5-2022

Understanding and Improving the System: The Effects of Weighting on the Accuracy of Political Polling in Arkansas

Beck Williams
University of Arkansas, Fayetteville

Understanding and Improving the System: The Effects of Weighting on the Accuracy of Political Polling in Arkansas

An Honors Thesis submitted in partial fulfillment of the requirements
for Honors Studies in Political Science

By
Beck Williams

Spring 2022

Political Science
J. William Fulbright College of Arts and Sciences
The University of Arkansas

Acknowledgements

To begin, I would like to thank Dean Todd Shields of the J. William Fulbright College of Arts and Science, whose advice and mentorship were invaluable as thesis director. Dr. Samantha Robinson has also been a fantastic mentor with great advice regarding the statistical processes throughout the duration of this project. Thank you also to Dr. Angie Maxwell for her eagerness to join the thesis committee and offer her expertise in southern politics.

Additionally, I would like to acknowledge Dr. Noah Pittman, who first exposed me to the discourse surrounding political polling through his Campaign 2020 Honors College Forum. As much of this project involved the usage of R, I thank Drs. Jyotishka Datta and Avishek Chakraborty for introducing me to this software through their statistics courses. Thank you as well to Lane Schmidt, whose administrative help through scheduling and swift communication has allowed for a smooth project.

My time completing this thesis, as well as the full extent of my undergraduate career, have been greatly benefited by the consistent support of the Fulbright College Honors Program, including Jill Geisler Wheeler and Mary Ellen Hartford. Finally, this project would not exist without the administrators of the Arkansas Poll—many thanks to Dr. Janine A. Parry and the Diane D. Blair Center of Southern Politics and Society for their continued work in providing Arkansans with this consistently professional and important survey.

Table of Contents

Chapter 1 – Introduction

Election history in the United States is full of shocking and unexpected moments which shifted the political playing field—so much so, in fact, that attempting to pinpoint a specific election as the "most shocking" proves quite difficult. One may point to the early 19th Century in which the electoral college resulted in a tie between Thomas Jefferson and his apparent running mate Aaron Burr, while others may cite the famous "Dewey Defeats Truman!" newspaper headline which proved false in 1948. Modern voters might be more likely to point out the 2000 Election which remained hotly contested for weeks after voting had already closed.

Despite these contenders, an argument can be made for one much more recent: the 2016 Presidential Election. It is no secret that the results of this election were more than unexpected. Not only did it involve one of the most unorthodox presidential nominees in recent history, but political polling at the time was more advanced and modernized than ever before. The fact nearly every major media outlet and pollster predicted the election results incorrectly, and by such wide margins, proved to be an earthquake in the polling community.

FiveThirtyEight, one of the most trusted and publicized polling aggregates in the industry today, also made the early call for Clinton. Through compiling pre-election polls from every state in the country, FiveThirtyEight produced their final forecast as voting began on November 8th, giving Clinton a 71.4% chance of ascending to the presidency while giving her opponent, Republican Nominee Donald Trump, a mere 28.6% chance. While this may not seem too extreme, their predictions in individual states were disheartening as well. FiveThirtyEight assigned at least 75% to Clinton's odds in Pennsylvania, Wisconsin, and Michigan—all states which eventually went to Trump. As it turns out, FiveThiryEight was rather modest in their overall prediction; many other major media outlets gave Clinton odds between 80-90%.[1]

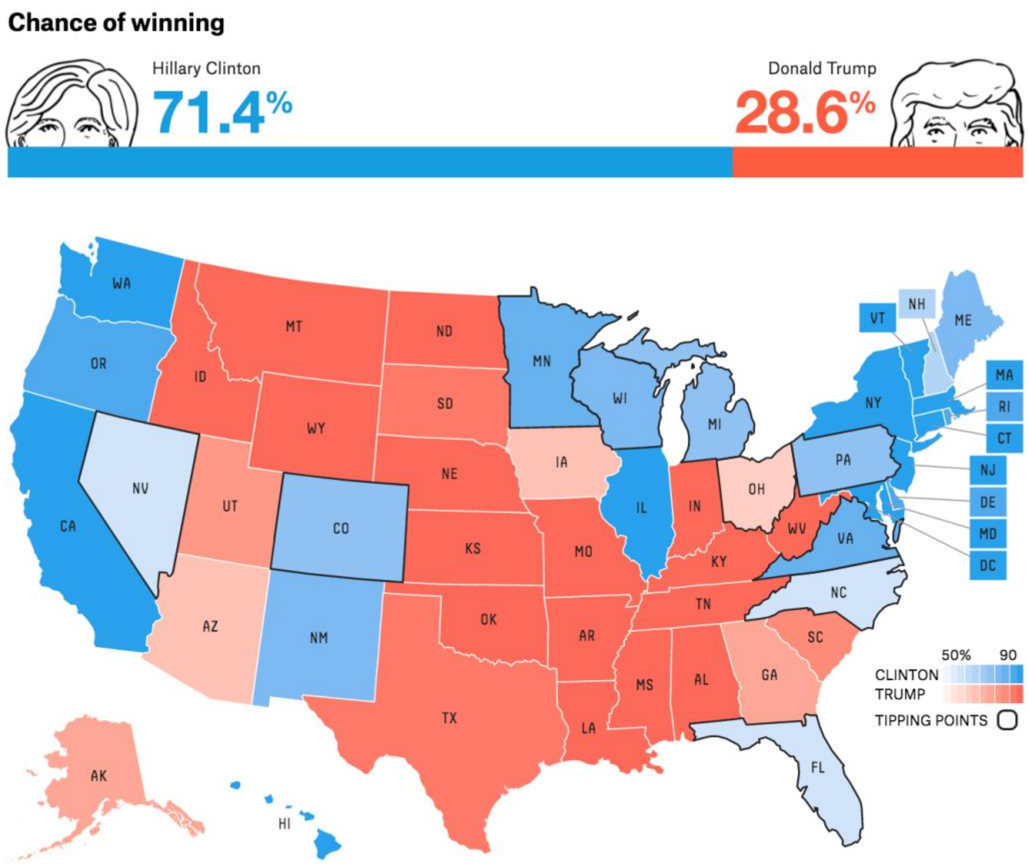

Figure 1.1: FiveThirtyEight Final Election Forecast – 2016 Election Coverage.[2]

Because of what was largely seen as a "failure" of political polling across the board, pre-election polls were placed under a microscope like never before in the following years.[3] The

[1] Kennedy, et al., "Evaluation of 2016 Election Polls."
[2] FiveThirtyEight, "2016 Election Forecast."
[3] Kenett, et al., "Election Polls."

potential failure of political polls is a tense thought for anyone from everyday voters to those directly involved with the election process. This is due to polling's role as a cornerstone of our political process. Political scientists rely on accurate polls in order to correctly understand voter behavior.[4] Politicians themselves use pre-election polls to assess campaign strategy, focusing their efforts and resources on those areas in which they find a lack of support.[5] News media in particular emphasizes polling as a means to identify election winners to their viewers days, weeks, or even months in advance of the actual elections.

Over the years, statistical strategies have been implemented in order to achieve the most accurate polls possible, with an invigorated focus drawn to the subject since the 2016 Election. These strategies include those which shape the sampling process, but in many cases, they focus instead on manipulation of previously collected survey data. Adjusting a survey's sampling process can only increase accuracy to a certain degree, so these after-the-fact methods, such as weighting and imputation, can be especially important for accurate polling results.

This thesis will explore the subject of polling accuracy by focusing on one state in particular: Arkansas. While Arkansas has not necessarily been cited as a "swing state" in recent elections (or in much of its history at all), it presents an interesting environment for polling examination in which our conclusions may yield potential for application in other areas of the country. To complete our study, we will explore the modern polling environment and analyze increasingly popular strategies for accurate results. Then, we will apply these strategies, namely weighting, to the sample of The Arkansas Poll, one of the only two state-specific public opinion polls in Arkansas which is utilized by polling aggregates including FiveThirtyEight. The goal is then to inspect how our manipulation of the survey sample changes Arkansas Poll results in

[4] Hillygus, "Evolution of Election Polling."
[5] Ibid.

recent years, both in terms of presidential election prediction and public opinion on hot-button issues. Finally, we will conclude with an analysis that examines the potential for improvement moving forward, both within The Arkansas Poll and other public opinion polls across the country.

Chapter 2 – The State of Political Polling Today

To develop the focus of our study, we will begin by diving into the current state of polling within the United States. First, we will examine the performance of political polling as an industry in the context of the most recent presidential elections. This will allow us to gain better insight into the catalysts behind recent polling inaccuracies while also displaying the full extent to which such inaccuracies occurred. Then, we will reel our concentration toward Arkansas through an evaluation of the state's own polling environment. The information presented here can then guide our experiments in weighting the Arkansas Poll.

2.1 Polling Performance in Recent Elections

Despite what we have already covered regarding the 2016 election, polling is not a misguided or "broken" medium for gaining political information. The reason it might appear that way, in actuality, is that a political upset stands out more than a correct prediction because of the high rate of correct predictions in the polling community.[6] In other words, an affirmed election prediction is hardly news because it has been reported as the likely outcome for weeks before the election; a "failure" of polling, on the other hand, is going to garner a much greater rate of media attention due to its sheer surprise.[7]

Actually, the term "failure" is a considerably negative one considering the 2016 polls. While they may appear to have failed because they gave greater odds to Clinton than Trump

[6] Kennedy, "Can We Still Trust Polls?"
[7] Kenett, et al., "Election Polls."

(sometimes to an extreme), the winner of the election is only a piece of the larger picture. One piece of consolation is the fact that Clinton won the national popular vote as predicted—albeit by 2% rather than the predicted 3%.[8] This was a good sign that national polls were performing well and accurately predicting within their margin of error, even reaching historical records of popular vote accuracy.[9]

The reason for the missed prediction, then, came down to a matter of state polls. Presidential elections in the U.S. are decided not by popular vote but by the Electoral College's "winner take all" system in each of the individual states. Therefore, when FiveThirtyEight or other polling aggregates calculates the election odds, they are largely considering state polls, rather than national polls, for those designated as "swing states."[10] For 2016, this emerged as a problem in a number of key midwestern states—namely Michigan, Wisconsin, and Pennsylvania, each of which was slated to go to Clinton.[11] Polling in these states suffered a few major problems which then contributed to the upset. First of these was an unprecedented number of undecided voters in each state who did not make their decisions until a few days before the election, causing a significant rift from the prediction when a large majority swung to Trump.[12] The other major problem was a failure to collect demographically representative samples, specifically on the education front. More specifically, most of the state polls in these swing states oversampled college graduates for whatever reason, therefore polling non-college-educated voters at a significantly lower rate than they showed up to the polls on election day.[13]

---

[8] Ibid.
[9] Dean, et al., "Field Guide to Polling."
[10] Ibid.
[11] Ibid.
[12] Kennedy, et al., "Evaluation of 2016 Election Polls."
[13] Ibid.

The good news involving the latter of these state poll problems is that underrepresentation of specific demographic groups can be addressed after the survey has been completed. The problem was not the undersampling but instead the failure to account for such. In recent years, extra focus has been placed on fixing unrepresentative samples through strategies such as weighting and imputation (both of which are explored further in the following chapters).[14]

Four years later, the 2020 Presidential Election did not create the same "accuracy frenzy" that the previous election had. This is unsurprising considering that most polling aggregates, including both FiveThirtyEight and RealClearPolitics, attributed higher odds to the winner and therefore correctly predicted the election (with FiveThirtyEight giving Joe Biden, the Democratic nominee, an 89% chance of winning—over 10 percentage points more than Clinton in 2016).[15] However, this is not immediately indicative of a more accurate polling year than that of 2016. As it turns out, whereas the accuracy of national polling reached a historic high in 2016, it was considerably lower in 2020.[16] Both FiveThirtyEight and RealClearPolitics listed Biden at an advantage of over 7 percentage points, but he only outperformed Trump by less than 5.[17] With the recency of this election, many are still unsure of the reasons behind this decrease in national polling accuracy. In the very least, this result is important to note in that it shows our political polling did not get necessarily "better" between those elections, and that it instead is continuing to evolve.

---

[14] Dean, et al., "Field Guide to Polling."
[15] FiveThirtyEight, "2020 Election Forecast."
[16] Panagopoulos, "Polls and Elections."
[17] Ibid.

2.2 Political Polling in Arkansas

Many in the political community are familiar with the phrase “all politics is local.” Regardless of its validity, there is at least merit in the fact that state polls, not national ones, determine presidential elections. As discussed, this was displayed in 2016 when national polls correctly identified the popular vote victor while state polls led the community astray. When we examine a specific state, an entirely new snapshot of the election process is produced.

Looking at Arkansas specifically creates a unique situation. Unlike much larger states, such as California and Texas, or those states which tend to swing to either party such as Wisconsin and Pennsylvania, Arkansas has been a completely red state for the entirety of the 21st Century at least in terms of presidential elections. Pre-election polling, then, is not exactly as emphasized as elsewhere. For the entire state of Arkansas, polling aggregate RealClearPolitics utilizes only two polls to predict elections: The Arkansas Poll and a yearly poll conducted by Talk Business & Politics in conjunction with Hendrix College.[18] FiveThirtyEight similarly only uses three polls, with the two previously mentioned pollsters being the only ones that are specific to Arkansas.[19]

Published every year, the Arkansas Poll in particular covers a lot of ground in terms of public opinion. Not only does it survey a random sample of Arkansans on their voting preferences during election years, but it also gathers information on major political issues of concern to the Arkansas population. Its goal, then, is to accurately reflect the public opinion of the state population through its yearly summary report. In doing this, it is difficult to measure the accuracy of general public opinion, but on the question of presidential choice, the Arkansas Poll does have one retroactive accuracy measure: the election results. The Arkansas Poll has a solid

---

[18] RealClearPolitics, “Arkansas: Trump vs Biden.”
[19] FiveThirtyEight, “Trump is Very Likely to Win Arkansas.”

track record in terms of predicting elections—not just in choosing the winner, but also in calling the percentages within their specified margin of error.

Even with this track record, the Arkansas Poll is still assigned a rating of "B+" by FiveThirtyEight, which weights its pollsters based on the letter grade they receive. This letter grade is affected by a number of different factors, namely of which is track record, but also includes survey design and something the Arkansas Poll lacks: the presence of additional weighting to match a sample to its population.[20] The remainder of this thesis will explore the potential for improvement within the Arkansas Poll with a specific eye toward weighting the survey sample to achieve the goal of accurately reflecting the population.

[20] Silver, "How FiveThirtyEight Calculates Pollster Ratings."

Chapter 3 – The Arkansas Poll & Potential for Improvement

"The mission of the Arkansas Poll is to supply timely, accurate, and impartial public opinion information on matters of policy and politics to public officials, researchers, students, and the public."[21] This quotation, taken directly from the current website of The Arkansas Poll, demonstrates the importance of the poll in our state. Being one of only two state-exclusive public opinion polls in Arkansas, The Arkansas Poll has a responsibility to conduct their survey on a yearly basis and to follow the standards agreed upon in the political polling community.[22]

As noted already, The Arkansas Poll has been fairly accurate in its election predictions of previous years. Some might respond that at this point in the state's history, it is not hard to predict the winner of a nationally elected office in Arkansas: all U.S. House of Representatives and Senate seats have been held by Republicans since 2014, and similarly, no Democratic candidate for president has received Arkansas's electoral votes since Bill Clinton in 1996.[23] In other words, although this has not been the norm for all of its history, Arkansas is a "red state" right now and looks as though it will continue to be in the coming years. However, The Arkansas Poll has not only been accurate in its prediction of election winners, but in both 2020 and 2016, it was within its margin of error in measuring the vote percentage of each presidential candidate.

To put these observations in context, we need to dive deeper into The Arkansas Poll. In this chapter, we begin by exploring the protocols of The Arkansas Poll, including its presentation of findings and the typical conduct of the survey itself. Then, we will examine the Poll's

---

[21] "The Arkansas Poll" Website.
[22] RealClearPolitics, "Arkansas: Trump vs Biden."
[23] 270 to Win, "Arkansas Presidential Election Voting History."

performance in recent years, both as an election predictor (with specific focus on presidential elections) and as a reflection of the Arkansas population. Once this is established, we will continue by examining survey design broadly as to identify potential statistical strategies which can be employed for the Poll's improvement.

3.1 The Arkansas Poll – A Closer Look

In the conducting of any survey, the goal is to represent a population using a group from within that population, which constitutes the survey sample. It is from this sample that we derive our statistics, or estimates which seek to describe certain attributes of the population. The makeup of this sample is therefore of prime importance in surveying, because if it does not reflect the views and opinions of the population proportionately, then it will provide researchers with insufficient or inaccurate estimations of population parameters. Therefore, if we are to measure the potential for accuracy improvement in The Arkansas Poll, we must first understand the survey design behind the Poll.

Since the first conducted Arkansas Poll in 1999, the survey data has been gathered through a random sample of phone numbers.[24] Survey conductors will call randomly selected phone numbers of Arkansas adults and proceed to ask the potential respondent to participate in the survey; then, they would conduct the phone interview according to a set question protocol and record their answers. It should be noted that not every person contacted would participate in the call. In fact, more often than not, the potential respondent refuses; in the first edition of The Arkansas Poll, there were over 3,500 calls made for an end result of 885 respondents, whereas in the 2020 Poll, there were 804 completed phone interviews with a response rate of 46%.[25] The

---

[24] Parry, et al., *The Arkansas Poll, 1999: A Summary Report.*
[25] Parry, *The Arkansas Poll, 2020: Summary Report.*

Arkansas Poll always shoots for a sample size of at least 800 respondents, and each of the phone interviews is conducted by an outside organization, which in recent cases has always been Issues and Answers Network, Inc.[26]

Telephone interviews have long been one of the simplest and most effective methods of conducting surveys, but that is not to say it is without pitfalls. As it currently stands, while there are definitely standards and best practices, there is no widely agreed-upon method for conducting surveys—all methods are going to fail to reach select groups and demographics within populations, and all methods will fall prey to some form of bias. In telephone polls, the major pitfall right now is the declining rate of participation via landline phones and the increasing rate of response via cell phone, which introduces a potential sampling discrepancy when certain demographics gravitate strongly toward one of those two types of phones.[27] Even if we ignore these trends, telephone surveys still tend to reach a different audience than surveys conducted via the internet, so there will persist discrepancies between the sample and the population nonetheless.

That being said, The Arkansas Poll specifically has one partial remedy to this situation: it acknowledges its shortcomings. The summary report of each edition of The Arkansas Poll states the following: "To assess the representativeness of the sample drawn for the poll, the Arkansas Poll team publishes what most polling organizations do not: a comparison of survey respondents' key demographic characteristics to those of the state as a whole."[28] Following this statement, the summary report includes a chart comparing demographics percentages of the sample versus demographic percentages of the state population, shown below.

[26] Ibid.
[27] Keiding and Louis, "Web-Based Enrollment."
[28] Parry, *The Arkansas Poll, 2020: Summary Report.*

	2020 Arkansas Poll Sample	State of Arkansas
Average Age Category (for 18+ pop.)	64	49
Married	55%	49%**
Gender		
Male	45%	49%
Female	55%	51%
Educational Attainment		
High school graduates	93%	85%
College graduates	32%	22%
Median Household Income Category	$35,001-$50,000	$42,336
Race/Ethnicity		
White	81%	73%
Black or African American	9%	16%
Multi-Ethnic	4%	2%
Hispanic	2%	8%*
Native American	2%	1%
Asian	<1%	2%
Something Else, Don't Know, Refused	2%	-

Figure 3.1: 2020 Arkansas Poll Demographics Comparison.[29]

3.2 Understanding the Problem

In their mission to "supply timely, accurate, and impartial public opinion information," the Arkansas Poll has embraced an objective of representing the opinions of the state of Arkansas. The Poll is surely timely; it is conducted once per year so that results are published on a yearly basis and can be compared across time, and the conduct/publication of the survey in October means that the results are published at the height of election season during even-numbered years. The Poll is also impartial, being that its interviews are conducted by an outside organization and the questions are written in such a way that they follow survey design best practices, never misleading respondents or favoring a specific answer. Our question, then, lies on the "accurate" part of the mission—how can the Arkansas Poll be determined as accurate, and can we see that it has been so?

[29] Ibid.

To begin analyzing this question, we can first look to the Arkansas Poll's history as an election pollster—that is, how well they have been able to reflect the voting population of Arkansas to poll an election before it takes place. As mentioned in Chapter 1, this element of political polling is what garners the most media attention, especially during election season, and for good reason: the American people want to know the election results before they even happen, and pre-election polls are necessary for making these predictions. For this measure, we will be focusing specifically on presidential elections as these are the most high-profile questions asked in the Arkansas Poll, and furthermore, unlike issue questions, we have an actual parameter measure for this: the election results themselves.

Now, as already stated, Arkansas presidential elections are not hard to predict in terms of the winner as the state has not gone to a Democratic candidate since the 20th Century. It should come as no surprise that the Arkansas Poll has correctly called the winner every time. That being said, the Arkansas Poll has indeed been very consistent in its accuracy of voting percentages. Presented below are the Arkansas Poll's results compared to the actual election results for the past three presidential elections:

Year	AR Poll (reported)	Actual Results (rounded)	Difference (AR-Actual)	AR Poll Error Margin
2020	Democrat: 32% Republican: 65% Other: 3%	Democrat: 35% Republican: 62% Other: 3%	Democrat: -3% Republican: +3% Other: 0%	+/- 3.9%
2016	Democrat: 36% Republican: 59% Other: 4%	Democrat: 34% Republican: 60% Other: 6%	Democrat:+2% Republican: -1% Other: -2%	+/- 4.1%
2012	Democrat: 31% Republican: 58% Other: 11%	Democrat: 37% Republican: 61% Other: 2%	Democrat: -6% Republican: -3% Other: +9%	+/- 4.0%

Figure 3.2: 2012, 2016, and 2020 Arkansas Poll/Election Comparisons.[30] [31] [32]

In each of these past elections, the Arkansas Poll has predicted the election results within a single-digit percentage point. In both 2016 and 2020, their prediction even fell within their margin of error. From the looks of it, there may be evidence to show that the Arkansas Poll's coverage of the presidential election using very likely voters may even be growing more accurate given the results of 2012 compared to those of 2016 and 2020, but with only these three elections, we cannot make immediate conclusions. It should also be noted that the presidential election results cannot actually be a perfect measure of accuracy for these predictions—the Arkansas Poll is conducted in October and therefore is predicting the percentage of candidate support among the Arkansas voting population at that time rather than on election day, and therefore cannot account for changes between the survey and the election such as last-minute decisions by undecided voters. Despite this, the Arkansas Poll is still used as a factor in election

[30] Parry, *The Arkansas Poll, 2020: Summary Report.*
[31] Parry, *The Arkansas Poll, 2016: Summary Report.*
[32] Parry, *The Arkansas Poll, 2012: Summary Report.*

prediction for polling aggregates such as FiveThirtyEight, and the election results are the closest benchmark we have to an actual measure of the population parameter for this question.

We must also devote ample attention to the fact that this question of presidential choice in the Arkansas Poll does not use the entire respondent sample, but instead, only the respondents marked as “very likely voters.” While it is the Arkansas Poll’s goal to survey the public opinion of the entire Arkansas population, it is in the best interest of any election pollster to only include the answers of likely voters in the sample used for election-specific questions. Doing so is the best way to ensure results that will be closest to the actual election as accounting for those respondents who do not intend to vote could significantly skew the data. Unfortunately, we do not have population parameters describing the characteristics of this “likely voting population” the same way that we have demographics of the entire population through the U.S. Census, so while only using likely voters for the election-specific questions will typically provide more accurate results, we have no way of knowing whether the sample of likely voters matches the likely voting population of Arkansas.

We have examined the Arkansas Poll’s performance as an election pollster, but that only covers 3-5 questions (presidential choice, senator choice, ballot measures) on any given election year’s corresponding poll. We must also consider the rest of the poll, which reports on views and opinions covering various political issues using the entire sample rather than just the likely voters. This means that we must now ask: is the Arkansas Poll achieving its goal of accurately reflecting the entire state population through its survey sample? The inclusion of the demographic comparison chart in every yearly report is a worthwhile step in reaching this goal—it adds an extra layer of context to the report by allowing readers and researchers to analyze the findings with an understanding of the discrepancies between the sample and population demographics. For example, a reader of the 2020 Arkansas Poll report might note that while 77% of the sample has been noted to feel that Arkansas is heading in the right direction, that may not match the actual population total since 32% of the sample consists of college graduates while only 22% of the state are college graduates (and an entire list of other demographic differences).

In order to accurately reflect the views of the Arkansas population, one would hope that the Arkansas Poll sample would demographically match the state's population, but achieving that level of sample accuracy tends to be quite difficult when completing a random sample of a political survey such as this. While normally a large enough sample size would negate this problem, there are many political trends which may contribute to this difficulty. For example, recent years in particular have shown a growing distrust in media (and specifically in surveys and pollsters) by some groups, but an eagerness to embrace said surveys in other groups. In 2019, a study performed by the University of Arkansas found that high levels of trust in journalists had hit a historic low in the state of Arkansas, with many of those responsible for rating journalists with low-trust being Republicans.[33] On top of this, a report from *Data for Progress* polling firm found those citizens identifying as liberals were not only more trusting of polls but much eager to seek them out than their conservative counterparts.[34] While we may not be able to pinpoint the exact problem in our modern political environment contributing to these polling hardships, we can clearly see that despite large sample sizes, they continue to persist.

Because of this practically unlimited potential for problems in creating an accurate sample while remaining random in respondent selection, the Arkansas Poll has opted for years to instead publish their sample as is along with their comparison of state population demographics. However, these sample discrepancies are not a new subject in statistics; in fact, these problems

---

[33] Parry, et al., "Journalists and the Red-State Voter."
[34] McAuliffe, et al., *2020 Polling Retrospective*.

have been such a point of discretion in surveys that statisticians have developed methods for their potential remedy. Now, we will examine the survey process with an eye toward the Arkansas Poll in hopes of finding an applicable method for addressing what these yearly samples lack in representation.

3.3 Survey Design 101

In statistics, “survey design” is the subject of creating polls and surveys in such ways that they reduce potential biases which may arise in the sampling process. Major topics in survey design include the wording of survey questions, the method in which the survey is delivered or conducted, and the process used for obtaining the sample.[35] With our study of the Arkansas Poll’s representation of the state’s population, we are primarily concerned with the survey sample.

At its most basic level, a survey is used to estimate an unknown population measure, or parameter, from a sample of the target population. The same measure of interest is obtained from the sample and then serves as our estimate for the target population parameter; however, this can only be accurate assuming the sample is representative of the population.[36] When it comes to the Arkansas Poll, we want the sample to reflect the population because the goal of the Poll is to provide public opinion information for the entire population, but in statistical terms, a survey’s results become less statistically significant the further their sample is from the population in terms of key demographics, meaning it becomes harder to trust the survey’s results.[37]

[35] Laaksonen, 27.
[36] Laaksonen, 50.
[37] Laaksonen, 113.

To address the concerns of sampling design, best practices have been established across the statistics community as to what makes a "good" sample. Firstly, as has been stated, a good sample is one that is representative. Statistical consultant and writer Sharon Lohr explains that we cannot know whether a sample is a perfect representation of the population without first measuring the entire target population. Instead, she describes this representativeness as "the sense that characteristics of interest in the population can be estimated from the sample with a known degree of certainty."[38] Note that with the Arkansas Poll, there actually does exist a measurement of the population, the U.S. Census for Arkansas, meaning that we can always compare the sample to the overall population (as the Poll does in each yearly issue). Even with this need for representativeness, however, surveyors need to be careful to avoid "judgment samples" in which respondents are deliberately selected, rather than randomly, in order to garner a sample that matches the population.[39] This introduces unnecessary bias which can skew the data and lead to a lack of precision. Surveys must also avoid overcoverage and undercoverage, in which samples include those units outside of the population or do not include certain groups within the population, respectively.[40]

Another practice common to political surveys specifically is the selection of likely voters within the sample. We have discussed already that the Arkansas Poll designates certain respondents as "very likely voters" (referred to from now on simply as "likely voters" for simplicity) for a few of its election-specific questions, including that of presidential choice. The method used for likely voter selection across political surveys varies, but for the Arkansas Poll, one question determines a respondent's inclusion in the category: "How likely are you to vote in

[38] Lohr, 3.
[39] Lohr, 5.
[40] Lohr, 6.

the election next month?" Those respondents answering "very likely" are then designated as likely voters with their answers contributing to the election-specific questions, while all other answers are left out of the likely voter pool; in other words, the likely voters in the Arkansas Poll are self-identified as such. Other polling organizations have more rigorous standards for their selection of likely voters—for example, Pew Research uses an entire list of questions regarding voting habits and knowledge of candidates in addition to a self-identifying question to give their respondents "likely voting scores" which determine their eligibility. That being said, as Pew Research is a much more expansive organization with more time and resources than the Arkansas Poll, it is understandable why they are capable of doing so. Furthermore, the Arkansas Poll's method, while less extensive, has not been a faltering point seeing as the Poll made election predictions within their margin of error in both 2016 and 2020.

Thus far, we have examined the best practices of survey design with a specific eye toward sampling in the context of the Arkansas Poll, but there is one major area of sampling error which we have yet to mention, largely because it constitutes the main cause of sample bias not only in the Arkansas Poll but in most political surveys: nonresponse bias. Nonresponse is a type of bias created by the lack of answers from specific selected respondents, either entirely or for specific questions. When large numbers of contacted individuals are not responding to the survey, this can have a major effect on the results of the poll. This is frequently what causes sample-population discrepancies such as in the Arkansas Poll, and it tends to happen particularly in political surveys due to their subject matter. According to statistician Seppo Laaksonen, nonresponse can occur "if a potential respondent is not sufficiently motivated to participate…or he or she does not like the questions in a questionnaire or considers them incorrect or invalid."[41]

[41] Laaksonen, 28.

Given the previously mentioned study regarding growing media and pollster distrust, this is obviously a problem that would affect the Arkansas Poll specifically.

Whatever the reasoning for its occurrence, nonresponse can be a major contributor of survey bias seriously affecting results. This is made especially so if there are patterns among the non-responding units, such as specific demographic groups being more likely to refuse participation and therefore being underrepresented in the sample.[42] Even worse, accounting for nonresponse in the survey design is nearly impossible; we have already discussed the dangers of a judgment sample in which the surveyor hand-picks the respondents to fit a representation, and increasing the sample size cannot remedy the situation since nonresponse will continue at the same rate.[43] Luckily, in an effort to overcome the unavoidable problems of nonresponsive and unrepresentative samples, statisticians have long made use of one specific strategy created specifically for better representing a target population. This statistical strategy is known as weighting, and it constitutes our main opportunity for improvement of the Arkansas Poll's sample.

3.4 Weighting 101

In our brief introduction to survey design, we discussed multiple methods of accounting for potential sampling bias that are factored into the methods and processes of sample selection, but in our mission to apply greater accuracy strategies to the Arkansas Poll, "sample weighting" is the strategy which gives us the most opportunity for changing the results in a meaningful way. The first thing to note regarding weighting is that this survey strategy happens after the sample has already been polled and is technically not an element of survey design. Instead, weighting is

[42] Lohr, 331.
[43] Ibid.

part of the post-survey analysis which helps to develop our findings from the survey's raw results.

Weighting can be defined as the process of fitting our sample to the population in a way that makes results more representative of key demographics. In other words, weighting is a strategy for making the sample "look like" the population and observing the changes in the survey results which follow. We project the demographic proportions of the target population onto those of the sample, and the answers of respondents will either increase or decrease in their effect on the result depending on what happens in this projection.[44] To do this, every respondent in the sample is assigned an individual "weight" according to their demographics. These weights correspond to the respondent's probability of selection for the survey—the less likely a respondent is to be selected for the sample from the population according to their demographics, the higher their assigned weight will be, and vice versa for those more likely to be selected. These weights then determine how much a particular respondent's survey answers are worth across the entire sample, so those with higher weights will affect the results more than those with lower weights in order to account for over- and under-representation. Assuming best practices and standards are followed (discussed in Chapter 4), these results should be more accurate to the actual target population parameter of interest.

Weighting processes can vary between surveys and fields of statistics, but the general procedure follows as such:

1. The survey is conducted and meets the necessary standards required of the respective field.

[44] Valliant and Dever, *Survey Weights*, 11.

