Andreas Landgrebe
Laboratory Assignment # 2: Procedural Programming and File Processing System
3. Answers to all of the questions posed, with supporting output and evidence as appropriate.

1. What are the names of the attributes in the data set?
   The names of the attributes in the data set are season, size, speed, $mxPH$ (maximum PH level), MnO_2 (Manganese dioxide), Cl (Chlorine), NO_3 (Nitrate), NH_4 (Ammonium), oPO_4, PO_4 (Phosphate), and $Chla$ (Chlorophyll). Then there are concentration levels of 7 harmful algae.

2. How many attributes are in the data set?
   There are 11 variables in the data data and concentration level of 7 harmful algae.

3. How many rows are in the data set?
   There are 18 rows in the data set.

4. Using the mean and the median values of the frequencies for algae a1, is this algae more likely to bloom in rivers classified as "small", "medium", or "large"?
   From using the mean and the median values below:
   > tapply(a1,size,mean)
   large medium small
   11.35333 11.26786 27.14507
   > tapply(a1,size,median)
   large medium small
   2.40 3.65 19.40
   The a1 algae is more likely to bloom in small rivers

5. Using the mean and median values of the frequencies for algae a2 and a3, are these algae more likely to bloom in "small", "medium", or "large" rivers?
   From using the mean and median values below:
   > tapply(a2,size,mean)
   large medium small
   10.137778 7.861905 5.283099
   > tapply(a2,size,median)
   large medium small
   9.70 2.75 0.00
   > tapply(a3,size,mean)
   large medium small

3.722222 5.536905 3.229577
> tapply(a3,size,median)
large medium small
1.8 2.0 1.0
The a2 and a3 are more likely to bloom in medium size rivers.

6. Using the mean and median values of the frequencies for algae a2 and a3,
   are these algae more likely to bloom in "low", "medium", or "high" speed?
   Using the mean and median values below:

> tapply(a1,list(size,speed), mean) high   low   medium
large 23.62857 5.80 11.757143
medium 14.28235 11.32 8.317143
small 32.41860 35.50 18.437037

> tapply(a1,list(size,speed), median)
high   low   medium
large 18.10 1.4 2.5
medium 5.05 6.5 2.4
small 29.70 35.5 5.3

> tapply(a2,list(size,speed), mean)
high   low   medium
large 5.500000 13.188235 9.214286
medium 5.500000 8.593333 9.842857
small 3.127907 0.000000 8.911111

> tapply(a2,list(size,speed), median)
high   low   medium
large 1.7 12.0 7.0
medium 1.5 4.1 4.0
small 0.0 0.0 2.9

> tapply(a3,list(size,speed), mean)
high   low   medium
large 6.185714 1.035294 5.076190
medium 5.597059 5.606667 5.448571
small 3.802326 0.000000 2.437037

> tapply(a3,list(size,speed), median)
high   low   medium
large 4.00 0 3.7
medium 1.55 2 2.1
small 1.20 0 1.0
The a2 and a3 algae are more likely to bloom at high river speeds.

7. What is the relationship between the mean value of chlorophyll and the amount of "a1", "a2", and "a3" in the rivers? For instance, if a river has a "high" amount of chlorophyll, does this mean that it will contain a "high" or a "low" amount of each algae? Why do you think this is the case?

   The relationship between the mean of value of cl and the amount "a1", "a2", and "a3" in the rivers vary depending on the different algae. However, there are some trend that seemed to stay consistent throughout this experiment. If the chlorophyll level in the river was high, then the amount of "a1", "a2", and "a3" varied greatly in each river for this experiment. I believe this is the case because there are several contributing factors to consider when looking at a specific river. The first aspect to look at is the season. The season can make a different on the amount of algae in the river. Some algae may like cold weather so they will arise in the winter and some like the warm weather so they might arise in the summer. Another contributing factor that will arise is the size of the river. Each river that was examined was either small, medium or large. Each harmful algae may like large environment and some are small. The last contributing factor is the speed of the current in the river. Some algae would like the current to be slow and calm and some may like it to be quick. As in the classes that we have in Computer Science at Alden Hall, the answer to big questions is it depends.