

# 2014 General Election Predicted Turnout

Christian Riboldi

# The Problem



# Solutions

## Polling

### Pros

- Localization of messages
- When done well they can usually provide very accurate results
- Very well documented and studied procedures

### Cons

- Expensive – Time and Money
- Hidden Variables
- Potentially biased

## Machine Learning

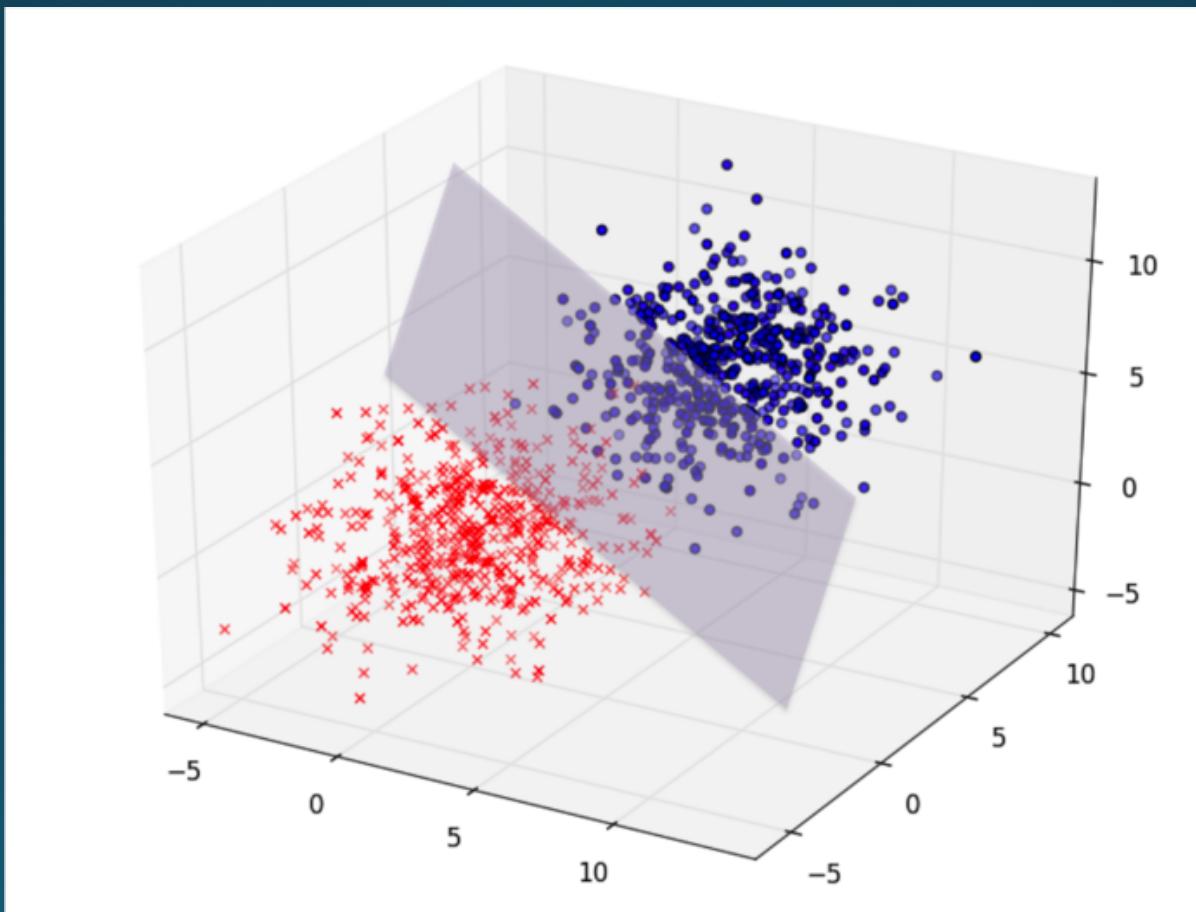
### Pros

- Repeatable
- Unbiased
- Results can improve over time
- Inexpensive

### Cons

- Needs massive amounts of data
- Is a very complex problem
- Takes a lot of tuning
- Can be woefully inaccurate

