Date:-9/2/23 Vamsi Manubothu-192011429

LAB ASSIGNMENT-2 FEB 9

ITA0443-STATISTICS WITH R PROGRAMMING FOR REAL TIME PROBLEMS

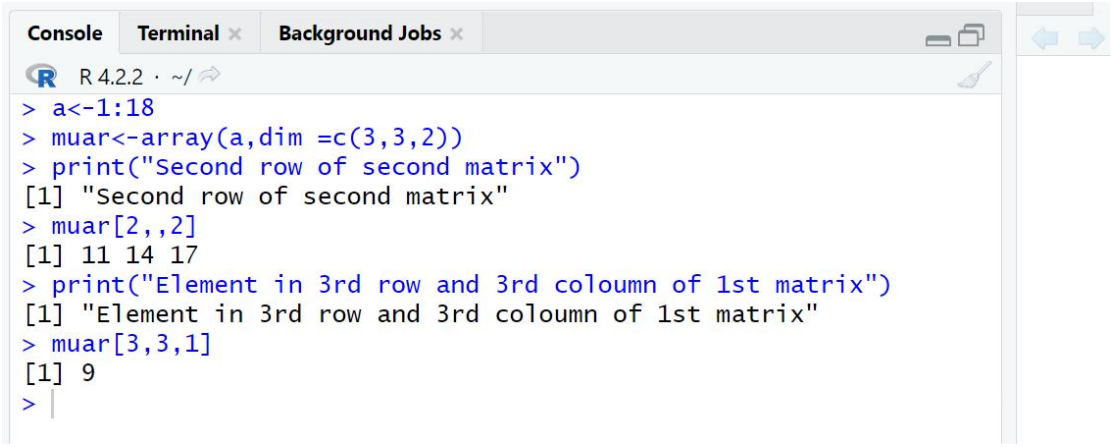
GITHUBLINK:-https://github.com/Vamsim29/ITA0443-STATISTICS-WITH-R-PROGRAMMING

1.Write a R program to create an array of two 3x3 matrices each with 3 rows and 3 columns from
two given two vectors. Print the second row of the second matrix of the array and the element in
the 3rd row and 3rd column of the 1st matrix.

PROGRAM:-

```
a<-1:18
muar<-array(a,dim =c(3,3,2))
print("Second row of second matrix")
muar[2,,2]
print("Element in 3rd row and 3rd coloumn of 1st matrix")
muar[3,3,1]
```

OUTPUT:-

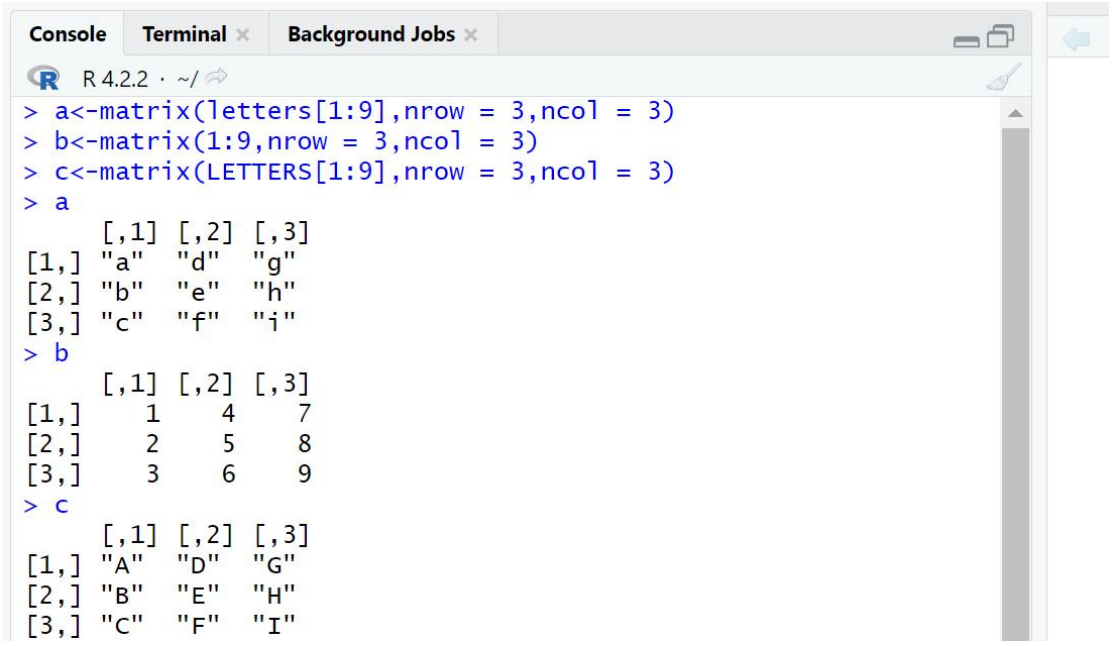


2. Write a R program to combine three arrays so that the first row of the first array is followed by the first row of the second array and then first row of the third array.