2. After the raw results have been gathered, it is time to analyze the sample by comparing it to the target population. For this to be possible, it is necessary to obtain key demographic information (such as gender, age, race, etc.) of the target population which corresponds to the known demographics of the survey sample. Where a population census is unavailable for the target population, estimates may suffice given that they are of high quality. Without available population information, weighting will not be possible. Once the key demographic variables are decided, the sample and target populations are compared according to their proportion of each. If the proportions differ by a significant amount, then weighting will likely be necessary for more accurate results. Keep in mind that a “significant amount” is not necessarily an established or agreed-upon percentage difference but is left to the discretion of the surveyor.
3. If deemed necessary, weighting can now begin. Required data includes the target population demographics, sample demographics, and sample responses (raw survey results). Individual weights are calculated and assigned for every respondent based on their demographics. There are a variety of ways in which these weights can be calculated (discusses in Chapter 4), but they are typically based on each respondent’s probability of inclusion in the sample and how far apart the sample and population proportions are for the corresponding variables.
4. Weighted results are calculated with each respondent being multiplied by the factor of their assigned weight. Now, the demographic proportions of the sample and target population should match, and we have entirely new survey results for this weighted sample.

5. The weighted survey results are analyzed as a more representative estimate of the target population.

Weighting, as a statistical strategy, is recommended across the field of surveying in the common instance of nonresponse creating unrepresentative samples. When faced with high levels of nonresponse, Chang and Butar (2012) note that simply ignoring to effectively manipulate the data can create poor survey quality and inaccurate results. One might then be compelled to address this problem in the survey design, but as already discussed, nonresponse is a common problem that undoubtedly arises in any survey, just to varying degrees.[45] Pew Research Center has advised researchers on the characteristics of "good" and "bad" polls, noting that those of lesser quality tended to be the polls which took no steps to adjust or weight their samples to fit the target population. Not only is this useful to know, but the information guiding this came from Pew Research Center's study of 2016 state polls similar to (and potentially even including) the Arkansas Poll.[46] Furthermore, as noted earlier, the polling aggregate FiveThirtyEight gives its highest pollster ratings to those surveys which use statistical strategies such as weighting to report an accurate sample.[47]

We now know that weighting is one of the most frequently used methods for creating a more representative survey sample when such a goal cannot be met with strategies of survey design. It is also a respected and encouraged step in the political polling process according to some of the field's most credible sources. Therefore, in looking for a way to make the Arkansas Poll's sample more reflective of the overall state population, our new goal is to weight the

---

[45] Chang and Butar, "Weighting Methods in Survey Sampling."
[46] Dean, et al., "Field Guide to Polling."
[47] Silver, "How FiveThirtyEight Calculates Pollster Ratings."

Arkansas Poll's results. Despite the best efforts of the Arkansas Poll to achieve an unbiased survey design, nonresponse still arises within the sample as it happens to do in most major surveys. With all of the necessary data on hand, we will weight the Arkansas Poll from the most recent presidential elections, 2016 and 2020, to see if the results become more accurate to the actual election numbers. Then, we can examine the changes caused by weighting to many of the Arkansas Poll's most prominent issue questions. While we know that the Arkansas Poll has been very accurate to the election results in recent years, our goal is to observe the effects caused by making its sample more representative of the Arkansas population.

3.5 Further Application

Before we begin the weighting process, it is necessary to note the state-specific nature of this experiment. We are strictly focusing on Arkansas, which although similar to some other states in terms of demographics, should still be seen as its own unique location for the application of this polling strategy. That being said, the results from these trials will likely give us some hint as to how weighting may affect the polling results of other states or even national-level polls. This study is therefore important to the state of Arkansas's polling accuracy but should be applicable to other states as a jumping-off point for their own polling studies.

Chapter 4 – Exploring the Methodology

From the beginning, we have been examining the Arkansas Poll and its dedication throughout the past couple decades to creating an accurate reflection of the public opinion of Arkansas. Now that we have established weighting as a viable and recommended statistical strategy for increasing accuracy in political polls, our goal is to weight the results of the Arkansas Poll and observe the effects. The hope, of course, is that this creates a more accurate reflection of the state population in terms of sample demographics.

Before the weighting process can begin, it is necessary to explore the methodology behind weighting survey samples. Weighting is a very common strategy in surveying with lots of literature and trials to its name, so determining the best process for Arkansas Poll-specific trials requires a number of significant choices. Therefore, in order to create the best possible system for accurate results with the least chance for introduction of additional bias, we will explore subjects including the choice of demographic variables, methods of weight calculation, and accuracy to be prepared for our experiment.

4.1 Choice of Variables

Of these many considerations on the table during the weighting process, perhaps one of the most important is the choice of which key demographics will be our variables for calculating weights. These are the variables by which we measure the differences between sample and population proportions. For example, we may compare a sample and its target population in terms of gender. If the sample is overwhelmingly male but the target population is known to be relatively even, then weighting by gender may be a good choice.

In any instance of weighting, these variables need to be known measures. In other words, they need to be readily accessible for the respondents. This is why, during the survey design stage, it is important to include a number of background questions regarding demographics and not necessarily pertaining to the actual subject matter of the survey. In our same example using gender, the point of the survey may have been to estimate a population's support for a new immigration law. Despite the fact that there is no immediately recognizable connection between gender and support for a specific immigration law, asking the question in the survey provides us with a variable that can later be used for weighting. Of course, these demographic variables must also be known (or estimated with high confidence) for the entire target population—without such, there is no way to calculate the weights.

As long as they are known measures for the population and the sample, these variables can be anything which may affect a respondent's answer to our survey question of interest. Therefore, these variables by which we weight can be either categorical (discrete) variables or continuous variables. In statistics, a categorical variable is one for which there is a set number of potential values or answers—for example, a respondent's race or gender. Continuous variables, on the other hand, can take the form of any quantitative value without having a limited number of potential responses, such as weight or body temperature. Categorical variables are more frequently used and are typically much easier to deal with when weighting; most demographics that are recorded in social and political surveys are those which fall into categories, such as the aforementioned gender and race.[48] Plus, when a variable has a fixed number of potential

[48] Gelman, "Struggles with Survey Weighting."

responses, then weighting by that variable alone will produce the same weight for every respondent in each respective category. Continuous variables are much less frequently used because they are rarer and more difficult for weighting but can in many instances produce more accurate survey results when continuous-specific weighting methods are utilized (discussed in Section 4.2).

When it comes to political polls, choice of variables is paramount. There is discussion across the field of political polling regarding which variables will produce the most accurate results, and in 2018, Pew Research Center conducted a number of studies to determine such. After weighting by both demographic variables (age, race, gender, etc.) and political variables (party identification, etc.), it was found that while demographic variables provided a consistent effect, the results when weighted by political variables tended to be more extreme but also more varied.[49] Not only that, but the statistical methods of weighting always had less effect than did the choice of variables.[50] Furthermore, looking back on the accuracy of 2016 pre-election polls, it is widely agreed that a lack of weighting by education, a demographic variable rather than political, is what led to much of the polling bias seen that year—and that weighting by education level would have likely provided more accurate results.[51]

While Arkansas is not historically known for its representation of different demographic backgrounds, the state has been quickly diversifying in recent years. In fact, the state's demographic representation is constantly changing, with trends pointing toward increases in the minority population every year. The U.S. Census's "diversity index," a measure of the state's racial and ethnic minority population, rose a whopping 8.2 percentage points between 41.6% in

---

[49] Mercer, et al., "Weighting Online Opt-In Samples."
[50] Ibid.
[51] Dean, et al., "Field Guide to Polling."

2010 to 49.8% in 2020.[52] This identifies the need for weighting in a state where demographics are changing at such a substantial rate. With that growing diversity across multiple demographics, we will likely be weighting with multiple variables on more than one occasion. Luckily, there are multiple weighting methods designed specifically with this in mind. When possible, it is typically recommended to do so in order to get the most effective weights for each respondent in the sample.[53]

Now, in selecting the variables to be used in our own experiment, we have an established set of expectations to guide the decision. First, as the most basic requirement, these variables need to be readily available—in other words, they need to be known for the sample respondents, and they need to be known for the state of Arkansas as a whole (most likely meaning "included in the U.S. Census"). Next, while this may be obvious, it is worth noting that these variables need to be those which could potentially influence a respondent's opinion on political issues and therefore affect their answers to the poll. Luckily, most of the widely collected demographic variables fall into this category. Given both our knowledge of changing racial demographics in Arkansas and the conclusions regarding education level's effect in 2016, both race and education need to be included as variables.

Using these requirements, the following variables have been selected for inclusion in weighting trials:

1. Congressional District
2. Gender
3. Age
4. Race

[52] U.S. Census, *Arkansas Population Topped 3 Million.*
[53] Mercer, et al., "Weighting Online Opt-In Samples."

5. Educational Attainment
6. Income Category

These variables constitute some of the most readily available choices, each of which is covered by both the Arkansas Poll sample and the U.S. Census. Each of these variables also differs by at least a significant amount between the Arkansas Poll sample and the state's population demographics (full comparison tables are given in Appendix A). Furthermore, the idea of political variables was also discussed in addition to demographic variables; however, the lack of consistency among their effects on the data along with their lack of availability in the Census made them less ideal. Moving forward, the goal is to weight using all six of these variables individually, all at once, and using different combinations.

4.2 Weighting Methods

As mentioned in the previous section, Pew Research Center's study of the 2016 elections established variable choice as the most important aspect of weighting. While this may be true, this does not mean we can turn a blind eye to the subject of weighting methods. There is a multitude of developed methods for calculating sample weights, and the decision of which to use for our trials with the Arkansas Poll is by all means a necessary consideration. We will discuss a few of the most common methods in order to decide which fits best with our weighting goals.

As it was discussed earlier, the calculation of sample weights produces an individual weight for each respondent. This weight is should reflect the chances that an individual with the same demographics is chosen for the sample from the population while also factoring in how well-represented those demographics are within the sample. For example, if Group A makes up a larger proportion of the population than it does in the sample, then respondents in Group B will be given a weight greater than 1; similarly, if Group B is overrepresented in the sample, its respondents will receive weights less than 1. Doing so allows the respondents from underrepresented groups to contribute more to the sample in order to make up for the lack of representation. With this process, it is important to note the relationship between a respondent and the other members of their weighted demographic: weighting allows respondents from underrepresented groups to account for the missing respondents from their demographic group, but we are not assuming that group is homogenous in their views. Such an assumption would defeat the purpose of weighting and polling in general—instead, we are acknowledging the potential and often very real existence of trends throughout demographic groups.

The most basic method of weighting is known as cell-by-cell weighting, or simply "cell weighting." For this method, we compare a variable (or combination of variables) by its proportions in both the sample and population, then find the individual weight for those respondents by dividing the population total by the sample total.[54] By doing this, we will create weights greater than 1 for variables with larger population proportions than sample proportions (underrepresented variables) and weights below 1 for variables with larger sample proportions than population proportions (overrepresented variables). A variable with equal proportions will return a weight of exactly 1.

Cell weighting is typically utilized for its simplicity. While it is a quick process when weighting by one variable, it remains relatively simple for variable combinations. Let us assume that we are weighting by two demographic variables (X and Y) and that each variable is categorical with three potential responses (1, 2, and 3). We can create tables of the sample and population proportions, as so:

[54] Kalton and Flores-Cervantes, "Weighting Methods."

Sample Proportions					*Population Proportions*			
	X1	**X2**	**X3**			**X1**	**X2**	**X3**
Y1	20%	5%	8%		**Y1**	25%	6%	8%
Y2	30%	10%	7%		**Y2**	20%	5%	9%
Y3	13%	6%	1%		**Y3**	5%	8%	14%

Figure 4.1: Cell-by-Cell Weighting Example Proportions (shading corresponds to the following paragraph's examples).

In this scenario, the intersection of X1 and Y1 is weighted according to both proportions, the intersection of X2 and Y1 is weighted, and so on for all nine cells. Respondents of X1 and Y1 are underrepresented and therefore have a weight of 25/20 = 1.25, respondents of X2 and Y2 are overrepresented and receive a weight of 5/10 = 0.5, and respondents of X3 and Y1 are represented exactly and are given weights of 8/8 = 1. This allows the sample to conform to population proportions and become more representative.

However, more representative is not necessarily indicative of more accurate results. When samples are weighted using cell-by-cell, major proportion differences can create unreasonably large weights which can increase variance.[55] As we will discuss later in the chapter, increases in variance mean that our answers are less precise, or in other words, we lose confidence as the range of potential answers becomes much larger. In the above example, the respondents of demographics X3 and Y3 receive weights of 14, an incredibly large number in the

[55] Ibid.

context of sample weights. This constitutes one of the most notable disadvantages to cell weighting, especially considering that our primary concern is of a state poll which often sees large proportion differences. Not only this, but cell weighting using these cross-sections of variables may create too many potential demographic groups with many of them not being large enough to be significantly sampled at all.[56] Furthermore, weighting through this method would be impossible using the Arkansas Poll since our population totals come from the U.S. Census which includes totals for our variables of choice but not for the combinations of those variables. Given this, we will need to use a method with less potential for variability and with which we can use the available Census data.

"Raking" is an alternative and more complex method of calculating sample weights which accounts for those disadvantages of cell-by-cell weighting. Also known as iterative proportional fitting, raking can account for multiple variables without the need for known population totals of those cross-combinations.[57] This is done by fitting the sample proportions to the population proportions using one variable at a time, applying the relative proportion of each potential response in a variable to the sample based on those proportions in the population.[58] In our earlier example using the combination of two variables X and Y (Figure 4.1), the sample row proportions would first be conformed to the population row proportions, then the sample column proportions would be conformed to the population column proportions, and the process would be repeated until the sample reaches "convergence" by matching those totals in the population.[59] Keeping the two-variable example in mind, this method of weighting gets its name from an analogy to gardening in which soil is "raked" in multiple directions and repeated until the soil is

[56] Lumley, 139.
[57] Ibid.
[58] Battaglia, et al., "Practical Considerations in Raking."
[59] Kalton and Flores-Cervantes, "Weighting Methods."

smooth.[60] The actual process of weighting is obviously much more involved and technical than that of cell-by-cell weighting. For this reason, complex algorithms are designed to make raking possible using systems such as R or Stata.[61]

The benefits of raking are numerous. Most of the potential problems introduced in cell-by-cell weighting are ruled out in raking; variable proportions in the population only need to be known for single variables even in the case of weighting by combinations, and the more drawn-out method of conforming proportions includes less potential for increasing variance. Raking is therefore especially good for weighting combinations of variables, making it especially useful for our experiment. Pew Research Center has also previously supported raking as one of the best strategies for weighting political polls--it is their weighting method of choice, and their earlier-mentioned study of the 2016 Election concluded that raking worked just as well as other complex methods.[62]

Whereas raking is a weighting method specifically designed for use with categorical variables, there exist a multitude of other weighting methods which can be more readily applied to continuous variables. Linear weighting and generalized regression (GREG) weighting provide algorithms for such usage, but seeing as our variables are categorical (with age broken into age groups by decade, as is given in the U.S. Census), raking provides the best and most widely used method for our situation.[63]

[60] Battaglia, et al., "Tips and Tricks for Raking."
[61] Kolenikov, "Calibrating Survey Data."
[62] Mercer, et al., "Weighting Online Opt-In Samples."
[63] Kalton and Flores-Cervantes, "Weighting Methods."

4.3 Imputation

As it turns out, weighting is not the only strategy used to adjust for nonresponse in survey samples. Earlier, we defined nonresponse as the missing data created by a lack of answers from respondents who were selected but only partially included or completely dropped from the sample, with this problem occurring in almost all surveys regardless of the target population.[64] This means that we can then split the problem of nonresponse into two separate categories.

The first category, unit nonresponse, encapsulates the missing data resulting from selected individuals who do not respond to the survey or portions of the population whom the survey intends but ultimately fails to reach.[65] Alternatively, this is when entire "units," or representative respondents, are missing from the sample. This is the category of nonresponse which frequently creates a disparity between the sample and population proportions—those groups suffering greater nonresponse are going to be underrepresented. Given this effect, we address unit nonresponse through weighting.

The other category is item nonresponse, a form of missing data which takes place when units in the sample provide partial responses.[66] When significant numbers of respondents answer some of the questions but fail to respond to others, item nonresponse is created. This can be for a multitude of reasons including participant opposition to a question or interviewer confusion. For example, a respondent to the Arkansas Poll might answer all of their demographic questions but fail to weigh in on their opinion of gun control legislation. Normally, this particular respondent would be removed from the sample when reporting this question. However, statisticians often make use of another form of nonresponse adjustment alternative to weighting.

[64] Valliant, 31.
[65] Lumley, 185.
[66] Ibid.

This strategy, called imputation, addresses item nonresponse by using the existing answers of partial respondents to predict the answers missing from those respondents. When a respondent is missing an answer for a particular item, a replacement value for that answer is derived from other respondents who "look alike," meaning they share similar demographics and answers to other questions in the survey. This replacement value is then imputed, or added, to the missing item. Not only does this allow surveyors to account for the potential bias created by this form of nonresponse, but it creates a "clean" set of data for ease of observations and calculations.[67] While both strategies account for nonresponse, weighting applies to unit nonresponse by projecting population proportions onto the sample proportions and giving more weight to existing responses, while imputation applies to item nonresponse not by giving more weight to any answers but by inserting missing answers to create a "complete" sample.[68]

As with weighting, there are multiple methods for selecting the imputed answers. Deductive imputation is used for logically deciding upon an imputed answer, such as matching those with similar aspects within the same survey or using previous answers which logically rule out later responses.[69] Continuous variables may lead surveyors to simply resort to using the mean for said variable in every missing response, while some surveys will use data collected in other surveys to inform the imputation of their own.[70]

Imputation has a number of benefits in its potential usage. As has been emphasized, it is especially good for reducing the effects of bias associated with item nonresponse. It has a number of potential uses, not only for when data is completely missing but also when potential

[67] Lohr, 346.
[68] Valliant, 11.
[69] Lohr, 347.
[70] Lohr, 350.

items are only partially answered or are assuredly incorrect for some reason or another.[71] Plus, the aforementioned idea of a "clean" data set without holes is alluring to any data scientist.

It is also necessary to consider the numerous potential drawbacks of imputation in certain situations. For starters, imputation can be quite useful to the institutions conducting the surveys, but outside researchers tend to have less luck in creating accurate imputations—they typically do not have the same level of information available regarding the respondents as the insiders and furthermore are less familiar with the data collection process.[72] High amounts of item nonresponse across multiple variables are typically handled through predictive software because they would be impossible to be dealt with by hand, but many of these imputation software suffer the same "outsider" problem and do not create accurate results.[73] With the high amount of imputed answers which tend to be general or close to the mean, variance almost always decreases during imputation which may be inaccurate to the actual to the true variance.[74]Each of these disadvantages are widened when we consider the detrimental effects of "bad" imputation. Methods of imputation need to be precise and accurate, and when they are neither, this strategy can severely harm a survey's results.

With the extreme potential for negative effects, the decision to utilize imputation for more accurate results can be demonstrated as such: imputation should only be used when a sample with imputed values would be more valuable than it would be without.[75] If no notable difference will accompany the results when imputation is implemented, then there is no reason to do so since the costs would be too high.

---

[71] Laaksonen, 157.
[72] Laaksonen, 158.
[73] Laaksonen, 171
[74] Lohr 350.
[75] Laaksonen, 158.

In the case of our weighting the Arkansas Poll, this reasoning implies that imputation would do more harm than good. We would be imputing outside the original surveying institution, and lack of professional experience with imputation means we would need to utilize a generalized imputation software. Furthermore, as we will discuss in the next chapter, our sample retains a significantly large size even when listwise deletion is used to remove those respondents suffering item nonresponse. While it is important to delve into this subject when nonresponse adjustment is on the table, our conclusions in such show that this experiment is not an ideal situation for the implementation of imputation. Instead, weighting takes center stage.

4.4 Measuring Accuracy

The entire purpose of surveys is to estimate a population measure that is currently unknown; if we knew the parameter, there would be no reason for a statistical estimate. Therefore, it may seem somewhat paradoxical to "measure the accuracy" of a survey's results since there is no benchmark for comparison. It is hard to know whether a survey is accurate to the true measure it seeks to replicate, and we typically view the most accurate surveys to be those which have the least amount of potential or observable bias.

In the case of the Arkansas Poll, this is not completely true. As it so happens, we have a population measure for one aspect of this experiment, that being the Arkansas Poll's question of presidential choice. Since the Arkansas Poll is conducted in October and we are observing the Polls of 2016 and 2020, the presidential election in November serves as the accuracy measure. Once the election results are announced, the Arkansas Poll results can be compared side-by-side with the Arkansas election numbers, showing just how close the Poll came to an accurate estimate. For us, this means that when we weight the Arkansas Poll results, we can compare our weighted results to those same election numbers and gain an understanding of potential improvement. That being said, it is important to note that this "accuracy measure" of the Arkansas Poll's presidential choice question is not, in actuality, a measure of the Poll's accuracy to the exact numbers of support at the time of the survey—instead, it is a measure of the Poll's accuracy as an election predictor. The actual population measure of presidential candidate support likely changes, at least slightly, between the survey's completion in October and Election Day in November. Regardless of this, there is no population measure for that specific moment in October, so the election results remain the best possible measure of accuracy for this question.

When considering our weighting of the Arkansas Poll's presidential choice question, we must also confront the distinction between the target population and the demographic parameters available. For our weighting trials, the best available demographic information describing the population comes from the U.S. Census. The Census has exact data for the state of Arkansas's overall population every ten years, and every year in between, the Census Bureau publishes population estimates using a variety of governmental data sources tracking such changes. This constitutes the population information which will be used to weight the sample, but a potential problem arises when we consider the target population of the presidential choice question. As mentioned in the previous chapter, the Arkansas Poll surveys a random sample of Arkansans with the hope of reflecting the state's population, but for the question of presidential choice, only those respondents considered "very likely voters" are considered. For this question in particular, the target population is the likely voters rather than the entire state population. This likely voting population probably does not line up exactly with the overall state population demographically, so is it a problem to apply information about the overall population to this sample? It certainly is not ideal, but known demographics of the state's likely voting population are unavailable as there is no Census question identifying likely voters. Therefore, we work with the resources available and can by all means test whether this process can still yield more accurate results.

When we turn away from the presidential choice and instead to the issue questions, we are immediately confronted by a lack of an accuracy measure. There is no election data or population measures for questions regarding a respondent's general opinion on gun control or abortion the same way there is for a respondent's support for Donald Trump's presidential run. So, the purpose of the experiment shifts when these issue questions are considered. Rather than trying to make the results more accurate, we can take what was learned from the presidential choice weighting and apply it to our weighting of these questions for the purpose of observing their effects alone.

Despite this, we must still ask: is there an alternative or partial method of determining significance of weighted issue question results? In fact, there are a few. For starters, there is no target population dichotomy when weighting the issue questions. Respondents are no longer removed from the sample based on the likelihood of their voting—these questions are not purposed with predicting an election but instead reflecting the public opinion of the entire population. The entire state population is the target population, and our Census information reflects that target population. Because of this, we can have more confidence that our weights are allowing the sample to be more accurate to the population. However, this is not a quantitative measure but merely a factor reducing bias.

If we want to discern a quantitative measure describing statistical significance of our weights, we can look to the data variance. Variance is an important aspect of any survey's results, and it is always going to be affected during the weighting process. In order to explore this, we need to establish the important statistical discussion of precision versus accuracy. Simply put, accuracy, as it has been used thus far, refers to the closeness of our estimation to the true population value. When we weight the Arkansas Poll sample, we are attempting to increase the accuracy of the Poll. Where accuracy refers to the closeness of an estimation to the true measure, precision on the other hand refers to the closeness of individual measurements or estimations. Concerning surveys, accuracy is about finding the best estimate for the population parameter, and precision describes our confidence in that estimation.

At its most simple, weighting introduces new data into the survey, with that new data being the population parameters for demographics. With this introduction of entirely new data, the survey results see an increase in variance, which in turn expands our confidence interval. Therefore, when we weight survey data for a better estimation of the population parameter, we are sacrificing precision for accuracy.[76] Ideally, any surveyor's goal should be to increase both accuracy and precision whenever possible, but with the potential trade-off that comes with weighting, we must ensure that whatever increase in accuracy we have to gain is worth the loss of confidence.

There is no set or agreed-upon increase in variance that is "too large," but it should be noted that an increase in variance, and thus a decrease in precision, creates less statistically significant results. Instead of naming a set increase as a cutoff, we need to remain cognizant of the variance increase which accompanies each weighting trial.

[76] DeBell and Krosnick, *Computing Weights for the ANES Survey Data.*

Chapter 5 – Experimentation & Data

After studying the Arkansas Poll and its processes in previous years, we have recognized the potential for a more reflective survey sample within the Poll through the use of weighting. We have explored the potential for improvement and considered the many choices necessary to complete this experiment. Now, the goal can be stated as such: we will perform weighting on the Arkansas Polls of both 2016 and 2020 through the use of U.S. Census data to inform our demographic weights. We will first weight the questions of presidential choice, which can be compared for accuracy, before moving on to the issue questions with informed variable choices to observe the effects of weighting on public opinion.

This chapter consists of a complete record of the experiment itself. Section 5.1 describes the methodology for each of the major steps in the process. Sections 5.2 and 5.3 list the full reports of the results from each of the weighting trials. While some brief observations and takeaways may be pointed out in these sections, the bulk of analysis can be found in Chapter 6.

5.1 Methodology

Sources of Data

In order to create an informative weighting process, it is necessary to clearly define the sources of data and target populations we seek to reflect.[77] We need three sources of data according to the parameters of our experiment: sample data (The Arkansas Poll), target population

[77] Valliant, 152.

demographics (U.S. Census), and target population accuracy measures (2016 and 2020 Presidential Election Totals). It is also necessary to note that while U.S. Census data will be used for the target population demographics, our target population differs in both the presidential choice weighting and the issue weighting: for presidential choice, the target population is Arkansas's likely voters, whereas for issue weighting, the target population is the general public. Because of data availability, we will continue with this data discrepancy but stay aware of the potential for bias.

The Arkansas Poll is published yearly following its completion in October. It is most often referenced using its summary report, which is comprised of the overall results for each question, some information regarding the polling process, and the sample information mentioned in Chapter 3. The Arkansas Poll also posts its entire sample for every year to its respective website, labeling each respondent with an identification number and listing each of their answers in an SPSS file.[78] This full sample is readily accessible and can be downloaded for use in statistical software such as R or Stata—therefore, it is what we will be using in our experiment. General information regarding the samples from 2016 and 2020 is given here:

Year of Poll	Total Respondents	Likely Voters
2020	804	562
2016	800	504

Figure 5.1: General Information of Arkansas Poll Samples, 2020 and 2016.[79] [80]

[78] "The Arkansas Poll" Website.
[79] Parry, *The Arkansas Poll, 2020: Summary Report.*
[80] Parry, *The Arkansas Poll, 2016: Summary Report.*

The U.S. Census is a measure of the entire United States population and is conducted once per decade. However, to keep an accurate flow of data year-to-year, the U.S. Census Bureau publishes Census “estimates” every year determined by the most recent Census combined with additional population information. Because these estimates come from the country’s highest authority on population information, these estimates will be taken as truth for the purpose of the experiment; they are likely to be closer to the actual population at the time of these specific Arkansas Polls than the previous decade’s U.S. Census because of the rapidly growing and changing Arkansas population. For the 2016 Arkansas Poll, demographics will be weighted according to the 2016 Census Estimates. For the 2020 Arkansas Poll, the situation is a bit more convoluted: while 2020 Census data has since been published, it would not have been available during October of 2020, and we will therefore be using the most recent Census data that had been available at the time, which was the 2019 Census Estimates. This circles back to the purpose of the experiment: if we want to replicate the Arkansas Poll’s potential for greater accuracy, we need to use the information that would have been available to them during the survey. The exact Census numbers for our selected variables are featured in Appendix A in comparison tables with Arkansas Poll sample information.

2016 and 2020 Presidential Election results will be used as accuracy measures when weighting the presidential choice questions in each Arkansas Poll. These numbers are widely reported and known, although we originally obtained them from Politico.[81] The numbers are available in Appendix A.

[81] Politico, “Arkansas Presidential Results – 2020.”

Chosen Variables

The following variables have been chosen for weighting the Arkansas Poll:

1. Congressional District
2. Gender
3. Age
4. Race
5. Educational Attainment
6. Income Category

Variables were chosen based on availability and relevance. Each of these six chosen variables could be found in both the Arkansas Poll's published sample data and the U.S. Census for Arkansas. In Chapter 4, it was noted that those variables which will be used for weighting should be only those that have differences across the sample and target populations. To get an idea of this ahead of time, we created tables for comparing these variables from the Arkansas Poll Sample, the Arkansas Poll likely voters, and the Census (available in Appendix A). Across both 2020 and 2016, the Arkansas Poll Samples (both full and divided into likely voters) each differed from the actual population totals by at least a few percentage points, warranting their inclusion in the weighting trials.

All of the selected variables fall into the "demographics" category, while the aforementioned "political" variables are notably missing. While studies such as the Pew Research polling study in 2018 have seen significant results utilizing variables such as party identification, political variables are largely unincluded in the U.S. Census.[82] For consistency's sake, we want to weight only by known population values published by the U.S. Census Bureau.

[82] Mercer, et al., "Weighting Online Opt-In Samples."

As for expectations, it is probable that weighting by education will create an increase in accuracy of the presidential choice and potentially the issue questions. Pew Research cited this as a typically under-sampled group in 2016 for which weighting would have accounted.[83] Each of the other variables with major sample-population discrepancies, namely age and race, will likely have the largest effects on the results—meaning not only that they may provide a better insight into Arkansan public opinion, but that we need to be especially attentive of the design effect causing variance increase. Those with the smallest discrepancies (and less notable trends), namely gender and congressional district, will be noted for exclusion in further trials in the occurrence that their effect on the presidential choice is not significant.

Weighting Method

Our research pointed to raking as the best, most accessible method for carrying out weighting of Arkansas Poll data. Its ability to weight according to multiple variables at once without creating abnormally large sample weights assures this. Furthermore, Pew Research found that it worked just as well, if not better, than other complex weighting methods used in their political poll weighting trials.[84] While an ideal trial might test according to a variety of weighting methods, our primary concern is the variables; the aforementioned Pew Research study reported the variables having a larger effect than the weighting method across multiple trials.[85]

To perform the weighting, all statistical procedures will be done via the programming language R. This programming language is widely used and supported for all varieties of

[83] Ibid.
[84] Ibid.
[85] Ibid.

advanced statistical computing, notable among them being weighting. The code for all computations done in this paper, including weighting, statistical comparison, and graph printing, is included in Appendix C.

As R is one of the most—if not the most—widely used statistical programming language that exists, there are an abundance of raking packages which can be used to perform the weighting trials. In order to choose which we would use, it was necessary to select a raking package which was recommended and meant for a political poll such as the Arkansas Poll. We decided to use the "anesrake" weighting package created by American National Election Studies (ANES). One of the most trusted organizations in the field of polling and elections, ANES receives funding from the National Science Foundation and seeks to serve public opinion research needs across the United States.[86] In 2010, ANES researchers published a report regarding the weighting needs and best practices of public opinion researchers in the United States, citing the immediate need for a standardized weighting algorithm meant specifically for public opinion polling.[87] This birthed the "anesrake" weighting package, part of the ANES Weighting Algorithm (AWA). The AWA is commonly used in political polls for the purpose of weighting through raking and allows for a streamlined approach to the process in R. With the program readily available, AWA gave us the perfect method for raking and weighting the Arkansas Poll sample.[88]

[86] *American National Election Studies* Website.
[87] Debell and Krosnick, *Computing Weights for the ANES Survey Data.*
[88] Pasek, *ANES Weighting Algorithm*

Missing Data

Our weighting experiment is an attempt to account for unit nonresponse within the Arkansas Poll sample. Item nonresponse, on the other hand, will be prevalent no matter what. There are two potential strategies for dealing with such: imputation and listwise deletion.

As discussed in Chapter 4, we are choosing not to use imputation during this experiment. As we would be performing imputation outside the original surveying institution and using a computational package in R to do so, research shows that this would likely do more harm than good. Instead, we will be utilizing listwise deletion as our method of choice for dealing with missing items. In this method, rather than keeping the respondents with missing data and filling in potential answers, we will be removing those with specific missing variables from the sample of any weighting trial which utilizes said variable. The initial fear in this method is that tossing out a large amount of data due to deletion could bias the results; that fear, however, is largely made up for by the fact that our sample sizes are large enough so as not to be dangerously lowered through deletion of respondents with missing data. To make sure this subject is not forgotten, we will print the total number of sample respondents included in every trial during the experimentation process in order to note any trials with particular smaller samples.

On the subject of paying special attention, a major potential for bias is introduced if listwise deletion disproportionately affects a certain group.[89] For instance, imagine that we are completing a weighting trial for the variable of race where the sample is equal parts male and female. If those 75% of the respondents who failed to answer the question about race are female, then we are removing a disproportionate amount of female respondents, thereby biasing the results toward the male respondents. To ensure that we are cognizant of these potential effects,

[89] Valliant, 31

we carried out multiple comparison studies using R. For both the 2020 and 2016 Arkansas Polls, we performed listwise deletion for each individual variable in order to create new samples without missing data, then compared these samples to the original according to our chosen demographics. In doing so, we were lucky to not find any conclusive patterns or trends in those respondents suffering item nonresponse. The resulting bias is therefore negligible. Information regarding these pattern tables is available in Appendix A.

