

# Homework 15

Stephanie Neul

10/9/2021

1. Load up the file `DsSurveySample.csv`. It should be in the Day 15 Homework folder. Then display the first 6 rows of the data. This is survey data that we looked earlier in the year.

```
ds.survey.data <- read.csv("DsSurveySample.csv")
head(ds.survey.data)
```

```
##   id.Number Heart.Rate zero.tolerance Political What.is.your.favorite.subject.
## 1         1         75             No         2                English
## 2         2         75             No         1                Language
## 3         3         65             No         2                Science
## 4         4         60             No         5                Language
## 5         5         64             Yes         2                 Math
## 6         6         50             Yes         2                English
##   ready.in.the.morning
## 1                    15
## 2                    25
## 3                    25
## 4                    15
## 5                    20
## 6                    20
```

2. We will look at the `Political` vector. While this is currently a numeric vector, it really should be a factor; right now we can't really tell what it shows.

- a. Change the `Political` vector into a factor.

```
ds.survey.data$Political = as.factor(ds.survey.data$Political)
class(ds.survey.data$Political)
```

```
## [1] "factor"
```

- b. Change the names of the levels so that instead of numbers, we have words representing the political positions (from 1 to 5, the numbers represent Very Liberal, Somewhat Liberal, Moderate, Somewhat Conservative, Very Conservative).

```
levels(ds.survey.data$Political) = c("Very Liberal", "Somewhat Liberal", "Moderate", "Somewhat Conservative", "Very Conservative")
levels(ds.survey.data$Political)
```

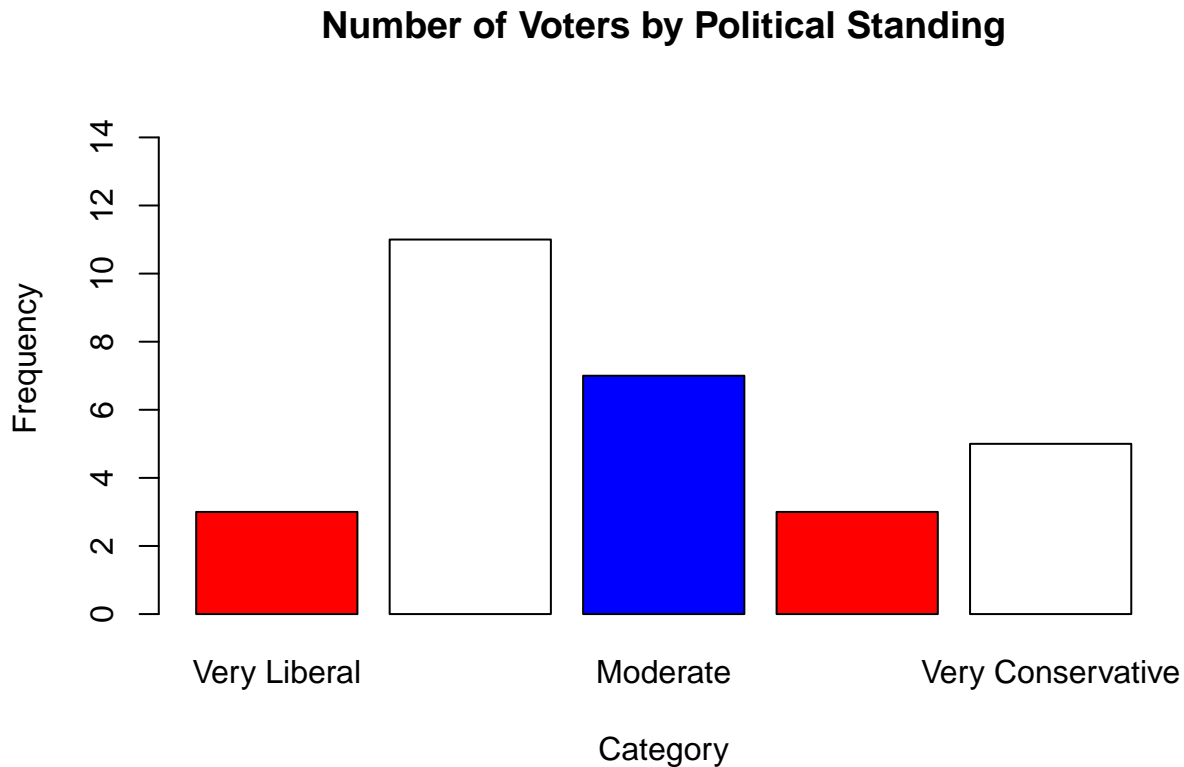
```
## [1] "Very Liberal"          "Somewhat Liberal"      "Moderate"
## [4] "Somewhat Conservative" "Very Conservative"
```

- c. Reorder the levels so that they are in the correct order. Note that the process is the same as when we reordered the Taylor Swift albums in the Day 15 notes.

