Homework 2

**Malik Hassan Qayyum**

Data 502

Iris dataset

Excel – one pivot table representing each species averages for all four attributes separately

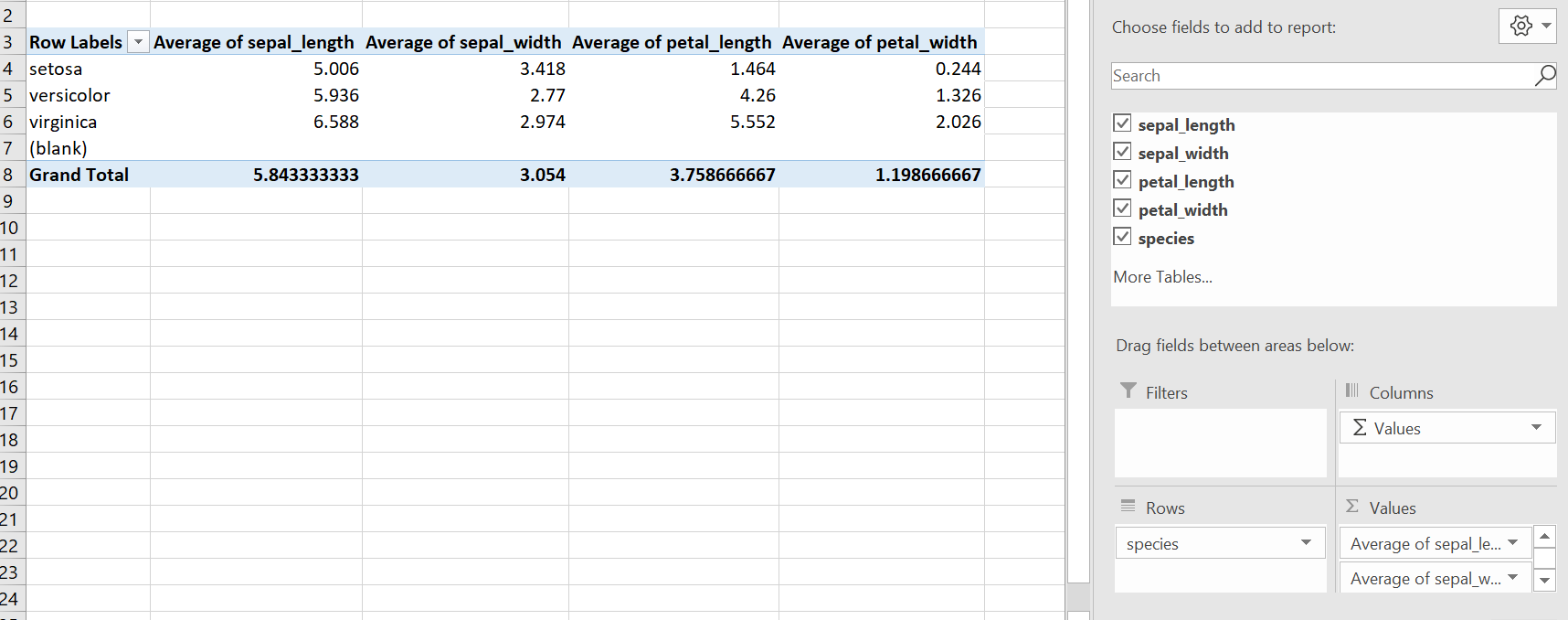
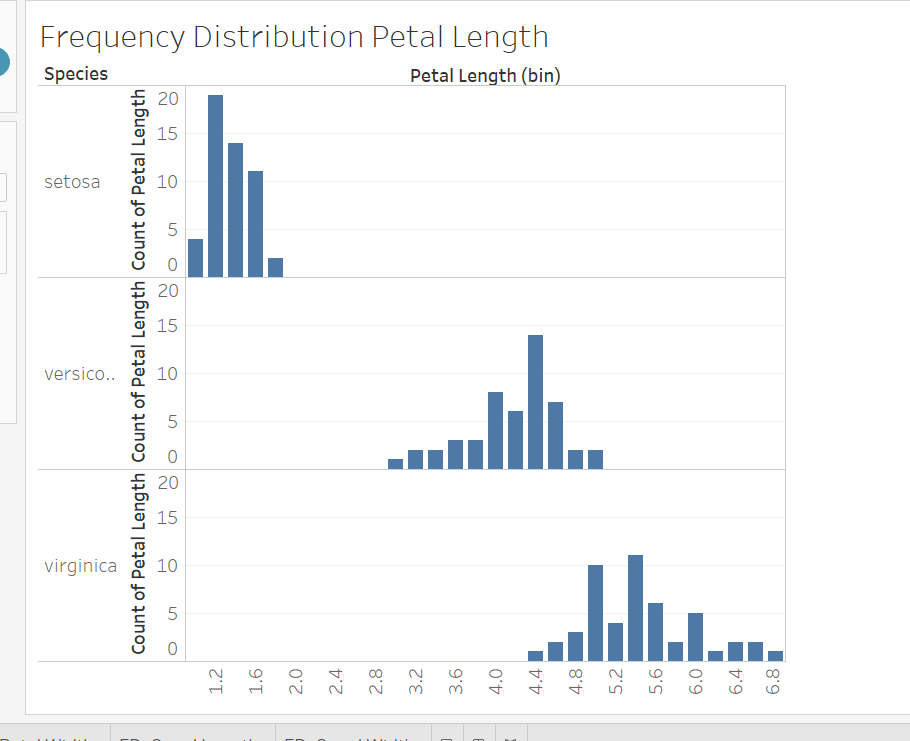
