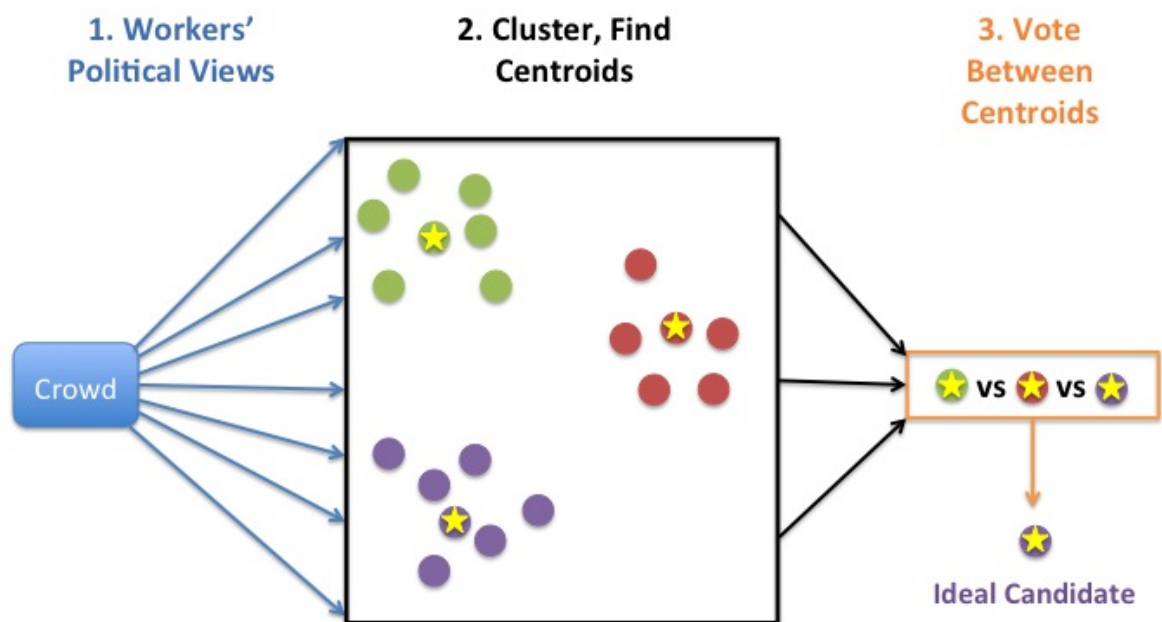


Crowdsourced Candidates

Ben Gitles, Abhishek Gadiraju, Chenyang (Ray) Lei
NETS 213 Final Project

The overall structure of our project is shown in the diagram below. We cluster on past political survey data to find ideal candidates based on the median voter theorem¹, which are the centroids of the survey taker clusters. We then create tasks on Crowdfunder to re-vote on the centroids. Below we provide a picture demo of each of the three main steps: **aggregate survey data, clustering on survey taker and vote on ideal candidate.**