PROGRAM:-

```
a<-matrix(letters[1:9],nrow = 3,ncol = 3)
b<-matrix(1:9,nrow = 3,ncol = 3)
c<-matrix(LETTERS[1:9],nrow = 3,ncol = 3)
a
b
c
m = matrix(t(cbind(a,b,c)),byrow = 9,ncol = 3)
m
```

OUTPUT:-



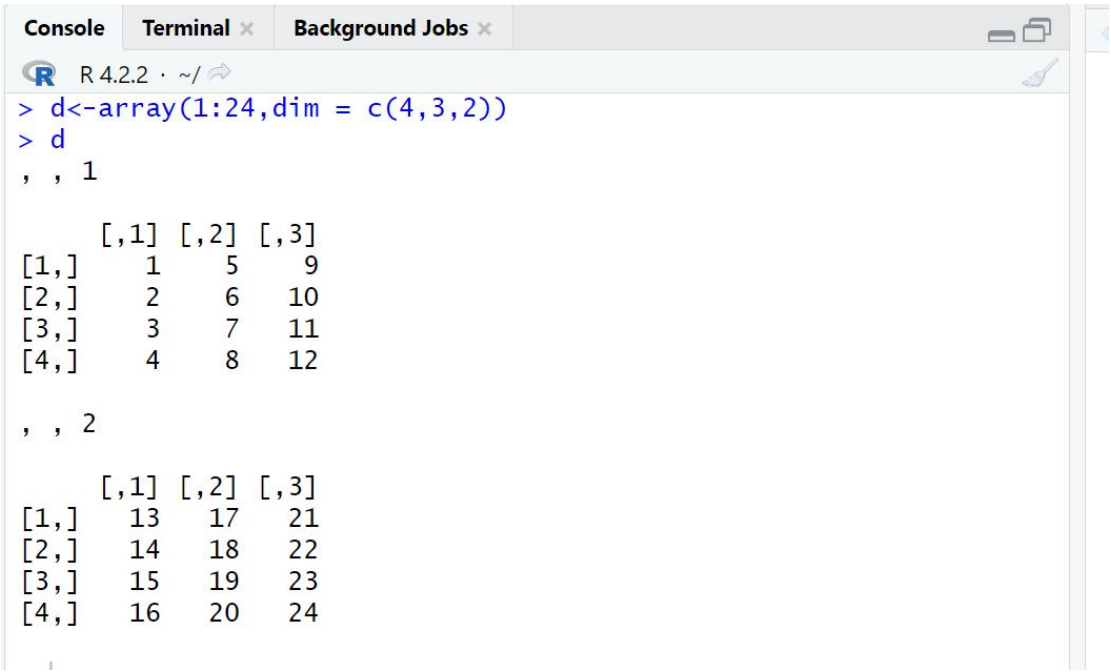

```
> m = matrix(t(cbind(a,b,c)),byrow = 9,ncol = 3)
> m
      [,1] [,2] [,3]
 [1,] "a"  "d"  "g"
 [2,] "1"  "4"  "7"
 [3,] "A"  "D"  "G"
 [4,] "b"  "e"  "h"
 [5,] "2"  "5"  "8"
 [6,] "B"  "E"  "H"
 [7,] "c"  "f"  "i"
 [8,] "3"  "6"  "9"
 [9,] "C"  "F"  "I"
```

3.Write a R program to create an array using four given columns, three given rows, and two given tables and display the content of the array.

