Sub : Foundations for Data Analytics

Name: Gudi Varaprasad

Reg.No: 19BCE7048

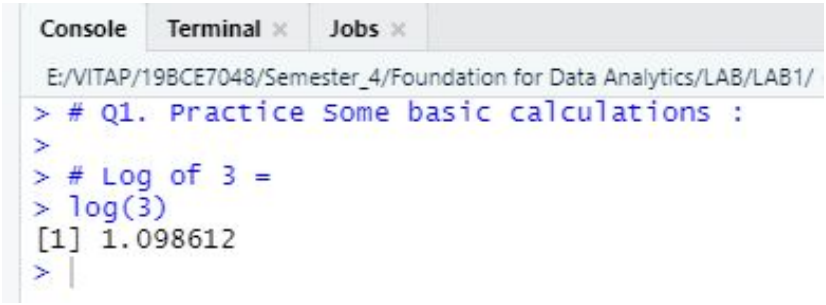
Slot No: L33+L34

Assignment 1: **Date: 02/02/2021**

Q1. Practice Some basic calculations.

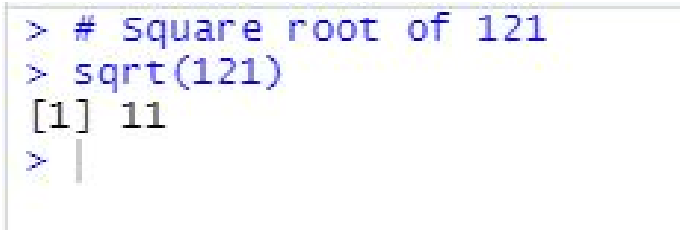
Log of 3

> log(3)



Square root of 121

> sqrt(121)



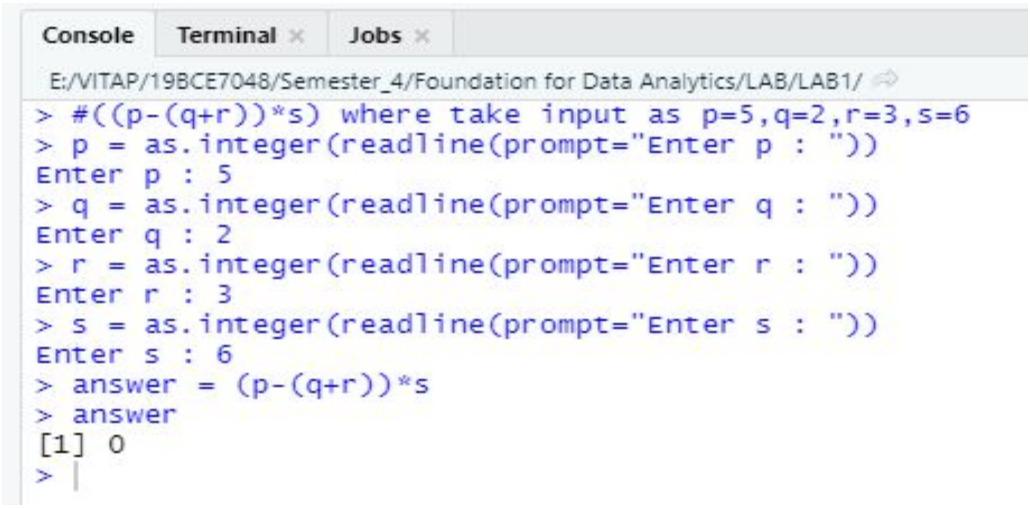
Power : ((p-(q+r))*s) where take input as p=5, q=2, r=3, s=6

> p = as.integer(readline(prompt="Enter p = "))

> q = as.integer(readline(prompt="Enter q = "))

> r = as.integer(readline(prompt="Enter r = "))

> s = as.integer(readline(prompt="Enter s = "))



Q2. Abhisekh is buying the number of baskets where each basket contains n number of eggs. Take input of the number of baskets, the number of eggs in each basket and the cost of each egg. Write a R program for calculating the cost of the total number of eggs.

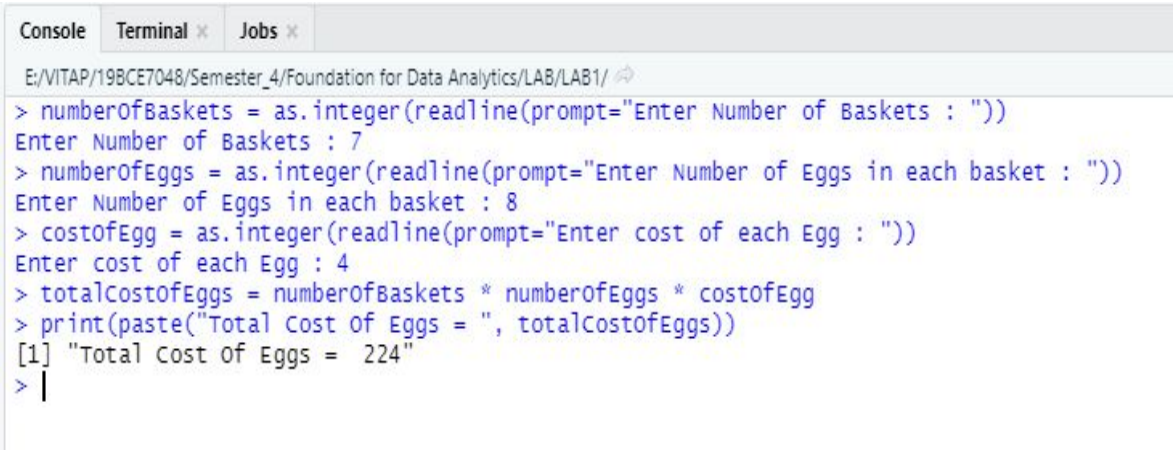
> numberOfBaskets = as.integer(readline(prompt="Enter Number of Baskets : "))

