Lab 1: Question 2: Are Democratic voters more enthusiastic about Joe Biden or Kamala Harris?

Elaine Chang, Dom Dillingham, Jesse Miller, Michael Wang

1. Importance and Context

In this analysis, we will be exploring whether Democratic voters are more enthusiastic about Joe Biden or Kamala Harris. With a contentious Democratic primary, we believe this exploration can help explore whether the selection of Kamala Harris as Vice President may have made the Democratic ballot more or less positive to those voting Democratic. There is often much discussion about the selection of Vice President on the ballot to make the ballot more appealing to a wider base of voters and this analysis can help determine whether the selection of Kamala Harris helped make the combined ballot more appealing to Democratic voters.

2. Description of Data

The data for this analysis is drawn from the 2020 Pre-Election Data by the American National Election Studies (ANES). This data was collected based on a variety of survey methods to sample the 231 million US citizens that are of voting age. An exploration of how representative this sample is of the US population and the independence of survey respondents is explored later in this analysis.

For this analysis, we will operationalize a Democratic voter to be whether the respondent had voted, intends to vote, or has a preference for a Democratic presidential candidate in the 2020 election and who intends to vote in the upcoming election. The former field, field `V201075x`, is a summary field calculated by the ANES based on whether the respondent has voted and, if not, whom they prefer. The latter field, field `V201019` indicates whether or not the respondent intends to register for elections for the 2020 election. The combination of which allows us to narrow our scope just to Democratic voters. Further, we will operationalize a respondent's preference for Joe Biden or Kamala Harris, fields `V201151` and `V201153`, respectively, with a respective `Feeling Thermometer` where the respondent ranks their feeling on a scale from 0 to 100, where responses greater than 50 indicate favorable feelings, less than 50 indicate unfavorable, and 50 being indifferent.

The operationalization of a Democratic voter is one of interpretation. Instead of an individual's intent or preference to vote, we could also have operationalized with a respondent's party of registration or their personal identification. Given that not all future voters were registered at the time of the survey and that they can vote for a candidate outside of their party, we believe the selected variable is more appropriate to define an individual as a Democratic voter or not. Further, an individual who identifies with one party may vote for another party due to the candidate pool. Therefore, a Democratic voter should be one intending to vote for a Democrat. To operationalize a voter, we use the individual's intent or register or if they have already registered to serve as a proxy for an individual voting in the upcoming election.

The use of the `Feeling Thermometer` does appear to be an appropriate variable for this study. An individual who is more enthusiastic about one politician over another will likely feel more warm about said politician, resulting in a higher thermometer for that politician.

For this analysis, the total dataset was just filtered down to those voting for a Democratic presidential candidate. Further, any erroneous values for either the Joe Biden or Kamala Harris thermometer are removed.

From Table 1 below we see that more individuals prefer Biden over Harris when the individual selected a thermometer reading that was not the same between the two politicians.

Table 1: Data Distribution by Politician

Record Count	Biden Greater Than Harris	Harris Greater Than Biden	Biden and Harris Equal
7526	2332	1925	3269

Given the thermometer scale is on an ordinal scale, differences between observations do not have a direct meaning but it is informative to observe the relative count of the spread between thermometer readings for both politicians. From the graph below we confirm our observations from above that most respondents provide the same response for Joe Biden and Kamala Harris. We also observe that the distribution is tight,