PROGRAM:-

```
d<-array(1:24,dim = c(4,3,2))
d
```

OUTPUT:-

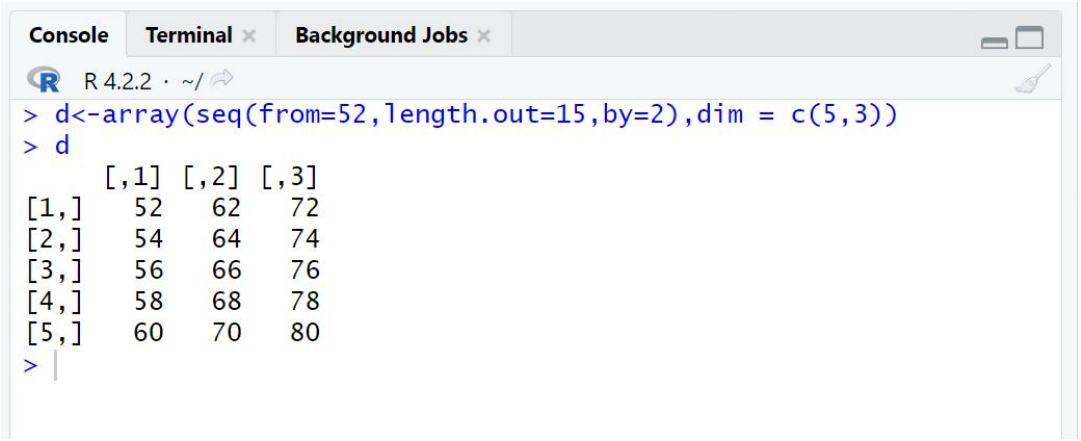


4. Write a R program to create a two-dimensional 5x3 array of sequence of even integers greater than 50.

PROGRAM:-

```
d<-array(seq(from=52,length.out=15,by=2),dim = c(5,3))
d
```

OUTPUT:-



Use Below Data frame from question 5 to 9

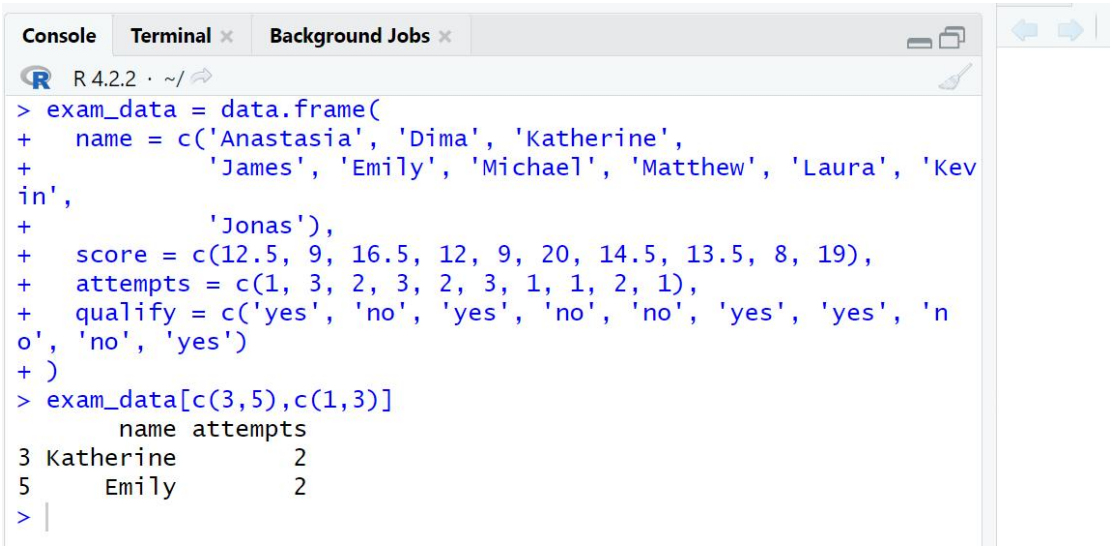
```
exam_data = data.frame(
name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin',
'Jonas'),
score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),
attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),
qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')
)
```

5. Write a R program to extract 3 rd and 5 th rows with 1 st and 3 rd columns from a given data frame

PROGRAM:-

```
exam_data = data.frame(
 name = c('Anastasia', 'Dima', 'Katherine',
        'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin',
        'Jonas'),
 score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),
 attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),
 qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')
)
exam_data[c(3,5),c(1,3)]
```

