## Example 4

The first step is to read in the data.

dat<-read.csv("example4.csv")

The next step is to check out the structure

str(dat)

## 'data.frame': 60 obs. of 2 variables:  
## $ major : Factor w/ 2 levels "engineering",..: 1 2 1 2 1 2 1 2 1 2 ...  
## $ gender: Factor w/ 2 levels "female","male": 1 1 1 1 1 1 1 1 2 1 ...

As we can see dat is a data frame with two variables: gender and major.

Now I'll create a table that shows how many people fell into each category.

table(dat$major, dat$gender)

##   
## female male  
## engineering 11 19  
## psychology 18 12

From examining the table, it appears that there might more female psych students and more male engineering statements.

NOTE: after I created the table, I ran the code to be able to interpret the output. I also do this for the statistical analyses. I recommend that you do this as you won't be able to draw the appropriate conclusions or report the right data unless you go back and forth running code and then interpreting the results.

To determine if the data support the appearance of a relationship between gender and major, I performed a chi-square test.

chisq.test(dat$major,dat$gender)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: dat$major and dat$gender  
## X-squared = 2.4027, df = 1, p-value = 0.1211

According to the results of the chi-square test, major and gender are independent of one another Chi-square(1) = 2.4, p = .12.