SDS | Page No. | Date / /

Bayesian Computing

Name: Ruksha A. Patel
Sapid: 60004210126
Branch: Computer Engineering
Div: C2

Experiment no. 1

Aim: To perform Data Exploration using R.

Theory:

Introduction to R:

R is a rich environment for statistical computing and has many capabilities to explore data in its base package. In addition, R contains a collection of functions for simulating and summarizing the familiar one-parameter probability distributions. It is widely used by statisticians, data scientists and researchers in various fields to analyze data, perform statistical modeling, and create informative visualizations. R's popularity stems from its flexibility, extensive functionality and active community of users.

Core Features of R are as follows:

1. It is an interpreted language
2. It handles vectorised operations efficiently
3. It provides rich statistical functionality.
4. It offers extensive graphics capabilities
5. It has a vibrant community of users of developers.

Some common functions in R used for data exploration are as follows:-

1. One can read data into R by using the 'read.table' command
2. The 'attach()' command is used to make variables visible in the R environment.
3. One can tally different responses for a categorical variable,
4. One can graph label frequencies with a bar graph by the 'barplot' command.

Date / /

5. A simple way to summarize quantitative variables is by the 'summary' command", which gives a variety of descriptive statistics about the variable.

[Note : The task in point 3 can be performed using 'table' command]

R in Bayesian Computing :-

R plays a crucial role in Bayesian computing by providing tools for building models. performing computations, and generating visualizations. R is particularly useful in MCMC implementation, model simulations, Bayesian inferences and ~~diagon~~ diagnostics and visualization of posterior distributions.

Conclusion :-

In this experiment, we used R programming to explore a dataset of students characteristics and preferences. We displayed initial data, counted gender and height frequencies, analyzed and plotted drinking preferences, calculated sleep hours, compared sleep hours by gender, summarized haircut choices, plotted histogram for DuD ownership, and constructed barplots for height variation. This proves the versatility and power of R as a ~~tot~~ tool for data explorations and analysis and for providing variable insights in data. In conclusion, we learned how to use R programming to manipulate, summarize and visualize a student dataset, we also gained some insights into students behavior and preferences.

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Engineering

SAP ID.: 60004210126 Name: Preksha Ashok Patel

Batch: C2-1

Subject: Bayesian Computing Laboratory Semester: VII

Experiment No. 1

AIM:- To Perform Data Exploration using R.

Tasks:

Exploring student's dataset for :

1. Displaying the first ten rows of the data frame.
2. What is your gender?
3. What is your height in inches?
4. Displaying the drinking preferences of the students and plotting their frequencies using a bar chart.
5. Displaying the summary of hours of sleep and plot histogram.
6. Plotting a boxplot of the hours of sleep for each level of gender.
7. Summary of haircut for each level of gender.
8. Constructing a histogram of DVDs and display the summary.
9. Constructing a frequency table of the individual values of DVDs that were observed.
10. Constructing parallel boxplots of the heights using the Gender variable.

Code:

Loading the dataset

```
library('LearnBayes')
data(studentdata)
```

Performing Data Exploration:

```
1.  print(studentdata[1:10,])

2.  table(studentdata$Gender)

3.  table(studentdata$Height)

4.  table(studentdata$Drink)
barplot(table(studentdata$Drink),xlab="Drink",ylab="Count")
```

```
5.  hours.of.sleep = studentdata$WakeUp - studentdata$ToSleep
summary(hours.of.sleep) hist(hours.of.sleep,main="")

6.  boxplot(hours.of.sleep~studentdata$Gender,ylab="Hours of Sleep")

7.  female.Haircut=studentdata$Haircut[studentdata$Gender=="female"]
summary(female.Haircut)
male.Haircut=studentdata$Haircut[studentdata$Gender=="male"]
summary(male.Haircut)

8.  hist(studentdata$Dvds) print(summary(studentdata$Dvds))

9.  print(table(studentdata$Dvds)) barplot(table(studentdata$Dvds))

10. boxplot(studentdata$Height~studentdata$Gender)
```

Output:

1.

