Finolex Academy of Management and Technology, Ratnagiri

Department of Information Technology

Subject:	**R Programming Lab. (ITL804)**		
Class:	**BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20**		
Name of Student:	**Kazi Jawwad A Rahim**		
Roll No:	**28**	**Date of performance (DOP) :**	
Assignment/Experiment No:	**01**	**Date of checking (DOC) :**	
Title: Program to demonstrate basic functionality of R such as- data types, characters, strings, factors, helps, accessing packages.			
Marks:		**Teacher's Signature:**	

1. Aim: To understand basics functionality of R software.

2. Prerequisites:

1. Basics of programming disciplines.

3. Hardware Requirements:

1. PC with minimum 2GB RAM

4. Software Requirements:

1. Windows / Linux OS.
2. R version 3.6 or higher

5. Learning Objectives:

1. To understand R software as a software development platform.
2. To understand elementary building blocks of R software such as- data types, character, string, factors, helps, packages.

6. Learning Objectives Applicable: LO 1

7. Program Outcomes Applicable: PO 1

8. Program Education Objectives Applicable: PEO 1

OUTPUT:

Data Types:

```
1)      x=5
        mode(x)
        >> numeric
2)      x=5.5
        mode(x)
        >> numeric
3)      x=”Jawwad”
        mode(x)
        >> character
4)      x=TRUE
        mode(x)
        >> logical
5)      x=6+4i
        mode(x)
        >> complex
6)      x=’Jawwad’
        mode(x)
        >> character
```

Relational Operators:

```
A=6     B=8
> A>B
[1] FALSE
> A>=B
[1] FALSE
> A<B
[1] TRUE
> A<=B
[1] TRUE
> A==B
[1] FALSE
> A!=B
[1] TRUE
```

Arithmetic Operators:

```
A=6     B=8
> A+B
[1] 14
> A-B
[1] -2
> A*B
[1] 48
> A/B
[1] 0.75
> A%%B
[1] 6
> A%/%B
[1] 0
```

Logical Operators:

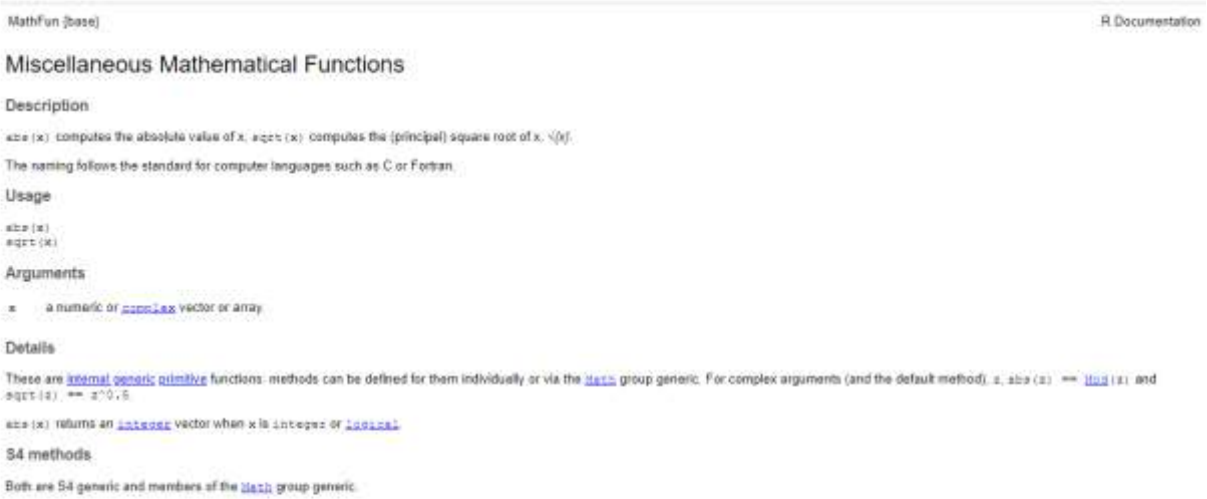
```
> A&B
[1] TRUE
> A&&B
[1] TRUE
> A||B
[1] TRUE
> A|B
[1] TRUE
```

Factors:

```
> d=c(4,1,6)
>
f=factor(d,levels=1:7,labels=c("Monday","Tuesday","Wednesday","Thursday","Friday","Saturday","Sunday
"))
> f[1]
[1] Thursday
Levels: Monday Tuesday Wednesday Thursday Friday Saturday Sunday
```

Help:

help(sqrt)



MathFun {base} R Documentation

Miscellaneous Mathematical Functions

Description

`abs(x)` computes the absolute value of x, `sqrt(x)` computes the (principal) square root of x, √{x}.

The naming follows the standard for computer languages such as C or Fortran.

Usage

```
abs(x)
sqrt(x)
```

Arguments

`x` a numeric or `complex` vector or array.

Details

These are internal generic primitive functions: methods can be defined for them individually or via the `Math` group generic. For complex arguments (and the default method), `z`, `abs(z) == Mod(z)` and `sqrt(z) == z^0.5`.

`abs(x)` returns an `integer` vector when x is `integer` or `logical`.