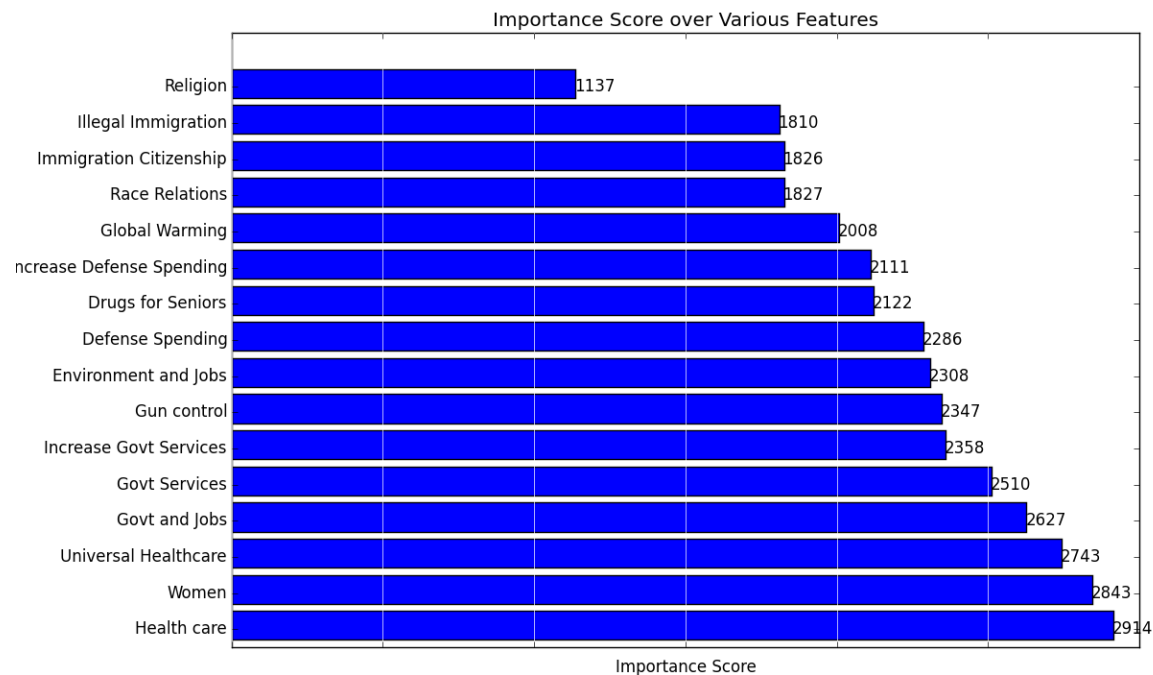
¹ The median voter theorem states that "a majority rule voting system will select the outcome most preferred by the median voter"

Aggregate Survey Data:

The political survey is quite comprehensive and contains more than 60 unique question topics. We first want to know which set of features we should focus on for our clustering model.

Through studying the survey results, we realized that many questions have an importance score chosen by the survey takers. We then aggregated the questions on the importance scores across surveys and ranked the questions with their total importance scores. The graph below shows the final rank.

We chose among the top ones the questions we want to focus on. They are **party, ideology, defense spending, health care, government and jobs and Gun Control**. Even though these are not the top five by importance score, they represent the most important jobs that also cover a wide variety of issues that are in the voter's interest to consider.

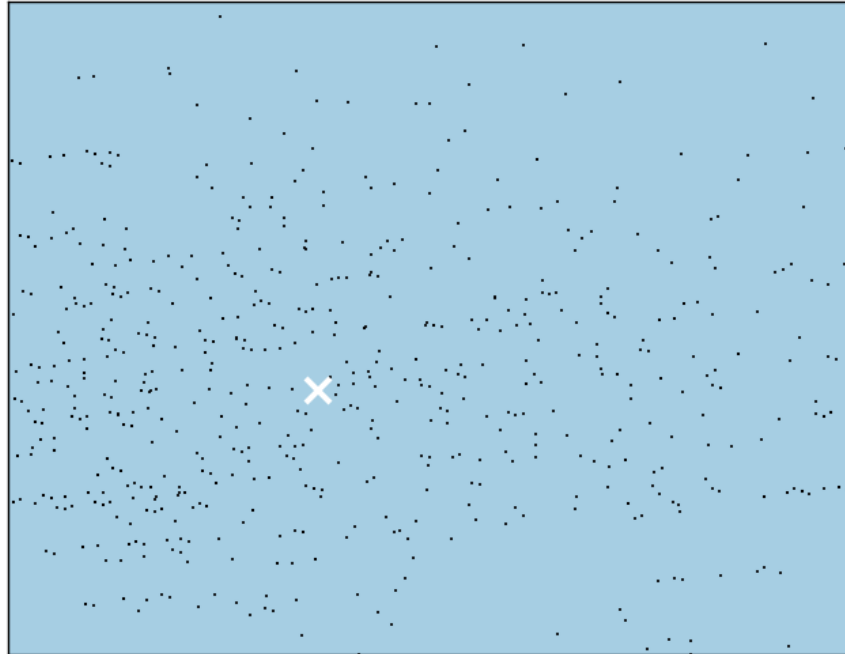


Clustering on Survey Takers:

We then run K-means clustering on the survey takers with the questionnaire results as their feature vector to get the “ideal candidates”.