	Student	Height	Gender	Shoes	Number	Dvds	ToSleep	WakeUp	Haircut	Job	Drink
1	1	67	female	10	5	10	-2.5	5.5	60	30.0	water
2	2	64	female	20	7	5	1.5	8.0	0	20.0	pop
3	3	61	female	12	2	6	-1.5	7.5	48	0.0	milk
4	4	61	female	3	6	40	2.0	8.5	10	0.0	water
5	5	70	male	4	5	6	0.0	9.0	15	17.5	pop
6	6	63	female	NA	3	5	1.0	8.5	25	0.0	water
7	7	61	female	12	3	53	1.5	7.5	35	20.0	water
8	8	64	female	25	4	20	0.5	7.5	25	0.0	pop
9	9	66	female	30	3	40	-0.5	7.0	30	25.0	water
10	10	65	male	10	7	22	2.5	8.5	12	0.0	milk

2.

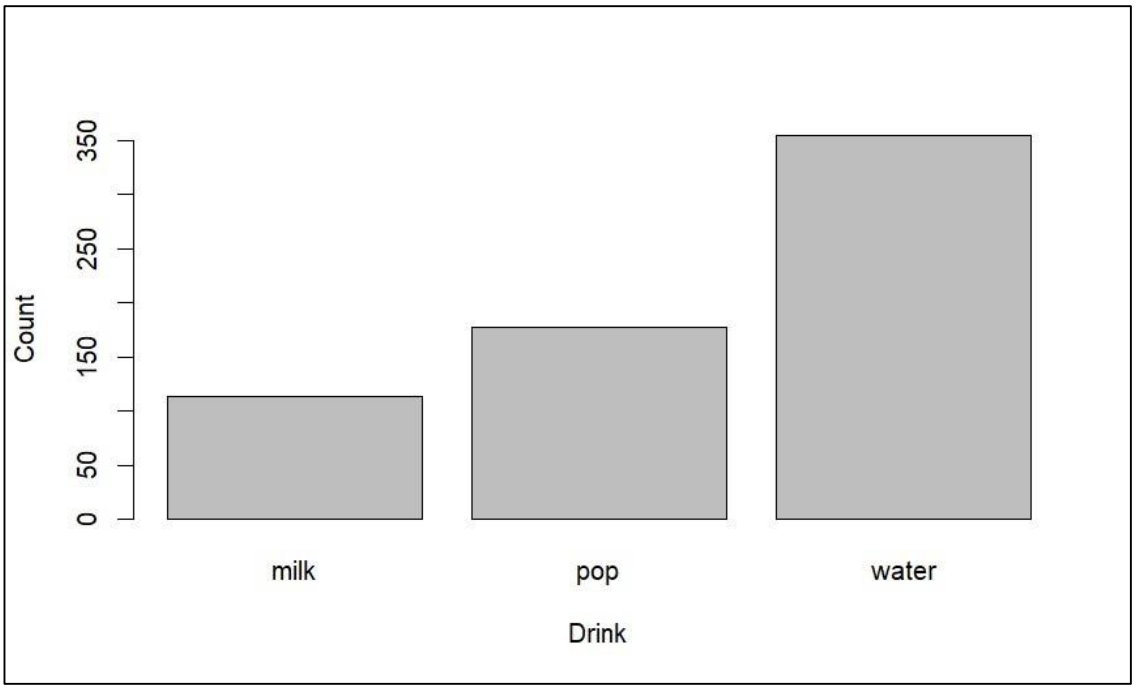
female	male
435	222

3.

54	55	56	57.75	58	58.5	59	59.75	60	60.5	61	61.5	62
2	1	3	1	2	1	3	1	18	3	23	2	49
62.5	63	63.5	64	64.5	65	65.5	66	66.5	66.929	67	67.5	68
2	41	1	66	3	52	2	51	3	1	54	4	45
68.5	69	69.5	70	70.5	71	71.5	71.75	72	72.5	73	74	75
1	28	3	46	2	36	1	1	41	1	17	16	7
75.5	76	77	78	79	84							
1	5	2	2	2	1							

