Which defining characteristic gives the most uniform voting outputs? A data story

Serena Yuan

Introduction

The elections in recent years (starting from 2000) have been met with much surprise by voters and policy-makers, especially the election in 2016 where many did not expect Trump to win.

The main question in this data story is: which characteristics of the voter should be used in the future to predict who they vote for in the future?

In other words, in which categories are there more differences in voting preference, and in which categories are there less differences in voting preference?

This can help those running in the future to know which communities to target and which communities based on identity, region, or work status, to rely on for votes. This study considers all political parties that were voted for, and states where each party has an advantage based on categories the voters fall within.

We experiment in answering this question by dividing characteristics into three relevant groups. These groups combine interconnected categories into the same group. The first group is based on identity such as gender and race. The second group is based on regions such as where they grew up and how urbanized the location of their home is. The third group is based on job status, such as what kind of work they do, and if they are worried about finding or losing jobs.

We experiment by faceting by different characteristics. This helps us understand which characteristics of voters lead to voting for various parties, when given a constraint. We experiment with faceting based on different variables to note where there is a majority of votes for one party.

Preprocessing of Data

We process variables for data analysis.

Add the variables as facts:

- "year": year of the record
- "turnout": turnout (voted, registered but did not vote, not registered and did not vote)
- "vote": which party they voted for
- "race": their race
- "gender": their gender
- "language": language the interview was conducted in
- "interview": mode of interview (telephone, internet, etc.)

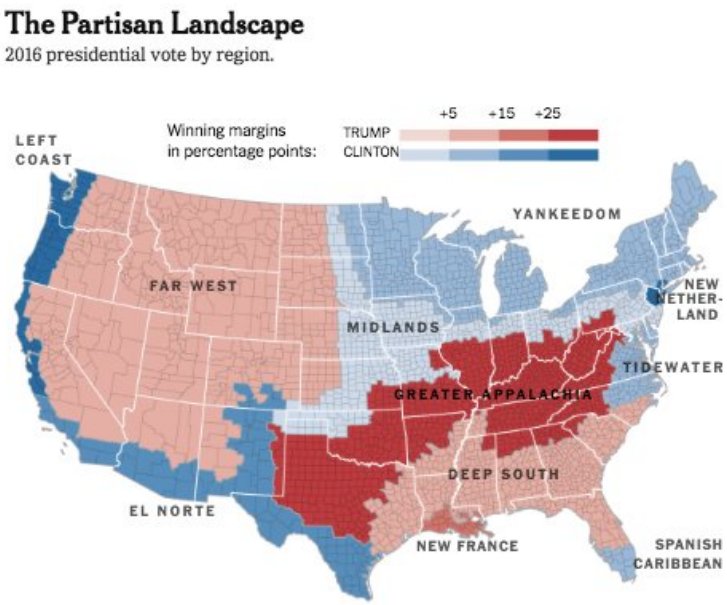


Figure 1: 2016 political landscape divided into regions

- "urbanism": the urbanism of where they are located (central, suburban, rural)
- "region": census region (northeast, north central, south, west)
- "work": work status (working, temp laid off, unemployed, retired, disabled, homemaker, student)
- "grewup" : type of community where grew up (farm, country not farm, small city, medium city, large city, suburb of large city, very large city, suburb of very large city)
- "job": worried about losing/finding job in the future (A lot, somewhat, not much at all)
- "outofwork": laid off / out of work in past 6 months? (Yes, no)

We can compare the results by grouping together characteristics:

1. race, language, mode of interview, gender
2. region, urbanism, grewup
3. work, job, outofwork

