VIT-AP UNIVERSITY, ANDHRA PRADESH

CSE2047 – Data Analytics - Lab Sheet : 7

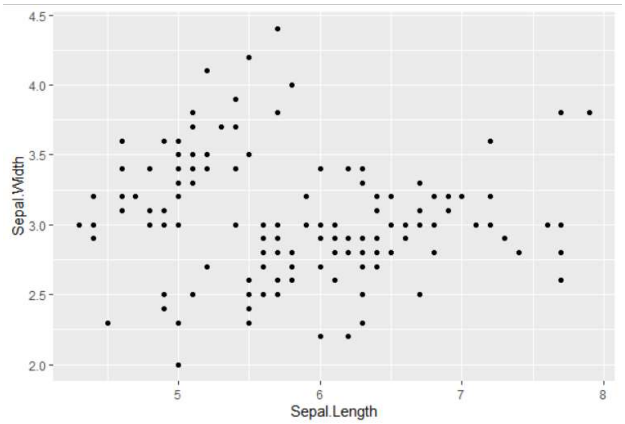
Academic year: 2020-2021 **Branch/ Class:** B.Tech/M.Tech
Semester: Fall **Date:**
Faculty Name: Prof. S.Gopikrishnan **School:** SCOPE
Student name: Valiveti Manikanta bhuvanesh **Reg. no.: 19BCD7088**

LAB 7

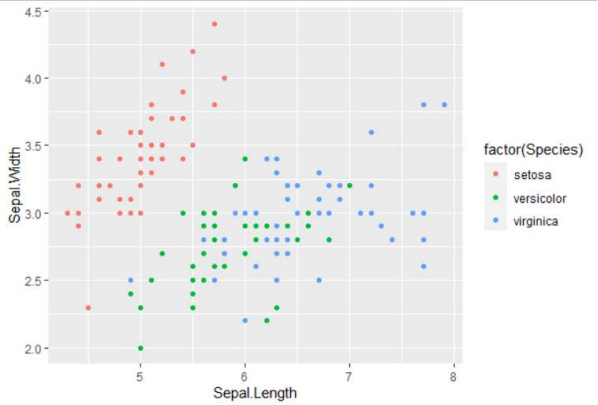
Use Iris dataset to show visual exploratory data Analysis in R

Part-1 (Ref: https://www.guru99.com/r-scatter-plot-ggplot2.html)

1. **Draw Basic scatter plot between SepalLengthCm and SepalWidthCm.**
   df<-iris
   ggplot(df, aes(x = Sepal.Length, y = Sepal.Width)) + geom_point()

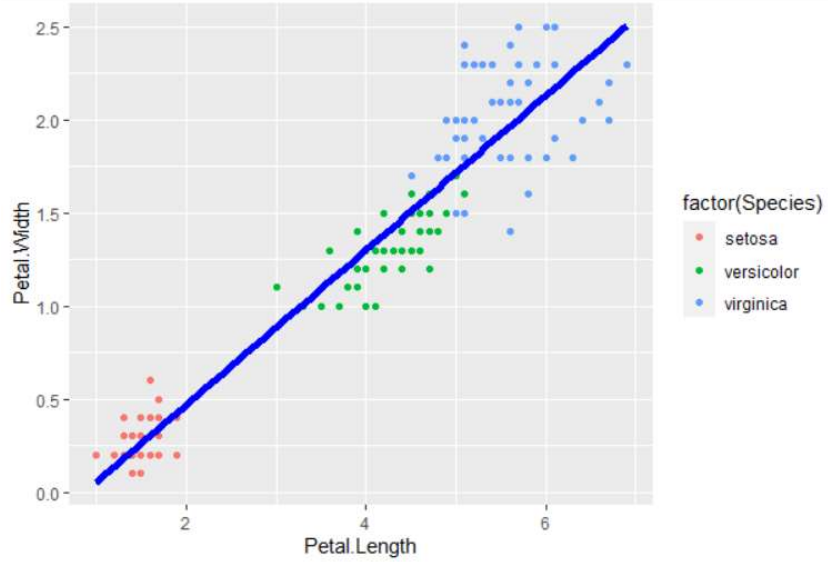


2. **Visualize Scatter plot with color between group SepalLengthCm and SepalWidthCm and group by Species.**
   ggplot(df, aes(x = Sepal.Length, y = Sepal.Width)) + geom_point(aes(color = factor(Species)))