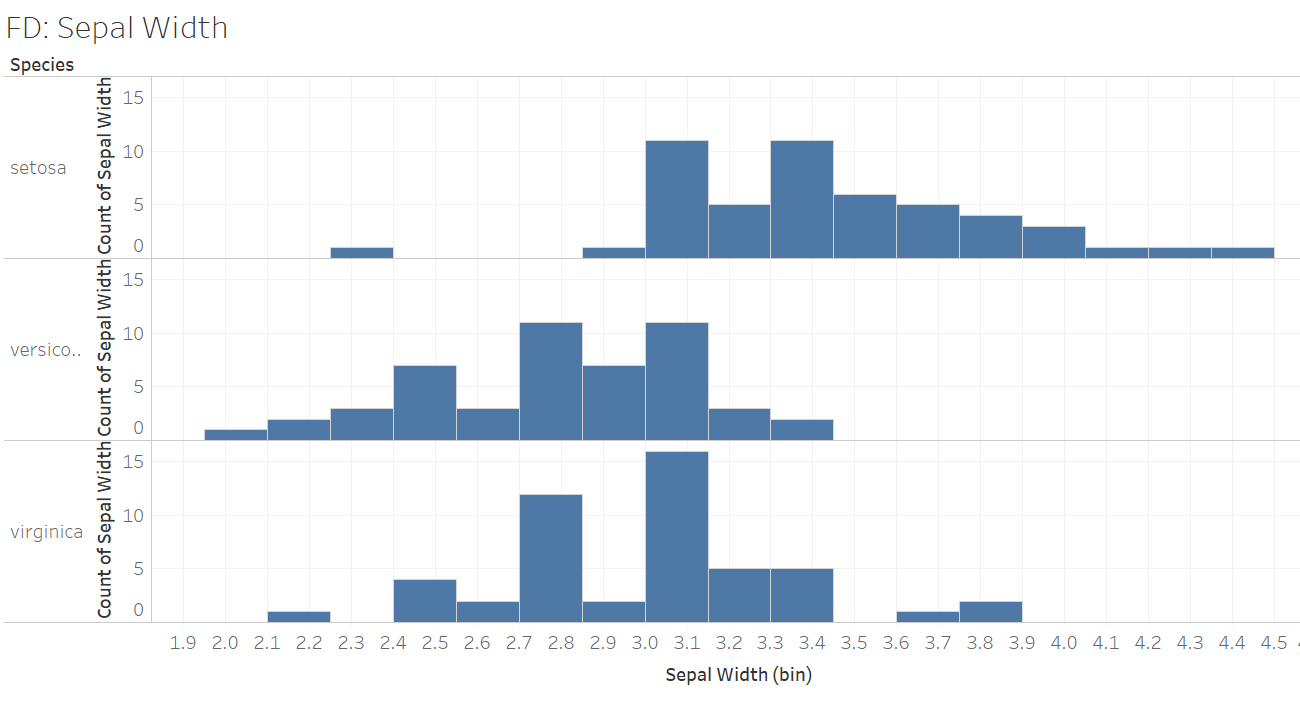
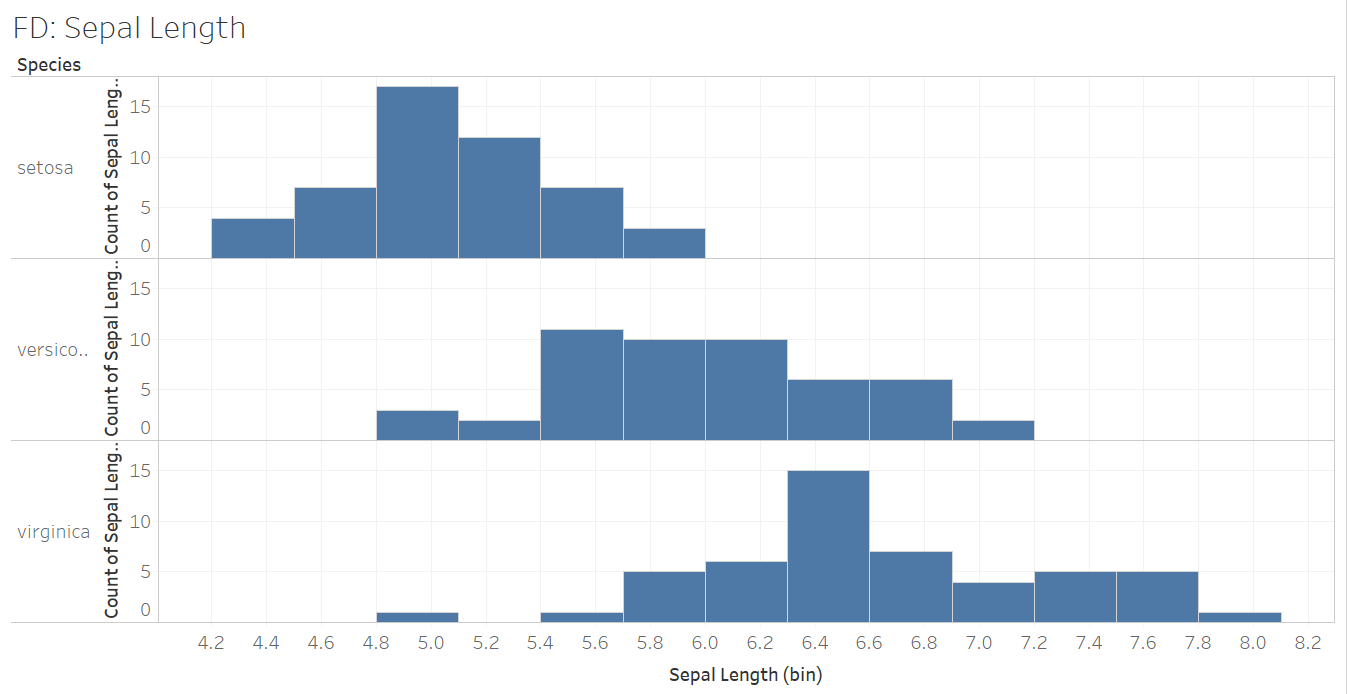
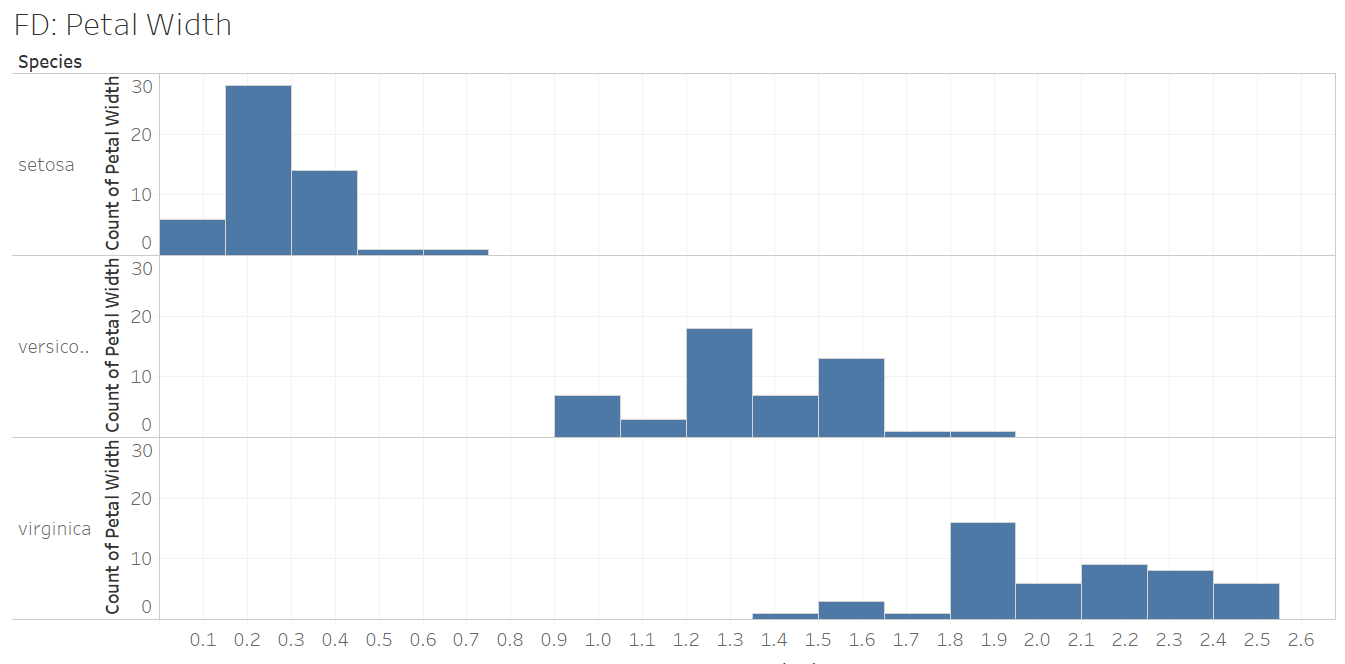


