Non- and private-voting trends: Kaiser survey data 12-2020

DACSS 601 Final Paper

Steve Linberg

23 January 2021

Introduction

In both the 2016 and 2020 U.S. presidential elections, there was a great deal of discussion in the media and among political pundits about the challenges of predicting the outcomes. Nearly every major news source and polling organization reported a significant polling advantage for Democratic candidate Hillary Clinton in the 2016 election, right up to election day, with most forecasts at or above a 90% likelihood of a Clinton victory. Most major polls (with some early exceptions) showed her with a comfortable lead throughout the primary season. Republican Donald Trump's victory in 2016 was a startling surprise for many observers (and voters).

Similarly, in 2020, most major polls and forecsts expected, and predicted, strong Democratic gains in the House and Senate, along with capturing the White House (in what was commonly referred to as the incoming "blue wave"). While Democrats did win the presidency, it was by a narrower margin than some predicted; gains in the Senate were smaller than thought to be likely (Democrats gained 4 seats, but 5 races thought to be competitive remained in Republican hands[1]), and Democrats actually lost 11 seats in the House. Once again, the strength of Republican turnout was stronger than many pundits (and statisticians) expected.

One popular theory that has sought to explain this discrepancy between predicted and actual outcomes is the idea of the "hidden voter," where polls fail to take into account a significant portion of the voting population because they either (a) do not participate in polls, or (b) do not respond truthfully when asked which candidate(s) they support. One possible reason for this could involve an aversion to unwanted social pressure from openly supporting controversial candidates. Put simply, the theory was that a lot of voters secretly supported, and voted for, Donald Trump, but did not openly support him or answer truthfully in polls.

There has been a great deal of speculation about what might explain thus undercounting or underestimating of support for conservative candidates in the national elections since 2016. Theories abound, from methodological critiques of polling methodologies to rather more fantastical suspicions of deliberate sabotage by poll respondents in an effort to discredit polling and the media in general. In this paper, we will look at a more benign theory: that conservative voters are less likely to disclose their political ideology or voting choices, creating an underestimation of their electoral power that is not accounted for by polling.

We will use a December 2020 study by the Kaiser Family Foundation which looks at opinions on a range of questions relating to COVID-19, the 2020 election, and aspects of political ideology. Although the scope of the survey is broad, we will focus on respondents who either declined to divulge who they had voted for, what their ideology is, or both, and see if we can find data to support the notion that these voters are likely to be more conservative than liberal.

[1] Iowa, Main, Montana, North and South Carolina: see https://www.nytimes.com/interactive/2020/11/03/us/elections/results-senate.html

Data

The data source for this paper is the Kaiser Family Foundation poll *December 2020 Kaiser Health Tracking Poll/COVID-19 Vaccine Monitor*[2], which may be seen at:

https://ropercenter.cornell.edu/ipoll/study/31118130

It consists of 1676 observations of 168 variables, resulting from a total of approximately 40 questions and conducted by telephone. It includes an oversample of prepaid ("pay as you go") telephone numbers (25% of the total number of cell phone numbers dialed). The majority of the questions are concerned with the Affordable Care Act, COVID-19, the 2020 U.S. presidential election, and political ideology.