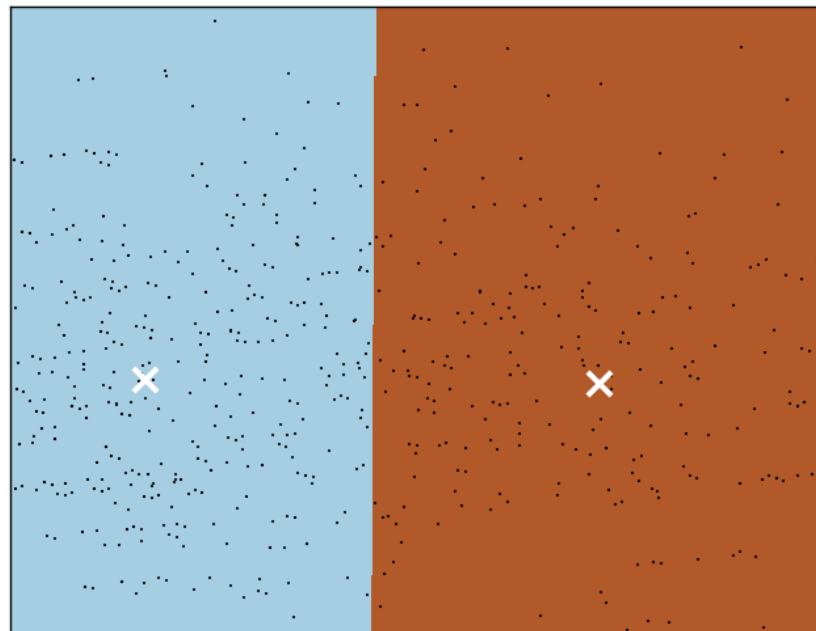
We first ran the clustering with $k=1$, which is the center of mass of the entire dataset:

K-means clustering on the candidate dataset (PCA-reduced data)
Centroids are marked with white cross



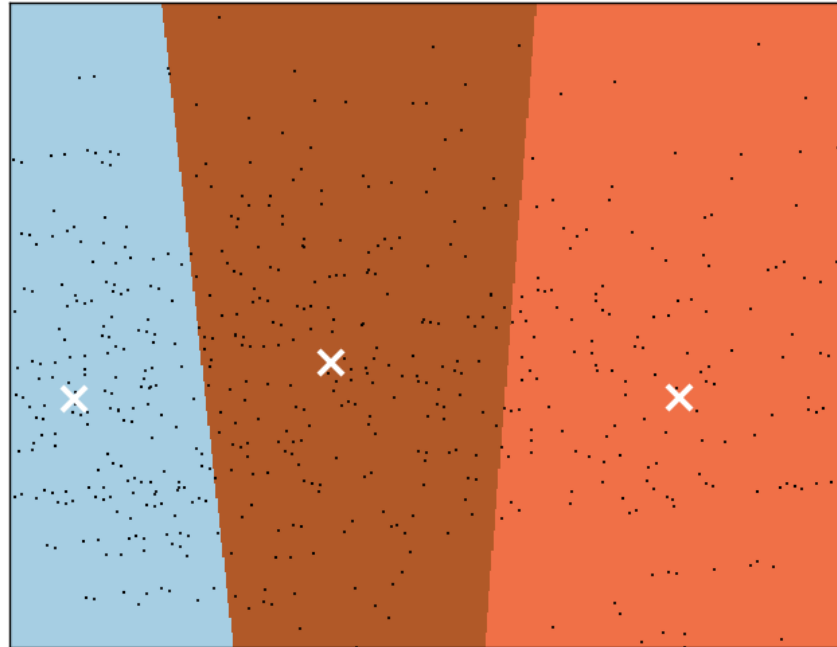
We also ran clustering with $k=2$, which is similar to the US presidential election with two candidates:

K-means clustering on the candidate dataset (PCA-reduced data)
Centroids are marked with white cross



We then ran clustering with $k=3$, to better represent the complete dataset:

K-means clustering on the candidate dataset (PCA-reduced data)
Centroids are marked with white cross



After reviewing the clustering results, we decided to use the results from clustering with $k=3$ as our ideal candidates, because the results from other clusters with different k values overlaps a lot with the $k=3$ results. Notice that the points in the PCA reduced dataset do not seem to have any natural clustering. This indicates that our dataset was more diverse in opinion, which is something that we specifically looked for when choosing our features.

