in more
or fewer government services?
Proportion of NV
per response
0.5
0.4
0.3
0.2
0.1
0.0
False
True
Responded "Don't Know"
Do Democrats believe in more
or fewer government services?
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Gvmt should
provide
fewer services
Gvmt should
provide
more services
Ordinal Responses
Do Republicans believe in more
or fewer government services?
Proportion of NV
per response
0.5
0.4
0.3
0.2
0.1
0.0
False
True
Responded "Don't Know"
Do Republicans believe in more
or fewer government services?
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Gvmt should
provide
fewer services
Gvmt should
provide
more services

More “Don’t Know” Responses = Probably a Non-Voter

President's Job Performance Equal Across Voters and Non-Voters



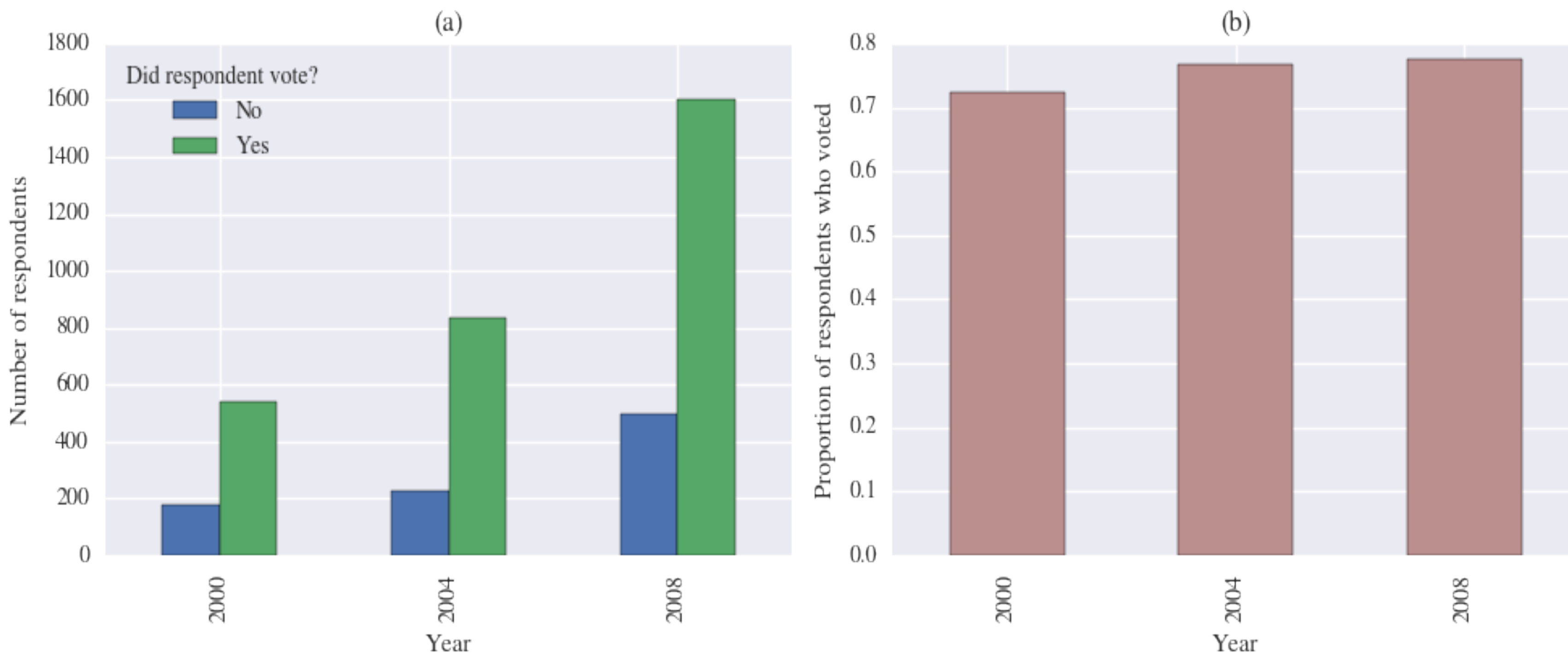
Closeness of Elections Do Not Matter



Non-Voters Are

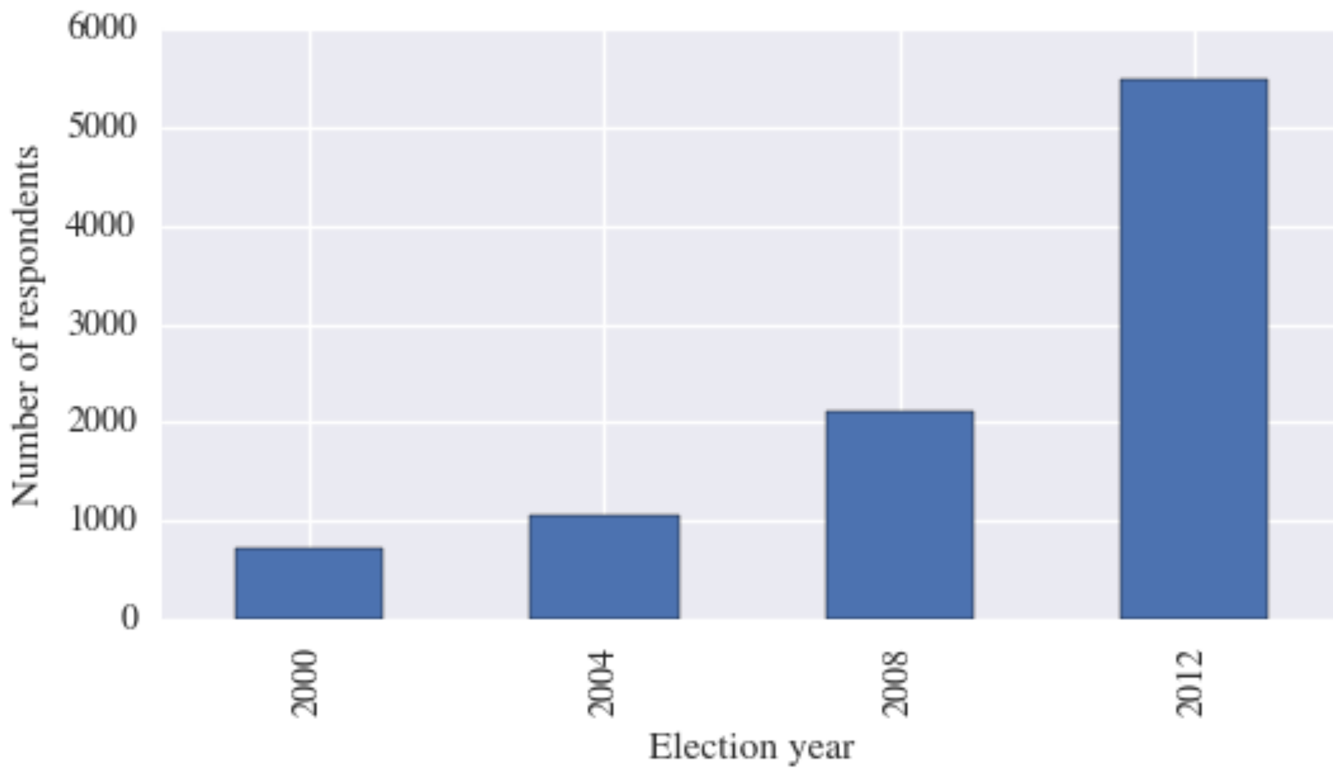
Younger	More apathetic	More Hispanic	More pro-government	More culturally conservative
Less educated	Less religious	Less engaged	Less partisan	Less wealthy

Problem: Unbalanced Data



- Solution: evaluate models using F1-score rather than accuracy

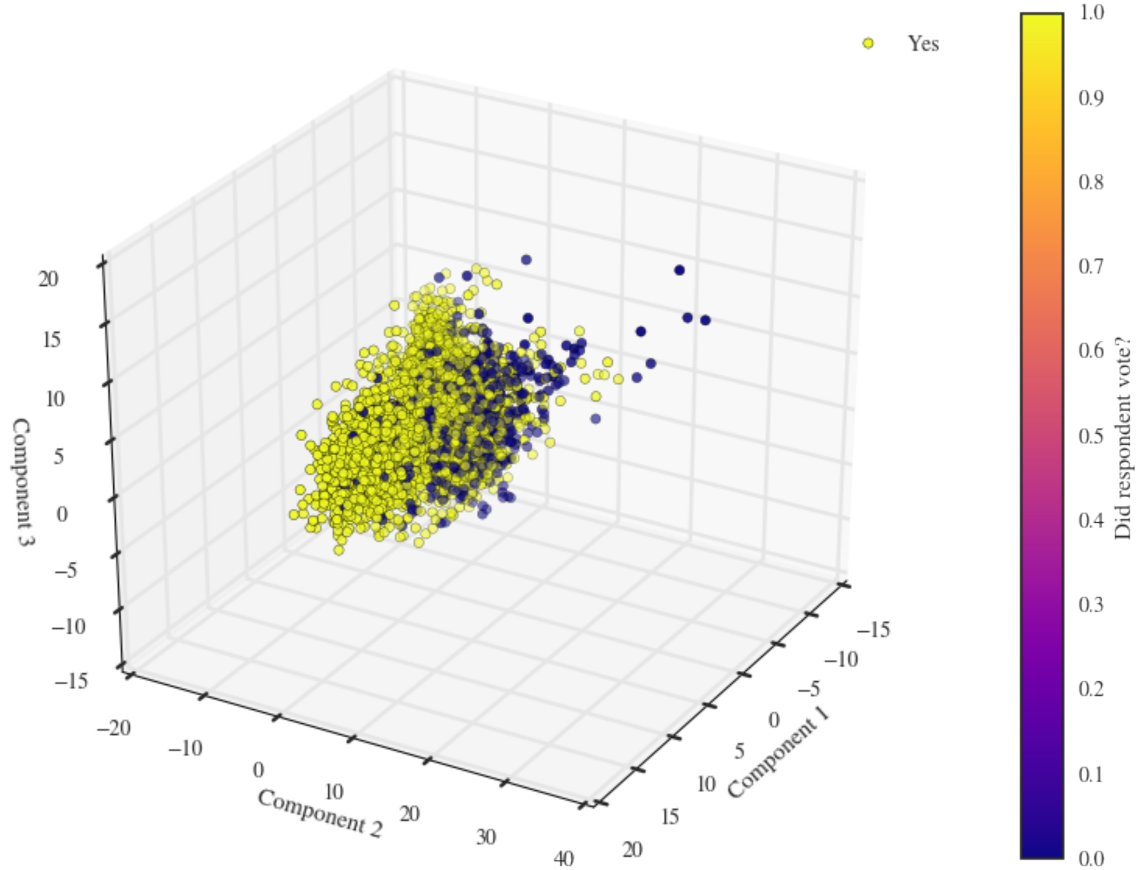
Problem: Unbalanced Data



- Varying amounts of data available for training on different target years
- Solution: automate feature selection with recursive feature elimination