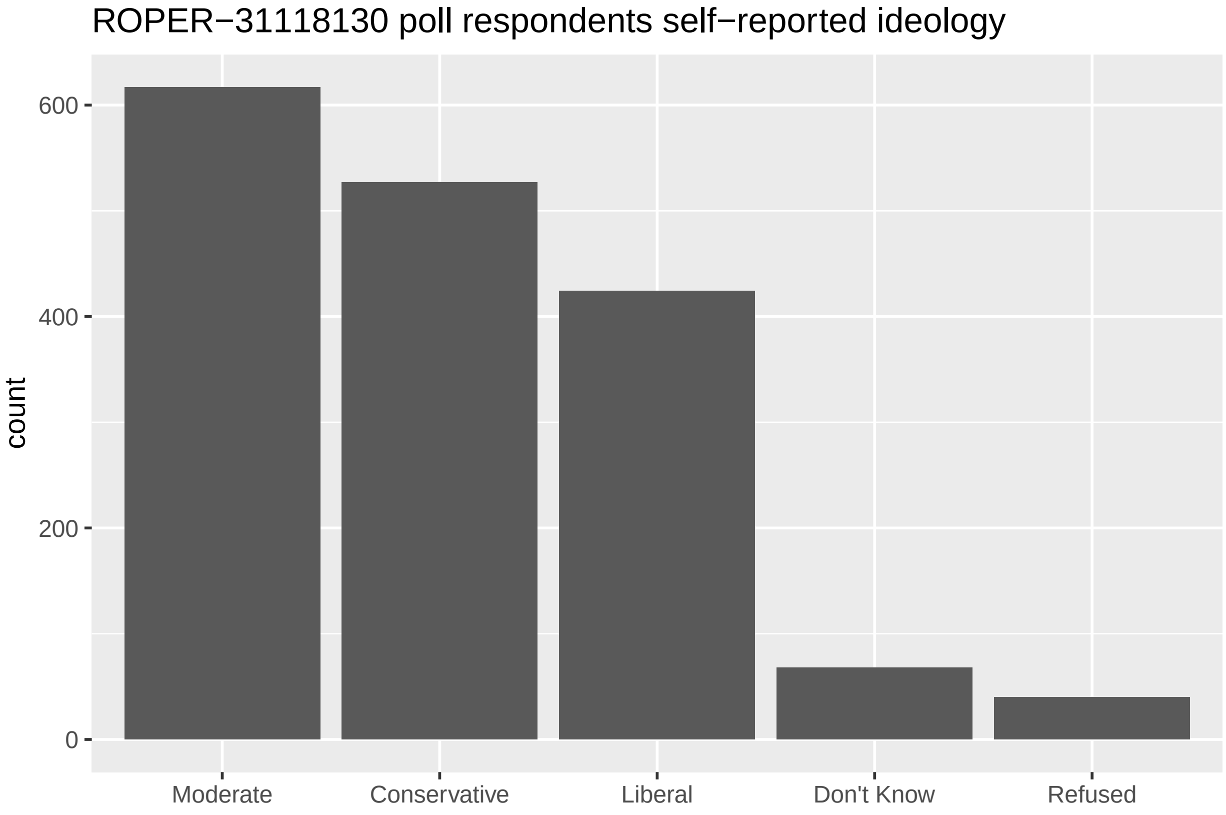
Two key variables, `ideology` and `voted2`, provide an immediate sense of the overall political leanings of respondents. `ideology` is a response to the question:

> Would you say your views in most political matters are liberal, moderate, or conservative?

The options given for response were:

- Liberal
- Moderate
- Conservative
- **(DO NOT READ)** Don't know
- **(DO NOT READ)** Refused

Given these options, the breakdown of respondents' answers is:



A small number of respondents either didn't know or refused to disclose their political ideology, but among the 1568 that did, 617 (39.3%) identified as Moderate, 527 (33.6%) identified as Conservative, and 424 (27%)

[2] Henry J. Kaiser Family Foundation. Kaiser Family Foundation Poll: December 2020 Kaiser Health Tracking Poll/COVID-19 Vaccine Monitor, 2020 [Dataset]. Roper #31118130, Version 2. SSRS [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research [distributor]. doi:10.25940/ROPER-31118130

identified as Liberal.

As this is a self-reported description, we must be careful in drawing too deeply from this variable in isolation, as the labels Moderate, Conservative and Liberal are not defined by the survey questions, but they do suggest a reasonable balance of political ideologies if we consider them as roughly "Center," "Right" and "Left" (respectively).

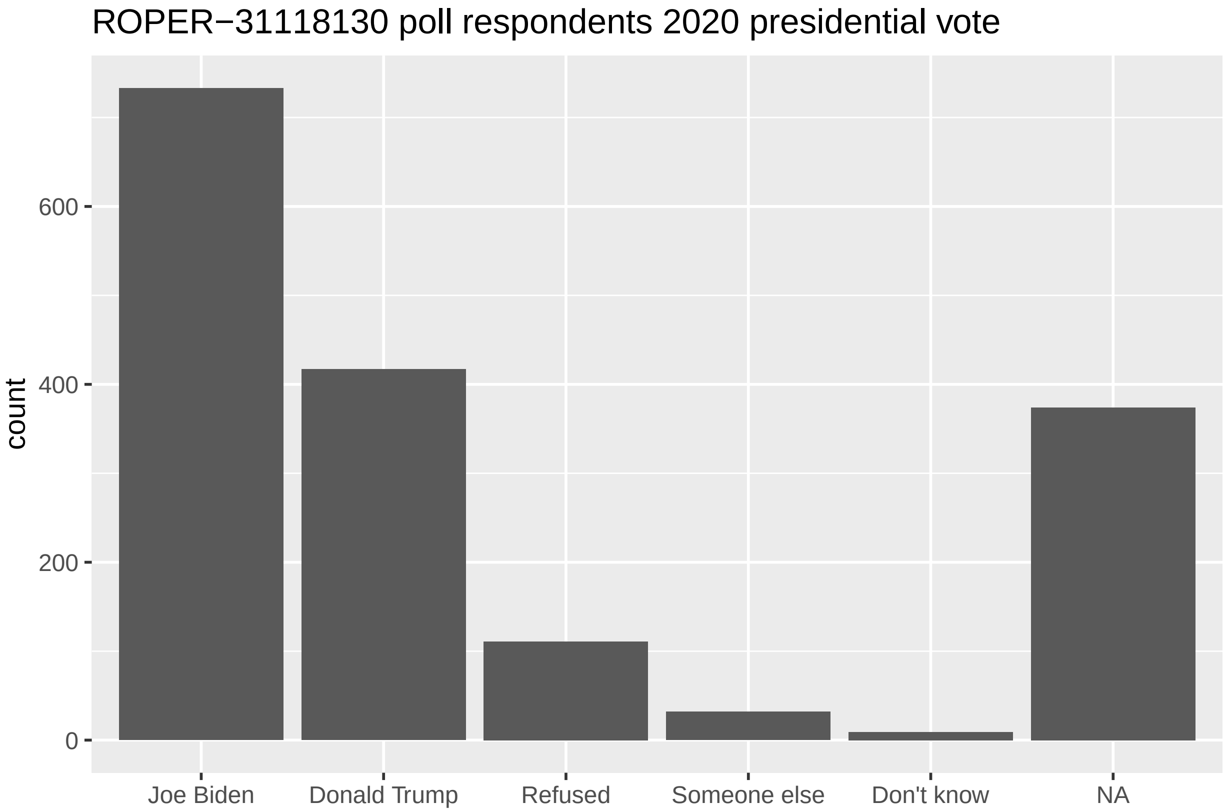
(A note on terminology: in this paper, the terms Moderate, Conservative and Liberal are capitalized when referring to survey categories, rather than the general meanings or usage of the words.)

