## Answer Q1.

iris\_new = read.csv("C:/Temp/input/iris.data")

head(iris\_new)

X5.1 X3.5 X1.4 X0.2 Iris.setosa

1 4.9 3.0 1.4 0.2 Iris-setosa

2 4.7 3.2 1.3 0.2 Iris-setosa

3 4.6 3.1 1.5 0.2 Iris-setosa

4 5.0 3.6 1.4 0.2 Iris-setosa

5 5.4 3.9 1.7 0.4 Iris-setosa

6 4.6 3.4 1.4 0.3 Iris-setosa

colnames(iris\_new) <- c('Sepal.Length', 'Sepal.Width','Petal.Length', 'Petal.Width', 'Species')

head(iris\_new)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1 4.9 3.0 1.4 0.2 Iris-setosa

2 4.7 3.2 1.3 0.2 Iris-setosa

3 4.6 3.1 1.5 0.2 Iris-setosa

4 5.0 3.6 1.4 0.2 Iris-setosa

5 5.4 3.9 1.7 0.4 Iris-setosa

6 4.6 3.4 1.4 0.3 Iris-setosa

mean( head(iris\_new, 5)$Sepal.Length)

[1] 4.92

The mean of first 5 rows is 4.92

install.packages("stringr")

library(stringr)

iris\_new$Species <- str\_replace(iris\_new$Species, 'Iris-', '')

head(iris\_new)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1 4.9 3.0 1.4 0.2 setosa

2 4.7 3.2 1.3 0.2 setosa

3 4.6 3.1 1.5 0.2 setosa

4 5.0 3.6 1.4 0.2 setosa

5 5.4 3.9 1.7 0.4 setosa

6 4.6 3.4 1.4 0.3 setosa

iris\_new$Species <- tools::toTitleCase(iris\_new$Species)

head(iris\_new)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1 4.9 3.0 1.4 0.2 Setosa

2 4.7 3.2 1.3 0.2 Setosa

3 4.6 3.1 1.5 0.2 Setosa

4 5.0 3.6 1.4 0.2 Setosa

5 5.4 3.9 1.7 0.4 Setosa

6 4.6 3.4 1.4 0.3 Setosa

## Answer Q2.

iris\_sub <- iris\_new[ iris\_new$Sepal.Length < 6.4 & iris\_new$Petal.Length > 5.1, ]

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

100 6.3 3.3 6.0 2.5 virginica

103 6.3 2.9 5.6 1.8 virginica

134 6.1 2.6 5.6 1.4 virginica

136 6.3 3.4 5.6 2.4 virginica

148 6.2 3.4 5.4 2.3 virginica

nrow(iris\_sub)

[1] 5

The iris\_sub dataset has 5 observations

mean(iris\_sub$Sepal.Width)

[1] 3.12

mean(iris\_sub$Petal.Length)

[1] 5.64

The mean of sepal.width column in iris\_sub is 3.12 and the mean of petal.length column in the same dataset is 5.64

## Answer Q3.

par(mfrow = c(2, 2))

boxplot( iris\_new$Sepal.Length ~ iris\_new$Species,

xlab = "Species",

xaxt = "n",

col = 2:6,

main = "Sepal.Length",

ylab = "Sepal.Length",

las=2)

axis(1, at = 1:3, labels = c("Setosa", "Versicolor", "Virginica"))

boxplot(iris\_new$Sepal.Width ~ iris\_new$Species ,

xlab = "Species",

xaxt = "n",

col = 2:6,

main = "Sepal.Width",

ylab = "Sepal.Width",

las=2)

axis(1, at = 1:3, labels = c("Setosa", "Versicolor", "Virginica"))

boxplot(iris\_new$Petal.Length ~ iris\_new$Species ,

xlab = "Species",

xaxt = "n",

col = 2:6,

main = "Petal.Length",

ylab = "Petal.Length",

las=2)

axis(1, at = 1:3, labels = c("Setosa", "Versicolor", "Virginica"))

boxplot(iris\_new$Petal.Width ~ iris\_new$Species ,

xlab = "Species",

xaxt = "n",

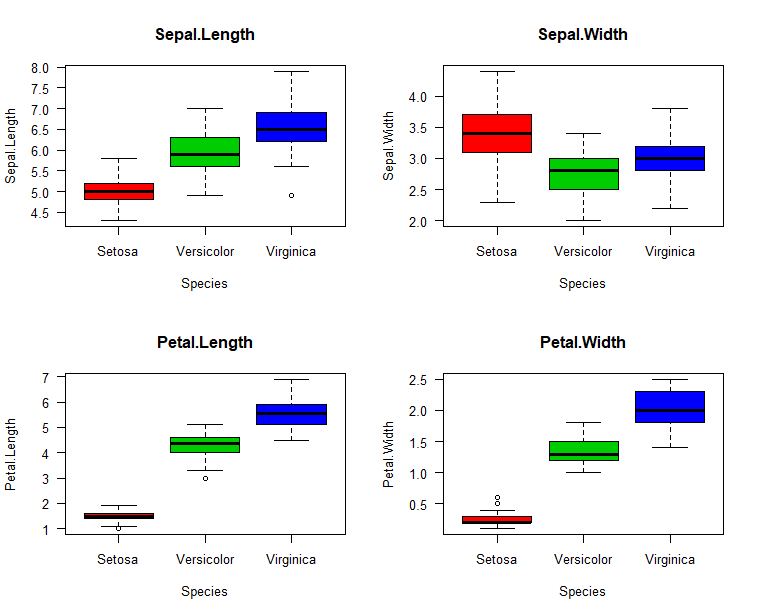
col = 2:6,

main = "Petal.Width",

ylab = "Petal.Width",

las=2)

axis(1, at = 1:3, labels = c("Setosa", "Versicolor", "Virginica"))