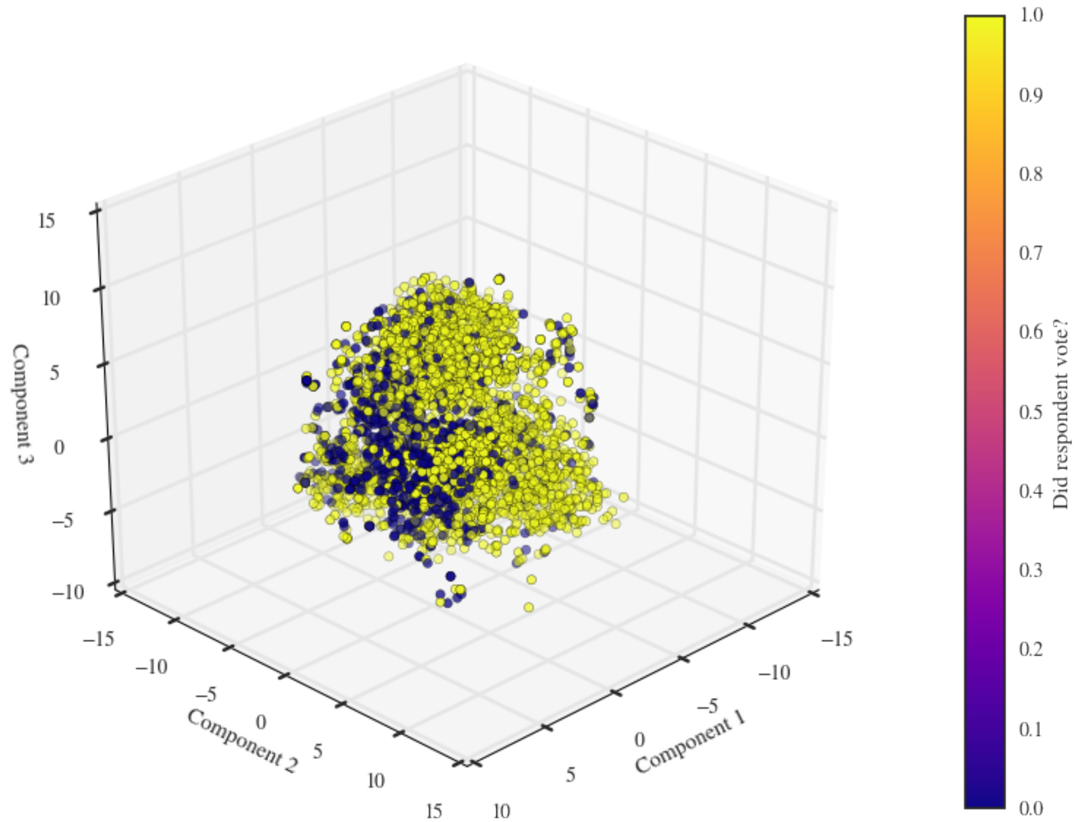
Mixed Signals

Principal component analysis



Mixed Signals

t-Distributed Stochastic Neighbor
Embedding



Training and Testing Models

	Target year of test set								
	2004			2008			2012		
	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.
Logistic Regression	0.612	0.604	0.620	0.614	0.531	0.727	0.595	0.606	0.585
Adaptive Boosting	0.611	0.694	0.546	0.617	0.522	0.699	0.584	0.532	0.565
Bernoulli Naïve Bayes	0.636	0.647	0.624	0.618	0.525	0.749	0.538	0.655	0.456
Support Vector Machine	0.659	0.703	0.620	0.618	0.518	0.764	0.542	0.448	0.687