Vote on Ideal Candidate:

With the three ideal candidates we generated in the previous step, we designed our HIT interface to ask the Crowdfunder workers vote on them again with their information provided. The resulting candidate with the highest votes is then our ideal candidate. Below is the screenshot of our HIT interface.

Candidates	Party	Ideology	Defense Spending 1: Govt should decrease defense spending 7: Govt should increase defense spending	Health Care 1: Govt insurance plan 7: Private insurance plan	Government and Jobs 1: Govt should see to jobs and standard of living 7: Govt should let each person get ahead on own	Gun Control 1: More difficult for people to buy a gun 7: Make it easier for people to buy a gun
Candidate A	Other	Slightly liberal	3	4	4	4
Candidate B	Democrat	Liberal	2	2	3	1

Which statement would Candidate A agree with more?

- ☐ The government should provide health care for all citizens.
- ☐ All citizens should purchase private health insurance.
- ☐ Candidate A would be indifferent between the two statements above.

1

Vote for your preferred candidate

- ☐ Candidate A
- ☐ Candidate B

1

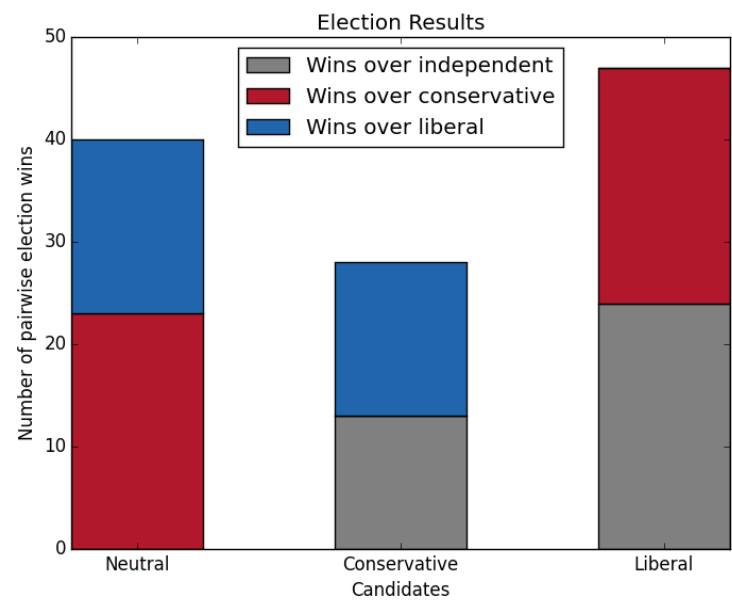
Explain your choice (optional)

What sound does a kitty cat make?