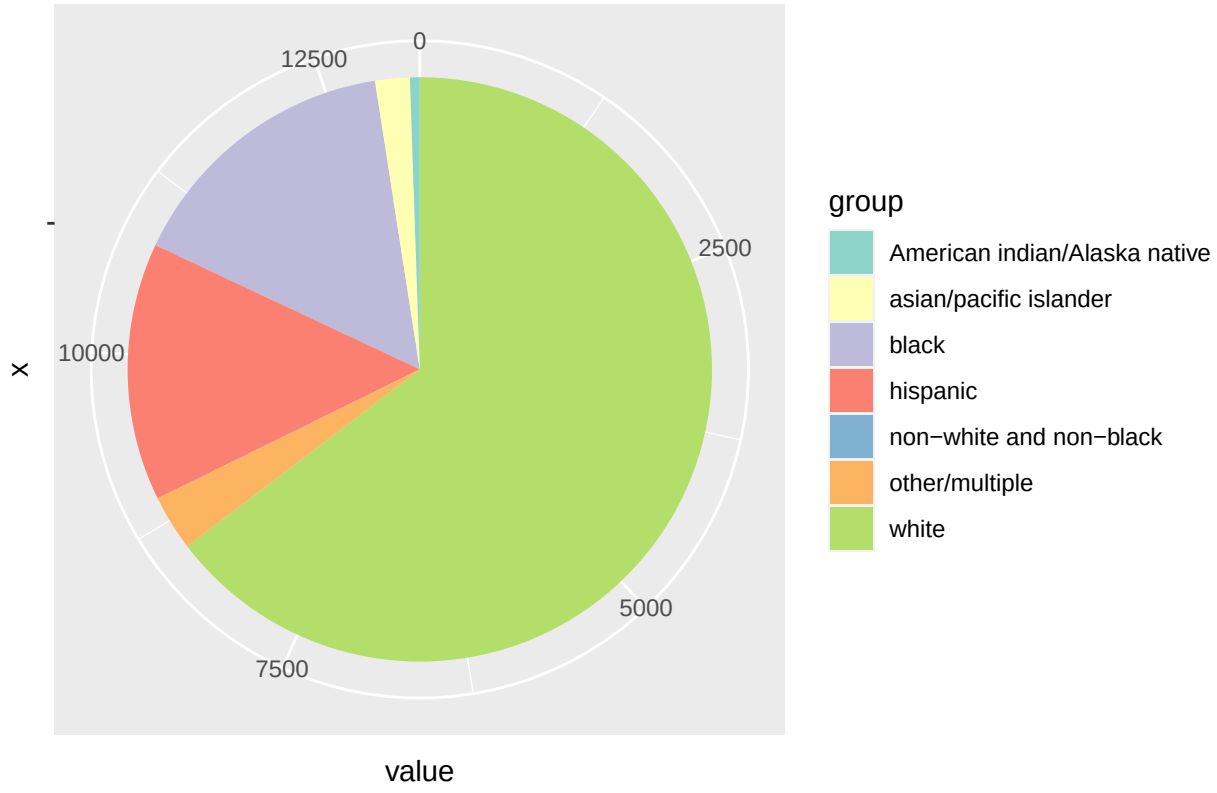
So we group the data into different dataframes based on groups that combine relevant characteristics:

1. select(year, vote, race, language, interview, gender, turnout)
2. select(year, vote, region, urbanism, grewup, turnout)
3. select(year, vote, work, job, outofwork, turnout)

