mid-term-q26

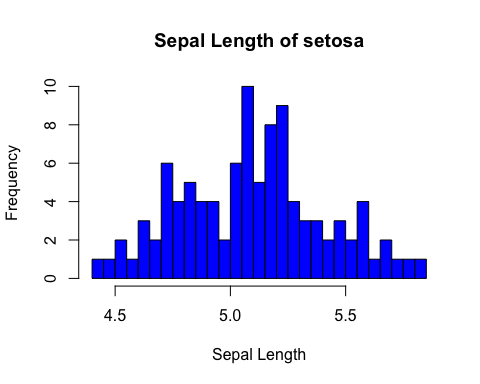
Pradeep Paladugula

2/23/2020

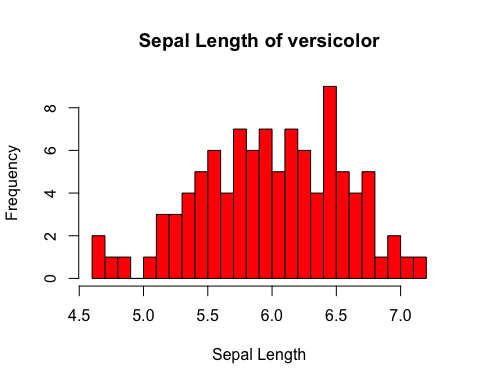
library('ggplot2')  
library('e1071')  
data <- read.csv('iris\_exams.csv', header = TRUE)  
diff\_species = sapply(data, levels)$Species  
diff\_species

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

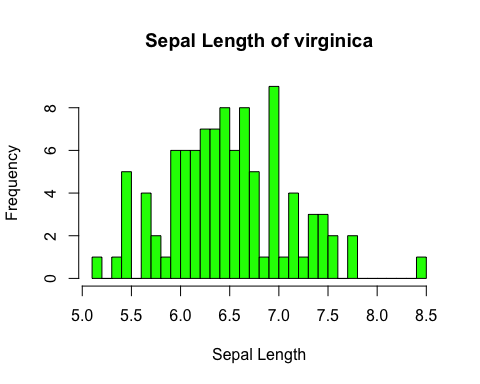
#histSetosa <- ggplot(data, aes(data[data$Species=="setosa",]$Sepal.Length))  
#histSetosa + geom\_histogram(binwidth = 0.05, color = 'pink') + labs(xlab = 'Sepal Length', ylab = 'Frequency', title = 'Sepal Length of setosa')  
hist(data[data$Species=="setosa",]$Sepal.Length, main = 'Sepal Length of setosa', xlab = 'Sepal Length', ylab = 'Frequency', col = 'blue', border = 'black', breaks = 30)



hist(data[data$Species=="versicolor",]$Sepal.Length, main = 'Sepal Length of versicolor', xlab = 'Sepal Length', ylab = 'Frequency', col = 'red', border = 'black', breaks = 30)



hist(data[data$Species=="virginica",]$Sepal.Length, main = 'Sepal Length of virginica', xlab = 'Sepal Length', ylab = 'Frequency', col = 'green', border = 'black', breaks = 30)



setosalength <- data[data$Species=="setosa",]$Sepal.Length  
  
Versicolorlength <- data[data$Species=="versicolor",]$Sepal.Length  
  
Virginicalength <- data[data$Species=="virginica",]$Sepal.Length  
  
Setosa = c(mean = mean(setosalength), sd = sd(setosalength), quartile = quantile(setosalength), min=min(setosalength), max = max(setosalength), skewness = skewness(setosalength), kurtosis = kurtosis(setosalength))  
  
Versicolor = c(mean = mean(Versicolorlength), sd = sd(Versicolorlength), quartile = quantile(Versicolorlength), min=min(Versicolorlength), max = max(Versicolorlength), skewness = skewness(Versicolorlength), kurtosis = kurtosis(Versicolorlength))  
  
Verginica = c(mean = mean(Virginicalength), sd = sd(Virginicalength), quartile = quantile(Virginicalength), min=min(Virginicalength), max = max(Virginicalength), skewness = skewness(Virginicalength), kurtosis = kurtosis(Virginicalength))  
  
specices\_stat = data.frame(Setosa, Versicolor, Verginica)  
specices\_stat

## Setosa Versicolor Verginica  
## mean 5.0938686 5.9860688 6.4913343  
## sd 0.3144154 0.5519908 0.6079287  
## quartile.0% 4.4165896 4.6498671 5.1965548  
## quartile.25% 4.8489891 5.5776984 6.0850477  
## quartile.50% 5.0957080 6.0111028 6.4884913  
## quartile.75% 5.2810494 6.4375376 6.9067087  
## quartile.100% 5.8435807 7.1235462 8.4776301  
## min 4.4165896 4.6498671 5.1965548  
## max 5.8435807 7.1235462 8.4776301  
## skewness 0.1113925 -0.2683246 0.3416658  
## kurtosis -0.5035743 -0.4665129 0.1918683

Interpretation: The Virginica has the highest Sepal Length among three flower species.  
setosa has the less standard devisation, low skewness among the spcises, and from the histogrma graph its seems like having reatively normal distribution. Versicolor seems to have a low negative skewness. Virginica has the highest standard deviation, hight skewness to the right, positive kurtosis and also has the longest sepal length than other two. From the Virginica histogram Graph itseems like has the outliers.