POLS 7012: Practice Midterm (Answer Key)

Joe Ornstein

Create an R project, an R script, load tidyverse, and read in the ANES 2019 Pilot data from Problem Set 1.

```
# Load the ANES dataset
load('data/anes_pilot_2019.RData')
```

Tidy the Data

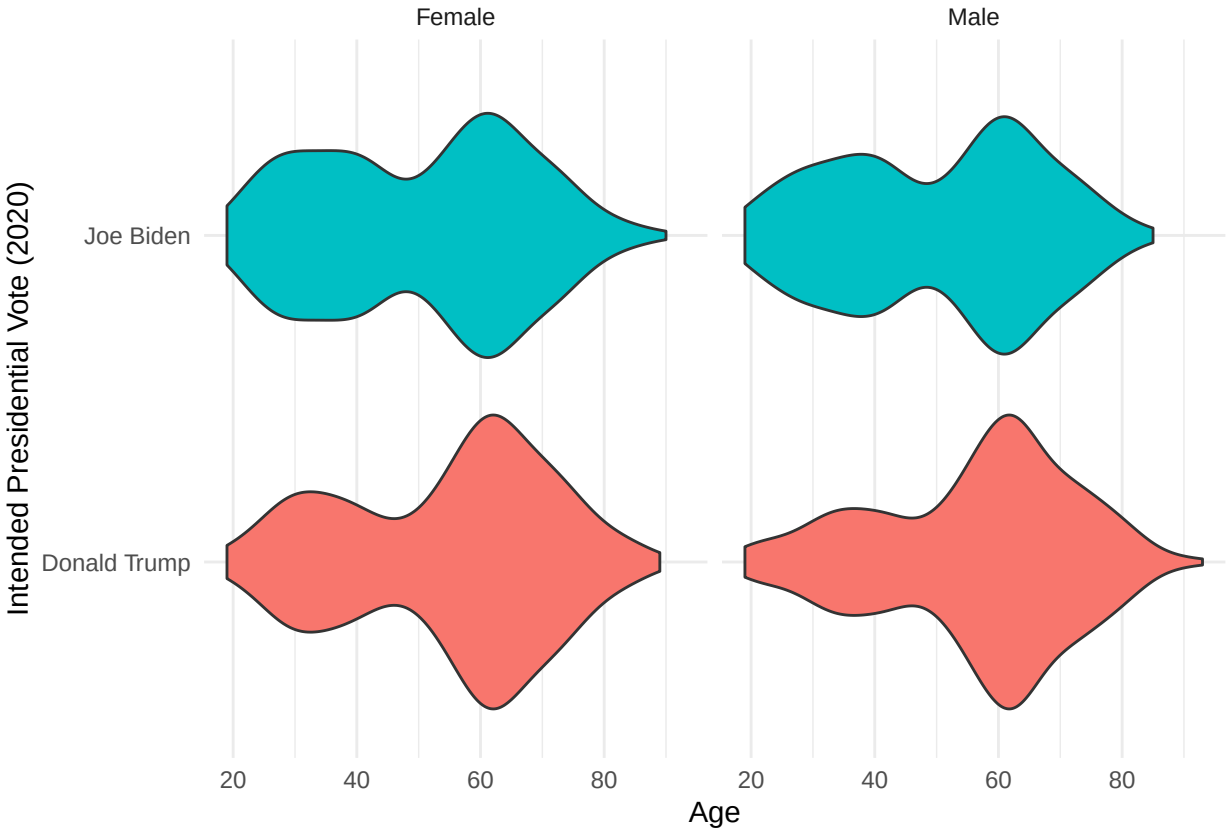
1. Recode the `vote20jb` variable with the actual responses rather than the numeric codes (see the codebook).
2. Recode gender as "Male" or "Female" instead of 1 or 2.
3. Create a new variable called `age`, equal to 2019 minus the respondent's birth year.
4. Filter out the respondents who say they will "probably not vote".
5. Select only the `caseid`, `gender`, `age`, and `vote20jb` variables.
6. Save the cleaned up dataset to your `data/` folder as a .csv file called `anes_2019_pilot_clean.csv`.

```
data <- data %>%
  mutate(vote20jb = case_when(vote20jb == 1 ~ 'Donald Trump',
                              vote20jb == 2 ~ 'Joe Biden',
                              vote20jb == 3 ~ 'Someone Else',
                              vote20jb == 4 ~ 'Probably Not Vote'),
         gender = if_else(gender == 1, 'Male', 'Female'),
         age = 2019 - birthyr) %>%
  filter(vote20jb != 'Probably Not Vote') %>%
  select(caseid, gender, age, vote20jb)

write_csv(data, 'data/anes_2019_pilot_clean.csv')
```