Visualize proportions with respect to defining characteristics

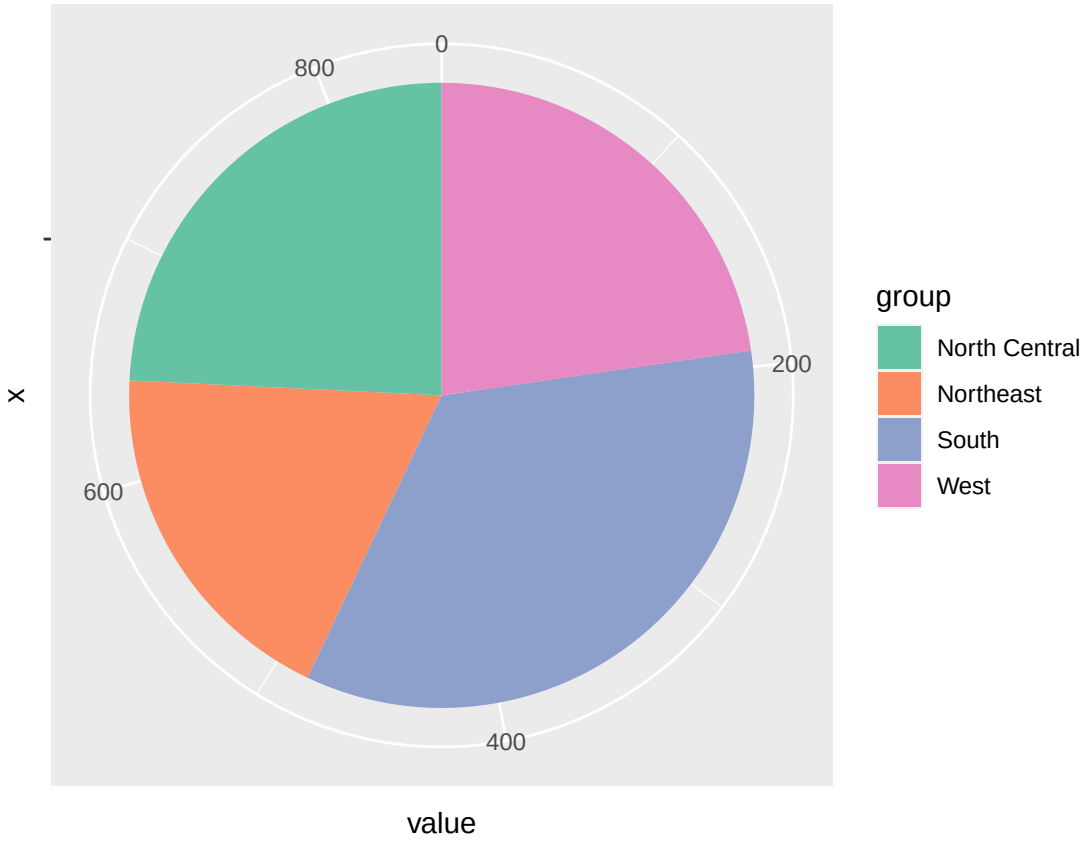
In this section, we visualize proportions with pie plots, based on the defining characteristics taken from each group: race, region, and work status.

We can count the number of rows for each race to further understand not only the proportions within each category but the proportions across categories.