We can see one outlier on Sepal.Length for Virginica, one outlier on Petal.Lenght for Versicolor and two outliers on Petal.Width for Setosa.

## Answer Q4.

library(ggplot2)

par(mfrow = c(1, 2))

# Change point shapes and colors

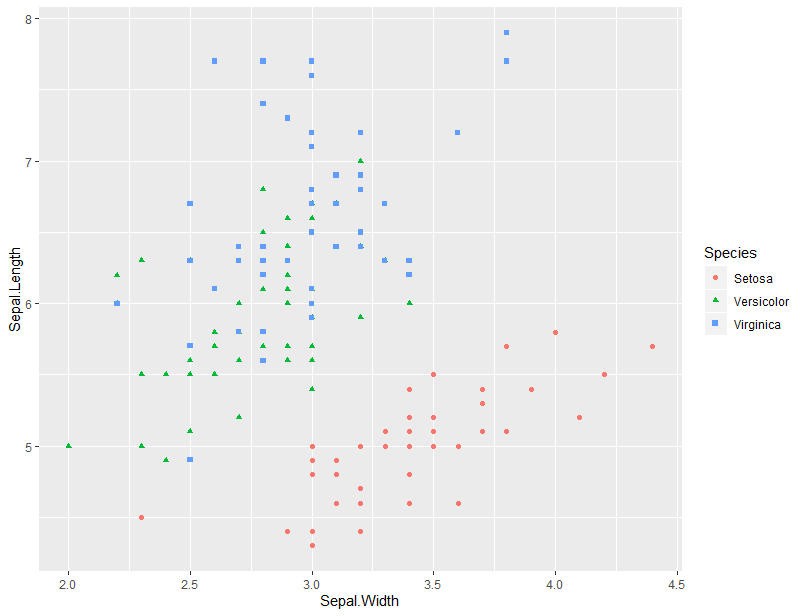
ggplot(iris\_new, aes(x=Sepal.Width, y=Sepal.Length, shape=Species, color=Species)) +

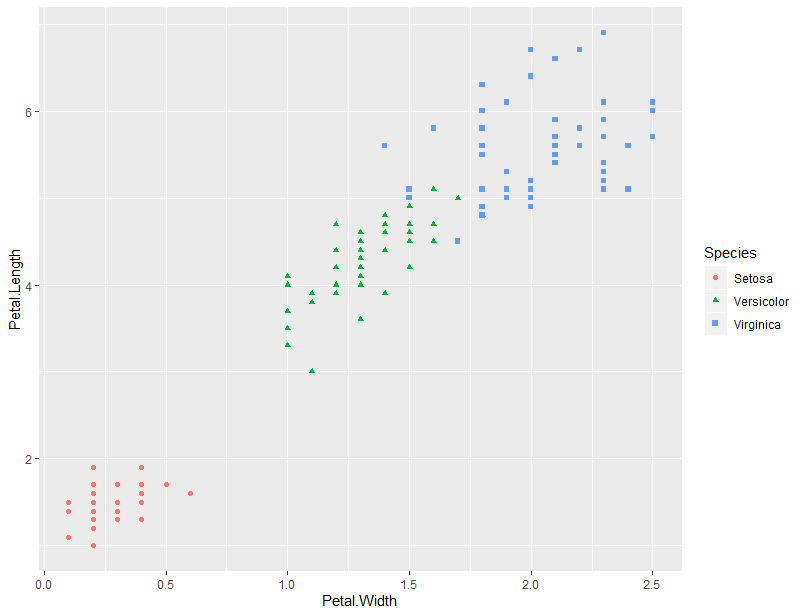
geom\_point()

# Change point shapes and colors

ggplot(iris\_new, aes(x=Petal.Width, y=Petal.Length, shape=Species, color=Species)) +

geom\_point()





## Answer Q5.

iris\_sub <- iris\_new[ , c(1,3,5) ]

head(iris\_sub)

Sepal.Length Petal.Length Species

1 4.9 1.4 setosa

2 4.7 1.3 setosa

3 4.6 1.5 setosa

4 5.0 1.4 setosa

5 5.4 1.7 setosa

6 4.6 1.4 setosa

iris\_sub[ iris\_sub$Sepal.Length == 5.7 & iris\_sub$Petal.Length == 4.1, ]$Species

[1] "Versicolor"