OUTPUT:-

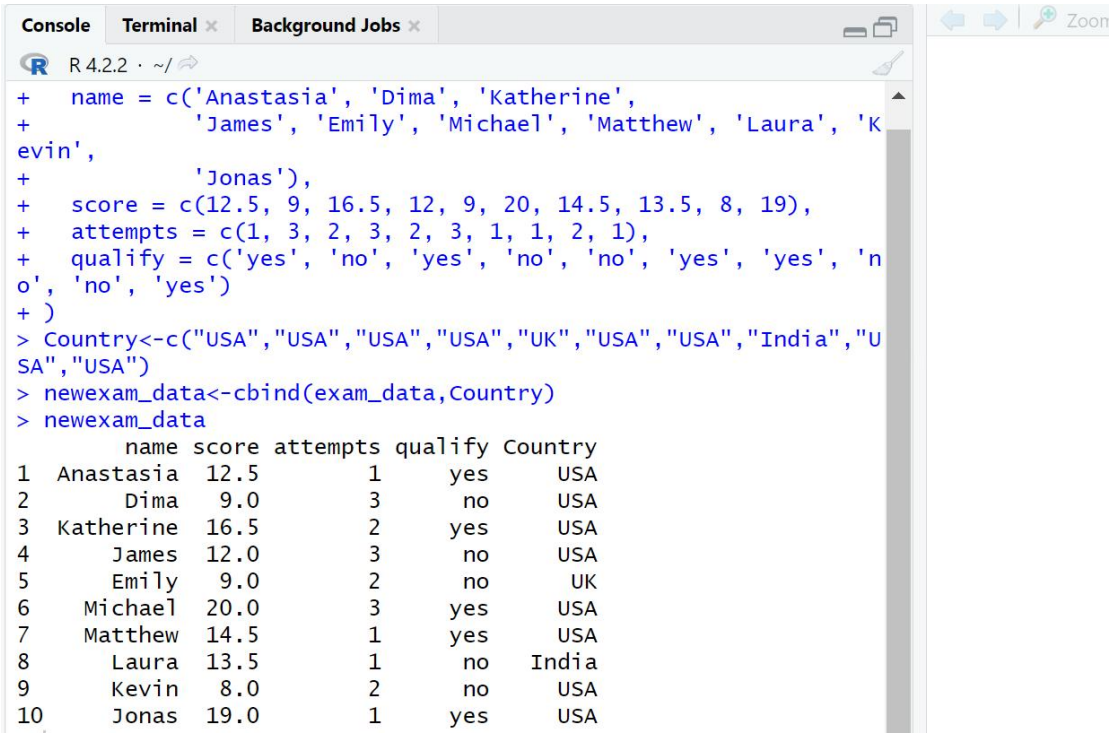


6. Write a R program to add a new column named country in a given data frame

PROGRAM:-

```
exam_data = data.frame(
 name = c('Anastasia', 'Dima', 'Katherine',
        'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin',
        'Jonas'),
 score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),
 attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),
 qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')
)
Country<-c("USA","USA","USA","USA","UK","USA","USA","India","USA","USA")
newexam_data<-cbind(exam_data,Country)
newexam_data
```

OUTPUT:-

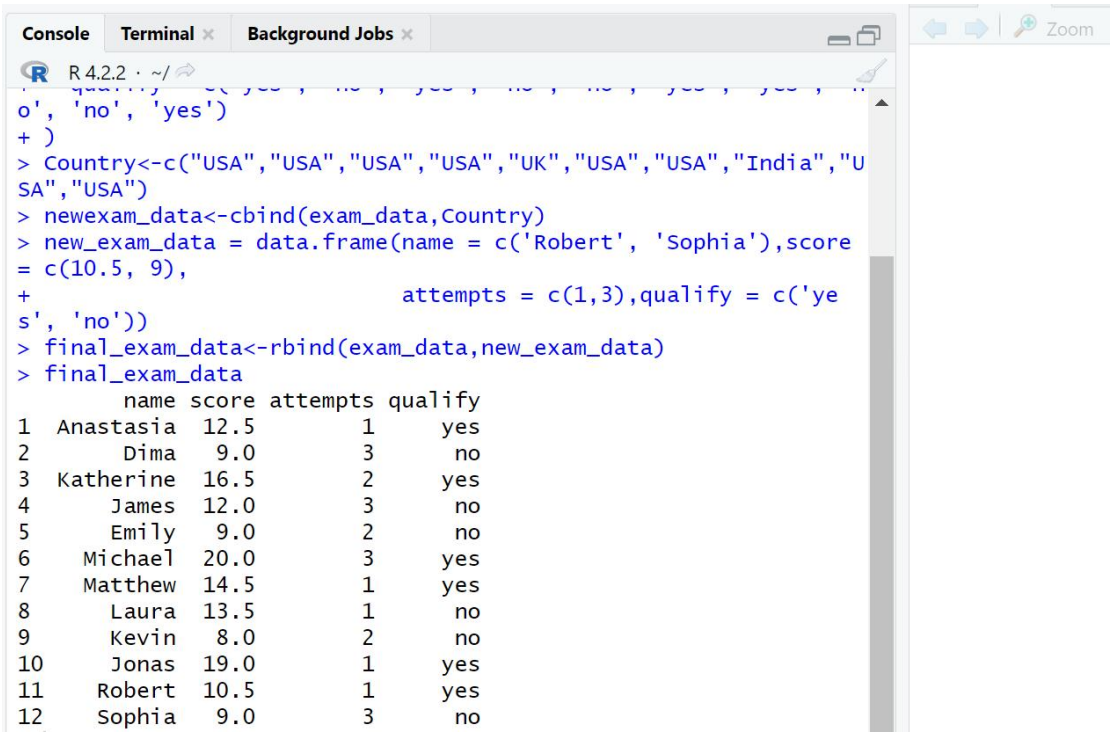


7.Write a R program to add new row(s) to an existing data frame
new_exam_data = data.frame(name = c('Robert', 'Sophia'),score = c(10.5, 9), attempts = c(1, 3) ,qualify = c('yes', 'no'))

PROGRAM:-

```
exam_data = data.frame(
 name = c('Anastasia', 'Dima', 'Katherine',
       'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin',
       'Jonas'),
 score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),
 attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),
 qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')
)
Country<-c("USA","USA","USA","USA","UK","USA","USA","India","USA","USA")
newexam_data<-cbind(exam_data,Country)
new_exam_data = data.frame(name = c('Robert', 'Sophia'),score = c(10.5, 9),
                  attempts = c(1,3),qualify = c('yes', 'no'))
final_exam_data<-rbind(exam_data,new_exam_data)
final_exam_data
```