# Gaussian Process Regression



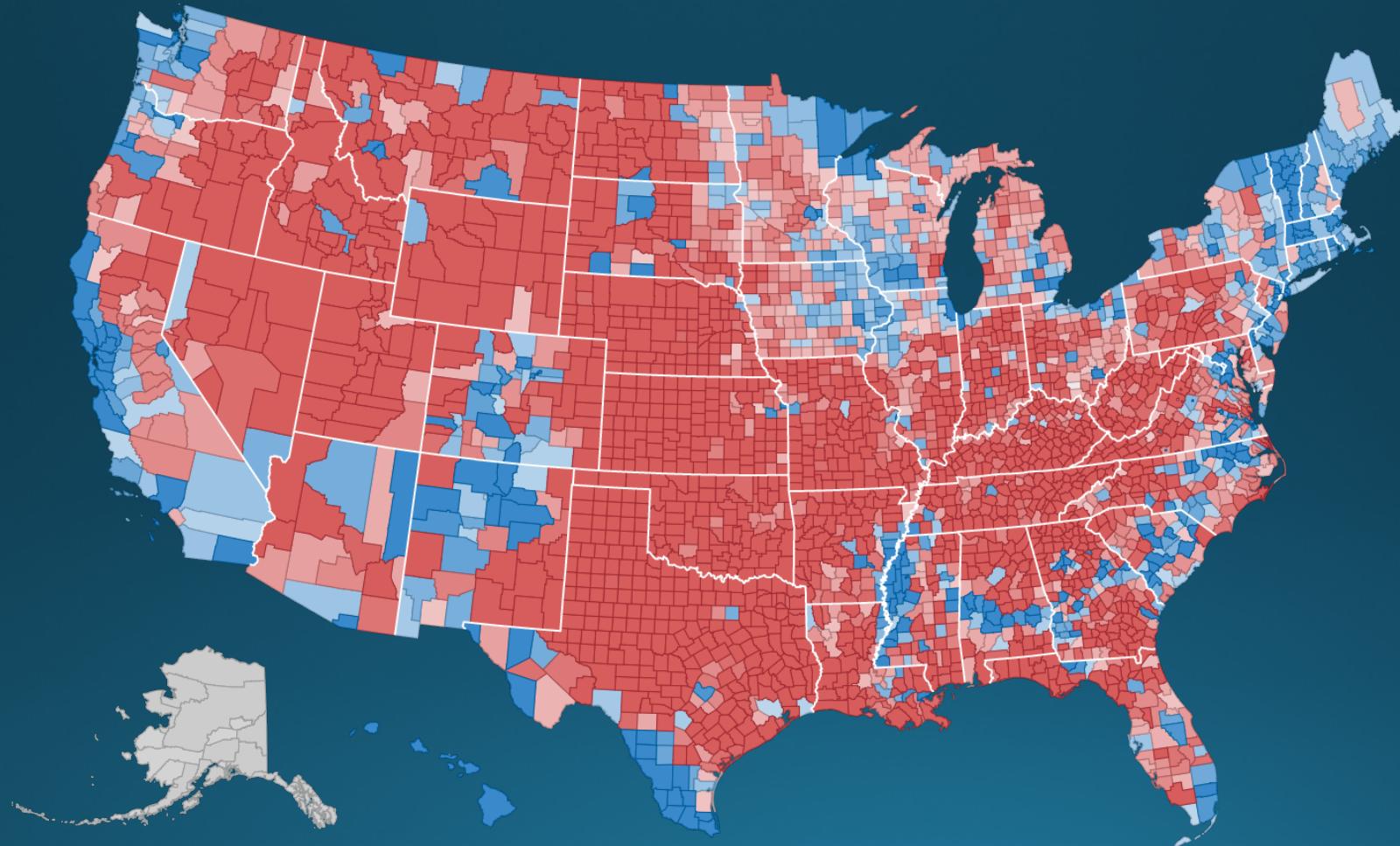
# Probabilistic Classifier

- Create Landmarks using sklearn
- Generate the K matrix for the subset of regressors
- Use the large scale GPR algorithm to generate  $\mu'$  for each data point
- $\mu'$  is a list of probabilities
- Pick a probability threshold to classify the potential voter into a voter or non-voter