A look at the `voted2` variable yields some interesting observations. The text of the question is

> In the election for U.S. president, did you vote for (Donald Trump) or (Joe Biden), or someone else?

The response options were:

- Donald Trump
- Joe Biden
- Someone else (Specify)
- **(DO NOT READ)** Don't know
- **(DO NOT READ)** Refused



There are three noteworthy observations here:

1. **A strong preference towards Joe Biden** Despite a somewhat balanced `ideology` variable, of the 1302 respondents who voted in the election, 733 (56.3%) voted for Joe Biden, 417 (32%) voted for Donald Trump, and 111 (8.5%) refused to answer the question (see below). Joe Biden won the national popular vote with 51.3% to Donald Trump's 46.9%, which means the results from this survey disproportionately represent Joe Biden voters as compared the national totals.

2. **A significant number of non-votes** 374 respondents (22.3%) of the total did not vote ("NA") in the presidential election, and (77.7%) did. While the visual impression of nonvotes in the diagram above

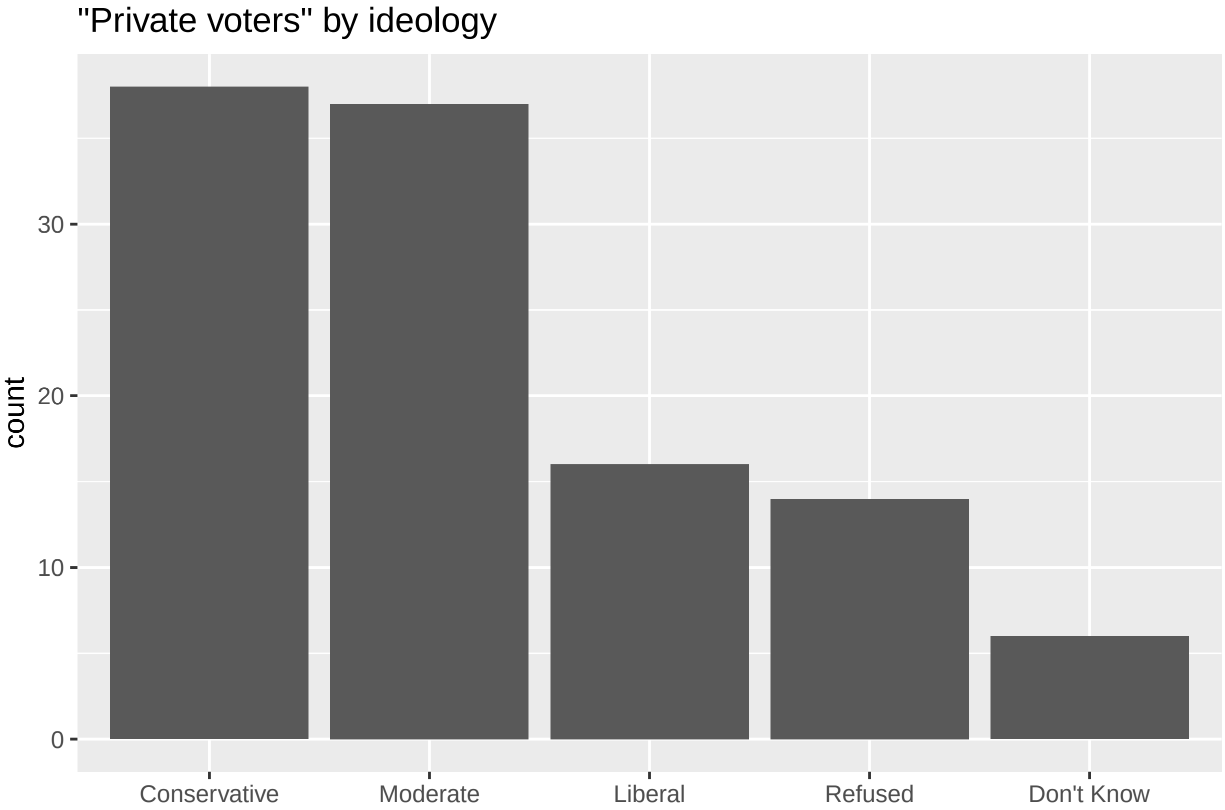
seems high, the respondents to this poll still significantly exceeded the national turnout rate of 66.7%.[3]

