Week 2: Organizing Data

Example 1

Professor Nick Horton's website contains a data set of student survey result. We use the function read.csv to open the file remotely. The head function shows the first part of the dataframe, and names returns the variable names

```
dat <- read.csv(url("https://nhorton.people.amherst.edu/is5/data/Student_survey.csv"))
head(dat)
```

```
##      Sex Do.you.believe.in.God Pick.Random.Number Height
## 1 Female              Not sure                  6     71
## 2   Male                    No                  2     66
## 3   Male                   Yes                  9     73
## 4 Female                    No                  6     67
## 5   Male                   Yes                  7     71
## 6   Male              Not sure                  9     75
##                          Hand Dates FB.Friends Weight Drinks Varsity Songs
## 1  Predominantly Left Handed     1        314    138      0     Yes  1564
## 2 Predominantly Right Handed     2       1228    130      0      No    97
## 3 Predominantly Right Handed     1       1189    183      0     Yes  1397
## 4 Predominantly Right Handed     1          0    125      0      No  2241
## 5 Predominantly Right Handed     0        709    245      0      No  1299
## 6 Predominantly Right Handed     0       1072    161      0     Yes  1718
##         Diet                 Politics.9Cat Politics.numeric Politics.3Cat
## 1   Omnivore               2. Very Liberal                2       Liberal
## 2 Vegetarian               2. Very Liberal                2       Liberal
## 3  Carnivore     7. Moderatly Conservative                7  Conservative
## 4   Omnivore         3. Moderately Liberal                3       Liberal
## 5   Omnivore 5. Independent/Middle of Road                5      Moderate
## 6 Vegetarian         3. Moderately Liberal                3       Liberal
```

```
names(dat)
```

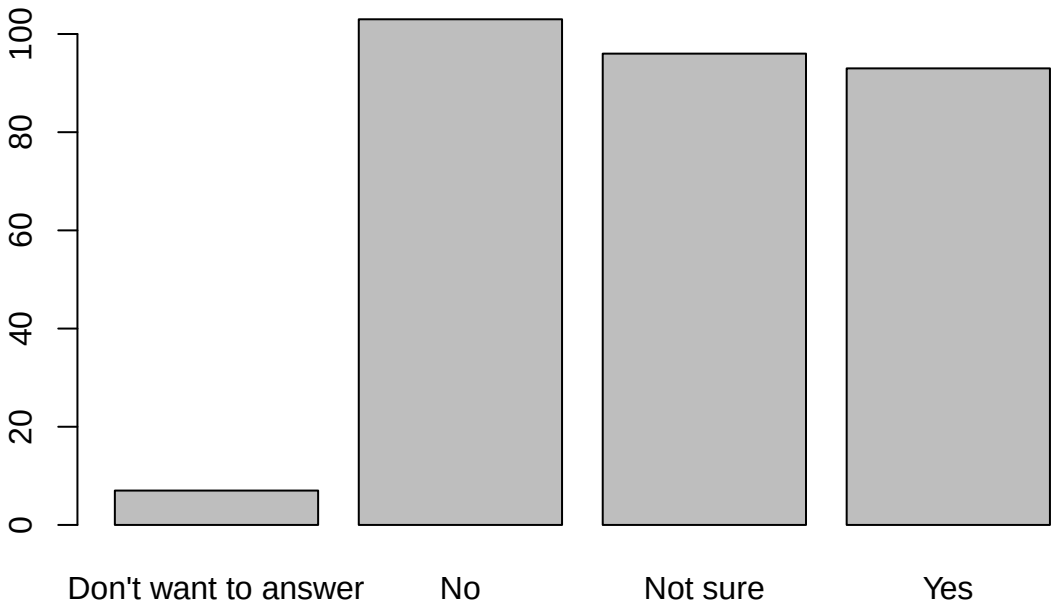
```
##  [1] "Sex"                   "Do.you.believe.in.God" "Pick.Random.Number"
##  [4] "Height"                "Hand"                  "Dates"
##  [7] "FB.Friends"            "Weight"                "Drinks"
## [10] "Varsity"               "Songs"                 "Diet"
## [13] "Politics.9Cat"         "Politics.numeric"      "Politics.3Cat"
```