with the majority of responses between -25 and +25 of one another though more respondents have a more positive response to Joe Biden than Kamala Harris.
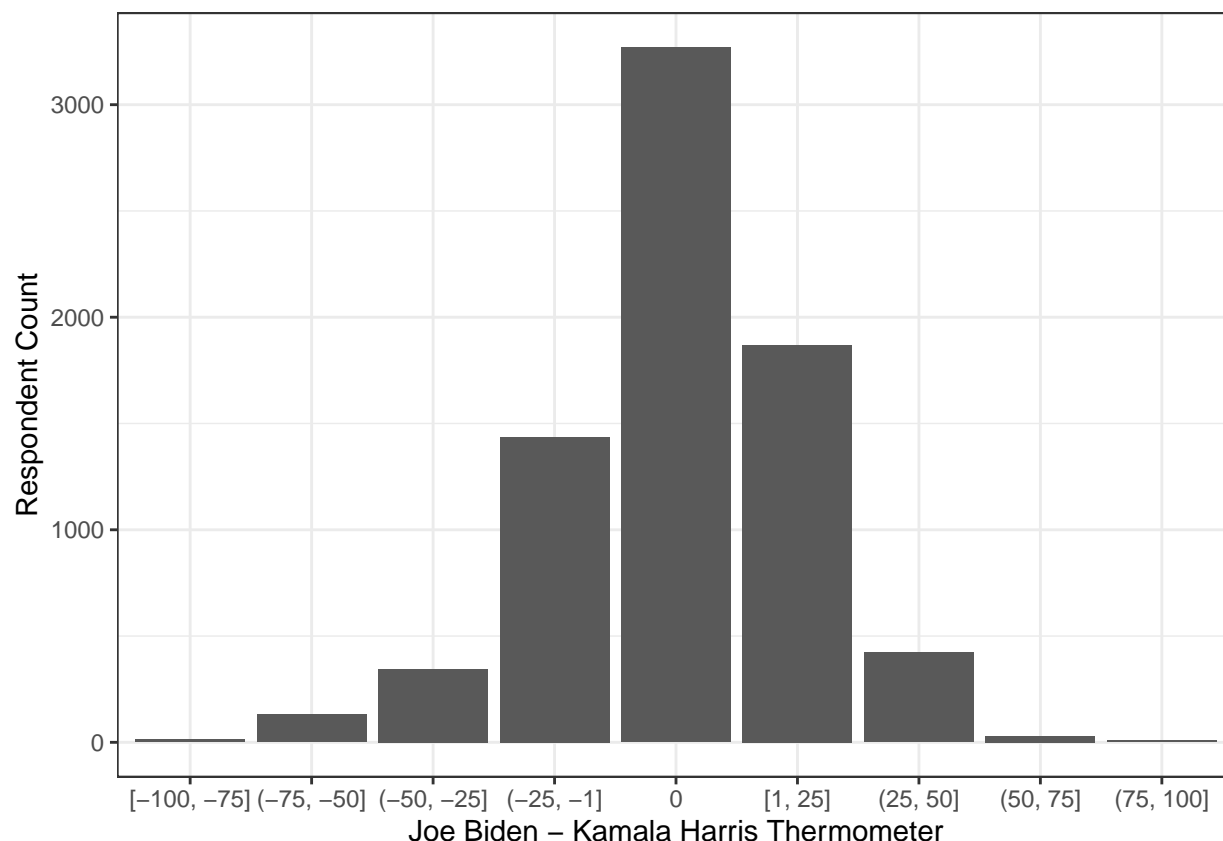

Figure 1: Difference Between Joe Biden and Kamala Harris

3. Most appropriate test

The most appropriate test for this analysis a Sign Test. For this problem, a paired statistical test is preferred as we are not controlling for the influence of other factors that may impact the difference such as age and historical voting patterns. Within the range of paired tests, the Sign Test passes or nearly misses the assumptions required for this test. The assumptions are:

1. Paired dependent samples - Each of the drawings are paired dependent samples as the same individual would be asked both questions.
2. Data is drawn IID - This data does not fully align with the properties of IID. From the properties of the survey, some individuals are more likely to be sampled and weights to account for this sampling bias were provided. For this analysis, these weights were ignored.

While the additional power of the Wilcoxon Signed Rank Test would be preferred, this test requires our data to be on a metric scale which this thermometer reading is not. Our hypothesis test then is:

- H0: There is no difference in median of the signed differences between the Joe Biden and Kamala Harris thermometer

- Ha: There is a difference in median of the signed differences between the Joe Biden and Kamala Harris thermometer

4. Test, results and interpretation

```r
binomial_test <- dat %>%
  filter(!is.na(diff)) %>%
  mutate(counter = 1) %$%
  binom.test(x = sum(diff), n = sum(counter), alternative = 'two.sided')
```

The result of the statistical test indicates that, with an alpha-level of 0.05 on a two-sided test, we reject the null hypothesis, given the p-value of 0.00088, that there is no difference in the thermometer between Joe Biden and Kamala Harris. An alpha-level of 0.05 was selected given this is a common alpha-level selected throughout the industry. Further, the two-sided test was selected for a more conservative test given there is no *a priori* knowledge that Joe Biden or Kamala Harris would be more preferred.

We observe that Joe Biden is preferred to Kamala Harris 53.4% of the time for Democratic voters, excluding any ties. This practical effect size, 3.4% nominally relative to an expected 50%, is small. While we have seen this to be statistically different from the analysis above, the large proportion of respondents who responded the same for Joe Biden and Kamala Harris are ignored by this statistical test. In summary, the small observed nomial effect size (3.4%) in favor of Joe Biden, the democratic ballot was not made drastically more positive by the addition of Kamala Harris.

5. Limitations

There are many limitations to this analysis, highlighted below:

- For this analysis, we are ignoring the weights to correct for sampling bias that are provided with this dataset. Our analysis is therefore biased to the same extent that the sample is biased.
- As previously mentioned, the statistical test selected does not account for high density of respondents who selected the same response for both Joe Biden and Kamala Harris.