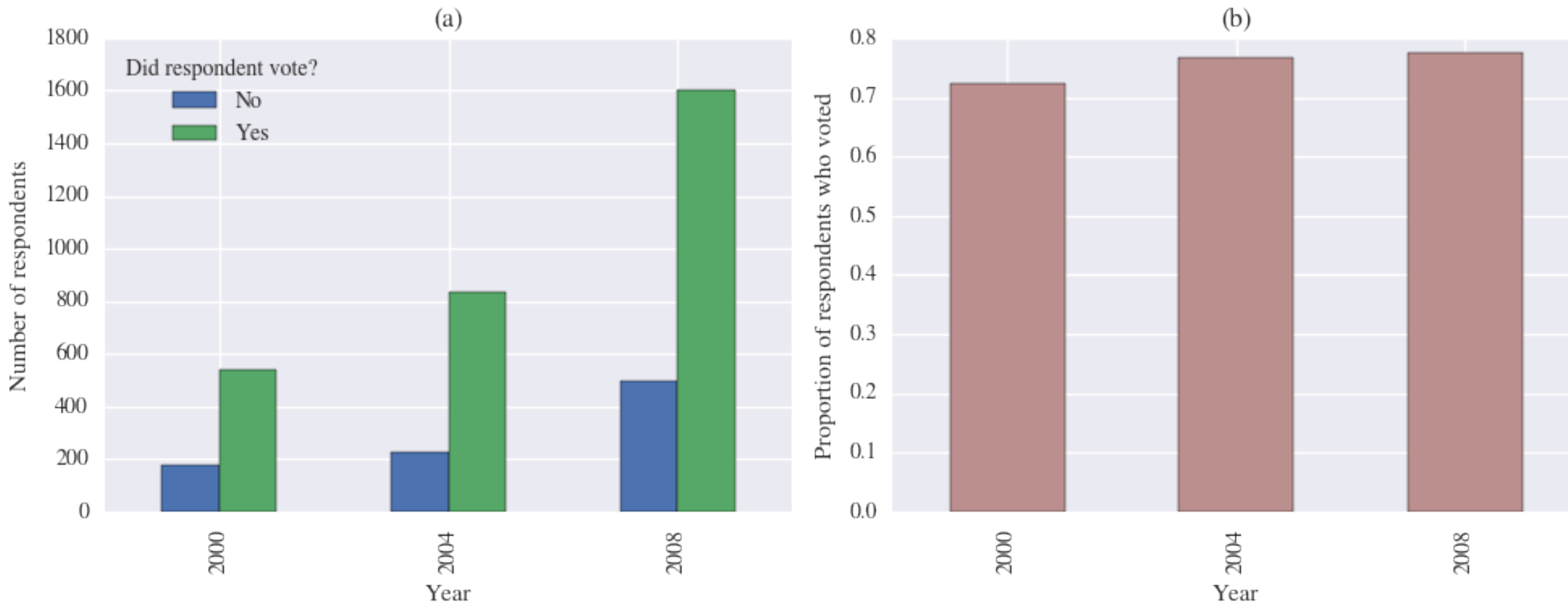
- Cross-validated training scores between 0.7 to 0.75
- No individual model appears to consistently outperform all others

Soft Voting Classifier

	Target year		
	2004	2008	2012
Training F1	0.726	0.765	0.784
Test F1	0.619	0.647	0.611

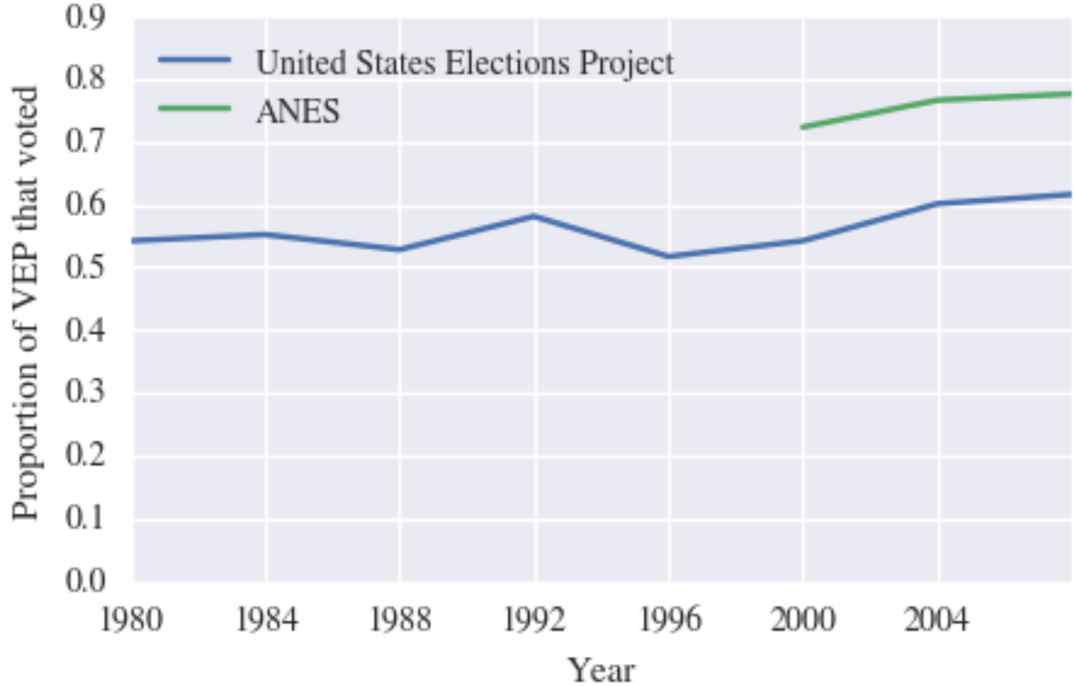
- Leverages the strength of each individual model
- Outputs prediction based on weighted averages of each model's classification probability
- Improves test F1-score for 2008 and 2012