OUTPUT:-

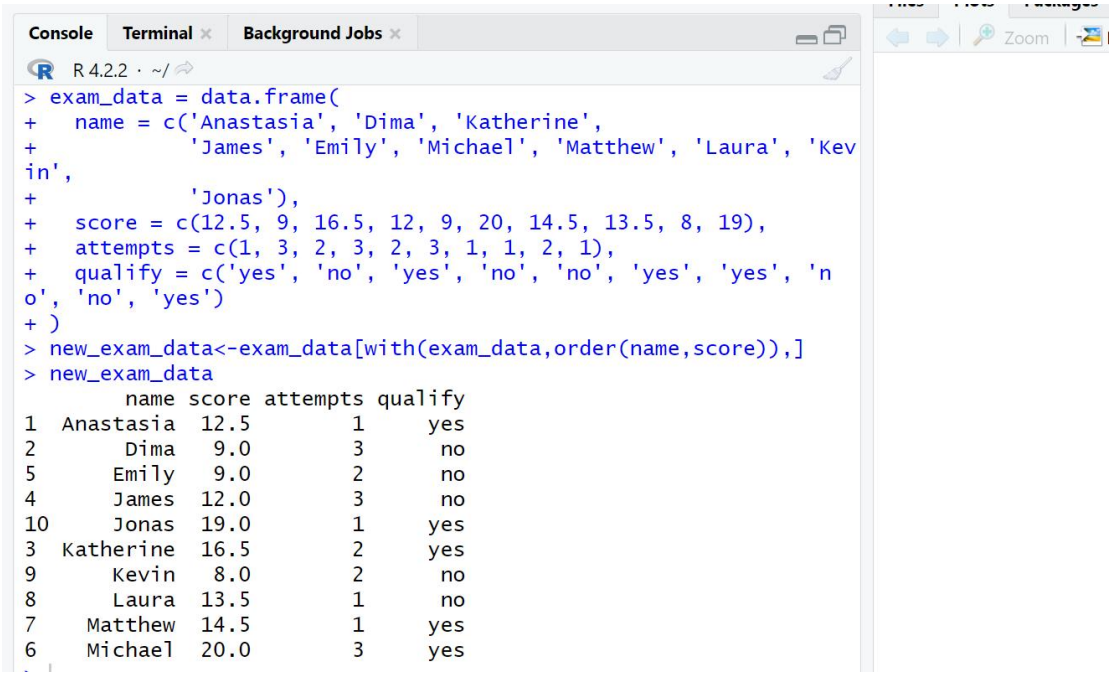


8.Write a R program to sort a given data frame by name and score

PROGRAM:-

```
exam_data = data.frame(
 name = c('Anastasia', 'Dima', 'Katherine',
       'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin',
       'Jonas'),
 score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),
 attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),
 qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')
)
new_exam_data<-exam_data[with(exam_data,order(name,score)),]
new_exam_data
```

OUTPUT:-

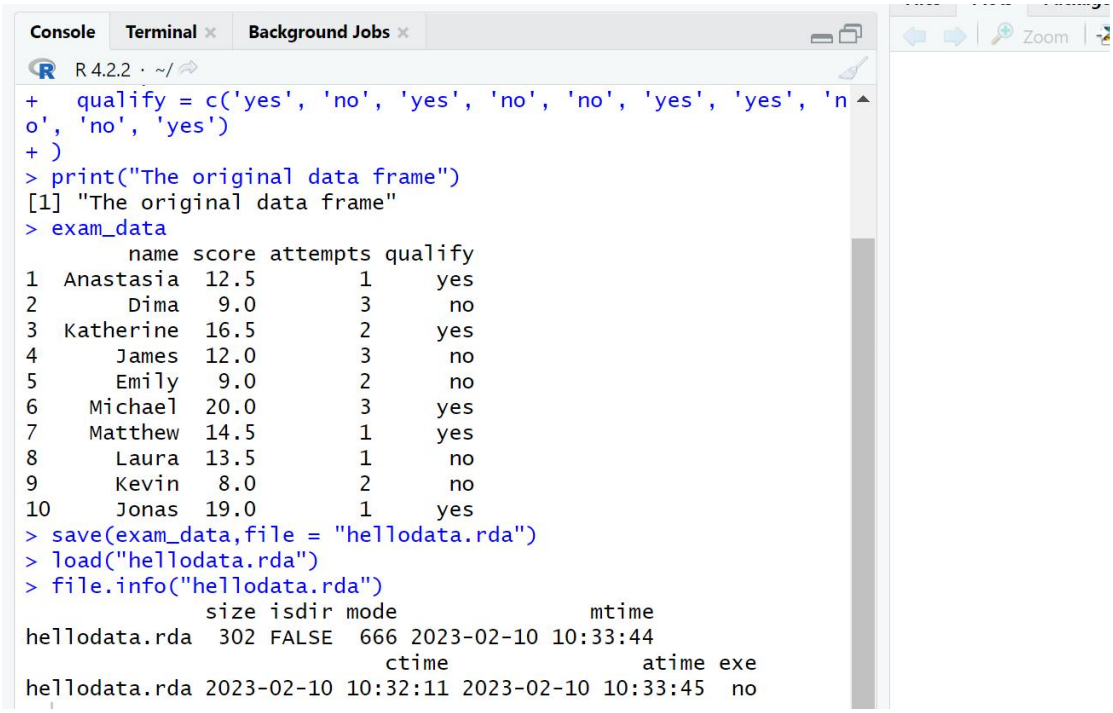


9.Write a R program to save the information of a data frame in a file and display the information of the file.