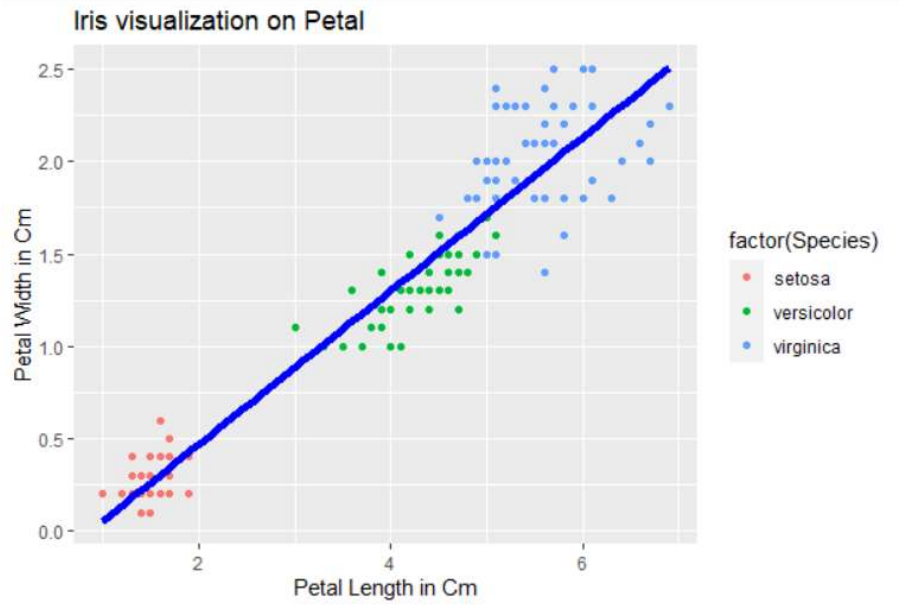
3. **Visualize Scatter plot with added fitted values between PetalLengthCm and PetalWidthCm and use Linear regression for fitted line.**

```
mygraph<-ggplot(df, aes(x = Petal.Length, y = Petal.Width)) +
geom_point(aes(color = factor(Species))) + stat_smooth(method = "lm",col =
"#0000FF",se = FALSE,size = 2)
mygraph
```



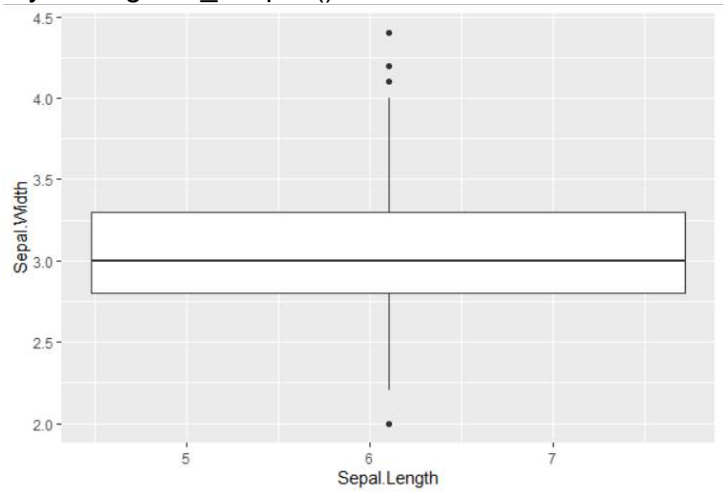
4. **Add the following information to the above drawn graph**
   a. **Add a title as “Iris visualization on Petal”**
   b. **Rename x-axis as “Petal Length in Cm” and y-axis as “Petal Width in Cm”**

```
mygraph + labs(title = "Iris visualization on Petal",x="Petal Length in
Cm",y="Petal Width in Cm"
```

