CSC 315, Fall 2016
Lab #3: Associations

You should create a single R script that covers all problems, and your script should have 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 a hardcopy of the Notebook in when the assignment is due.

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. Does there appear to be a relationship between Income and happiness?

	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: http://pastebin.com/raw/QDSga7qF

3. Construct a contingency table, along with a table showing the appropriate conditional proportions, that would answer the question: *In our class, is there an association between whether someone is a dog person or a cat person, and whether they prefer to fight "100 horse-sized ducks" or "1 duck-sized horse"?* Construct a stacked bar graph to help answer this question. Make sure the bars and y-axis are labeled appropriately, and your chart has an appropriate title and legend. Set the argument

   args.legend = list(x=.95, y = .8, cex = .7)

   where *x* is the *x*-value of the right border of the legend, *y* is the y-value of the top of the legend, and *cex* is a scaling factor for the legend text. The above settings should produce a nice looking legend in your Notebook. Note: do not

worry about the box that surrounds the legend text; it will be formatted correctly when you create a notebook.

4. Based on your analysis, what is your answer to the question: *Are cat people more likely to prefer to fight 100 duck-sized horses than 1 horse-sized duck?* Why or why not?

5. Construct a scatterplot of Alcohol Consumption vs. College GPA, and calculate the correlation. Make sure you correctly label the *x*- and *y*-axes and give the scatterplot an appropriate title. Based on these results, how would you describe the association between Alcohol Consumption and College GPA?

6. Construct a scatterplot of HS GPA vs. College GPA, and calculate the correlation. Make sure you correctly label *x*- and *y*-axes and give the scatterplot an appropriate title. Based on these results, how would you describe the association between HS and College GPA?

7. Fit a linear model that predicts College GPA from HS GPA, and add the regression line to the scatterplot. Find and interpret the *y*-intercept. Find and interpret the slope.

8. Predict the college GPA for a student with a HS GPA of 3.0.

9. The *cereal* dataset contains sugar, caloric, and additional information for 77 different cereals. Import the cereal data set, which is available here: http://pastebin.com/raw/0G6DrHyC

10. Construct a scatterplot for sugar and calorie content, and calculate the correlation. Is there a positive or negative association between sugar and calorie content?

11. Find the linear regression line that predicts calorie content from sugar content (which is in grams), and add the regression line to the scatterplot.

12. How many calories would you expect a cereal to have, on average, if it had 10 grams of sugar? How many calories would you expect a cereal to have, on average, if it had 25 grams of sugar? (Remember that if the prediction would be an extrapolation, you should say so and not make this prediction). What is the residual from an observation of 5 grams of sugar and 70 calories?

13. Create two vectors of length 20 that have a correlation of -1, and show this correlation.