Lab 1: Question 1

Elaine Chang, Dom Dillingham, Jesse Miller, Michael Wang

```r
library(ggplot2)
library(tidyverse) # if you want more, but still core, toolkit
library(magrittr)
library(reshape2)
library(knitr)

# Load dplyr last for any package issues
library(dplyr)
```

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 may have made the Democratic ballot more positive to those voting Demomcratic.

Description of Data

For this analysis, we will operationalize a Democratic voter to be whether the respondent had voted, intention to vote, or preference for a Democratic presidential candidate. This field is a summary field calcuated by the American National Election Studies. Further, we will operationalize a respondent's preference for Joe Biden or Kamala Harris with a respective `Feeling Thermometer` where the respondent ranks their feeling on a scale from 0 to 100.

The operationalization of a Democratic voter is one of interpretation. Instead of an individuals intent or preference to vote, we could also have operationalized with a respondent's party of regristration. While not all voters are registered, we beleive the selected variable is more appropriate.

The use of the `Feeling Thermometer` does appear to be an appropriate variable for this study. This will allow for better distinction between an individual's perception of each candidate.

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 the difference in thermometer between Joe Biden and Kamala Harris is very close, with Joe Biden being higher by just 1.6 points. While there does appear to be preference for Joe Biden, it is not by much.

Table 1: Data Distribution by Candidate

Record Count	Average Kamala Harris Thermometer	Average Joe Biden Thermometer	Average Difference
3974	75.06	76.64	1.58

From the table below, we see that, along with a small difference in mean, there isn't a strong visual difference between the distributions of thermometer by Joe Biden and Kamala Harris though Kamala Harris has a higher distribution near 50.

```
## Warning: attributes are not identical across measure variables; they will
## be dropped
```