3. **A significant number of "Refused" (to answer) responses** As noted above, 111 respondents (8.5% of the total who voted) refused to answer who they voted for. Even if it turned out that 100% of these respondents had voted for Trump, it would still only result in a total turnout of 40.6%, well short of his 46.9% national total. The Trump vote is still under-represented in this poll.

This final observation, however, does not imply that the distribution of ideology within the group of voters who refused to answer who they voted for does not fit the national average, and this is the group we are going to investigate further, with the goal of trying to get a sense of how these respondents might have voted based on answers to other questions from the survey.

Observation 1: The ideology of "Refused" voters

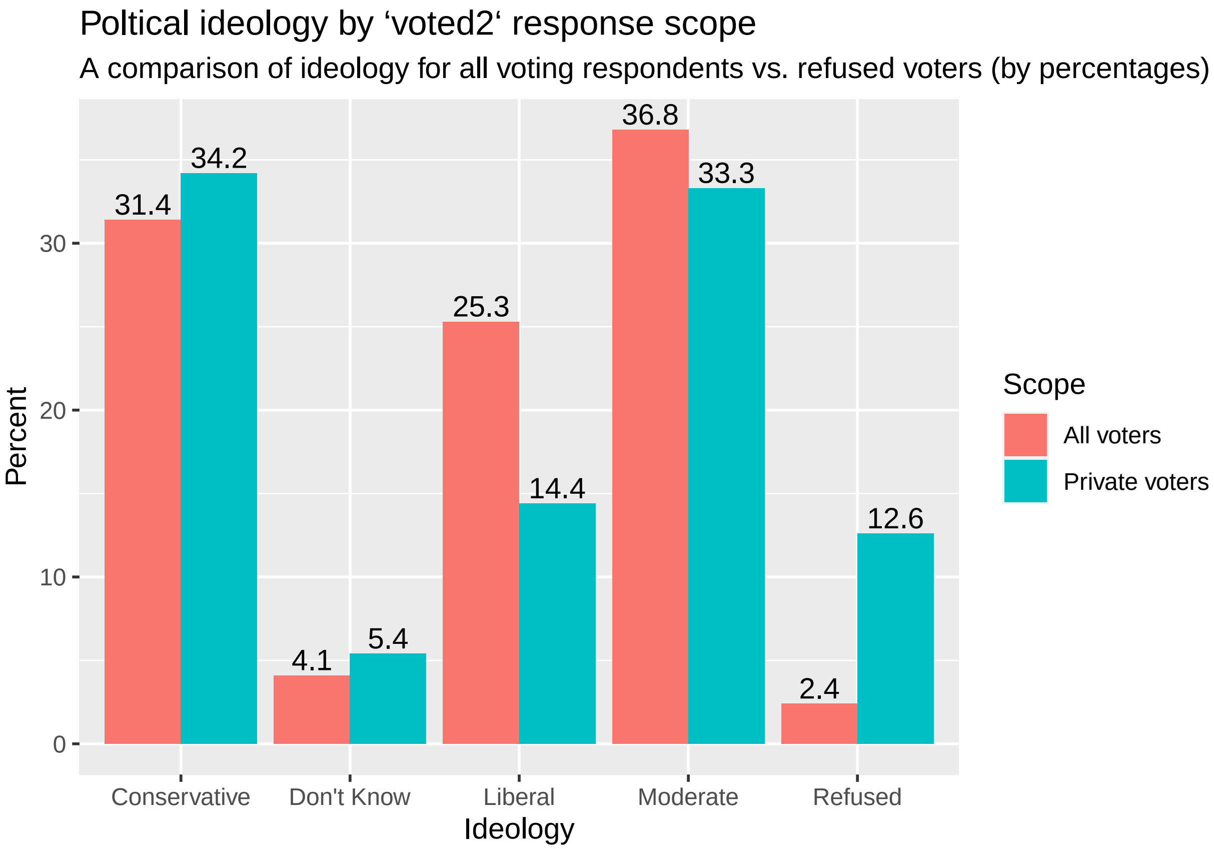
The most obvious observation to make would be: how do the voters who refused to answer who they voted for for president in the 2020 election (hereafter "private voters") self-report their political views?