PROGRAM:-

```
exam_data = data.frame(
 name = c('Anastasia', 'Dima', 'Katherine',
       'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin',
       'Jonas'),
 score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),
 attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),
 qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')
)
print("The original data frame")
exam_data
save(exam_data,file = "hellodata.rda")
load("hellodata.rda")
file.info("hellodata.rda")
```

OUTPUT:-

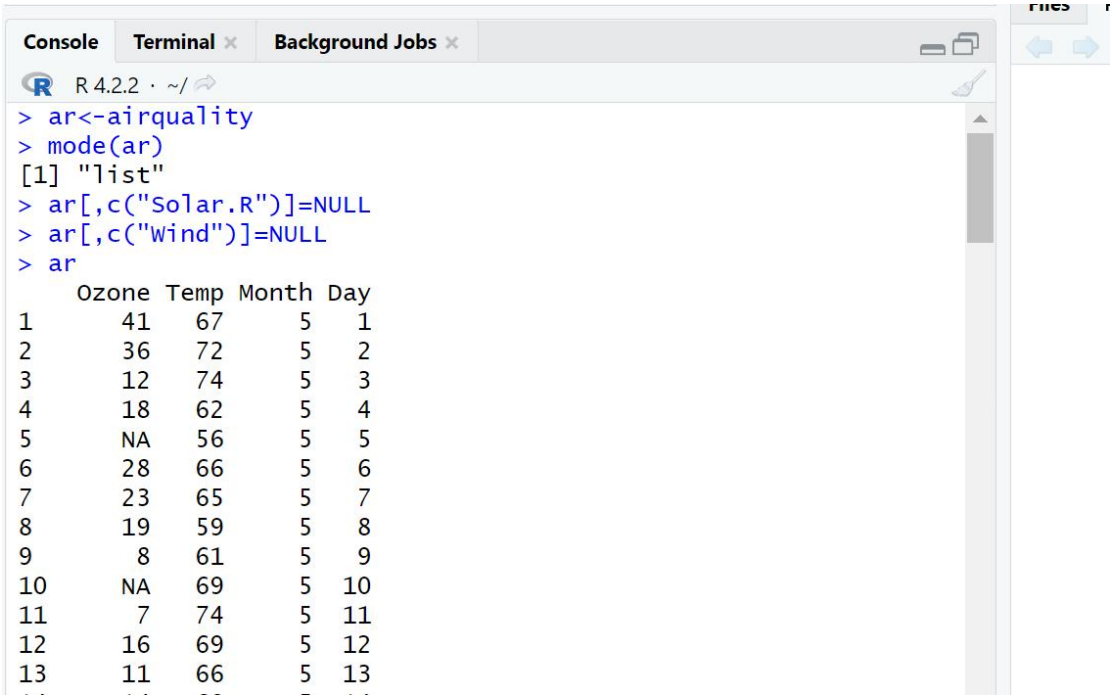


10. Write a R program to call the (built-in) data airquality. Check whether it is a data frame or not? Order the entire data frame by the first and second column. Remove the variables 'Solar.R' and 'Wind' and display the data frame.

PROGRAM:-

```
ar<-airquality
mode(ar)
ar[,c("Solar.R")]=NULL
ar[,c("Wind")]=NULL
ar
```