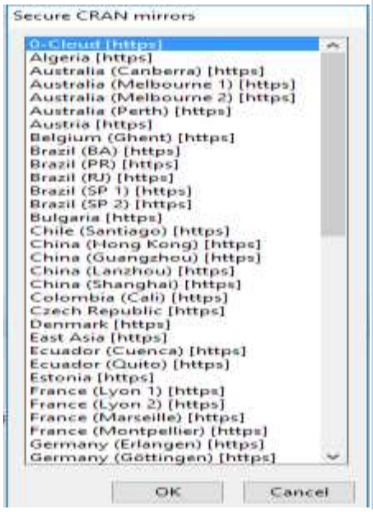
S4 methods

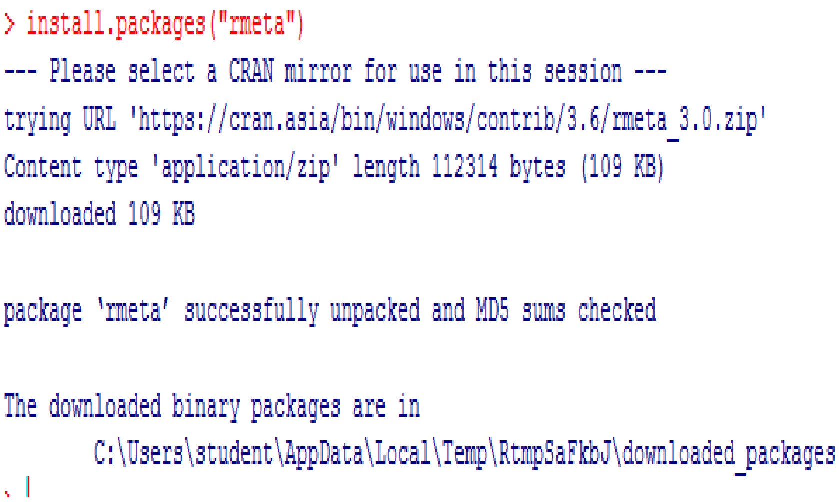
Both are S4 generic and members of the `Math` group generic.

Packages:

```
> install.packages("rmeta")
```

Select mirror





```
> install.packages("rmeta")
--- Please select a CRAN mirror for use in this session ---
trying URL 'https://cran.asia/bin/windows/contrib/3.6/rmeta_3.0.zip'
Content type 'application/zip' length 112314 bytes (109 KB)
downloaded 109 KB


package 'rmeta' successfully unpacked and MD5 sums checked


The downloaded binary packages are in
        C:\Users\student\AppData\Local\Temp\RtmpSaFkbJ\downloaded_packages
> |
```

Learning Outcomes:

1. We understood R software as a software development platform.
2. We understood elementary building blocks of R software such as- data types, character, string, factors, helps, packages.

Conclusion:

We have successfully demonstrated installation of R along with introduction to R and basic building blocks of R.

13. Experiment/Assignment Evaluation

<table>
<tr><th colspan="6">Experiment/Assignment Evaluation:</th></tr>
<tr><th>Sr. No.</th><th colspan="3">Parameters</th><th>Marks obtained</th><th>Out of</th></tr>
<tr><td><b>1</b></td><td colspan="3">Technical Understanding (Assessment may be done based on Q &amp; A <b><u>or</u></b> any other relevant method.) Teacher should mention the other method used -</td><td></td><td>6</td></tr>
<tr><td><b>2</b></td><td colspan="3">Neatness/presentation</td><td></td><td>2</td></tr>
<tr><td><b>3</b></td><td colspan="3">Punctuality</td><td></td><td>2</td></tr>
<tr><td colspan="2"><b>Date of performance (DOP)</b></td><td></td><td><b>Total marks obtained</b></td><td></td><td><b>10</b></td></tr>
<tr><td colspan="2"><b>Date of checking (DOC)</b></td><td></td><td colspan="3"><b>Signature of teacher</b></td></tr>
</table>

References:

1. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What is R?
2. How is R different than Python?
3. What are different data-types in R?
4. How to define a string in R?
5. What is factor data class in R?
6. How to take help in R?
7. How to load packages and libraries in R?

Finolex Academy of Management and Technology, Ratnagiri

Department of Information Technology

Subject:	**R Programming Lab. (ITL804)**		
Class:	**BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20**		
Name of Student:	**Kazi Jawwad A Rahim**		
Roll No:	**28**	**Date of performance (DOP) :**	
Assignment/Experiment No:	**02**	**Date of checking (DOC) :**	
Title: Program to demonstrate data structures such as- vectors, matrix, list and data frames.			
Marks:		**Teacher’s Signature:**	

1. Aim: To understand the use of vectors, matrix, list and data frames in R.

2. Prerequisites:

1. Basics of R programming.

3. Hardware Requirements:

1. PC with minimum 2GB RAM

4. Software Requirements:

1. Windows / Linux OS.
2. R version 3.6 or higher

5. Learning Objectives:

1. To understand vectors, matrix and lists.
2. To understand *data frames* which are mainly required for data analysis in R.

6. Learning Objectives Applicable: LO 1, LO 2

7. Program Outcomes Applicable: PO 1

8. Program Education Objectives Applicable: PEO 1, PEO 2

Vectors:

```
> x=c(1,2,3,4,5,6)
> x
[1] 1 2 3 4 5 6
> x=1:7
> x
[1] 1 2 3 4 5 6 7
```

Matrix:

```
A=matrix(nrow=2,ncol=3,data=c(9,2,1,7,5,4))
print(A)
B=t(A)
print(B)
print(A%*%B)
```

OUTPUT:

```
> source('G:/Practicals/R/EXP2/Second.R')
     [,1] [,2] [,3]
[1,]    9    1    5
[2,]    2    7    4
     [,1] [,2]
[1,]    9    2
[2,]    1    7
[3,]    5    4
     [,1] [,2]
[1,]  107   45
[2,]   45   69
```

List:

```
a=list(3,1,"Hello",4.1,TRUE,c(3,1,5),-3+4i)
print(a[[1]])
```

OUTPUT:

```
> source('G:/Practicals/R/EXP2/Second.R')
[1] 3
```

Data Frames:

```
fr=data.frame(1:3,c("Mahesh","Ganesh","Mangesh"),c(21,22,23))
colnames(fr)=c("Roll No.","Name","Age")
print(fr)
print(rownames(fr))
```

OUTPUT:

```
> source('G:/Practicals/R/EXP2/Second.R')
  Roll No.    Name Age
1        1  Mahesh  21
2        2  Ganesh  22
3        3 Mangesh  23
[1] "1" "2" "3"
```

Learning Outcomes Achieved:

1. We understood vectors, matrix and lists.
2. We understood *data frames* which are mainly required for data analysis in R.