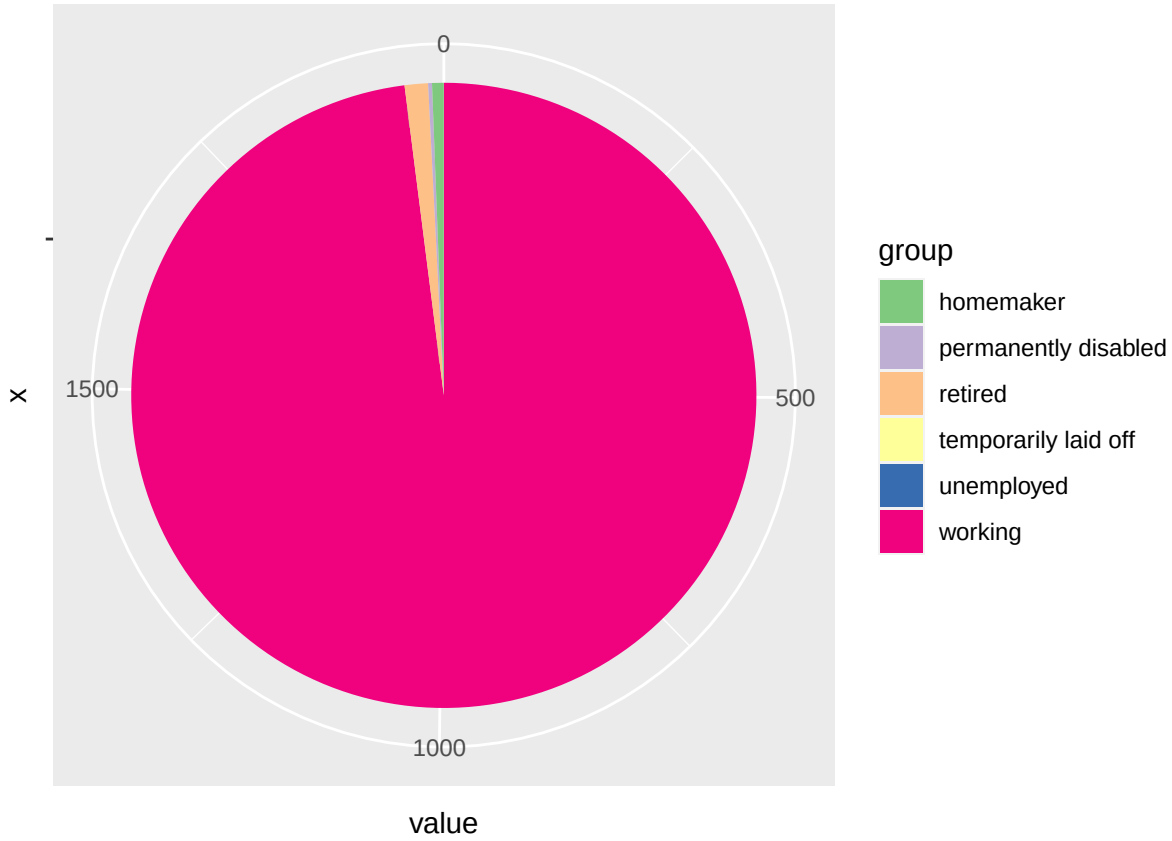


For race, white non-hispanic consists of a very large majority of people sampled. This is followed by black non-hispanic, hispanic, other / multiple races non-hispanic, asian or pacific islander, American indian or Alaska native, and non-white and non-black.

Region:



From above, we see that the regions where the survey participants came from are distributed quite evenly, but slightly more from the South.



Work:

From the pie chart above, we see that the majority of survey respondents recorded from 2000 to 2016 were working, followed by temporarily laid off and homemaker.

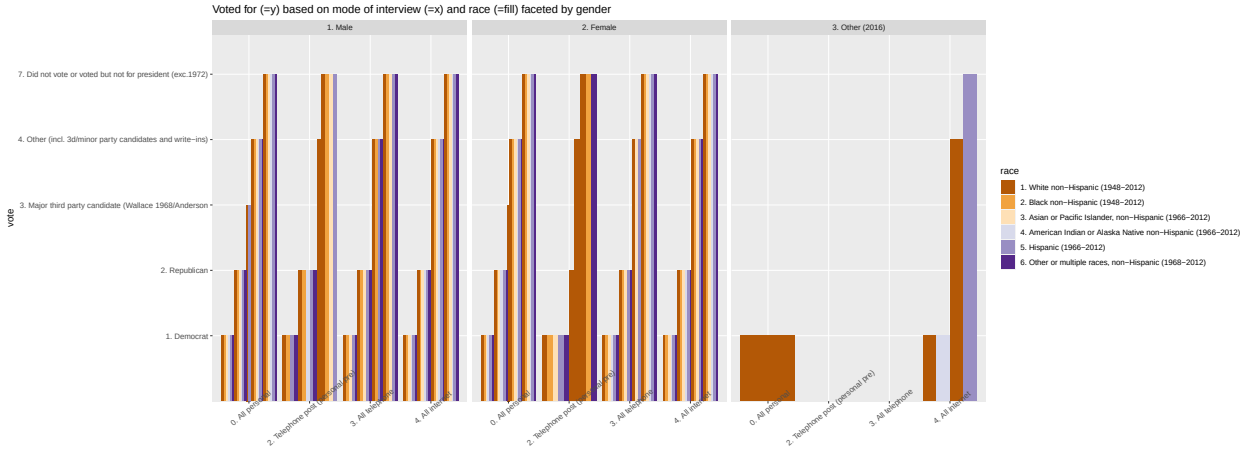
Main Results and Analyses

In the following analyses, we experiment changing the variable that is faceted by. For each group, we produce two plots. The first plot illustrates the independent variable as who they voted for (vote), and the second plot facets by who they voted for (vote). The two plots let us visualize unique conclusions as a result of varying the variables.

In the plots of group 1, we always use race as fill. In the plots of group 2, we always use region as fill. In the plots of group 3, we always use work status as fill.