OUTPUT:-

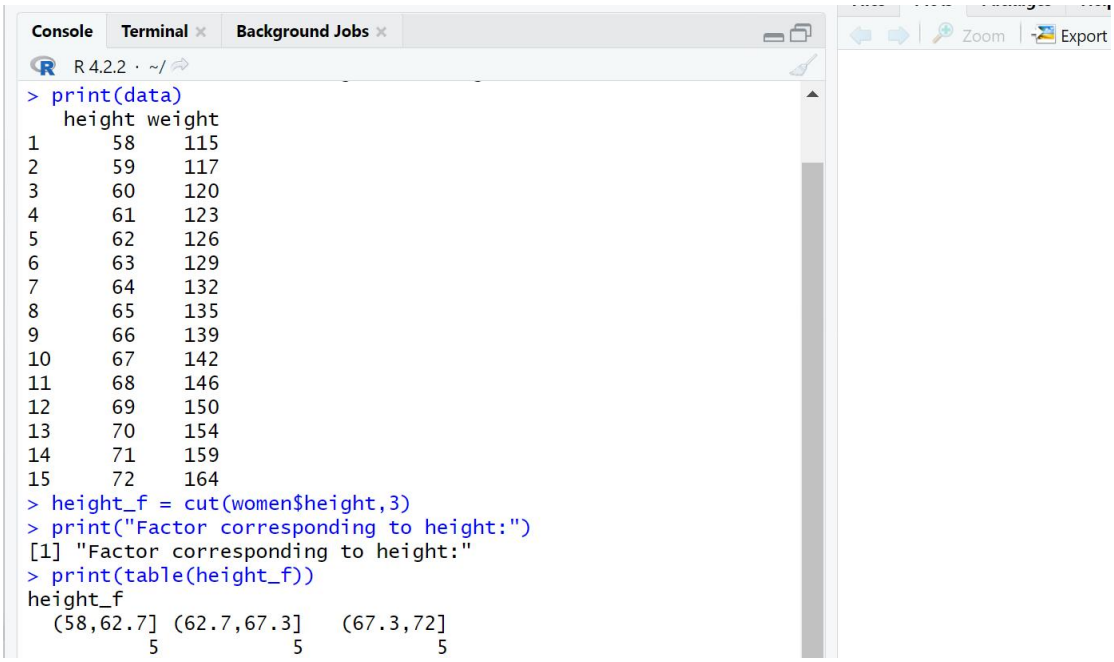


11.Write a R program to create a factor corresponding to height of women data set , which inbuild in R, contains height and weights for a sample of women.

PROGRAM:-

```
data = women
print("Women data set of height and weights:")
print(data)
height_f = cut(women$height,3)
print("Factor corresponding to height:")
print(table(height_f))
```

OUTPUT:-



12. Write a R program to extract the five of the levels of factor created from a random sample from the LETTERS (Part of the base R distribution.)

PROGRAM:-

```
data<-sample(LETTERS,size = 20,replace = TRUE)
print("printing the original data")
data
fac<-factor(data)
print("printing original factors")
fac
print("only five levels")
print(table(data[1:5]))
```

OUTPUT:-

```
> data<-sample(LETTERS,size = 20,replace = TRUE)
> print("printing the original data")
[1] "printing the original data"
> data
 [1] "P" "T" "Y" "O" "V" "O" "R" "F" "R" "Y" "M" "N" "W" "Q"
[15] "L" "Q" "G" "C" "B" "Z"
> fac<-factor(data)
> print("printing original factors")
[1] "printing original factors"
> fac
 [1] P T Y O V O R F R Y M N W Q L Q G C B Z
Levels: B C F G L M N O P Q R T V W Y Z
> print("only five levels")
[1] "only five levels"
> print(table(data[1:5]))

O P T V Y
1 1 1 1 1
>
```

13.Iris dataset is a very famous dataset in almost all data mining, machine learning courses, and it has been an R build-in dataset. The dataset consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features(variables) were measured from each sample, they are the length and the width of sepal and petal, in centimetres. Perform the following EDA steps .

(i) Find dimension, Structure, Summary statistics, Standard Deviation of all features.

(ii)Find mean and standard deviation of features groped by three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor)

(iii)Find quantile value of sepal width and length

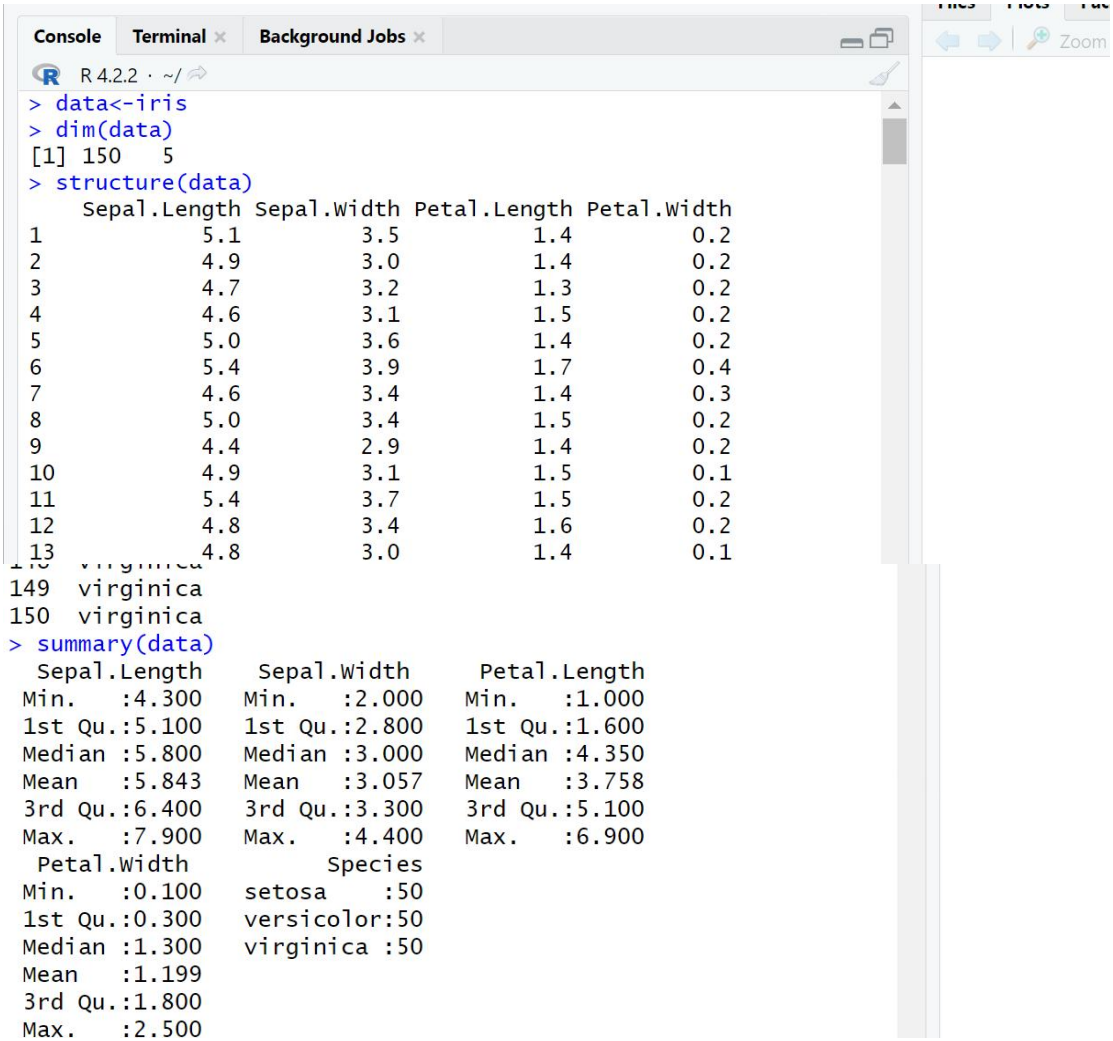
(iV)create new data frame named iris1 which have a new column name Sepal.Length.Cate that categorizes “Sepal.Length” by quantile