Conclusion:

We have successfully demonstrated vectors, matrix, list and data frames in R.

13. Experiment/Assignment Evaluation

<table>
<tr><td colspan="6"><b>Experiment/Assignment Evaluation:</b></td></tr>
<tr><td><b>Sr. No.</b></td><td colspan="3"><b>Parameters</b></td><td><b>Marks obtained</b></td><td><b>Out of</b></td></tr>
<tr><td><b>1</b></td><td colspan="3">Technical Understanding (Assessment may be done based on Q & A <b>or</b> any other relevant method.) Teacher should mention the other method used -</td><td></td><td>6</td></tr>
<tr><td><b>2</b></td><td colspan="3">Neatness/presentation</td><td></td><td>2</td></tr>
<tr><td><b>3</b></td><td colspan="3">Punctuality</td><td></td><td>2</td></tr>
<tr><td colspan="2"><b>Date of performance (DOP)</b></td><td></td><td><b>Total marks obtained</b></td><td></td><td><b>10</b></td></tr>
<tr><td colspan="2"><b>Date of checking (DOC)</b></td><td></td><td colspan="3"><b>Signature of teacher</b></td></tr>
</table>

References:

1. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What is vector in R ?
2. How to create matrix in R ?
3. What is difference between vector and list?
4. How is the data-frame different than matrix?
5. What is importance of data-frames in R?

Finolex Academy of Management and Technology, Ratnagiri

Department of Information Technology

Subject:	**R Programming Lab. (ITL804)**		
Class:	**BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20**		
Name of Student:	**Kazi Jawwad A Rahim**		
Roll No:	**28**	**Date of performance (DOP) :**	
Assignment/Experiment No:	**03**	**Date of checking (DOC) :**	
Title: Program to demonstrate flow control instructions and functions			
Marks:		**Teacher's Signature:**	

1. Aim: To understand the use of various flow control instructions and functions in R.

2. Prerequisites:

1. Basics of R programming, various data structures used in R etc.

3. Hardware Requirements:

1. PC with minimum 2GB RAM

4. Software Requirements:

1. Windows / Linux OS.
2. R version 3.6 or higher

5. Learning Objectives:

1. To understand decision and loop control instructions.
2. To understand function definition and calling to it.

6. Learning Objectives Applicable: LO 1

7. Program Outcomes Applicable: PO 1, PO 2

8. Program Education Objectives Applicable: PEO 2

OUTPUT:

IF ELSE Example:

```
age=as.numeric(readline("Enter age"))
gender=readline("Enter Gender")
if(age>=60 && gender=="M"){
  print("Available for Concession")
}else if(age>=45 && gender=="F"){
  print("Available for Concession");
}else{
  print("Noot avaialable for Concession")
}
```

OUTPUT:

```
> source('D:/JK/If Else.R')
Enter age60
Enter GenderM
[1] "Available for Concession"
```

SWITCH:

```
day=as.numeric(readline("Enter Day Number\n"))
y=switch(day,"Monday","Tuesday","Wednesday","Thursday","Friday","Saturday","Sunday")
print(y)
```

OUTPUT:

```
> source('D:/JK/Switch.R')
Enter Day Number
5
[1] "Friday"
```

For:

```
for(i in 1:10){
  print(i)
}
```

OUTPUT:

```
> source('D:/JK/For.R')
[1] 1
[1] 2
[1] 3
[1] 4
[1] 5
[1] 6
[1] 7
[1] 8
[1] 9
[1] 10
```

While:

```
i=1
while(i<=5){
  print(i)
  i=i+1
}
```

OUTPUT:

```
> source('D:/JK/While.R')
[1] 1
[1] 2
```

```
[1] 3
[1] 4
[1] 5
```

Repeat:

```
i=1
repeat{
  print(i)
  i=i+1
  if(i>5){
    break
  }
}
```

OUTPUT:

```
> source('D:/JK/Repeat.R')
[1] 1
[1] 2
[1] 3
[1] 4
[1] 5
```

Function:

```
area = function(l,w){
  a=l*w
  return(a)
}
print(area(3,5))
```

OUTPUT:

```
> source('D:/JK/Function.R')
[1] 15
```

Double Function:

```
volume=function(r,l){
  area=function(r){
    a=r*r
    return(a)
  }
  v=area(l)*3.14*r*l
  return(v)
}
print(volume(3,5))
```

OUTPUT:

```
> source('D:/JK/Double Function.R')
[1] 1177.5
```

Learning Outcomes Achieved:

1. We understood decision and loop control instructions.
2. We understood function definition and calling to it.

Conclusion:

We have successfully demonstrated the loop instructions like If-Else, Switch, For, While, Repeat and functions and double functions.

