Lab handout (Week 3)

Problem 1

Let's investigate the amount of sugar (measured in grams, g), and type (for adults vs children) in breakfast cereals. Use the following R command to import the data.

```
cereal<-read.csv("http://sites.williams.edu/bklingen/files/2015/05/cereal.csv")
```

Answer the problem a) through i). Show your work to receive full credit.

a) Use `head( )` and identify the names of all *four* variables in the data set (*remember, R is case-sensitive*). Also state whether each variable is categorical or quantitative. You may use `str()` to study type of variable.

b) Use `table()` to create a summary table of `Type` (C for children / A for Adult's cereal). Identify the number of children's cereals and adult's cereals in the data set.

c) Use `boxplot(cereal$Sugar)` to construct a boxplot of `Sugar`. Describe the overall shape of the distribution.
Use `main=" "` option to add a title and `ylab=" "` to add a vertical axis title.

d) Use `summary()` command to find the five number summary of sugar amount.

e) Find the mean and standard deviation of sugar content, and interpret.

f) For a variable that is not extremely skewed / with no extreme outliers, it is known that almost all observations fall within 3 standard deviations from the mean.
Based on the mean and standard deviation from the previous question, what is the interval that almost all sugar contents of breakfast cereal fall within?

g) Use the following R command to construct a histogram for the sugar amount with appropriate titles. How many cereals have a sugar contend more than 15g?

```
hist(cereal$Sugar, main="add your title", xlab="add your axis title")
```

h) Now construct a histogram of sugar content *with 10 breaks*. Describe the overall shape of the distribution (such as unimodal, bimodal, uniform, symmetric, or right-skewed/left-skewed).

i) Construct a side-by-side boxplot of sugar amount by type. Use the R command below. Describe the shape of distribution for adult cereal (Type 'A')and children (Type 'C') cereal.

```
boxplot(cereal$Sugar ~ cereal$Type)
```

Problem 2

For each of the following variables, what would you think the likely shape of the distribution is?

a. Salary of employees of a large company

b. Heights of adult male

Problem 3

Identify each of the following variables as categorical or quantitative. If quantitative, continuous or discrete.

a. How to commute to work (drive, bus, walk, bike, et.)

b. The number of people in line at a box office to purchase theater tickets

c. The weight of a dog

d. Area code from your cell phone number