I am unsure what this question is asking. Are the levels not already in order?

- d. Make a good (labeled, titled, colors) barplot of the `Political` factor. You should be able to see on your barplot that you correctly renamed and ordered the levels.

```
barplot(table(ds.survey.data$Political),
  main = "Number of Voters by Political Standing",
  col = c("red", "white", "blue"),
  ylim = c(0, 15),
  ylab = "Frequency",
  xlab = "Category")
```



2. Run this code to re-import the data so we can work with the numerical Political vector again:

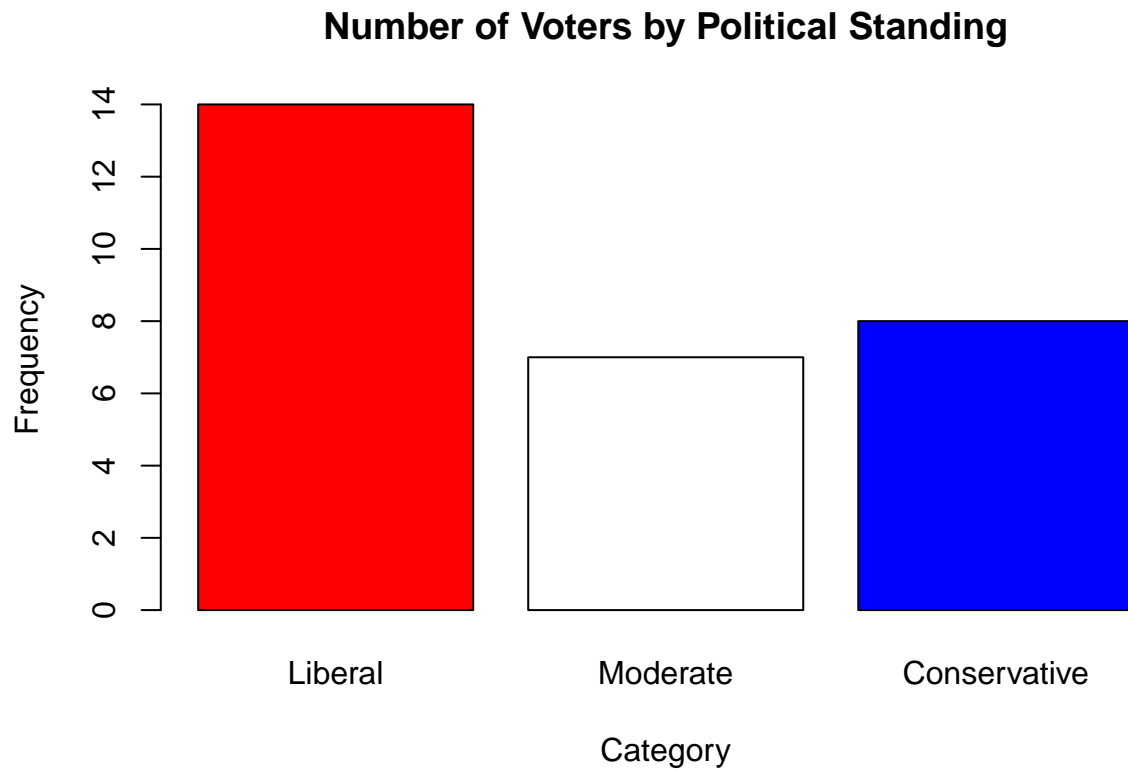
```
ds.survey.data <- read.csv("DsSurveySample.csv")
```

a. If we don't care about the degree to which one is liberal and conservative and just want to group all of liberals together and all of the conservatives together we could combine the 1s with the 2s from the original data and the 4s with the 5s. Then we would only have three groups (Liberals, Moderates, Conservatives). Do this now using the *cut* function on the Political vector. Make sure to assign labels to each of the three levels.

```
breaks.for.groups = c(0, 2, 3, 6)
ds.survey.data$Political = cut(ds.survey.data$Political, breaks = breaks.for.groups, right = TRUE)
levels(ds.survey.data$Political) = c("Liberal", "Moderate", "Conservative")
```

b. Make a barplot of the Political vector showing that you have combined the levels. There should be just three bars if you did this correct.