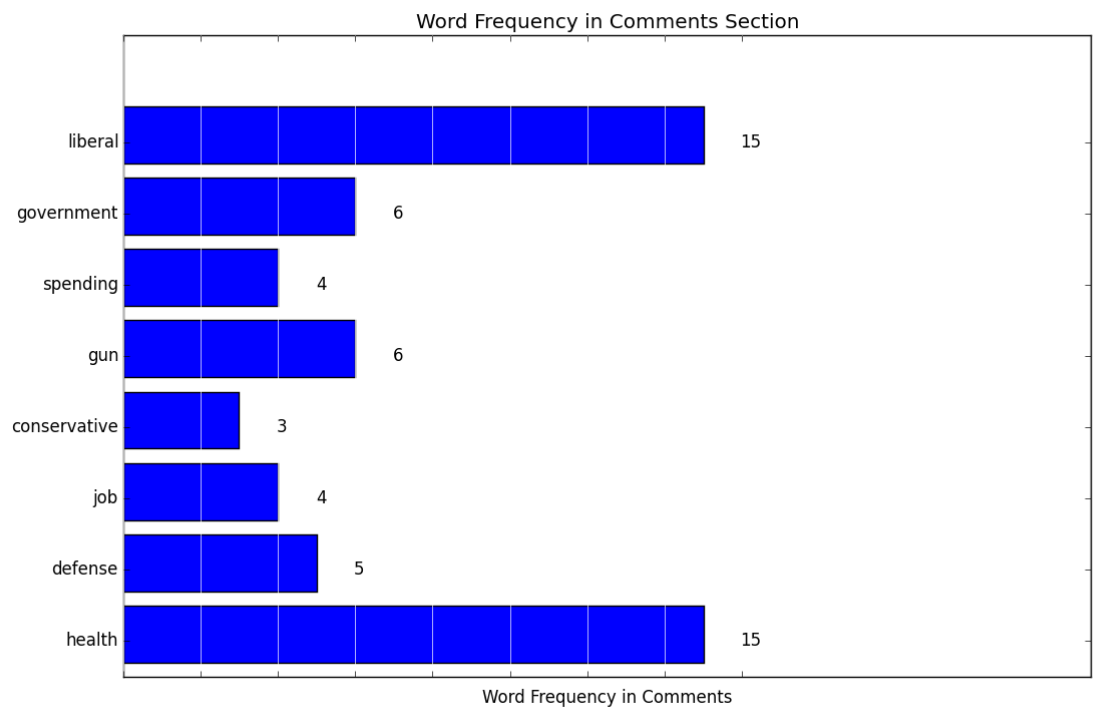
- ☐ bark
- ☐ meow
- ☐ moo
- ☐ roar

1

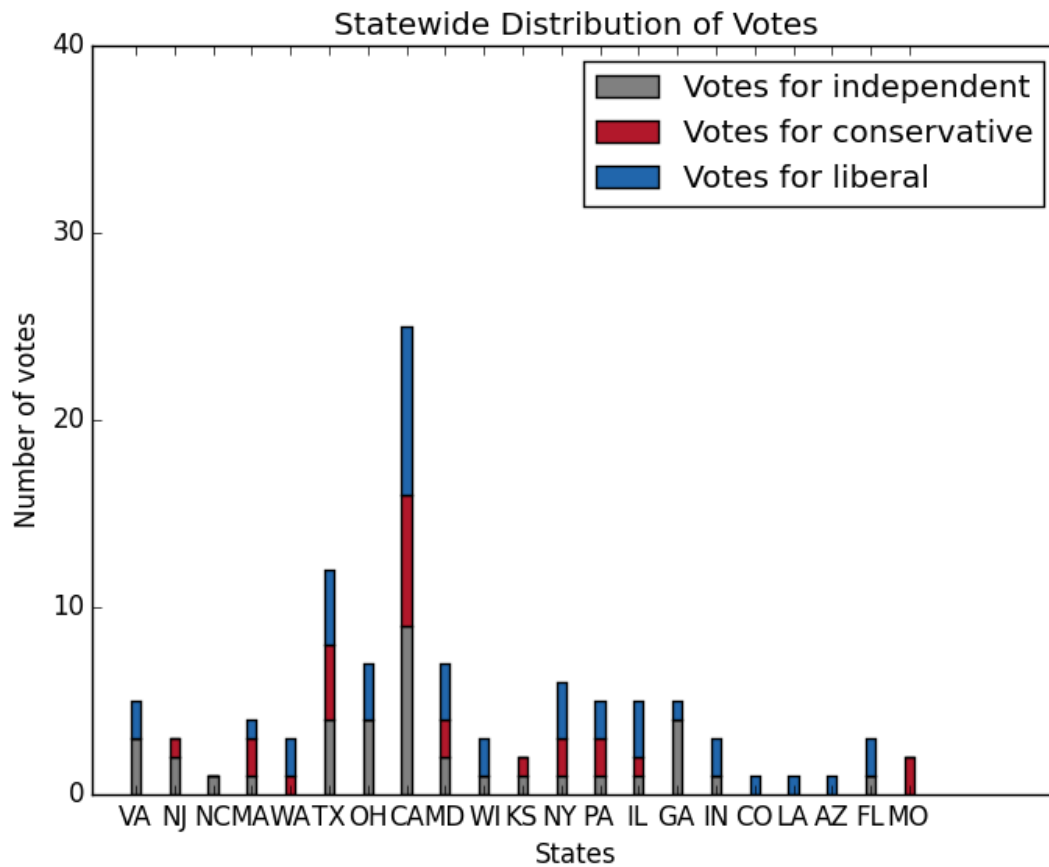
Resulting Graphs



The results of each of the pairwise elections, accumulated



Frequency of words in the optional free response section with the prompt, “Explain your choice”



State from which voters' IP addresses were located. While California clearly has the most, this is not surprising because California has an extremely large population relative to other states.