$$K = \begin{bmatrix} K_{mm} & K_{mn} \\ K_{nm} & K_{nn} \end{bmatrix}$$

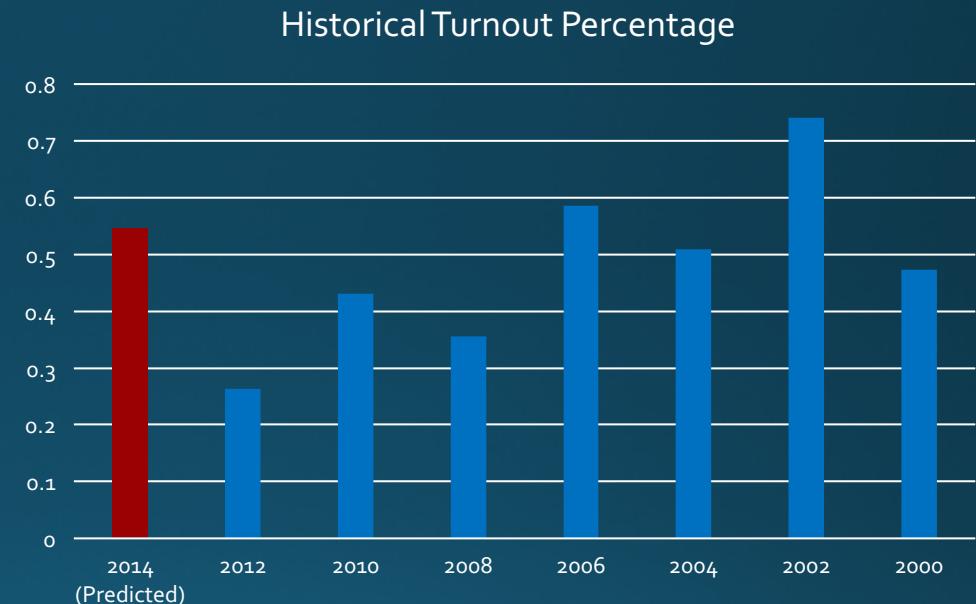
$$\mu' = K(X_m, x_{\text{test}})^T (K_{mn}K_{nm} + \sigma^2 K_{mm})^{-1} K_{mny}$$

# The Results



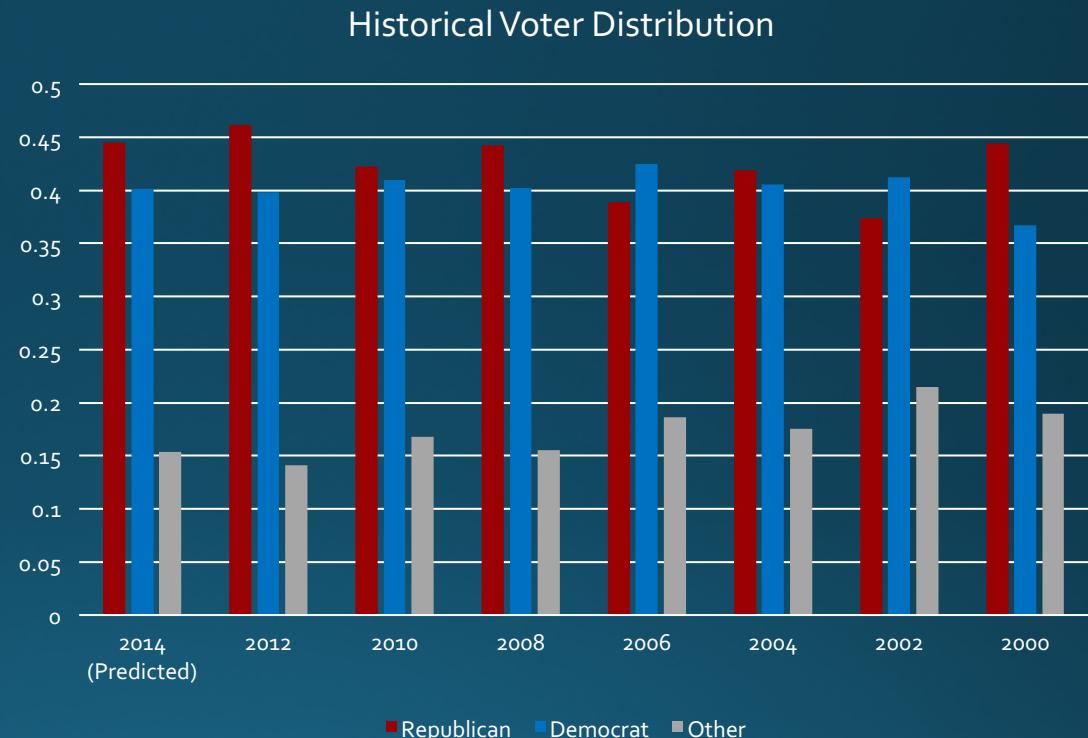
# Historical Context

- Historically this sample has turned had an average turnout of 48%
- Low of 26% in 2012
- High of 74% in 2002
- Our predicted voter turn out for 2014 is 54%

