DATA 606 Data Project Proposal

Josh Iden

Data Preparation

```
# load data
library(tidyverse)

url = "https://raw.githubusercontent.com/fivethirtyeight/data/master/non-voters/nonvoters_data.csv"
df = read.csv(url)

head(df)
```

Research question

Is education level predictive of voter attitudes amongst non-voters?

Cases

The dataset contains survey data from **5,836 respondents**

Data collection

The data was collected from polling done by Ipsos for FiveThirtyEight for the story Why Many Americans Don't Vote. Details about the data collection can be found here

Type of study

This study is observational, relying on data taken from a random sample survey.

Data Source

The data is courtest of FiveThirtyEight

Dependent Variable

The response variables are different attitudes about elections, ie, "Do elections matter?", "How easy or difficult is it to vote?", and current voter registration.

Independent Variable(s)

The independent variable is level of education

Relevant summary statistics

Provide summary statistics for each the variables. Also include appropriate visualizations related to your research question (e.g. scatter plot, boxplots, etc). This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

```
cols = c("educ","Q5","Q16","Q20")
df_sub = df[cols]
head(df_sub)
```

```
##                     educ Q5 Q16 Q20
## 1                College  1   1   1
## 2                College  1   2   1
## 3                College  1   1   1
## 4           Some college  1   4   1
## 5 High school or less  1   1   1
## 6 High school or less  2  -1   2
```

```
ggplot(df_sub, aes(educ)) +
  geom_bar(stat="count")
```
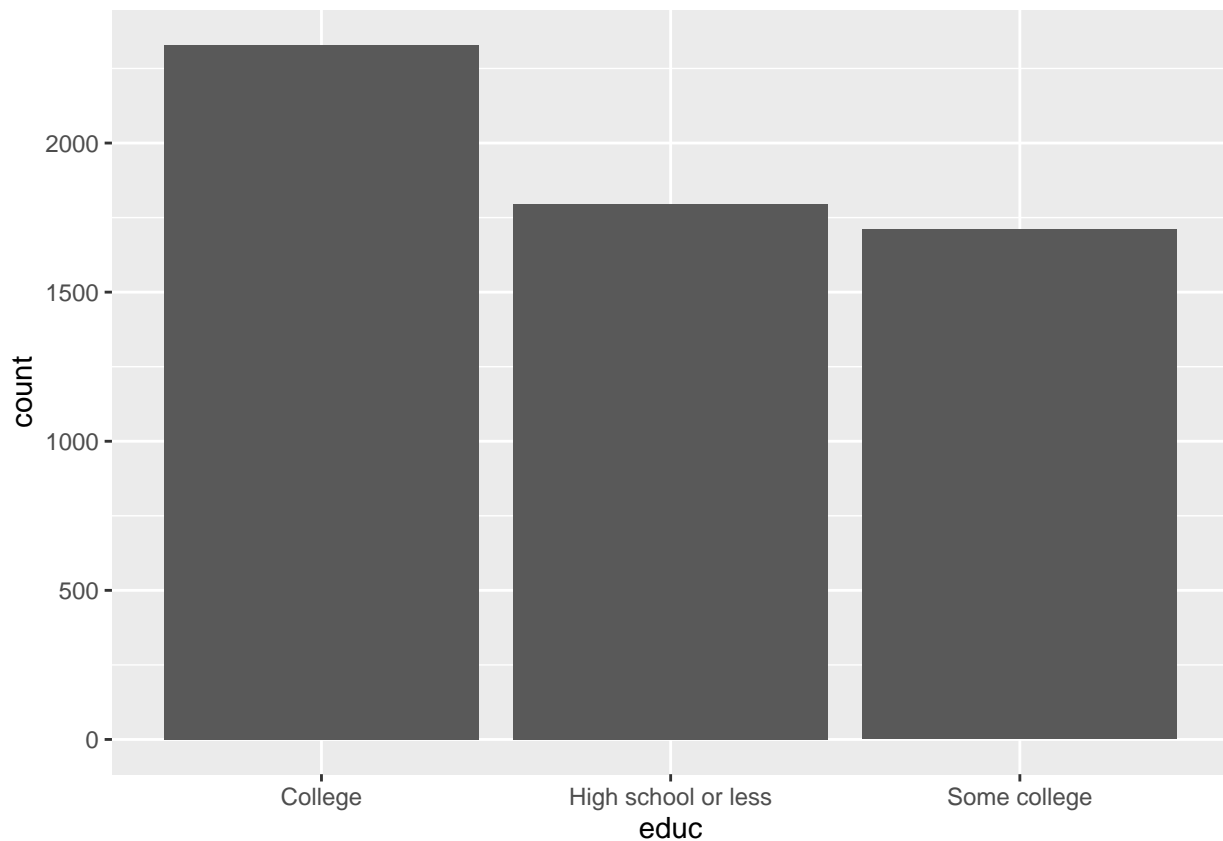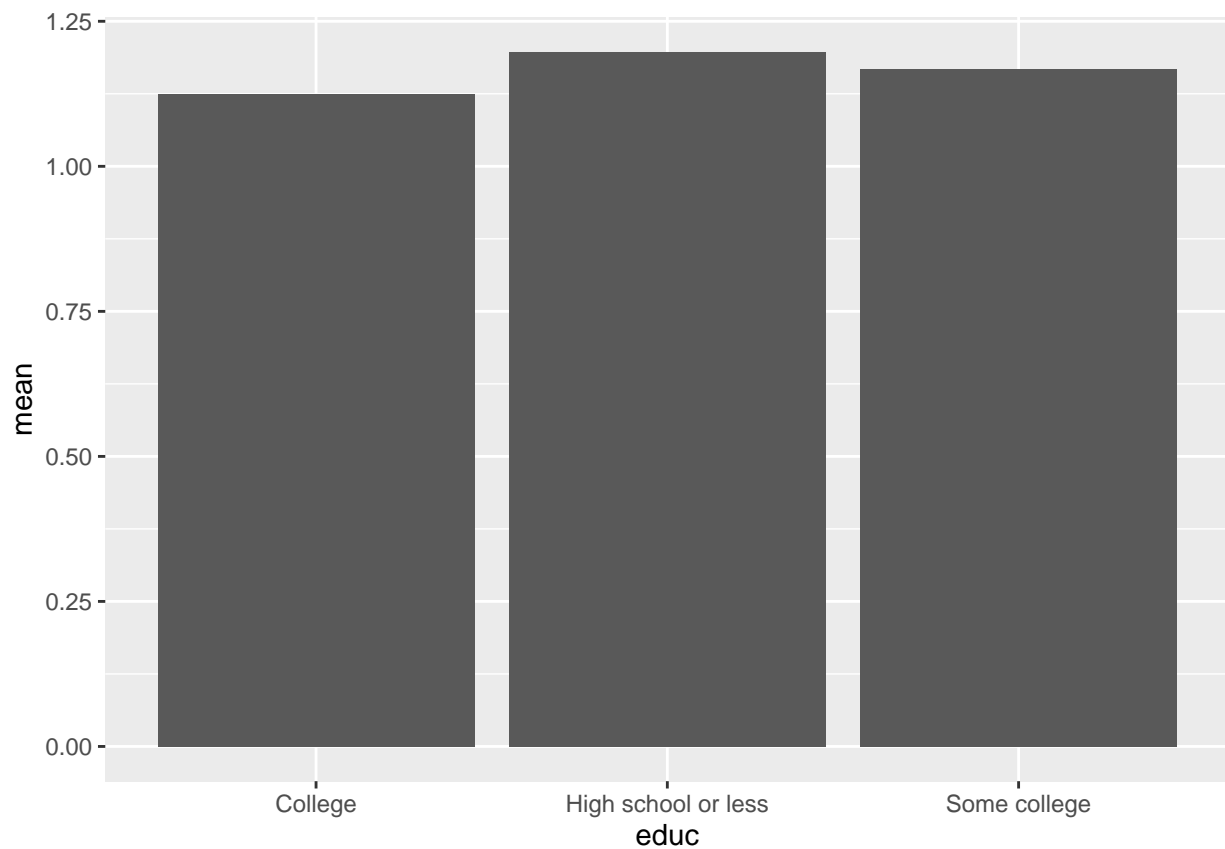

Does it matter who wins elections? 1 = Yes, 2 = No

```
Q5 = df_sub |>
  select(c(educ, Q5)) |>
```

```
  group_by(educ) |>
  summarize(mean = mean(Q5))

Q5
```

```
## # A tibble: 3 x 2
##   educ                mean
##   <chr>              <dbl>
## 1 College             1.12
## 2 High school or less 1.20
## 3 Some college        1.17
```

```
ggplot(Q5, aes(x=educ, y=mean)) +
  geom_bar(stat="identity")
```