Violin of Thermometer

Difference Between Kamala Harris & Joe Biden
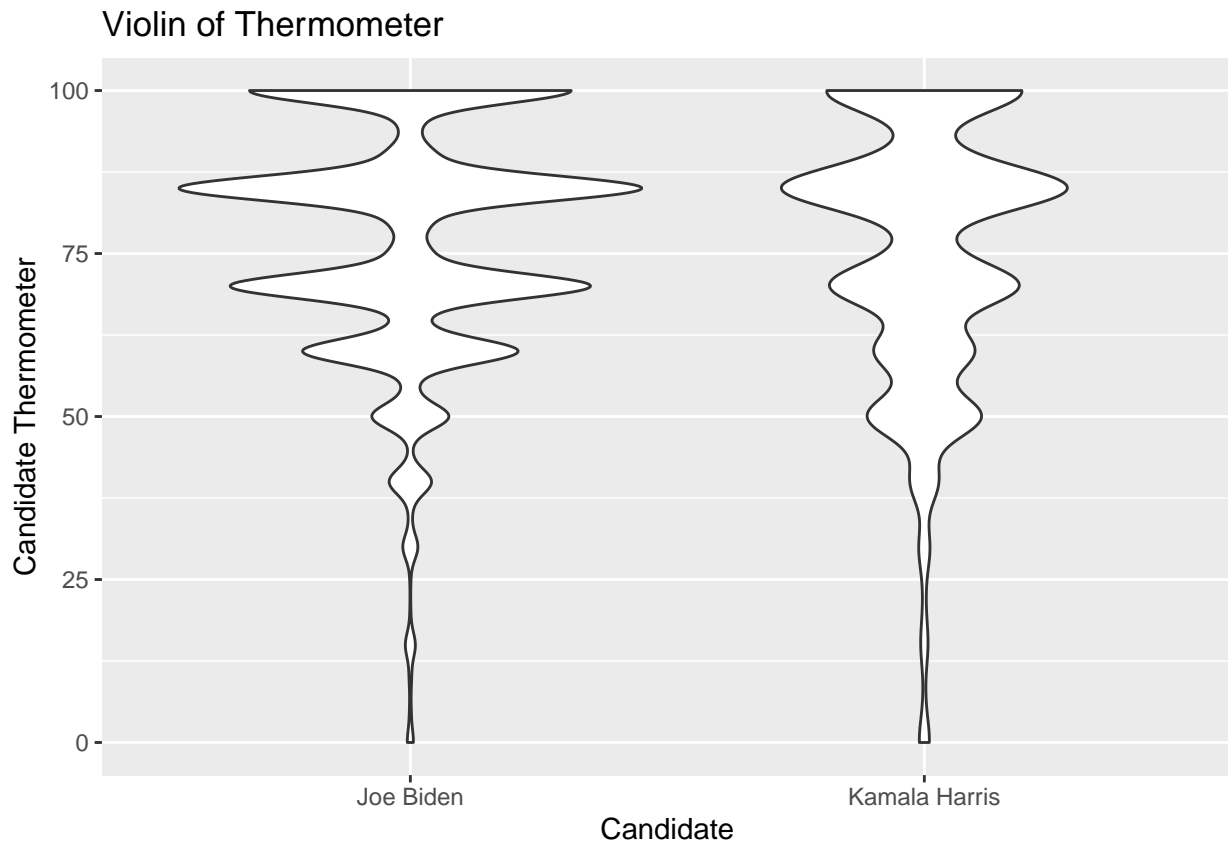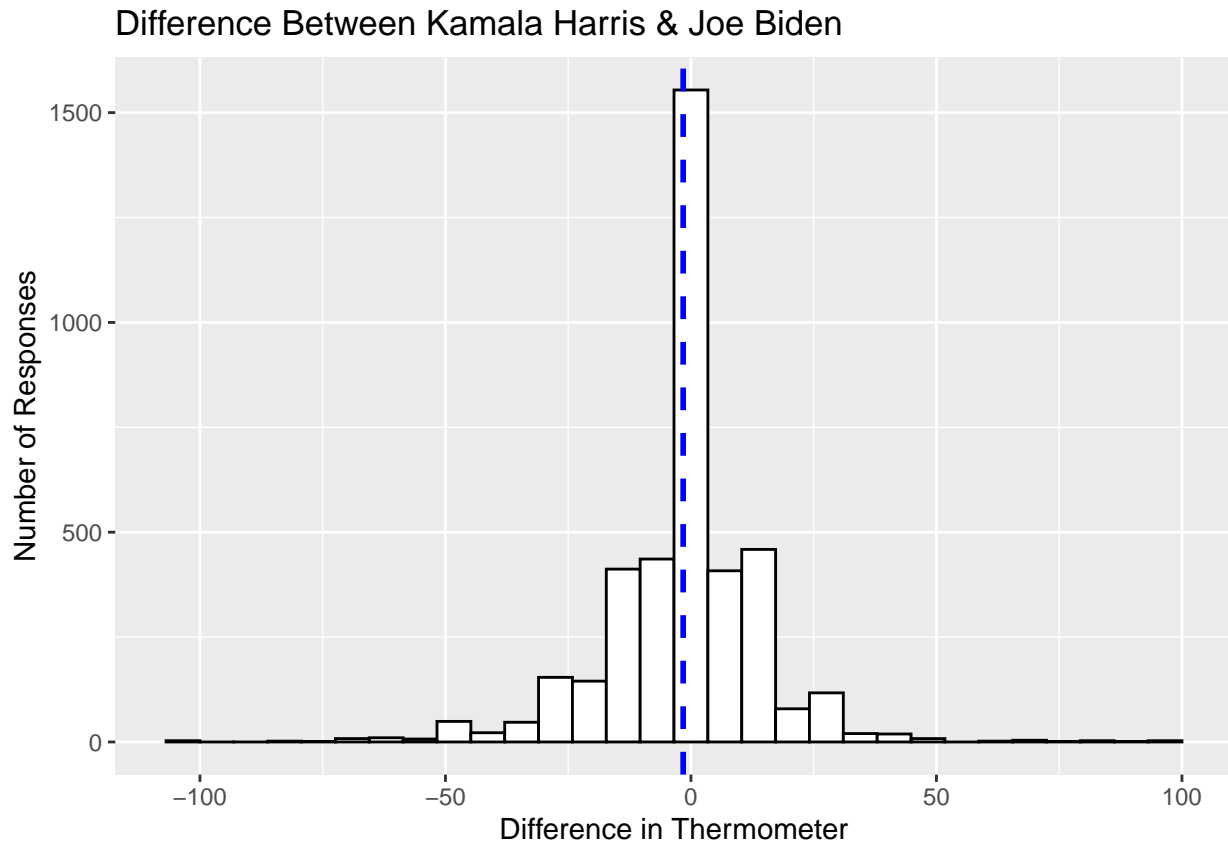
Most appropriate test

The most appropriate test for this analysis a Wilcoxon Signed Rank Test. For this problem, a paired statistical test is prefered as we are not controlling for the influence of other factors that may impact the difference such as age and historical voting patterns, which a paired test will help account for. Within the range of paired test, the Wilcoxon Signed Rank Test pass all the tests assumptions while providing the most power. The assumptions are:

- Metric scale - The thermometer is a metric scale?
- Paired dependent samples - Each of the drawings are paired dependent samples as the same individual would be asked both questions.
- Difference between samples is symmetric about mean - From the graph above, it does appear that the data is symmetric about the mean.
- Paid is drawn IID - ?

Test, results and interpretation

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dat$Kamala_Meter and dat$Biden_Meter
## V = 1293058, p-value = 3.531e-08
## alternative hypothesis: true location shift is not equal to 0
```

The result of the statisical test indicate that, with an alpha-level of 0.05, we reject the null hypothesis that the difference bewteen thermometers of Joe Biden and Kamala Harris are 0.

With the small measured difference between Kamala Harris and Joe Biden, while there is a statisically significant difference between the politicians, the practical significance is small.