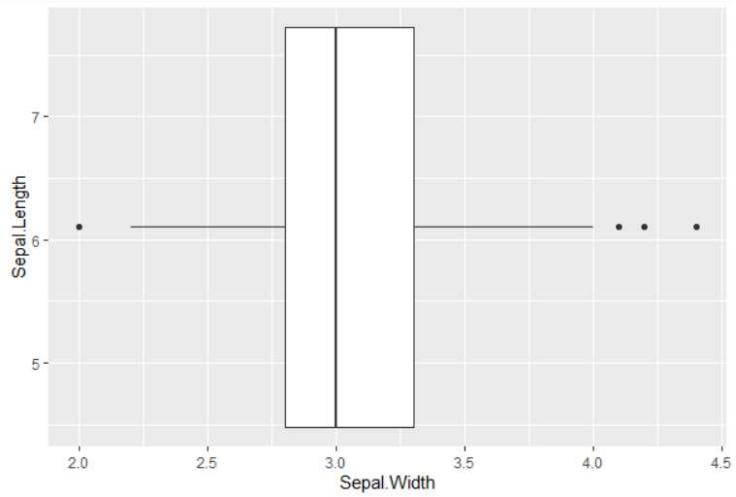


Part-2 (Ref: https://www.guru99.com/r-boxplot-tutorial.html)

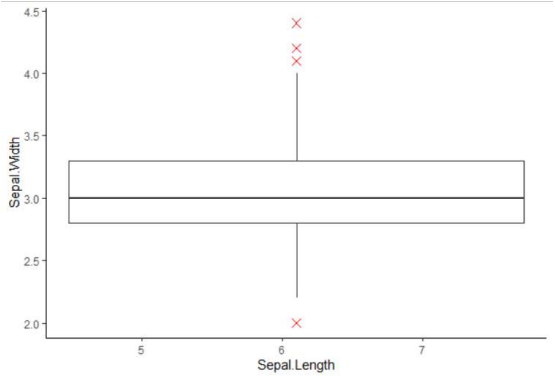
1. **Visualize the Basic box plot on species wise weight data**
   mybox<-ggplot(df, aes(x = Sepal.Length, y = Sepal.Width))
   mybox + geom_boxplot()



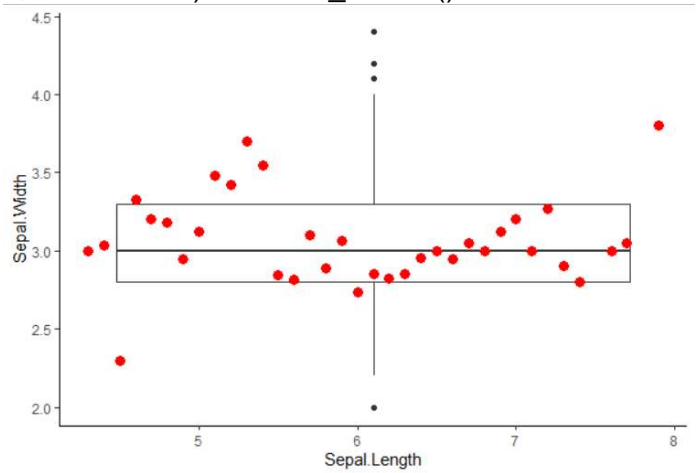
2. **Change side of the graph which you have plotted in question 1**
   mybox + geom_boxplot()+ coord_flip()