Chosen Questions

Weighting will be performed on the Arkansas Polls of 2020 and 2016 for both the presidential choice question and a selection of issue questions that gauge public opinion. Each of the questions presented in the Arkansas Poll has a selection of answers from which the respondent is told to choose, but in the case that they refuse to answer or respond that they are unsure, we will count the respondent as missing the answer and use listwise deletion to remove them from the sample for that question.

The question of presidential choice is straightforward: respondents are asked which of the presidential candidates they are most likely to vote for, and their options as follows:

1. *Donald Trump, the Republican*
2. *Joe Biden, the Democrat (in the 2016 Arkansas Poll, this is "Hillary Clinton, the Democrat")*
3. *Another Candidate (specify)*[90]

For the issue questions, it was necessary to select questions that were asked in both the 2016 and 2020 Arkansas Polls for the sake of consistency and comparison within our

[90] *The Arkansas Poll* Website, "2020 Arkansas Poll Protocol."

experiment. We also wanted to select popular “wedge” issues that are widely discussed, appear frequently in media, and on which most respondents would at least be partially informed. The following questions were chosen and assigned shorthand titles for the duration of the experiment:

1. ABORTION – “Do you favor laws that would make it more difficult for a woman to get an abortion, favor laws that would make it easier to get an abortion, or should no change be made to existing abortion laws?”
    a. More Difficult
    b. Easier
    c. No Change
2. GUN CONTROL – “In general, would you say that favor stricter gun control, less strict gun control, or should no change be made to the existing gun control laws?”
    a. Stricter
    b. Less Strict
    c. No Change
3. CLIMATE CHANGE – “Do you think global warming, or climate change, will pose a serious threat to you or your way of life in your lifetime?”
    a. Yes
    b. No
    c. Not a Problem (this is not a listed answer, but some respondents will give it unprompted)
4. DIRECTION – “Overall, do you feel that Arkansas is generally headed in the right direction or the wrong direction?”
    a. Right

b. Wrong[91]

Each of these questions appears in both the 2016 and 2020 Arkansas Polls, and none of them suffer from too large of an item nonresponse rate to be considered insignificant results. While the question of “the biggest problem facing Arkansas” was considered for inclusion, its results have historically seen little difference year-to-year, with most Arkansans responding “the economy” on a yearly basis. Therefore, focus was instead given to these other four issue questions.

Measures of Accuracy and Precision

The presidential choice question involves what the issue questions do not: a measure of accuracy. We can directly compare the Arkansas Poll raw and weighted results to the presidential elections numbers from the respective years. As we discussed in Chapter 3, the Arkansas Poll for both 2016 and 2020 was within its margin of error in predicting the presidential election percentages, so it is already fairly accurate; however, weighted results being closer to the actual election numbers may be indicative of an increase in accuracy. While this method may not be a foolproof designator, it is at least a point of comparison which is unavailable for the issue questions.

Rather than measure accuracy, the issue questions will be observed for the magnitude of the data effect produced by weighted results. The larger the difference in raw results versus weighted results, the larger effect denoted by that particular variable or variable combination.

In terms of precision estimation, all weighted result reports for both presidential choice and issue questions will include the “general design effect” given by the “anesrake” procedure in

[91] *The Arkansas Poll* Website, “2020 Arkansas Poll Protocol.”

R. This number denotes the increase in variance endured by the data when each particular weighting procedure is carried out. If the general design effect is 1.5, this indicates a variance increase of 50%, and a general design effect of 3.0 denotes a variance increase of 200%. The margin of error for the 2020 Arkansas Poll is 3.9 percentage points for the presidential choice question and 3.4 percentage points the issue questions, so if weighting the presidential choice question results in a general design effect of 2.0, the variance increases by 100% and the margin of error grows from 3.9% to 7.8%--a very sizable increase. The margin of errors for each Arkansas Poll in question are printed here for reference:

	Margin of Error – Pres.	**Margin of Error - Issue**
2020 Arkansas Poll	+/- 3.9%	+/- 3.4%
2016 Arkansas Poll	+/- 4.1%	+/- 3.5%

Figure 5.2: Arkansas Poll Margin of Error with 95% Confidence for both Presidential Choice and Issue Questions.[92] [93]

Order of Weighting and Defining Variable Effects

After all of the necessary data is correctly implemented into R, we will begin our experiment by performing weighting on the questions of presidential choice from both 2020 and 2016. We weight first for all six variables (this trial will likely see the highest increase in variance, but it is necessary to include for the sake of a complete experiment). Then, we will weight according to each of the six individual variables—six individual weighting trials. After

[92] Parry, *The Arkansas Poll, 2016: Summary Report.*
[93] Parry, *The Arkansas Poll, 2020: Summary Report.*

looking at the results of the above trials, we will determine a few more combinations of variables by which to weight according to their effects on the results. Considering both years, this alone requires at least 16 individual weighting trials.

Once the presidential choice results have been weighted, we will briefly observe the effects which each variable and variable combination had on the sample and the results. This will allow us to determine which variables and variable combinations should be used in weighting the issue questions. The discussion of these variable combinations will be given at the beginning of Section 5.3. Then, the weighting of the issue questions (all four questions for both years) can commence.

It should be noted that in our determination of variable effects, it is necessary for us to have a set effect measurement. We will be given a variance measurement in the general design effect, printed in every weighting trial by the anesrake package in R. For defining a trial's effect, we need to look at presidential choice and the issue questions individually.

For the presidential choice question, we define the effects according to the change in "percent difference," which refers to the difference in percentage points between poll results (weighted or raw) and the actual election results. These effects can be categorized into one of three options: *more accurate*, *more extreme*, or *slight change*. Their definitions follow:

- *More Accurate:* after weighting, the percent difference decreases by more than 0.75% for at least one candidate while the percent difference for the other candidate does not increase.
- *More Extreme:* after weighting, the percent difference increases by more than 0.75% for at least one candidate while the percent difference for the other candidate does not decrease.

- *Slight Change*: anything other than the above categories.

In observing the effects of the issue questions, we are no longer looking for accuracy but instead looking for change. Therefore, the weighting effects will be categorized based on the magnitude of the change in a question's response percentages. More specifically, these effects will be defined in certain cases as "shifts," in such instances as one response decreases by a certain amount by which another response increases. These changes/shifts are defined as follows:

- *Large Change/Shift:* 4-6% difference in percentages, raw vs weighted
- *Medium Change/Shift:* 2-4% difference in percentages, raw vs weighted
- *Slight Change/Shift:* <2% difference in percentages, raw vs weighted

These change categories will allow us to easily separate variable/question combinations by their effect. Following the completion of issue question weighting, we will divide the variable/question combinations among the three categories to at a glance deduce where the largest effect can be seen and whether there emerge any patterns between them.

Finally, once the weighting trials have been completed and the initial observations have been categorized, a full analysis of the results will take place in Chapter 6.

5.2 Results – Weighting Presidential Choice, 2020 and 2016

The weighting trials for presidential choice were conducted according to the following order:

All 2020 trials were completed first, followed by 2016. We began, in 2020, by weighting all of the variables together and then by weighting with variable individually (7 total trials). Then, we chose to conduct three more trials using variable combinations, with those trials being Age, Race, Education, and Income; Income and Age; and Education and Race. The category of

Age, Race, Education, and Income was chosen due to both gender and congressional district's relatively minute effect on the results, so weighting according to the four more influential variables without the non-influential ones could reduce bias but encourage accuracy. The other two combinations were chosen because of their similar effects when weighted individually—age and income both produced seemingly more accurate results, while race and education both produced seemingly more extreme results (with "more accurate" and "more extreme" both pertaining to the definitions in Section 5.1). Once these 10 trials had been completed, the 2016 trials were carried out according to the exact same variable categories. All of the R code necessary for weighting is available for replication in Appendix C.

All of the data tables for presidential choice weighting will be presented in this section. Each individual trial will be presented in a table such as this:

Variable(s):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	--.--%	-2.57%	--.--%
Trump:	65.30%	62.40%	--.--%	2.90%	--.--%
Other:	2.49%	2.82%	--.--%	-0.33%	--.--%
Respondents: ---			Gen. Design Effect: (Variance Indicator)		
Weighting Effect: (Regardless of Gen. Design Effect)					

Figure 5.3: Example Table – Presidential Choice Weighting.

The first two result columns, the Arkansas Poll's results for their likely voting sample and the actual election results, both shaded dark, will remain the same across each table for a single year and are included for ease of comparison. The percent difference columns, shaded lightly, are used to determine the effect of the variables through weighting. A negative percent difference indicates that the survey results (either raw or weighted) are lower than the actual election results, whereas a positive percent difference indicates the opposite. The Arkansas Poll to Actual Results column will also remain the same across each trial for the respective years. The respondent total indicates the respondents which remained in the sample for the weighting trial following listwise deletion. Also, the general design effect is included in each table as an indicator of variance increase throughout the trials. Finally, the broad weighting effect, either *More Accurate*, *More Extreme*, or *Slight Change* (as defined in Section 5.1) is included at the bottom of every table.

The resulting tables are printed below.

2020 Weighting Trials – Presidential Choice:

All Variables:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	36.63%	-2.57%	1.85%
Trump:	65.30%	62.40%	60.46%	2.90%	-1.94%
Other:	2.49%	2.82%	2.92%	-0.33%	0.10%
Respondents: 508			Gen. Design Effect: 3.101397		
Weighting Effect: *More Accurate*					

Gender Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.14%	-2.57%	-2.64%
Trump:	65.30%	62.40%	65.38%	2.90%	2.98%
Other:	2.49%	2.82%	2.48%	-0.33%	-0.34%
Respondents: 562			Gen. Design Effect: 1.001344		
Weighting Effect: *Slight Change*					

Congressional District Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.36%	-2.57%	-2.42%
Trump:	65.30%	62.40%	65.06%	2.90%	2.66%
Other:	2.49%	2.82%	2.59%	-0.33%	-0.23%
Respondents: 548			Gen. Design Effect: 1.009808		
Weighting Effect: *Slight Change*					

Race Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	39.66%	-2.57%	4.88%
Trump:	65.30%	62.40%	58.18%	2.90%	-4.22%
Other:	2.49%	2.82%	2.16%	-0.33%	-0.66%
Respondents: 557			Gen. Design Effect: 1.375674		
Weighting Effect: *More Extreme*					

Age Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.44%	-2.57%	-2.34%
Trump:	65.30%	62.40%	63.58%	2.90%	1.18%
Other:	2.49%	2.82%	3.98%	-0.33%	1.16%
Respondents: 551			Gen. Design Effect: 1.672223		
Weighting Effect: *More Accurate*					

Education Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	30.34%	-2.57%	-4.44%
Trump:	65.30%	62.40%	67.28%	2.90%	4.88%
Other:	2.49%	2.82%	2.37%	-0.33%	-0.45%
Respondents: 561			Gen. Design Effect: 1.518178		
Weighting Effect: *More Extreme*					

Income Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.68%	-2.57%	-2.10%
Trump:	65.30%	62.40%	64.58%	2.90%	2.18%
Other:	2.49%	2.82%	2.74%	-0.33%	-0.08%
Respondents: 533			Gen. Design Effect: 1.009051		
Weighting Effect: *More Accurate*					

Age, Race, Education, Income (Gender and District Removed)					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	36.29%	-2.57%	1.51%
Trump:	65.30%	62.40%	61.17%	2.90%	-1.23%
Other:	2.49%	2.82%	2.53%	-0.33%	-0.29%
Respondents: 521			Gen. Design Effect: 2.900476		
Weighting Effect: *More Accurate*					

Income and Age (produced accurate results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.52%	-2.57%	-2.26%
Trump:	65.30%	62.40%	63.36%	2.90%	0.96%
Other:	2.49%	2.82%	4.12%	-0.33%	1.30%
Respondents: 526			Gen. Design Effect: 1.641491		
Weighting Effect: *More Accurate*					

Education and Race (produced extreme results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	38.37%	-2.57%	3.59%
Trump:	65.30%	62.40%	59.63%	2.90%	-2.77%
Other:	2.49%	2.82%	2.00%	-0.33%	-0.82%
Respondents: 556			Gen. Design Effect: 1.991669		
Weighting Effect: *Slight Change*					

Figures 5.4-5.13: 2020 Weighting Trials – Presidential Choice.

2016 Weighting Trials – Presidential Choice

All Variables:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	34.44%	2.66%	0.79%
Trump:	59.33%	60.57%	61.12%	-1.24%	0.55%
Other:	4.37%	5.78%	4.44%	-1.41%	-1.34%
Respondents: 373			Gen. Design Effect: 4.542968		
Weighting Effect: *More Accurate*					

Gender Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	36.07%	2.66%	2.42%
Trump:	59.33%	60.57%	59.38%	-1.24%	-1.19%
Other:	4.37%	5.78%	4.55%	-1.41%	-1.23%
Respondents: 504			Gen. Design Effect: 1.006177		
Weighting Effect: *Slight Change*					

Congressional District Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	36.69%	2.66%	3.04%
Trump:	59.33%	60.57%	59.00%	-1.24%	-1.57%
Other:	4.37%	5.78%	4.31%	-1.41%	-1.47%
Respondents: 498			Gen. Design Effect: 1.006596		
Weighting Effect: *Slight Change*					

Race Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	41.21%	2.66%	7.56%
Trump:	59.33%	60.57%	55.22%	-1.24%	-5.35%
Other:	4.37%	5.78%	3.57%	-1.41%	-2.21%
Respondents: 492			Gen. Design Effect: 2.260936		
Weighting Effect: *More Extreme*					

Age Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	33.80%	2.66%	0.15%
Trump:	59.33%	60.57%	59.92%	-1.24%	-0.65%
Other:	4.37%	5.78%	6.28%	-1.41%	0.50%
Respondents: 491			Gen. Design Effect: 1.719485		
Weighting Effect: *More Accurate*					

Education Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	35.68%	2.66%	2.03%
Trump:	59.33%	60.57%	60.63%	-1.24%	0.06%
Other:	4.37%	5.78%	3.69%	-1.41%	-2.09%
Respondents: 501			Gen. Design Effect: 1.281187		
Weighting Effect: *More Accurate*					

Income Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	36.17%	2.66%	2.52%
Trump:	59.33%	60.57%	59.34%	-1.24%	-1.23%
Other:	4.37%	5.78%	4.49%	-1.41%	-1.29%
Respondents: 388			Gen. Design Effect: 1.01335		
Weighting Effect: *Slight Change*					

Age, Race, Education, Income (Gender and District Removed)					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	33.68%	2.66%	0.03%
Trump:	59.33%	60.57%	62.33%	-1.24%	1.76%
Other:	4.37%	5.78%	3.99%	-1.41%	-1.79%
Respondents: 376			Gen. Design Effect: 4.409438		
Weighting Effect: *Slight Change*					

Income and Age (produced accurate results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	34.99%	-2.57%	1.34%
Trump:	59.33%	60.57%	58.42%	2.90%	-2.15%
Other:	4.37%	5.78%	6.59%	-0.33%	0.81%
Respondents: 383			Gen. Design Effect: 1.604377		
Weighting Effect: *More Accurate*					

Education and Race (produced extreme results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	45.51%	-2.57%	11.86%
Trump:	59.33%	60.57%	55.33%	2.90%	-5.24%
Other:	4.37%	5.78%	3.17%	-0.33%	-2.61%
Respondents: 490			Gen. Design Effect: 2.868935		
Weighting Effect: *More Extreme*					

Figures 5.14-5.23: 2016 Weighting Trials – Presidential Choice.

5.3 Results – Weighting Issue Questions, 2020 and 2016

The results of presidential choice weighting were relatively mixed, both in terms of the variable effects and the variance increases—some trials saw more accurate or more extreme results, and these did not necessarily coincide with a certain magnitude of general design effect. One undeniable observation, however, was the low effects of gender and congressional district, both in terms of weighting effect and general design effect. Given the mixed effects of other variables versus the consistently low effects of gender and congressional district, we removed trials involving those two variables from the issue question weighting. Therefore, the following variable categories were used for issue weighting:

1. Age
2. Income
3. Age and Income
4. Race
5. Education
6. Race and Education
7. All Variables (Age, Income, Race, and Education – NOT Gender or District)

The weighting trials for presidential choice were conducted according to the following order:

All 2020 trials were completed first, followed by 2016. Within both years, there were four issue questions to weight: Abortion, Gun control, Climate Change, and Direction. In each case, we weighted by one question at a time, performing trials for each variable category in the 2020 Abortion question, then the 2020 Gun Control Question, then 2020 Climate Change, 2020 Direction, and repeated with 2016. We finished by categorizing each trial by its effects.

Based on the effects observed from the presidential choice weighting, we made a number of hypotheses regarding potential effects in issue weighting. Age, Income, and all four variables together tended to point toward more accurate results, so there is basis for the same expectation when weighting by issue. Race and Education, however, produced more extreme results (except for 2016 Education), meaning a similar effect is possible in the issue weighting. Variance increases were also quite large in certain trials but fairly low in others; typically, those results which were categorized as more extreme also produced greater general design effects, so we can expect the same to happen in the following issue weighting trials.

It is also important to note that the nature of these trials is different than those previously. Whereas the presidential choice trials were measured for accuracy, these issue question trials are measured for their effect magnitude. In other words, we are now observing how much the variable weights change the survey results, and furthermore, what those effects might mean for Arkansas public opinion and the accuracy of the Arkansas Poll in each question.

With both 2020 and 2016, all four chosen issue questions, and 7 different variable categories, we completed 56 issue weighting trials in total. To efficiently report the findings, each of the trials will be listed as a row of a larger "effect table," which will list the effect category and effect magnitude for each year/question/variable combination. For further reference, full tables for each individual trial are given in Appendix B. Two "effect tables" are presented in this section—one for 2020 and one for 2016. The "Weighting Effect" listed for each variable/question combination corresponds to the definitions given in Section 5.1.

The effect tables are presented below:

2020 Weighting Trials – Issue Questions

Abortion - 2020			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*large shift to "Easier"*	741	1.600102
Income	*slight change*	718	1.020411
Age & Income	*large shift to "Easier"*	706	1.632359
Race	*medium shift to "Easier"*	750	1.247403
Education	*slight change*	759	1.197067
Race & Education	*medium decrease for "More Difficult," slight increase for "Easier"*	749	1.472398
All Variables	*medium shift to "Easier"*	696	2.429891

Gun Control - 2020			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*medium increase in "Less Strict," slight increase in "Stricter," large decrease in "No Change"*	764	1.594978
Income	*slight change*	744	1.015819
Age & Income	*medium increase in "Less Strict," slight increase in "Stricter," large decrease in "No Change"*	731	1.610003
Race	*large increase in "Stricter," medium decrease in "Less Strict" and "No change"*	772	1.247388
Education	*slight change*	784	1.181413
Race & Education	*large increase in "Stricter," medium decrease in "Less Strict" and "No change"*	771	1.455007
All Variables	*medium increase in "Stricter" and "Less Strict," large decrease in "No Change"*	720	2.363860

Climate Change - 2020			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*large shift to "Yes"*	767	1.562697
Income	*slight change*	742	1.017391
Age & Income	*large shift to "Yes"*	729	1.582889
Race	*large shift to "Yes"*	776	1.233719
Education	*slight shift to "No"*	788	1.182054
Race & Education	*medium shift to "Yes"*	775	1.442296
All Variables	*large shift to "Yes"*	718	2.324929

Direction - 2020			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*large shift to "Right"*	752	1.592463
Income	*slight change*	729	1.020737
Age & Income	*large shift to "Right"*	718	1.622487
Race	*slight/medium shift to "Wrong"*	760	1.232487
Education	*slight shift to "Right"*	771	1.197821
Race & Education	*slight shift to "Wrong"*	759	1.457294
All Variables	*large shift to "Right"*	708	2.358094

Figures 5.24-5.27: 2020 Weighting Trial Summaries – Issue Questions.

2016 Weighting Trials – Issue Questions

Abortion - 2016			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*slight change*	683	1.473350
Income	*slight change*	551	1.028747
Age & Income	*slight change*	540	1.434633
Race	*slight change*	699	1.762027
Education	*slight change*	712	1.193939
Race & Education	*slight change*	693	2.070869
All Variables	*medium decrease to "Easier," medium increase to "No Change"*	527	2.773603

Gun Control - 2016			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*medium shift to "Less Strict"*	729	1.45636
Income	*slight change*	578	1.02846
Age & Income	*medium shift to "Less Strict"*	567	1.41357
Race	*medium/slight shift to "Stricter"*	745	1.81597
Education	*slight change*	759	1.18733
Race & Education	*medium/slight change to "Stricter"*	740	2.09438
All Variables	*medium decrease in "Stricter," slight decrease in "Less strict," medium increase in "No Change"*	554	2.74874

Climate Change - 2016			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*large shift to "Yes"*	683	1.443005
Income	*slight change*	548	1.030334
Age & Income	*medium shift to "Yes"*	538	1.419846
Race	*large shift to "Yes"*	695	1.767024
Education	*slight/medium shift to "No"*	711	1.149263
Race & Education	*large shift to "Yes"*	693	2.009392
All Variables	*large shift to "Yes"*	528	2.71327

Direction - 2016			
Variable(s)	**Weighting Effect**	**Respondents**	**GDE**
Age	*slight change*	640	1.45817
Income	*slight change*	513	1.01805
Age & Income	*slight change*	503	1.39526
Race	*slight change*	649	1.86607
Education	*slight change*	661	1.17839
Race & Education	*slight change*	643	2.21074
All Variables	*medium shift to "Wrong"*	490	2.58369

Figures 5.28-5.31: 2016 Weighting Trial Summaries – Issue Questions.

Chapter 6 – Analyzing the Results

With over 70 different trials covering 2 separate polling years, 5 different poll questions, and 10 different variable weighting groups, the different combinations and measurements make the results of this experiment a complex statistical environment. While we have created brief observations for each individual trial in the previous chapter, directly comparing results of varying combinations is not an immediately intuitive task. To make it more manageable, we defined and assigned terms such as “more accurate” or “large shift” to the findings in order to establish a standard for measurement, but these observational terms do not account for the general design effect which if too high renders the trial obsolete. Although, as discussed earlier, “too high” does not correspond to a widely agreed-upon number or level of variance increase. Therefore, we will work through the results of our weighting trials while remaining attentive of the bias introduced by the general design effect.

For our analysis, we begin by first working through the weighting trials for presidential choice, both in 2020 and 2016. While we already came to some immediate conclusions regarding which variables created which effects, this will give us the chance to pick out a few of the most significant and determine their wider context in terms of our findings. Then, we move into a similar discussion involving the issue questions where we address our earlier hypotheses and consider the most notable trials. Finally, we can use our conclusions from these discussions to build an understanding of the Arkansas Poll’s accuracy in the past two presidential election years.

6.1 Analyzing Presidential Choice Weighting

Prior to our analysis of the presidential choice weighting, we need to once more note the discrepancy regarding the target population. Whereas the Arkansas Poll typically seeks to represent the state population as a whole, the target population in this instance is the likely voters. The potential problem arises with our population information provided by the U.S. Census, which has no measure for likely voters. The trade-off, of course, is that we have an accuracy measure for the question of presidential choice, which is the election numbers from the respective years. We can therefore discern whether weighting by the Census population variables creates results that are more accurate or extreme. This does not necessarily mean that these variables would create the same effect every time across different years and polls, but instead, they may be indicative of useful variables for weighting moving forward.

We begin our presidential choice analysis with the 2020 trials. Variance increases tended to range from moderate to large across the board. Generally, the more variables that were used, the higher the general design effect and therefore the increase in variance. This was especially so when variables which alone produced opposite effects were grouped together—for example, pairing either race or education (both of which produced more extreme results) with either age or income (both of which produced more accurate results). Luckily, when weighted alone (in no combination with other variables), no variable produced greater than a 70% increase in variance. This indicates that the confidence intervals, which were originally 3.9 percentage points before weighting, grow to no more than 6.6 percentage points. While this is a significant increase, it should not immediately rule out our results—in fact, they are still fairly manageable, and the trade-off is worthwhile considering the results in multiple cases. For the combination of income and age, the resulting variance increase is similar (64% increase). However, other variable combinations are much less statistically significant because of their extreme design effects. The combination of education and race creates a 99% increase in variance, while weighting by all variables produced a whopping 210% increase in variance. Because of this, the weighted results from these categories are largely obsolete because of the immense variance involved. Therefore, we will specifically focus on the single-variable trials as well as the two-variable combinations.

In exploring the effects of each of the variables, we should begin with what was already stated in Section 5.3: gender and congressional district both had little to no effect on the results of the presidential choice. Because of this, we can assume that neither variable will produce any other effect across the other issue questions, and both variables will therefore be ignored moving forward. It was by all means necessary to include both in the experiment, but as they produce no conclusive results, we will direct focus to the four remaining variables and their combinations.

There are three variable categories with reasonable design effects which created more accurate results. By this, we mean that the weighted results were significantly closer to the actual election results than the raw survey results. Those categories were Age Only, Income Only, and the combination of Age and Income. Age contributed the most to variable increase with the Age Only category resulting in a 67% increase in variance, whereas the Income Only category saw a completely negligible increase in variance (less than 1%). The combination resulted in a slightly lower variance increase than Age Only with a 64% increase, and its results were also of the more accurate variety.

More extreme results, on the other hand, were those in which the weighted results differed from the election numbers by a significantly greater amount than the raw survey results. The variable categories which resulted in this outcome were Race Only (37% variance increase) and Education Only (51% variance increase). The combination category Race and Education, however, saw a potentially too high variance increase to be relevant (99% increase), but it should be noted that its results were also more extreme, just to a lesser degree.

Moving to 2016, we can see many similarities and a few significant differences in the weighting results. Like the 2020 trials, we will begin with an examination of the variance effects. Overall, the 2016 weighting trials saw much larger variance increases than in 2020. As a summary, whereas weighting by all variables in 2020 produced a 200% increase in variance, doing the same in 2016 produced a 350% increase. Because of this, the categories of All Variables and All Variables Besides Gender and Congressional District both suffer such massive design effects that their results have no statistical significance. For the individual variable trials, variance rates were similar to those of 2020. All were completely manageable save for Race Only, which saw a 126% increase in variance—a testament to the extremity of the variable's influence, which we will see now as we discuss variable effects.

Age Alone remains one of the variable categories which produced the most accurate results in 2016, reducing each answer to the presidential choice question to within a percentage point of the election results. Weighting based on Income and Age together also resulted in more accurate results, but this was largely due to the Age factor; Income produced little effect on its own, which is in contrast to its more accurate results in 2016. A surprising aspect of the 2016 results was that Education Alone actually had the opposite effect compared to 2020. Rather than produce more extreme results, it produced more accurate ones, becoming the only variable to completely "flip" variable effect between 2016 and 2020. Furthermore, it did so with only a 28% increase in variance.

Race Alone was consistent with its effect from 2020, creating more extreme results once again during the 2016 trials. However, its variance, as mentioned earlier, increased by 126%.

That being said, because the results are more extreme, the massive variance increase is unsurprising and not necessarily a detriment to the statistical significance in this case. Race and Education intuitively has an even greater extreme effect, seeing as its variance increases by 186% and it involves weighting by two variables which weighted alone have opposite effects.

Taking both years together, we can focus on the consistencies to identify our larger conclusions. As expected, weighting by more variables always resulted in a greater increase in variance, especially when the variables had opposite effects alone. Gender and congressional district both had such minute effects that their weighted results are completely negligible and insignificant. Race created more extreme results in both years, indicating that it is the most influential variable to be used for weighting, but while it may make the sample more representative of the population, it does not create more accurate results. Therefore, race may be a better variable to be used for the issue questions than for presidential choice. Education had opposing effects in 2016 and 2020, so it is hard to classify its effect overall despite the aforementioned Pew Research study pointing to education weighting as a major reducer of bias.[94] Perhaps 2020 was simply an anomaly for weighting by education, but we cannot know for sure given the extent of this experiment.

Finally, if we are to consider a potential "golden combination" for accuracy when weighting, Age was consistently produced the most accurate results in both years. This is not to say that age "predicts" how Arkansans will vote, but instead that making the Arkansas Poll sample more representative of the population's age parameters resulted in more accurate results in 2016 and 2020. Income may also be a worthy addition to our "more accurate" variables. Overall, our results are not conclusive enough to identify a golden combination for accuracy, but

[94] Mercer, et al., "Weighting Online Opt-In Samples."

for now, we know that Age and Income's weighting results in 2016 and 2020 may be indicative of producing more accurate results in future election years.

6.2 Analyzing Issue Question Weighting

Moving from presidential choice weighting to issue weighting, our expectations from the beginning of Section 5.3 were largely based on the conclusions found in the previous trials: we expected to see the most extreme results in the cases of Race and Education. This potential for extremity is where we focus in the new trials—we no longer have an accuracy measure for the issue questions like we had for presidential choice. Therefore, we are no longer focusing on the potential for accuracy but rather the effect that each variable category has across the four selected issue questions.

We begin, as we did during presidential choice, with an examination of the variance increases. One of the more convenient aspects of weighting the issue questions is that variable groups in a particular year usually experienced very similar increases across all four questions, such as the variable category Age Only experiencing 55%-60% variance increase each time. This narrows down the variance categories significantly and allows us to take note of larger groups, such as All Variables in both 2016 and 2020, which always shows a variance increase of at least 130%. For 2020, every other category besides All Variables has a variance increase of 70% or early, so staying consistent with the standards set during our presidential choice analysis, these variances are manageable, and we can consider the results significant—just less so than had they been lower. 2016, on the other hand, saw higher design effects in many categories. While most of their categories stayed relatively low similar to those in 2020, Race and the combination Race and Education both had high variances, with Race being in the 75%-90% range and Race and

Education typically following a 100-120% variance increase. These categories should then be taken with an extra grain of salt in terms of their statistical significance, but it at least coincides with the variance increases we observed in the presidential choice trials.

Given the 52 total trials present in the issue question portion of the experiment, it is easiest to observe the data through grouping. We should first note that, as expected, no questions were "flipped." In other words, no answer that was originally the minority for its question became the majority due to weighting by any particular variable category. Instead, each trial was able to fall into one of the previously set definitions for effect magnitude. Below, we print two tables which categorize the various trials into their effect groups, one for 2020 and another for 2016. Note that these tables do not mark a "shift" in data as the term is used in the tables of Section 5.3, and furthermore, that variance increases of above 70% are marked accordingly.

2020 Issue Weighting Trials – Grouped by Effect Magnitude			
	Most Change (Large Effect)	**Medium Change (Medium Effect)**	**Slight Change (Small Effect)**
Abortion	Age Age/Income	Race Race/Education All Variables*	Income Education
Gun Control	Age Age/Income Race Race/Education All Variables*		Income Education
Climate Change	Age Age/Income Race All Variables*	Race/Education	Income Education
Direction	Age Age/Income All Variables*	Race	Income Education Race/Education
			**variance increased by more than 70%*

Figure 6.1: 2020 Issue Weighting Trials – Grouped by Effect Magnitude.

2016 Issue Weighting Trials – Grouped by Effect Magnitude			
	Most Change (Large Effect)	**Medium Change (Medium Effect)**	**Slight Change (Small Effect)**
Abortion		All Variables*	Age Income Age/Income Race* Education Race/Education*
Gun Control		Age Age/Income Race* Race/Education* All Variables*	Income Education
Climate Change	Age Race Race/Education* All Variables*	Age/Income Education	Income
Direction		All Variables*	Age Income Age/Income Race* Education Race/Education*
			**variance increased by more than 70%*

Figure 6.2: 2016 Issue Weighting Trials – Grouped by Effect Magnitude.

These grouping tables reveal a number of observations regarding the results, some expected and some surprising. Age Only, as well as Age and Income, created large effects in every question of 2020; however, Income Only always resulted in slight change, meaning the Age and Income category only created such large effects, as seen earlier, because of the presence of Age as a variable. This is surprising considering that Age Alone made the presidential choice more accurate but typically not more extreme. The largest effects during presidential choice weighting had been those created by Race Only and Race and Education, and despite our expectation that the same would happen in these trials, Race and Race/Education produced large effects on the results in less than half of the trials. Education Only and Income Only both resulted in slight changes every time (except for Education Only producing a medium effect in the 2016 Climate Change trial). Weighting by All Variables tended to also have large effects, especially in 2020, but once again always resulted in extreme variance increases and is not a producer of statistically significant results.