Problem: Social Desirability Bias



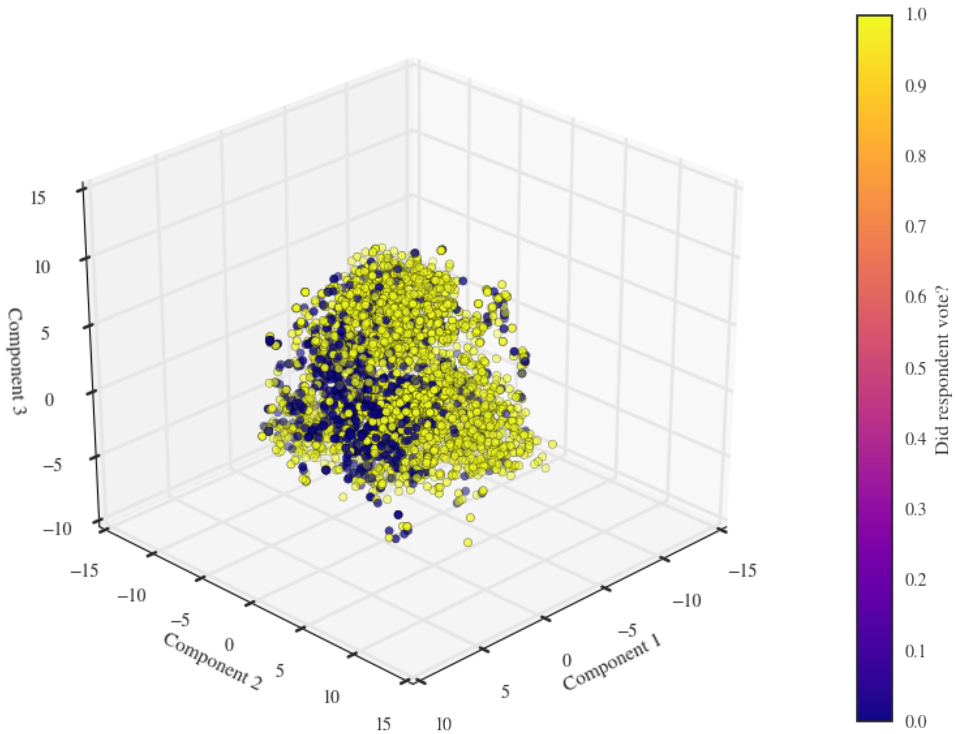
- Self-reported vote participation is un-reliable
- Many respondents claimed to have voted despite not voting

Problem: Social Desirability Bias



- Ground truth is unreliable
- Solution: require recall above threshold

Dealing with Unreliable Respondents



- Rig models to aggressively identify non-voters
- False positives preferred to false negatives
- After training a model, find lowest classification threshold that gives recall > 0.8
- Adjust threshold on test set classification

Calculating Polling Results

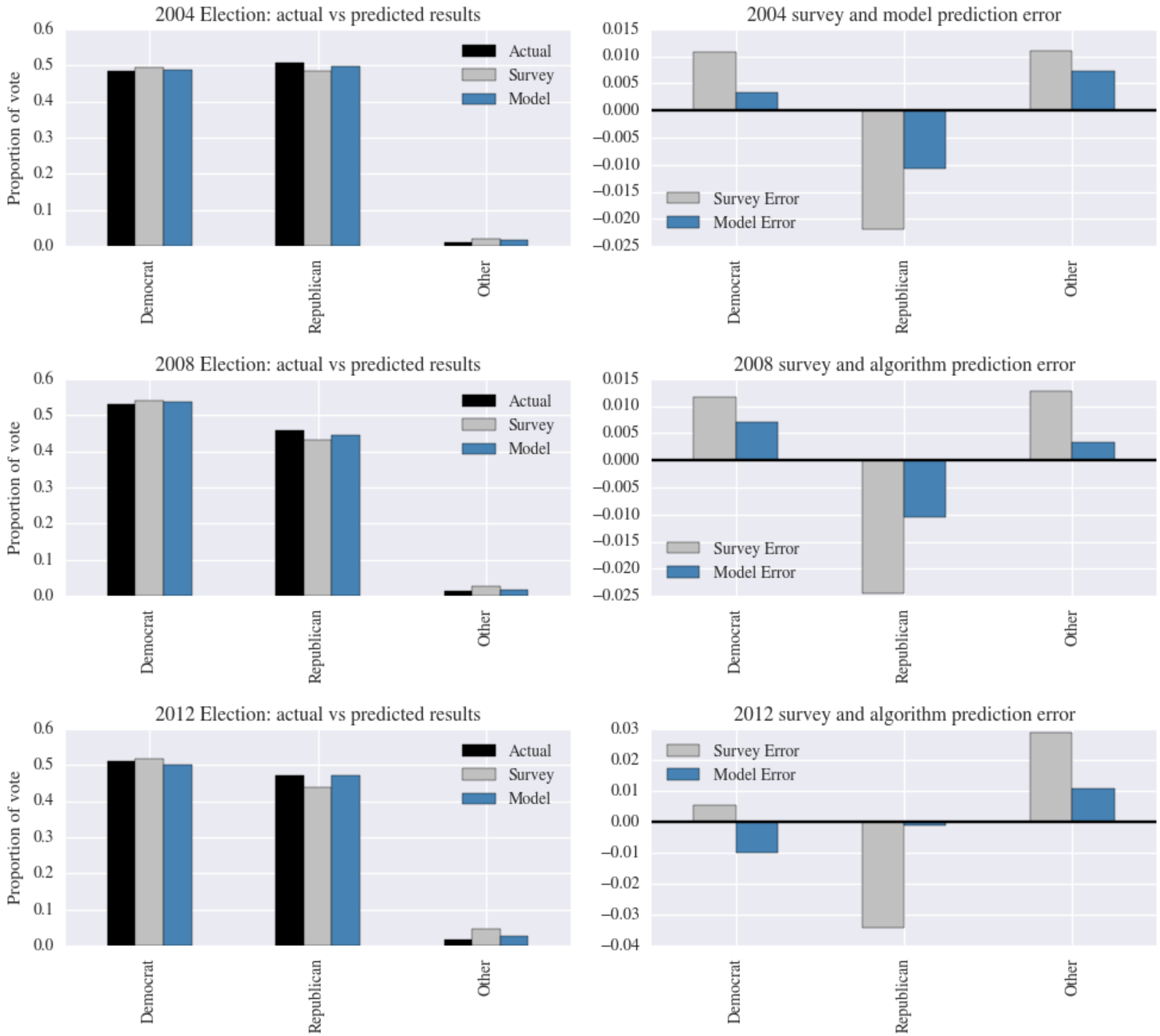
- Without non-voter filter:

$$P_j = \frac{\sum_{i=1}^{m} w_i \delta_{v_i j}}{\sum_{i=1}^{m} w_i (1 - \delta_{v_i n})} \text{ ,where } \delta_{v_i j} = \begin{cases} 0, & \text{if } v_i \neq j. \\ 1, & \text{if } v_i = j. \end{cases}$$

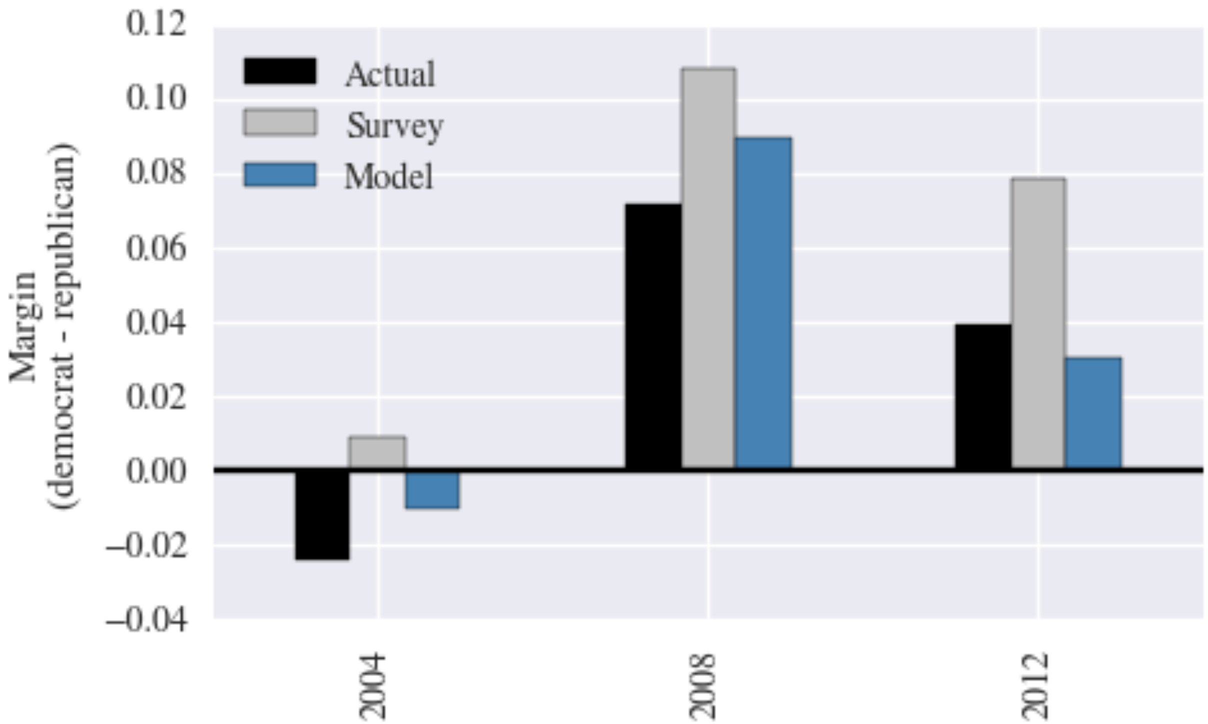
- With non-voter filter:

$$P'_j = \frac{\sum_{i=1}^{m} w_i \delta_{v_i j} (1 - z_i)}{\sum_{i=1}^{m} w_i (1 - \delta_{v_i n})(1 - z_i)} \text{ ,where } \delta_{v_i j} = \begin{cases} 0, & \text{if } v_i \neq j. \\ 1, & \text{if } v_i = j. \end{cases}$$

