CSC 315, Fall 2019
Lab #3: Associations

Create a single R script that covers all problems, and contains a heading similar to the following:

```
##########################################################
##  Garrett Dancik
##  CSC 315, Lab #3
##########################################################
```

Your answer to each question should be numbered in a comment in your R script. When your script is complete, create a Notebook and turn in a hardcopy of the Notebook when the assignment is due.

Note: all graphs must be given an appropriate title, x-axis label, and y-axis label, and ggplot must be used for all graphs unless specified otherwise

1. Construct the following contingency table in *R*, along with a table showing the appropriate conditional proportions, where *Income* is the explanatory variable and *Happiness* is the response variable. Answer the question: does there appear to be a relationship between Income and happiness? Why or why not? Note: Your table should include the values only with row names of "Above average", "Average", and "Below average", and column names of "Not Too Happy", "Pretty Happy", and "Very Happy". There is no way to format the R table to include the labels in bold (Happiness and Income).

	Happiness		
Income	Not Too Happy	Pretty Happy	Very Happy
Above average	26	233	164
Average	117	473	293
Below average	172	383	132

2. Import our class survey data, which is available here: https://gdancik.github.io/CSC-315/data/datasets/CSC-315_survey.csv

3. Construct a stacked bar graph that shows whether there is an association between whether someone heard Yanny or Laurel and whether their preference was to fight 1 horse-sized duck or 100 duck-sized horses. In our class, does there appear to be an association between these two variables? If so, describe the association.

4. The code below generates a contingency table for the *YannyOrLaurel* and *Fight* variables (this code assumes that your survey is called *survey*). Run this code (change *survey* to the name of your table if it is different).

```
t <- table(survey$Fight, survey$YannyOrLaurel)
```

   (a) Convert *t* to a table of conditional proportions, conditional on whether a student prefers to fight 1 horse-sized duck or 100 duck-sized horses. What percent of those who heard Yanny prefer to fight 1 horse-sized duck?

   (b) Convert *t* to a table of conditional proportions, conditional on whether a student heard Yanny or Laurel. What percent of those who prefer to fight 1 horse-sized duck heard Yanny?

5. Construct a scatterplot of HS GPA vs. College GPA, so that College GPA would be predicted from HS GPA, and add the regression line from the corresponding linear model.

6. Calculate the correlation and describe the association between HS and College GPA.

7. Repeat (5), but color code the points by Gender and add separate regression lines for Females and Males. This is accomplished by adding `color = Gender` to the *ggplot* aesthetics, adding `se = FALSE` (to remove the confidence intervals), and `fullrange = TRUE` to the geom_smooth layer. Setting *fullrange* says to extend the regression line across the entire plot, rather covering only the data values. Answer the questions below based on the graph.

   a. How do the slopes of the lines compare to each other, and what does this indicate?

   b. How do the y-intercepts of the lines compare to each other, and what does this indicate?

8. The *mtcars* dataset contains data on 32 cars extracted from the 1974 *Motor Trend US* magazine. This dataset is available in *R* in the data.frame *mtcars*, and can be viewed using the code below:

```
View(mtcars)
```

The two variables we will examine are *wt*, the weight of the car in thousands of pounds, and *mpg*, the gas mileage in miles per gallon from road tests. Additional information about the dataset can be found by typing *?mtcars* in the *R* console. Construct a scatterplot that predicts gas mileage from the vehicle's weight, and add the corresponding regression line. Describe the relationship between weight and miles per gallon based on these results.

9. Find the linear regression line that predicts miles per gallon from weight. Find and interpret the *y*-intercept. Find and interpret the slope.

10. Based on this set of cars (in 1974), what would you predict the miles per gallon to be for a car that weighed 3000 pounds? What would you predict the miles per gallon to be for a car that weighed 7000 pounds? (Remember that if the prediction would be an extrapolation, you should say so and not make this prediction).