Bootcamp R Worksheet

Data Science Dojo

To warm up your R skills for the bootcamp, please follow the instructions below to write the R code.

Navigate your working directory in R to the Datasets folder in the main bootcamp repository.

```r
#setwd("~/bootcamp/Datasets")
getwd()
```

```
## [1] "C:/Users/Brian/SharePoint/Data Science Dojo - Documents/5-Day Bootcamp/Course Work Material/Int
```

Read the iris data set using read.csv.

```r
iris.data <- read.csv("Iris_Data.csv")
```

Display the first few rows of the iris data.

```r
head(iris.data)
```

```
##   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
```

Rename the column name Species to Type.

```r
colnames(iris.data)[5] <- "Type"
```

Display the first 5 rows and last 3 columns of the iris data frame.

```r
iris.data[1:5, 3:5]
```

```
##   Petal.Length Petal.Width   Type
## 1          1.4         0.2 setosa
## 2          1.4         0.2 setosa
## 3          1.3         0.2 setosa
## 4          1.5         0.2 setosa
## 5          1.4         0.2 setosa
```
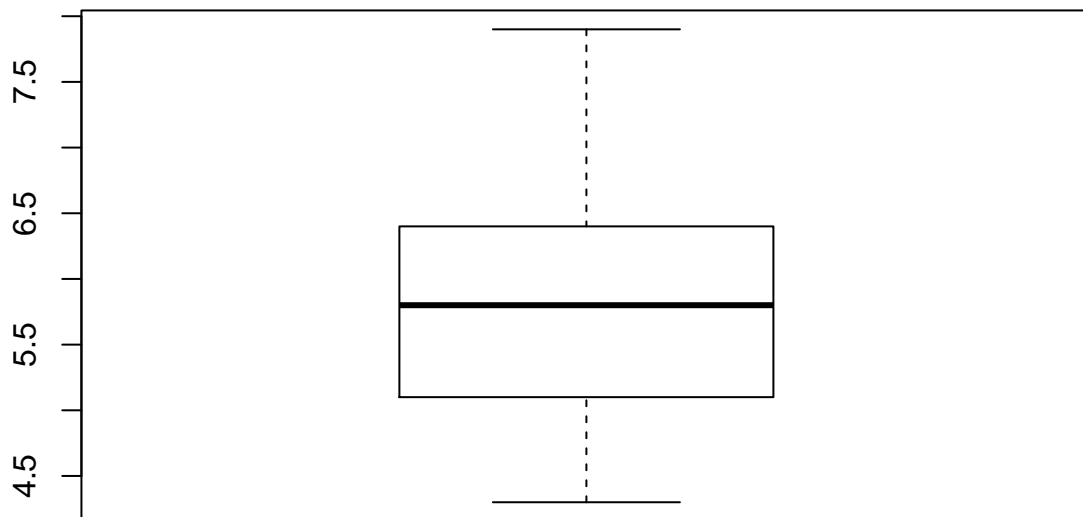
What is the data type of each column in this data frame of iris data?

```
str(iris.data)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Type        : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```
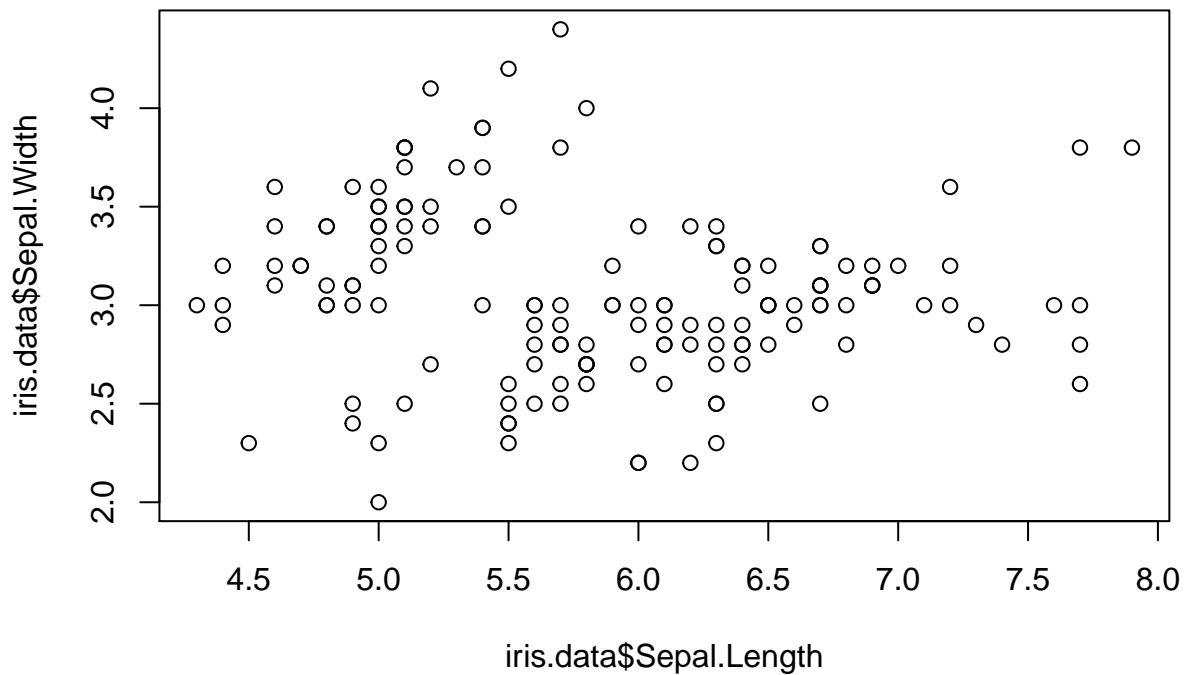
Draw a box plot of Sepal Length

```
boxplot(iris.data$Sepal.Length)
```


Draw a scatter plot of Sepal Length vs Sepal Width.

```
plot(iris.data$Sepal.Length, iris.data$Sepal.Width)
```

Create a new column in the iris data frame which is the sum of Sepal Length and Sepal Width.

```
iris.data[,"Sepal.Sum"] <- iris.data[,"Sepal.Length"] + iris.data[,"Sepal.Width"]
```

What are the means, medians, and standard deviations of the four predictor columns in this data frame?

```
summary(iris.data)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##   Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##   1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##   Median :5.800   Median :3.000   Median :4.350   Median :1.300
##   Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##   3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##   Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##            Type        Sepal.Sum
##   setosa    :50   Min.   : 6.800
##   versicolor:50   1st Qu.: 8.300
##   virginica :50   Median : 8.850
```

```
##                         Mean   : 8.901
##                         3rd Qu.: 9.575
##                         Max.   :11.700
```

```
sapply(as.list(iris.data[,1:4]), sd)
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##    0.8280661    0.4358663    1.7652982    0.7622377
```

Display the pair-wise correlations between the features of the iris data set.

```
pairs(iris.data)
```