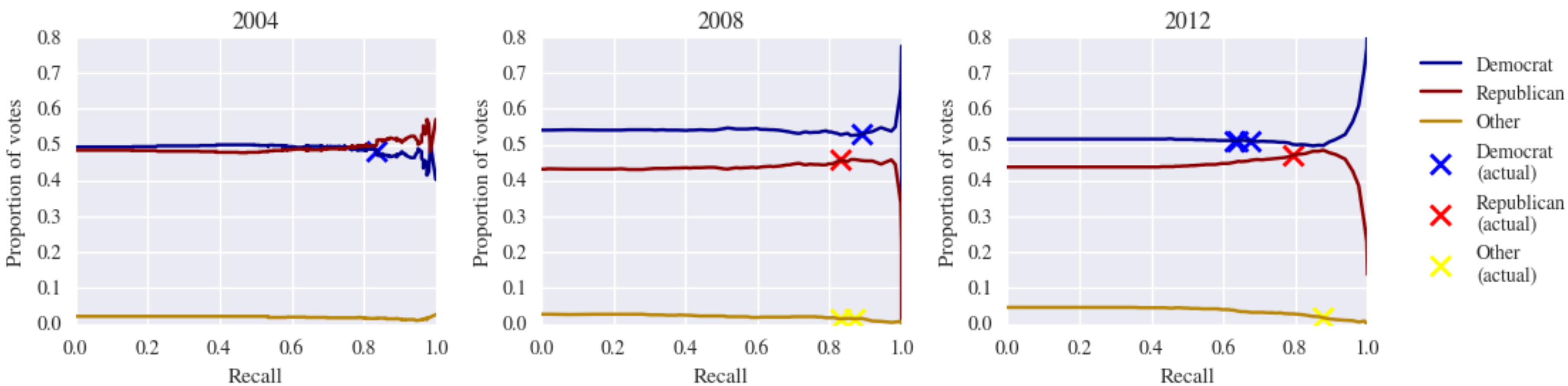
- Where P_j is the proportion of voting respondents who supported candidate j, w_i and v_i in {n, 1, 2, 3} are the sampling weights and vote intentions of the i^{th} respondent, n in v denotes the intention of the respondent not to vote, and z_i in {0,1} denotes the prediction as to whether the i^{th} respondent voted.


2004 Election: actual vs predicted results
Proportion of vote
Actual
Survey
Model
Democrat
Republican
Other
2004 survey and model prediction error
Survey Error
Model Error
2008 Election: actual vs predicted results
2008 survey and algorithm prediction error
2012 Election: actual vs predicted results
2012 survey and algorithm prediction error

Using a Filter Improves Predicted Margins

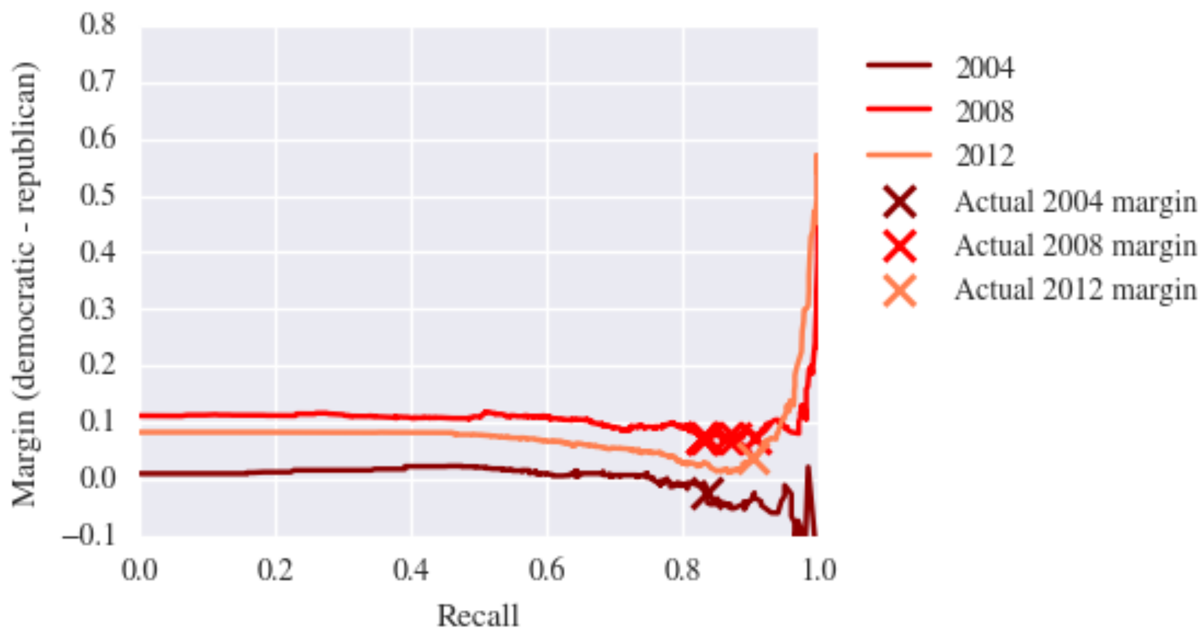


Recall Analysis



- A recall of 0 is equivalent to not using the non-voting filter
- 0.8 appears to be a good recall requirement, but other values within 0.6-0.9 would also adjust polls in correct direction

Recall Analysis



- Vote margin predictions are also robust against different recall requirements

Conclusion

- A soft voting classifier appears to perform best for classifying voters and non-voters from a survey
- Using soft voting classifiers to serve as a non-voting filter improves the accuracy of national polls
- Machine learning may be a viable model for developing likely voter models