Despite all of these variable-specific results, the most notable observation from these tables is the difference in effect magnitude between the two years. Even a quick glance at the above tables shows that large and medium effects were much more frequent in 2020 than 2016. To quantify this observation, 2020 saw 14 large effect trials, 5 medium effect trials, and 9 slight change trials. 2016 contrarily saw 4 large effect trials (all in the Climate Change question), 9 medium effect trials, and 15 slight change trials. This difference is quite extreme, but we should not forget that it completely hinges on our definitions of large, medium, and slight effects as were laid out in Chapter 5. Recall that we defined large effects to be trials involving a 4-6% difference in the raw percentages and the weighted percentages, medium effects having 2-4% difference, and slight change being less than 2%. This puts the groupings above into perspective: weighting in 2020 most often resulted in a 4-6% change in results, whereas weighting in 2016 saw mostly results with changes of less than 2%. Therefore, when we use the term “large effect,” we are still referring to percentage differences that are within the single digits. While this may seem odd to some observers, viewing the results with this perspective highlights the major difference that exists between the 2020 and 2016 Arkansas Polls and the extent to which their results change through weighting.

6.3 Notable Observations and Conclusions

After summarizing the most important aspects observable throughout these results, we can piece together the larger picture contextualizing the effect of weighting the Arkansas Poll. One potential route for observations that has yet to be considered is the political nature of the weighting shifts. By this, we mean the potential for assigning certain answers throughout the issue questions with either "liberal" or "conservative" labels depending on the policies with which they generally coincide. In doing so, we can then discern whether or not certain variables shift the results toward one direction or the other across all of the questions, therefore showing that a certain viewpoint shines through in greater capacity when weighting is performed according to specific variables. Sadly, this does not appear to be the case, at least for the data we have from 2020 and 2016. For instance, our most influential variable category, Age Only, shifts the data toward the more "liberal" option in some questions but then to "conservative" in others in the 2020 Poll—this category creates a large increase in the percentage that respond favoring laws making it easier to get abortions (a generally liberal policy) but also creates a large increase in the percentage that respond favoring a "less strict" system of gun control (a generally conservative policy). While this may be an interesting point to investigate in future Arkansas Polls, it does not appear to be conclusive during 2020 or 2016.

Another important consideration is that which we have already referenced: the changes to the results following weighting were not incredibly extreme in either case. To the general observer, none of our weights seem to be drastically affecting the results of the Arkansas Poll.

No questions are “flipped” to one side or the other due to weighting by one or more variables—the results tend to just shift one way or the other by a single-digit percentage. The largest changes from weighting were those that shift the percentages by around 6%-7%, not enough to change the outlook on the state’s overall public opinion by any means. This consideration then begs the question: is there any worth in weighting the Arkansas Poll? Given these general effects, what is the point in weighting the Arkansas Poll year-to-year only to see the results change by a maximum of 6%?

As it turns out, there very much is reason to do so, and it circles back to our consideration of the differences in effects between the two years. While the results of the Arkansas Poll are not drastically changing as a result of weighting, there was a significantly greater rate of change in 2020 than there had been in 2016. As the entire basis for this experiment, weighting is a statistical strategy used to make survey results more accurate by allowing the sample to better reflect the population. If a survey’s results change drastically due to weighting, this implies that the survey was inaccurate in its representation of the target population since the sample needed to be weighted heavily in order to reflect said population. Similarly, a survey whose results only change slightly implies that the sample already generally reflects the population. Focusing on the Arkansas Poll, the greater effects on the results in 2020 show us that 2016 was more likely an accurate representation of the Arkansas population at the time than the 2020 poll. More importantly, this may indicate a loss of accuracy between the years 2016 and 2020, and a potential loss of accuracy should be a point of concern for the Arkansas Poll.

This is brought up not because an accuracy loss is necessarily confirmed, but that any developing bias such as this is worth noting—especially for a political poll which seeks to accurately represent the state’s population. We must emphasize that only two years were tested, and a more explanatory picture could be gained through an exploration of other recent Arkansas Polls such as those in 2017-2019. However, given the concern arising from our study, we reinforce the importance of weighting not as an accuracy increaser but as an accuracy indicator. While it may not be worth the variance increases to weight the Arkansas Poll results before publishing them, it would be by all means worthwhile to be carried out each year to determine the Poll's performance in representing the target population. Greater effects through weighting can indicate further deviations from the population demographics, and if the effects continue to grow across years of the Poll, the surveyors can note a significant and consistent loss of accuracy worth addressing. Such a process would be beneficial to the continued accuracy of one of Arkansas's only major public opinion polls.

Chapter 7 – Conclusion & Next Steps

The 2016 Presidential Election marked a focal point for political polling in the United States, and its aftermath created widespread skepticism regarding the accuracy of such pre-election surveys. This project originally emerged as part of the growing national interest surrounding polling accuracy, and the intention was to measure Arkansas's involvement in this phenomenon. Instead, once we delved into the Arkansas Poll and its current track record in election prediction, we learned that the most practical measures for improving poll accuracy were those which accounted for sampling errors such as nonresponse. The subject then evolved into a statistical study regarding the effect and importance of weighting on the Arkansas Poll with the goal of potential future application in other statewide public opinion surveys.

Given this goal, we identified those weighting strategies most important in political polling and created an experiment to measure their potential effect. The first round of trials involved weighting voter choice presented by the Arkansas Poll—specifically, the presidential choice question using the 2016 and 2020 Arkansas Polls. Available Census data were utilized for raking the sample according to demographic variables that were shared by our data frames. The findings of presidential choice weighting then informed our strategies as we experimented with the issue questions in both polls, selecting four questions largely considered to be "wedge issues" in Arkansas's current political environment.

While significant increases in variance meant that a number of our trials were obsolete, the resulting experiments still presented us with a number of important findings. Age, rather than educational attainment, seemed to be the variable that most frequently created more accurate results when weighting the presidential choice question, while race as a variable had the strongest opposite effect. Our weighting of the issue questions never produced extremely different results or “flipped” any questions to a different answer, but it did provide us with one vital finding: we saw far greater impact from weighting in 2020 than in 2016 across all variables and issue questions. The implication of this finding is that the 2016 sample was a better representation of the population than that of 2020 since the results changed by a greater amount due to weighting in 2020.

This point leads directly into our recommendations for the future of the Arkansas Poll. Weighting is likely not necessary, at least in the immediate future, for publication within the Poll; results change typically only by a single percentage point, and because Arkansas tends to be fairly one-sided on most issues, researchers and politicians would be unlikely to view the results any differently. However, as an institution which seeks to accurately reflect the overall population of the state, the Arkansas Poll would benefit greatly by performing in-house weighting as its own measure of accuracy. The increase in change due to weighting between 2016 and 2020 is not necessarily indicative of a loss in accuracy, but it should at the very least warrant concern. If the Arkansas Poll were to continue this system of weighting, they could identify a growing disparity between their raw and weighted results as a potential indicator of unrepresentative samples.

Despite including over 70 different weighting trials covering 2 years of polling, 6 different demographic variables, and 5 different polling questions, our experiment was by no means a comprehensive study. For findings with greater significance, we recognize the need for continued research and recommend that future studies focus on the years between 2016 and 2020. While those years were chosen for our experiment because of the presidential choice accuracy measure, we found that the magnitude of change due to weighting could be a significant indicator of accuracy in itself. Weighting the Arkansas Poll for the years between 2016 and 2020 could provide a more complete picture of this loss of accuracy and whether or not it has been a consistent process.

Despite the potential for further confirmation, the results of this study can prove incredibly useful to the Arkansas Poll in the coming years. Following a period of intense polling skepticism, we have viewed that, while the Arkansas Poll may be hitting the mark in terms of presidential election prediction, there are concerns regarding the accuracy of the sample's representativeness. As long as surveyors are cognizant of the potential design effects, we find that weighting can be an impactful strategy not just for creating more accurate results but for measuring the accuracy between polling years. Not only will this strategy be beneficial for the Arkansas Poll, but its potential for application across other state polls is practically limitless.

In facing the challenges arising from an era of polling skepticism, we have seen weighting allow us to place poll results in a greater context of accuracy. While the immediate consequences of this may seem small, they are immensely important for the state of polling and election prediction. The greater increases we see in state poll accuracy, the greater frequency of correct election predictions we will see, and the more confident Americans will be in political polling institutions moving forward.

References

"2016 Arkansas Presidential Election Results." *Politico.* December 13, 2016. https://www.politico.com/2016-election/results/map/president/arkansas/.

"2016 Election Forecast." *FiveThirtyEight.* November 8, 2016. https://projects.fivethirtyeight.com/2016-election-forecast/.

"2020 Arkansas Poll Protocol." *Past Arkansas Poll Results - Fulbirght College of Arts and Sciences.* October, 2020. https://fulbright.uark.edu/departments/political-science/partners/arkpoll/2020-protocol.pdf.

"2020 Election Forecast." *FiveThirtyEight.* November 3, 2020. https://projects.fivethirtyeight.com/2020-election-forecast/.

American National Election Studies. n.d. https://electionstudies.org/.

"Arkansas 2016 Election Forecast." *FiveThirtyEight.* November 8, 2016. https://projects.fivethirtyeight.com/2016-election-forecast/arkansas/.

Arkansas Population Topped 3 Million in 2020. August 25, 2021. https://www.census.gov/library/stories/state-by-state/arkansas-population-change-between-census-decade.html.

"Arkansas Presidential Election Voting History." *270 To Win.* n.d. https://www.270towin.com/states/Arkansas.

"Arkansas Presidential Results - 2020." *Politico.* January 6, 2021. https://www.politico.com/2020-election/results/arkansas/.

"Arkansas: Trump vs. Biden." *RealClearPolitics.* 2020. https://www.realclearpolitics.com/epolls/2020/president/ar/arkansas_trump_vs_biden-7213.html.

Battaglia, Michael P, David Izrael, David Hoaglin, and Martin R Frankel. "Tips and Tricks for Raking Survey Data (a.k.a. Survey Balancing)." *American Association for Public Opinion Research* (2004).

Battaglia, Michael P, David Izrael, David Hoaglin, and Martin R Frankel. "Practical Considerations in Raking Survey Data." *Survey Practice* 2, no.5 (2009).

Brace, Paul, Kellie Sims-Butler, Kevin Arceneuax, and Martin Johnson. "Public Opinion in the American States: New Perspectives Using National Survey Data." *American Journal of Political Science* 46, no. 1 (2002): 173-189.

Chang, Chiao-chih, and Ferry Butar Butar. "Weighting Methods in Survey Sampling." *Section on Survey Research Methods* (2012).

Deane, Claudia, Courtney Kennedy, and Scott Keeter. "A Field Guide to Polling: Election 2020 Edition." *Pew Research Center*. November 19, 2019.

Debell, Matthew, and Jon A. Krosnick. *Computing Weights for the American National Election Study Survey Data.* ANES Technical Report Series, Ann Arbor and Palo Alto: American National Election Study, 2009.

Gelman, Andrew. "Struggles with Survey Weighting and Regression Modelling." *Statistical Science* 22, no. 2 (2007): 153-164.

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. *Applied Survey Data Analysis, Second Edition.* Boca Raton: CRC Press, 2017.

Hillygus, D. Sunshine. "The Evolution of Election Polling in the United States." *Public Opinion Quarterly* 75, no. 5 (2011): 962-981.

Hoek, Janet, and Philip Gendall. "Factors Affecting Poll Accuracy: An Analysis of Undecided Respondents." *Marketing Bulletin* 8, no. 1 (1997): 1-14.

Kalton, Graham, and Ismael Flores-Cervantes. "Weighting Methods." *Journal of Official Statistics* 19, no. 2 (2003): 81-97.

Keeter, Scott. "Determining Who Is a 'Likely Voter'." *Pew Research Center*. August 29, 2012.

Kenett, Ron S., Danny Pfeffermann, and David M. Steinberg. "Election Polls - A Survey, A Critique, and Proposals." *Annual Review of Statistics and Its Applications* 5, no. 1 (2018): 1-24.

Kennedy, Courtney. "Can We Still Trust Polls?" *Pew Research Center*. May 14, 2018.

Kennedy, Courtney, Andrew Mercer, Scott Keeter, Nick Hatley, Kylie McGeeney, and Alejandra Gimenez. "Evaluating Online Nonprobability Surveys." *Pew Research Center*. May 2, 2016.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, et al. "An Evaluation of the 2016 Election Polls in the United States." *Public Opinion Quarterly* 82, no. 1 (2018): 1-33.

Kieding, Niels, and Thomas A. Louis. "Web-Based Enrollment and Other Types of Self-Selection in Surveys and Studies: Consequences for Generalizability." *Annual Review of Statistics and Its Application* 5, no. 2 (2018): 25-47.

Kolenikov, Stanislav. "Calibrating Survey Data using Iterative Proportional Fitting (Raking)." *The Stata Journal* 14, no. 1 (2014).

Kolenikov, Stas. "Post-Stratification or Non-Response Adjustment?" *Survey Practice* 9, no. 3 (2016).

Kreuter, Frauke. "What Surveys Really Say." *Nature*. December 08, 2021.

Laaksonen, Seppo. *Survey Methodology and Missing Data.* Cham: Springer, 2018.

Lax, Jeffrey R., and Justin H. Phillips. "How Should We Estimate Public Opinion in The States." *American Journal of Political Science* 53, no. 1 (2009): 107-121.

Lohr, Sharon L. *Sampling: Design and Analysis, Second Edition.* Boca Raton: CRC Press, 2019.

Lumley, Thomas. *Complex Surveys: A Guide to Analysis Using R.* Hoboken: John Wiley & Sons, 2010.

McAuliffe, Colin, Johannes Fischer, Charlotte Swasey, Jason Katz-Brown, Jason Ganz, Gustavo Sanchez, and Sean McElwee. *2020 Polling Retrospective.* Data for Progress. 2021.

Mercer, Andrew, Arnold Lau, and Courtney Kennedy. "For Weighting Online Opt-In Samples, What Matters Most?" *Pew Research Center*. January 26, 2018.

Panagopoulos, Costas. "Polls and Elections: Accuracy and Bias in the 2020 U.S. General Election Polls." *Presidential Studies Quarterly* 51, no. 1 (2021).

Parry, Janine A. *The Arkansas Poll, 2012: Summary Report.* Survey Report, Fayetteville (2012): University of Arkansas.

Parry, Janine A. *The Arkansas Poll, 2014: Summary Report.* Survey Report, Fayetteville (2014): University of Arkansas.

Parry, Janine A. *The Arkansas Poll, 2016: Summary Report.* Survey Report, Fayetteville (2016): University of Arkansas.

Parry, Janine A. *The Arkansas Poll, 2020: Summary Report.* Survey Report, Fayetteville (2020): University of Arkansas.

Parry, Janine A., Raymond McCaffrey, Michael Duffy, Jr., and Bobbie J. Foster. "Journalists and the Red-State Voter: Perceptions of Journalism Credibility at a Statewide Level in the

U.S. and the Trump Effect." *Southwestern Mass Communication Journal* 34, no. 2 (2019): 1-15.

Parry, Janine A., Will Miller, Bill Schreckhise, and Todd Shields. *The Arkansas Poll, 1999: A Summary Report.* Survey Report, Fayetteville (1999): University of Arkansas.

Pasek, John. *ANES Weighting Algorithm: A Description.* Stanford (2010): Stanford University.

Silver, Nate. "How FiveThirtyEight Calculates Pollster Ratings." *FiveThirtyEight*. Septemeber 25, 2014.

"The Arkansas Poll." *Fulbright College of Arts and Sciences.* n.d. https://fulbright.uark.edu/departments/political-science/partners/arkansas-poll.php.

"Trump is Very Likely to Win Arkansas - 2020 Election Forecast." *FiveThirtyEight.* November 3, 2020. https://projects.fivethirtyeight.com/2020-election-forecast/arkansas/.

Valliant, Richard, and Jill A. Dever. *Survey Weights: A Step-by-Step Guide to Calculation.* College Station: Stata Press, 2018.

Valliant, Richard, Jill A. Dever, and Frauke Kreuter. *Practical Tools for Designing and Weighting Survey Samples, Second Edition.* Cham: Springer, 2018.

Appendix A: Raw Data and Miscellaneous Tables

A.1 Arkansas Poll Data

The Arkansas Poll web page on the University of Arkansas website contains all of the information obtained regarding the Poll and its data. The landing page with a general description of the poll, as well as links to each individual poll and their data, can be found at https://fulbright.uark.edu/departments/political-science/partners/arkansas-poll.php.

Each year, the Arkansas Poll publishes three components: a summary report of its results, the questionnaire/protocol used during the surveying, and a full SPSS sav file of its raw data for every respondent. All three can be found for each individual year, including 2016 and 2020, on the Arkansas Poll webpage.

A.2 Census Data

All of the demographic data used for our weighting trials were obtained from the U.S. Census Estimates from 2016 and 2019. Data was obtained from the U.S. Census Website, and the following tables were utilized for our specific variables:

- 2019 Estimates for Arkansas:
    - S0101: Age and Sex
    - DP02: Selected Social Characteristics in the United States
    - S1501: Educational Attainment
    - S1901: Income in the Past 12 Months (In 2019 Inflation-Adjusted Dollars)
- 2016 Estimates for Arkansas:
    - DP05: ACS Demographic and Housing Estimates

- DP02: Selected Social Characteristics in the United States
- S1501: Educational Attainment
- S1901: Income in the Past 12 Months (In 2016 Inflation-Adjusted Dollars)

A.3 Election Results Data

The presidential election results in Arkansas for both 2016 and 2020 are readily known and available across a number of election reporters, such as Politico, The New York Times, and CNN. The Arkansas Secretary of State reports them here: https://www.sos.arkansas.gov/elections/for-election-results.

Because these results are factually known, we report them below:

Presidential Election Results in Arkansas:		
	2020 Results:	**2016 Results:**
Trump:	62.40%	60.57%
Biden/Clinton: (2016/2020)	34.78%	33.65%
Other:	2.82%	5.78%

A.4 Variable Comparison Tables

As mentioned in Chapter 5.1, our selection of variables to be used during weighting was guided by our comparison tables. The tables compared the sample demographics of the Arkansas Poll (both likely voters and total sample) against those of the Arkansas population for all six variables. Note that for these comparison tables, listwise deletion was utilized beforehand, even for the totally samples, to remove the missing data. Furthermore, while we created tables both with numerical totals and percentages, only the percentage totals are pasted below as they present the easiest avenue for comparison. The comparison tables are printed below:

2020 Comparison Tables

Presidential Choice:			
	AR Poll 2020 (LV):	*AR Poll 2020 (Full Sample):*	*Actual:*
Biden:	32%	31.97%	34.78%
Trump:	65%	56.72%	62.40%
Other:	3%	11.32%	2.82%

Congressional District:			
	AR Poll 2020 (LV):	*AR Poll 2020 (Full Sample):*	*Census 2019 Estimates:*
District 1:	26.02%	26.37%	24.14%
District 2:	22.49%	21.67%	25.43%
District 3:	28.44%	29.77%	26.79%
District 4:	23.05%	22.19%	23.64%

Gender:			
	AR Poll 2020 (LV):	*AR Poll 2020 (Full Sample):*	*Census 2019 Estimates:*
Male:	46.44%	45.15%	48.27%
Female:	53.56%	54.85%	51.73%

Race:			
	AR Poll 2020 (LV):	*AR Poll 2020 (Full Sample):*	*Census 2019 Estimates:*
White:	84.20%	82.28%	71.97%
Black or African American:	7.90%	8.61%	15.41%
Hispanic:	1.62%	2.41%	7.69%
Asian:	0.36%	0.38%	1.52%
Native American:	1.97%	2.28%	0.54%
Multi-Ethnic:	3.95%	4.05%	2.36%
Other:	NA	NA	0.52%

Educational Attainment:			
	AR Poll 2020 (LV):	*AR Poll 2020 (Full Sample):*	*Census 2019 Estimates:*
No high school:	0.36%	1.00%	4.07%
Some high school:	4.81%	5.99%	8.36%
High school graduate:	24.60%	25.81%	34.83%
Some college including business or trade school [ASSOCIATES DEGREE]:	35.65%	35.04%	31.14%
College graduate [BACHELOR'S DEGREE]:	18.72%	17.96%	14.35%
Graduate or professional degree [MASTERS OR DOCTORATE DEGREE]:	15.86%	14.21%	7.25%

Age:			
	AR Poll 2020 (LV):	*AR Poll 2020 (Full Sample):*	*Census 2019 Estimates:*
18-24:	3.63%	4.10%	12.18%
25-34:	5.63%	6.03%	16.67%
35-44:	8.53%	8.59%	16.27%
45-54:	10.89%	11.41%	15.63%
55-64:	21.42%	20.90%	16.65%
65-74:	24.68%	23.97%	13.09%
75-84:	19.96%	19.23%	7.00%
85+:	5.26%	5.77%	2.53%

Income:			
	AR Poll 2020 (LV):	*AR Poll 2020 (Full Sample):*	*Census 2019 Estimates:*
$15,000 or Less:	11.44%	15.87%	13.10%
$15,001 - $25,000	11.63%	14.02%	12.10%
$25,001 - $35,000	12.76%	12.17%	11.50%
$35,001 - $50,000	14.45%	13.89%	14.10%
$50,001 - $75,000	17.07%	15.48%	18.30%
$75,001 - $100,000	13.88%	12.30%	11.50%
$100,001 +	18.76%	16.27%	19.30%

2016 Comparison Tables

Presidential Choice:			
	AR Poll 2016 (LV):	*AR Poll 2016 (Full Sample):*	*Actual:*
Clinton:	36%	30.63%	33.65%
Trump:	59%	45.38%	60.57%
Other:	4%	24.00%	5.78%

Congressional District:			
	AR Poll 2016 (LV):	*AR Poll 2016 (Full Sample):*	*Census 2016 Estimates:*
District 1:	26.91%	28.90%	24.44%
District 2:	22.29%	20.15%	25.37%
District 3:	26.51%	26.24%	26.16%
District 4:	24.30%	24.71%	24.03%

Gender:			
	AR Poll 2016 (LV):	*AR Poll 2016 (Full Sample):*	*Census 2016 Estimates:*
Male:	45.24%	44.63%	49.15%
Female:	54.76%	55.38%	50.85%

Race:			
	AR Poll 2016 (LV):	*AR Poll 2016 (Full Sample):*	*Census 2016 Estimates:*
White:	83.94%	85.25%	72.79%
Black or African American:	10.37%	9.40%	15.47%
Hispanic:	0.41%	0.65%	7.18%
Asian:	0.20%	0.13%	1.38%
Native American:	2.24%	1.83%	0.54%
Multi-Ethnic:	2.85%	2.74%	2.24%
Other/DK/Refused:	NA	NA	0.39%

Educational Attainment:			
	AR Poll 2016 (LV):	*AR Poll 2016 (Full Sample):*	*Census 2016 Estimates:*
No high school:	1.20%	1.54%	4.66%
Some high school:	6.39%	9.62%	10.46%
High school graduate:	24.35%	27.95%	36.24%
Some college including business or trade school [ASSOCIATES DEGREE]:	26.15%	24.74%	26.85%
College graduate [BACHELOR'S DEGREE]:	26.35%	22.82%	14.19%
Graduate or professional degree [MASTERS OR DOCTORATE DEGREE]:	15.57%	13.33%	7.60%

Age:			
	AR Poll 2016 (LV):	*AR Poll 2016 (Full Sample):*	*Census 2016 Estimates:*
18-24:	3.05%	4.79%	12.73%
25-34:	6.92%	7.58%	16.77%
35-44:	7.54%	8.51%	16.00%
45-54:	13.24%	13.70%	16.62%
55-64:	19.35%	18.75%	16.58%
65-74:	26.27%	23.27%	12.46%
75-84:	16.90%	16.49%	6.38%
85+:	6.72%	6.91%	2.45%

Income:			
	AR Poll 2016 (LV):	*AR Poll 2016 (Full Sample):*	*Census 2016 Estimates:*
$15,000 or Less:	14.43%	19.49%	15.40%
$15,001 - $25,000	11.60%	11.86%	12.70%
$25,001 - $35,000	12.63%	12.88%	11.90%
$35,001 - $50,000	17.01%	16.78%	15.20%
$50,001 - $75,000	17.78%	15.93%	18.30%
$75,001 - $100,000	12.89%	11.36%	10.40%
$100,001 +	13.66%	11.69%	16.10%

A.5 Missing Data Patterns

In Chapter 5.1, we mentioned that it was necessary to compare the respondents with missing data across demographics to ensure that significant patterns did not exist which could create bias. For example, we needed to be sure that when we used listwise deletion to remove those with missing data from a specific variable, no groups were disproportionately decreased due to deletion. To do this, we created a data frame for every variable that included those respondents who were missing that variable. Then, we compared the variable demographics of each of those samples to the original full sample. No significant patterns were found.

The tables were printed and analyzed in R, and can be done so again using the replication code in Appendix C. The code for these pattern tables is available in Part 3 (line 444) of the R code file.

Appendix B: Full Results of Weighting Trials

In Chapter 5.2, we included the full data tables for our presidential weighting trials. However, for the issue question weighting in Chapter 5.3, we opted to include summary tables rather than the data for all 52 trials. Full data tables for every trial, both presidential choice and issue question, are included in this appendix.

B.1 Presidential Choice Weighting Results

2020 Weighting Trials – Presidential Choice:

All Variables:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	36.63%	-2.57%	1.85%
Trump:	65.30%	62.40%	60.46%	2.90%	-1.94%
Other:	2.49%	2.82%	2.92%	-0.33%	0.10%
Respondents: 508			Gen. Design Effect: 3.101397		
Weighting Effect: *More Accurate*					

Gender Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.14%	-2.57%	-2.64%
Trump:	65.30%	62.40%	65.38%	2.90%	2.98%
Other:	2.49%	2.82%	2.48%	-0.33%	-0.34%
Respondents: 562			Gen. Design Effect: 1.001344		
Weighting Effect: *Slight Change*					

Congressional District Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.36%	-2.57%	-2.42%
Trump:	65.30%	62.40%	65.06%	2.90%	2.66%
Other:	2.49%	2.82%	2.59%	-0.33%	-0.23%
Respondents: 548			Gen. Design Effect: 1.009808		
Weighting Effect: *Slight Change*					

Race Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	39.66%	-2.57%	4.88%
Trump:	65.30%	62.40%	58.18%	2.90%	-4.22%
Other:	2.49%	2.82%	2.16%	-0.33%	-0.66%
Respondents: 557			Gen. Design Effect: 1.375674		
Weighting Effect: *More Extreme*					

Age Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.44%	-2.57%	-2.34%
Trump:	65.30%	62.40%	63.58%	2.90%	1.18%
Other:	2.49%	2.82%	3.98%	-0.33%	1.16%
Respondents: 551			Gen. Design Effect: 1.672223		
Weighting Effect: *More Accurate*					

Education Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	30.34%	-2.57%	-4.44%
Trump:	65.30%	62.40%	67.28%	2.90%	4.88%
Other:	2.49%	2.82%	2.37%	-0.33%	-0.45%
Respondents: 561			Gen. Design Effect: 1.518178		
Weighting Effect: *More Extreme*					

Income Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.68%	-2.57%	-2.10%
Trump:	65.30%	62.40%	64.58%	2.90%	2.18%
Other:	2.49%	2.82%	2.74%	-0.33%	-0.08%
Respondents: 533			Gen. Design Effect: 1.009051		
Weighting Effect: *More Accurate*					

Age, Race, Education, Income (Gender and District Removed)					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	36.29%	-2.57%	1.51%
Trump:	65.30%	62.40%	61.17%	2.90%	-1.23%
Other:	2.49%	2.82%	2.53%	-0.33%	-0.29%
Respondents: 521			Gen. Design Effect: 2.900476		
Weighting Effect: *More Accurate*					

Income and Age (produced accurate results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	32.52%	-2.57%	-2.26%
Trump:	65.30%	62.40%	63.36%	2.90%	0.96%
Other:	2.49%	2.82%	4.12%	-0.33%	1.30%
Respondents: 526			Gen. Design Effect: 1.641491		
Weighting Effect: *More Accurate*					

Education and Race (produced extreme results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Biden:	32.21%	34.78%	38.37%	-2.57%	3.59%
Trump:	65.30%	62.40%	59.63%	2.90%	-2.77%
Other:	2.49%	2.82%	2.00%	-0.33%	-0.82%
Respondents: 556			Gen. Design Effect: 1.991669		
Weighting Effect: *Slight Change*					

2016 Weighting Trials – Presidential Choice

All Variables:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	34.44%	2.66%	0.79%
Trump:	59.33%	60.57%	61.12%	-1.24%	0.55%
Other:	4.37%	5.78%	4.44%	-1.41%	-1.34%
Respondents: 373			Gen. Design Effect: 4.542968		
Weighting Effect: *More Accurate*					

Gender Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	36.07%	2.66%	2.42%
Trump:	59.33%	60.57%	59.38%	-1.24%	-1.19%
Other:	4.37%	5.78%	4.55%	-1.41%	-1.23%
Respondents: 504			Gen. Design Effect: 1.006177		
Weighting Effect: *Slight Change*					

Congressional District Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	36.69%	2.66%	3.04%
Trump:	59.33%	60.57%	59.00%	-1.24%	-1.57%
Other:	4.37%	5.78%	4.31%	-1.41%	-1.47%
Respondents: 498			Gen. Design Effect: 1.006596		
Weighting Effect: *Slight Change*					

Race Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	41.21%	2.66%	7.56%
Trump:	59.33%	60.57%	55.22%	-1.24%	-5.35%
Other:	4.37%	5.78%	3.57%	-1.41%	-2.21%
Respondents: 492			Gen. Design Effect: 2.260936		
Weighting Effect: *More Extreme*					

Age Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	33.80%	2.66%	0.15%
Trump:	59.33%	60.57%	59.92%	-1.24%	-0.65%
Other:	4.37%	5.78%	6.28%	-1.41%	0.50%
Respondents: 491			Gen. Design Effect: 1.719485		
Weighting Effect: *More Accurate*					

Education Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	35.68%	2.66%	2.03%
Trump:	59.33%	60.57%	60.63%	-1.24%	0.06%
Other:	4.37%	5.78%	3.69%	-1.41%	-2.09%
Respondents: 501			Gen. Design Effect: 1.281187		
Weighting Effect: *More Accurate*					

Income Only:					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	36.17%	2.66%	2.52%
Trump:	59.33%	60.57%	59.34%	-1.24%	-1.23%
Other:	4.37%	5.78%	4.49%	-1.41%	-1.29%
Respondents: 388			Gen. Design Effect: 1.01335		
Weighting Effect: *Slight Change*					

Age, Race, Education, Income (Gender and District Removed)					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	33.68%	2.66%	0.03%
Trump:	59.33%	60.57%	62.33%	-1.24%	1.76%
Other:	4.37%	5.78%	3.99%	-1.41%	-1.79%
Respondents: 376			Gen. Design Effect: 4.409438		
Weighting Effect: *Slight Change*					

Income and Age (produced accurate results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	34.99%	-2.57%	1.34%
Trump:	59.33%	60.57%	58.42%	2.90%	-2.15%
Other:	4.37%	5.78%	6.59%	-0.33%	0.81%
Respondents: 383			Gen. Design Effect: 1.604377		
Weighting Effect: *More Accurate*					

Education and Race (produced extreme results individually):					
	AR Poll (Likely Voters):	Actual Election Results:	Weighted AR Poll Results:	% Difference (AR Poll to Actual)	% Difference (Weighted to Actual)
Clinton:	36.31%	33.65%	45.51%	-2.57%	11.86%
Trump:	59.33%	60.57%	55.33%	2.90%	-5.24%
Other:	4.37%	5.78%	3.17%	-0.33%	-2.61%
Respondents: 490			Gen. Design Effect: 2.868935		
Weighting Effect: *More Extreme*					

B.2 Issue Question Weighting Results

2020 - Abortion

Age – Abortion, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 741
More Difficult:	47.77%	42.89%	-4.88%	
Easier:	15.65%	21.70%	6.05%	
No Change:	36.57%	35.41%	-1.17%	Gen. Design Effect: 1.600102

Income – Abortion, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 718
More Difficult:	46.80%	47.26%	0.46%	
Easier:	16.30%	16.15%	-0.15%	
No Change:	36.91%	36.59%	-0.31%	Gen. Design Effect: 1.020411

Age and Income – Abortion, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 706
More Difficult:	46.88%	42.73%	-4.15%	
Easier:	16.15%	21.60%	5.45%	
No Change:	36.97%	35.67%	-1.29%	Gen. Design Effect: 1.632359

Race – Abortion, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 750
More Difficult:	47.60%	45.05%	-2.55%	
Easier:	15.47%	17.48%	2.02%	
No Change:	36.93%	37.47%	0.53%	Gen. Design Effect: 1.247403

Education – Abortion, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 759
More Difficult:	47.69%	48.12%	0.43%	
Easier:	15.68%	14.54%	-1.14%	
No Change:	36.63%	37.34%	0.71%	Gen. Design Effect: 1.197067

Race and Education – Abortion, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 749
More Difficult:	47.53%	45.12%	-2.41%	
Easier:	15.49%	16.10%	0.62%	
No Change:	36.98%	38.77%	1.79%	Gen. Design Effect: 1.472398

All Variables – Abortion, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 696
More Difficult:	46.70%	42.80%	-3.90%	
Easier:	16.09%	19.90%	3.80%	
No Change:	37.21%	37.31%	0.09%	Gen. Design Effect: 2.429891