Visualize and Summarize the Data

1. Visualize the age distribution for Biden vs. Trump voters, faceting by gender.

```
figure1 <- data %>%
  filter(vote20jb %in% c('Donald Trump', 'Joe Biden')) %>%
  ggplot +
  geom_violin(mapping = aes(x=age, y=vote20jb, fill = vote20jb)) +
  facet_wrap(~gender) +
  labs(x = 'Age', y = 'Intended Presidential Vote (2020)') +
  theme_minimal() +
  theme(legend.position = 'none')

figure1
```



2. Group by `vote20jb` and `gender`. What is the mean age and number of respondents for each group?

```
data %>%
  group_by(vote20jb, gender) %>%
  summarize(mean_age = age %>% mean %>% round(1),
            num = n()) %>%
  kable # this is just to make it look fancy in RMarkdown
```

vote20jb	gender	mean_age	num
Donald Trump	Female	55.1	579
Donald Trump	Male	56.7	694
Joe Biden	Female	50.4	696
Joe Biden	Male	50.3	592
Someone Else	Female	43.5	181
Someone Else	Male	46.0	140

Make Inferences

1. Compute a 95% confidence interval for the average age.

```
# Approach 1: Take the sample mean plus or minus 2 standard errors
sample_mean <- mean(data$age)
standard_error <- sd(data$age) / sqrt(nrow(data))

confidence_interval <- c(sample_mean - 1.96 * standard_error,
                         sample_mean + 1.96 * standard_error)
```

```
confidence_interval
```

```
## [1] 51.56400 52.80311
```

```
# Approach 2: Confidence Interval reported with the t.test function
t.test(data$age)
```

```
##
##  One Sample t-test
##
## data:  data$age
## t = 165.09, df = 2881, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  51.56375 52.80336
## sample estimates:
## mean of x
##  52.18355
```

2. Is the difference in age between Trump and Biden supporters statistically significant? Compute a p-value and 95% confidence interval.

```
test1 <- data %>%
  filter(vote20jb %in% c('Donald Trump', 'Joe Biden')) %>%
  t.test(age ~ vote20jb, data = .)

test1
```

```
##
##  Welch Two Sample t-test
##
## data:  age by vote20jb
## t = 8.6957, df = 2557.5, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.388284 6.943648
## sample estimates:
## mean in group Donald Trump    mean in group Joe Biden
##                   55.98429                   50.31832
```

```
# I didn't show you this before, but if you save the test to an object,
# you can extract the relevant values like so.
test1$p.value
```

```
## [1] 6.049895e-18
```

```
test1$conf.int
```

```
## [1] 4.388284 6.943648
## attr(,"conf.level")
## [1] 0.95
```

3. Are men significantly more likely than women to support Trump? Compute the associated p-value.

```
test2 <- data %>%
  filter(vote20jb %in% c('Donald Trump', 'Joe Biden')) %>%
  select(vote20jb, gender) %>%
  table %>%
```

```
  chisq.test

test2
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  .
## X-squared = 18.399, df = 1, p-value = 1.792e-05
```

```
# Here's the expected table if the two variables were independent
test2$expected
```

```
##               gender
## vote20jb         Female     Male
##   Donald Trump 633.7661 639.2339
##   Joe Biden    641.2339 646.7661
```

```
# Here's the actual table
test2$observed
```

```
##               gender
## vote20jb       Female Male
##   Donald Trump    579  694
##   Joe Biden       696  592
```

```
# Here's the p-value
test2$p.value
```

```
## [1] 1.791662e-05
```