Further Study

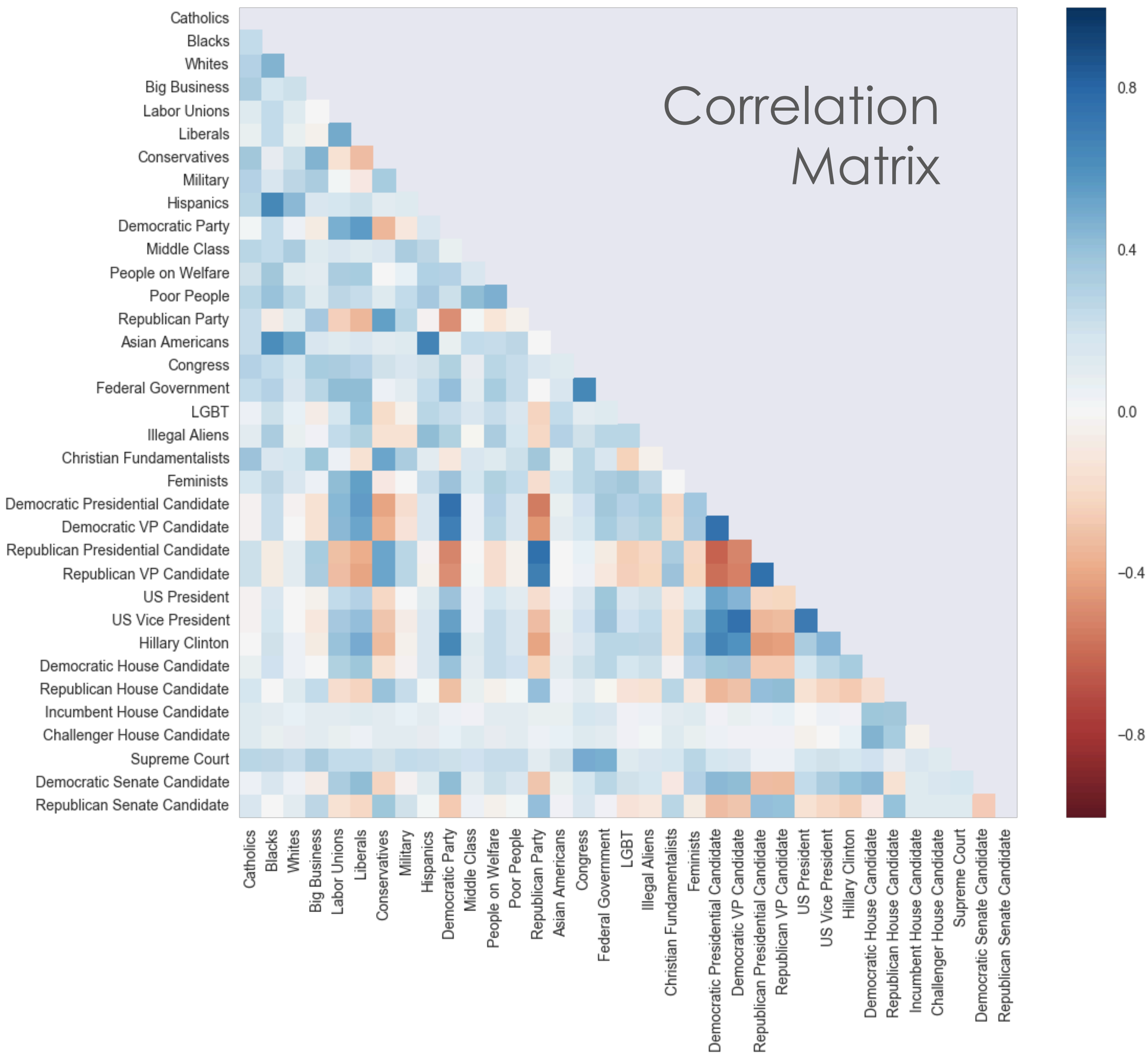
- Compare performance of machine-learning based models against current heuristic models (i.e. Perry-Gallup Index)
- Apply the same technique to recently released 2016 ANES data
- Explore whether this method works on data for which ground truth is reliable (i.e. extending data with state voter registries)
- Perform similar studies for state and local level polling

A Detour: Partisanship in America

How divided are we?

Thermometer features

- ANES surveys contain a set of "thermometer" questions
  - Measure negative/positive sentiment
  - Response is an integer value
    - 0 = negative sentiment
    - 100 = positive sentiment
- Remaining analyses incorporates <u>sampling weights</u> for population inference



Correlation Matrix
Catholics
Blacks
Whites
Big Business
Labor Unions
Liberals
Conservatives
Military
Hispanics
Democratic Party
Middle Class
People on Welfare
Poor People
Republican Party
Asian Americans
Congress
Federal Government
LGBT
Illegal Aliens
Christian Fundamentalists
Feminists
Democratic Presidential Candidate
Democratic VP Candidate
Republican Presidential Candidate
Republican VP Candidate
US President
US Vice President
Hillary Clinton
Democratic House Candidate
Republican House Candidate
Incumbent House Candidate
Challenger House Candidate
Supreme Court
Democratic Senate Candidate
Republican Senate Candidate
0.8
0.4
0.0
−0.4
−0.8

Identity vs. Ideology

Thermometer Pair	Correlation Coefficient
Democratic vs. Republican Presidential Candidate	-0.61752
Democratic vs. Republican VP Candidate	-0.530565
Democratic vs. Republican Party	-0.469287
Liberals vs. Conservatives	-0.306816
Democratic vs. Republican Senate Candidate	-0.254292
LGBT vs. Christian Fundamentalists	-0.230465
Democratic vs. Republican House Candidate	-0.159287
Big Business vs. Labor Unions	0.015196

- Expected anti-correlated pairs only moderately correlated at best
- Correlation weaker for parties and weaker still for ideology
- Down ballot races far less divisive
- Temporary polarization
- Politics is tribalism, not ideology