Lab 2

July 16, 2019

1 Lab 2: Comparing Means | w203 Statistics for Data Science

1.1 Prepared by Sophia Ayele and Navya Sandadi

2 Introduction

We use the 2018 ANES Pilot Survey dataset to address five research questions. For each question, we operationalize the concepts in the questions, conduct exploratory analysis, select and conduct an appropriate hypothesis test and interpret the results in terms of statistical and practical significane.

The research questions generalize to US voters, but since this study does not represent a true random sample of the population of US voters, our conclusions only apply to the 2018 ANES Pilot Survey respondents.

2.1 Question 1: Do US voters have more respect for the police or for journalists?

2.1.1 Operationalization

> To analyze respect for police and journalists, we consider variables representing ratings given by voters to police and journalists based on a feeling thermometer. This is because there isn't a direct question asked on the survey about respect. However, we are not implying a direct correlation between ratings and respect. It is possible that even though people rate police highly, they may not highly respect them. Considering the set of questions asked on the survey, we agreed that it is reasonable to assume that ratings for police and journalists can be used for measuring respect. As the same set of people rated police and journalists, we regard the 2 variables as dependent.

Relevant questions asked on the survey: - [ftpolice] How would you rate the police? - [ftjournal] How would you rate journalists?

2.1.2 Exploratory Data Analysis

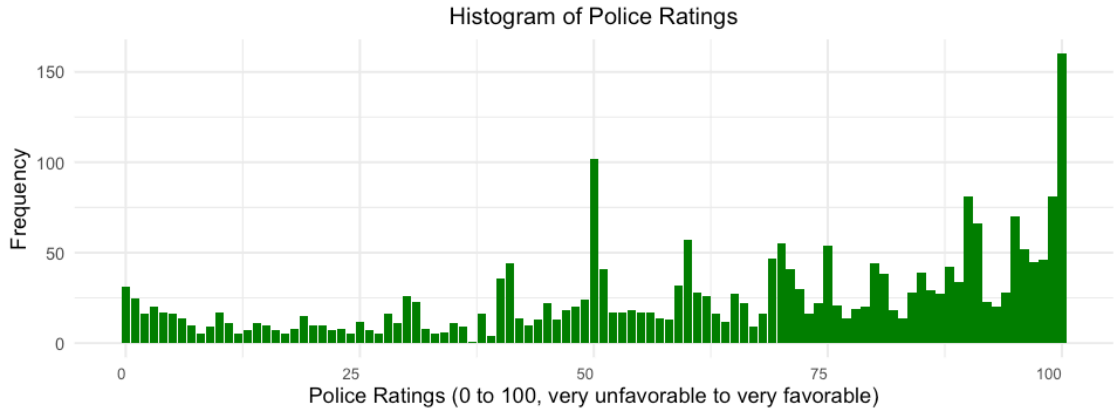
Data Transformation & Justification for Removed Values > On examining the variables, [ftpolice] and [ftjournal], we found that there are some values in [ftjournal] that are less than 0, specifically 2 observation of -7, which correspond to "no answer". Thus, we subset the data to remove the 2 observations.

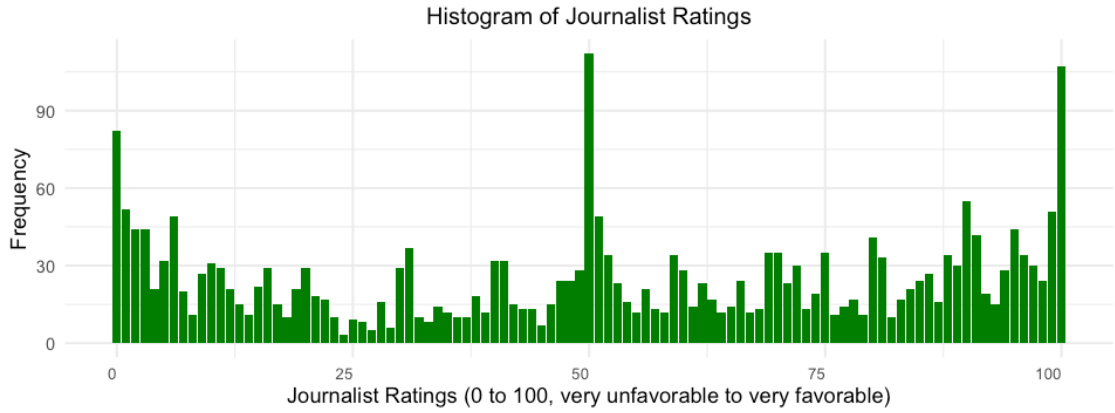
Sanity Checks > We looked at the summary of the variables (ftpolice and ftjournal) to check for abnormal minimum and maximum values. There are no NA values for the variables. Both

variables have 2500 observations that match the row count of the overall dataset. Although 99 can sometimes represent missing values in some data sets, we assume that it is a valid value here because of the range of the feeling thermometer (0 to 100), and because the inapplicable values, errors and missing values (or "no answer") are specifically denoted by -1, -4 and -7 respectively.