> numberOfEggs = as.integer(readline(prompt="Enter Number of Eggs in each basket : "))

> costOfEgg = as.integer(readline(prompt="Enter cost of each Egg : "))

> totalCostOfEggs = numberOfBaskets * numberOfEggs * costOfEgg

> print(paste("Total Cost of Eggs = ", totalCostOfEggs))

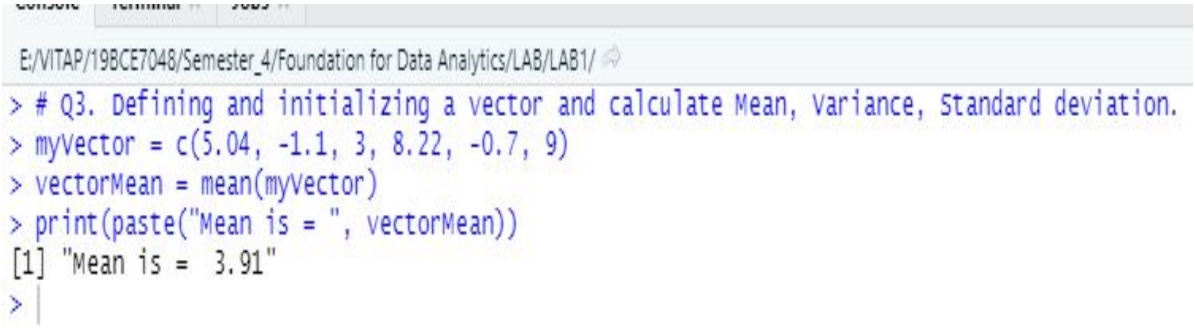


Q3. Defining and initializing a vector and calculating Mean, Variance, Standard deviation.

> myVector = c(5.04, -1.1, 3, 8.22, -0.7, 9)

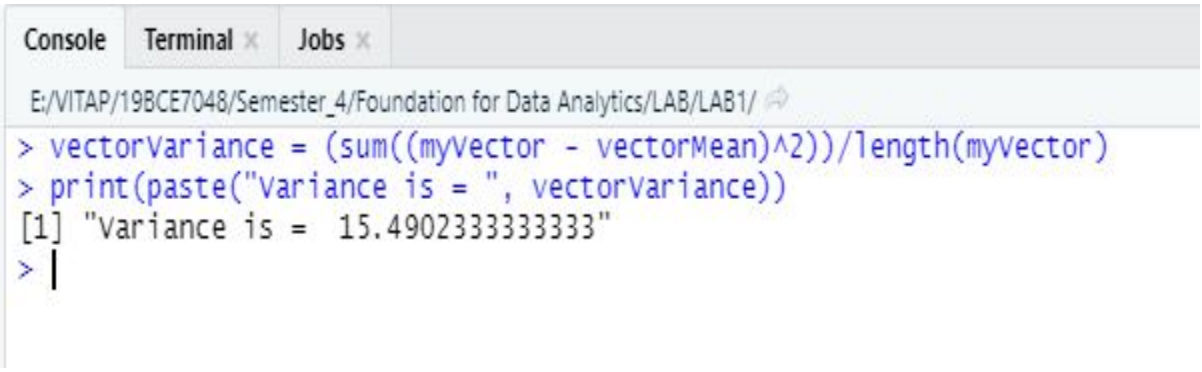
> vectorMean = mean(myVector)

> print(paste("Mean is = ", vectorMean))



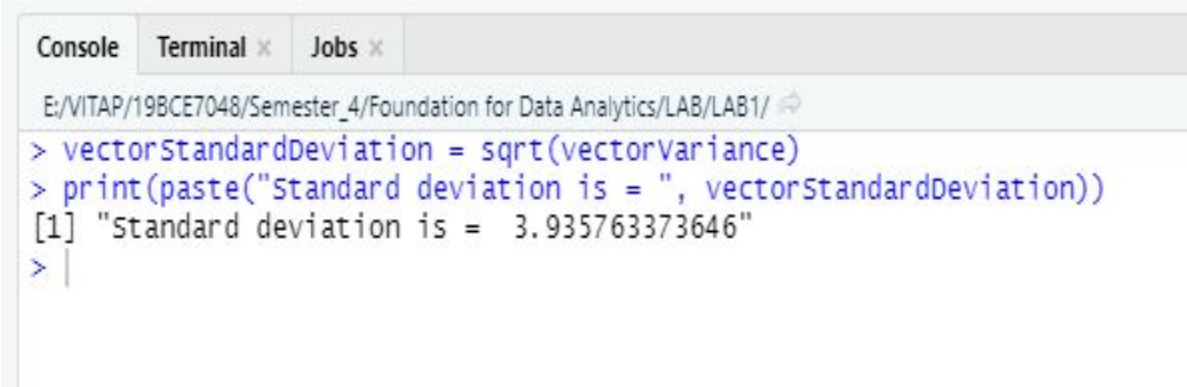
> vectorVariance = (sum((myVector - vectorMean)^2))/length(myVector)

> print(paste("Variance is = ", vectorVariance))



> vectorStandardDeviation = sqrt(vectorVariance)

> print(paste("Standard deviation is = ", vectorStandardDeviation))



