Lab 3

Salim M'jahad msm2243

February 1, 2018

Instructions

Before Lab 3 is due, make sure that you upload an RMarkdown file to the canvas page (this should have a .Rmd extension) as well as the PDF (or HTML) output after you have knitted the file (this will have a .pdf extension or .html). Note that since you have already knitted this file, you should see both a **Lab3__UNI.pdf** and a **Lab3__UNI.Rmd** file in your UN2102 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.

```
## 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"
```

head() shows the first part of the iris dataset

names() returns the names of the variables in our dataset

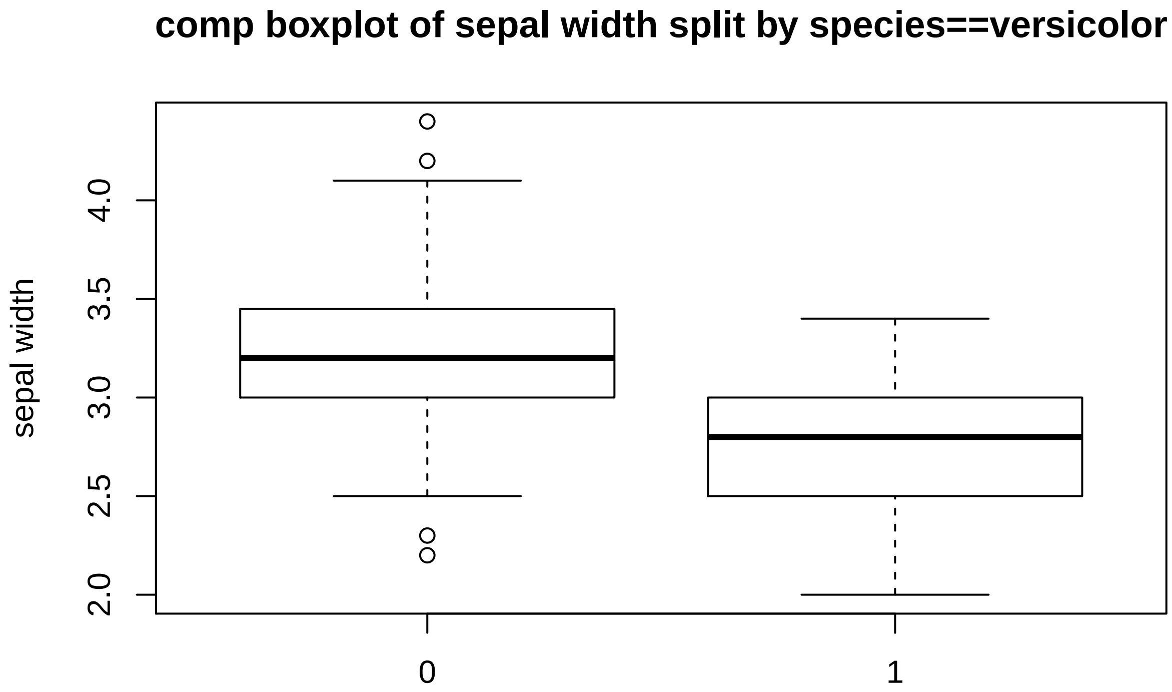
2) In one line of code, create a vector named **Versicolor** which contains a 1 if the species is versicolor and contains a 0 otherwise. Use the function **ifelse()** to accomplish this task.

```
#-- R code goes here ----
Versicolor <- ifelse(iris[,"Species"]=="versicolor", 1, 0)
print(Versicolor)
```

```
##   [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [36] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [71] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0
## [106] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [141] 0 0 0 0 0 0 0 0 0 0
```

3) Based on the vector **Versicolor** defined in Part 2, construct a comparative boxplot of sepal width split by whether of 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="comp boxplot of sepal width split by species==versicolor",ylab="sep
```



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.

```
#-- R code goes here ----
mean(subset(iris, Species=="setosa")$Sepal.Length)
```

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

5) Run the following code and briefly explain what the function **tapply()** is accomplishing.

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

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

This function is separating the Sepal.Length column of iris into the categories present in the Species column. Then, it's applying the function mean on each category seperately.

6) Write a loop that 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")
colnames(MeanFlowers)
```

```
## [1] "setosa"     "versicolor" "virginica"
```

```
# Loop

for (j in measurements) {
  tmp <- tapply(iris[[j]],iris$Species,mean)
  for (i in colnames(MeanFlowers)) {
    MeanFlowers[j, i] <- tmp[i]
  }
}
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
```