2020 – Gun Control

Age – Gun Control, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 764
Stricter:	32.07%	32.98%	0.91%	
Less Strict:	17.28%	21.22%	3.95%	
No Change:	50.65%	45.80%	-4.86%	Gen. Design Effect: 1.594978

Income – Gun Control, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 744
Stricter:	32.80%	32.59%	-0.20%	
Less Strict:	16.67%	16.63%	-0.04%	
No Change:	50.54%	50.78%	0.24%	Gen. Design Effect: 1.015819

Age and Income – Gun Control, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 731
Stricter:	32.56%	33.06%	0.50%	
Less Strict:	16.83%	20.57%	3.75%	
No Change:	50.62%	46.37%	-4.25%	Gen. Design Effect: 1.610003

Race – Gun Control, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 772
Stricter:	32.51%	37.65%	5.14%	
Less Strict:	16.45%	14.78%	-1.67%	
No Change:	51.04%	47.57%	-3.47%	Gen. Design Effect: 1.247388

Education – Gun Control, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 784
Stricter:	32.40%	31.20%	-1.20%	
Less Strict:	17.09%	17.08%	-0.01%	
No Change:	50.51%	51.72%	1.21%	Gen. Design Effect: 1.181413

Race and Education – Gun Control, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 771
Stricter:	32.56%	36.88%	4.32%	
Less Strict:	16.47%	15.11%	-1.36%	
No Change:	50.97%	48.01%	-2.96%	Gen. Design Effect: 1.455007

All Variables – Gun Control, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 720
Stricter:	32.50%	34.89%	2.39%	
Less Strict:	16.53%	20.05%	3.52%	
No Change:	50.97%	45.06%	-5.91%	Gen. Design Effect: 2.36386

2020 – Climate Change

Age – Climate Change, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 767
Yes:	36.38%	42.71%	6.33%	
No:	62.58%	56.12%	-6.47%	
Not a Problem:	1.04%	1.18%	0.14%	Gen. Design Effect: 1.562697

Income – Climate Change, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 742
Yes:	37.20%	36.54%	-0.66%	
No:	61.86%	62.53%	0.67%	
Not a Problem:	0.94%	0.93%	-0.02%	Gen. Design Effect: 1.017391

Age and Income – Climate Change, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 729
Yes:	36.63%	42.00%	5.38%	
No:	62.41%	56.80%	-5.61%	
Not a Problem:	0.96%	1.20%	0.24%	Gen. Design Effect: 1.582889

Race – Climate Change, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 776
Yes:	37.24%	41.67%	4.43%	
No:	61.73%	57.46%	-4.27%	
Not a Problem:	1.03%	0.87%	-0.16%	Gen. Design Effect: 1.233719

Education – Climate Change, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 788
Yes:	37.18%	35.58%	-1.61%	
No:	61.68%	63.18%	1.50%	
Not a Problem:	1.14%	1.25%	0.10%	Gen. Design Effect: 1.182054

Race and Education – Climate Change, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 775
Yes:	37.29%	40.06%	2.77%	
No:	61.68%	58.94%	-2.73%	
Not a Problem:	1.03%	1.00%	-0.04%	Gen. Design Effect: 1.442296

All Variables – Climate Change, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 718
Yes:	36.63%	41.11%	4.48%	
No:	62.40%	57.45%	-4.95%	
Not a Problem:	0.97%	1.44%	0.47%	Gen. Design Effect: 2.324929

2020 – Direction

Age – Direction, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 752
Right:	80.98%	75.90%	-5.08%	Gen. Design Effect: 1.592463
Wrong:	19.02%	24.10%	5.08%	

Income – Direction, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 729
Right:	80.66%	80.63%	-0.02%	Gen. Design Effect: 1.020737
Wrong:	19.34%	19.37%	0.03%	

Age and Income – Direction, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 718
Right:	80.92%	76.32%	-4.60%	Gen. Design Effect: 1.622487
Wrong:	19.08%	23.68%	4.60%	

Race – Direction, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 760
Right:	80.53%	79.06%	-1.46%	Gen. Design Effect: 1.232487
Wrong:	19.47%	20.94%	1.46%	

Education – Direction, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 771
Right:	80.54%	81.40%	0.85%	Gen. Design Effect: 1.197821
Wrong:	19.46%	18.60%	-0.85%	

Race and Education – Direction, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 759
Right:	80.50%	79.99%	-0.51%	Gen. Design Effect: 1.457294
Wrong:	19.50%	20.01%	0.51%	

All Variables – Direction, 2020				
	AR Poll 2020:	Weighted Results:	Percent Difference:	Respondents: 708
Right:	80.93%	76.59%	-4.34%	Gen. Design Effect: 2.358094
Wrong:	19.07%	23.41%	4.34%	

2016 – Abortion

Age – Abortion, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 683
More Difficult:	50.81%	50.23%	-0.58%	
Easier:	13.62%	13.72%	0.10%	
No Change:	35.58%	36.05%	0.47%	Gen. Design Effect: 1.47335

Income – Abortion, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 551
More Difficult:	50.45%	50.34%	-0.12%	
Easier:	13.97%	14.11%	0.13%	
No Change:	35.57%	35.55%	-0.02%	Gen. Design Effect: 1.028747

Age and Income – Abortion, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 540
More Difficult:	50.56%	49.57%	-0.98%	
Easier:	13.89%	14.00%	0.11%	
No Change:	35.56%	36.43%	0.88%	Gen. Design Effect: 1.434633

Race – Abortion, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 699
More Difficult:	50.36%	48.99%	-1.37%	
Easier:	14.16%	13.78%	-0.39%	
No Change:	35.48%	37.24%	1.76%	Gen. Design Effect: 1.762027

Education – Abortion, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 712
More Difficult:	49.86%	51.80%	1.94%	
Easier:	14.19%	13.10%	-1.08%	
No Change:	35.96%	35.10%	-0.86%	Gen. Design Effect: 1.193939

Race and Education – Abortion, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 693
More Difficult:	49.93%	49.84%	-0.09%	
Easier:	14.29%	12.91%	-1.37%	
No Change:	35.79%	37.25%	1.46%	Gen. Design Effect: 2.070869

All Variables – Abortion, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 527
More Difficult:	50.47%	50.34%	-0.13%	
Easier:	14.23%	11.63%	-2.60%	
No Change:	35.29%	38.02%	2.73%	Gen. Design Effect: 2.773603

2016 – Gun Control

Age – Gun Control, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 729
Stricter:	32.24%	29.63%	-2.61%	
Less Strict:	14.13%	16.93%	2.80%	
No Change:	53.64%	53.45%	-0.19%	Gen. Design Effect: 1.45636

Income – Gun Control, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 578
Stricter:	32.35%	32.07%	-0.29%	
Less Strict:	12.80%	12.49%	-0.32%	
No Change:	54.84%	55.45%	0.60%	Gen. Design Effect: 1.02846

Age and Income – Gun Control, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 567
Stricter:	32.10%	29.42%	-2.68%	
Less Strict:	12.87%	15.33%	2.46%	
No Change:	55.03%	55.24%	0.22%	Gen. Design Effect: 1.41357

Race – Gun Control, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 745
Stricter:	32.35%	34.21%	1.86%	
Less Strict:	13.96%	12.22%	-1.74%	
No Change:	53.69%	53.57%	-0.12%	Gen. Design Effect: 1.81597

Education – Gun Control, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 759
Stricter:	32.41%	32.22%	-0.19%	
Less Strict:	14.10%	14.28%	0.18%	
No Change:	53.49%	53.51%	0.02%	Gen. Design Effect: 1.18733

Race and Education – Gun Control, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 740
Stricter:	32.30%	33.40%	1.10%	
Less Strict:	13.92%	12.26%	-1.66%	
No Change:	53.78%	54.35%	0.56%	Gen. Design Effect: 2.09438

All Variables – Gun Control, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 554
Stricter:	32.13%	29.83%	-2.30%	
Less Strict:	13.00%	11.86%	-1.14%	
No Change:	54.87%	58.31%	3.44%	Gen. Design Effect: 2.74874

2016 – Climate Change

Age – Climate Change, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 683
Yes:	28.26%	32.89%	4.63%	
No:	70.28%	65.72%	-4.55%	
Not a Problem:	1.46%	1.38%	-0.08%	Gen. Design Effect: 1.443005

Income – Climate Change, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 548
Yes:	29.38%	29.00%	-0.38%	
No:	69.16%	69.62%	0.46%	
Not a Problem:	1.46%	1.37%	-0.09%	Gen. Design Effect: 1.030334

Age and Income – Climate Change, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 538
Yes:	29.55%	33.45%	3.90%	
No:	68.96%	65.14%	-3.82%	
Not a Problem:	1.49%	1.41%	-0.07%	Gen. Design Effect: 1.419846

Race – Climate Change, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 695
Yes:	27.77%	34.39%	6.62%	
No:	70.79%	64.27%	-6.52%	
Not a Problem:	1.44%	1.34%	-0.09%	Gen. Design Effect: 1.767024

Education – Climate Change, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 711
Yes:	28.27%	26.50%	-1.77%	
No:	70.32%	72.02%	1.70%	
Not a Problem:	1.41%	1.48%	0.07%	Gen. Design Effect: 1.149263

Race and Education – Climate Change, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 693
Yes:	27.71%	33.40%	5.70%	
No:	70.85%	65.28%	-5.58%	
Not a Problem:	1.44%	1.32%	-0.12%	Gen. Design Effect: 2.009392

All Variables – Climate Change, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 528
Yes:	29.36%	36.81%	7.45%	
No:	69.13%	61.66%	-7.47%	
Not a Problem:	1.52%	1.54%	0.02%	Gen. Design Effect: 2.71327

2016 - Direction

Age – Direction, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 640
Right:	68.44%	67.83%	-0.61%	Gen. Design Effect: 1.45817
Wrong:	31.56%	32.17%	0.61%	

Income – Direction, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 513
Right:	69.20%	68.92%	-0.28%	Gen. Design Effect: 1.01805
Wrong:	30.80%	31.08%	0.28%	

Age and Income – Direction, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 503
Right:	69.18%	68.11%	-1.08%	Gen. Design Effect: 1.39526
Wrong:	30.82%	31.89%	1.08%	

Race – Direction, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 649
Right:	68.88%	67.41%	-1.46%	Gen. Design Effect: 1.86607
Wrong:	31.12%	32.59%	1.46%	

Education – Direction, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 661
Right:	68.84%	69.69%	0.85%	Gen. Design Effect: 1.17839
Wrong:	31.16%	30.31%	-0.85%	

Race and Education – Direction, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 643
Right:	69.05%	68.94%	-0.11%	Gen. Design Effect: 2.21074
Wrong:	30.95%	31.06%	0.11%	

All Variables – Direction, 2016				
	AR Poll 2016:	Weighted Results:	Percent Difference:	Respondents: 490
Right:	69.18%	66.66%	-2.52%	Gen. Design Effect: 2.58369
Wrong:	30.82%	33.34%	2.52%	

Appendix C: R Code for Replication

All R code used for this thesis is printed in the remainder of this document. The code is divided into 7 parts, as follows:

Begin Code Here:

```
# PAPER TITLE: Understanding and Improving the System:
# The Effects of Weighting on the Accuracy of Political Polling in Arkansas
# AUTHOR: Beck Williams
# THESIS DIRECTOR: Dr. Todd Shields, University of Arkansas, J. William Fulbright College of Arts and
Sciences
# FINAL EDIT DATE: April 8th, 2022#####

# TABLE OF CONTENTS:
### PART 1: Coding the 2020 Arkansas Poll Data -       line 29    ###
### PART 2: Coding the 2016 Arkansas Poll Data -       line 240   ###
### PART 3: Checking for Patterns in Missing Data -    line 444   ###
### PART 4: Weighting the 2020 Presidential Choice -   line 625   ###
### PART 5: Weighting the 2016 Presidential Choice -   line 921   ###
### PART 6: Weighting the 2020 Issue Questions -       line 1220  ###
### PART 7: Weighting the 2016 Issue Questions -       line 1712  ###

#Open packages
library(foreign)
library(sjmisc)
library(plyr)
library(tidyverse)
library(readstata13)
library(haven)
library(weights)
```

```
library(anesrake)

#Set Working Directory
setwd("/Users/beck/Downloads")

##### PART 1: Coding The 2020 Arkansas Poll Data #####

# Read 2020 AR Poll Data
AR2020 <- read.dta13("Arkansas_Poll_2020.dta", generate.factors=TRUE)
View(AR2020)
### At this point, the entire Arkansas Poll 2020 data is stored in 'AR2020'

# Change the name of specific columns so that we may use them in the code:
names(AR2020)[names(AR2020) == '_v2'] <- "LV"
names(AR2020)[names(AR2020) == '_v1'] <- "caseid"

# Select variables
myvars <- c("q49", "q2", "q99a", "Age", "q21", "q20", "LV", "q36", "caseid")
##Variable Key:
#'q2' = County (recoded later as district)
#'q49' = Gender
#'q99a' = Presidential Choice
#'Age' = Age of Respondent
#'q21' = Race
#'q20' = Education (highest level achieved)
#'q36' = Household Income
#'LV' = Likely Voters
#'caseid' = Case ID (respondent identifier)
AR2020.brief <- AR2020[myvars]
view(AR2020.brief)
### This 'AR2020.brief" contains only the variables of concern for our code

# Label variables
AR2020.brief$pres_vote <- AR2020.brief$q99a
AR2020.brief$gender    <- AR2020.brief$q49
AR2020.brief$county    <- AR2020.brief$q2
AR2020.brief$age       <- AR2020.brief$Age
AR2020.brief$race      <- AR2020.brief$q21
AR2020.brief$education <- AR2020.brief$q20
AR2020.brief$income    <- AR2020.brief$q36
AR2020.brief$LV        <- AR2020.brief$LV
AR2020.brief$caseid    <- AR2020.brief$caseid

# Recode variables
AR2020.brief$pres_vote <- ifelse(AR2020.brief$pres_vote == "Joe Biden [BUY din], the Democrat", 0, 0)
AR2020.brief$gender <- ifelse(AR2020.brief$gender == "Female", 0, 0)
AR2020.brief$race <- ifelse(AR2020.brief$race == "White", 1, 0)
AR2020.brief$education <- ifelse(AR2020.brief$q20 == "No high school", 0, 0)

# Coding for gender 1 = female, 2 = male, 0 = NA/refused
AR2020.brief$gender[AR2020.brief$q49 == "Female"]   <- 1
AR2020.brief$gender[AR2020.brief$q49 == "Male"]     <- 2

# Coding vote for Trump 1, Biden 2, and others=3
AR2020.brief$pres_vote[AR2020.brief$q99a == "Donald Trump, the Republican"]       <- 1
AR2020.brief$pres_vote[AR2020.brief$q99a == "Joe Biden [BUY din], the Democrat"]  <- 2
```

```
AR2020.brief$pres_vote[AR2020.brief$q99a == "Another candidate [specify]:"]      <- 3

# Coding race for categories included in Census and ARPoll
AR2020.brief$race[AR2020.brief$q21 == "White"]                     <- 1
AR2020.brief$race[AR2020.brief$q21 == "Black or African-American"]  <- 2
AR2020.brief$race[AR2020.brief$q21 == "Hispanic"]                  <- 3
AR2020.brief$race[AR2020.brief$q21 == "Asian"]                     <- 4
AR2020.brief$race[AR2020.brief$q21 == "Native American"]           <- 5
AR2020.brief$race[AR2020.brief$q21 == "Multi-Ethnic"]              <- 6
AR2020.brief$race[AR2020.brief$q21 == "Other"]                     <- 7

# Coding education level
AR2020.brief$education[AR2020.brief$q20 == "No high school"]                                         <- 1
AR2020.brief$education[AR2020.brief$q20 == "Some high school"]                                       <- 2
AR2020.brief$education[AR2020.brief$q20 == "High school graduate"]                                   <- 3
AR2020.brief$education[AR2020.brief$q20 == "Some college including business or trade school [ASSOCIATES
DEGREE]"]  <- 4
AR2020.brief$education[AR2020.brief$q20 == "College graduate [BACHELORS DEGREE]"]
<- 5 #grouped with below
AR2020.brief$education[AR2020.brief$q20 == "Some graduate school"]                                   <- 5
#grouped with above
AR2020.brief$education[AR2020.brief$q20 == "Graduate or professional degree [MASTERS OR DOCTORATE
DEGREE]"]       <- 6

# Coding age by decades
AR2020.brief$age[  AR2020.brief$age == "18" | AR2020.brief$age == "19" | AR2020.brief$age == "20" |
                   AR2020.brief$age == "21" | AR2020.brief$age == "22" | AR2020.brief$age == "23" |
                   AR2020.brief$age == "24"]                                      <- 1
AR2020.brief$age[  AR2020.brief$age == "25" | AR2020.brief$age == "26" | AR2020.brief$age == "27" |
                   AR2020.brief$age == "28" | AR2020.brief$age == "29" | AR2020.brief$age == "30" |
                   AR2020.brief$age == "31" | AR2020.brief$age == "32" | AR2020.brief$age == "33" |
                   AR2020.brief$age == "34"]                                      <- 2
AR2020.brief$age[  AR2020.brief$age == "35" | AR2020.brief$age == "36" | AR2020.brief$age == "37" |
                   AR2020.brief$age == "38" | AR2020.brief$age == "39" | AR2020.brief$age == "40" |
                   AR2020.brief$age == "41" | AR2020.brief$age == "42" | AR2020.brief$age == "43" |
                   AR2020.brief$age == "44"]                                      <- 3
AR2020.brief$age[  AR2020.brief$age == "45" | AR2020.brief$age == "46" | AR2020.brief$age == "47" |
                   AR2020.brief$age == "48" | AR2020.brief$age == "49" | AR2020.brief$age == "50" |
                   AR2020.brief$age == "51" | AR2020.brief$age == "52" | AR2020.brief$age == "53" |
                   AR2020.brief$age == "54"]                                      <- 4
AR2020.brief$age[  AR2020.brief$age == "55" | AR2020.brief$age == "56" | AR2020.brief$age == "57" |
                   AR2020.brief$age == "58" | AR2020.brief$age == "59" | AR2020.brief$age == "60" |
                   AR2020.brief$age == "61" | AR2020.brief$age == "62" | AR2020.brief$age == "63" |
                   AR2020.brief$age == "64"]                                      <- 5
AR2020.brief$age[  AR2020.brief$age == "65" | AR2020.brief$age == "66" | AR2020.brief$age == "67" |
                   AR2020.brief$age == "68" | AR2020.brief$age == "69" | AR2020.brief$age == "70" |
                   AR2020.brief$age == "71" | AR2020.brief$age == "72" | AR2020.brief$age == "73" |
                   AR2020.brief$age == "74"]                                      <- 6
AR2020.brief$age[  AR2020.brief$age == "75" | AR2020.brief$age == "76" | AR2020.brief$age == "77" |
                   AR2020.brief$age == "78" | AR2020.brief$age == "79" | AR2020.brief$age == "80" |
                   AR2020.brief$age == "81" | AR2020.brief$age == "82" | AR2020.brief$age == "83" |
                   AR2020.brief$age == "84"]                                      <- 7
AR2020.brief$age[  AR2020.brief$age == "85" | AR2020.brief$age == "86" | AR2020.brief$age == "87" |
                   AR2020.brief$age == "88" | AR2020.brief$age == "89" | AR2020.brief$age == "90" |
                   AR2020.brief$age == "91" | AR2020.brief$age == "92" | AR2020.brief$age == "93" |
                   AR2020.brief$age == "94"| AR2020.brief$age == "95"]                     <- 8
```

```
# Coding counties into congressional districts
AR2020.brief$congdist <- as.character(AR2020.brief$q2)
AR2020.brief$congdist[AR2020.brief$congdist=="Arkansas [ARK-in-saw]" | AR2020.brief$congdist=="Baxter" |
        AR2020.brief$congdist=="Chicot" | AR2020.brief$congdist=="Clay" |
        AR2020.brief$congdist=="Cleburne" | AR2020.brief$congdist=="Craighead" |
        AR2020.brief$congdist=="Crittenden" | AR2020.brief$congdist=="Cross" |
        AR2020.brief$congdist=="Desha" | AR2020.brief$congdist=="Fulton" |
        AR2020.brief$congdist=="Greene" | AR2020.brief$congdist=="Independent" |
        AR2020.brief$congdist=="Izard" | AR2020.brief$congdist=="Jackson" |
        AR2020.brief$congdist=="Jefferson" | AR2020.brief$congdist=="Lawrence" |
        AR2020.brief$congdist=="Lee" | AR2020.brief$congdist=="Lincoln" |
        AR2020.brief$congdist=="Lonoke" | AR2020.brief$congdist=="Mississippi" |
        AR2020.brief$congdist=="Monroe" | AR2020.brief$congdist=="Philips" |
        AR2020.brief$congdist=="Poinsett" | AR2020.brief$congdist=="Prairie" |
        AR2020.brief$congdist=="Randolph" | AR2020.brief$congdist=="Searcy" |
        AR2020.brief$congdist=="Sharp" | AR2020.brief$congdist=="St. Francis" |
        AR2020.brief$congdist=="Stone" | AR2020.brief$congdist=="Woodruff"]   <- 1

AR2020.brief$congdist[AR2020.brief$congdist=="Conway" | AR2020.brief$congdist=="Faulkner" |
        AR2020.brief$congdist=="Perry" | AR2020.brief$congdist=="Pulaski" |
        AR2020.brief$congdist=="Saline" | AR2020.brief$congdist=="Van Buren" |
        AR2020.brief$congdist=="White" ]  <- 2

AR2020.brief$congdist[AR2020.brief$congdist=="Benton" | AR2020.brief$congdist=="Boone" |
        AR2020.brief$congdist=="Carroll" | AR2020.brief$congdist=="Crawford" |
        AR2020.brief$congdist=="Marion" | AR2020.brief$congdist=="Newton" |
        AR2020.brief$congdist=="Pope" | AR2020.brief$congdist=="Searcy" |
        AR2020.brief$congdist=="Sebastian" | AR2020.brief$congdist=="Washington" ] <- 3

AR2020.brief$congdist[AR2020.brief$congdist=="Ashley" | AR2020.brief$congdist=="Bradley" |
        AR2020.brief$congdist=="Calhoun" | AR2020.brief$congdist=="Clark" |
        AR2020.brief$congdist=="Cleveland" | AR2020.brief$congdist=="Columbia" |
        AR2020.brief$congdist=="Crawford" | AR2020.brief$congdist=="Dallas" |
        AR2020.brief$congdist=="Drew" | AR2020.brief$congdist=="Franklin" |
        AR2020.brief$congdist=="Garland" | AR2020.brief$congdist=="Grant" |
        AR2020.brief$congdist=="Hempstead" | AR2020.brief$congdist=="Hot Springs" |
        AR2020.brief$congdist=="Howard" | AR2020.brief$congdist=="Jefferson" |
        AR2020.brief$congdist=="Johnson" | AR2020.brief$congdist=="Lafayette" |
        AR2020.brief$congdist=="Little River" | AR2020.brief$congdist=="Logan" |
        AR2020.brief$congdist=="Madison" | AR2020.brief$congdist=="Miller" |
        AR2020.brief$congdist=="Montgomery" | AR2020.brief$congdist=="Nevada" |
        AR2020.brief$congdist=="Newton" | AR2020.brief$congdist=="Ouachita" |
        AR2020.brief$congdist=="Pike" | AR2020.brief$congdist=="Polk" |
        AR2020.brief$congdist=="Scott" | AR2020.brief$congdist=="Sebastian" |
        AR2020.brief$congdist=="Sevier" | AR2020.brief$congdist=="Union" |
        AR2020.brief$congdist=="Yell" ] <- 4

# Coding household income
AR2020.brief$income <- as.character(AR2020.brief$income)
AR2020.brief$income[AR2020.brief$income == "$7,500 or less"]         <- 1 #15,0000 or less
AR2020.brief$income[AR2020.brief$income == "$7,501 to $15,000"]      <- 1 #15,0000 or less
AR2020.brief$income[AR2020.brief$income == "$15,001 to $25,000"]     <- 2
AR2020.brief$income[AR2020.brief$income == "$25,001 to $35,000"]     <- 3
AR2020.brief$income[AR2020.brief$income == "$35,001 to $50,000"]     <- 4
AR2020.brief$income[AR2020.brief$income == "$50,001 to $75,000"]     <- 5
```

```
AR2020.brief$income[AR2020.brief$income == "$75,001 to $100,000"]     <- 6
AR2020.brief$income[AR2020.brief$income == "$100,001 or over"]        <- 7
AR2020.brief$income[AR2020.brief$income == "[DO NOT READ] Dont know"] <- 0
AR2020.brief$income[AR2020.brief$income == "[DO NOT READ] Refused"]   <- 0
AR2020.brief$income <- as.integer(AR2020.brief$income)

# New version of AR2020 Data Frame
View(AR2020.brief)
### Note that this data frame contains all respondents. Next, we will remove those who are not Likely Voters.

# Create table of ONLY Likely Voters
AR2020.brief.LVandNA <- subset(AR2020.brief, LV == "Selected")
AR2020.brief.LV <- subset(AR2020.brief.LVandNA, pres_vote == "1" | pres_vote == "2" | pres_vote == "3")
view(AR2020.brief.LV)
### The data frame 'AR2020.brief.LV' contains only the likely voters  who answered the presidential choice
question.
### Each of the following tables that includes "LV" in the name only includes Likely Voters.
### If no "LV" in name, then the table includes all respondents.

# Print tables for all variables
table_pres_vote <- table(AR2020.brief$pres_vote)
table_pres_vote
prop.table(table_pres_vote)
table_pres_vote_LV <- table(AR2020.brief.LV$pres_vote)
table_pres_vote_LV
prop.table(table_pres_vote_LV)
table_congdist <- table(AR2020.brief$congdist)
table_congdist
table_congdist_LV <- table(AR2020.brief.LV$congdist)
table_congdist_LV
table_gender <- table(AR2020.brief$gender)
table_gender
table_gender_LV <- table(AR2020.brief.LV$gender)
table_gender_LV
table_race <- table(AR2020.brief$race)
table_race
table_race_LV <- table(AR2020.brief.LV$race)
table_race_LV
table_age <- table(AR2020.brief$age)
table_age
table_age_LV <- table(AR2020.brief.LV$age)
table_age_LV
table_education <- table(AR2020.brief$education)
table_education
table_education_LV <- table(AR2020.brief.LV$education)
table_education_LV
table_income <- table(AR2020.brief$income)
table_income
table_income_LV <- table(AR2020.brief.LV$income)
table_income_LV



##### PART 2: Coding The 2016 Arkansas Poll Data #####

# Read 2016 AR Poll Data
```

```
AR2016 <- read_sav("2016-data.sav")
View(AR2016)
###At this point, the entire Arkansas Poll 2016 data is stored in 'AR2016'

# Select variables
myvars2 <- c("Q49", "Q2", "Q99A", "Q48", "Q21", "Q20", "Q47", "Q36", "ID")
##Variable Key:
#'Q2' = County (recoded later as district)
#'Q49' = Gender
#'Q99A' = Presidential Choice
#'Q48' = Birth Year of Respondent
#'Q21' = Race
#'Q20' = Education (highest level achieved)
#'Q47' = Likely Voters
#'Q36' = Household Income
#'ID' = Case ID (respondent identifier)
AR2016.brief <- AR2016[myvars2]
view(AR2016.brief)
### This 'AR2016.brief" contains only the variables of concern for our code
AR2016.brief <- as.data.frame(AR2016.brief)

# Label variables
AR2016.brief$pres_vote <- AR2016.brief$Q99A
AR2016.brief$gender    <- AR2016.brief$Q49
AR2016.brief$county    <- AR2016.brief$Q2
AR2016.brief$age       <- AR2016.brief$Q48
AR2016.brief$race      <- AR2016.brief$Q21
AR2016.brief$education <- AR2016.brief$Q20
AR2016.brief$LV        <- AR2016.brief$Q47
AR2016.brief$income    <- AR2016.brief$Q36
AR2016.brief$caseid    <- AR2016.brief$ID

# Coding gender
AR2016.brief$gender[AR2016.brief$gender == 2]   <- 3 #(female)
AR2016.brief$gender[AR2016.brief$gender == 1]   <- 2 #(male)
AR2016.brief$gender[AR2016.brief$gender == 3]   <- 1 #(female)


# Coding education level
AR2016.brief$education[AR2016.brief$Q20 == 1]   <- 1
AR2016.brief$education[AR2016.brief$Q20 == 2]   <- 2
AR2016.brief$education[AR2016.brief$Q20 == 3]   <- 3
AR2016.brief$education[AR2016.brief$Q20 == 4]   <- 4
AR2016.brief$education[AR2016.brief$Q20 == 5]   <- 5 #grouped with below
AR2016.brief$education[AR2016.brief$Q20 == 6]   <- 5 #grouped with above
AR2016.brief$education[AR2016.brief$Q20 == 7]   <- 6

# Coding age by decades
AR2016.brief$age[  AR2016.brief$age == 1992 | AR2016.brief$age == 1993 | AR2016.brief$age == 1994 |
               AR2016.brief$age == 1995 | AR2016.brief$age == 1996 | AR2016.brief$age == 1997 |
               AR2016.brief$age == 1998]                                <- 1 #18-24
AR2016.brief$age[  AR2016.brief$age == 1991 | AR2016.brief$age == 1990 | AR2016.brief$age == 1989 |
               AR2016.brief$age == 1988 | AR2016.brief$age == 1987 | AR2016.brief$age == 1986 |
               AR2016.brief$age == 1985 | AR2016.brief$age == 1984 | AR2016.brief$age == 1983 |
               AR2016.brief$age == 1982]                                <- 2 #25-34
AR2016.brief$age[  AR2016.brief$age == 1981 | AR2016.brief$age == 1980 | AR2016.brief$age == 1979 |
```

```
                    AR2016.brief$age == 1978 | AR2016.brief$age == 1977 | AR2016.brief$age == 1976 |
                    AR2016.brief$age == 1975 | AR2016.brief$age == 1974 | AR2016.brief$age == 1973 |
                    AR2016.brief$age == 1972]                                   <- 3 #35-44
AR2016.brief$age[  AR2016.brief$age == 1971 | AR2016.brief$age == 1970 | AR2016.brief$age == 1969 |
                    AR2016.brief$age == 1968 | AR2016.brief$age == 1967 | AR2016.brief$age == 1966 |
                    AR2016.brief$age == 1965 | AR2016.brief$age == 1964 | AR2016.brief$age == 1963 |
                    AR2016.brief$age == 1962]                                   <- 4 #45-54
AR2016.brief$age[  AR2016.brief$age == 1961 | AR2016.brief$age == 1960 | AR2016.brief$age == 1959 |
                    AR2016.brief$age == 1958 | AR2016.brief$age == 1957 | AR2016.brief$age == 1956 |
                    AR2016.brief$age == 1955 | AR2016.brief$age == 1954 | AR2016.brief$age == 1953 |
                    AR2016.brief$age == 1952]                                   <- 5 #55-64
AR2016.brief$age[  AR2016.brief$age == 1951 | AR2016.brief$age == 1950 | AR2016.brief$age == 1949 |
                    AR2016.brief$age == 1948 | AR2016.brief$age == 1947 | AR2016.brief$age == 1946 |
                    AR2016.brief$age == 1945 | AR2016.brief$age == 1944 | AR2016.brief$age == 1943 |
                    AR2016.brief$age == 1942]                                   <- 6 #65-74
AR2016.brief$age[  AR2016.brief$age == 1941 | AR2016.brief$age == 1940 | AR2016.brief$age == 1939 |
                    AR2016.brief$age == 1938 | AR2016.brief$age == 1937 | AR2016.brief$age == 1936 |
                    AR2016.brief$age == 1935 | AR2016.brief$age == 1934 | AR2016.brief$age == 1933 |
                    AR2016.brief$age == 1932]                                   <- 7 #74-85
AR2016.brief$age[  AR2016.brief$age == 1931 | AR2016.brief$age == 1930 | AR2016.brief$age == 1929 |
                    AR2016.brief$age == 1928 | AR2016.brief$age == 1927 | AR2016.brief$age == 1926 |
                    AR2016.brief$age == 1925 | AR2016.brief$age == 1924 | AR2016.brief$age == 1923 |
                    AR2016.brief$age == 1922 | AR2016.brief$age == 1921 | AR2016.brief$age == 1920 |
                    AR2016.brief$age == 1919 | AR2016.brief$age == 1918 | AR2016.brief$age == 1917 |
                    AR2016.brief$age == 1916 | AR2016.brief$age == 1915 | AR2016.brief$age == 1914 |
                    AR2016.brief$age == 1913 | AR2016.brief$age == 1912]              <- 8 #85+

# Coding counties into congressional districts
AR2016.brief$congdist <- as.character(AR2016.brief$county)
AR2016.brief$congdist[AR2016.brief$congdist=="1" | AR2016.brief$congdist=="2" |
                    AR2016.brief$congdist=="50" | AR2016.brief$congdist=="3" |
                    AR2016.brief$congdist=="4" | AR2016.brief$congdist=="5" |
                    AR2016.brief$congdist=="6" | AR2016.brief$congdist=="7" |
                    AR2016.brief$congdist=="55" | AR2016.brief$congdist=="8" |
                    AR2016.brief$congdist=="9" | AR2016.brief$congdist=="10" |
                    AR2016.brief$congdist=="11" | AR2016.brief$congdist=="12" |
                    AR2016.brief$congdist=="62" | AR2016.brief$congdist=="13" |
                    AR2016.brief$congdist=="14" | AR2016.brief$congdist=="64" |
                    AR2016.brief$congdist=="15" | AR2016.brief$congdist=="16" |
                    AR2016.brief$congdist=="17" | AR2016.brief$congdist=="18" |
                    AR2016.brief$congdist=="19" | AR2016.brief$congdist=="20" |
                    AR2016.brief$congdist=="21" | AR2016.brief$congdist=="23" |
                    AR2016.brief$congdist=="24" | AR2016.brief$congdist=="22" |
                    AR2016.brief$congdist=="25" | AR2016.brief$congdist=="26"]   <- 1

AR2016.brief$congdist[AR2016.brief$congdist=="27" | AR2016.brief$congdist=="28" |
                    AR2016.brief$congdist=="29" | AR2016.brief$congdist=="30" |
                    AR2016.brief$congdist=="31" | AR2016.brief$congdist=="32" |
                    AR2016.brief$congdist=="33" ]  <- 2

AR2016.brief$congdist[AR2016.brief$congdist=="35" | AR2016.brief$congdist=="36" |
                    AR2016.brief$congdist=="37" | AR2016.brief$congdist=="38" |
                    AR2016.brief$congdist=="42" | AR2016.brief$congdist=="43" |
                    AR2016.brief$congdist=="44" | AR2016.brief$congdist=="23" |
                    AR2016.brief$congdist=="45" | AR2016.brief$congdist=="46" ] <- 3
```

```
AR2016.brief$congdist[AR2016.brief$congdist=="47" | AR2016.brief$congdist=="48" |
                    AR2016.brief$congdist=="49" | AR2016.brief$congdist=="51" |
                    AR2016.brief$congdist=="52" | AR2016.brief$congdist=="53" |
                    AR2016.brief$congdist=="38" | AR2016.brief$congdist=="54" |
                    AR2016.brief$congdist=="56" | AR2016.brief$congdist=="39" |
                    AR2016.brief$congdist=="57" | AR2016.brief$congdist=="58" |
                    AR2016.brief$congdist=="59" | AR2016.brief$congdist=="60" |
                    AR2016.brief$congdist=="61" | AR2016.brief$congdist=="62" |
                    AR2016.brief$congdist=="40" | AR2016.brief$congdist=="63" |
                    AR2016.brief$congdist=="65" | AR2016.brief$congdist=="66" |
                    AR2016.brief$congdist=="41" | AR2016.brief$congdist=="67" |
                    AR2016.brief$congdist=="68" | AR2016.brief$congdist=="69" |
                    AR2016.brief$congdist=="43" | AR2016.brief$congdist=="70" |
                    AR2016.brief$congdist=="71" | AR2016.brief$congdist=="72" |
                    AR2016.brief$congdist=="73" | AR2016.brief$congdist=="45" |
                    AR2016.brief$congdist=="74" | AR2016.brief$congdist=="75" |
                    AR2016.brief$congdist=="34" ] <- 4

# Coding household income
AR2016.brief$income[AR2016.brief$income == 1]   <- 1 #15,0000 or less
AR2016.brief$income[AR2016.brief$income == 2]   <- 1 #15,0000 or less
AR2016.brief$income[AR2016.brief$income == 3]   <- 2
AR2016.brief$income[AR2016.brief$income == 4]   <- 3
AR2016.brief$income[AR2016.brief$income == 5]   <- 4
AR2016.brief$income[AR2016.brief$income == 6]   <- 5
AR2016.brief$income[AR2016.brief$income == 7]   <- 6
AR2016.brief$income[AR2016.brief$income == 8]   <- 7
AR2016.brief$income[AR2016.brief$income == 98]  <- 0
AR2016.brief$income[AR2016.brief$income == 99]  <- 0


# New version of AR2016 Data Frame
View(AR2016.brief)
### Note that this data frame contains all respondents. Next, we will remove those who are not Likely Voters.

# Create table of ONLY Likely Voters
AR2016.brief.LVandNA <- subset(AR2016.brief, LV == 1)
AR2016.brief.LV <- subset(AR2016.brief.LVandNA, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
view(AR2016.brief.LV)
### The data frame 'AR2020.brief.LV' contains only the likely voters who answered the presidential choice question.
### Each of the following tables that includes "LV" in the name only includes Likely Voters.
### If no "LV" in name, then the table includes all respondents.

# Print tables for all variables
table_pres_vote_2016 <- table(AR2016.brief$pres_vote)
table_pres_vote_2016
prop.table(table_pres_vote_2016)
table_pres_vote_2016_LV <- table(AR2016.brief.LV$pres_vote)
table_pres_vote_2016_LV
prop.table(table_pres_vote_2016_LV)
table_congdist_2016 <- table(AR2016.brief$congdist)
table_congdist_2016
prop.table(table_congdist_2016)
table_congdist_2016_LV <- table(AR2016.brief.LV$congdist)
table_congdist_2016_LV
```

```
prop.table(table_congdist_2016_LV)
table_gender_2016 <- table(AR2016.brief$gender)
table_gender_2016
prop.table(table_gender_2016)
table_gender_2016_LV <- table(AR2016.brief.LV$gender)
table_gender_2016_LV
prop.table(table_gender_2016_LV)
table_race_2016 <- table(AR2016.brief$race)
table_race_2016
prop.table(table_race_2016)
table_race_2016_LV <- table(AR2016.brief.LV$race)
table_race_2016_LV
prop.table(table_race_2016_LV)
table_age_2016 <- table(AR2016.brief$age)
table_age_2016
prop.table(table_age_2016)
table_age_2016_LV <- table(AR2016.brief.LV$age)
table_age_2016_LV
prop.table(table_age_2016_LV)
table_education_2016 <- table(AR2016.brief$education)
table_education_2016
prop.table(table_education_2016)
table_education_2016_LV <- table(AR2016.brief.LV$education)
table_education_2016_LV
prop.table(table_education_2016_LV)
table_income_2016 <- table(AR2016.brief$income)
table_income_2016
prop.table(table_income_2016)
table_income_2016_LV <- table(AR2016.brief.LV$income)
table_income_2016_LV
prop.table(table_income_2016_LV)



##### PART 3: Checking for Patterns in Missing Data #####

### Because we will utilize listwise deletion rather that imputation, we need to check that there are
### no patterns within the missing data, i.e. no demographics or variables correlate with item non-response.

# Create data frame with only the variables we need
vars2020 <- c("caseid", "pres_vote", "gender", "age", "race", "education", "income", "congdist")
AR2020.Full <- AR2020.brief[vars2020]
view(AR2020.Full)

# Create data frames containing respondents with missing info for corresponding variables
### There is no missing data for gender, so no new data frame is created to look for patterns
AR2020.Missing.age <- subset(AR2020.Full, caseid == 1388 | caseid == 2870 | caseid == 346 | caseid == 659 |
                             caseid == 750 | caseid == 838 | caseid == 2051 | caseid == 2302 | caseid == 2538 |
                             caseid == 2876 | caseid == 3983 | caseid == 4050 | caseid == 4080 | caseid == 4221 |
                             caseid == 4399 | caseid == 4506 | caseid == 4613 | caseid == 4718 | caseid == 4740 |
                             caseid == 4772 | caseid == 5059 | caseid == 5260 | caseid == 5356 | caseid == 5687)
AR2020.Missing.race <- subset(AR2020.Full, race != 1 & race != 2 & race != 3 & race != 4 & race != 5
                            & race != 6)
AR2020.Missing.education <- subset(AR2020.Full, education != 1 & education != 2
                                 & education != 3 & education != 4 & education != 5 & education != 6
                                 & education != 7)
```

```
AR2020.Missing.income <- subset(AR2020.Full, income != 1 & income != 2 & income != 3
                       & income != 4 & income != 5 & income != 6 & income != 7)
AR2020.Missing.congdist <- subset(AR2020.Full, congdist != 1 & congdist != 2
                        & congdist != 3 & congdist != 4)
### We find here that there are only 2 respondents missing for education, so that will not be tested either.
### Now, we test Age, Race, Income, and District to make sure we aren't seeing patterns.

# Age
### (respondents missing Age compared by the other 5 variables)
table_gender_2020 <- table(AR2020.Full$gender)
table_gender_2020
prop.table(table_gender_2020)
table_gender_2020_missing <- table(AR2020.Missing.age$gender)
table_gender_2020_missing
prop.table(table_gender_2020_missing)

table_congdist_2020 <- table(AR2020.Full$congdist)
table_congdist_2020
prop.table(table_congdist_2020)
table_congdist_2020_missing <- table(AR2020.Missing.age$congdist)
table_congdist_2020_missing
prop.table(table_congdist_2020_missing)

table_race_2020 <- table(AR2020.Full$race)
table_race_2020
prop.table(table_race_2020)
table_race_2020_missing <- table(AR2020.Missing.age$race)
table_race_2020_missing
prop.table(table_race_2020_missing)

table_education_2020 <- table(AR2020.Full$education)
table_education_2020
prop.table(table_education_2020)
table_education_2020_missing <- table(AR2020.Missing.age$education)
table_education_2020_missing
prop.table(table_education_2020_missing)

table_income_2020 <- table(AR2020.Full$income)
table_income_2020
prop.table(table_income_2020)
table_income_2020_missing <- table(AR2020.Missing.age$income)
table_income_2020_missing
prop.table(table_income_2020_missing)

# Race
### (respondents missing Race compared by the other 5 variables)
table_gender_2020 <- table(AR2020.Full$gender)
table_gender_2020
prop.table(table_gender_2020)
table_gender_2020_missing <- table(AR2020.Missing.race$gender)
table_gender_2020_missing
prop.table(table_gender_2020_missing)

table_congdist_2020 <- table(AR2020.Full$congdist)
table_congdist_2020
prop.table(table_congdist_2020)
```

```
table_congdist_2020_missing <- table(AR2020.Missing.race$congdist)
table_congdist_2020_missing
prop.table(table_congdist_2020_missing)

table_age_2020 <- table(AR2020.Full$age)
table_age_2020
prop.table(table_age_2020)
table_age_2020_missing <- table(AR2020.Missing.race$age)
table_age_2020_missing
prop.table(table_age_2020_missing)

table_education_2020 <- table(AR2020.Full$education)
table_education_2020
prop.table(table_education_2020)
table_education_2020_missing <- table(AR2020.Missing.race$education)
table_education_2020_missing
prop.table(table_education_2020_missing)

table_income_2020 <- table(AR2020.Full$income)
table_income_2020
prop.table(table_income_2020)
table_income_2020_missing <- table(AR2020.Missing.race$income)
table_income_2020_missing
prop.table(table_income_2020_missing)

# Income
### (respondents missing Income compared by the other 5 variables)
table_gender_2020 <- table(AR2020.Full$gender)
table_gender_2020
prop.table(table_gender_2020)
table_gender_2020_missing <- table(AR2020.Missing.income$gender)
table_gender_2020_missing
prop.table(table_gender_2020_missing)

table_congdist_2020 <- table(AR2020.Full$congdist)
table_congdist_2020
prop.table(table_congdist_2020)
table_congdist_2020_missing <- table(AR2020.Missing.income$congdist)
table_congdist_2020_missing
prop.table(table_congdist_2020_missing)

table_age_2020 <- table(AR2020.Full$age)
table_age_2020
prop.table(table_age_2020)
table_age_2020_missing <- table(AR2020.Missing.income$age)
table_age_2020_missing
prop.table(table_age_2020_missing)

table_education_2020 <- table(AR2020.Full$education)
table_education_2020
prop.table(table_education_2020)
table_education_2020_missing <- table(AR2020.Missing.income$education)
table_education_2020_missing
prop.table(table_education_2020_missing)

table_race_2020 <- table(AR2020.Full$race)
```

```
table_race_2020
prop.table(table_race_2020)
table_race_2020_missing <- table(AR2020.Missing.income$race)
table_race_2020_missing
prop.table(table_race_2020_missing)

# District
### (respondents missing District compared by the other 5 variables)
table_gender_2020 <- table(AR2020.Full$gender)
table_gender_2020
prop.table(table_gender_2020)
table_gender_2020_missing <- table(AR2020.Missing.congdist$gender)
table_gender_2020_missing
prop.table(table_gender_2020_missing)

table_income_2020 <- table(AR2020.Full$income)
table_income_2020
prop.table(table_income_2020)
table_income_2020_missing <- table(AR2020.Missing.congdist$income)
table_income_2020_missing
prop.table(table_income_2020_missing)

table_age_2020 <- table(AR2020.Full$age)
table_age_2020
prop.table(table_age_2020)
table_age_2020_missing <- table(AR2020.Missing.congdist$age)
table_age_2020_missing
prop.table(table_age_2020_missing)

table_education_2020 <- table(AR2020.Full$education)
table_education_2020
prop.table(table_education_2020)
table_education_2020_missing <- table(AR2020.Missing.congdist$education)
table_education_2020_missing
prop.table(table_education_2020_missing)

table_race_2020 <- table(AR2020.Full$race)
table_race_2020
prop.table(table_race_2020)
table_race_2020_missing <- table(AR2020.Missing.congdist$race)
table_race_2020_missing
prop.table(table_race_2020_missing)

### No significant patterns are found.



##### PART 4: Weighting the 2020 Presidential Choice #####

### At this point, the weighting will continue as follows:
### First, we weight the voting according to demographics. This includes LV only.
### The goal is to look for which demographics give either the most accurate results or biggest changes.
### Then, we apply those weights to a few specific opinion questions from the AR Poll (Section 7)

# Create table of only necessary variables
vars2020 <- c("caseid", "pres_vote", "gender", "age", "race", "education", "income", "congdist")
```

```
AR2020.Vars <- AR2020.brief.LV[vars2020]
view(AR2020.Vars)

# Now Create a separate table for each variable, combined with presidential choice, and remove missing data
gender2020 <- c("caseid", "pres_vote", "gender")
AR2020.gender <- AR2020.brief.LV[gender2020]
age2020 <- c("caseid", "pres_vote", "age")
AR2020.age.all <- AR2020.brief.LV[age2020]
AR2020.age <- na.omit(AR2020.age.all)
race2020 <- c("caseid", "pres_vote", "race")
AR2020.race.all <- AR2020.brief.LV[race2020]
AR2020.race <- subset(AR2020.race.all, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
                | race == 6)
education2020 <- c("caseid", "pres_vote", "education")
AR2020.education.all <- AR2020.brief.LV[education2020]
AR2020.education <- subset(AR2020.education.all, education == 1 | education == 2
                  | education == 3 | education == 4 | education == 5 | education == 6)
income2020 <- c("caseid", "pres_vote", "income")
AR2020.income.all <- AR2020.brief.LV[income2020]
AR2020.income <- subset(AR2020.income.all, income == 1 | income == 2 | income == 3
                 | income == 4 | income == 5 | income == 6 | income == 7)
congdist2020 <- c("caseid", "pres_vote", "congdist")
AR2020.congdist.all <- AR2020.brief.LV[congdist2020]
AR2020.congdist <- subset(AR2020.congdist.all, congdist == 1 | congdist == 2
                 | congdist == 3 | congdist == 4)

# Now combine these tables so that there is no missing data
AR2020.Vars.nomiss1 <- subset(AR2020.Vars, gender == 1 | gender == 2)
AR2020.Vars.nomiss2 <- subset(AR2020.Vars.nomiss1, age == 1 | age == 2 | age == 3 | age == 4 | age == 5
                     | age == 6 | age  == 7 | age == 8)
AR2020.Vars.nomiss3 <- subset(AR2020.Vars.nomiss2, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2020.Vars.nomiss4 <- subset(AR2020.Vars.nomiss3, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
                     | race == 6)
AR2020.Vars.nomiss5 <- subset(AR2020.Vars.nomiss4, education == 1 | education == 2
                     | education == 3 | education == 4 | education == 5 | education == 6
                     | education == 7)
AR2020.Vars.nomiss6 <- subset(AR2020.Vars.nomiss5, income == 1 | income == 2 | income == 3
                     | income == 4 | income == 5 | income == 6 | income == 7)
AR2020.Vars.nomiss <- subset(AR2020.Vars.nomiss6, congdist == 1 | congdist == 2
                    | congdist == 3 | congdist == 4)
AR2020.Vars.nomiss$pres_vote <- as.integer(AR2020.Vars.nomiss$pres_vote)
AR2020.Vars.nomiss$gender <- as.integer(AR2020.Vars.nomiss$gender)
AR2020.Vars.nomiss$age <- as.integer(AR2020.Vars.nomiss$age)
AR2020.Vars.nomiss$race <- as.integer(AR2020.Vars.nomiss$race)
AR2020.Vars.nomiss$education <- as.integer(AR2020.Vars.nomiss$education)
AR2020.Vars.nomiss$income <- as.integer(AR2020.Vars.nomiss$income)
AR2020.Vars.nomiss$congdist <- as.integer(AR2020.Vars.nomiss$congdist)

### Now, we will weight this individually, a few at a time, and all at once.

### We begin weighting "all at once," since this is the quickest way to see results.
### Keep in mind that later, we will remove the variable weights for those that are not discrepancies,
### like gender and congressional district, which are both fairly close in Census and sample.


### First, we need to input the Census data into our "target." It is listed here as comments,
```

```
### then inputted into code directly after:
#District:
#District 1 (1) -> .2414
#District 2 (2) -> .2543
#District 3 (3) -> .2679
#District 4 (4) -> .2364
#Gender:
#Female (1)    -> .5173
#Male   (2)    -> .4827
#Race:
#White        (1)  -> .7197
#Black        (2)  -> .1541
#Hispanic     (3)  -> .0769
#Asian        (4)  -> .0152
#Native Amer. (5)  -> .0054
#Multi-Ethnic (6)  -> .0236
#Educational Attainment:
#No HS        (1)  -> .0407
#Some HS      (2)  -> .0836
#HS Deg       (3)  -> .3483
#Some Col     (4)  -> .3114
#Col Deg      (5)  -> .1435
#Grad Deg     (6)  -> .0725
#Age:
#18-24  (1) -> .1218
#25-34  (2) -> .1667
#35-44  (3) -> .1627
#45-54  (4) -> .1563
#55-64  (5) -> .1665
#65-74  (6) -> .1309
#75-84  (7) -> .0700
#85+    (8) -> .0253
#Income:
#< 15       (1)  -> .1310
#15 - 25    (2)  -> .1210
#25 - 35    (3)  -> .1150
#35 - 50    (4)  -> .1410
#50 - 75    (5)  -> .1830
#75 - 100   (6)  -> .1150
#> 100      (7)  -> .1930

# Read 2019 Census Estimates
gender <- c(.5173,.4827)
names(gender) <- c(1, 2)
age <- c(.1218,.1667,.1627,.1563,.1665,.1309,.0700,.0253)
names(age) <- c(1, 2, 3, 4, 5, 6, 7, 8)
race <- c(.7197,.1541,.0769,.0152,.0054,.0236)
names(race) <- c(1, 2, 3, 4, 5, 6)
education <- c(.0407,.0836,.3483,.3114,.1435,.0725)
names(education) <- c(1, 2, 3, 4, 5, 6)
income <- c(.1310,.1210,.1150,.1410,.1830,.1150,.1930)
names(income) <- c(1, 2, 3, 4, 5, 6, 7)
congdist <- c(.2414,.2543,.2679,.2364)
names(congdist) <- c(1, 2, 3, 4)