Assignment 2: **Date: 09/02/2021**

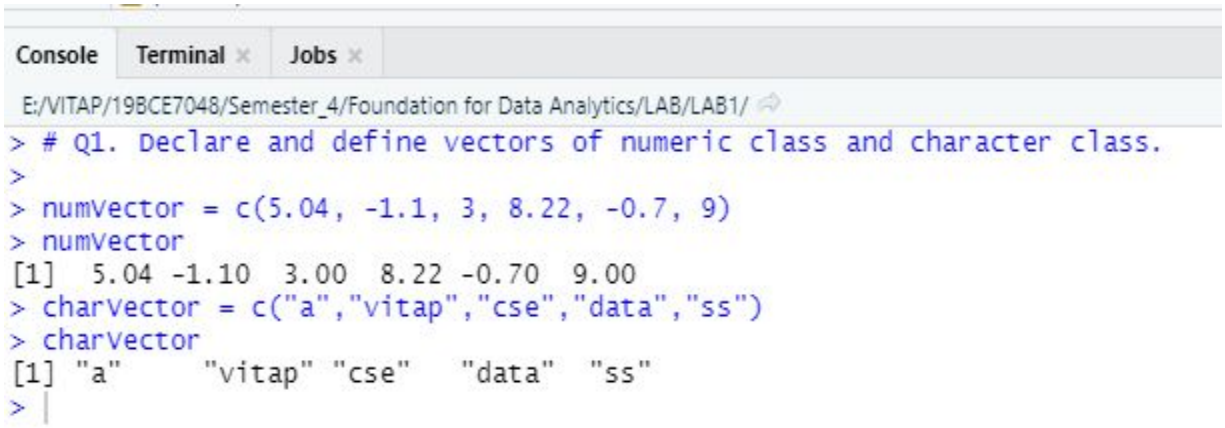
Q1. Declare and define vectors of numeric class and character class.

> numVector = c(5.04, -1.1, 3, 8.22, -0.7, 9)

> numVector

> charVector = c("a","vitap","cse","data","ss")

> charVector



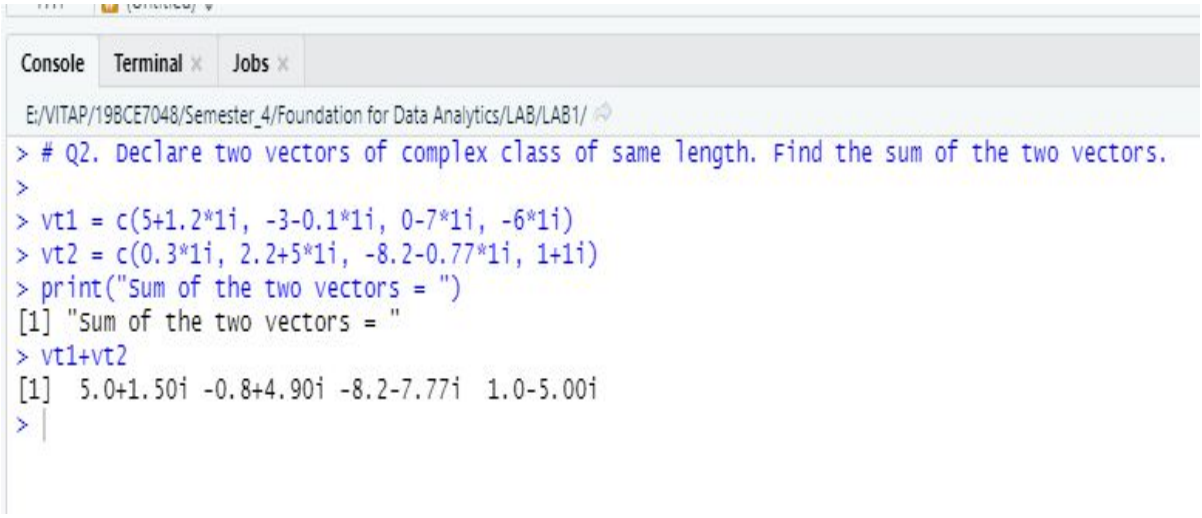
Q2. Declare two vectors of complex class of same length. Find the sum of the two vectors.

> vt1 = c(5+1.2*1i, -3-0.1*1i, 0-7*1i, -6*1i)

> vt2 = c(0.3*1i, 2.2+5*1i, -8.2-0.77*1i, 1+1i)

> print("Sum of the two vectors = ")

> vt1+vt2



Q3. Declare two vectors of numeric class of same length. Find the sum of the two vectors and assign to another vector. Find the average of the resultant vector.

> vect1 = c(5.04, -1.1, 3, -0.7, 9)

> vect2 = c(-1, -0.33, 2.26, 7, -3.1)

> vect3 = vect1+vect2

> print("Sum of the two vectors = ")

> vect3

> avgvect3 = sum(vect3)/length(vect3)

> print(paste("Average of resultant vector = ", avgvect3))

```
>
> vect1 = c(5.04, -1.1, 3, -0.7, 9)
> vect2 = c(-1, -0.33, 2.26, 7, -3.1)
>
> vect3 = vect1+vect2
> print("Sum of the two vectors = ")
[1] "Sum of the two vectors = "
> vect3
[1]  4.04 -1.43  5.26  6.30  5.90
>
> avgvect3 = sum(vect3)/length(vect3)
> print(paste("Average of resultant vector = ", avgvect3))
[1] "Average of resultant vector =  4.014"
>
```

Q4. Declare and define vectors of integer and numeric class. Find the mean, median, mod, variance and standard deviation for both the classes separately.