1. Identity, language, and mode of interview

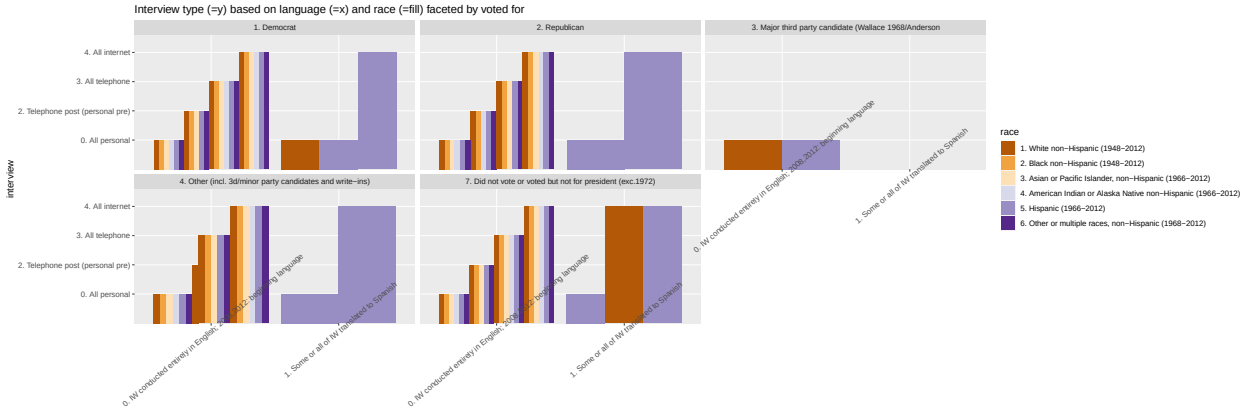
Consider the votes based on racial and sexual identity, language of interview, and mode of interview.



From above, there is greatest uniformity for other gender of white non-hispanic race, where 100 percent in the "all personal" interview category with these characteristics voted for democrat. For the other gender, there were also participants in the "all internet" interview style who primarily voted for democrat and other party (and did not vote).

For male and female genders, it seems that all races voted for different parties and it is difficult to find a party that was much more popular than other parties.

We can plot it again, faceting by turnout:



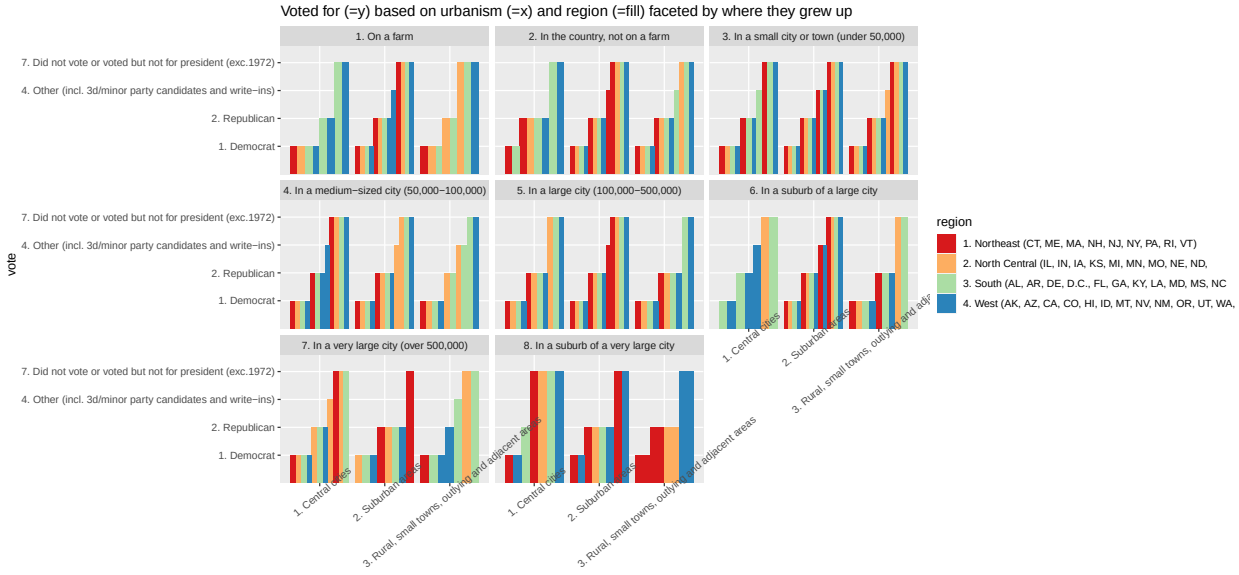
We facet by voted for party in the above plot. We see that for major third party candidate, the voters are either white or hispanic, and they conducted this interview in English.

Those who conducted the interview in Spanish voted uniformly for Republicans, and they happened to be all hispanic.

For democrats and republicans and language of interview as English, there seem to be a variety of modes of interview and all the races who voted for them, and it is difficult to find out which race or mode of interview contributed most to the votes towards democrats and republicans. However, for democrats and republicans, when the language was in Spanish, there is more similarity in the race of the voters. For democrats, the race are either hispanic or white non-hispanic. For republicans in the Spanish language category, the race was all hispanic. The mode of interviews for Spanish were also either all personal or all internet, for democrats and republicans.