Relationship Among Variables > We compare the means of the 2 samples. The average rating of police is 64.6 and the average rating of journalists is 52.3, with the difference in means being 12.3. There is more variance in [ftjournal] observations than in [ftpolice] observations. The histograms (seen below) show that the distributions are not normal. There is an especially high skew towards the very beginning and the very end of the distribution. Both the distributions are very similar except for rather high frequency of low ratings for journalists.

In [44]:





2.1.3 Hypothesis Test Selection

We selected a Paired two-tailed t-test (parametric test) for the following reasons:

- [ftpolice] and [ftjournal] are interval variables (based on a feeling thermometer).

- We assume dependence due to the fact that the ratings for police and journalists were given by the same set of people.
- There is no specific direction in the research question, so we choose a 2-tailed test.
- Parametric tests can perform well when the spread of each group is different.
- Even though we have non-normal distribution, because of large sample size (2498 observations), we can apply Central Limit Theorem and thus, parametric tests are generally valid for large samples.
- Parametric tests usually have more statistical power than nonparametric tests. Thus, we are more likely to detect a significant effect when one truly exists.

2.1.4 Hypothesis Test

In [8]:

```
Paired t-test

data:  Data_1$ftpolice and Data_1$ftjournal
t = 13.711, df = 2497, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 10.58776 14.12160
sample estimates:
mean of the differences
               12.35468
```

2.1.5 Conclusion

Statistical Significance: > We can reject the null hypothesis that the difference between the means (of ratings) is equal to zero. The paired 2-tailed t-test has a p-value far less than 0.05, thus, making the result statistically significant.

Practical Significance: > The result is practically significant as well since there is a significant difference of 12.3 between the mean in police ratings and mean in journalist ratings, with police ratings having the higher mean value.

We can conclude that the voters who responded to the 2018 ANES survey have more respect for police than for journalists. We are considering ratings given by voters to imply respect. This result, however, cannot be generalized for the entire population of US voters as the survey respondents were not randomly selected.

2.2 Question 2: Are Republican voters older or younger than Democratic voters?

2.2.1 Operationalization

The concepts asked about on the survey align fairly well with the research question. Respondents were randomly assigned a group and asked one of two versions of the political affiliation question (see below). The survey also collected respondents' year of birth (see below). Year of birth can be used to calculate an approximate age since we do not have the day and month of birth. Lastly, the

research questions uses the word "voters", which we operationalize to mean survey respondents, regardless of voting behavior. A more appropriate research question might be: "Are 2018 ANES respondents that identify as Republicans older or younger than those that identify as Democrats"? Relevant questions asked on survey: * [pid1d] Generally speaking, do you usually think of yourself as a Democrat, a Republican, an independent, or what? * [pid1r] Generally speaking, do you usually think of yourself as a Republican, a Democrat, an independent, or what? * [birthyr] In what year were you born?