We can see from this that a majority of respondents who did not disclose which presidential candidate they voted for self-identified as either Conservative or Moderate.

To take a deeper look at this tendency, we can convert the ideological share of each group of voters - all voters, and "private voters" - into percentages of their respective sums, and compare the percentages. The following chart shows the difference in these percentages; we can see that a higher percentage of "private voters" self-report as Conservative than all voters do, a smaller percentage of them self-report as Moderate than all voters, and a significantly smaller percentage self-report as Liberal.

[3] see https://www.statista.com/statistics/1184621/presidential-election-voter-turnout-rate-state/

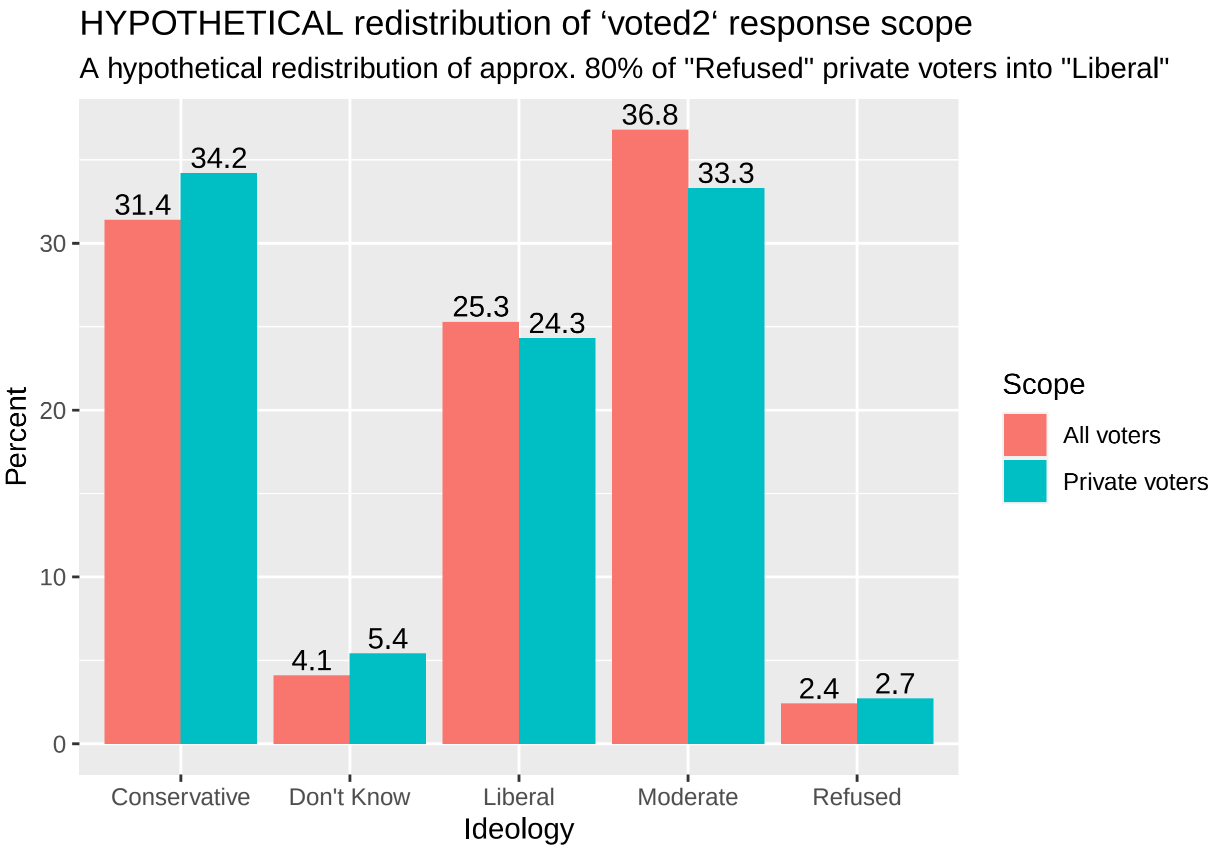


These differences are statistically significant (`chisq.test`: (chiˆ2(4) = 42.09, p < .001); `fisher.test`: (two.sided), $p = 5.3389748 \times 10^{-6}$), and appear to show that voters who refused to disclose who they voted for were more likely to self-report as Conservative, (34% to 31%) and significantly less likely to report as Liberal (14% to 25%). This supports our initial hypothesis that voters who refuse to disclose their ideology are likelier to identify as Conservative.