We use the table function to construct the frequency distribution of the response to the question "Do you believe in God?" in the example below, and the barplot function to build a bar chart.

```
tb1 <- table(dat$Do.you.believe.in.God)
tb1
```

```
##
## Don't want to answer                   No             Not sure
##                    7                  103                   96
##                  Yes
##                   93
```

```
barplot(tb1)
```



A relative frequency table displays percentages or proportions rather than the counts in each category.

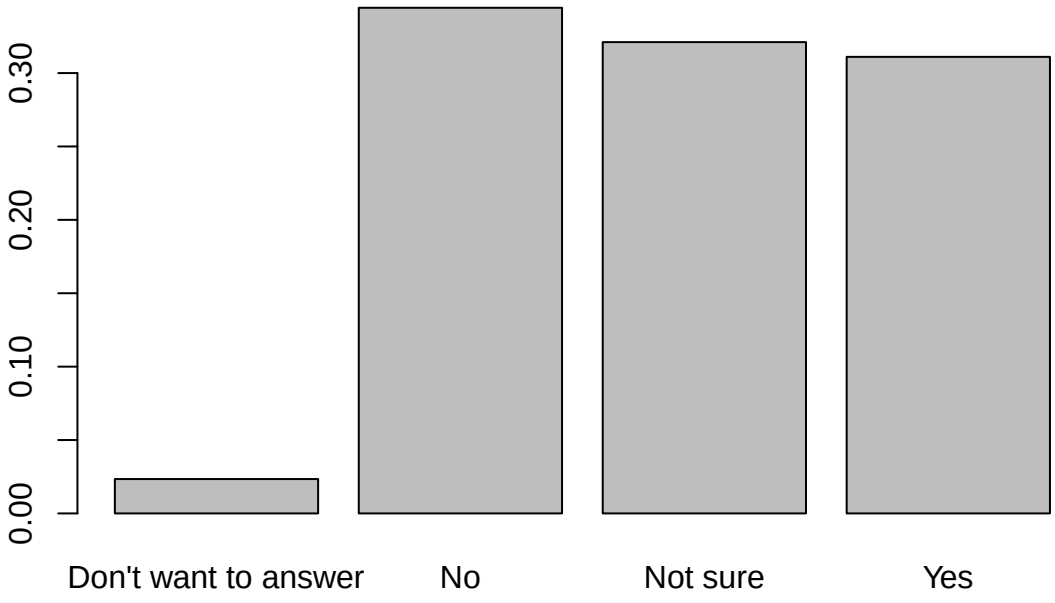
```
n <- sum(tb1)
n
```

```
## [1] 299
```

```
tb2 <- tb1/n
tb2
```

```
##
## Don't want to answer                   No             Not sure
##           0.02341137           0.34448161           0.32107023
##                  Yes
##           0.31103679
```

```
barplot(tb2)
```