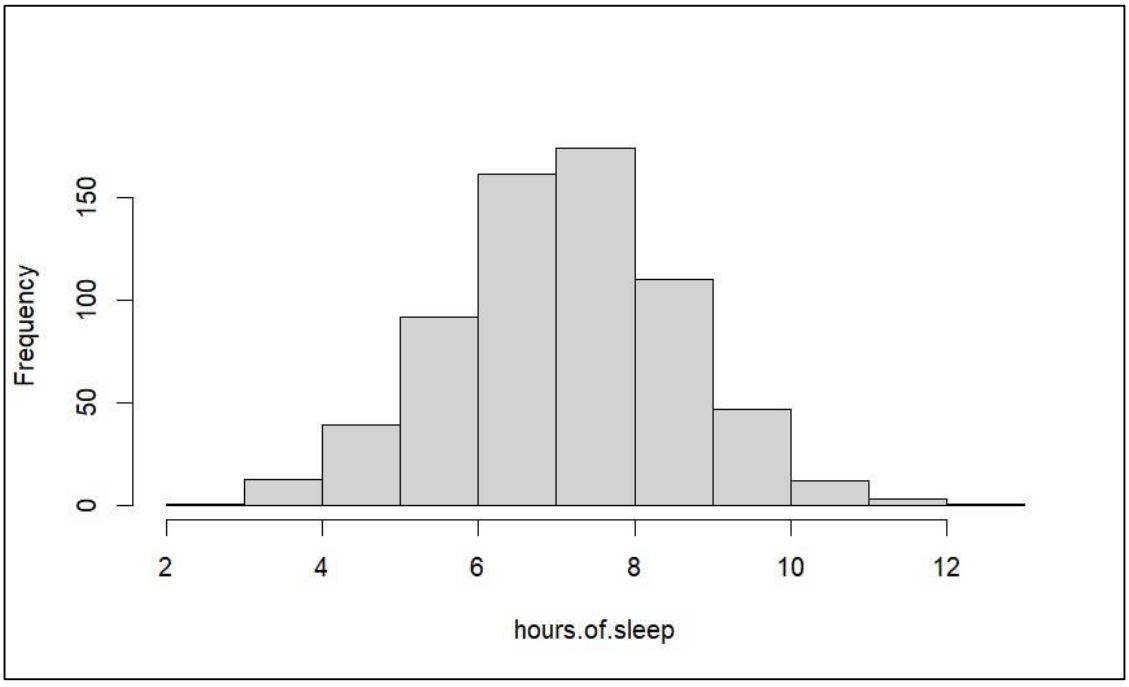
4.