Observation 2: Hidden data in "Refused" ideology category

However, there is still one problem with this conclusion: As we can see in the chart, 12.6% of "private voters" also declined to share their ideology, a far higher percentage than that of the total of all voters. With perfect knowledge, we could reduce the "Refused" column down to a level similar to the total percentage of 2.4%, and redistribute most of among the other categories. Although such a scenario would seem to be very unlikely, what if it were true that most of the voters who refused to disclose their ideology were in fact Liberal? If we were to move the bulk of the "Refused" voters into the Liberal column, we get a chart with fewer overall differences.

The following chart represents ***this hypothetical scenario, and not the actual data from the study***; it is for visualization and analysis only. We will make an assumption that 11 of the 14 "private voters" who refused to disclose their ideology are in fact Liberal, and adjust the data accordingly. This would represent approximately 80% of the total of the "private voters".



With this hypothetical redistribution of 9.9 percentage points from “Refused” into Liberal, we get results that are much closer in balance to the ideologies of all voters. These differences would no longer be statistically significant (`chisq.test`: (chi^2(4) = 1.14, p = .888), and this hypothetical data would suggest that there is *no difference* in ideology between voters who did, and did not, reveal who they voted for in the election.

It is very important to note that this far, ***there is no evidence whatsoever*** that there is a significant hidden population of Liberal votes lurking within the actual 12.6% of voters who refused to disclose who they voted for; however, until we rule that out, we need to be cautious in drawing conclusions about the ideologies of “Private” voters in this study.

This merits further investigation. How can we show that this hypothetical scenario - that a majority of the voters who refused to disclose either their political ideology or who they voted for are not actually Conservative? One way would be to compare the answers to other questions on the survey which correlate strongly with, or against, responses from Conservative and Liberal voters.

One limiting factor here is that we are now down to a very small subset of our initial data; we are looking at 14 out of 1676 voters, 0.84% of the total. This is a small enough sample that statistical tests and graphs will be of dubious value; its small size increases the statistical potential of unusually strong trends within it, contributing to the overall difficulty of drawing conclusions from it, but also making an unbalancing effect stemming from it more possible than a larger set might.

Observation 3: Some manual examinations do lean Liberal

In this final observation, we extract the rows of responses for the 14 voters who disclosed neither who they voted for, nor their ideology, and compare their responses to that of the overall data set.

Question 1: The ACA Question 1 refers to the Affordable Care Act. The text of the question reads:

> What would you like to see the next presidential administration and Congress do when it comes to the health care law?

The offered responses are:

- Build on what the law does
- Keep the law as it is
- Scale back what the law does
- Repeal the entire law
- **(DO NOT READ)** None of these/Something else (Vol.)
- **(DO NOT READ)** Don't know
- **(DO NOT READ)** Refused

Looking at the results for the entire survey, we can see a tendency for strong support for "Build on what the law" does from Liberal respondents., while "Repeal the entire law" is stronger among Conservative respondents (all values in percentages by ideology):

```
##
##                                         Conservative Don't Know Liberal Moderate
##   Build on what the law does                   26.57      38.24   72.41    58.35
##   Don't know                                    6.64      17.65    3.77     4.70
##   Keep the law as it is                        11.20      27.94   12.74    14.75
##   None of these/Something else (Vol.)           2.28       2.94    1.42     2.27
##   Refused                                       0.00       2.94    0.00     0.49
##   Repeal the entire law                        39.85       8.82    6.37    11.67
##   Scale back what the law does                 13.47       1.47    3.30     7.78
##
##                                         Refused
##   Build on what the law does              40.00
##   Don't know                              12.50
##   Keep the law as it is                   15.00
##   None of these/Something else (Vol.)      5.00
##   Refused                                  7.50
##   Repeal the entire law                   20.00
##   Scale back what the law does             0.00
```

This shows strong Liberal support for the response "Build on what the law does" (72%), while "Keep the law as it is" lacks strong support among both Liberals and Conservatives (and more among "Don't Know", but this group only represents approximately 4% of the respondents total).

If we look at the answers to this question among our small, 14-voter sample, we see:

```
##
## Build on what the law does                 Don't know
##                      35.71                       7.14
##      Keep the law as it is                    Refused
##                      28.57                       7.14
##      Repeal the entire law
##                      21.43
```

This shows close-to-even support for "Build on what the law does", "Keep the law as it is," and "Repeal the entire law". The overall results could be interpreting as tilting slightly towards an overall sentiment consistent with a Liberal distribution. Although the sample size is small, this is our first indication that it is indeed possible that our (small) hidden population of ideology tilts Liberal.