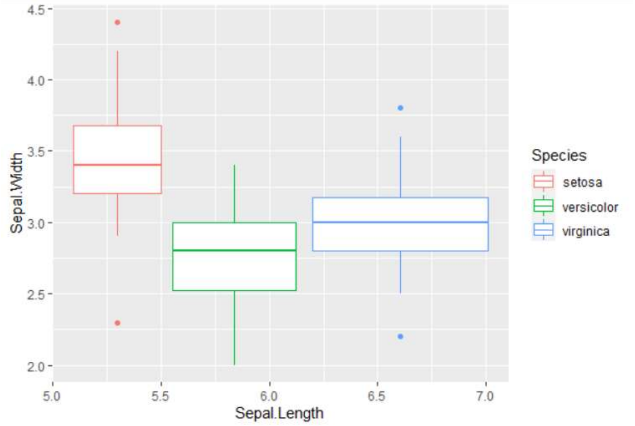
3. **Visualize the outliers of weight data in different colour**
   mybox + geom_boxplot(outlier.colour = "red",outlier.shape = 4,outlier.size = 3)
   + theme_classic()



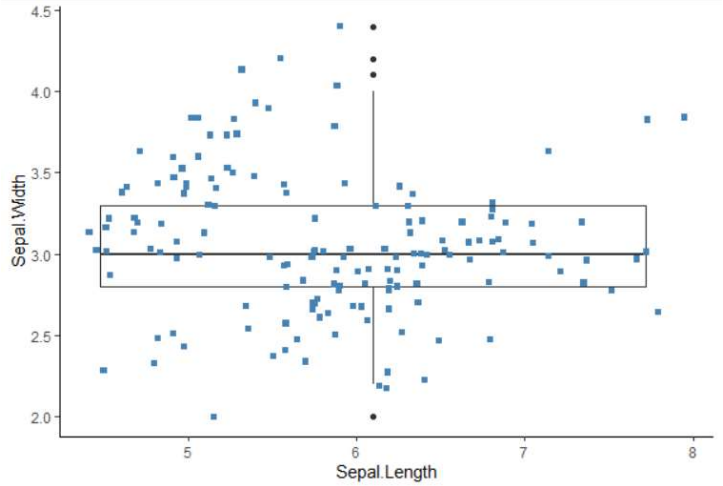
4. **Add the summary statistic on the box plot drawn in question 1.**
   mybox + geom_boxplot() + stat_summary(fun.y = mean,geom = "point", size = 3,color = "red") + theme_classic()



5. **Change the colour of the box based on season**
   ggplot(df, aes(x = Sepal.Length, y = Sepal.Width,color=Species))+ geom_boxplot()

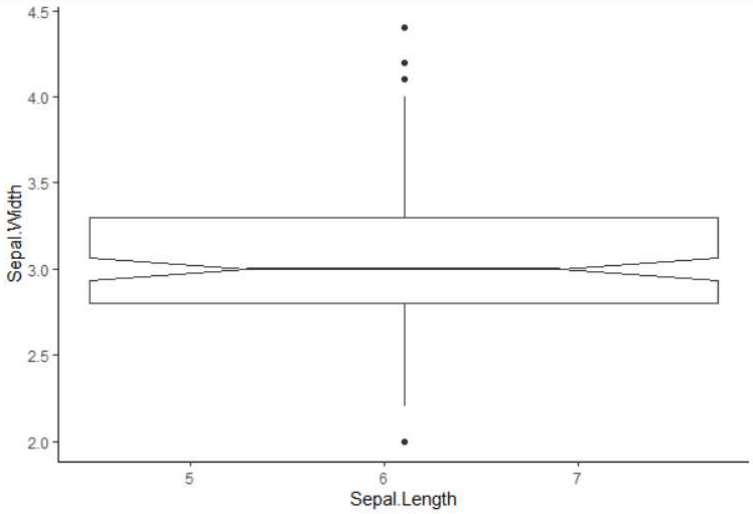


6. **Visualize the Box Plot with Jittered Dots**
   mybox + geom_boxplot() + geom_jitter(shape = 15,color = "steelblue",position = position_jitter(width = 0.21)) +theme_classic()