milk	pop	water
113	178	355

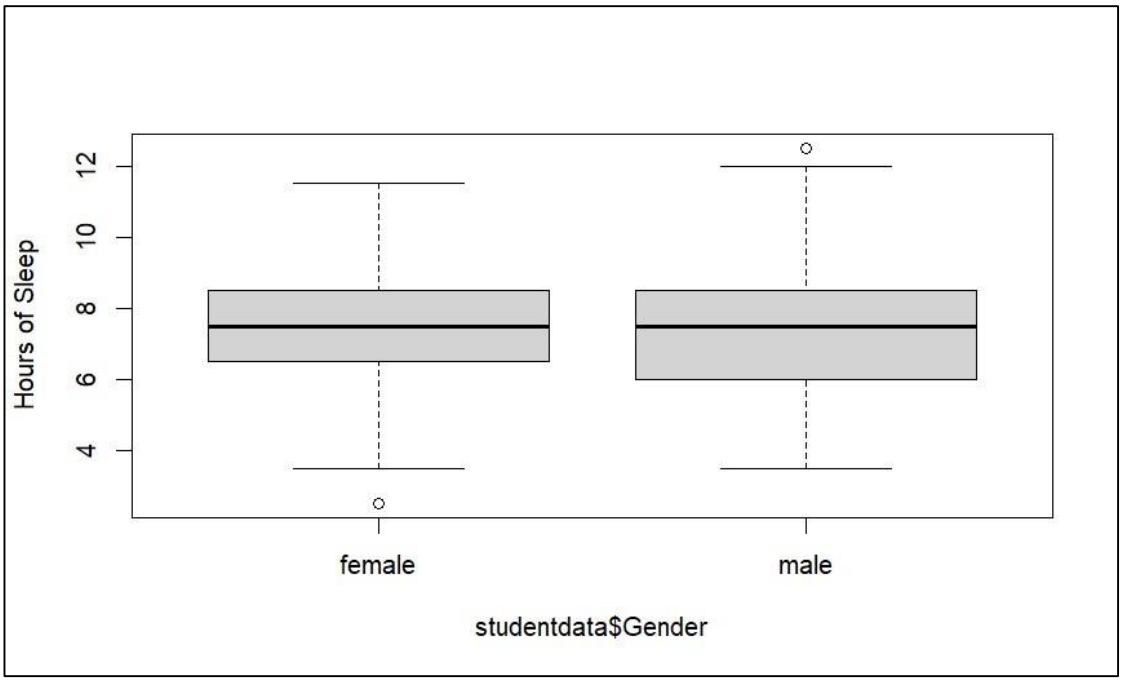


5.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  2.500   6.500   7.500   7.385   8.500  12.500       4
```



6.

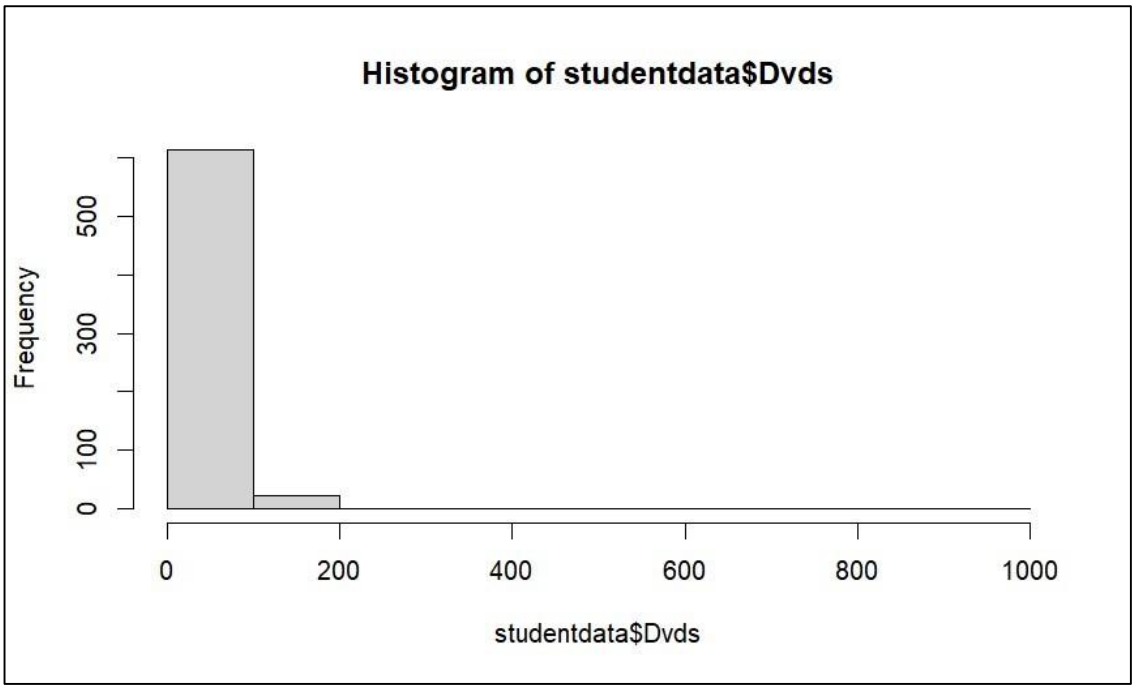


7.

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.00   15.00   25.00   34.08   45.00  180.00      19
```

