**Bangabandhu Sheikh Mujibur Rahman Agricultural University**
**EDGE_Batch-11**
**Quiz Exam**
**Marks: 20 Time: 90 minutes**
**Name: Jahidul Islam**
**Reg. No:21-05-5738 Faculty: Agriculture**

**Note**: Submit the completed file to rabiulauwul@bsmrau.edu.bd with subject ***EDGE11_Quiz_Your registration number_ Dept.***

1. **Short Questions** **(6*1=06)**
   a) In R, you can use install.packages() to install a package from CRAN.
   b) To check the structure of an object in R, the function str() is used.
   c) To subset a data frame by selecting specific rows and columns, the $ operator is used.
   d) In R, the summary() function provides a summary of key descriptive statistics
   e) In R, the na.omit() function can be used to remove missing values (NA) from a vector x.
   f) The residuals of a regression model are the differences between the observed values and the fitted values predicted by the model.

2. For the ***iris*** data**:** **(7)**
   a) Calculate descriptive statistics ($\boldsymbol{median \pm SD, mean, CV}$) for each numeric variable in a single table.
   b) Construct boxplots with ggplot2 package for each variable by ***Species*** categories with color aesthetic and interpret your results.

3. For the provided dataset of "***vegitables"***, answer the following questions: **(7)**
   a) Identify missing values in each variable and impute them using the mean values of the corresponding variables.
   b) Fit a suitable multiple linear regression model for the dataset and interpret your findings.

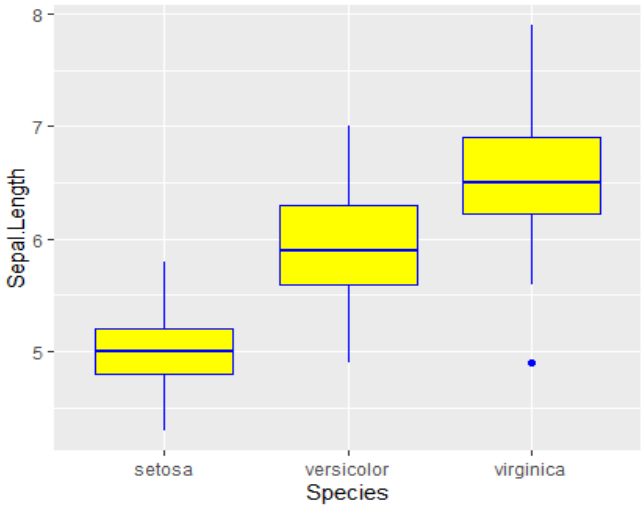
**ANS: 2. a)**

```
x<-iris
str(x)
me1<-median(x$Sepal.Length)
me2<-median(x$Sepal.Width)
me3<-median(x$Petal.Length)
```

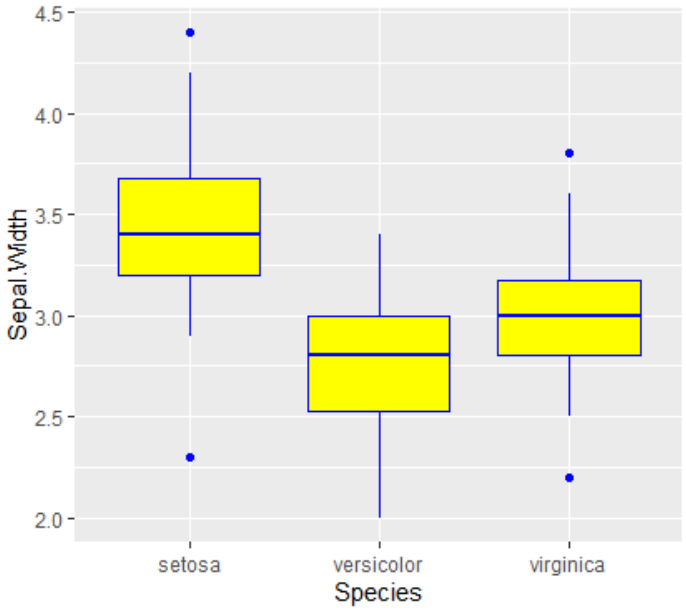
```
me4<-median(x$Petal.Width)
A1<-me1+sd1
A2<-me2+sd2
A3<-me3+sd3
A4<-me4+sd4

A5<-me1-sd1
A6<-me2-sd2
A7<-me3-sd3
A8<-me4-sd4

M1<-mean(x$Sepal.Length)
M2<-mean(x$Sepal.Width)
M3<-mean(x$Petal.Length)
M4<-mean(x$Petal.Width)

sd1<-sd(x$Sepal.Length)
sd2<-sd(x$Sepal.Width)
sd3<-sd(x$Petal.Length)
sd4<-sd(x$Petal.Width)
cv1<-CV<-(sd1/M1*100)
cv2<-CV<-(sd2/M2*100)
cv3<-CV<-(sd3/M3*100)
cv4<-CV<-(sd4/M4*100)

table<-table(A1,A2,A3,A4,A5,A6,A7,A8,M1,M2,M3,M4,cv1,cv2,cv3,cv4)
```

A1 =6.62806612797786 , A2= 3.4358662849367 , A3 = 6.11529823325947, A4 = 2.06223766896035, A5 = 4.97193387202214, A6 = 2.5641337150633, A7 = 2.58470176674053, A8 = 0.537762331039654, M1 = 5.84333333333333, M2 = 3.05733333333333, M3 = 3.758, M4 = 1.19933333333333, cv1 = 14.171125977944, cv2 = 14.2564201353041, cv3 = 46.974407484286, cv4 = 63.5551141434419

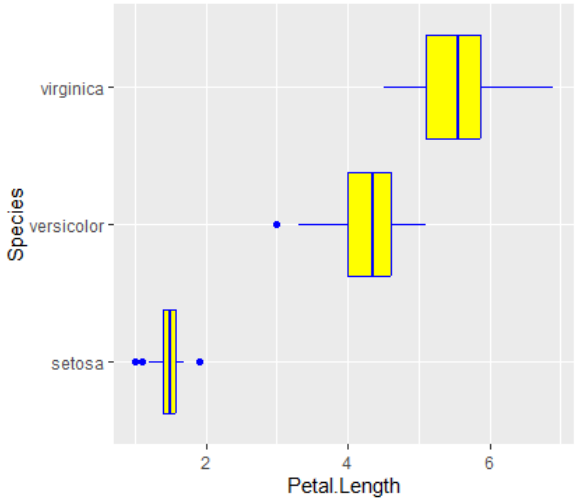
b)




Setosa has the smallest sepals, Versicolor has intermediate-sized sepals, and Virginica has the largest sepals. There is no overlap between Setosa and the other two species, while Versicolor and Virginica overlap somewhat in their ranges




The boxplot shows that Setosa has the widest sepals on average, Versicolor the narrowest, and Virginica falls in between. Setosa also exhibits the greatest variation in Sepal Width, with outliers present in both Setosa and Virginica.




Setosa has the shortest petals with minimal variation, Virginica has the longest petals with the widest range, and Versicolor falls in between. There is no overlap between Setosa and the other two species, while some overlap exists between Versicolor and Virginica.

2) setwd("C:/Data Analysis")

Y<-read.csv("Question.csv")

mean(Y$Length.of.vine..cm.,na.rm = TRUE)

mean(Y$Length.of.vine..cm.)

Y$Length.of.vine..cm.[is.na(Y$Length.of.vine..cm.)]<-

  mean(Y$Length.of.vine..cm.,na.rm = TRUE)