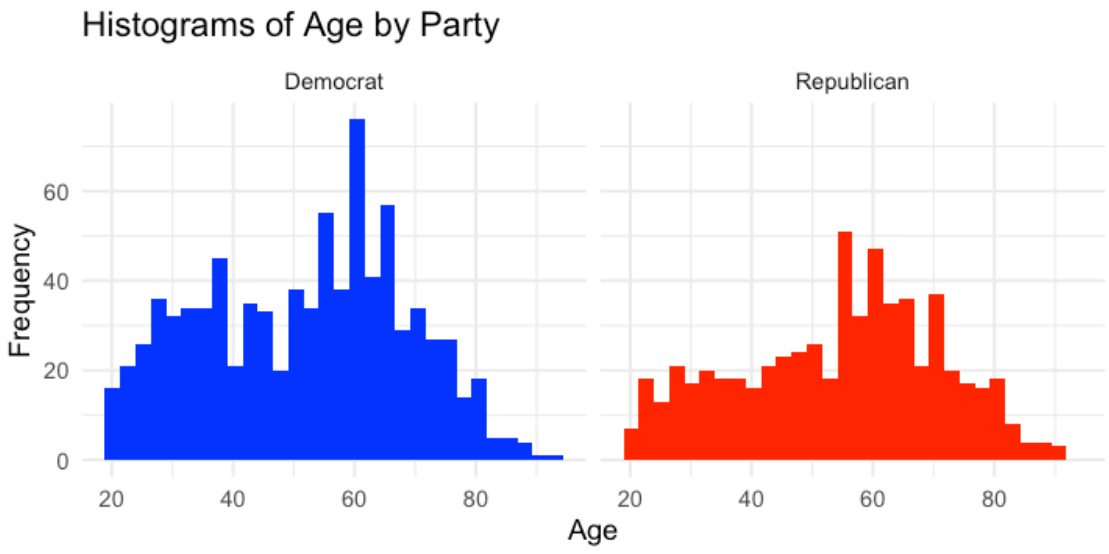
2.2.2 Exploratory Data Analysis

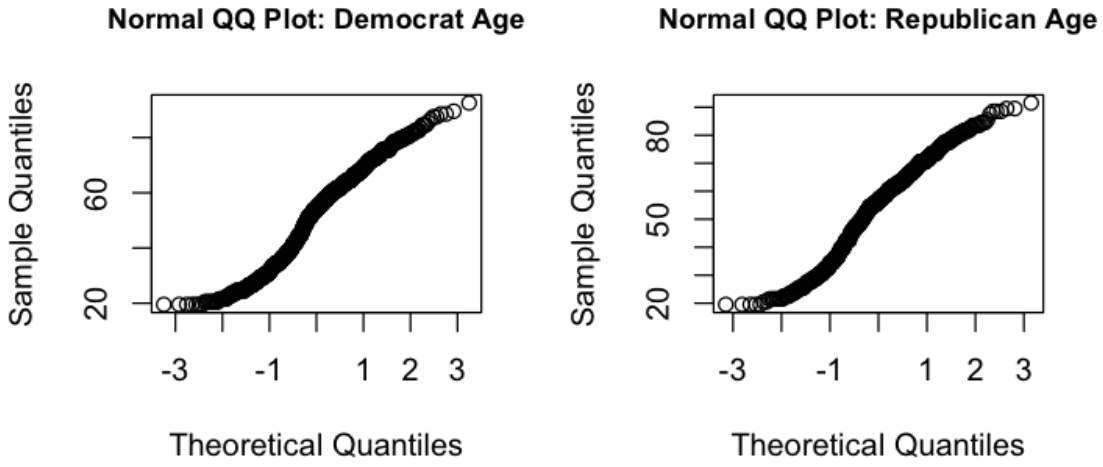
Data Transformations & Justification for Removed Values Since respondents were randomly assigned to two questions about party affiliation, we began by creating a single variable with all of the party information. Next, we subset the data to only include respondents that identify as Republican or Democrat. Since the data only contains year of birth, we calculated approximte date of birth using the first day of the first month of each year.

Sanity Checks After subsetting the data, we checked for missing birth year. We found no missing birth year data. The sample sizes of the groups are fairly large, but unbalanced: 857 Democrats vs. 609 Republicans. Both distributions of age by party are truncated on the left and have a right skew (see histograms below). To further examine their normality, we plotted normal qq plots for the two distributions below. These plots show that both distributions are approximately normal in the middle, but less normal at the tails. All age observations were within a reasonable range. The minimum age for both groups was 19 and the maximum age was approximately 90.

Relationship Among Variables On average, Republican respondents are 2.6 years older than Democrat respondents (54.39 vs. 51.78). While the histograms below show a higher proportion of Democrats between the ages of 25 and 50, the distributions are otherwise quite similar. This is reflected in the similar standard deviation for both groups. The standard deviation of the Republican age distribution is 16.83 vs. 16.84 for the Democrat age distribution.

In [45]:





2.2.3 Hypothesis Test Selection

We selected a Welch two-sided t-test because of the following assumptions:

- This test assumes normality of the sample means (even though our distributions are not normal, the CLT tell us that the sample means will have an approximately normal distribution)
- This test is for parametric data (approximate age is a continuous variable)
- This test can be used to compare the mean of two groups
- A two-sided test is our best option because our research question does not hypothesize about directionality

The two-sided t-test also assumes random sampling. Although our data was not generated through a random sample, we decided that this test was still the most appropriate although results need to be interpreted with caution.