> intVector = c(5L, -2L, 3L, 5L, -8L, 5L, 2L, 9L, 5L)

> intVector

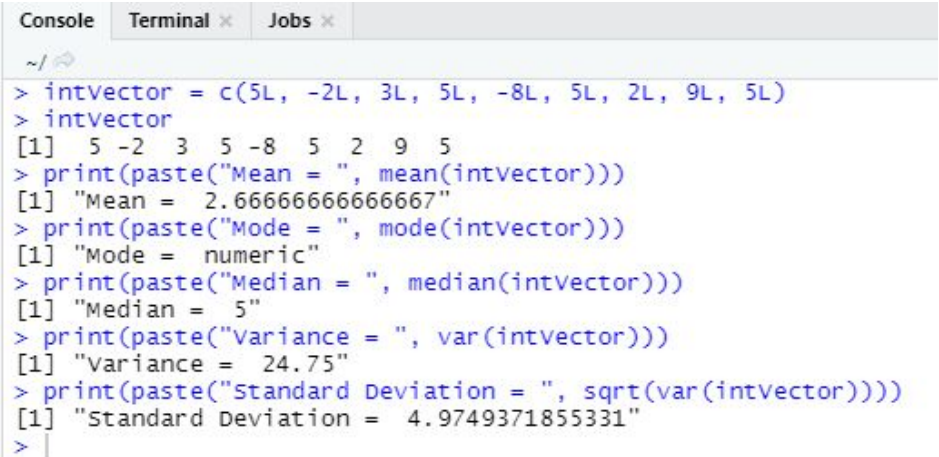
> print(paste("Mean = ", mean(intVector)))

> print(paste("Mode = ", mode(intVector)))

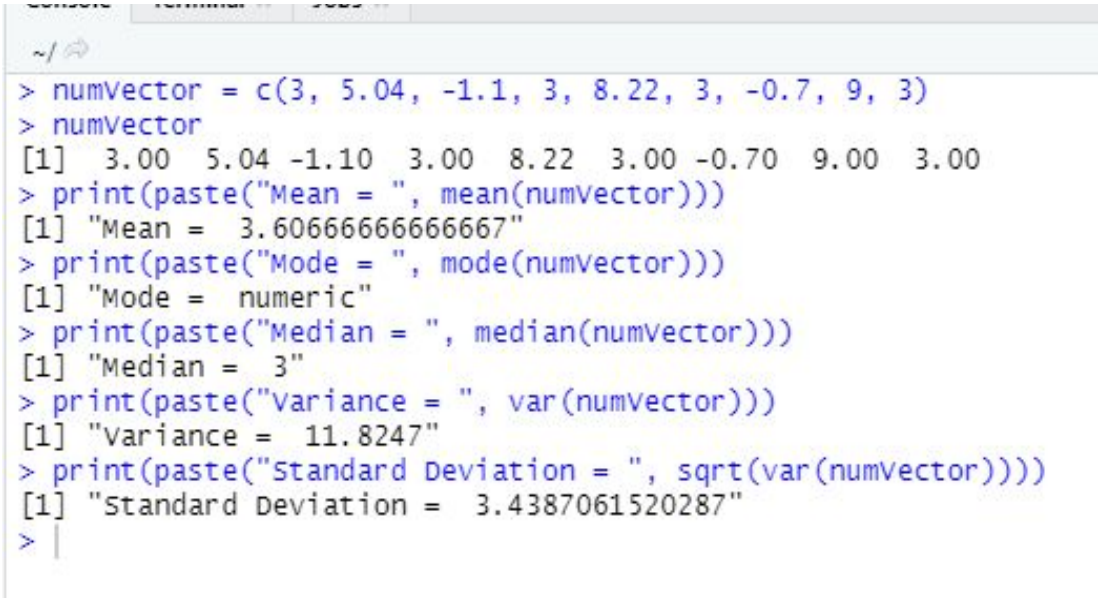
> print(paste("Median = ", median(intVector)))

> print(paste("Variance = ", var(intVector)))

> print(paste("Standard Deviation = ", sqrt(var(intVector))))

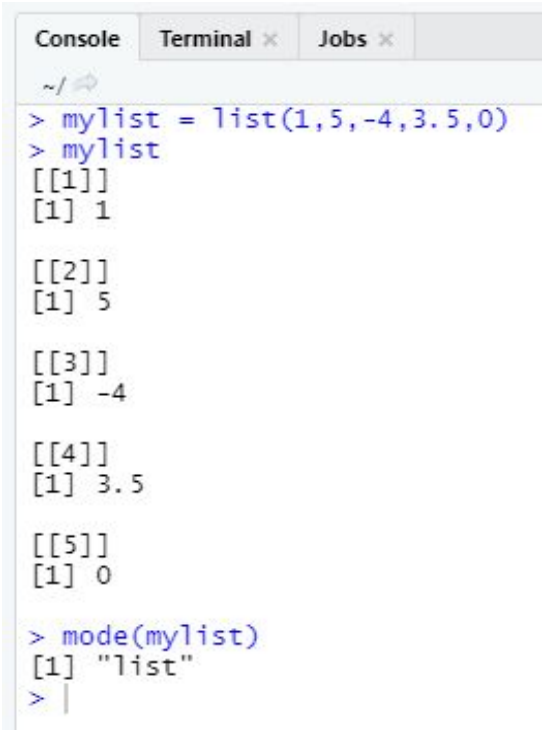


```
> numVector = c(3, 5.04, -1.1, 3, 8.22, 3, -0.7, 9, 3)
> numVector
> print(paste("Mean = ", mean(numVector)))
> print(paste("Mode = ", mode(numVector)))
> print(paste("Median = ", median(numVector)))
> print(paste("Variance = ", var(numVector)))
> print(paste("Standard Deviation = ", sqrt(var(numVector))))
```