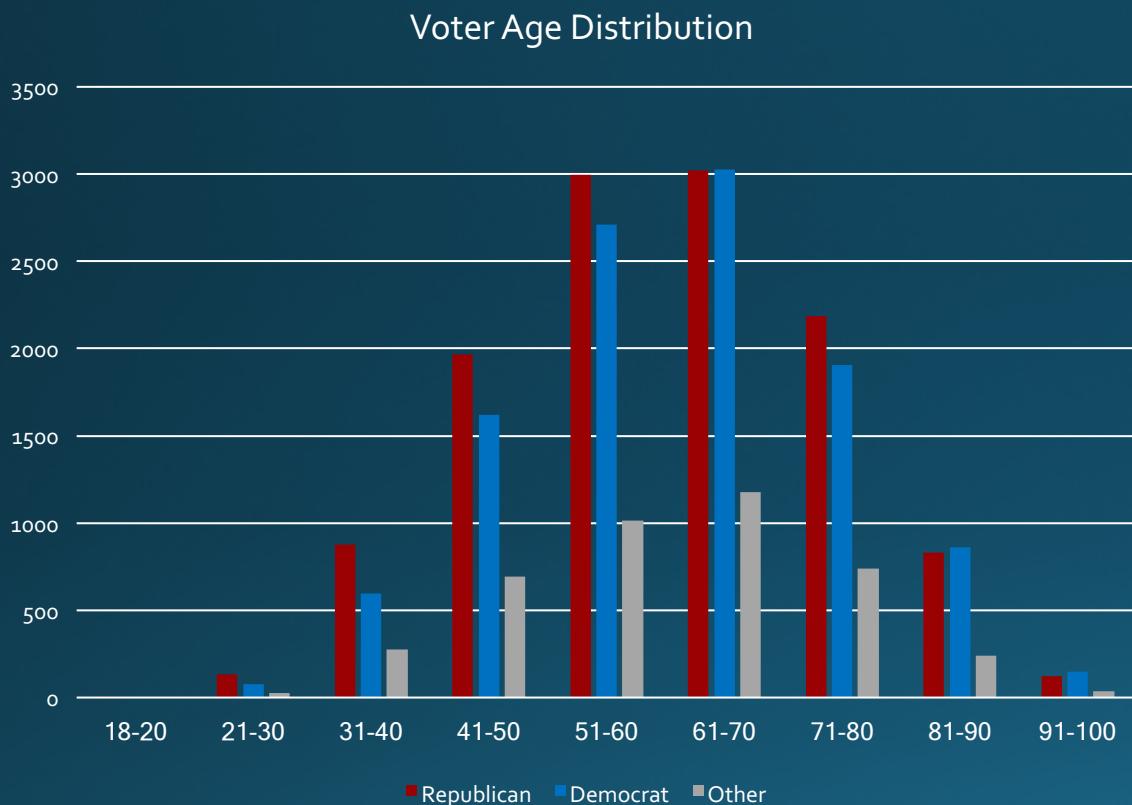


# Historical Distribution

- Our predicted turnout down partisan lines is similar to past general elections
- 44.5% Republican
- 40.1% Democrat
- 15.4% Other



