Let's investigate the amount of sugar and salt (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 e). Show your work to receive full credit.

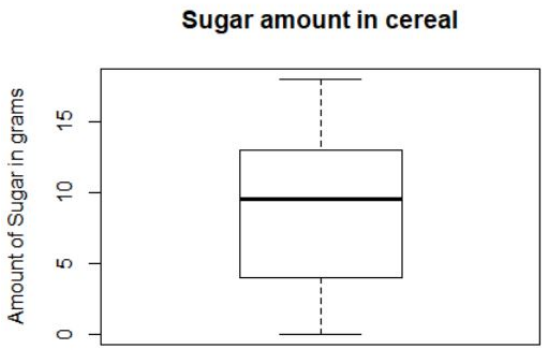
a) Use `head( )` and identify the names of all *four* variables in the data set (*remember, R is case-sensitive*). Also state whether each variable is categorical or quantitative. You may use `str()` to study type of variable.

Name of the four variables are: Cereal (categorical), Sodium (quantitative), Sugar (quantitative), Type (categorical).

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.

```
> table(cereal$Type)
 A  C
10 10
```

c) Construct a boxplot of `Sugar`. Describe the overall shape of the distribution. Use R to find the five number summary of sugar amount.



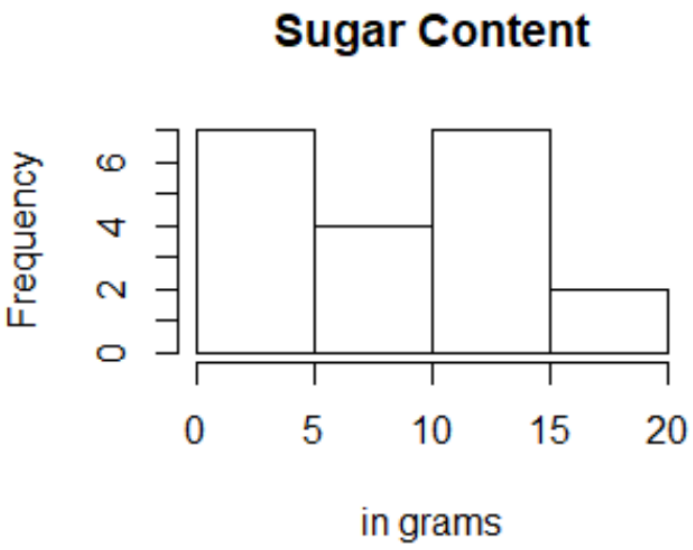
Based on the boxplot, it seems that overall distribution is relatively uniform. For the middle 50%, the distribution is slightly left-skewed.

```
> summary(cereal$Sugar)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    4.00    9.50    8.75   12.50   18.00
```

Five summary numbers are :

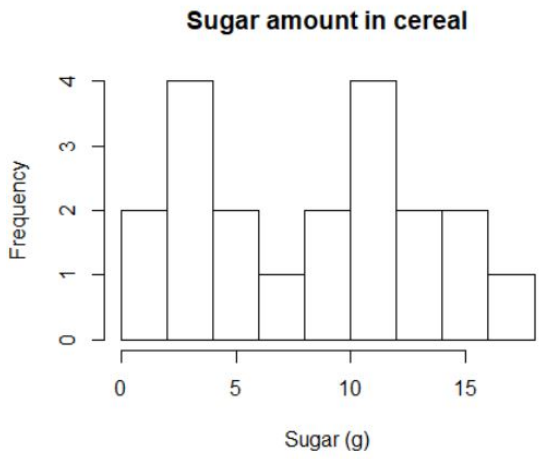
Min: 0, Q1: 4, Median (Q2): 9.5, Q3: 12.50, Max: 18.

d) Construct a histogram for the sugar amount with appropriate titles. How many cereals have a sugar contend more than 15g?



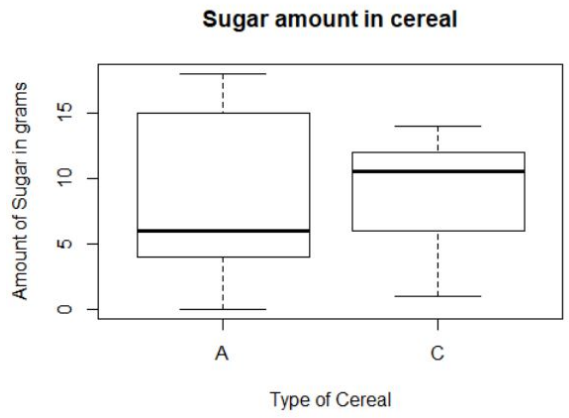
2 cereals have a sugar contend more than 15g.

e) 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).



Distribution is bimodal. Relatively symmetric.

f) Construct a side-by-side boxplot of sugar amount by type. Include the screen shot of the plot in your homework submission. Discribe the shape of distribution for adult cereal (Type 'A')and children (Type 'C') cereal.



Children's cereals tend to have more sugar (median around 11 grams) than adults' (median around 6 grams). Both distributions are skewed. Adult cereals have right-skewed distribution and children cereals have left-skewed distribution.

- Textbook Problem 2.29 Median versus mean. For each of the following variables, would you use the mean or median for describing the center of the distribution? Why? (Think about the likely shape of

the distribution.)

a. Salary of employees of a large university

b. Time spent on a difficult exam

c. Scores on a standardized test

a. Median (The distribution would be right-skewed)
b. Median (The distribution would be left-skewed as many students would spend the whole exam time period whereas only a few might finish/give up early.)
c. Mean (The distribution would be symmetric.)