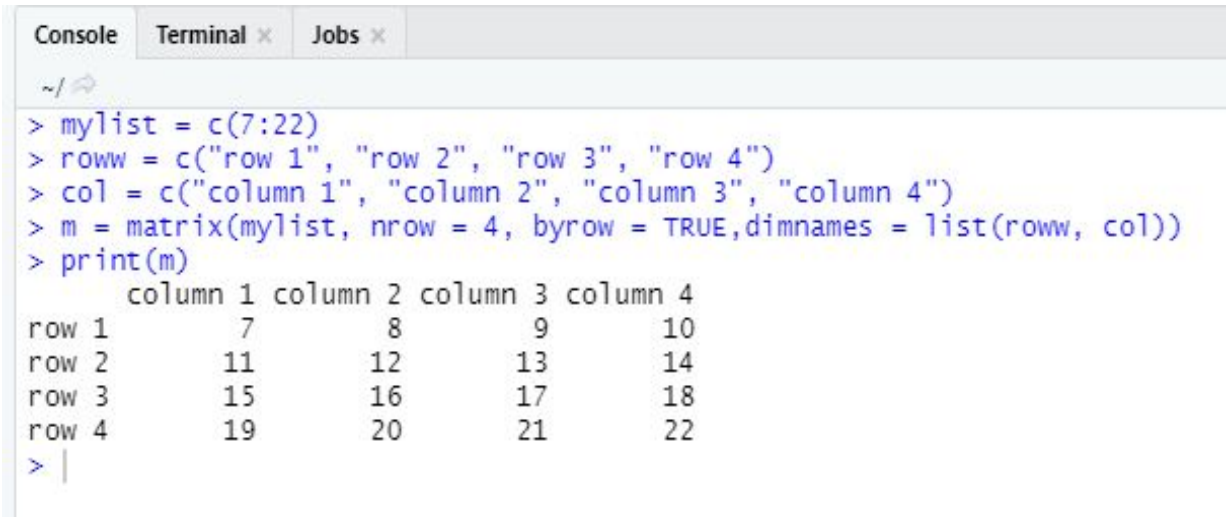
Assignment 3: **Date: 16/02/2021**

Q1. Declare and define list and find the mode of it.



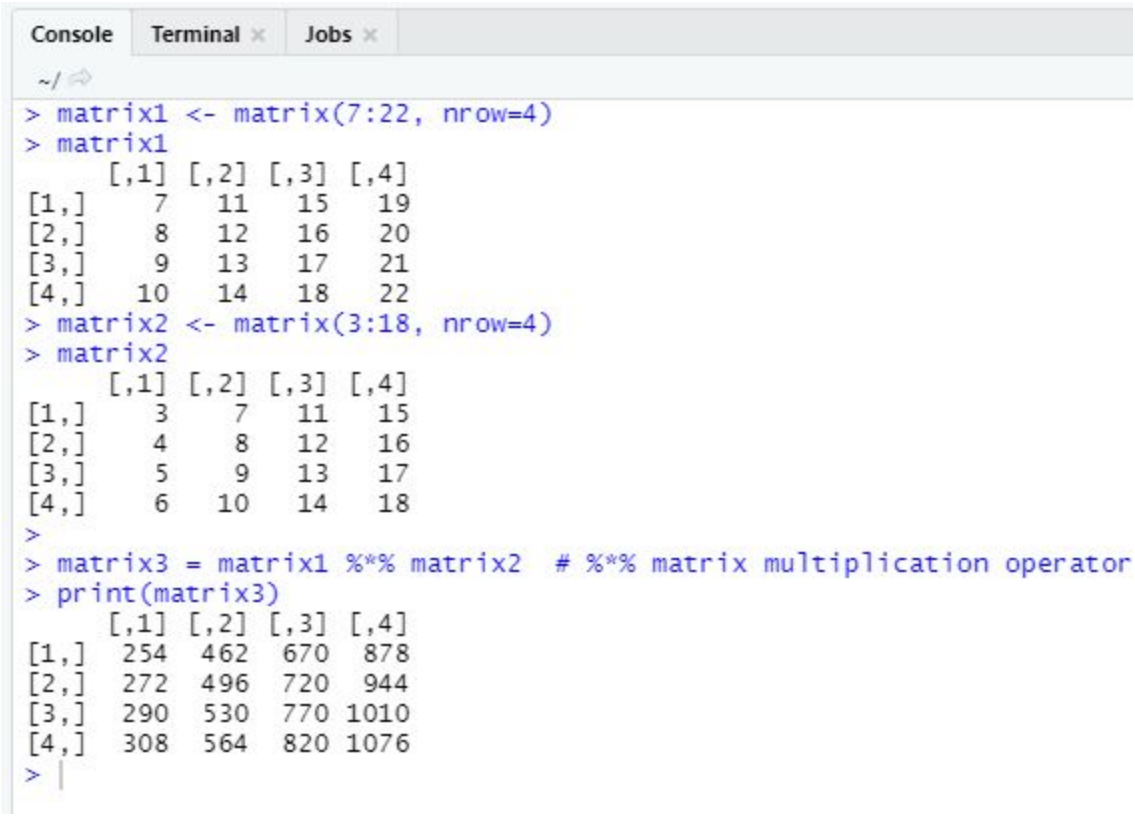
```
Console  Terminal  Jobs
~/
> mylist = list(1,5,-4,3.5,0)
> mylist
[[1]]
[1] 1

[[2]]
[1] 5

[[3]]
[1] -4

[[4]]
[1] 3.5

[[5]]
[1] 0

> mode(mylist)
[1] "list"
>
```

Q2. Declare two vectors and convert it to a matrix.