We can see from a high level, proportionally, those with less education are more skeptical of the difference voting makes
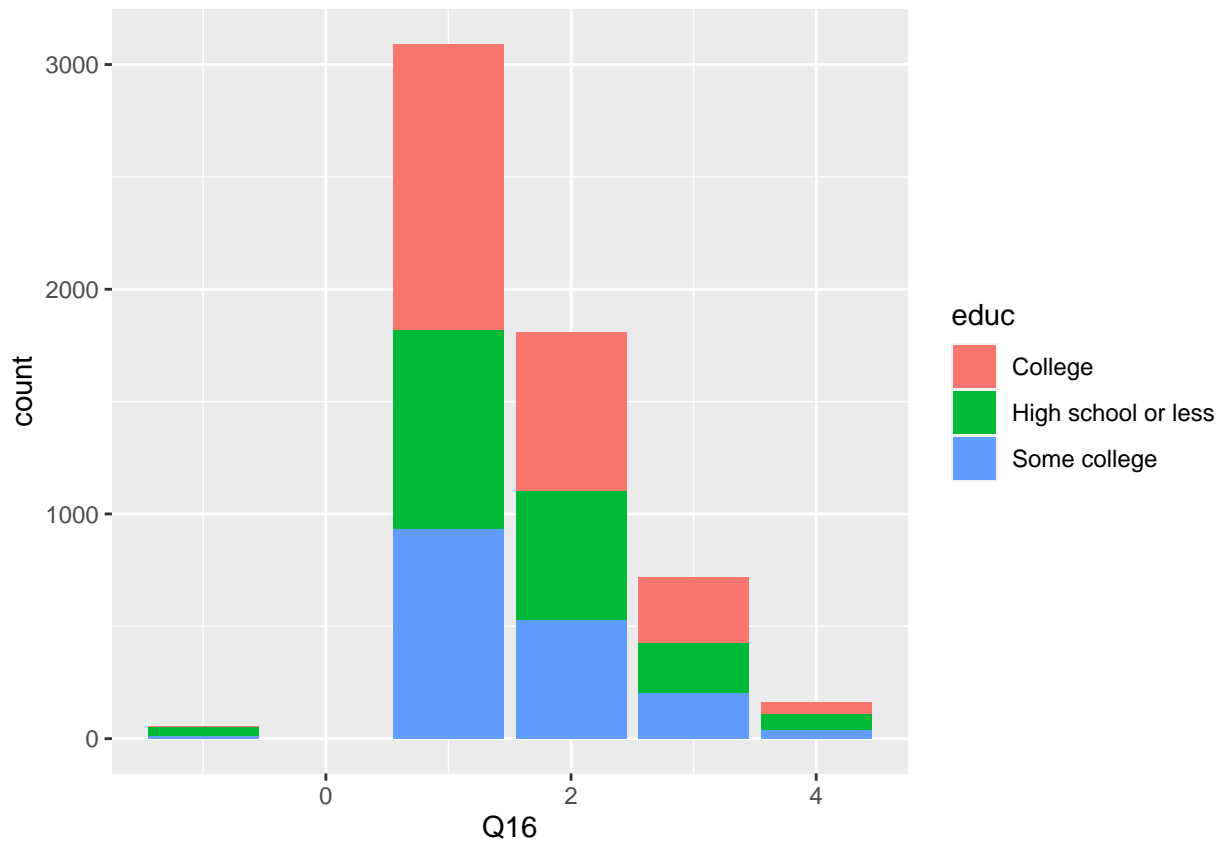
In general, how easy or difficult is it to vote in elections? 1= Very Easy, 4 = Very Difficult

```
Q16 <- df_sub |>
  select(c(educ, Q16)) |>
  group_by(educ) |>
  summarize(mean = mean(Q16))

Q16
```

```
## # A tibble: 3 x 2
##   educ               mean
##   <chr>             <dbl>
## 1 College            1.62
## 2 High school or less  1.64
## 3 Some college       1.60
```

```r
ggplot(df_sub, aes(Q16, fill=educ)) +
  geom_bar(stat="count")
```