7. **Plot Notched Box Plot**

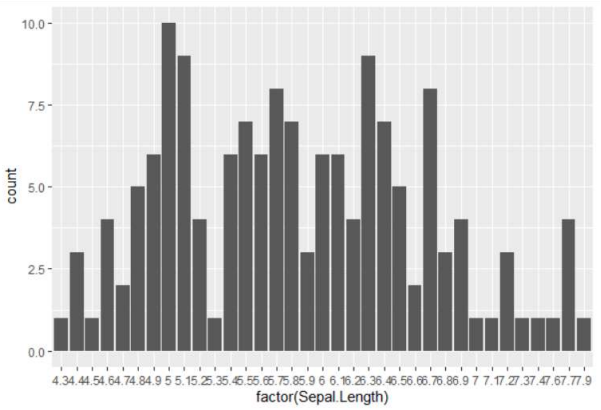
mybox + geom_boxplot(notch = TRUE) + theme_classic()



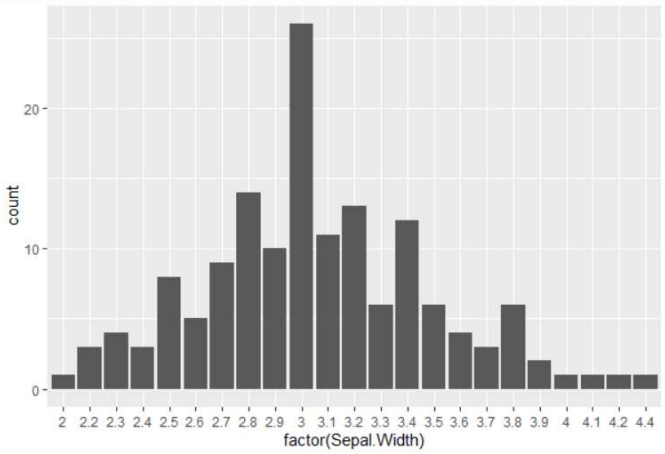
Part-3 (Ref: https://www.guru99.com/r-bar-chart-histogram.html)

1. **Show individual geom bar plot on factors of SepalLengthCm, SepalWidthCm, SepalLengthCm and SepalWidthCm**

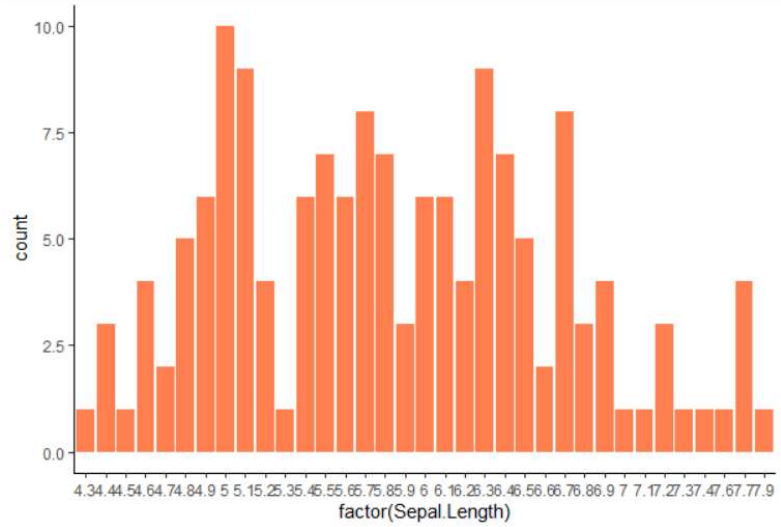
mybar<-ggplot(df,aes(x=factor(Sepal.Length)))
mybar+geom_bar()