2. Regions and locations

Consider the votes based on regions, urbananism or urban quality, and where the voters grew up.



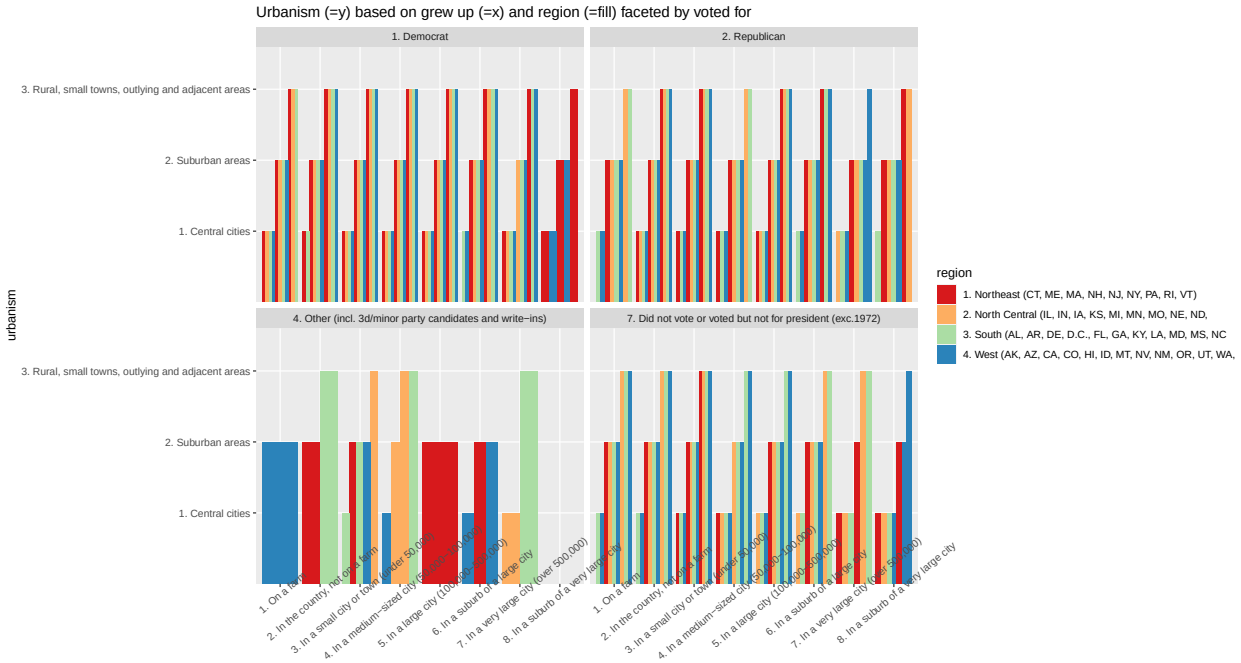
From the plot above, we see that there seems to be an even proportion based on where the voter grew up and based on their region towards which party they voted for where the y variable = vote, x variable = urbanism, fill = region, and it is faceted by where they grew up.

For those who grew up in a very large city (facet 7), those who still live in cities in the Northeast will vote for democrats (or else not vote). For those who grew up in a very large city, those who now live in suburban areas in the Northeast will vote for republicans (or else not vote).

For those who grew up in a suburb of a very large city (facet 8), those in the South will uniformly vote for republicans (or else not vote). In the same category of where they grew up, those in the North Central regions also vote for republicans (or else not vote).

For those who grew up on a farm (facet 1) and are in the South, they will vote for democrat or republican (or not vote). The people who grew up on a farm rarely vote for third party candidates, except for a small proportion living in the suburban areas in the West.

It looks like there is much variation in preferences for vote of people who grew up in small, medium and large cities than in farms, very large cities, and suburbs of very large cities (more y levels).