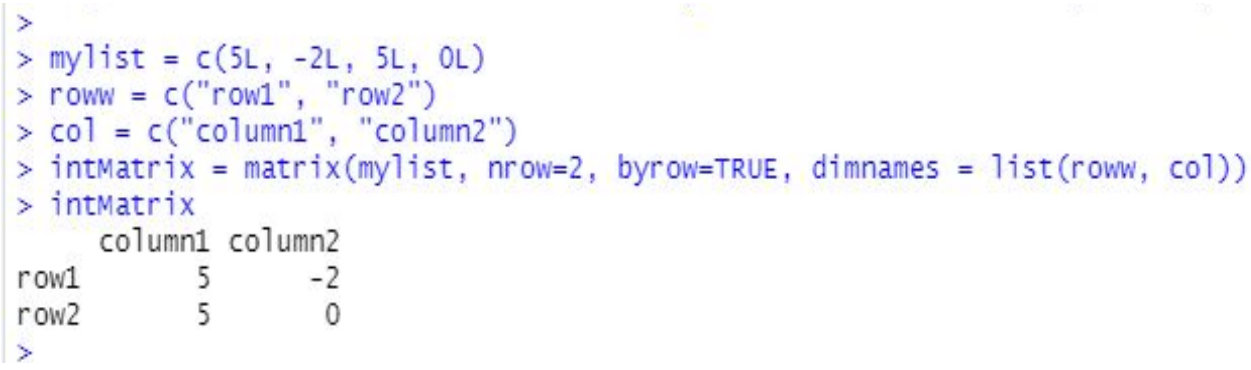
```
Console  Terminal  Jobs
~/
> mylist = c(7:22)
> roww = c("row 1", "row 2", "row 3", "row 4")
> col = c("column 1", "column 2", "column 3", "column 4")
> m = matrix(mylist, nrow = 4, byrow = TRUE,dimnames = list(roww, col))
> print(m)
      column 1 column 2 column 3 column 4
row 1        7        8        9       10
row 2       11       12       13       14
row 3       15       16       17       18
row 4       19       20       21       22
>
```

Q3. Declare two matrices. Find the sum and multiplication of the two matrices and store it in the 3rd matrix.



```
Console   Terminal ×   Jobs ×
~/
> matrix1 <- matrix(7:22, nrow=4)
> matrix1
     [,1] [,2] [,3] [,4]
[1,]    7   11   15   19
[2,]    8   12   16   20
[3,]    9   13   17   21
[4,]   10   14   18   22
> matrix2 <- matrix(3:18, nrow=4)
> matrix2
     [,1] [,2] [,3] [,4]
[1,]    3    7   11   15
[2,]    4    8   12   16
[3,]    5    9   13   17
[4,]    6   10   14   18
>
> matrix3 = matrix1 %*% matrix2  # %*% matrix multiplication operator
> print(matrix3)
     [,1] [,2] [,3] [,4]
[1,]  254  462  670  878
[2,]  272  496  720  944
[3,]  290  530  770 1010
[4,]  308  564  820 1076
>
```

Q4. Declare and define the matrix of integer class. Find the mean, median, mod, variance and standard deviation with respect to row and col.



```
>
> mylist = c(5L, -2L, 5L, 0L)
> roww = c("row1", "row2")
> col = c("column1", "column2")
> intMatrix = matrix(mylist, nrow=2, byrow=TRUE, dimnames = list(roww, col))
> intMatrix
     column1 column2
row1       5      -2
row2       5       0
>
```

Mean :

```
>
> print(paste("Mean of 1st row = ", mean(intMatrix[1,])))
[1] "Mean of 1st row =  1.5"
> print(paste("Mean of 2nd row = ", mean(intMatrix[2,])))
[1] "Mean of 2nd row =  2.5"
>
> print(paste("Mean of 1st column = ", mean(intMatrix[,1])))
[1] "Mean of 1st column =  5"
> print(paste("Mean of 2nd column = ", mean(intMatrix[,2])))
[1] "Mean of 2nd column =  -1"
>
```

Mode :

```
>
> print(paste("Mode of 1st row = ", class(intMatrix[1,])))
[1] "Mode of 1st row =  integer"
> print(paste("Mode of 2nd row = ", class(intMatrix[2,])))
[1] "Mode of 2nd row =  integer"
>
> print(paste("Mode of 1st column = ", class(intMatrix[,1])))
[1] "Mode of 1st column =  integer"
> print(paste("Mode of 2nd column = ", class(intMatrix[,2])))
[1] "Mode of 2nd column =  integer"
>
```

Median :

```
>
> print(paste("Median of 1st row = ", median(intMatrix[1,])))
[1] "Median of 1st row =  1.5"
> print(paste("Median of 2nd row = ", median(intMatrix[2,])))
[1] "Median of 2nd row =  2.5"
>
> print(paste("Median of 1st column = ", median(intMatrix[,1])))
[1] "Median of 1st column =  5"
> print(paste("Median of 2nd column = ", median(intMatrix[,2])))
[1] "Median of 2nd column =  -1"
>
```

Variance :

```
>
> print(paste("Variance of 1st row = ", var(intMatrix[1,])))
[1] "Variance of 1st row =  24.5"
> print(paste("Variance of 2nd row = ", var(intMatrix[2,])))
[1] "Variance of 2nd row =  12.5"
>
> print(paste("Variance of 1st column = ", var(intMatrix[,1])))
[1] "Variance of 1st column =  0"
> print(paste("Variance of 2nd column = ", var(intMatrix[,2])))
[1] "Variance of 2nd column =  2"
>
```

Standard deviation :

```
>
> print(paste("Standard Deviation of 1st row = ", sd(intMatrix[1,])))
[1] "Standard Deviation of 1st row =  4.94974746830583"
> print(paste("Standard Deviation of 2nd row = ", sd(intMatrix[2,])))
[1] "Standard Deviation of 2nd row =  3.53553390593274"
>
> print(paste("Standard Deviation of 1st column = ", sd(intMatrix[,1])))
[1] "Standard Deviation of 1st column =  0"
> print(paste("Standard Deviation of 2nd column = ", sd(intMatrix[,2])))
[1] "Standard Deviation of 2nd column =  1.4142135623731"
>
```