2.2.4 Hypothesis Test

In [10]:

```
Welch Two Sample t-test

data:  age by party
t = -2.939, df = 1309.7, p-value = 0.00335
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
```

```
 -4.3723921 -0.8718651
sample estimates:
mean in group 1 mean in group 2
       51.77036        54.39249
```

2.2.5 Conclusion

Statistical Significance We reject the null hypothesis that the difference in means between the two groups is equal to 0. The Welch two sample t-test has a p-value of 0.00335, considerably lower than the commonly used 0.05 threshold. Under the null hypothesis there is a 0.335% chance of getting this t-statistic or something more extreme.

Practical Significance Although this result is statistically significant, the 2.6 year difference in ages is small enough that it is unlikely to impact policy making. Therefore we conclude that this result has little practical significance. As mentioned previously, this result is only relevant to the 2018 ANES Pilot Study population and cannot be generalized to the larger US population.

2.3 Question 3: Do a majority of independent voters believe that the federal investigations of Russian election interference are baseless?

2.3.1 Operationalization

> To analyze the belief of independent voters in terms of baselessness of the federal investigations, we looked at the variable [muellerinv]. This is the only variable that most directly measures the "baselessness" sentiment, as it asks voters if they approve or disapprove (or neither) of Robert Mueller's investigation. We dichotomized this variable as either disapproved or approved/neutral by dividing the respondents as voters who disapproved of the investigation or not, (i.e), combined responses of voters who either approved or were neutral towards the investigation into a single group. We used the variables [pid1d] and [pid1r] to filter out voters who are independent. These variables were used by survey respondents to identify themselves as independent.
>
> Other variables that are related to this question are [russia16] and [coord16] which indicate whether voters believe if the interference actually happened and if Trump campaign coordinated with the Russians. Even though voters believe that the interference actually happened, they may think that the investigation itself is baseless. There is another variable [ftmueller] that asks voters to rate Special Counsel Robert Mueller. This again has little to do with the investigation itself. Thus, we do not consider the related variables for answering the research question.

The variables considered are: - [pid1d] - Generally speaking, do you usually think of yourself as a Democrat, a Republican, an independent, or what? - [pid1r] - Generally speaking, do you usually think of yourself as a Republican, a Democrat, an independent, or what? - [muellerinv] - Do you approve, disapprove, or neither approve nor disapprove of Robert Mueller's investigation of Russian interference in the 2016 election? - [russia16] Do you think the Russian government probably interfered in the 2016 presidential election to try to help Donald Trump win, or do you think this probably did not happen? - [coord16] Do you think Donald Trump's 2016 campaign

probably coordinated with the Russians, or do you think his campaign probably did not do this? - [ftmueller] How would you rate Special Counsel Robert Mueller?

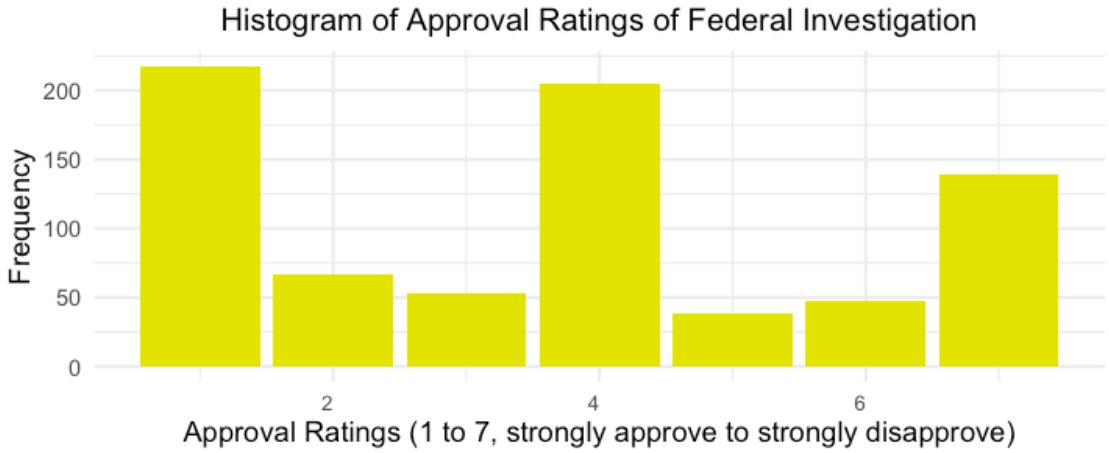
2.3.2 Exploratory Data Analysis

Data Transformation & Justification for Removed Values > On examining the variable [mueller-inv], we found 1 observation that had the value of -7 that corresponds to "no answer". We subset the data to remove that observation. We further subset the data to filter for independents, by applying rules on pid1d and pid1r.

