Lab Worksheet 1

This worksheet uses the `iris` data set available in R. This data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica:*

```
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Question 1: (3 pts) Is there a difference in sepal length between species setosa and species virginica? Perform a t test and discuss your results (1-2 sentences).

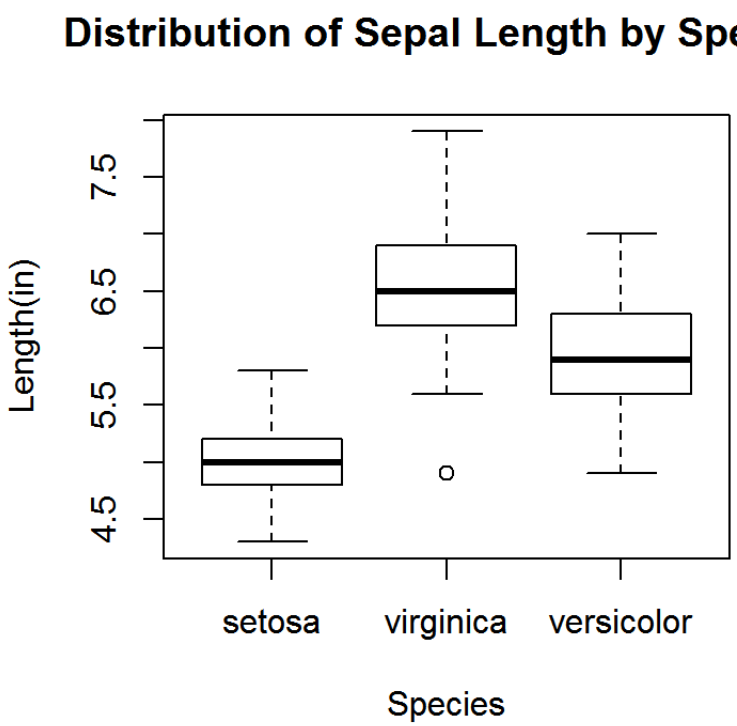
```
# R code goes here
setosaSpecies <- iris[iris$Species=="setosa",]
virginicaSpecies <- iris[iris$Species =="virginica",]
versicolorSpecies <- iris[iris$Species =="versicolor",]
t.test(setosaSpecies$Sepal.Length, virginicaSpecies$Sepal.Length)
```

```
##
##  Welch Two Sample t-test
##
## data:  setosaSpecies$Sepal.Length and virginicaSpecies$Sepal.Length
## t = -15.3862, df = 76.516, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.78676 -1.37724
## sample estimates:
## mean of x mean of y
##     5.006     6.588
```

Discussion goes here. # The t test shows that there is a p-value far less than 0.01 meaning there is substantial evidence against the null hypothesis. Therefore the differences in mean sepal length between the two species is more likely to be zero than statistically different

Question 2: (3 pts) Make side-by-side box plots of sepal length for the three species. Discuss what patterns you observe (1-2 sentences).

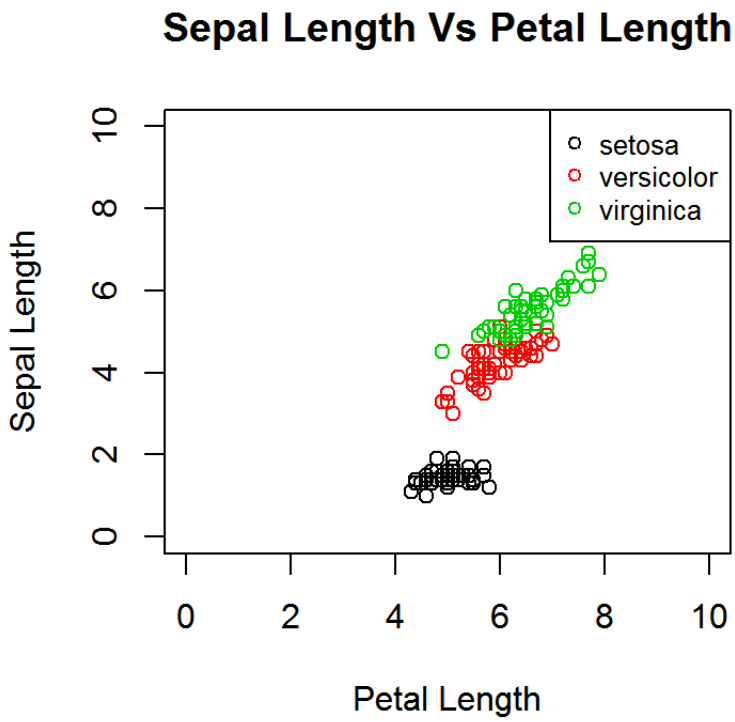
```
# R code goes here.
boxplot(setosaSpecies$Sepal.Length, virginicaSpecies$Sepal.Length,versicolorSpecies$Sepal.Length,name
s=c("setosa","virginica","versicolor"),main="Distribution of Sepal Length by Species",xlab = "Specie
s",ylab="Length(in)")
```



Discussion goes here. # The boxplots show that typically virginica has longer Sepal length distribution. Furthermore, this is the only species that contains an outlier in the data.Setosa has the smallest length distribution, and versicolor is in the middle.

Question 3: (4 pts) Make a scatter plot of sepal length vs. petal length for the three species. Make a single plot that shows the data for all three species at once, in different colors. Hint: To see all data in one plot, you will have two manually set the plot limits, using the `xlim` and `ylim` parameters of the `plot` function. Discuss your results (1-2 sentences).

```
# R code goes here.
plot(iris$Sepal.Length, iris$Petal.Length, col = iris$Species, xlab = "Petal Length", ylab = "Sepal L
ength",main = "Sepal Length Vs Petal Length",xlim=c(0,10),ylim=c(0,10))
legend('topright', legend = levels(iris$Species), col = 1:3, cex = 0.8, pch = 1)
```



Discussion goes here. #Results show Virginica has the highest sepal length to petal length ratio, however versicolors ratio is very similar. Setosa falls farthest behind with smallest ratio of Sepal to Petal Length