Assignment 4: **Date: 23/02/2021**

Q1. Declare a data frame named as Stu_Result of 15 students with attributes Reg.No, Name, M1, M2, M3, M4 and M5.

```
Console  Terminal  Jobs
E:/VITAP/19BCE7048/Semester_4/Foundation for Data Analytics/LAB/LAB3/
> # Q1. Declare a data frame named as Stu_Result of 15 students with attributes Reg.No, Name, M1, M2, M3, M4 and M5.
>
> RegNo = c(101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115)
> Name = c("Ramesh", "Sakshi", "Krishna", "Siri", "Alex", "Ramya", "Kiran", "Karthik", "Suresh", "Neha", "Rekha", "Mohan", "Prasad", "Radha", "Charan")
> M1 = c(87, 75, 83, 68, 62, 93, 85, 97, 78, 91, 93, 82, 100, 56, 51)
> M2 = c(53, 80, 41, 97, 66, 70, 61, 91, 82, 39, 55, 93, 98, 74, 77)
> M3 = c(66, 38, 44, 60, 50, 72, 78, 74, 97, 47, 33, 30, 96, 83, 53)
> M4 = c(75, 41, 97, 68, 55, 32, 30, 60, 80, 91, 62, 57, 94, 78, 56)
> M5 = c(39, 85, 53, 88, 51, 67, 94, 33, 37, 68, 72, 76, 100, 40, 88)
> Stu_Result = data.frame(RegNo, Name, M1, M2, M3, M4, M5)
> names(Stu_Result) = c('RegNo', 'Name', 'M1', 'M2', 'M3', 'M4', 'M5')
> Stu_Result
   RegNo    Name  M1 M2 M3 M4  M5
1    101  Ramesh  87 53 66 75  39
2    102  Sakshi  75 80 38 41  85
3    103 Krishna  83 41 44 97  53
4    104    Siri  68 97 60 68  88
5    105    Alex  62 66 50 55  51
6    106   Ramya  93 70 72 32  67
7    107   Kiran  85 61 78 30  94
8    108 Karthik  97 91 74 60  33
9    109  Suresh  78 82 97 80  37
10   110    Neha  91 39 47 91  68
11   111   Rekha  93 55 33 62  72
12   112   Mohan  82 93 30 57  76
13   113  Prasad 100 98 96 94 100
14   114   Radha  56 74 83 78  40
15   115  Charan  51 77 53 56  88
>
```

(a) Calculate the Average of Marks in another vector and append it to the Stu_Result.

```
> AverageMark = c(mean(Stu_Result$M1), mean(Stu_Result$M2), mean(Stu_Result$M3), mean(Stu_Result$M4), mean(Stu_Result$M5))
> Stu_Result$AverageMark <- AverageMark
> Stu_Result
   RegNo    Name  M1 M2 M3 M4  M5 AverageMark
1    101  Ramesh  87 53 66 75  39    80.06667
2    102  Sakshi  75 80 38 41  85    71.80000
3    103 Krishna  83 41 44 97  53    61.40000
4    104    Siri  68 97 60 68  88    65.06667
5    105    Alex  62 66 50 55  51    66.06667
6    106   Ramya  93 70 72 32  67    80.06667
7    107   Kiran  85 61 78 30  94    71.80000
8    108 Karthik  97 91 74 60  33    61.40000
9    109  Suresh  78 82 97 80  37    65.06667
10   110    Neha  91 39 47 91  68    66.06667
11   111   Rekha  93 55 33 62  72    80.06667
12   112   Mohan  82 93 30 57  76    71.80000
13   113  Prasad 100 98 96 94 100    61.40000
14   114   Radha  56 74 83 78  40    65.06667
15   115  Charan  51 77 53 56  88    66.06667
>
```

(b) **Then append another column as Grade with respect to the table given below. The resultant data frame name should be Updated_Stu_Result**

Average	Grade
>90	S
81-90	A
71-80	B
61-70	C
51-60	P
<=50	F

Console Terminal Jobs

E:/VITAP/19BCE7048/Semester_4/Foundation for Data Analytics/LAB/LAB3/

```
> Grades = c('C','C', 'C', 'B', 'P', 'C', 'C', 'B', 'B', 'C', 'C', 'C', 'S', 'C', 'C')
>
> Stu_Result$Grades = Grades
> Stu_Result
   RegNo     Name  M1 M2 M3 M4  M5 AverageScore Grades
1    101   Ramesh  87 53 66 75  39         64.0      C
2    102   Sakshi  75 80 38 41  85         63.8      C
3    103 Krishna  83 41 44 97  53         63.6      C
4    104     Siri  68 97 60 68  88         76.2      B
5    105     Alex  62 66 50 55  51         56.8      P
6    106    Ramya  93 70 72 32  67         66.8      C
7    107    Kiran  85 61 78 30  94         69.6      C
8    108 Karthik  97 91 74 60  33         71.0      B
9    109   Suresh  78 82 97 80  37         74.8      B
10   110     Neha  91 39 47 91  68         67.2      C
11   111    Rekha  93 55 33 62  72         63.0      C
12   112    Mohan  82 93 30 57  76         67.6      C
13   113   Prasad 100 98 96 94 100         97.6      S
14   114    Radha  56 74 83 78  40         66.2      C
15   115   Charan  51 77 53 56  88         65.0      C
>
> Updated_Stu_Result = Stu_Result
> Updated_Stu_Result
   RegNo     Name  M1 M2 M3 M4  M5 AverageScore Grades
1    101   Ramesh  87 53 66 75  39         64.0      C
2    102   Sakshi  75 80 38 41  85         63.8      C
3    103 Krishna  83 41 44 97  53         63.6      C
4    104     Siri  68 97 60 68  88         76.2      B
5    105     Alex  62 66 50 55  51         56.8      P
6    106    Ramya  93 70 72 32  67         66.8      C
7    107    Kiran  85 61 78 30  94         69.6      C
8    108 Karthik  97 91 74 60  33         71.0      B
9    109   Suresh  78 82 97 80  37         74.8      B
10   110     Neha  91 39 47 91  68         67.2      C
11   111    Rekha  93 55 33 62  72         63.0      C
12   112    Mohan  82 93 30 57  76         67.6      C
13   113   Prasad 100 98 96 94 100         97.6      S
14   114    Radha  56 74 83 78  40         66.2      C
15   115   Charan  51 77 53 56  88         65.0      C
>
```