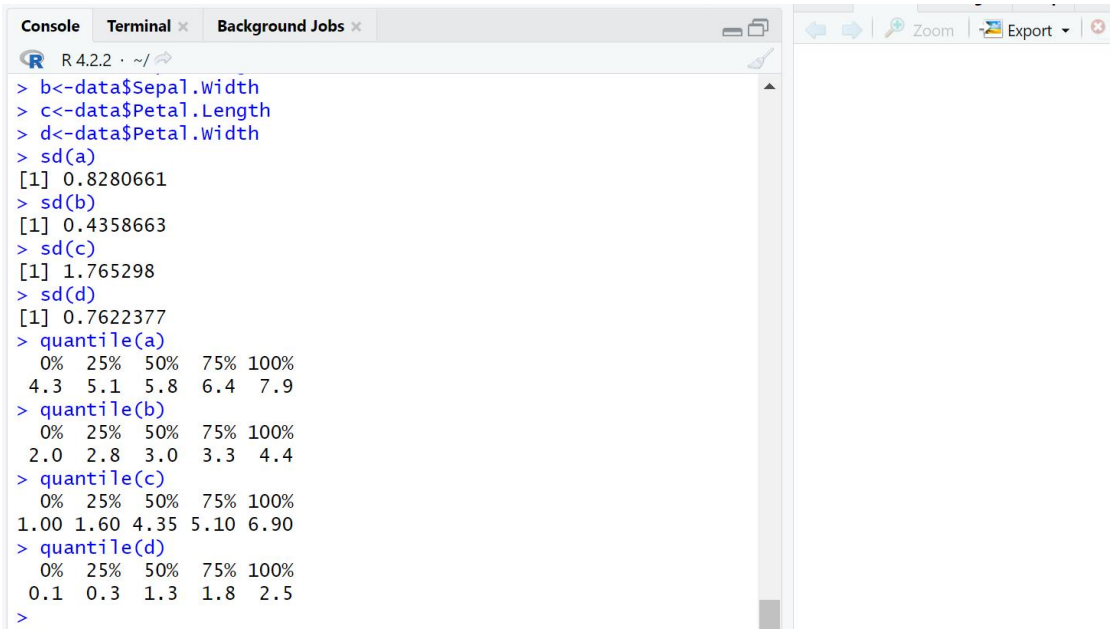
(V) Average value of numerical varialbes by two categorical variables: Species and Sepal.Length.Cate:

PROGRAM:-

```
data<-iris
dim(data)
structure(data)
summary(data)
a<-data$Sepal.Length
b<-data$Sepal.Width
c<-data$Petal.Length
d<-data$Petal.Width
sd(a)
sd(b)
sd(c)
sd(d)
quantile(a)
quantile(b)
quantile(c)
quantile(d)
```

OUTPUT:-





14.Titanic Casualties – Use the standard 'Titanic' dataset which is part of R Base to answer the following questions.

(i). Use an appropriate apply function to get the sum of males vs females aboard.

(ii). Get a table with the sum of survivors vs sex.

(iii). Get a table with the sum of passengers by sex vs age

PROGRAM:-

```
data<-Titanic
aa<-data$Male
bb<-data$Female
aa
bb
```