target <- list(gender, age, race, education, income, congdist)
names(target) <- c("gender", "age", "race", "education", "income", "congdist")
```

```
# Weighting by All Variables
weight.2020.allvars <- anesrake(target, AR2020.Vars.nomiss, AR2020.Vars.nomiss$caseid, cap = 9999999999,
                    type = "pctlim", pctlim = 0.01, force1 = TRUE)
weight.2020.allvars
weight.summary.2020.allvars <- summary(weight.2020.allvars)
weight.summary.2020.allvars
AR2020.Vars.nomiss$weight <- weight.2020.allvars$weightvec
view(AR2020.Vars.nomiss)
wpct(AR2020.Vars.nomiss$pres_vote, AR2020.Vars.nomiss$weight)

# Weighting by individual variables

# Weighting by Gender individually
AR2020.gender$pres_vote <- as.integer(AR2020.gender$pres_vote)
AR2020.gender$gender <- as.integer(AR2020.gender$gender)
target.gender <- list(gender)
names(target.gender) <- c("gender")
weight.2020.gender <- anesrake(target.gender, AR2020.gender, AR2020.gender$caseid, cap = 9999999,
                   type = "pctlim", pctlim = 0.01, force1 = TRUE)
weight.2020.gender
wieght.summary.2020.gender <- summary(weight.2020.gender)
wieght.summary.2020.gender
AR2020.gender$weight <- weight.2020.gender$weightvec
view(AR2020.gender)
wpct(AR2020.gender$pres_vote, AR2020.gender$weight)

# Weighting by District Individually
AR2020.congdist$pres_vote <- as.integer(AR2020.congdist$pres_vote)
AR2020.congdist$congdist <- as.integer(AR2020.congdist$congdist)
target.congdist <- list(congdist)
names(target.congdist) <- c("congdist")
weight.2020.congdist <- anesrake(target.congdist, AR2020.congdist, AR2020.congdist$caseid, cap = 9999999,
                    type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.congdist
wieght.summary.2020.congdist <- summary(weight.2020.congdist)
wieght.summary.2020.congdist
AR2020.congdist$weight <- weight.2020.congdist$weightvec
view(AR2020.congdist)
wpct(AR2020.congdist$pres_vote, AR2020.congdist$weight)

# Weighting by Race Individually
AR2020.race$pres_vote <- as.integer(AR2020.race$pres_vote)
AR2020.race$race <- as.integer(AR2020.race$race)
target.race <- list(race)
names(target.race) <- c("race")
weight.2020.race <- anesrake(target.race, AR2020.race, AR2020.race$caseid, cap = 9999999,
                  type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.race
wieght.summary.2020.race <- summary(weight.2020.race)
wieght.summary.2020.race
AR2020.race$weight <- weight.2020.race$weightvec
view(AR2020.race)
wpct(AR2020.race$pres_vote, AR2020.race$weight)

# Weighting by Age Individually
```

```
AR2020.age$pres_vote <- as.integer(AR2020.age$pres_vote)
AR2020.age$age <- as.integer(AR2020.age$age)
target.age <- list(age)
names(target.age) <- c("age")
weight.2020.age <- anesrake(target.age, AR2020.age, AR2020.age$caseid, cap = 9999999,
                    type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.age
wieght.summary.2020.age <- summary(weight.2020.age)
wieght.summary.2020.age
AR2020.age$weight <- weight.2020.age$weightvec
view(AR2020.age)
wpct(AR2020.age$pres_vote, AR2020.age$weight)

# Weighting by Education Individually
AR2020.education$pres_vote <- as.integer(AR2020.education$pres_vote)
AR2020.education$education <- as.integer(AR2020.education$education)
target.education <- list(education)
names(target.education) <- c("education")
weight.2020.education <- anesrake(target.education, AR2020.education, AR2020.education$caseid, cap = 9999999,
                          type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.education
wieght.summary.2020.education <- summary(weight.2020.education)
wieght.summary.2020.education
AR2020.education$weight <- weight.2020.education$weightvec
view(AR2020.education)
wpct(AR2020.education$pres_vote, AR2020.education$weight)

# Weighting by Income Individually
AR2020.income$pres_vote <- as.integer(AR2020.income$pres_vote)
AR2020.income$income <- as.integer(AR2020.income$income)
target.income <- list(income)
names(target.income) <- c("income")
weight.2020.income <- anesrake(target.income, AR2020.income, AR2020.income$caseid, cap = 9999999,
                       type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.income
wieght.summary.2020.income <- summary(weight.2020.income)
wieght.summary.2020.income
AR2020.income$weight <- weight.2020.income$weightvec
view(AR2020.income)
wpct(AR2020.income$pres_vote, AR2020.income$weight)

# Weighting by combinations of variables

# Weighting by Race, Age, Education, and Income
AR2020.RAEI <- AR2020.Vars[c("pres_vote", "race", "age", "income", "education", "caseid")]
AR2020.RAEI <- subset(AR2020.RAEI, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2020.RAEI <- subset(AR2020.RAEI, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
               | race == 6)
AR2020.RAEI <- subset(AR2020.RAEI, age == 1 | age == 2 | age == 3 | age == 4 | age == 5
               | age == 6 | age  == 7 | age == 8)
AR2020.RAEI <- subset(AR2020.RAEI, education == 1 | education == 2
               | education == 3 | education == 4 | education == 5 | education == 6
               | education == 7)
AR2020.RAEI <- subset(AR2020.RAEI, income == 1 | income == 2 | income == 3
               | income == 4 | income == 5 | income == 6 | income == 7)
AR2020.RAEI$pres_vote <- as.integer(AR2020.RAEI$pres_vote)
```

```
AR2020.RAEI$race <- as.integer(AR2020.RAEI$race)
AR2020.RAEI$age <- as.integer(AR2020.RAEI$age)
AR2020.RAEI$education <- as.integer(AR2020.RAEI$education)
AR2020.RAEI$income <- as.integer(AR2020.RAEI$income)
AR2020.RAEI$caseid <- as.integer(AR2020.RAEI$caseid)
targets.RAEI <- list(race, age, income, education)
names(targets.RAEI) <- c("race", "age", "income", "education")
weight.2020.RAEI <- anesrake(targets.RAEI, AR2020.RAEI, AR2020.RAEI$caseid, cap = 9999999,
                             type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.RAEI
wieght.summary.2020.RAEI <- summary(weight.2020.RAEI)
wieght.summary.2020.RAEI
AR2020.RAEI$weight <- weight.2020.RAEI$weightvec
view(AR2020.RAEI)
wpct(AR2020.RAEI$pres_vote, AR2020.RAEI$weight)

#Weighting by Income and Age
AR2020.AI <- AR2020.Vars[c("pres_vote", "age", "income", "caseid")]
AR2020.AI <- subset(AR2020.AI, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2020.AI <- subset(AR2020.AI, age == 1 | age == 2 | age == 3 | age == 4 | age == 5
                    | age == 6 | age  == 7 | age == 8)
AR2020.AI <- subset(AR2020.AI, income == 1 | income == 2 | income == 3
                    | income == 4 | income == 5 | income == 6 | income == 7)
AR2020.AI$pres_vote <- as.integer(AR2020.AI$pres_vote)
AR2020.AI$age <- as.integer(AR2020.AI$age)
AR2020.AI$income <- as.integer(AR2020.AI$income)
AR2020.AI$caseid <- as.integer(AR2020.AI$caseid)
targets.AI <- list(age, income)
names(targets.AI) <- c("age", "income")
weight.2020.AI <- anesrake(targets.AI, AR2020.AI, AR2020.AI$caseid, cap = 9999999,
                           type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.AI
wieght.summary.2020.AI <- summary(weight.2020.AI)
wieght.summary.2020.AI
AR2020.AI$weight <- weight.2020.AI$weightvec
view(AR2020.AI)
wpct(AR2020.AI$pres_vote, AR2020.AI$weight)

#Weighting by Race and Education
AR2020.RE <- AR2020.Vars[c("pres_vote", "race", "education", "caseid")]
AR2020.RE <- subset(AR2020.RE, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2020.RE <- subset(AR2020.RE, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
                    | race == 6)
AR2020.RE <- subset(AR2020.RE, education == 1 | education == 2
                    | education == 3 | education == 4 | education == 5 | education == 6
                    | education == 7)
AR2020.RE$pres_vote <- as.integer(AR2020.RE$pres_vote)
AR2020.RE$race <- as.integer(AR2020.RE$race)
AR2020.RE$education <- as.integer(AR2020.RE$education)
AR2020.RE$caseid <- as.integer(AR2020.RE$caseid)
targets.RE <- list(race, education)
names(targets.RE) <- c("race", "education")
weight.2020.RE <- anesrake(targets.RE, AR2020.RE, AR2020.RE$caseid, cap = 9999999,
                           type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2020.RE
wieght.summary.2020.RE <- summary(weight.2020.RE)
```

```
wieght.summary.2020.RE
AR2020.RE$weight <- weight.2020.RE$weightvec
view(AR2020.RE)
wpct(AR2020.RE$pres_vote, AR2020.RE$weight)



##### PART 5: Weighting the 2016 Presidential Choice #####

### We repeat the exact same process as above, only this time for the 2016 AR Poll and Election.
### First, we weight the voting according to demographics. This includes LV only.
### The goal is to look for which demographics give either the most accurate results or biggest changes.
### Then, we apply those weights to a few specific opinion questions from the AR Poll (Section 8)

# Create table of only necessary variables
vars2016 <- c("caseid", "pres_vote", "gender", "age", "race", "education", "income", "congdist")
AR2016.Vars <- AR2016.brief.LV[vars2016]
view(AR2016.Vars)

# Now Create a separate table for each variable, combined with presdiential choice, and remove missing data
gender2016 <- c("caseid", "pres_vote", "gender")
AR2016.gender <- AR2016.brief.LV[gender2016]
age2016 <- c("caseid", "pres_vote", "age")
AR2016.age.all <- AR2016.brief.LV[age2016]
AR2016.age <- subset(AR2016.age.all, age == 1 | age == 2 | age == 3 | age == 4 | age == 5
              | age == 6 | age == 7 | age == 8)
race2016 <- c("caseid", "pres_vote", "race")
AR2016.race.all <- AR2016.brief.LV[race2016]
AR2016.race <- subset(AR2016.race.all, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
              | race == 6)
education2016 <- c("caseid", "pres_vote", "education")
AR2016.education.all <- AR2016.brief.LV[education2016]
AR2016.education <- subset(AR2016.education.all, education == 1 | education == 2
                 | education == 3 | education == 4 | education == 5 | education == 6
                 | education == 7)
income2016 <- c("caseid", "pres_vote", "income")
AR2016.income.all <- AR2016.brief.LV[income2016]
AR2016.income <- subset(AR2016.income.all, income == 1 | income == 2 | income == 3
               | income == 4 | income == 5 | income == 6 | income == 7)
congdist2016 <- c("caseid", "pres_vote", "congdist")
AR2016.congdist.all <- AR2016.brief.LV[congdist2016]
AR2016.congdist <- subset(AR2016.congdist.all, congdist == 1 | congdist == 2
                 | congdist == 3 | congdist == 4)

# Now combine these tables so that there is no missing data
AR2016.Vars.nomiss1 <- subset(AR2016.Vars, gender == 1 | gender == 2)
AR2016.Vars.nomiss2 <- subset(AR2016.Vars.nomiss1, age == 1 | age == 2 | age == 3 | age == 4 | age == 5
                  | age == 6 | age  == 7 | age == 8)
AR2016.Vars.nomiss3 <- subset(AR2016.Vars.nomiss2, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2016.Vars.nomiss4 <- subset(AR2016.Vars.nomiss3, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
                  | race == 6)
AR2016.Vars.nomiss5 <- subset(AR2016.Vars.nomiss4, education == 1 | education == 2
                  | education == 3 | education == 4 | education == 5 | education == 6
                  | education == 7)
AR2016.Vars.nomiss6 <- subset(AR2016.Vars.nomiss5, income == 1 | income == 2 | income == 3
                  | income == 4 | income == 5 | income == 6 | income == 7)
```

```
AR2016.Vars.nomiss <- subset(AR2016.Vars.nomiss6, congdist == 1 | congdist == 2
                             | congdist == 3 | congdist == 4)
AR2016.Vars.nomiss$pres_vote <- as.integer(AR2016.Vars.nomiss$pres_vote)
AR2016.Vars.nomiss$gender <- as.integer(AR2016.Vars.nomiss$gender)
AR2016.Vars.nomiss$age <- as.integer(AR2016.Vars.nomiss$age)
AR2016.Vars.nomiss$race <- as.integer(AR2016.Vars.nomiss$race)
AR2016.Vars.nomiss$education <- as.integer(AR2016.Vars.nomiss$education)
AR2016.Vars.nomiss$income <- as.integer(AR2016.Vars.nomiss$income)
AR2016.Vars.nomiss$congdist <- as.integer(AR2016.Vars.nomiss$congdist)
AR2016.Vars.nomiss$caseid <- as.integer(AR2016.Vars.nomiss$caseid)

# Now, we will weight this individually, a few at a time, and all at once.

### We begin weighting "all at once," since this is the quickest way to see results.
### Keep in mind that later, we will remove the variable weights for those that are not discrepancies,
### like gender and congressional district, which are both fairly close in Census and sample.

### First, we need to input the Census data into our "target." It is listed here as comments,
### then inputted into code directly after:
#District:
#District 1 (1) -> .2444
#District 2 (2) -> .2537
#District 3 (3) -> .2616
#District 4 (4) -> .2403
#Gender:
#Female (1)    -> .5085
#Male   (2)    -> .4915
#Race:
#White        (1)  -> .7279
#Black        (2)  -> .1547
#Hispanic     (3)  -> .0718
#Asian        (4)  -> .0138
#Native Amer. (5)  -> .0054
#Multi-Ethnic (6)  -> .0224
#Educational Attainment:
#No HS        (1)  -> .0466
#Some HS      (2)  -> .1046
#HS Deg       (3)  -> .3624
#Some Col     (4)  -> .2685
#Col Deg      (5)  -> .1419
#Grad Deg     (6)  -> .0760
#Age:
#18-24  (1) -> .1273
#25-34  (2) -> .1677
#35-44  (3) -> .1600
#45-54  (4) -> .1662
#55-64  (5) -> .1658
#65-74  (6) -> .1246
#75-84  (7) -> .0638
#85+    (8) -> .0245
#Income:
#< 15      (1)  -> .1540
#15 - 25   (2)  -> .1270
#25 - 35   (3)  -> .1190
#35 - 50   (4)  -> .1520
#50 - 75   (5)  -> .1830
```

```
#75 - 100   (6)  -> .1040
#> 100      (7)  -> .1610

# Read 2016 Census Estimates
gender16 <- c(.5085,.4915)
names(gender16) <- c(1, 2)
age16 <- c(.1273, .1677, .1600, .1662, .1658, .1246, .0638, .0245)
names(age16) <- c(1, 2, 3, 4, 5, 6, 7, 8)
race16 <- c(.7279, .1547, .0718, .0138, .0054, .0224)
names(race16) <- c(1, 2, 3, 4, 5, 6)
education16 <- c(.0466, .1046, .3624, .2685, .1419, .0760)
names(education16) <- c(1, 2, 3, 4, 5, 6)
income16 <- c(.1540, .1270, .1190, .1520, .1830, .1040, .1610)
names(income16) <- c(1, 2, 3, 4, 5, 6, 7)
congdist16 <- c(.2444, .2537, .2616, .2403)
names(congdist16) <- c(1, 2, 3, 4)
target16 <- list(gender16, age16, race16, education16, income16, congdist16)
names(target16) <- c("gender", "age", "race", "education", "income", "congdist")

# Weighting by All Variables
AR2016.Vars.nomiss <- as.data.frame(AR2016.Vars.nomiss)
weight.2016.allvars <- anesrake(target16, AR2016.Vars.nomiss, AR2016.Vars.nomiss$caseid, cap = 99999999,
                               type = "pctlim", pctlim = 0.01, force1 = TRUE)
weight.2016.allvars
weight.summary.2016.allvars <- summary(weight.2016.allvars)
weight.summary.2016.allvars
AR2016.Vars.nomiss$weight <- weight.2016.allvars$weightvec
view(AR2016.Vars.nomiss)
wpct(AR2016.Vars.nomiss$pres_vote, AR2016.Vars.nomiss$weight)

# Weighting by individual variables

# Weighting by Gender individually
AR2016.gender$pres_vote <- as.integer(AR2016.gender$pres_vote)
AR2016.gender$gender <- as.integer(AR2016.gender$gender)
target.gender16 <- list(gender16)
names(target.gender16) <- c("gender")
weight.2016.gender <- anesrake(target.gender16, AR2016.gender, AR2016.gender$caseid, cap = 9999999,
                              type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.gender
wieght.summary.2016.gender <- summary(weight.2016.gender)
wieght.summary.2016.gender
AR2016.gender$weight <- weight.2016.gender$weightvec
view(AR2016.gender)
wpct(AR2016.gender$pres_vote, AR2016.gender$weight)

# Weighting by District Individually
AR2016.congdist$pres_vote <- as.integer(AR2016.congdist$pres_vote)
AR2016.congdist$congdist <- as.integer(AR2016.congdist$congdist)
target.congdist16 <- list(congdist16)
names(target.congdist16) <- c("congdist")
weight.2016.congdist <- anesrake(target.congdist16, AR2016.congdist, AR2016.congdist$caseid, cap = 9999999,
                                type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.congdist
wieght.summary.2016.congdist <- summary(weight.2016.congdist)
wieght.summary.2016.congdist
```

```
AR2016.congdist$weight <- weight.2016.congdist$weightvec
view(AR2016.congdist)
wpct(AR2016.congdist$pres_vote, AR2016.congdist$weight)

# Weighting by Race Individually
AR2016.race$pres_vote <- as.integer(AR2016.race$pres_vote)
AR2016.race$race <- as.integer(AR2016.race$race)
target.race16 <- list(race16)
names(target.race16) <- c("race")
weight.2016.race <- anesrake(target.race16, AR2016.race, AR2016.race$caseid, cap = 9999999,
                    type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.race
wieght.summary.2016.race <- summary(weight.2016.race)
wieght.summary.2016.race
AR2016.race$weight <- weight.2016.race$weightvec
view(AR2016.race)
wpct(AR2016.race$pres_vote, AR2016.race$weight)

# Weighting by Age Individually
AR2016.age$pres_vote <- as.integer(AR2016.age$pres_vote)
AR2016.age$age <- as.integer(AR2016.age$age)
target.age16 <- list(age16)
names(target.age16) <- c("age")
weight.2016.age <- anesrake(target.age16, AR2016.age, AR2016.age$caseid, cap = 9999999,
                   type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.age
wieght.summary.2016.age <- summary(weight.2016.age)
wieght.summary.2016.age
AR2016.age$weight <- weight.2016.age$weightvec
view(AR2016.age)
wpct(AR2016.age$pres_vote, AR2016.age$weight)

#Weighting by Education Individually
AR2016.education$pres_vote <- as.integer(AR2016.education$pres_vote)
AR2016.education$education <- as.integer(AR2016.education$education)
target.education16 <- list(education16)
names(target.education16) <- c("education")
weight.2016.education <- anesrake(target.education16, AR2016.education, AR2016.education$caseid, cap =
9999999,
                         type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.education
wieght.summary.2016.education <- summary(weight.2016.education)
wieght.summary.2016.education
AR2016.education$weight <- weight.2016.education$weightvec
view(AR2016.education)
wpct(AR2016.education$pres_vote, AR2016.education$weight)

#Weighting by Income Individually
AR2016.income$pres_vote <- as.integer(AR2016.income$pres_vote)
AR2016.income$income <- as.integer(AR2016.income$income)
target.income16 <- list(income16)
names(target.income16) <- c("income")
weight.2016.income <- anesrake(target.income16, AR2016.income, AR2016.income$caseid, cap = 9999999,
                      type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.income
wieght.summary.2016.income <- summary(weight.2016.income)
```

```
wieght.summary.2016.income
AR2016.income$weight <- weight.2016.income$weightvec
view(AR2016.income)
wpct(AR2016.income$pres_vote, AR2016.income$weight)

#Weighting by combinations of variables

#Race, Age, Education, and Income
AR2016.RAEI <- AR2016.Vars[c("pres_vote", "race", "age", "income", "education", "caseid")]
AR2016.RAEI <- subset(AR2016.RAEI, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2016.RAEI <- subset(AR2016.RAEI, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
            | race == 6)
AR2016.RAEI <- subset(AR2016.RAEI, age == 1 | age == 2 | age == 3 | age == 4 | age == 5
            | age == 6 | age  == 7 | age == 8)
AR2016.RAEI <- subset(AR2016.RAEI, education == 1 | education == 2
            | education == 3 | education == 4 | education == 5 | education == 6
            | education == 7)
AR2016.RAEI <- subset(AR2016.RAEI, income == 1 | income == 2 | income == 3
            | income == 4 | income == 5 | income == 6 | income == 7)
AR2016.RAEI$pres_vote <- as.integer(AR2016.RAEI$pres_vote)
AR2016.RAEI$race <- as.integer(AR2016.RAEI$race)
AR2016.RAEI$age <- as.integer(AR2016.RAEI$age)
AR2016.RAEI$education <- as.integer(AR2016.RAEI$education)
AR2016.RAEI$income <- as.integer(AR2016.RAEI$income)
AR2016.RAEI$caseid <- as.integer(AR2016.RAEI$caseid)
targets.RAEI.16 <- list(race16, age16, income16, education16)
names(targets.RAEI.16) <- c("race", "age", "income", "education")
weight.2016.RAEI <- anesrake(targets.RAEI.16, AR2016.RAEI, AR2016.RAEI$caseid, cap = 9999999,
                type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.RAEI
wieght.summary.2016.RAEI <- summary(weight.2016.RAEI)
wieght.summary.2016.RAEI
AR2016.RAEI$weight <- weight.2016.RAEI$weightvec
view(AR2016.RAEI)
wpct(AR2016.RAEI$pres_vote, AR2016.RAEI$weight)

#Income and Age
AR2016.AI <- AR2016.Vars[c("pres_vote", "age", "income", "caseid")]
AR2016.AI <- subset(AR2016.AI, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2016.AI <- subset(AR2016.AI, age == 1 | age == 2 | age == 3 | age == 4 | age == 5
           | age == 6 | age  == 7 | age == 8)
AR2016.AI <- subset(AR2016.AI, income == 1 | income == 2 | income == 3
           | income == 4 | income == 5 | income == 6 | income == 7)
AR2016.AI$pres_vote <- as.integer(AR2016.AI$pres_vote)
AR2016.AI$age <- as.integer(AR2016.AI$age)
AR2016.AI$income <- as.integer(AR2016.AI$income)
AR2016.AI$caseid <- as.integer(AR2016.AI$caseid)
targets.AI.16 <- list(age16, income16)
names(targets.AI.16) <- c("age", "income")
weight.2016.AI <- anesrake(targets.AI.16, AR2016.AI, AR2016.AI$caseid, cap = 9999999,
               type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.AI
wieght.summary.2016.AI <- summary(weight.2016.AI)
wieght.summary.2016.AI
AR2016.AI$weight <- weight.2016.AI$weightvec
view(AR2016.AI)
```

```
wpct(AR2016.AI$pres_vote, AR2016.AI$weight)

#Race and Education
AR2016.RE <- AR2016.Vars[c("pres_vote", "race", "education", "caseid")]
AR2016.RE <- subset(AR2016.RE, pres_vote == 1 | pres_vote == 2 | pres_vote == 3)
AR2016.RE <- subset(AR2016.RE, race == 1 | race == 2 | race == 3 | race == 4 | race == 5
            | race == 6)
AR2016.RE <- subset(AR2016.RE, education == 1 | education == 2
            | education == 3 | education == 4 | education == 5 | education == 6
            | education == 7)
AR2016.RE$pres_vote <- as.integer(AR2016.RE$pres_vote)
AR2016.RE$race <- as.integer(AR2016.RE$race)
AR2016.RE$education <- as.integer(AR2016.RE$education)
AR2016.RE$caseid <- as.integer(AR2016.RE$caseid)
targets.RE.16 <- list(race16, education16)
names(targets.RE.16) <- c("race", "education")
weight.2016.RE <- anesrake(targets.RE.16, AR2016.RE, AR2016.RE$caseid, cap = 9999999,
                type = "pctlim", pctlim = 5, force1 = TRUE)
weight.2016.RE
wieght.summary.2016.RE <- summary(weight.2016.RE)
wieght.summary.2016.RE
AR2016.RE$weight <- weight.2016.RE$weightvec
view(AR2016.RE)
wpct(AR2016.RE$pres_vote, AR2016.RE$weight)



##### PART 6: Weighting the 2020 Issue Questions #####

### Now that we have found weighted results for the election using the identified Likely Voters,
### we weight the entire AR Poll Sample on four issue questions using the most significant
### variables and combinations found in the presidential weighting.
### For demographic targets, we use the same 2019 Census Estimates used for 2020 election weighting.

# Create data frames for the four main opinion questions and variables with DK/Ref removed

# Set variables
opinion.vars <- c("caseid", "age", "race", "education", "income")
AR2020.brief$caseid <- as.integer(AR2020.brief$caseid)
AR2020.brief$age <- as.integer(AR2020.brief$age)
AR2020.brief$race <- as.integer(AR2020.brief$race)
AR2020.brief$education <- as.integer(AR2020.brief$education)
AR2020.brief$income <- as.integer(AR2020.brief$income)

# Abortion Question Data Frame
Abortion2020.vars <- AR2020.brief[opinion.vars]
Abortion2020.vars$abortion <- AR2020$q13
Abortion2020.vars <- subset(Abortion2020.vars, abortion == "More difficult" | abortion == "Easier"
                 | abortion == "No change")

# Gun Control Question Data Frame
GunControl2020.vars <- AR2020.brief[opinion.vars]
GunControl2020.vars$guncontrol <- AR2020$q12
GunControl2020.vars <- subset(GunControl2020.vars, guncontrol == "Stricter" | guncontrol == "Less strict"
                 | guncontrol == "No change")
```

```
# Climate Change Question Data Frame
Climate2020.vars <- AR2020.brief[opinion.vars]
Climate2020.vars$climate <- AR2020$q93
Climate2020.vars <- subset(Climate2020.vars, climate == "Yes" | climate == "No"
                    | climate == "[DO NOT READ] Not a problem")

# Direction Question Data Frame
Direction2020.vars <- AR2020.brief[opinion.vars]
Direction2020.vars$direction <- AR2020$q31
Direction2020.vars <- subset(Direction2020.vars, direction == "Right" | direction == "Wrong")

### Now, we can begin weighting using these data frames. Note that in each trial, new data frames
### will be created from these larger sets for the individual trials.

# Weighting the Abortion Question (2020)

# Age alone (Abortion)
Abortion2020.age <- Abortion2020.vars[c("caseid", "age", "abortion")]
Abortion2020.age <- na.omit(Abortion2020.age)
target.Abortion2020.age <- list(age)
names(target.Abortion2020.age) <- c("age")
str(target.Abortion2020.age)
weight.Abortion2020.age <- anesrake(target.Abortion2020.age, Abortion2020.age, Abortion2020.age$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2020.age <- summary(weight.Abortion2020.age)
wieght.summary.Abortion2020.age
Abortion2020.age$weight <- weight.Abortion2020.age$weightvec
view(Abortion2020.age)
wpct(Abortion2020.age$abortion)
wpct(Abortion2020.age$abortion, Abortion2020.age$weight)

# Income alone (Abortion)
Abortion2020.income <- Abortion2020.vars[c("caseid", "income", "abortion")]
Abortion2020.income <- subset(Abortion2020.income, income != 0)
target.Abortion2020.income <- list(income)
names(target.Abortion2020.income) <- c("income")
str(target.Abortion2020.income)
weight.Abortion2020.income <- anesrake(target.Abortion2020.income, Abortion2020.income,
Abortion2020.income$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2020.income <- summary(weight.Abortion2020.income)
wieght.summary.Abortion2020.income
Abortion2020.income$weight <- weight.Abortion2020.income$weightvec
view(Abortion2020.income)
wpct(Abortion2020.income$abortion)
wpct(Abortion2020.income$abortion, Abortion2020.income$weight)

# Age and Income (Abortion)
Abortion2020.AI <- Abortion2020.vars[c("caseid", "income", "age", "abortion")]
Abortion2020.AI <- subset(Abortion2020.AI, income != 0)
Abortion2020.AI <- na.omit(Abortion2020.AI)
target.Abortion2020.AI <- list(age, income)
names(target.Abortion2020.AI) <- c("age", "income")
str(target.Abortion2020.AI)
weight.Abortion2020.AI <- anesrake(target.Abortion2020.AI, Abortion2020.AI, Abortion2020.AI$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
```

```
wieght.summary.Abortion2020.AI <- summary(weight.Abortion2020.AI)
wieght.summary.Abortion2020.AI
Abortion2020.AI$weight <- weight.Abortion2020.AI$weightvec
view(Abortion2020.AI)
wpct(Abortion2020.AI$abortion)
wpct(Abortion2020.AI$abortion, Abortion2020.AI$weight)

# Race alone (Abortion)
Abortion2020.race <- Abortion2020.vars[c("caseid", "race", "abortion")]
Abortion2020.race <- subset(Abortion2020.race, race != 0)
target.Abortion2020.race <- list(race)
names(target.Abortion2020.race) <- c("race")
str(target.Abortion2020.race)
weight.Abortion2020.race <- anesrake(target.Abortion2020.race, Abortion2020.race, Abortion2020.race$caseid,
                                     cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2020.race <- summary(weight.Abortion2020.race)
wieght.summary.Abortion2020.race
Abortion2020.race$weight <- weight.Abortion2020.race$weightvec
view(Abortion2020.race)
wpct(Abortion2020.race$abortion)
wpct(Abortion2020.race$abortion, Abortion2020.race$weight)

# Education alone (Abortion)
Abortion2020.education <- Abortion2020.vars[c("caseid", "education", "abortion")]
Abortion2020.education <- na.omit(Abortion2020.education)
target.Abortion2020.education <- list(education)
names(target.Abortion2020.education) <- c("education")
str(target.Abortion2020.education)
weight.Abortion2020.education <- anesrake(target.Abortion2020.education, Abortion2020.education,
                                          Abortion2020.education$caseid, cap = 9999999, type = "pctlim",
                                          pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2020.education <- summary(weight.Abortion2020.education)
wieght.summary.Abortion2020.education
Abortion2020.education$weight <- weight.Abortion2020.education$weightvec
view(Abortion2020.education)
wpct(Abortion2020.education$abortion)
wpct(Abortion2020.education$abortion, Abortion2020.education$weight)

# Race and Education (Abortion)
Abortion2020.RE <- Abortion2020.vars[c("caseid", "race", "education", "abortion")]
Abortion2020.RE <- subset(Abortion2020.RE, race != 0)
Abortion2020.RE <- na.omit(Abortion2020.RE)
target.Abortion2020.RE <- list(race, education)
names(target.Abortion2020.RE) <- c("race", "education")
str(target.Abortion2020.RE)
weight.Abortion2020.RE <- anesrake(target.Abortion2020.RE, Abortion2020.RE, Abortion2020.RE$caseid,
                                   cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2020.RE <- summary(weight.Abortion2020.RE)
wieght.summary.Abortion2020.RE
Abortion2020.RE$weight <- weight.Abortion2020.RE$weightvec
view(Abortion2020.RE)
wpct(Abortion2020.RE$abortion)
wpct(Abortion2020.RE$abortion, Abortion2020.RE$weight)

# All Four Variables (Abortion)
Abortion2020.RAEI <- Abortion2020.vars[c("caseid", "race", "age", "education", "income", "abortion")]
```

```
Abortion2020.RAEI <- subset(Abortion2020.RAEI, income != 0)
Abortion2020.RAEI <- subset(Abortion2020.RAEI, race != 0)
Abortion2020.RAEI <- na.omit(Abortion2020.RAEI)
target.Abortion2020.RAEI <- list(race, age, education, income)
names(target.Abortion2020.RAEI) <- c("race", "age", "education", "income")
str(target.Abortion2020.RAEI)
weight.Abortion2020.RAEI <- anesrake(target.Abortion2020.RAEI, Abortion2020.RAEI,
Abortion2020.RAEI$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2020.RAEI <- summary(weight.Abortion2020.RAEI)
wieght.summary.Abortion2020.RAEI
Abortion2020.RAEI$weight <- weight.Abortion2020.RAEI$weightvec
view(Abortion2020.RAEI)
wpct(Abortion2020.RAEI$abortion)
wpct(Abortion2020.RAEI$abortion, Abortion2020.RAEI$weight)

# Weighting the Gun Control Question (2020)

# Age alone (Gun Control)
GunControl2020.age <- GunControl2020.vars[c("caseid", "age", "guncontrol")]
GunControl2020.age <- na.omit(GunControl2020.age)
target.GunControl2020.age <- list(age)
names(target.GunControl2020.age) <- c("age")
str(target.GunControl2020.age)
weight.GunControl2020.age <- anesrake(target.GunControl2020.age, GunControl2020.age,
GunControl2020.age$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2020.age <- summary(weight.GunControl2020.age)
wieght.summary.GunControl2020.age
GunControl2020.age$weight <- weight.GunControl2020.age$weightvec
view(GunControl2020.age)
wpct(GunControl2020.age$guncontrol)
wpct(GunControl2020.age$guncontrol, GunControl2020.age$weight)

# Income alone (Gun Control)
GunControl2020.income <- GunControl2020.vars[c("caseid", "income", "guncontrol")]
GunControl2020.income <- subset(GunControl2020.income, income != 0)
target.GunControl2020.income <- list(income)
names(target.GunControl2020.income) <- c("income")
str(target.GunControl2020.income)
weight.GunControl2020.income <- anesrake(target.GunControl2020.income, GunControl2020.income,
GunControl2020.income$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2020.income <- summary(weight.GunControl2020.income)
wieght.summary.GunControl2020.income
GunControl2020.income$weight <- weight.GunControl2020.income$weightvec
view(GunControl2020.income)
wpct(GunControl2020.income$guncontrol)
wpct(GunControl2020.income$guncontrol, GunControl2020.income$weight)

# Age and Income (Gun Control)
GunControl2020.AI <- GunControl2020.vars[c("caseid", "income", "age", "guncontrol")]
GunControl2020.AI <- subset(GunControl2020.AI, income != 0)
GunControl2020.AI <- na.omit(GunControl2020.AI)
target.GunControl2020.AI <- list(age, income)
names(target.GunControl2020.AI) <- c("age", "income")
```

```
str(target.GunControl2020.AI)
weight.GunControl2020.AI <- anesrake(target.GunControl2020.AI, GunControl2020.AI,
GunControl2020.AI$caseid,
                                    cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2020.AI <- summary(weight.GunControl2020.AI)
wieght.summary.GunControl2020.AI
GunControl2020.AI$weight <- weight.GunControl2020.AI$weightvec
view(GunControl2020.AI)
wpct(GunControl2020.AI$guncontrol)
wpct(GunControl2020.AI$guncontrol, GunControl2020.AI$weight)

# Race alone (Gun Control)
GunControl2020.race <- GunControl2020.vars[c("caseid", "race", "guncontrol")]
GunControl2020.race <- subset(GunControl2020.race, race != 0)
target.GunControl2020.race <- list(race)
names(target.GunControl2020.race) <- c("race")
str(target.GunControl2020.race)
weight.GunControl2020.race <- anesrake(target.GunControl2020.race, GunControl2020.race,
GunControl2020.race$caseid,
                                      cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2020.race <- summary(weight.GunControl2020.race)
wieght.summary.GunControl2020.race
GunControl2020.race$weight <- weight.GunControl2020.race$weightvec
view(GunControl2020.race)
wpct(GunControl2020.race$guncontrol)
wpct(GunControl2020.race$guncontrol, GunControl2020.race$weight)

# Education alone (Gun Control)
GunControl2020.education <- GunControl2020.vars[c("caseid", "education", "guncontrol")]
GunControl2020.education <- na.omit(GunControl2020.education)
target.GunControl2020.education <- list(education)
names(target.GunControl2020.education) <- c("education")
str(target.GunControl2020.education)
weight.GunControl2020.education <- anesrake(target.GunControl2020.education, GunControl2020.education,
                                           GunControl2020.education$caseid, cap = 9999999, type = "pctlim",
                                           pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2020.education <- summary(weight.GunControl2020.education)
wieght.summary.GunControl2020.education
GunControl2020.education$weight <- weight.GunControl2020.education$weightvec
view(GunControl2020.education)
wpct(GunControl2020.education$guncontrol)
wpct(GunControl2020.education$guncontrol, GunControl2020.education$weight)

# Race and Education (Gun Control)
GunControl2020.RE <- GunControl2020.vars[c("caseid", "race", "education", "guncontrol")]
GunControl2020.RE <- subset(GunControl2020.RE, race != 0)
GunControl2020.RE <- na.omit(GunControl2020.RE)
target.GunControl2020.RE <- list(race, education)
names(target.GunControl2020.RE) <- c("race", "education")
str(target.GunControl2020.RE)
weight.GunControl2020.RE <- anesrake(target.GunControl2020.RE, GunControl2020.RE,
GunControl2020.RE$caseid,
                                    cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2020.RE <- summary(weight.GunControl2020.RE)
wieght.summary.GunControl2020.RE
GunControl2020.RE$weight <- weight.GunControl2020.RE$weightvec
```

```
view(GunControl2020.RE)
wpct(GunControl2020.RE$guncontrol)
wpct(GunControl2020.RE$guncontrol, GunControl2020.RE$weight)

# All Four Variables (Gun Control)
GunControl2020.RAEI <- GunControl2020.vars[c("caseid", "race", "age", "education", "income", "guncontrol")]
GunControl2020.RAEI <- subset(GunControl2020.RAEI, income != 0)
GunControl2020.RAEI <- subset(GunControl2020.RAEI, race != 0)
GunControl2020.RAEI <- na.omit(GunControl2020.RAEI)
target.GunControl2020.RAEI <- list(race, age, education, income)
names(target.GunControl2020.RAEI) <- c("race", "age", "education", "income")
str(target.GunControl2020.RAEI)
weight.GunControl2020.RAEI <- anesrake(target.GunControl2020.RAEI, GunControl2020.RAEI,
GunControl2020.RAEI$caseid,
                              cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2020.RAEI <- summary(weight.GunControl2020.RAEI)
wieght.summary.GunControl2020.RAEI
GunControl2020.RAEI$weight <- weight.GunControl2020.RAEI$weightvec
view(GunControl2020.RAEI)
wpct(GunControl2020.RAEI$guncontrol)
wpct(GunControl2020.RAEI$guncontrol, GunControl2020.RAEI$weight)

# Weighting the Climate Change Question (2020)

# Age alone (Climate Change)
Climate2020.age <- Climate2020.vars[c("caseid", "age", "climate")]
Climate2020.age <- na.omit(Climate2020.age)
target.Climate2020.age <- list(age)
names(target.Climate2020.age) <- c("age")
str(target.Climate2020.age)
weight.Climate2020.age <- anesrake(target.Climate2020.age, Climate2020.age, Climate2020.age$caseid,
                            cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2020.age <- summary(weight.Climate2020.age)
wieght.summary.Climate2020.age
Climate2020.age$weight <- weight.Climate2020.age$weightvec
view(Climate2020.age)
wpct(Climate2020.age$climate)
wpct(Climate2020.age$climate, Climate2020.age$weight)

# Income alone (Climate Change)
Climate2020.income <- Climate2020.vars[c("caseid", "income", "climate")]
Climate2020.income <- subset(Climate2020.income, income != 0)
target.Climate2020.income <- list(income)
names(target.Climate2020.income) <- c("income")
str(target.Climate2020.income)
weight.Climate2020.income <- anesrake(target.Climate2020.income, Climate2020.income,
Climate2020.income$caseid,
                             cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2020.income <- summary(weight.Climate2020.income)
wieght.summary.Climate2020.income
Climate2020.income$weight <- weight.Climate2020.income$weightvec
view(Climate2020.income)
wpct(Climate2020.income$climate)
wpct(Climate2020.income$climate, Climate2020.income$weight)

# Age and Income (Climate Change)
```