Surprisingly, those with some college have more favorable opinions regarding the ease of voting than those with college degrees, although respondents with high school educations or less find voting harder

Are you currently registered to vote? $1 = $ Yes, $2 = $ No

```r
Q20 <- df_sub |>
  select(c(educ, Q20)) |>
  group_by(educ) |>
  summarize(mean = mean(Q20))

Q20
```
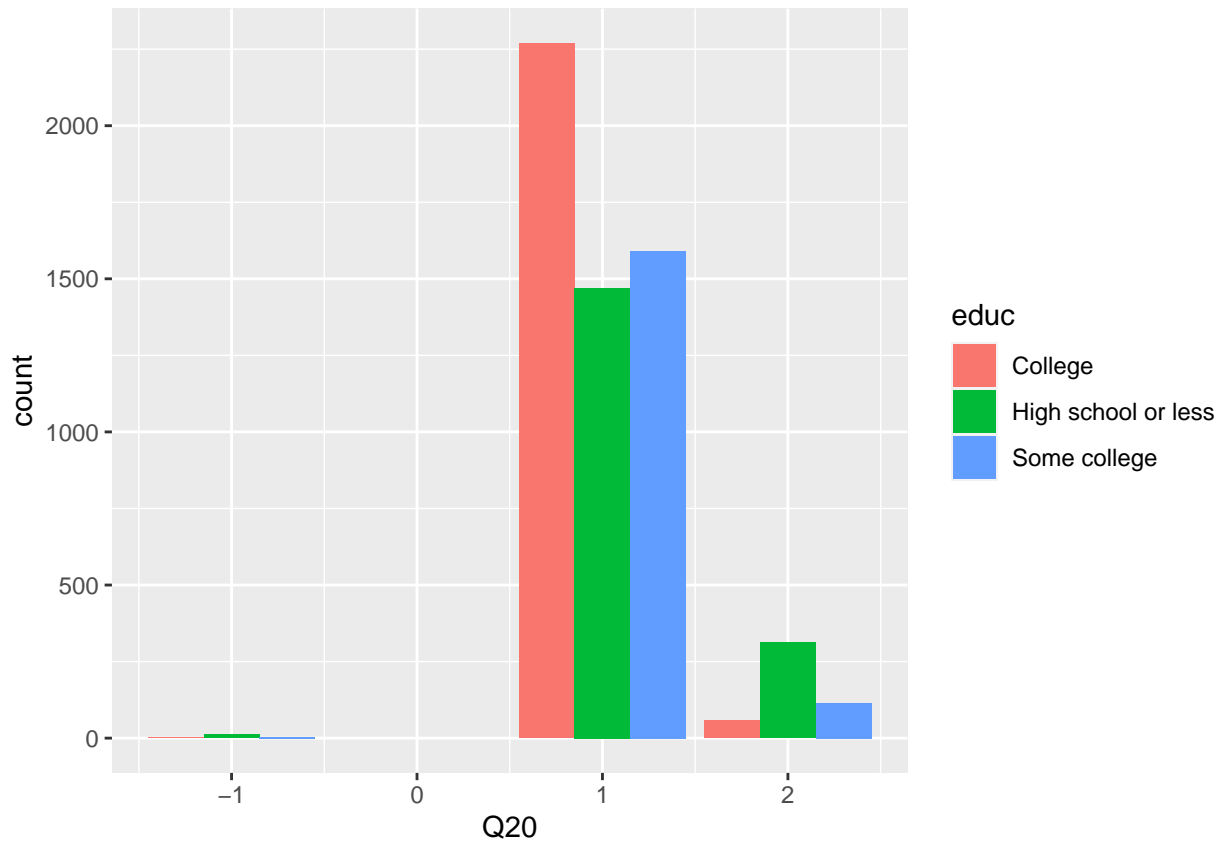
```
## # A tibble: 3 x 2
##   educ               mean
##   <chr>             <dbl>
## 1 College            1.02
```

```
## 2 High school or less  1.16
## 3 Some college         1.06
```

```
ggplot(df_sub, aes(x=Q20, fill=educ)) +
  geom_bar(stat="count", position="dodge")
```


We can see that college educated respondents have the highest proportion of voter registartion, followed by those with some college. Respondents with high school educations or less have the lowest proportion of voter registration

We can also see from the data above that there is likely a data entry issue with a -1 appearing in place of 1, to be addressed in tidying