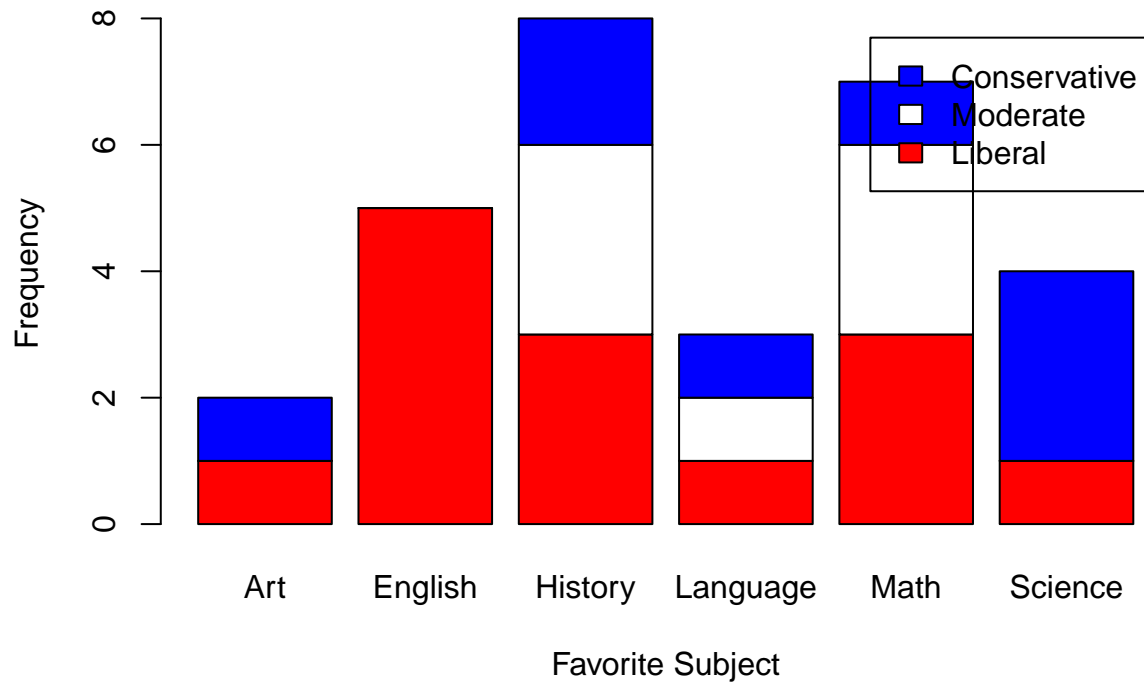
```
barplot(table(ds.survey.data$Political),
  main = "Number of Voters by Political Standing",
  col = c("red", "white", "blue"),
  ylab = "Frequency",
  xlab = "Category")
```



3. Make the following barplots using the `ds.survey.data`: *stacked bar chart using both the `Political` and `What.is.your.favorite.subject.` variables* You should use your `Political` vector with the new combined levels from problem 2b. *It is your choice what variable you put on the x-axis. Choose what makes most sense and looks best to you!* *side by side bar chart using both the `Political` and `What.is.your.favorite.subject.` variables.* *\*segmented bar graph showing the political breakdown within each favorite subject.*

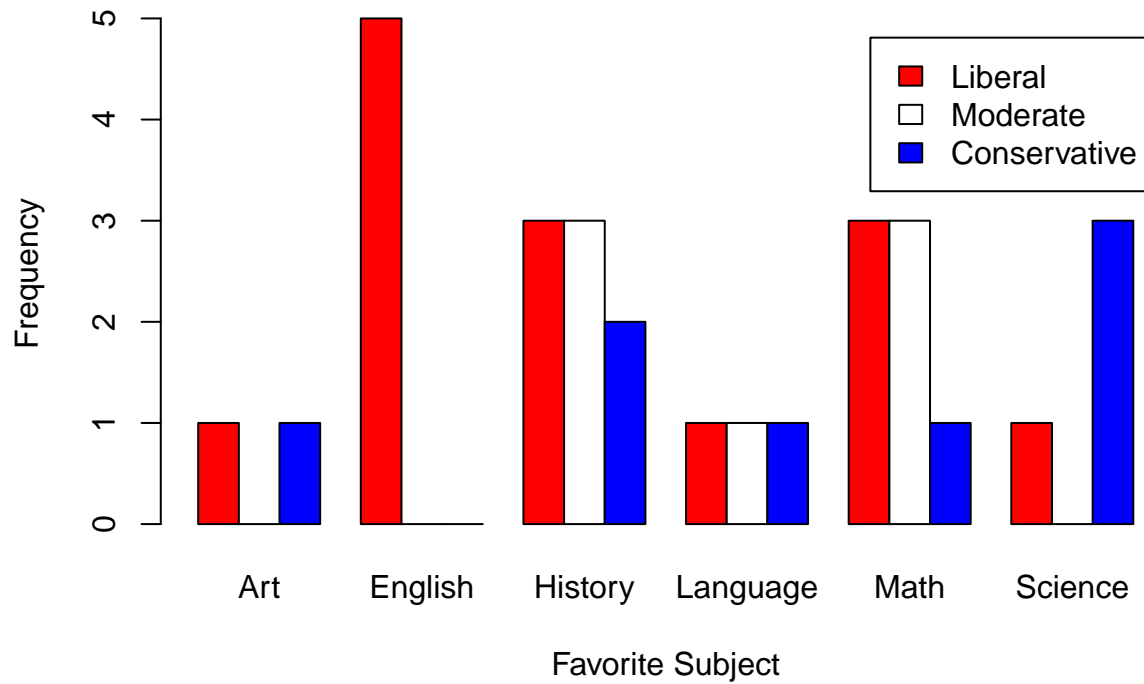
```
barplot(table(ds.survey.data$Political, ds.survey.data$What.is.your.favorite.subject.),
  main = "Types of Voters by Favorite Subject",
  xlab = "Favorite Subject",
  ylab = "Frequency",
  col = c("red", "white", "blue"),
  legend.text = TRUE)
```