Sanity Checks > For [pid1d] and [pid1r], as we filter the data only on [pid1d] = 3 and [pid1r] = 3 to subset for independent voters, we don't need to worry about any abnormal/missing values. There are no NA values for the variables. All variables have 2500 observations that match the row count of the overall dataset.

Relationship Among Variables > Among the 767 voters that identified themselves as independent, 224 voters disapproved of the investigation, 338 voters approved of the investigation, and 205 voters were neutral towards the investigation. As seen in the histogram below, the distribution of ratings for the federal investigation is not normal. It has high skew in the extreme values (approve extremely strongly, neutral and disapprove extremely strongly). This further advocates for us to dichotomize this variable.

In [55]:



2.3.3 Hypothesis Test Selection

We chose a 1-sample t-test of proportions for the following reasons: - [muellerinv] is an ordinal variable, which we dichotomized into voters who either disapproved or not (approved and neutral). - We have a single variable and each response is independent from the other. - Sample size is large (767 observations).

2.3.4 Hypothesis Test

In [14]:

```
approve or neutral         disapprove
               543                224
```

```
1-sample proportions test without continuity correction

data:  224 out of 767, null probability 0.5
X-squared = 132.67, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.2609673 0.3251992
sample estimates:
        p
0.2920469
```

2.3.5 Conclusion

Statistical Significance > We can reject the null hypothesis that 50% of the independents disapprove the investigations. The p-value is less than 0.05. Thus, the test is statistically significant.

Practical Significance > Only 29% of the independent voters disapprove of the investigations, which means that 71% of them either approve the investigation or are neutral about it, thus, making the result practically significant.

Hence, we can conclude that the majority of the independent voters who took the ANES survey do not believe that the federal investigations of Russian election interference are baseless.

2.4 Question 4: Was anger or fear more effective at driving increases in voter turnout from 2016 to 2018?

2.4.1 Operationalization

This reasearch question implies causality, however, we can only analyze correlations in the data since the data is not from a random sample. Conclusions also only pertain to the pilot study population and cannot be generalized to the larger US population. The research question assumes that fear/anger are associated with increases in voter turnout which may not necessarily be true. This question also uses the word "voters", which we operationalize mean survey respondents regardless of voting behavior.

The survey directly asks respondents whether they voted in 2016 and 2018 and their current level of fear and anger. However, there are two potential gaps between what is asked and the concepts that we are trying to study: (1) anger/fear are only measured in 2018 and we don't know people's level of anger/fear prior to 2016 or if they became more or less angry/fearful between 2016 and 2018, (2) the research question says "anger" and "fear", but does not indicate what those feelings are about. Given the nature of the survey, we assume that they relate to the direction of the country.

A more appropriate research question might be: "Are anger and fear about the direction of the

country in 2018 associated with increased voter turnout between 2016 and 2018 for pilot study respondents?" Relevant questions asked on survey:

- [turnout16] In 2016, the major candidates for president were Donald Trump for the Republicans and Hillary Clinton for the Democrats. In that election, did you definitely vote, definitely not vote, or are you not completely sure whether you voted?
- [turnout18] In the election held on November 6, did you definitely vote in person on election day, vote in person before Nov 6, vote by mail, did you definitely not vote, or are you not completely sure whether you voted in that election?
- Generally speaking, how do you feel about the way things are going in the country these days?
  - [geangry] How angry do you feel?
  - [geafraid] How afraid do you feel?

2.4.2 Exploratory Data Analysis

Data Transformations & Justification for Removed Values We began by subsetting the data to remove respondents that did not remember whether they voted in 2016 or 2018. These cases are not useful because we are unable to accurately track change. Next we created indicator variables for 2016 and 2018 voting (1 = voted; 0 = did not vote).

To analyze voting by anger and fear, we then subset the data into two dataframes: one with respondents that answered the anger question and one with respondents that answered the fear question. It was necessary to exclude people that did not answer these questions because these constructs are critical to the research question. Next we created indicator variables for anger/fear in the relevant dataframes (1 = somewhat/very/extremely; 0 = not at all/a little). We chose this cutoff to ensure more balanced samples and adequate cell size for out hypothesis test (see next section).