Question 5: Reporting on the seriousness of Coronavirus Question 5 on the survey reads:

> Thinking about what is said in the news, in your view is the seriousness of coronavirus (generally exaggerated), generally correct, or is it (generally underestimated)?

The offered responses are:

- Generally exaggerated
- Generally correct
- Generally underestimated
- **(DO NOT READ)** Don't know
- **(DO NOT READ)** Refused

The responses of the entire survey are:

```
##
##                            Conservative Don't Know Liberal Moderate Refused
##   Don't know                       2.66       7.35    4.48     1.46    5.00
##   Generally correct               21.82      41.18   49.29    44.08   32.50
##   Generally exaggerated           55.98      32.35   10.14    25.61   32.50
##   Generally underestimated        19.35      17.65   35.38    28.53   27.50
##   Refused                          0.19       1.47    0.71     0.32    2.50
```

We can see here that nearly 50% of Liberals respond with "Generally correct", an assessment shared by only 22% of Conservatives. Our small sample shows a 42% majority of responses of "Generally correct" among the respondents who did not disclose ideology or presidential choice:

```
##
##        Generally correct    Generally exaggerated Generally underestimated
##                    42.86                    28.57                    28.57
```

This sample, although small, continues to support the notion of a Liberal-direction ideological lean for these respondents.

Question 9: On mask-wearing Question 9 on the survey reads:

> Which comes closer to your view: wearing a mask to prevent the spread of COVID-19 (is a personal choice) OR wearing a mask (is part of everyone's responsibility to protect the health of others)?

The offered responses are:

- Wearing a mask is a personal choice
- Wearing a mask is part of everyone's responsibility to protect the health of others
- **(DO NOT READ)** Both (Vol.)
- **(DO NOT READ)** Neither (Vol.)
- **(DO NOT READ)** Don't know
- **(DO NOT READ)** Refused

The responses of the entire survey are:

```
##
##                                                                                        Conservative
##   Both (Vol.)                                                                                  1.52
##   Don't know                                                                                   0.38
##   Neither (Vol.)                                                                               1.33
##   Refused                                                                                      0.00
##   Wearing a mask is a personal choice                                                         42.50
##   Wearing a mask is part of everyone's responsibility to protect the health of others        54.27
##
##                                                                                        Don't Know
##   Both (Vol.)                                                                                2.94
##   Don't know                                                                                 1.47
##   Neither (Vol.)                                                                             0.00
##   Refused                                                                                    0.00
##   Wearing a mask is a personal choice                                                       16.18
##   Wearing a mask is part of everyone's responsibility to protect the health of others      79.41
##
##                                                                                        Liberal
##   Both (Vol.)                                                                             1.89
##   Don't know                                                                              0.00
##   Neither (Vol.)                                                                          0.24
##   Refused                                                                                 0.00
##   Wearing a mask is a personal choice                                                     5.66
##   Wearing a mask is part of everyone's responsibility to protect the health of others   92.22
##
##                                                                                        Moderate
##   Both (Vol.)                                                                              1.94
##   Don't know                                                                               0.00
##   Neither (Vol.)                                                                           0.00
##   Refused                                                                                  0.16
##   Wearing a mask is a personal choice                                                     15.56
##   Wearing a mask is part of everyone's responsibility to protect the health of others    82.33
##
##                                                                                        Refused
##   Both (Vol.)                                                                            10.00
##   Don't know                                                                              5.00
##   Neither (Vol.)                                                                          2.50
##   Refused                                                                                 0.00
##   Wearing a mask is a personal choice                                                    22.50
##   Wearing a mask is part of everyone's responsibility to protect the health of others   60.00
```

This shows that Liberals overwhelmingly (92%) choose the response "Wearing a mask is part of everyone's responsibility to protect the health of others", whereas Conservatives are close to split between this response and "Wearing a mask is a personal choice". Our small survey shows:

```
##
##                                                                           Both (Vol.)
##                                                                                 14.29
##                                                   Wearing a mask is a personal choice
##                                                                                 21.43
## Wearing a mask is part of everyone's responsibility to protect the health of others
##                                                                                 64.29
```

So once again, we see a strong - though not overwhelming - Liberal tilt in the responses.

The answers to these questions indicate that there could indeed be a Liberal ideological bias - possibly even a strong one - among the respondents who chose not to reveal either their ideology or their presidential choice.

We saw above that if 80% of the respondents in this category were of (a hidden/unreported) Liberal ideology, it would be enough to balance out the overall ideological responses among voters who refused to disclose who they voted for, and *that* would be sufficient to prevent us from drawing the conclusion that it tends to be conservative voters who are more likely to conceal who they vote for.