```
Climate2020.AI <- Climate2020.vars[c("caseid", "income", "age", "climate")]
Climate2020.AI <- subset(Climate2020.AI, income != 0)
Climate2020.AI <- na.omit(Climate2020.AI)
target.Climate2020.AI <- list(age, income)
names(target.Climate2020.AI) <- c("age", "income")
str(target.Climate2020.AI)
weight.Climate2020.AI <- anesrake(target.Climate2020.AI, Climate2020.AI, Climate2020.AI$caseid,
                cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2020.AI <- summary(weight.Climate2020.AI)
wieght.summary.Climate2020.AI
Climate2020.AI$weight <- weight.Climate2020.AI$weightvec
view(Climate2020.AI)
wpct(Climate2020.AI$climate)
wpct(Climate2020.AI$climate, Climate2020.AI$weight)

# Race alone (Climate Change)
Climate2020.race <- Climate2020.vars[c("caseid", "race", "climate")]
Climate2020.race <- subset(Climate2020.race, race != 0)
target.Climate2020.race <- list(race)
names(target.Climate2020.race) <- c("race")
str(target.Climate2020.race)
weight.Climate2020.race <- anesrake(target.Climate2020.race, Climate2020.race, Climate2020.race$caseid,
                 cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2020.race <- summary(weight.Climate2020.race)
wieght.summary.Climate2020.race
Climate2020.race$weight <- weight.Climate2020.race$weightvec
view(Climate2020.race)
wpct(Climate2020.race$climate)
wpct(Climate2020.race$climate, Climate2020.race$weight)

# Education alone (Climate Change)
Climate2020.education <- Climate2020.vars[c("caseid", "education", "climate")]
Climate2020.education <- na.omit(Climate2020.education)
target.Climate2020.education <- list(education)
names(target.Climate2020.education) <- c("education")
str(target.Climate2020.education)
weight.Climate2020.education <- anesrake(target.Climate2020.education, Climate2020.education,
                  Climate2020.education$caseid, cap = 9999999, type = "pctlim",
                  pctlim = 5, force1 = TRUE)
wieght.summary.Climate2020.education <- summary(weight.Climate2020.education)
wieght.summary.Climate2020.education
Climate2020.education$weight <- weight.Climate2020.education$weightvec
view(Climate2020.education)
wpct(Climate2020.education$climate)
wpct(Climate2020.education$climate, Climate2020.education$weight)

# Race and Education (Climate Change)
Climate2020.RE <- Climate2020.vars[c("caseid", "race", "education", "climate")]
Climate2020.RE <- subset(Climate2020.RE, race != 0)
Climate2020.RE <- na.omit(Climate2020.RE)
target.Climate2020.RE <- list(race, education)
names(target.Climate2020.RE) <- c("race", "education")
str(target.Climate2020.RE)
weight.Climate2020.RE <- anesrake(target.Climate2020.RE, Climate2020.RE, Climate2020.RE$caseid,
                cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2020.RE <- summary(weight.Climate2020.RE)
```

```
wieght.summary.Climate2020.RE
Climate2020.RE$weight <- weight.Climate2020.RE$weightvec
view(Climate2020.RE)
wpct(Climate2020.RE$climate)
wpct(Climate2020.RE$climate, Climate2020.RE$weight)

# All Four Variables (Climate Change)
Climate2020.RAEI <- Climate2020.vars[c("caseid", "race", "age", "education", "income", "climate")]
Climate2020.RAEI <- subset(Climate2020.RAEI, income != 0)
Climate2020.RAEI <- subset(Climate2020.RAEI, race != 0)
Climate2020.RAEI <- na.omit(Climate2020.RAEI)
target.Climate2020.RAEI <- list(race, age, education, income)
names(target.Climate2020.RAEI) <- c("race", "age", "education", "income")
str(target.Climate2020.RAEI)
weight.Climate2020.RAEI <- anesrake(target.Climate2020.RAEI, Climate2020.RAEI, Climate2020.RAEI$caseid,
                                    cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2020.RAEI <- summary(weight.Climate2020.RAEI)
wieght.summary.Climate2020.RAEI
Climate2020.RAEI$weight <- weight.Climate2020.RAEI$weightvec
view(Climate2020.RAEI)
wpct(Climate2020.RAEI$climate)
wpct(Climate2020.RAEI$climate, Climate2020.RAEI$weight)

# Weighting the "Right or Wrong Direction" Question (2020)

# Age alone (Right or Wrong Direction)
Direction2020.age <- Direction2020.vars[c("caseid", "age", "direction")]
Direction2020.age <- na.omit(Direction2020.age)
target.Direction2020.age <- list(age)
names(target.Direction2020.age) <- c("age")
str(target.Direction2020.age)
weight.Direction2020.age <- anesrake(target.Direction2020.age, Direction2020.age, Direction2020.age$caseid,
                                     cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2020.age <- summary(weight.Direction2020.age)
wieght.summary.Direction2020.age
Direction2020.age$weight <- weight.Direction2020.age$weightvec
view(Direction2020.age)
wpct(Direction2020.age$direction)
wpct(Direction2020.age$direction, Direction2020.age$weight)

# Income alone (Right or Wrong Direction)
Direction2020.income <- Direction2020.vars[c("caseid", "income", "direction")]
Direction2020.income <- subset(Direction2020.income, income != 0)
target.Direction2020.income <- list(income)
names(target.Direction2020.income) <- c("income")
str(target.Direction2020.income)
weight.Direction2020.income <- anesrake(target.Direction2020.income, Direction2020.income,
Direction2020.income$caseid,
                                        cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2020.income <- summary(weight.Direction2020.income)
wieght.summary.Direction2020.income
Direction2020.income$weight <- weight.Direction2020.income$weightvec
view(Direction2020.income)
wpct(Direction2020.income$direction)
wpct(Direction2020.income$direction, Direction2020.income$weight)
```

```
# Age and Income (Right or Wrong Direction)
Direction2020.AI <- Direction2020.vars[c("caseid", "income", "age", "direction")]
Direction2020.AI <- subset(Direction2020.AI, income != 0)
Direction2020.AI <- na.omit(Direction2020.AI)
target.Direction2020.AI <- list(age, income)
names(target.Direction2020.AI) <- c("age", "income")
str(target.Direction2020.AI)
weight.Direction2020.AI <- anesrake(target.Direction2020.AI, Direction2020.AI, Direction2020.AI$caseid,
                    cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2020.AI <- summary(weight.Direction2020.AI)
wieght.summary.Direction2020.AI
Direction2020.AI$weight <- weight.Direction2020.AI$weightvec
view(Direction2020.AI)
wpct(Direction2020.AI$direction)
wpct(Direction2020.AI$direction, Direction2020.AI$weight)

# Race alone (Right or Wrong Direction)
Direction2020.race <- Direction2020.vars[c("caseid", "race", "direction")]
Direction2020.race <- subset(Direction2020.race, race != 0)
target.Direction2020.race <- list(race)
names(target.Direction2020.race) <- c("race")
str(target.Direction2020.race)
weight.Direction2020.race <- anesrake(target.Direction2020.race, Direction2020.race, Direction2020.race$caseid,
                    cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2020.race <- summary(weight.Direction2020.race)
wieght.summary.Direction2020.race
Direction2020.race$weight <- weight.Direction2020.race$weightvec
view(Direction2020.race)
wpct(Direction2020.race$direction)
wpct(Direction2020.race$direction, Direction2020.race$weight)

# Education alone (Right or Wrong Direction)
Direction2020.education <- Direction2020.vars[c("caseid", "education", "direction")]
Direction2020.education <- na.omit(Direction2020.education)
target.Direction2020.education <- list(education)
names(target.Direction2020.education) <- c("education")
str(target.Direction2020.education)
weight.Direction2020.education <- anesrake(target.Direction2020.education, Direction2020.education,
                        Direction2020.education$caseid, cap = 9999999, type = "pctlim",
                        pctlim = 5, force1 = TRUE)
wieght.summary.Direction2020.education <- summary(weight.Direction2020.education)
wieght.summary.Direction2020.education
Direction2020.education$weight <- weight.Direction2020.education$weightvec
view(Direction2020.education)
wpct(Direction2020.education$direction)
wpct(Direction2020.education$direction, Direction2020.education$weight)

# Race and Education (Right or Wrong Direction)
Direction2020.RE <- Direction2020.vars[c("caseid", "race", "education", "direction")]
Direction2020.RE <- subset(Direction2020.RE, race != 0)
Direction2020.RE <- na.omit(Direction2020.RE)
target.Direction2020.RE <- list(race, education)
names(target.Direction2020.RE) <- c("race", "education")
str(target.Direction2020.RE)
weight.Direction2020.RE <- anesrake(target.Direction2020.RE, Direction2020.RE, Direction2020.RE$caseid,
                    cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
```

```
wieght.summary.Direction2020.RE <- summary(weight.Direction2020.RE)
wieght.summary.Direction2020.RE
Direction2020.RE$weight <- weight.Direction2020.RE$weightvec
view(Direction2020.RE)
wpct(Direction2020.RE$direction)
wpct(Direction2020.RE$direction, Direction2020.RE$weight)

# All Four Variables (Right or Wrong Direction)
Direction2020.RAEI <- Direction2020.vars[c("caseid", "race", "age", "education", "income", "direction")]
Direction2020.RAEI <- subset(Direction2020.RAEI, income != 0)
Direction2020.RAEI <- subset(Direction2020.RAEI, race != 0)
Direction2020.RAEI <- na.omit(Direction2020.RAEI)
target.Direction2020.RAEI <- list(race, age, education, income)
names(target.Direction2020.RAEI) <- c("race", "age", "education", "income")
str(target.Direction2020.RAEI)
weight.Direction2020.RAEI <- anesrake(target.Direction2020.RAEI, Direction2020.RAEI,
Direction2020.RAEI$caseid,
                                cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2020.RAEI <- summary(weight.Direction2020.RAEI)
wieght.summary.Direction2020.RAEI
Direction2020.RAEI$weight <- weight.Direction2020.RAEI$weightvec
view(Direction2020.RAEI)
wpct(Direction2020.RAEI$direction)
wpct(Direction2020.RAEI$direction, Direction2020.RAEI$weight)


##### PART 7: Weighting the 2016 Issue Questions #####

### We repeat the process from Part 8. The same four variables and issue questions will be used.
### For demographic targets, we use the same 2016 Census Estimates used for 2016 election weighting.

# Create data frames for the four main issue questions and variables with DK/Ref removed

# Set variables
opinion.vars16 <- c("caseid", "age", "race", "education", "income")
AR2016.brief$caseid <- as.integer(AR2016.brief$caseid)
AR2016.brief$age <- as.integer(AR2016.brief$age)
AR2016.brief$race <- as.integer(AR2016.brief$race)
AR2016.brief$education <- as.integer(AR2016.brief$education)
AR2016.brief$income <- as.integer(AR2016.brief$income)

# Abortion Question Data Frame
Abortion2016.vars <- AR2016.brief[opinion.vars16]
Abortion2016.vars$abortion <- AR2016$Q13
Abortion2016.vars$abortion <- as.integer(Abortion2016.vars$abortion)
Abortion2016.vars$abortion[Abortion2016.vars$abortion == 1]  <- "More difficult"
Abortion2016.vars$abortion[Abortion2016.vars$abortion == 2]  <- "Easier"
Abortion2016.vars$abortion[Abortion2016.vars$abortion == 3]  <- "No change"
Abortion2016.vars <- subset(Abortion2016.vars, abortion == "More difficult" | abortion == "Easier"
                        | abortion == "No change")

# Gun Control Question Data Frame
GunControl2016.vars <- AR2016.brief[opinion.vars16]
GunControl2016.vars$guncontrol <- AR2016$Q12
GunControl2016.vars$guncontrol <- as.integer(GunControl2016.vars$guncontrol)
GunControl2016.vars$guncontrol[GunControl2016.vars$guncontrol == 1]  <- "Stricter"
```

```
GunControl2016.vars$guncontrol[GunControl2016.vars$guncontrol == 2]  <- "Less strict"
GunControl2016.vars$guncontrol[GunControl2016.vars$guncontrol == 3]  <- "No change"
GunControl2016.vars <- subset(GunControl2016.vars, guncontrol == "Stricter" | guncontrol == "Less strict"
                    | guncontrol == "No change")

# Climate Change Question Data Frame
Climate2016.vars <- AR2016.brief[opinion.vars16]
Climate2016.vars$climate <- AR2016$Q93
Climate2016.vars$climate <- as.integer(Climate2016.vars$climate)
Climate2016.vars$climate[Climate2016.vars$climate == 1]  <- "Yes"
Climate2016.vars$climate[Climate2016.vars$climate == 2]  <- "No"
Climate2016.vars$climate[Climate2016.vars$climate == 3]  <- "Not a problem"
Climate2016.vars <- subset(Climate2016.vars, climate == "Yes" | climate == "No"
                   | climate == "Not a problem")

# Direction Question Data Frame
Direction2016.vars <- AR2016.brief[opinion.vars16]
Direction2016.vars$direction <- AR2016$Q31
Direction2016.vars$direction <- as.integer(Direction2016.vars$direction)
Direction2016.vars$direction[Direction2016.vars$direction == 1]  <- "Right"
Direction2016.vars$direction[Direction2016.vars$direction == 2]  <- "Wrong"
Direction2016.vars <- subset(Direction2016.vars, direction == "Right" | direction == "Wrong")

### Now, we can begin weighting using these data frames. Note that in each trial, new data frames
### will be created from these larger sets for the individual trials.

# Weighting the Abortion Question (2016)

# Age alone (Abortion)
Abortion2016.age <- Abortion2016.vars[c("caseid", "age", "abortion")]
Abortion2016.age <- subset(Abortion2016.age, age != 9999 & age != 9998)
target.Abortion2016.age <- list(age16)
names(target.Abortion2016.age) <- c("age")
str(target.Abortion2016.age)
weight.Abortion2016.age <- anesrake(target.Abortion2016.age, Abortion2016.age, Abortion2016.age$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2016.age <- summary(weight.Abortion2016.age)
wieght.summary.Abortion2016.age
Abortion2016.age$weight <- weight.Abortion2016.age$weightvec
view(Abortion2016.age)
wpct(Abortion2016.age$abortion)
wpct(Abortion2016.age$abortion, Abortion2016.age$weight)

# Income alone (Abortion)
Abortion2016.income <- Abortion2016.vars[c("caseid", "income", "abortion")]
Abortion2016.income <- subset(Abortion2016.income, income != 0)
target.Abortion2016.income <- list(income16)
names(target.Abortion2016.income) <- c("income")
str(target.Abortion2016.income)
weight.Abortion2016.income <- anesrake(target.Abortion2016.income, Abortion2016.income,
Abortion2016.income$caseid,
                           cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2016.income <- summary(weight.Abortion2016.income)
wieght.summary.Abortion2016.income
Abortion2016.income$weight <- weight.Abortion2016.income$weightvec
view(Abortion2016.income)
```

```
wpct(Abortion2016.income$abortion)
wpct(Abortion2016.income$abortion, Abortion2016.income$weight)

# Age and Income (Abortion)
Abortion2016.AI <- Abortion2016.vars[c("caseid", "income", "age", "abortion")]
Abortion2016.AI <- subset(Abortion2016.AI, income != 0)
Abortion2016.AI <- subset(Abortion2016.AI, age != 9999 & age != 9998)
target.Abortion2016.AI <- list(age16, income16)
names(target.Abortion2016.AI) <- c("age", "income")
str(target.Abortion2016.AI)
weight.Abortion2016.AI <- anesrake(target.Abortion2016.AI, Abortion2016.AI, Abortion2016.AI$caseid,
                          cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2016.AI <- summary(weight.Abortion2016.AI)
wieght.summary.Abortion2016.AI
Abortion2016.AI$weight <- weight.Abortion2016.AI$weightvec
view(Abortion2016.AI)
wpct(Abortion2016.AI$abortion)
wpct(Abortion2016.AI$abortion, Abortion2016.AI$weight)

# Race alone (Abortion)
Abortion2016.race <- Abortion2016.vars[c("caseid", "race", "abortion")]
Abortion2016.race <- subset(Abortion2016.race, race != 7 & race != 8 & race != 9)
target.Abortion2016.race <- list(race16)
names(target.Abortion2016.race) <- c("race")
str(target.Abortion2016.race)
weight.Abortion2016.race <- anesrake(target.Abortion2016.race, Abortion2016.race, Abortion2016.race$caseid,
                            cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2016.race <- summary(weight.Abortion2016.race)
wieght.summary.Abortion2016.race
Abortion2016.race$weight <- weight.Abortion2016.race$weightvec
view(Abortion2016.race)
wpct(Abortion2016.race$abortion)
wpct(Abortion2016.race$abortion, Abortion2016.race$weight)

# Education alone (Abortion)
Abortion2016.education <- Abortion2016.vars[c("caseid", "education", "abortion")]
Abortion2016.education <- subset(Abortion2016.education, education != 7 & education != 8 & education != 9)
target.Abortion2016.education <- list(education16)
names(target.Abortion2016.education) <- c("education")
str(target.Abortion2016.education)
weight.Abortion2016.education <- anesrake(target.Abortion2016.education, Abortion2016.education,
                                 Abortion2016.education$caseid, cap = 9999999, type = "pctlim",
                                 pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2016.education <- summary(weight.Abortion2016.education)
wieght.summary.Abortion2016.education
Abortion2016.education$weight <- weight.Abortion2016.education$weightvec
view(Abortion2016.education)
wpct(Abortion2016.education$abortion)
wpct(Abortion2016.education$abortion, Abortion2016.education$weight)

# Race and Education (Abortion)
Abortion2016.RE <- Abortion2016.vars[c("caseid", "race", "education", "abortion")]
Abortion2016.RE <- subset(Abortion2016.RE, race != 7 & race != 8 & race != 9)
Abortion2016.RE <- subset(Abortion2016.RE, education != 7 & education != 8 & education != 9)
target.Abortion2016.RE <- list(race16, education16)
names(target.Abortion2016.RE) <- c("race", "education")
```

```
str(target.Abortion2016.RE)
weight.Abortion2016.RE <- anesrake(target.Abortion2016.RE, Abortion2016.RE, Abortion2016.RE$caseid,
                              cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2016.RE <- summary(weight.Abortion2016.RE)
wieght.summary.Abortion2016.RE
Abortion2016.RE$weight <- weight.Abortion2016.RE$weightvec
view(Abortion2016.RE)
wpct(Abortion2016.RE$abortion)
wpct(Abortion2016.RE$abortion, Abortion2016.RE$weight)

# All Four Variables (Abortion)
Abortion2016.RAEI <- Abortion2016.vars[c("caseid", "race", "age", "education", "income", "abortion")]
Abortion2016.RAEI <- subset(Abortion2016.RAEI, income != 0)
Abortion2016.RAEI <- subset(Abortion2016.RAEI, age != 9999 & age != 9998)
Abortion2016.RAEI <- subset(Abortion2016.RAEI, race != 7 & race != 8 & race != 9)
Abortion2016.RAEI <- subset(Abortion2016.RAEI, education != 7 & education != 8 & education != 9)
target.Abortion2016.RAEI <- list(race16, age16, education16, income16)
names(target.Abortion2016.RAEI) <- c("race", "age", "education", "income")
str(target.Abortion2016.RAEI)
weight.Abortion2016.RAEI <- anesrake(target.Abortion2016.RAEI, Abortion2016.RAEI,
Abortion2016.RAEI$caseid,
                              cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Abortion2016.RAEI <- summary(weight.Abortion2016.RAEI)
wieght.summary.Abortion2016.RAEI
Abortion2016.RAEI$weight <- weight.Abortion2016.RAEI$weightvec
view(Abortion2016.RAEI)
wpct(Abortion2016.RAEI$abortion)
wpct(Abortion2016.RAEI$abortion, Abortion2016.RAEI$weight)

# Weighting the Gun Control Question (2016)

# Age alone (Gun Control)
GunControl2016.age <- GunControl2016.vars[c("caseid", "age", "guncontrol")]
GunControl2016.age <- subset(GunControl2016.age, age != 9999 & age != 9998)
target.GunControl2016.age <- list(age16)
names(target.GunControl2016.age) <- c("age")
str(target.GunControl2016.age)
weight.GunControl2016.age <- anesrake(target.GunControl2016.age, GunControl2016.age,
GunControl2016.age$caseid,
                              cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2016.age <- summary(weight.GunControl2016.age)
wieght.summary.GunControl2016.age
GunControl2016.age$weight <- weight.GunControl2016.age$weightvec
view(GunControl2016.age)
wpct(GunControl2016.age$guncontrol)
wpct(GunControl2016.age$guncontrol, GunControl2016.age$weight)

# Income alone (Gun Control)
GunControl2016.income <- GunControl2016.vars[c("caseid", "income", "guncontrol")]
GunControl2016.income <- subset(GunControl2016.income, income != 0)
target.GunControl2016.income <- list(income16)
names(target.GunControl2016.income) <- c("income")
str(target.GunControl2016.income)
weight.GunControl2016.income <- anesrake(target.GunControl2016.income, GunControl2016.income,
GunControl2016.income$caseid,
                              cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
```

```
wieght.summary.GunControl2016.income <- summary(weight.GunControl2016.income)
wieght.summary.GunControl2016.income
GunControl2016.income$weight <- weight.GunControl2016.income$weightvec
view(GunControl2016.income)
wpct(GunControl2016.income$guncontrol)
wpct(GunControl2016.income$guncontrol, GunControl2016.income$weight)

# Age and Income (Gun Control)
GunControl2016.AI <- GunControl2016.vars[c("caseid", "income", "age", "guncontrol")]
GunControl2016.AI <- subset(GunControl2016.AI, income != 0)
GunControl2016.AI <- subset(GunControl2016.AI, age != 9999 & age != 9998)
target.GunControl2016.AI <- list(age16, income16)
names(target.GunControl2016.AI) <- c("age", "income")
str(target.GunControl2016.AI)
weight.GunControl2016.AI <- anesrake(target.GunControl2016.AI, GunControl2016.AI,
GunControl2016.AI$caseid,
                                   cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2016.AI <- summary(weight.GunControl2016.AI)
wieght.summary.GunControl2016.AI
GunControl2016.AI$weight <- weight.GunControl2016.AI$weightvec
view(GunControl2016.AI)
wpct(GunControl2016.AI$guncontrol)
wpct(GunControl2016.AI$guncontrol, GunControl2016.AI$weight)

# Race alone (Gun Control)
GunControl2016.race <- GunControl2016.vars[c("caseid", "race", "guncontrol")]
GunControl2016.race <- subset(GunControl2016.race, race != 7 & race != 8 & race != 9)
target.GunControl2016.race <- list(race16)
names(target.GunControl2016.race) <- c("race")
str(target.GunControl2016.race)
weight.GunControl2016.race <- anesrake(target.GunControl2016.race, GunControl2016.race,
GunControl2016.race$caseid,
                                     cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2016.race <- summary(weight.GunControl2016.race)
wieght.summary.GunControl2016.race
GunControl2016.race$weight <- weight.GunControl2016.race$weightvec
view(GunControl2016.race)
wpct(GunControl2016.race$guncontrol)
wpct(GunControl2016.race$guncontrol, GunControl2016.race$weight)

# Education alone (Gun Control)
GunControl2016.education <- GunControl2016.vars[c("caseid", "education", "guncontrol")]
GunControl2016.education <- subset(GunControl2016.education, education != 7 & education != 8 & education !=
9)
target.GunControl2016.education <- list(education16)
names(target.GunControl2016.education) <- c("education")
str(target.GunControl2016.education)
weight.GunControl2016.education <- anesrake(target.GunControl2016.education, GunControl2016.education,
                                          GunControl2016.education$caseid, cap = 9999999, type = "pctlim",
                                          pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2016.education <- summary(weight.GunControl2016.education)
wieght.summary.GunControl2016.education
GunControl2016.education$weight <- weight.GunControl2016.education$weightvec
view(GunControl2016.education)
wpct(GunControl2016.education$guncontrol)
wpct(GunControl2016.education$guncontrol, GunControl2016.education$weight)
```

```
# Race and Education (Gun Control)
GunControl2016.RE <- GunControl2016.vars[c("caseid", "race", "education", "guncontrol")]
GunControl2016.RE <- subset(GunControl2016.RE, race != 7 & race != 8 & race != 9)
GunControl2016.RE <- subset(GunControl2016.RE, education != 7 & education != 8 & education != 9)
target.GunControl2016.RE <- list(race16, education16)
names(target.GunControl2016.RE) <- c("race", "education")
str(target.GunControl2016.RE)
weight.GunControl2016.RE <- anesrake(target.GunControl2016.RE, GunControl2016.RE,
GunControl2016.RE$caseid,
                                   cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2016.RE <- summary(weight.GunControl2016.RE)
wieght.summary.GunControl2016.RE
GunControl2016.RE$weight <- weight.GunControl2016.RE$weightvec
view(GunControl2016.RE)
wpct(GunControl2016.RE$guncontrol)
wpct(GunControl2016.RE$guncontrol, GunControl2016.RE$weight)

# All Four Variables (Gun Control)
GunControl2016.RAEI <- GunControl2016.vars[c("caseid", "race", "age", "education", "income", "guncontrol")]
GunControl2016.RAEI <- subset(GunControl2016.RAEI, income != 0)
GunControl2016.RAEI <- subset(GunControl2016.RAEI, age != 9999 & age != 9998)
GunControl2016.RAEI <- subset(GunControl2016.RAEI, race != 7 & race != 8 & race != 9)
GunControl2016.RAEI <- subset(GunControl2016.RAEI, education != 7 & education != 8 & education != 9)
target.GunControl2016.RAEI <- list(race16, age16, education16, income16)
names(target.GunControl2016.RAEI) <- c("race", "age", "education", "income")
str(target.GunControl2016.RAEI)
weight.GunControl2016.RAEI <- anesrake(target.GunControl2016.RAEI, GunControl2016.RAEI,
GunControl2016.RAEI$caseid,
                                   cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.GunControl2016.RAEI <- summary(weight.GunControl2016.RAEI)
wieght.summary.GunControl2016.RAEI
GunControl2016.RAEI$weight <- weight.GunControl2016.RAEI$weightvec
view(GunControl2016.RAEI)
wpct(GunControl2016.RAEI$guncontrol)
wpct(GunControl2016.RAEI$guncontrol, GunControl2016.RAEI$weight)

# Weighting the Climate Change Question (2016)

# Age alone (Climate Change)
Climate2016.age <- Climate2016.vars[c("caseid", "age", "climate")]
Climate2016.age <- subset(Climate2016.age, age != 9999 & age != 9998)
target.Climate2016.age <- list(age16)
names(target.Climate2016.age) <- c("age")
str(target.Climate2016.age)
weight.Climate2016.age <- anesrake(target.Climate2016.age, Climate2016.age, Climate2016.age$caseid,
                                 cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2016.age <- summary(weight.Climate2016.age)
wieght.summary.Climate2016.age
Climate2016.age$weight <- weight.Climate2016.age$weightvec
view(Climate2016.age)
wpct(Climate2016.age$climate)
wpct(Climate2016.age$climate, Climate2016.age$weight)

# Income alone (Climate Change)
Climate2016.income <- Climate2016.vars[c("caseid", "income", "climate")]
```

```
Climate2016.income <- subset(Climate2016.income, income != 0)
target.Climate2016.income <- list(income16)
names(target.Climate2016.income) <- c("income")
str(target.Climate2016.income)
weight.Climate2016.income <- anesrake(target.Climate2016.income, Climate2016.income,
Climate2016.income$caseid,
                    cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2016.income <- summary(weight.Climate2016.income)
wieght.summary.Climate2016.income
Climate2016.income$weight <- weight.Climate2016.income$weightvec
view(Climate2016.income)
wpct(Climate2016.income$climate)
wpct(Climate2016.income$climate, Climate2016.income$weight)

# Age and Income (Climate Change)
Climate2016.AI <- Climate2016.vars[c("caseid", "income", "age", "climate")]
Climate2016.AI <- subset(Climate2016.AI, income != 0)
Climate2016.AI <- subset(Climate2016.AI, age != 9999 & age != 9998)
target.Climate2016.AI <- list(age16, income16)
names(target.Climate2016.AI) <- c("age", "income")
str(target.Climate2016.AI)
weight.Climate2016.AI <- anesrake(target.Climate2016.AI, Climate2016.AI, Climate2016.AI$caseid,
                  cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2016.AI <- summary(weight.Climate2016.AI)
wieght.summary.Climate2016.AI
Climate2016.AI$weight <- weight.Climate2016.AI$weightvec
view(Climate2016.AI)
wpct(Climate2016.AI$climate)
wpct(Climate2016.AI$climate, Climate2016.AI$weight)

# Race alone (Climate Change)
Climate2016.race <- Climate2016.vars[c("caseid", "race", "climate")]
Climate2016.race <- subset(Climate2016.race, race != 7 & race != 8 & race != 9)
target.Climate2016.race <- list(race16)
names(target.Climate2016.race) <- c("race")
str(target.Climate2016.race)
weight.Climate2016.race <- anesrake(target.Climate2016.race, Climate2016.race, Climate2016.race$caseid,
                   cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2016.race <- summary(weight.Climate2016.race)
wieght.summary.Climate2016.race
Climate2016.race$weight <- weight.Climate2016.race$weightvec
view(Climate2016.race)
wpct(Climate2016.race$climate)
wpct(Climate2016.race$climate, Climate2016.race$weight)

# Education alone (Climate Change)
Climate2016.education <- Climate2016.vars[c("caseid", "education", "climate")]
Climate2016.education <- subset(Climate2016.education, education != 7 & education != 8 & education != 9)
target.Climate2016.education <- list(education16)
names(target.Climate2016.education) <- c("education")
str(target.Climate2016.education)
weight.Climate2016.education <- anesrake(target.Climate2016.education, Climate2016.education,
                     Climate2016.education$caseid, cap = 9999999, type = "pctlim",
                     pctlim = 5, force1 = TRUE)
wieght.summary.Climate2016.education <- summary(weight.Climate2016.education)
wieght.summary.Climate2016.education
```

```
Climate2016.education$weight <- weight.Climate2016.education$weightvec
view(Climate2016.education)
wpct(Climate2016.education$climate)
wpct(Climate2016.education$climate, Climate2016.education$weight)

# Race and Education (Climate Change)
Climate2016.RE <- Climate2016.vars[c("caseid", "race", "education", "climate")]
Climate2016.RE <- subset(Climate2016.RE, race != 7 & race != 8 & race != 9)
Climate2016.RE <- subset(Climate2016.RE, education != 7 & education != 8 & education != 9)
target.Climate2016.RE <- list(race16, education16)
names(target.Climate2016.RE) <- c("race", "education")
str(target.Climate2016.RE)
weight.Climate2016.RE <- anesrake(target.Climate2016.RE, Climate2016.RE, Climate2016.RE$caseid,
                         cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2016.RE <- summary(weight.Climate2016.RE)
wieght.summary.Climate2016.RE
Climate2016.RE$weight <- weight.Climate2016.RE$weightvec
view(Climate2016.RE)
wpct(Climate2016.RE$climate)
wpct(Climate2016.RE$climate, Climate2016.RE$weight)

# All Four Variables (Climate Change)
Climate2016.RAEI <- Climate2016.vars[c("caseid", "race", "age", "education", "income", "climate")]
Climate2016.RAEI <- subset(Climate2016.RAEI, income != 0)
Climate2016.RAEI <- subset(Climate2016.RAEI, age != 9999 & age != 9998)
Climate2016.RAEI <- subset(Climate2016.RAEI, race != 7 & race != 8 & race != 9)
Climate2016.RAEI <- subset(Climate2016.RAEI, education != 7 & education != 8 & education != 9)
target.Climate2016.RAEI <- list(race16, age16, education16, income16)
names(target.Climate2016.RAEI) <- c("race", "age", "education", "income")
str(target.Climate2016.RAEI)
weight.Climate2016.RAEI <- anesrake(target.Climate2016.RAEI, Climate2016.RAEI, Climate2016.RAEI$caseid,
                         cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Climate2016.RAEI <- summary(weight.Climate2016.RAEI)
wieght.summary.Climate2016.RAEI
Climate2016.RAEI$weight <- weight.Climate2016.RAEI$weightvec
view(Climate2016.RAEI)
wpct(Climate2016.RAEI$climate)
wpct(Climate2016.RAEI$climate, Climate2016.RAEI$weight)

# Weighting the "Right or Wrong Direction" Question (2016)

# Age alone (Right or Wrong Direction)
Direction2016.age <- Direction2016.vars[c("caseid", "age", "direction")]
Direction2016.age <- subset(Direction2016.age, age != 9999 & age != 9998)
target.Direction2016.age <- list(age16)
names(target.Direction2016.age) <- c("age")
str(target.Direction2016.age)
weight.Direction2016.age <- anesrake(target.Direction2016.age, Direction2016.age, Direction2016.age$caseid,
                         cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2016.age <- summary(weight.Direction2016.age)
wieght.summary.Direction2016.age
Direction2016.age$weight <- weight.Direction2016.age$weightvec
view(Direction2016.age)
wpct(Direction2016.age$direction)
wpct(Direction2016.age$direction, Direction2016.age$weight)
```

```
# Income alone (Right or Wrong Direction)
Direction2016.income <- Direction2016.vars[c("caseid", "income", "direction")]
Direction2016.income <- subset(Direction2016.income, income != 0)
target.Direction2016.income <- list(income16)
names(target.Direction2016.income) <- c("income")
str(target.Direction2016.income)
weight.Direction2016.income <- anesrake(target.Direction2016.income, Direction2016.income,
Direction2016.income$caseid,
                                cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2016.income <- summary(weight.Direction2016.income)
wieght.summary.Direction2016.income
Direction2016.income$weight <- weight.Direction2016.income$weightvec
view(Direction2016.income)
wpct(Direction2016.income$direction)
wpct(Direction2016.income$direction, Direction2016.income$weight)

# Age and Income (Right or Wrong Direction)
Direction2016.AI <- Direction2016.vars[c("caseid", "income", "age", "direction")]
Direction2016.AI <- subset(Direction2016.AI, income != 0)
Direction2016.AI <- subset(Direction2016.AI, age != 9999 & age != 9998)
target.Direction2016.AI <- list(age16, income16)
names(target.Direction2016.AI) <- c("age", "income")
str(target.Direction2016.AI)
weight.Direction2016.AI <- anesrake(target.Direction2016.AI, Direction2016.AI, Direction2016.AI$caseid,
                              cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2016.AI <- summary(weight.Direction2016.AI)
wieght.summary.Direction2016.AI
Direction2016.AI$weight <- weight.Direction2016.AI$weightvec
view(Direction2016.AI)
wpct(Direction2016.AI$direction)
wpct(Direction2016.AI$direction, Direction2016.AI$weight)

# Race alone (Right or Wrong Direction)
Direction2016.race <- Direction2016.vars[c("caseid", "race", "direction")]
Direction2016.race <- subset(Direction2016.race, race != 7 & race != 8 & race != 9)
target.Direction2016.race <- list(race16)
names(target.Direction2016.race) <- c("race")
str(target.Direction2016.race)
weight.Direction2016.race <- anesrake(target.Direction2016.race, Direction2016.race, Direction2016.race$caseid,
                               cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2016.race <- summary(weight.Direction2016.race)
wieght.summary.Direction2016.race
Direction2016.race$weight <- weight.Direction2016.race$weightvec
view(Direction2016.race)
wpct(Direction2016.race$direction)
wpct(Direction2016.race$direction, Direction2016.race$weight)

# Education alone (Right or Wrong Direction)
Direction2016.education <- Direction2016.vars[c("caseid", "education", "direction")]
Direction2016.education <- subset(Direction2016.education, education != 7 & education != 8 & education != 9)
target.Direction2016.education <- list(education16)
names(target.Direction2016.education) <- c("education")
str(target.Direction2016.education)
weight.Direction2016.education <- anesrake(target.Direction2016.education, Direction2016.education,
                                  Direction2016.education$caseid, cap = 9999999, type = "pctlim",
                                  pctlim = 5, force1 = TRUE)
```

```
wieght.summary.Direction2016.education <- summary(weight.Direction2016.education)
wieght.summary.Direction2016.education
Direction2016.education$weight <- weight.Direction2016.education$weightvec
view(Direction2016.education)
wpct(Direction2016.education$direction)
wpct(Direction2016.education$direction, Direction2016.education$weight)

# Race and Education (Right or Wrong Direction)
Direction2016.RE <- Direction2016.vars[c("caseid", "race", "education", "direction")]
Direction2016.RE <- subset(Direction2016.RE, race != 7 & race != 8 & race != 9)
Direction2016.RE <- subset(Direction2016.RE, education != 7 & education != 8 & education != 9)
target.Direction2016.RE <- list(race16, education16)
names(target.Direction2016.RE) <- c("race", "education")
str(target.Direction2016.RE)
weight.Direction2016.RE <- anesrake(target.Direction2016.RE, Direction2016.RE, Direction2016.RE$caseid,
                            cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2016.RE <- summary(weight.Direction2016.RE)
wieght.summary.Direction2016.RE
Direction2016.RE$weight <- weight.Direction2016.RE$weightvec
view(Direction2016.RE)
wpct(Direction2016.RE$direction)
wpct(Direction2016.RE$direction, Direction2016.RE$weight)

# All Four Variables (Right or Wrong Direction)
Direction2016.RAEI <- Direction2016.vars[c("caseid", "race", "age", "education", "income", "direction")]
Direction2016.RAEI <- subset(Direction2016.RAEI, income != 0)
Direction2016.RAEI <- subset(Direction2016.RAEI, age != 9999 & age != 9998)
Direction2016.RAEI <- subset(Direction2016.RAEI, race != 7 & race != 8 & race != 9)
Direction2016.RAEI <- subset(Direction2016.RAEI, education != 7 & education != 8 & education != 9)
target.Direction2016.RAEI <- list(race16, age16, education16, income16)
names(target.Direction2016.RAEI) <- c("race", "age", "education", "income")
str(target.Direction2016.RAEI)
weight.Direction2016.RAEI <- anesrake(target.Direction2016.RAEI, Direction2016.RAEI,
Direction2016.RAEI$caseid,
                             cap = 9999999, type = "pctlim", pctlim = 5, force1 = TRUE)
wieght.summary.Direction2016.RAEI <- summary(weight.Direction2016.RAEI)
wieght.summary.Direction2016.RAEI
Direction2016.RAEI$weight <- weight.Direction2016.RAEI$weightvec
view(Direction2016.RAEI)
wpct(Direction2016.RAEI$direction)
wpct(Direction2016.RAEI$direction, Direction2016.RAEI$weight)
```