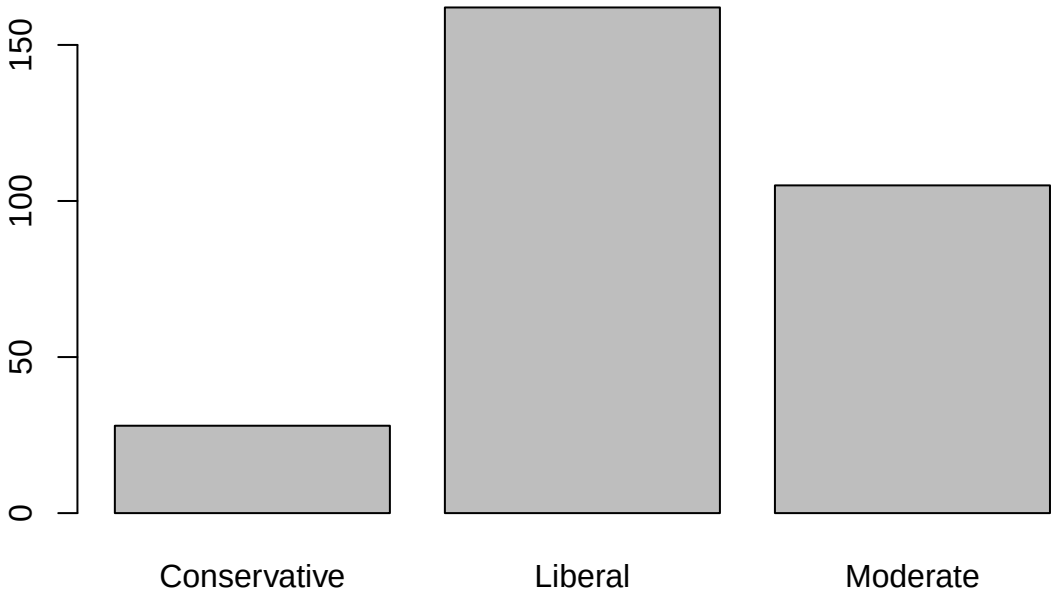
Example 2

We analyze students' self rating of their political inclination below.

```
tb3 <- table(dat$Politics.3Cat)
tb3
```

```
##
## Conservative      Liberal     Moderate
##           28          162          105
```

```
barplot(tb3)
```

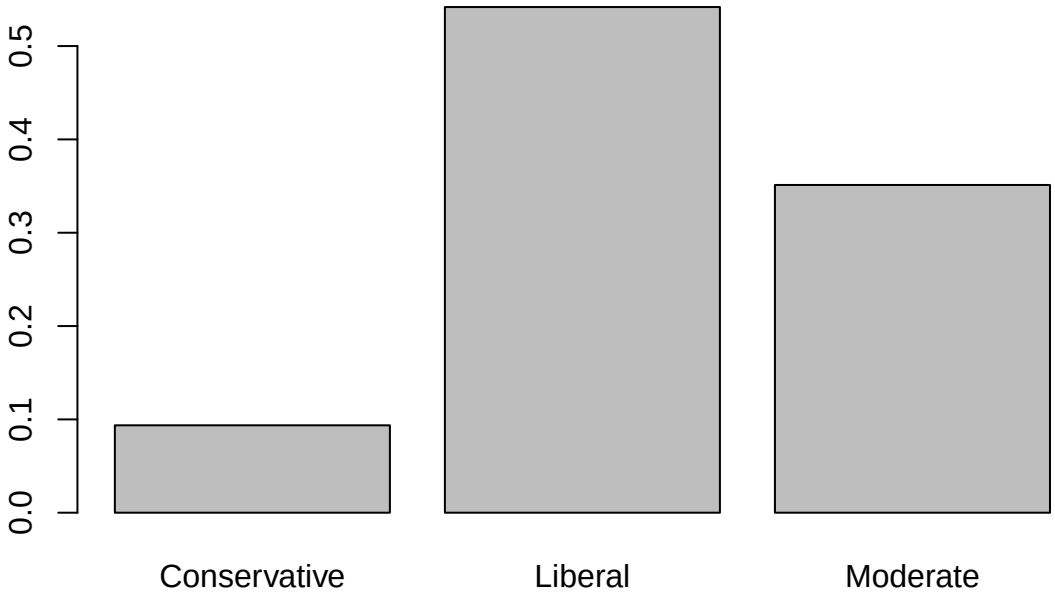



```
tb4 <- tb3/n
tb4
```

```
##
```

```
## Conservative      Liberal     Moderate
##   0.09364548   0.54180602   0.35117057
```

```
barplot(tb4)
```



Follow Up

Use the method demonstrated above to analyze the question on "How would you describe your diet?" The variable name is "Diet".