(c) Create another data frame named as New_Stu_Result with the attributes Reg.No, Name, M1, M2, M3, M4, M5, Average, and Grade.

```
> New_Stu_Result = data.frame(RegNo, Name, M1, M2, M3, M4, M5, AverageScore, Grades)
> names(New_Stu_Result) = c('RegNo', 'Name', 'M1', 'M2', 'M3', 'M4', 'M5', 'Average', 'Grade')
> New_Stu_Result
   RegNo    Name  M1 M2 M3 M4  M5 Average Grade
1    101  Ramesh  87 53 66 75  39    64.0     C
2    102  Sakshi  75 80 38 41  85    63.8     C
3    103 Krishna  83 41 44 97  53    63.6     C
4    104    Siri  68 97 60 68  88    76.2     B
5    105    Alex  62 66 50 55  51    56.8     P
6    106   Ramya  93 70 72 32  67    66.8     C
7    107   Kiran  85 61 78 30  94    69.6     C
8    108 Karthik  97 91 74 60  33    71.0     B
9    109  Suresh  78 82 97 80  37    74.8     B
10   110    Neha  91 39 47 91  68    67.2     C
11   111   Rekha  93 55 33 62  72    63.0     C
12   112   Mohan  82 93 30 57  76    67.6     C
13   113  Prasad 100 98 96 94 100    97.6     S
14   114   Radha  56 74 83 78  40    66.2     C
15   115  Charan  51 77 53 56  88    65.0     C
>
```

Q2. Create a data frame as given below

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.4	3.4	1.5	0.4	setosa
5.2	4.1	1.5	0.1	setosa
5.8	2.6	4	1.2	versicolor
5.1	3.5	1.4	0.3	setosa
6.3	2.5	4.9	1.5	versicolor
5.7	3.8	1.7	0.3	setosa
6.7	3.1	5.6	2.4	virginica
5.4	3.4	1.7	0.2	setosa
6.7	3.1	4.4	1.4	versicolor
5.5	3.5	1.3	0.2	setosa

Output :

```
Console  Terminal  Jobs
E:/VITAP/19BCE7048/Semester_4/Foundation for Data Analytics/LAB/LAB3/
> Sepal.Length = c(5.4, 5.2, 5.8, 5.1, 6.3, 5.7, 6.7, 5.4, 6.7, 5.5)
> Sepal.Width = c(3.4, 4.1, 2.6, 3.5, 2.5, 3.8, 3.1, 3.4, 3.1, 3.5)
> Petal.Length = c(1.5, 1.5, 4, 1.4, 4.9, 1.7, 5.6, 1.7, 4.4, 1.3)
> Petal.Width = c(0.4, 0.1, 1.2, 0.3, 1.5, 0.3, 2.4, 0.2, 1.4, 0.2)
> Species = c('setosa', 'setosa', 'versicolor','setosa', 'versicolor', 'setosa', 'virginica', 'setosa', 'versicolor', 'setosa')
>
> irisData = data.frame(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, Species)
> names(irisData) = c('Sepal.Length', 'Sepal.Width', 'Petal.Length', 'Petal.Width', 'Species')
> irisData
   Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1           5.4         3.4          1.5         0.4     setosa
2           5.2         4.1          1.5         0.1     setosa
3           5.8         2.6          4.0         1.2 versicolor
4           5.1         3.5          1.4         0.3     setosa
5           6.3         2.5          4.9         1.5 versicolor
6           5.7         3.8          1.7         0.3     setosa
7           6.7         3.1          5.6         2.4  virginica
8           5.4         3.4          1.7         0.2     setosa
9           6.7         3.1          4.4         1.4 versicolor
10          5.5         3.5          1.3         0.2     setosa
```

(a) What are the mean and median of the column “Sepal.Length”?

```
> print(paste("Mean of Sepal.Length = ", mean(irisData$Sepal.Length)))
[1] "Mean of Sepal.Length =  5.78"
> print(paste("Median of Sepal.Length = ", median(irisData$Sepal.Length)))
[1] "Median of Sepal.Length =  5.6"
>
```

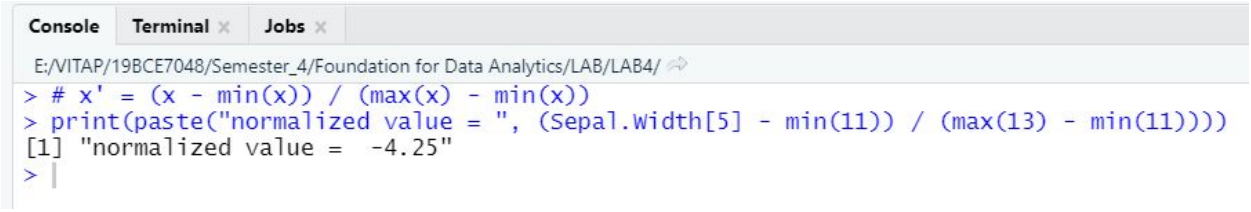
(b) What is the mode of the column “Species”?

```
> # (b) What is the mode of the column “Species”?
> print(paste("Mode of Species = ", mode(irisData$Species)))
[1] "Mode of Species =  character"
>
```

(c) What are standard deviation and variance of the column “Petal.Width”?

```
> # (c) What are standard deviation and variance of the column “Petal.Width”?
> print(paste("Variance of Petal.Width = ", var(irisData$Petal.Width)))
[1] "Variance of Petal.Width =  0.604444444444444"
> print(paste("Standard Deviation of Petal.Width = ", sd(irisData$Petal.Width)))
[1] "Standard Deviation of Petal.Width =  0.77746025264604"
>
```

(d) What is the normalized value of 2.5 in the column “Sepal.Width” using min-max normalization having new minimum value as 11 and new maximum value is 13.



(e) What is the normalized value of 2.5 in the column “Petal.length” using z-score normalization?