Will Merrow
11/18/19
INFO 640

Final Project Proposal

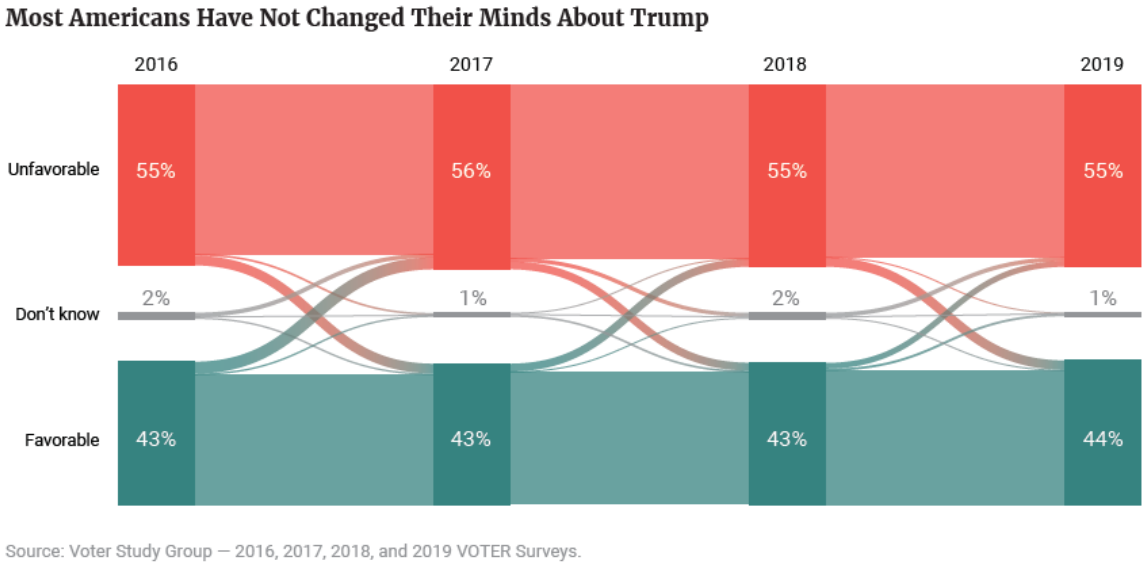
Research Question

My paper will attempt to describe the population of voters who have changed their opinions of Donald Trump since 2016. Using survey data from the Democracy Fund Voter Study Group, I will analyze the demographic and ideological characteristics of these groups, and how they differ from voters who did not change their opinions.

These questions are relevant to politics and public policy, and can help political analysts and observers understand which voter groups are most persuadable and how stable or unstable a president's support is with certain groups. The results could potentially be extrapolated to predict how voters will respond to future presidents, although more research would be needed to determine if the results are generalizable.

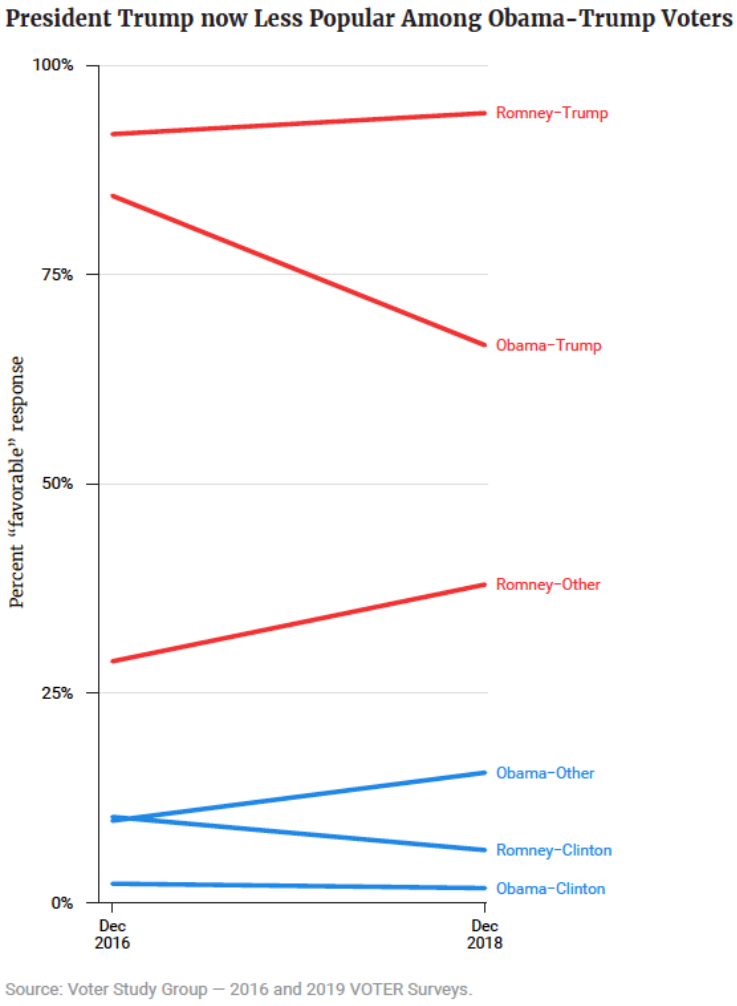
Existing Research

A previous analysis by Robert Griffin from the Voter Study Group analyzed this population using the same data, showing that most voters have not changed their opinions of Trump:



Griffin's analysis also segments the voters according to their votes in 2012 and 2016 (Obama-Clinton, Romney-Trump, Obama-Trump, Romney-Clinton) and finds that

Obama-Trump voters' opinions of Trump have become less favorable, while the other voter groups showed smaller changes that were not statistically significant:



Griffin's publication does not investigate changes in opinions of Trump beyond splitting voters into these six groups, although the charts produced are helpful for exploring the data and showing possible graphical approaches.

Dataset

The Democracy Fund Voter Study Group describes itself as a research collaboration, and conducts surveys of American voters in partnership with YouGov. The organization publishes both research and data. The 2018 data file, which I will use in my paper, contains individuals' responses to a large number of questions about political opinion asked of the same individuals in 2011, 2012, 2016, 2017, and 2018. 4,052 respondents participated in all years. The survey includes detailed demographic and ideological questions, many of which are consistent across survey years, amounting to 1,074 variables in total.

Planned Analysis

I see two possible ways of doing the analysis, and I am currently unsure which is preferable:

A. Group by opinion of Trump and whether it changed, and compare the demographic and ideological characteristics of each of those groups.
B. Group by demographic and ideological characteristics, and compare favorability of Trump and how it changed among the demographic/ideological groups (as in Griffin's line chart above).

Method A:

This would mean grouping respondents based on how their opinions of Trump have changed, and then looking at the demographic and ideological characteristics of these groups. I would divide respondents into the following four groups:

1. **Favorable to unfavorable**: opinion changed from very or somewhat favorable to very or somewhat unfavorable (n = around 500)
2. **Unfavorable to favorable**: opinion changed from very or somewhat unfavorable to very or somewhat favorable (n = around 500)
3. **Remained favorable**: opinion remained very or somewhat favorable
4. **Remained unfavorable**: opinion remained very or somewhat favorable.

I could then use a k-nearest neighbors algorithm to compare the demographic and ideological subgroups within each group.

Method B

Another possible method would be to first use k-nearest neighbors to group all respondents based on demographics and ideology, and then compare how the percent with a favorable opinion has changed over time for each group. This would be the same method used by Griffin to construct the above line chart, but with different groupings of voters.

Other considerations

I plan to analyze changes in opinion from 2016 to 2017 and, separately, from 2017 to 2018. The 2016 survey was conducted in December (after the election, but before Trump's inauguration).

It would be possible to analyze smaller changes in opinion, for example, very favorable to somewhat favorable, or to analyze those who responded to the favorability question with an answer of "don't know". However, my sense is that these groups will be considerably smaller and so focusing on the above groups (all favorable opinions vs. all unfavorable opinions) will be more useful.

For both of these methods, I would likely use the KNN algorithm to divide into demographic and ideological groups. I need to investigate this further, but my plan is to use five demographic indicators and twelve ideological indicators and have the algorithm determine the characteristics of each group, likely aiming for somewhere between five and ten groups.

Demographics:

- Age
- Race
- Gender
- Income
- Education

Ideologies:

- View that politics is a rigged game
- Importance of social security / Medicare
- Attitudes on foreign trade
- Attitudes on gender roles
- Pride in America
- Perception that “people like me” are in decline
- Attitude toward black people
- Feelings toward Muslims
- Attitudes on immigration
- Attitudes on moral issues
- Attitudes on economic inequality
- Attitudes toward government intervention

The ideological indicators are composite indices used by Lee Drutman in a Voter Study Group report that combine responses to individual questions in order to place respondents along a continuous scale from left to right. Recreating these indicators could be a small project in itself and may pose difficulties, so my approach may change but I am hoping to use these twelve indicators since they have already been deemed meaningful.

Challenges and Questions

I imagine a primary challenge will be determining the best possible method to use, as outlined above.

Another challenge will likely be the small sample sizes as the “favorable to unfavorable” and “unfavorable to favorable” groups are each only about 500 observations.

I currently have a few specific questions, which may be helpful to talk through:

1. Is KNN the best approach for dividing voters into demographic/ideological groupings? I imagine that as a fallback I could do a simpler predictive analysis

where I examine correlations with several different demographic/ideological indicators.
2. When should I apply the KNN algorithm? Would it make sense to group all observations this way, or apply it separately to different favorability groups?
3. When should I apply statistical significance tests?