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.00    0.00   12.00   10.54   15.00   75.00       1
```

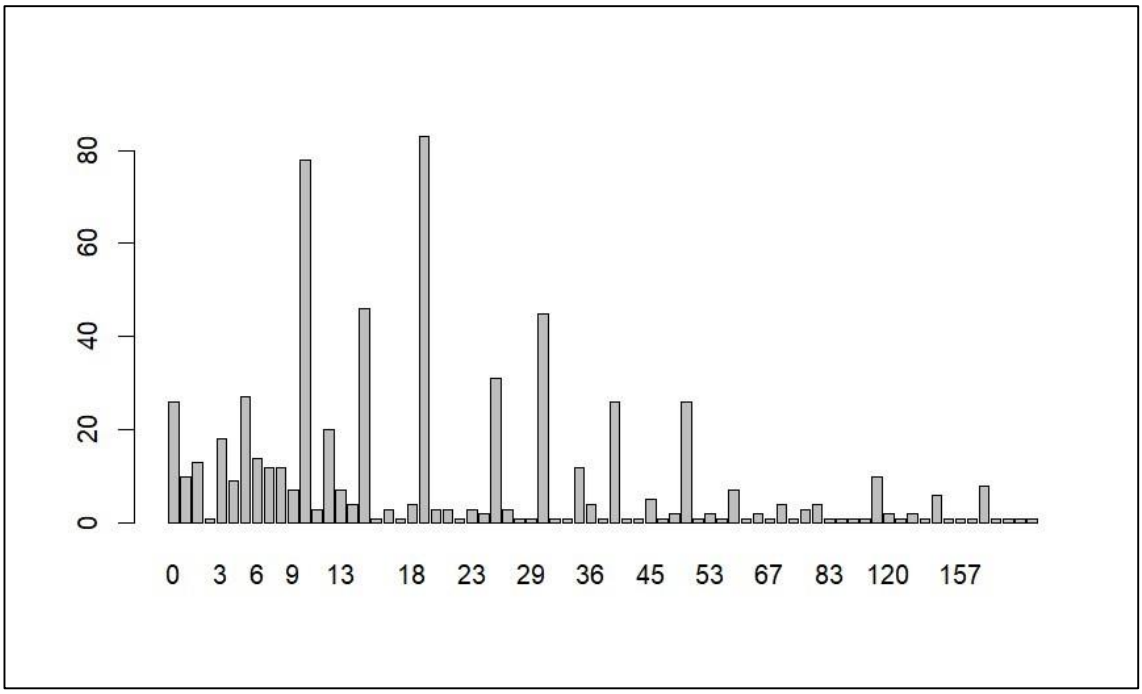
8.



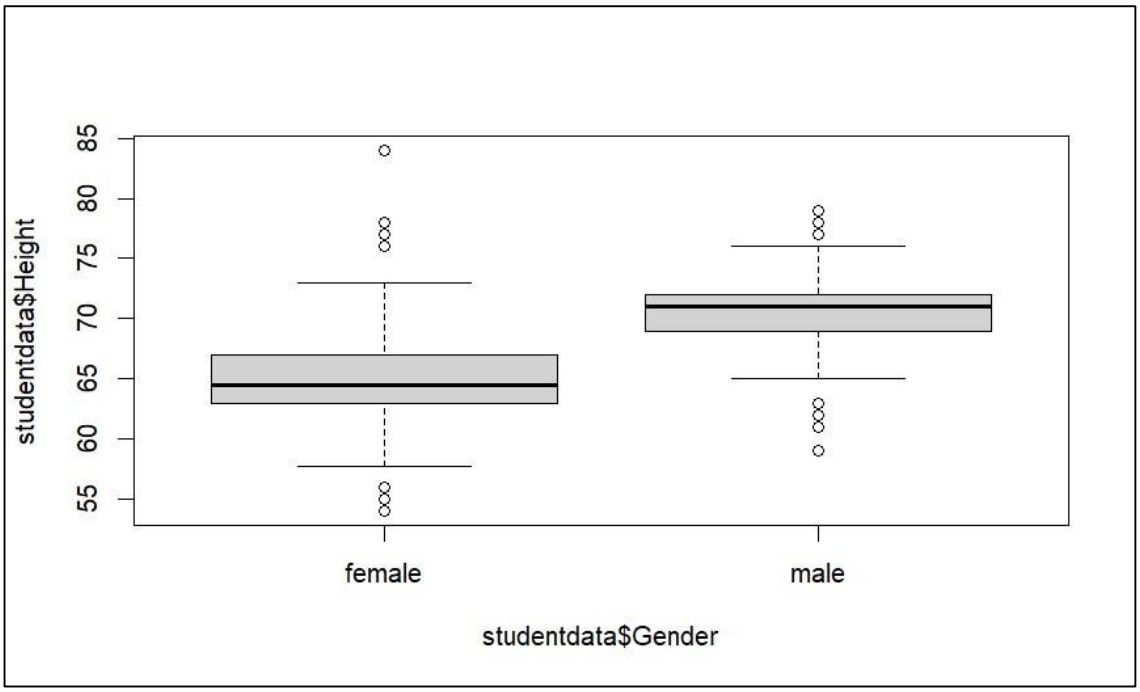

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.00   10.00   20.00   30.93   30.00 1000.00      16
```

9.

```
   0    1    2  2.5    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17
  26   10   13    1   18    9   27   14   12   12    7   78    3   20    7    4   46    1    3
17.5   18   20   21   22 22.5   23   24   25 27.5   28   29   30   31   33   35   36   37   40
   1    4   83    3    3    1    3    2   31    3    1    1   45    1    1   12    4    1   26
  41   42   45   46   48   50   52   53   55   60   62   65   67   70   73   75   80   83   85
   1    1    5    1    2   26    1    2    1    7    1    2    1    4    1    3    4    1    1
  90   97  100  120  122  130  137  150  152  157  175  200  250  500  900 1000
   1    1   10    2    1    2    1    6    1    1    1    8    1    1    1    1
```


0
20
40
60
80
0 3 6 9 13 18 23 29 36 45 53 67 83 120 157

10.


studentdata$Height
55
60
65
70
75
80
85
female
male
studentdata$Gender