

# Lab 2

*Christine Chong cc4190*

*May 25, 2017*

## Instructions

Make sure that you upload an RMarkdown file to the Canvas page (this should have a .Rmd extension) as well as the PDF output after you have knitted the file (this will have a .pdf extension). Note that since you have already knitted this file, you should see both a **Lab1\_UNI.pdf** and a **Lab1\_UNI.Rmd** file in your UN4206 folder. Click on the **Files** tab to the right to see this. The files you upload to the Canvas page should be updated with commands you provide to answer each of the questions below. You can edit this file directly to produce your final solutions.

## Background: Edgar Anderson's Iris Data

The R data description follows:

This famous (Fisher's or Anderson's) iris 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.

## Goal

The purpose of this lab is to preform simple filtering tasks, use the function **tapply()** and successfully write a loop.

## Tasks

- 1) Run the following code and briefly explain what the two functions are doing.

The **head** function returns the first parts of the vector iris. The **names** function returns the names in the vector iris.

```
## head function
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
```

```
## names function
names(iris)
```

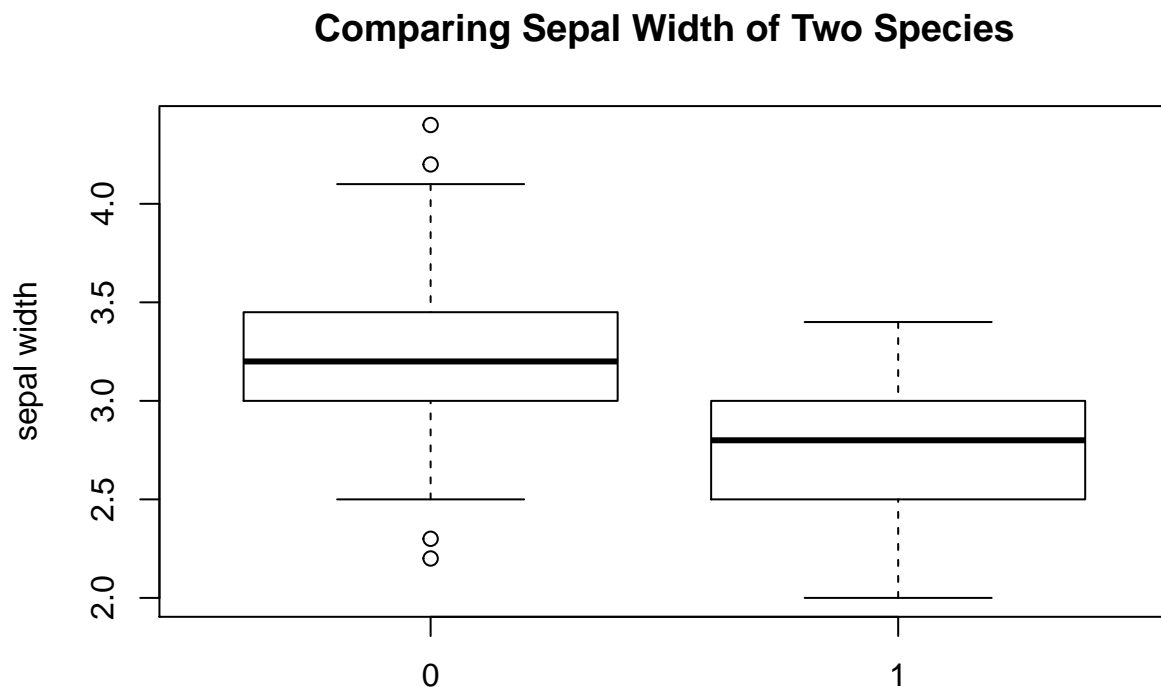
```
## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"
## [5] "Species"
```

- 2) In one line of code, create a vector named **Versicolor** which contains a 1 if the species is versicolor and contains a 0 otherwise. Use the function **ifelse()** to accomplish this task.

```
Versicolor <- ifelse(iris$Species == "versicolor", 1, 0)
```

- 3) Based on the vector **Versicolor** defined in Part 2, construct a comparative boxplot of sepal width split by whether or not the species is versicolor. The code is given below. Change the main title and y-label of the plot.

```
# Don't forget to uncomment the code below
boxplot(Sepal.Width~Versicolor,main="Comparing Sepal Width of Two Species",ylab="sepal width",data=iris)
```



- 4) Using filtering techniques and the function **mean()**, compute the mean sepal length of only the setosa species. Write only one line of code to accomplish this task.

```
mean(iris$Sepal.Length[iris$Species == "setosa"])
```

```
## [1] 5.006
```

- 5) Run the following code and briefly explain what the function **tapply()** is accomplishing. Here the **tapply()** function is finding means of the sepal lengths of each of species of flowers, which are setosa, versicolor and virginica.

```
tapply(iris$Sepal.Length,iris$Species,mean)
```

```
##      setosa versicolor  virginica
##      5.006      5.936      6.588
```

- 6) Write a loop that computes the mean of each quantitative variable split by each iris species. Store the computed means in a matrix named **MeanFlowers**. This matrix should have dimension 4 by 3. Also display the matrix after you run the loop. I started the loop below:

```
# define a matrix of zeros
MeanFlowers <- matrix(0,nrow=4,ncol=3)

# define a character vector corresponding to the numeric variable names
measurements <- c("Sepal.Length","Sepal.Width","Petal.Length","Petal.Width")

# name the rows and columns of the matrix MeanFlowers
rownames(MeanFlowers) <- measurements
colnames(MeanFlowers) <- c("setosa","versicolor","virginica")

#subsetting flowers
subsetflower1 <- subset(iris, iris$Species == "setosa")
subsetflower2 <- subset(iris, iris$Species == "versicolor")
subsetflower3 <- subset(iris, iris$Species == "virginica")

# Loop
for (j in measurements) {
  flower1<- mean((subsetflower1[, j]))
  flower2<- mean((subsetflower2[, j]))
  flower3<- mean((subsetflower3[, j]))
  vec <- c(flower1, flower2, flower3)
  MeanFlowers[j, ] <- vec
}
MeanFlowers
```

```
##           setosa versicolor virginica
## Sepal.Length  5.006      5.936      6.588
## Sepal.Width   3.428      2.770      2.974
## Petal.Length  1.462      4.260      5.552
## Petal.Width   0.246      1.326      2.026
```