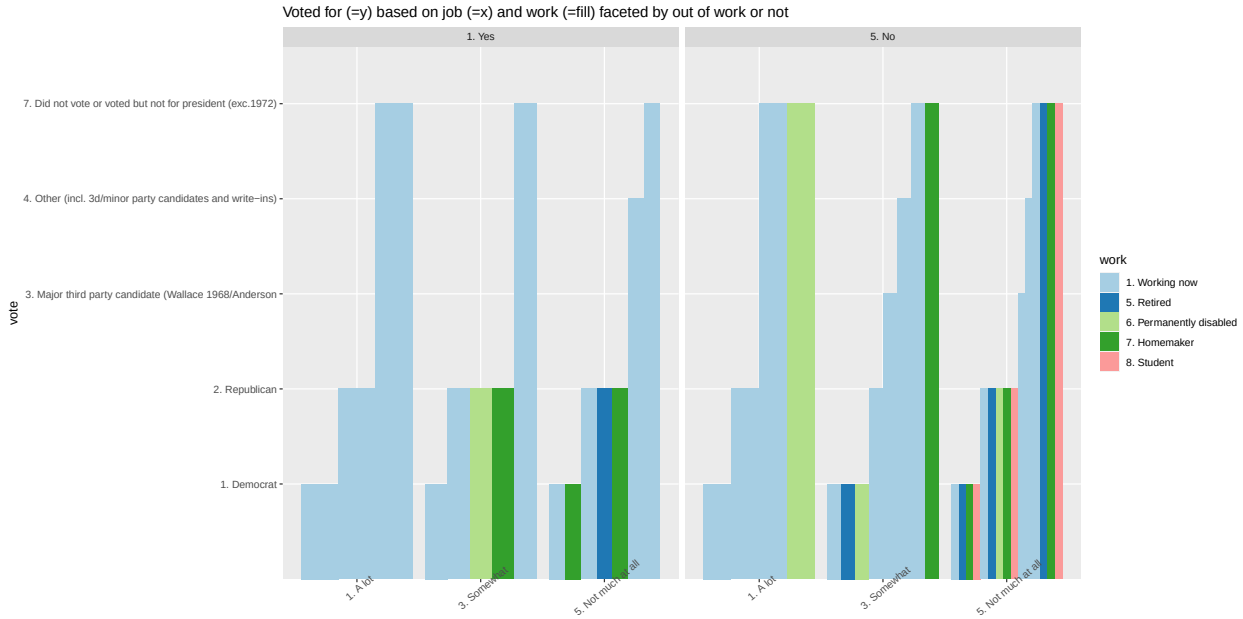
We see that democrats and republicans have voters from all regions and all kinds of places where they grew up.

For people who grew up in a suburb of a very large city, those who voted for democrats are only in the Northeast or West, and those who are currently in rural areas who voted for democrats are only in the Northeast. For people who grew up in a suburb of a very large city, those who are currently in rural areas who voted for republicans are in the Northeast and North Central regions.

For people who voted for Other/ minor 3rd party candidates, there is more uniformity based on where they grew up. If they grew up on a farm, they are also in suburban areas in the West. If they grew up in a very large city, if they are in central cities then they reside in the North Central region, and if they are in rural or small towns, they are in the South. If they grew up in a large city, these voters are in suburban areas in the Northeast. If these voters grew up in the country and not on a farm, they are either in Northeastern suburban areas or Southern rural / small towns.