Tableau – visual of the distribution of the data for each attribute by species.





R – 1) using the data setup from last week’s homework (when you were asked to restructure/split the original data), boxplots of each of the attributes for each species using the base R packages for this (add a y-axis label for measurement, title, and different color scheme than that used in lecture).

2) Then a second boxplot visual using ggplot.

Brief explanation of conclusions you can draw from these visualizations about the dataset. Pro/cons of the softwares, challenges you may have had with this assignment/softwares.

# install.packages("psych")

# install.packages('ggplot2', dependencies = TRUE)

# install.packages("vctrs", repos = "https://packagemanager.rstudio.com/cran/latest")

#################################################################################################

library(psych)

library(datasets)

library(ggplot2)

#################################################################################################

iris

describe(iris)

summary(iris)

Category  = split(iris, iris$Species)

#################################################################################################

# Regualar R BoxPlot

boxplot(Category$setosa[1:4], main='Setosa' ,ylab='Length in Centimeters', col='green' )

boxplot(Category$virginica[1:4], main='virginica' ,ylab='Length in Centimeters', col='green' )

boxplot(Category$versicolor[1:4], main='versicolor' ,ylab='Length in Centimeters', col='green' )

#################################################################################################

# ggplot Boxplot

ggplot(iris, aes(x=Sepal.Length, y=Species)) + geom\_boxplot()

ggplot(iris, aes(x=Sepal.Width, y=Species)) + geom\_boxplot()

ggplot(iris, aes(x=Petal.Length , y=Species)) + geom\_boxplot()

ggplot(iris, aes(x=Petal.Width, y=Species)) + geom\_boxplot()

#################################################################################################

##### Conclusion #####

#  - Sepal Measurements

# Sepal Length of Virginica is higher then other categories. However, this an outlier with a

# very small sepal length. Setosa has the smallest sepal length. However, Setosa has the

# largest sepal width. Versicolor and virginia have some overplaying measuremnt for sepal width.

#  - Petal Measurements

# Petal length of setosa is significantly smaller then other categories. Virginica has the

# largest and versicolor has 2nd largest petal lengths. Sepal width follows the same pattern.

# From Apparent looks virginca would be more slimmer and lengthier then  versicolor and Setosa.

# And Setoso would have more wide Sepals.

#################################################################################################