## Types of Voters by Favorite Subject



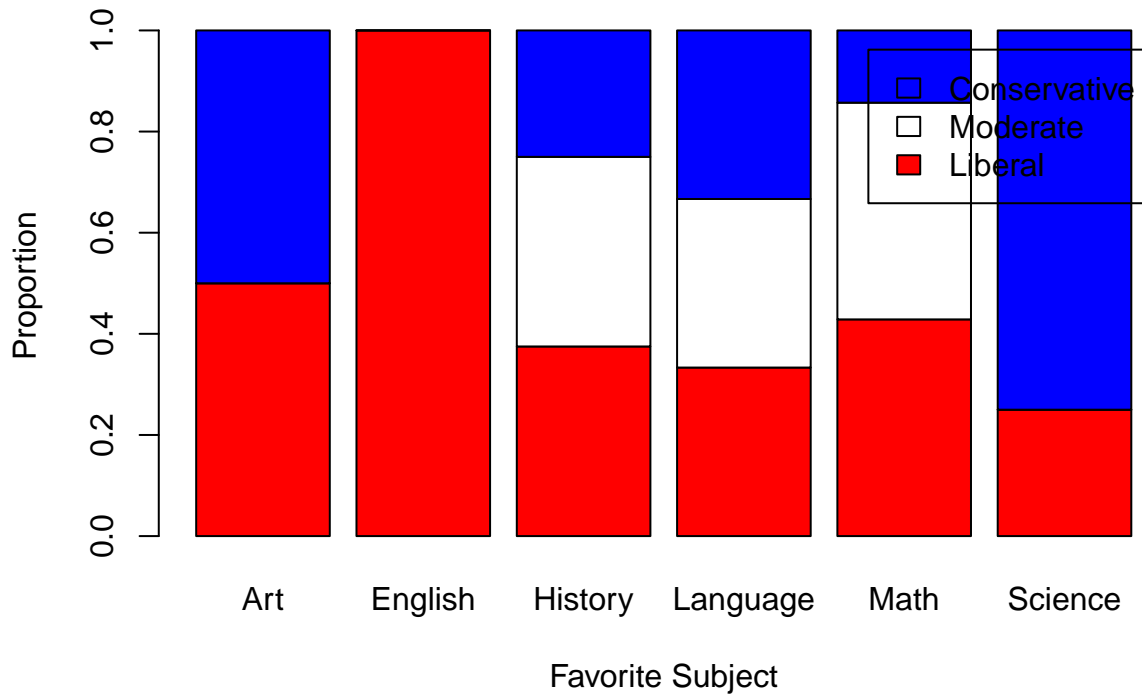
```
barplot(table(ds.survey.data$Political, ds.survey.data$What.is.your.favorite.subject.),
  main = "Types of Voters by Favorite Subject",
  xlab = "Favorite Subject",
  ylab = "Frequency",
  beside = TRUE,
  col = c("red", "white", "blue"),
  legend.text = TRUE)
```

## Types of Voters by Favorite Subject



```
barplot(prop.table(table(ds.survey.data$Political, ds.survey.data$What.is.your.favorite.subject.), 2),
        main = "Types of Voters by Favorite Subject",
        xlab = "Favorite Subject",
        ylab = "Proportion",
        col = c("red", "white", "blue"),
        legend.text = TRUE)
```

**Types of Voters by Favorite Subject**



Each graph should be in its own code chunk and should have labels and a title. After you make the graphs, pick which of the three you think best tells the story of the two variables. Explain why you picked that graph and what it tells you.

I think the third (segmented) graph best tells the story of the two variables. It shows how likely someone who has a certain subject as their favorite will be liberal/moderate/conservative. In other words, we can clearly see those who list art as their favorite subject have a 50/50 chance of being liberal or conservative, with absolutely none being moderate. Because we are using a segmented graph, it is easier to determine these ratios simply by looking.