Lastly, we created a new categorical variable in each dataframe with four voting behavior categories: increase (did not vote 2016, but voted 2018), decrease (voted 2016, but did not vote 2018), voted both years, did not vote either year.

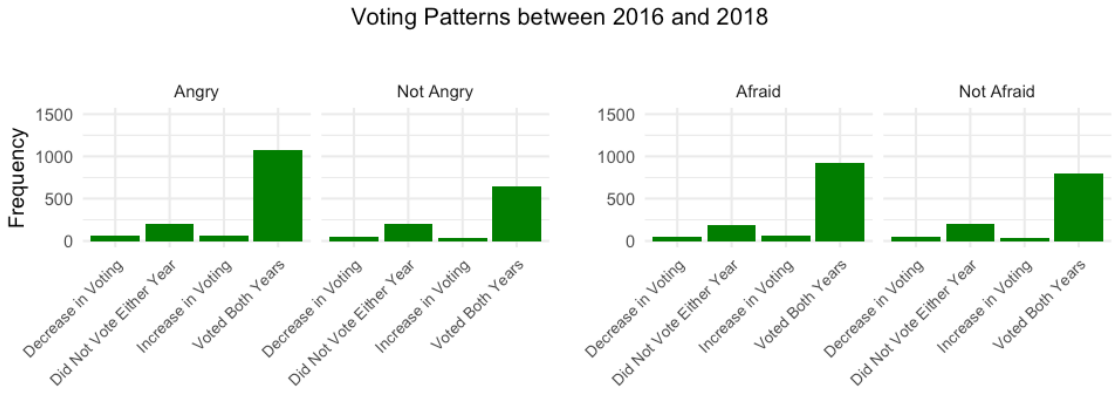
Sanity Checks We did not find missing data in the 2016 and 2018 voting variables. There were a small number of missing values in the anger and fear variables (3 and 6 observations, respectively). These missing values were removed during the subsetting process. The values for the anger and fear questions were all within the expected range (1 through 5).

Relationship Among Variables The table and corresponding bar charts below show that for each anger and fear category the largest proportion of respondents voted in both elections followed by respondents that did not vote in either election. Only a small proportion of respondents increased or decreased their voting between 2016 and 2018. Contrary to the assumptions of our research question, a higher number of angry voters decreased vs. increased their voting during the period (61 vs. 59). However, a higher number of afraid voters increased vs. decreased their voting (58 vs. 52). In both cases the differences are small and will need to be tested to ascertain whether they are statistically significant.

In [34]:

	Angry	Not Angry
Decrease in Voting	61	46
Did Not Vote Either Year	203	196
Increase in Voting	59	39
Voted Both Years	1082	643

	Afraid	Not Afraid
Decrease in Voting	52	55
Did Not Vote Either Year	194	204
Increase in Voting	58	38
Voted Both Years	929	796



2.4.3 Hypothesis Test Selection

We were initially seeking a test of differences in proportions for matched pairs that would allow us to test two dimensions: voting and anger or fear. We considered several tests, but found that each was only able to measure one dimension. We also considered testing the difference in the probabilities of starting vs. stopping voting by anger/fear. Unfortunately, we were unable to identify a test for the difference in differences in probabilites. Ultimately, by collapsing our data into four categories of voting behavior, we were able to eliminate one of the dimensions and the need to conduct a paired test.

We decided that the Chi Square Test was the most appropriate for our data for the following reasons:

- The test can be applied to categorical data with 2 or more categories
- The test evaluates the association between categories of two variables (in our case, difference in voting categories and anger/fear categories)
- The test assumes sufficient sample size (we have 2,000+ observations in both the anger and fear dataframes)

- The test assumes cell counts of 5 or more (our lowest cell count is 38)
- Observations are assumed to be independent (our collapsed difference data meets this assumption)

The Chi Square Test also assumes random sampling. Although our data was not generated through a random sample, we decided that this test was still the most appropriate although results must be interpreted with caution.