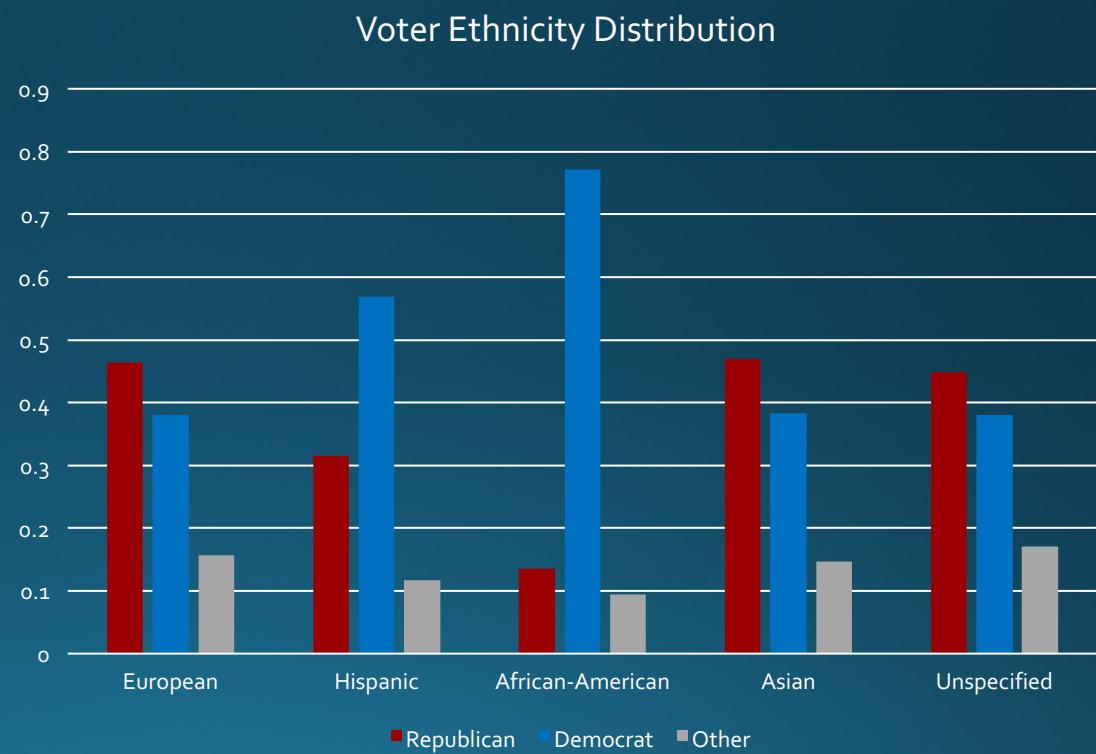
# Predicted Age Distribution



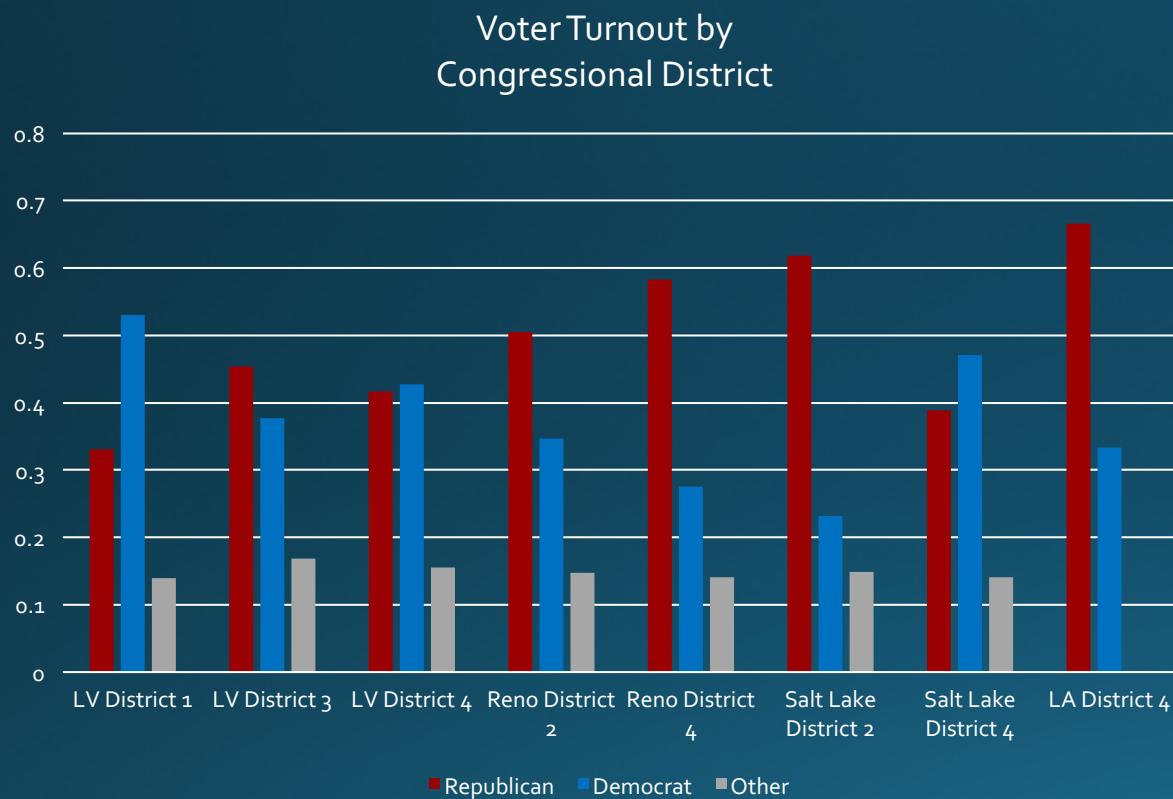
- Age distribution is normally distributed
- Consistent across all age groups

# Predicted Ethnic Distribution

- There is a strong correlation between Hispanic and African-American partisanship
- Other ethnicities lean Republican slightly



# Geographic Predictions



- Mapping this prediction onto corresponding congressional districts we can expect the following results
- Las Vegas Districts 3 and 4 are the most closely contested districts