In fact, with only 5 of the 11 "Refused" voters being redistributed added to the Liberal total, and the remaining 6 distributed in approximately equal proportions among the remaining choices (3 each to Conservative and Moderate), our `chisq.test` results are still chiˆ2(4) = 3.21, p = .524[4]. So a Liberal tilt in these voters is not

[4] for our purposes here, we retain 3 of the 14 voters in a "Refused" category, rather than redistributing them all among the other choices, in order to retain a percentage-wise balance with the full data set and avoiding creating a 0-value inflection point

even needed to prevent us from supporting our initial hypothesis, and the answers we saw to the questions above provide fairly strong support for the existence of this bias. All that is necessary is a redistribution - or reveal - of the "Refused" voters' hypothetical ideologies. It is the concealing of the ideologies, rather than their tilt, which creates the low p-value in the first `chisq.test` examining the differences in ideology between all of the respondents, and those who declined to disclose them.

Conclusion

This paper looked at the question of whether voters who chose not to reveal which presidential candidate they voted for, in data from a recent Kaiser Family Foundation survey, had a hidden ideological bias that could account for some of the difficulty in predicting the strength of Conservative turnout in recent elections.

Based on the ambiguity of the ideology of respondents who refused to answer who they voted for, and on the analysis of responses to other questions this subset of respondents answered, we cannot support the hypothesis – which a first-level examination of the data seemed to suggest – that voters who choose not to reveal who they vote for have any particular ideological bias compared to those who do.

It also does not *disprove* this possibility, of course. But these results could be surprising to those (including this author) who accepted the general concept of some degree of Conservative bias in undisclosed ideological preferences among voters. It is possible that such a bias exists, but this examination of this dataset does not support it.

Reflection

The process of writing this paper was rather grueling, chiefly because I went through the class essentially solo and online (during the January term), without peer support or other students to work with, so there were several "rabbit holes" I ended up going down (for several days apiece) that probably could have been avoided with a peer group. Professor Rolfe generously met with me once a week for questions and discussion, which was helpful, but I still think I would have gotten through with less wasted time if I'd been able to work actively with other students.

Some of the time-sinks would be expected in any circumstance of bootstrapping onself into a new area of knowledge; I have a long background in computer programming in various languages dating back to the early 1980s, but they are/were mostly low-level languages initially (C and assembly), and in more recent years, Perl, Objective-C, PHP and Swift, so I have become increasingly used to strongly-typed languages (or the ability to activate strong typing in more permissive languages), and R is pretty different in some important ways. Although the basics of the language are straightforward, some of the nuances about underlying data structures are quite complex (like the differences between sometimes-interchangeable vectors, data frames, tables and matrices, and their modes of access), and without strict typing, it's possible to make errors that are difficult to detect.

In particular, I got hung up pretty hard on the proper way to run `chisq.test` on frequency tables from the data, and initially was getting an erroneous result due to bad inputs rather than bad data. My case was specific enough that doing productive online searches was arduous and probably more confusing than helpful; I ended up rolling back to a completely manual data set and doing a lot of experimenting to first get the correct results, and then did a lot of work to reverse-engineer the manual data back to the proper references for the actual data set. Someone more experienced than me in R could probably have identified my issue immediately, but it took a long time to find on my own.

Another issue was getting stuck for a couple of days trying to work out whether the process of looking through the data for statistically significant distributions constituted an unintended form of data dredging, or "p-hacking", and whether I needed to be applying a remedy such as a Bonferroni correction to my results. This topic was not addressed in the class curriculum, but I had heard of it on my own previously and needed to learn more about it to make sure my results were valid. In the end, I concluded on my own that looking through different question data for statistically significant results was valid, and not the same thing as asking

for the `chisq.test` to factor in, which would skew the results more strongly than our approximation of 3.

the same question of narrower subsets of the same data looking for significant results, but this was not clear to me at the beginning.

Once I had those issues sorted out and began to increase my ability to query the data and look for trends, the process became more enjoyable, and I was surprised to come upon the conclusion that I did - or rather, to find a *lack* of statistical evidence to support my own bias. This is actually quite exciting to me intellectually, even if it feels like something of an ego-check on something I assumed was true about myself. It is an easy mistake to make - getting an initial result of statistical significance that confirms one's own bias is a difficult siren call to resist. It is not enough to get a low p-value; the meaning of the data must be considered as well, and important information can be hiding in it.

This is exactly why we need statistics and analytics in social science. I would not have had the mechanisms for looking at questions of my own bias prior to undertaking this paper (and was not expecting to find any in its execution).

I can see at this point how powerful deep fluency with data query languages like R, and with statistical principles, will be in future research. I have a lot to learn about R still, and about statistical methods, but I need no convincing of their utility. I look forward to continuing to develop both in future classes as I move through the DACCS program.