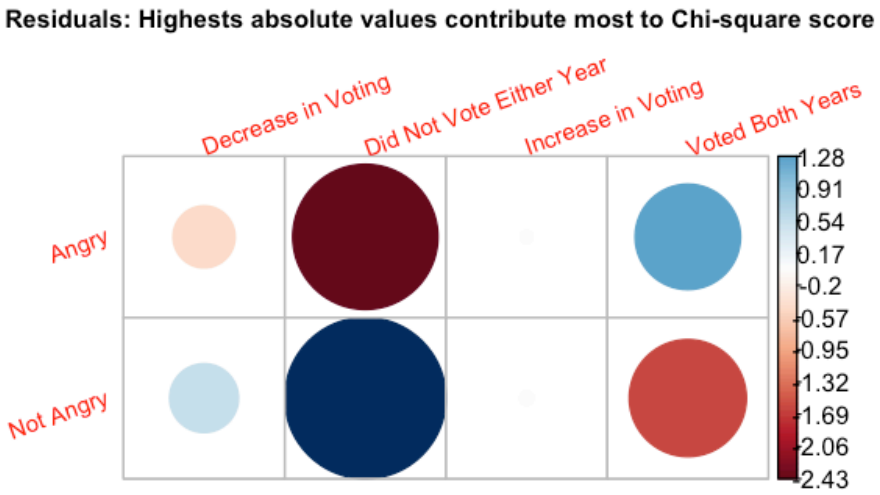
2.4.4 Hypothesis Test

In [53]:

```
Pearson's Chi-squared test

data:  q4_data_angry$angry_string and q4_data_angry$voting_cat_string
X-squared = 19.523, df = 3, p-value = 0.0002131
```

```
Pearson's Chi-squared test

data:  q4_data_afraid$afraid_string and q4_data_afraid$voting_cat_string
X-squared = 6.3531, df = 3, p-value = 0.09564
```

```
q4_data_angry$angry_string Decrease in Voting Did Not Vote Either Year
                 Angry                -0.442                   -2.430
                 Not Angry             0.545                    2.997

q4_data_angry$angry_string Increase in Voting Voted Both Years
                 Angry                -0.016           1.282
                 Not Angry             0.019          -1.581
```



2.4.5 Conclusion

Statistical Significance **Anger**: We reject the null hypothesis that there is no relationship between anger and voting behavior. The Pearson's Chi-Squared Test has a value of 19.523 which corresponds to a p-value of 0.0002131, condsiderably lower than the commonly used 0.05 threshold. This indicates that there is a relationship between the two anger categories and the voting behavior categories. In order to better understand which categories contribute most to the Chi-Square score, we calculated the residuals between the expected and observed values for each cell (see table and plot above). The formula for the residuals is shown below where r = residual, o = observed value, and e = expected value:

$$r = \frac{o - e}{\sqrt{e}}$$

We see that the cells for not voting at all and voting in both years are contributing most to the Chi-Square score. Interestingly, the cell for increase in voting and anger is contributing the least to the score (absolute value of 0.016). Given these results, we are unable to find an association between anger and increases in voting.

Fear: We are unable to reject the null hypothesis that there is no relationship between fear and voting behavior. The Pearson's Chi-squared test has a p-value of 0.096, considerably higher than the commonly used 0.05 threshold.

Practical Significance These results indicate that, among survey respondents, there is a relationship between anger and voting patterns, but not fear and voting patterns. The relationship between anger and voting patterns primarily relates to anger level and voting in both years or not voting in both years. Therefore, we conclude that neither anger nor fear is associated with increasing voting between 2016 and 2018. As mentioned earlier, these results only apply to the pilot survey population and cannot be generalized to the larger US population.

2.5 Question 5: Do Republican or Democratic survey respondents think that voter suppression is more of a problem?

2.5.1 Operationalization

This research question is topical because controversy around voter suppression is often in the news, including cuts to early voting and voter ID laws. There is also a common perception that Democrats think voter suppression is more of a problem than Republicans. We wanted to test whether this perception is true for the pilot survey population.

A potential limitation is the way that respondents were asked about their views on voter suppression. Respondents were randomly selected into two groups and given different questions (see below). The first question asks how often people that are "legally" allowed to vote are "stopped" from voting and the second question asks how often people that are "eligible" to vote are "denied" the right to vote. The different phrasing could elicit different responses from different people and these responses could vary by party. The party affiliation questions were also asked in two ways (see below), but the phrasing is more similar. Relevant questions asked on survey:

- [votestop1] How often are people who are legally allowed to vote stopped from voting?
- [votestop2] How often are people who are eligible to vote denied the right to vote?
- [pid1d] Generally speaking, do you usually think of yourself as a Democrat, a Republican, an independent, or what?
- [pid1r] Generally speaking, do you usually think of yourself as a Republican, a Democrat, an independent, or what?