3. Work, job, and out of work

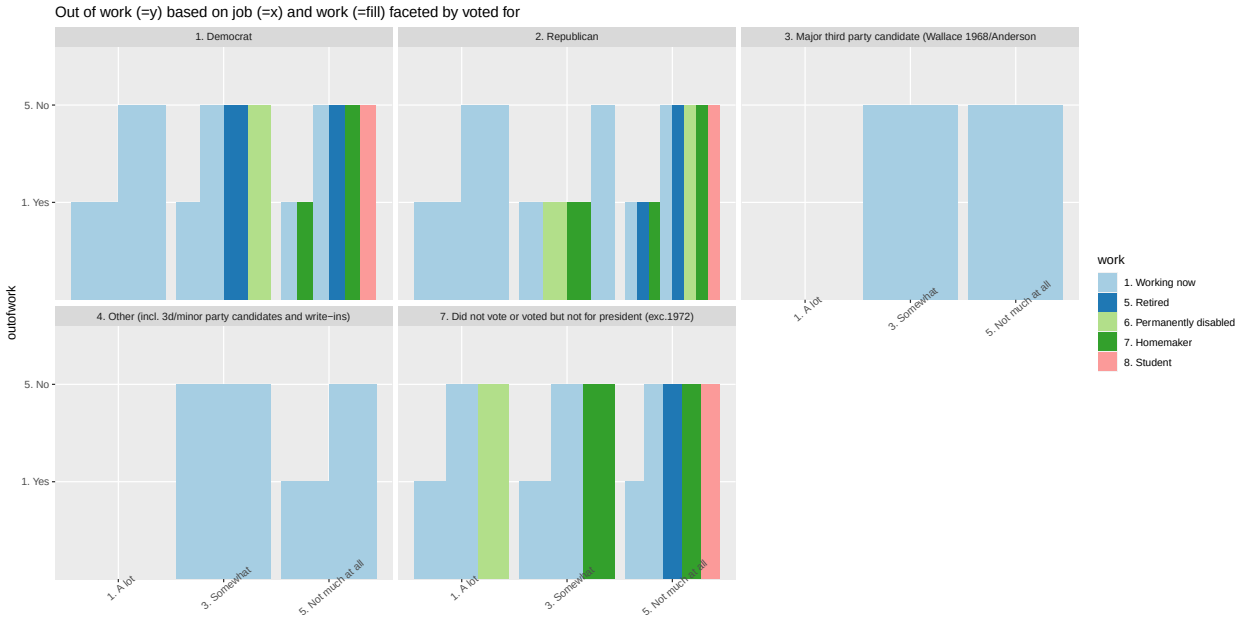
Consider the votes based on work status, whether they are worried about jobs, and if they were laid off or out of work within the past 6 months.



From the plot above, we see that for those who were out of work (or laid off) for the last 6 months (Yes) in the first plot, there are respondents who are worried "a lot" about losing/finding job in the future and who are working now (in red), there tends to be votes uniformly for either democrats or republicans; or else did not vote. We see that the homemakers in the first plot who are somewhat or not worreid about losing/finding a job in the future tend to vote mostly for republicans, and some who are not worried about jobs also vote for democrats.

For the second plot (for those who were not out of work for the last 6 months) where those are not out of work for the last 6 months and where the respondents are worreid "a lot" about losing/finding jobs in the future, they vote for republicans or democrats. We find that students and retired people who are not much worried about finding/losing jobs in the future also vote for republicans or democrats.

For major third party candidates, only people who are not out of work for the last 6 months who are working now would vote for them. This is similar pattern for minor 3rd party candidates / other, with additional people who were out of work for the last 6 months who are not much at all worried about losing or finding jobs.



We see there is most uniformity for faceted plot 3 for Major third party candidate where those are not out of work for the last months and who are somewhat and not much worried about losing/finding jobs are all working now.

There is also much uniformity for faceted plot 4 for Other/minor 3rd party candidates who are all working now, and mostly not out of work for the last 6 months and are somewhat or not much worried about losing or finding jobs in the future. This also includes those who were out of work for the last 6 months but who are working now, and who are not much worried about losing or finding jobs.

After these political parties, democrats have some uniformity. Observe those who were out of work for the last 6 months who voted for democrats: they are mostly working now, with some who are homemakers.

We compare democrats and republicans for the category of not much worried about losing/finding jobs in the future. Here, republicans have more variety of type of work (more colors). We also compare the category of somewhat worried about losing/finding jobs in the future. Here, democrats have more respondents who were not out of work in the last 6 months and these include retired people (unlike republican voters) but not homemakers (like republican voters).

Conclusion

We discuss which group (identity-related, region-related, or work status related) may tell us the most about homogeneous voting preferences based on years 2000 to 2016.

It seems that the work status related group, group 3, had more predictable or uniform results of who was voted for based on certain work-related categories than the region-related and identity-related groups. It seems that region-related group had the most variation and was not as predictable as the other groups. These results may imply that Americans have more similar voting patterns based on their work status and work-related characteristics such as being worried about finding jobs, or having lost their job in the last 6 months.

This study allows politicians to understand which groups are likely to vote for their party based on various constraints. For example, if they are in cities or rural areas, they know which groups will likely vote for them in those scenarios. Or, if they know how to speak Spanish, they can refer to the plots from group 1

where the language of the interview can be used predict which party they voted for. If the voters are worried about jobs, politicians can cater to them by promising various bills that can create more jobs.

Combined with the pie charts, politicians can use the information provided by the faceted figures strategically, and try to attract more people who conventionally did not vote for them, especially if the voters fall into the majority of people who voted as illustrated by the pie charts.