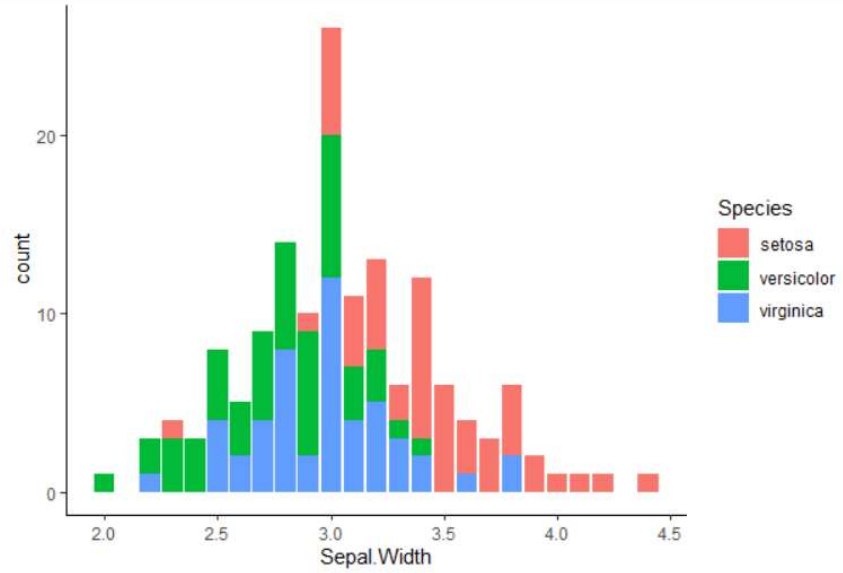
ggplot(df,aes(x=factor(Sepal.Width)))+geom_bar()



2. **Visualize the colour geom bar plot on factors of season data**
   mybar + geom_bar(fill = "coral")+theme_classic()

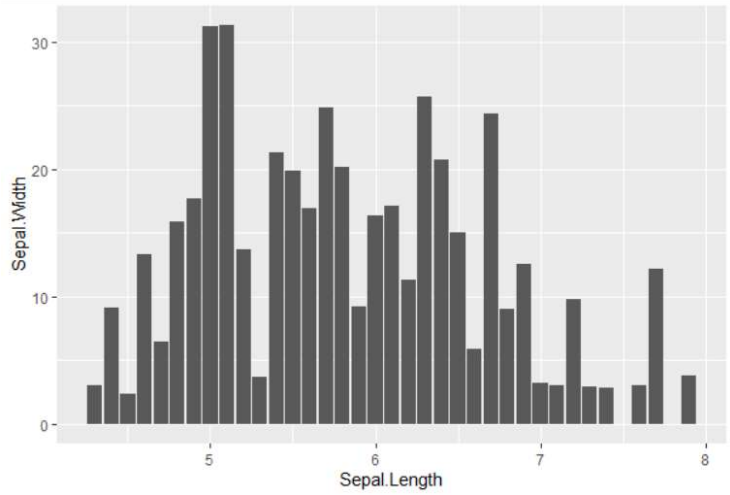


3. **Add species group in the bars which you have drawn in question 2.**
   mutate(Species = factor(Species, labels = c("setosa", "versicolor","virginica")),Sepal.Width = factor(Sepal.Width))
   ggplot(df, aes(x = Sepal.Width, fill = Species)) +geom_bar() +theme_classic()



4. **Create a basic histogram with season and weight data**

```
mutate(Sepal.Length = factor(Sepal.Length))
group_by(Species)
ggplot(df, aes(x = Sepal.Length, y = Sepal.Width)) +geom_bar(stat =
"identity")
```



5. **Change the colour and add labels to the graph which you drawn in question 4**

```
ggplot(df, aes(x = Sepal.Length, y = Sepal.Width,fill=Species))
+geom_bar(stat = "identity")+geom_text(aes(label =Sepal.Width
))+theme_classic()
```