2.5.2 Exploratory Data Analysis

Data Transformations & Justification for Removed Values Since respondents were randomly assigned to different questions, we began by combining the data to create one variable for party affiliation and one variable for voter suppression. Next, we subset the data to only include respondents that identified as Republicans or Democrats, the two groups relevant to our research question.

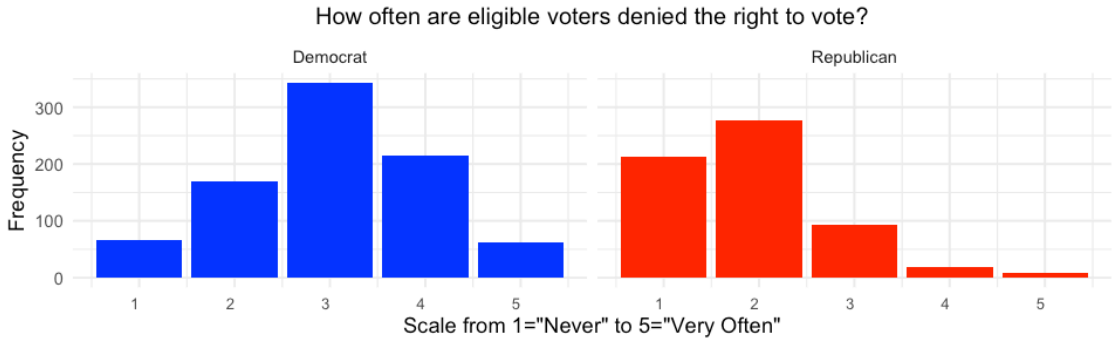
Sanity Checks There were no missing values in the voter suppresssion question and responses only included the expected values for the scale (1 through 5). We observed 150 missing values in the party affiliation quesiton, but, otherwise, values were the expected values for the question scale (1 through 4). The missing values were removed when the data was subset to only include Republicans and Democrats.

Relationship Among Variables The party groups are fairly large, but unbalanced: 875 Democrats and 609 Republicans. In the bar chart and table below, we visualize responses to the voter suppression variable. The distributions differ dramatically between the two parties. The largest proportion of Republicans responded that voter suppression 1="never" or 2="rarely" happens while the largest proportion of Democrats responded that it happens 3="occasionally" or 4="fairly often". Few Republicans responded that voter suppression happens more than "occassionaly". The distribution for Democrat voters is fairly normal, but there is a right skew for Republican voters.

In [54]:

'Number of Democrat observations: 857'
'Number of Republican observations: 609'

	Democrat	Republican
Never	67	212
Rarely	169	276
Occasionally	343	94
Fairly Often	215	19
Very Often	63	8



2.5.3 Hypothesis Test Selection

Because our data is non-parametric and ordinal, we decided that it best meets the assumptions of the Wilcoxon Rank-Sum Test. The Wilcoxon Rank-Sum Test assumes:

- Observations from both groups are independent (our two groups are independent)
- Responses are ordinal (the voter suppression variable is ordinal)
- Distributions of both populations are equal under the null hypothesis (this aligns well with our research question)
- This is a non-parametric test and does not assume that the data is normal (the Republican voter suppression distribution has a right skew)

2.5.4 Hypothesis Test

In [32]:

```
Wilcoxon rank sum test with continuity correction

data:  votestop by party
W = 415044, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
```

```
 1.000005 1.000056
sample estimates:
difference in location
              1.000051
```

2.5.5 Conclusion

Statistical & Practical Significance The Wilcoxon Rank Sum Test has a p-value of 0.00000000000000022, much lower than the commonly used 0.05 threshold. Therefore, we reject the null hypothesis that the distribution of the ranks is equal for both populations. The test also estimates that the median of the difference between a sample from the Republican and Democrat ranking distributions is approximately 1.000051. This means that if we were to draw repeatedly from the Democrat and Republican rankings we would observe a median difference between rankings of approximately one point, on average. We conclude that Democrat survey respondents think that voter suppression is more of a problem than Republican survey respondents.

While our findings cannot be generalized to the larger US population, they do have practical significance. As mentioned previously, there is a common perception that Democrats think voter suppression is more of a problem than Republicans. Our analysis confirms that this is true for the survey population.

Non-Technical Summary Among pilot survey respondents, Democrats think that voter suppression is more of a problem than Republicans. In fact, overall, Democrats selected rankings that are approximately one point higher than Republicans on the voter suppression Likert scale. Our results are statistically significant, which means that we can be reasonably sure that the differences that we see in survey responses are due to real differences in how these groups view voter suppression rather than just due to chance.