13. Experiment/Assignment Evaluation

<table>
<tr><td colspan="5"><b>Experiment/Assignment Evaluation:</b></td></tr>
<tr><td><b>Sr. No.</b></td><td colspan="2"><b>Parameters</b></td><td><b>Marks obtained</b></td><td><b>Out of</b></td></tr>
<tr><td><b>1</b></td><td colspan="2">Technical Understanding (Assessment may be done based on Q & A <b><u>or</u></b> any other relevant method.) Teacher should mention the other method used -</td><td></td><td>6</td></tr>
<tr><td><b>2</b></td><td colspan="2">Neatness/presentation</td><td></td><td>2</td></tr>
<tr><td><b>3</b></td><td colspan="2">Punctuality</td><td></td><td>2</td></tr>
<tr><td><b>Date of performance (DOP)</b></td><td></td><td><b>Total marks obtained</b></td><td></td><td><b>10</b></td></tr>
<tr><td><b>Date of checking (DOC)</b></td><td></td><td colspan="3"><b>Signature of teacher</b></td></tr>
</table>

References:

1. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What are decision control instructions ?
2. What are loop control instructions ?
3. Compare flow control instructions in R with flow control instructions in Python ?
4. How to define function in R?
5. Can I shuffle arguments of the functions while calling it?

Finolex Academy of Management and Technology, Ratnagiri

Department of Information Technology

Subject:	**R Programming Lab. (ITL804)**		
Class:	**BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20**		
Name of Student:	**Kazi Jawwad A Rahim**		
Roll No:	**28**	**Date of performance (DOP) :**	
Assignment/Experiment No:	**04**	**Date of checking (DOC) :**	
Title: Exploratory data analysis such as- Range, summary, mean, variance, median, standard deviation, histogram, boxplot, scatterplot			
Marks:		**Teacher's Signature:**	

1. Aim: To understand the exploratory data analysis and the methods required to do it in R.

2. Prerequisites:

1. Basics of R programming, various data structures, functions etc.

3. Hardware Requirements:

1. PC with minimum 2GB RAM

4. Software Requirements:

1. Windows / Linux OS.
2. R version 3.6 or higher

5. Learning Objectives:

1. To understand decision and loop control instructions.
2. To understand function definition and calling to it.

6. Learning Objectives Applicable: LO 3. LO 4

7. Program Outcomes Applicable: PO 2, PO 3

8. Program Education Objectives Applicable: PEO 2, PEO 3

Range:

```
a=c(1,5,7,8,9,6,4,9,5,8,9,6)
range(a)
```

OUTPUT:

```
[1] 1 9
```

Summary:

```
a=c(1,5,7,8,9,6,4,9,5,8,9,6)
summary(a)
```

OUTPUT:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   5.000   6.500      6.417   8.250    9.000
```

Mean:

```
a=c(1,5,7,8,9,6,4,9,5,8,9,6)
mean(a)
```

OUTPUT:

```
[1] 6.416667
```

Mode:

```
a=c(1,5,7,8,9,6,4,9,5,8,9,6)
table(a)
```

OUTPUT:

```
a
1 4 5 6 7 8 9
1 1 2 2 1 2 3
=> Mode=9
```

Median:

```
a=c(1,5,7,8,9,6,4,9,5,8,9,6)
median(a)
```

OUTPUT:

```
[1] 6.5
```

Variance:

```
a=c(1,5,7,8,9,6,4,9,5,8,9,6)
var(a)
```

OUTPUT:

```
[1] 5.901515
```

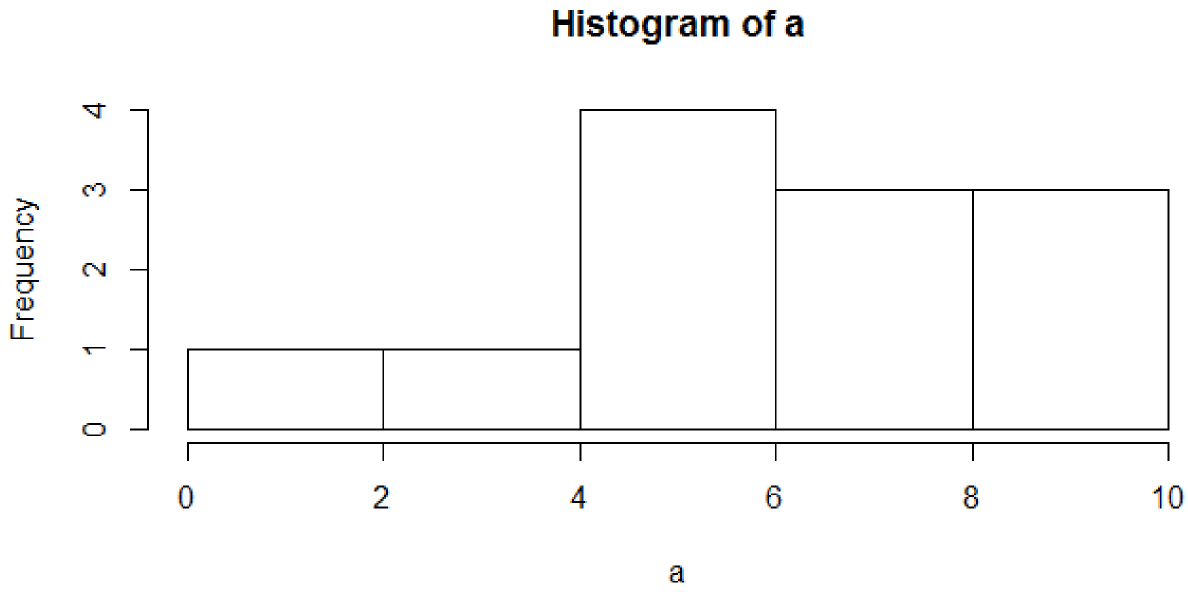
Standard Deviation:

```
a=c(1,5,7,8,9,6,4,9,5,8,9,6)
sqrt(var(a))
```

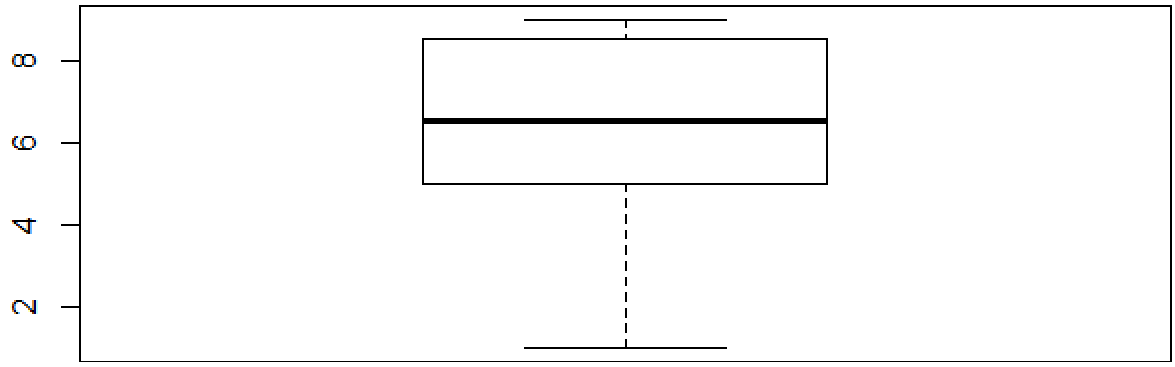
OUTPUT:

```
[1] 2.429303
```

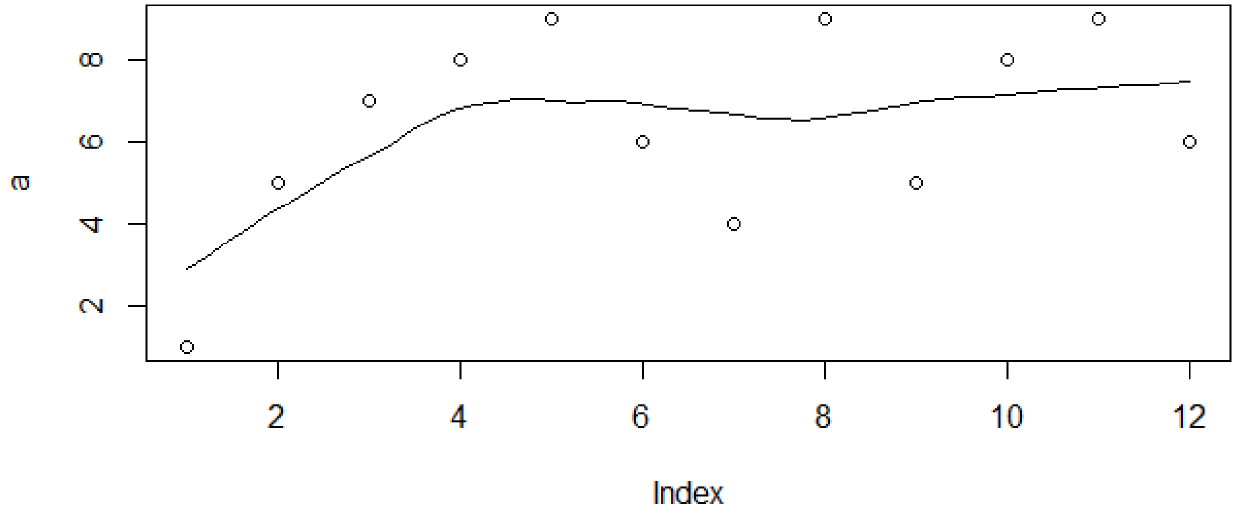
Histogram:



Boxplot:



Scatterplot:



Learning Outcomes Achieved:

1. We understood decision and loop control instructions.
2. We understood the function definition and it's calling.

Conclusion:

We have successfully demonstrated the data analysis such as- Range, summary, mean, variance, median, standard deviation, histogram, boxplot, scatterplot.

13. Experiment/Assignment Evaluation

<table>
<tr><td colspan="6"><b>Experiment/Assignment Evaluation:</b></td></tr>
<tr><td><b>Sr. No.</b></td><td colspan="3"><b>Parameters</b></td><td><b>Marks obtained</b></td><td><b>Out of</b></td></tr>
<tr><td><b>1</b></td><td colspan="3">Technical Understanding (Assessment may be done based on Q & A <b><u>or</u></b> any other relevant method.) Teacher should mention the other method used -</td><td></td><td>6</td></tr>
<tr><td><b>2</b></td><td colspan="3">Neatness/presentation</td><td></td><td>2</td></tr>
<tr><td><b>3</b></td><td colspan="3">Punctuality</td><td></td><td>2</td></tr>
<tr><td colspan="2"><b>Date of performance (DOP)</b></td><td></td><td><b>Total marks obtained</b></td><td></td><td><b>10</b></td></tr>
<tr><td colspan="2"><b>Date of checking (DOC)</b></td><td></td><td colspan="3"><b>Signature of teacher</b></td></tr>
</table>

References:

1. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What is exploratory data analysis ?
2. What is summary of the data ?
3. What is importance of median of the data collection ?
4. What is histogram? Why is it important in data?
5. What information does the box plot provides?
6. List various R library functions used in exploratory data analysis.

Finolex Academy of Management and Technology, Ratnagiri

Department of Information Technology

Subject:	**R Programming Lab. (ITL804)**		
Class:	**BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20**		
Name of Student:	**Kazi Jawwad A Rahim**		
Roll No:	**28**	**Date of performance (DOP) :**	
Assignment/Experiment No:	**05**	**Date of checking (DOC) :**	
Title: Working with graphics and tables			
Marks:		**Teacher's Signature:**	

1. Aim: To understand the exploratory data analysis and the methods required to do it in R.

2. Prerequisites:

1. Basics of R programming, various data structures for data sets.

3. Hardware Requirements:

1. PC with minimum 2GB RAM

4. Software Requirements:

1. Windows / Linux OS.
2. R version 3.6 or higher

5. Learning Objectives:

1. To understand various graphical visualization of data sets.
2. To understand the use of tables.

6. Learning Objectives Applicable: LO 5

7. Program Outcomes Applicable: PO 4, PO 5

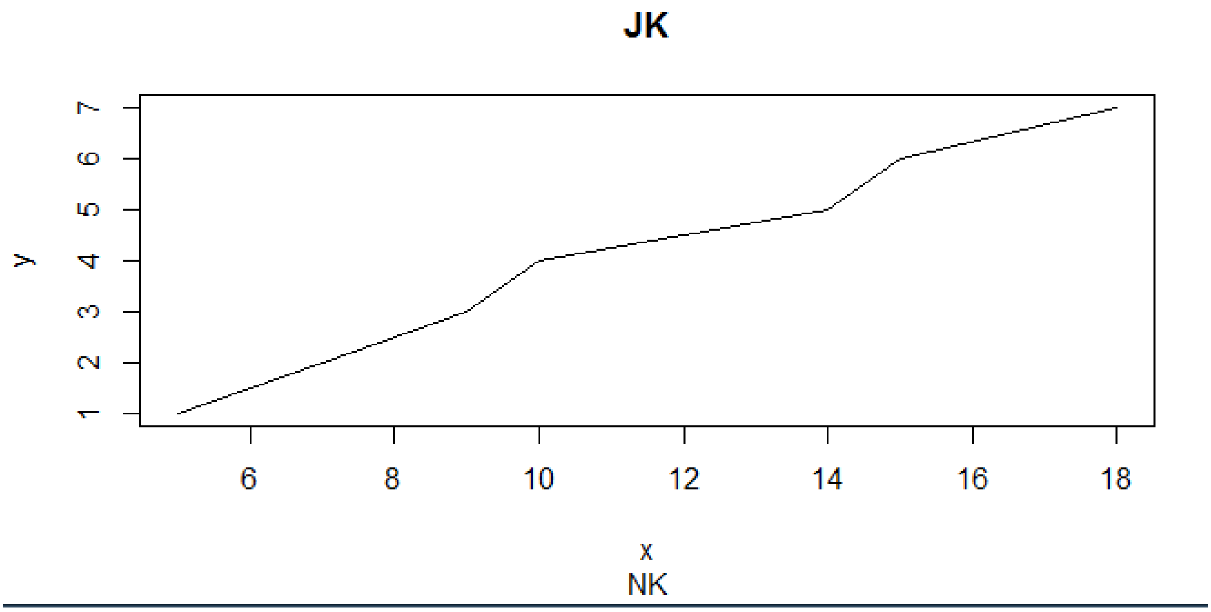
8. Program Education Objectives Applicable: PEO 3, PEO 4

10. Results:

Plot:

```
x=c(5,7,9,10,14,15,18)
y=c(1,2,3,4,5,6,7)
plot(x,y,'l',main="JK",sub="NK")
```

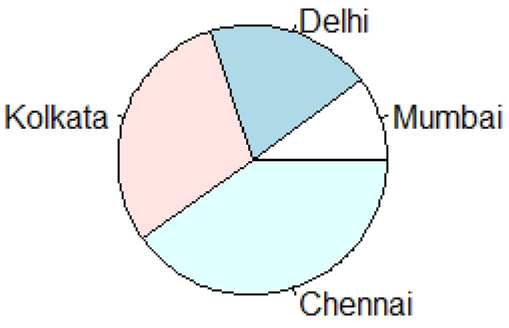
OUTPUT:



Pie Charts:

```
x1=c(21,42,63,84)
labels=c("Mumbai","Delhi","Kolkata","Chennai")
pie(x1,labels)
```

OUTPUT:



Pairs:

```
x2=1:5
y2=x2**2
z2=c(2,1,3,9,2)
A=cbind(x2,y2,z2)
pairs(A)
```

OUTPUT:

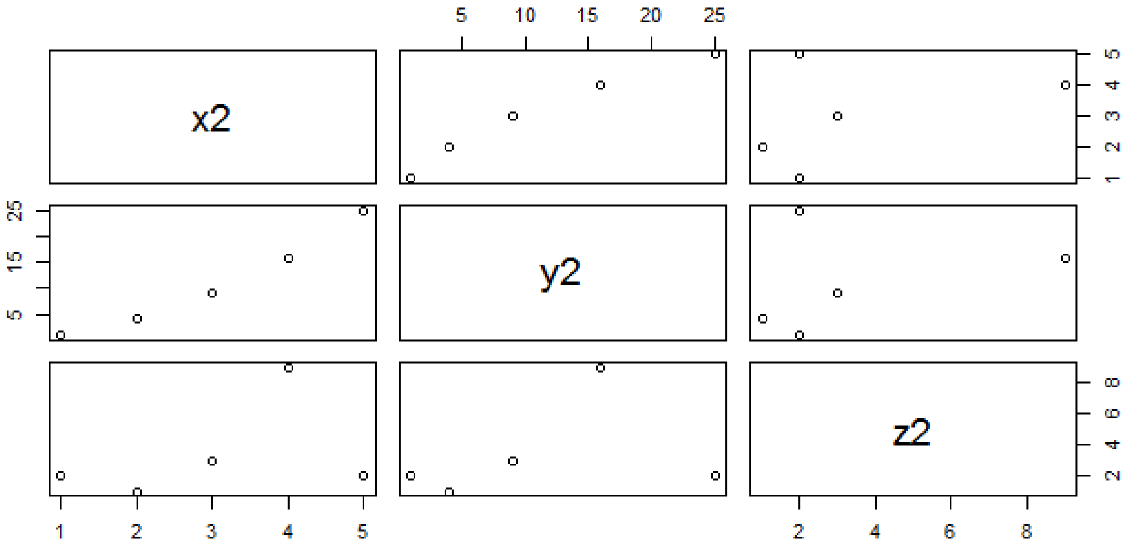
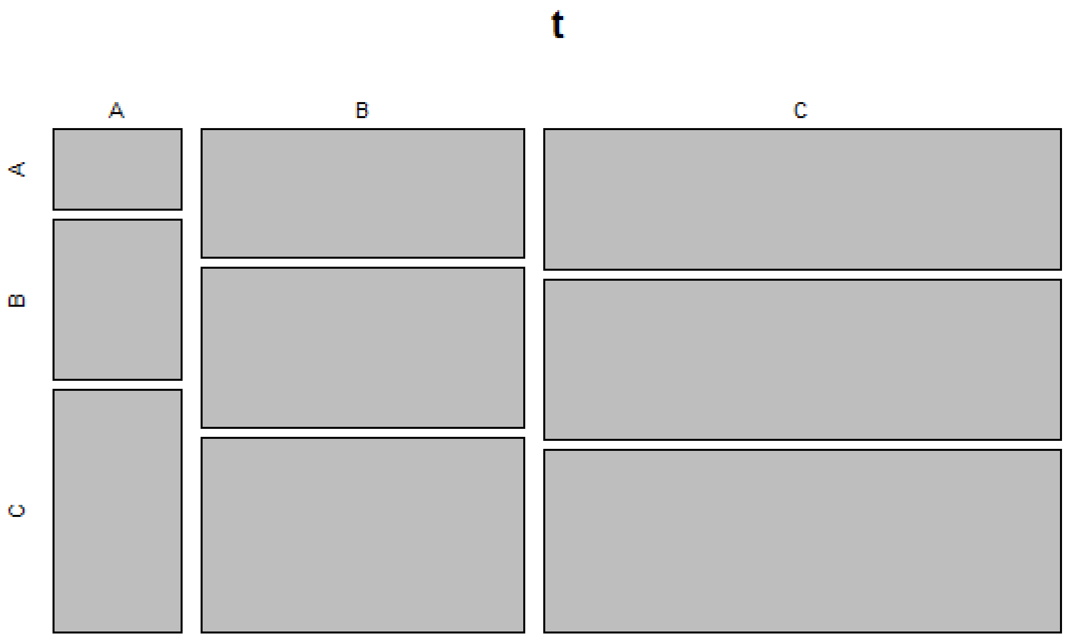


Table:

```
B=matrix(c(1:9),nrow=3,byrow=TRUE)
t=as.table(B)
print(t)
plot(t)
```

OUTPUT:

```
  A B C

A 1 2 3

B 4 5 6

C 7 8 9
```



11. Learning Outcomes Achieved:

1. We understood various graphical visualization of data sets.
2. We understood the use of tables.

12. Conclusion:

We have successfully demonstrated the exploratory data analysis and the methods required to do it in R. We have also demonstrated various graphics methods such as scatterplots, pairs, pie charts and Tables.

13. Experiment/Assignment Evaluation

<table>
<tr><td colspan="6"><b>Experiment/Assignment Evaluation:</b></td></tr>
<tr><td><b>Sr. No.</b></td><td colspan="3"><b>Parameters</b></td><td><b>Marks obtained</b></td><td><b>Out of</b></td></tr>
<tr><td><b>1</b></td><td colspan="3">Technical Understanding (Assessment may be done based on Q &amp; A <b><u>or</u></b> any other relevant method.) Teacher should mention the other method used -</td><td></td><td>6</td></tr>
<tr><td><b>2</b></td><td colspan="3">Neatness/presentation</td><td></td><td>2</td></tr>
<tr><td><b>3</b></td><td colspan="3">Punctuality</td><td></td><td>2</td></tr>
<tr><td colspan="2"><b>Date of performance (DOP)</b></td><td></td><td><b>Total marks obtained</b></td><td></td><td><b>10</b></td></tr>
<tr><td colspan="2"><b>Date of checking (DOC)</b></td><td></td><td colspan="3"><b>Signature of teacher</b></td></tr>
</table>

References:

1. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What are different data visualization command and functions in R?
2. What is table?
3. How table is different than data frame?

Finolex Academy of Management and Technology, Ratnagiri

Department of Information Technology

Subject:	**R Programming Lab. (ITL804)**		
Class:	**BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20**		
Name of Student:	**Kazi Jawwad A Rahim**		
Roll No:	**28**	**Date of performance (DOP) :**	
Assignment/Experiment No:	**06**	**Date of checking (DOC) :**	
Title: Working with larger data-sets and introduction to ggplot2 graphics.			
Marks:		**Teacher's Signature:**	

1. Aim: To understand the exploratory data analysis and the methods required to do it in R.

2. Prerequisites:

1. Data-frames, tables, basic graphical functions.

3. Hardware Requirements:

1. PC with minimum 2GB RAM

4. Software Requirements:

1. Windows / Linux OS.
2. R version 3.6 or higher

5. Learning Objectives:

1. To understand the sources of larger data sets..
2. To understand how the larger data-sets are maintained and managed.
3. To understand the basic usages of ggplot2 graphics package.

6. Learning Objectives Applicable: LO 3, LO 5

7. Program Outcomes Applicable: PO 4, PO 5

8. Program Education Objectives Applicable: PEO 4, PEO 6

10. Results:

```
setwd("f:/exp6")

fr = read.csv("data.csv")

print(fr)
```

```
>>>     Sr.   Name Age Gender Marks

1   1 Jawwad  21      M    80
2   2  Sahil  22      M    82
3   3 Aniket  22      M    84
4   4  Sagar  22      M    86
```

```
mode(fr)
[1] "list"
```

```
class(fr)
[1] "data.frame"
```

```
fr$Name
[1] Jawwad Sahil  Aniket Sagar
```

```
fr$Age
[1] 21 22 22 22
```

```
fr$Marks
[1] 80 82 84 86
```

```
mode(fr$Name)
[1] "numeric"
```

```
class(fr$Name)
[1] "factor"
```

After adding **"header = FALSE"** as a parameter in **read.csv(….)**, we got

```
  V1     V2  V3     V4    V5
1 Sr.   Name Age Gender Marks
2   1 Jawwad  21      M    80
3   2  Sahil  22      M    82
4   3 Aniket  22      M    84
5   4  Sagar  22      M    86
```

Now, I’m using a large data set **“lendingdata.csv”** of about 15 columns and 27518 rows.

```
fr = read.csv("lendingdata.csv")
```

```
ncol(fr)
[1] 15
```

```
nrow(fr)
[1] 27518
```

Now, I’m listing one of the columns data as follows

```
fr$country
   [1] Cambodia                    Philippines
   [3] Peru                    Tajikistan
   [5] Uganda                    Jordan
   [7] Tajikistan                  Cambodia
   [9] Nicaragua                   Nigeria
  [11] Colombia                    Nicaragua
  [13] Colombia                    Philippines
  [15] Ecuador                    Colombia
```

And so on

```
mode(fr$country)
[1] "numeric"
```

```
class(fr$country)
[1] "factor"
```

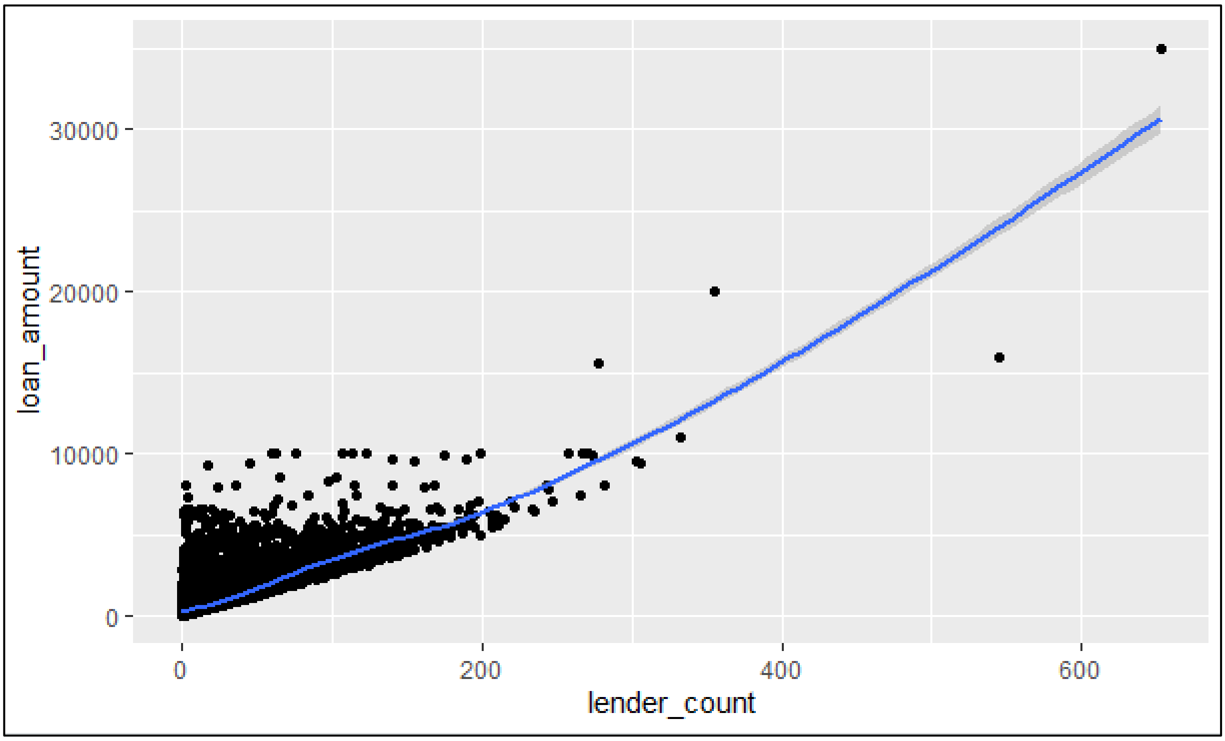
Now, for demonstrating the GGPlot, first we need to install the **ggplot2** package as

```
install.packages(ggplot2)
```

After successfully installing the ggplot2 package and its dependencies, I’m ready to demonstrate.

Source code:

```
library(ggplot2)

setwd("f:/exp6")

fr = read.csv("lendingdata.csv")

ggplot(fr,aes(x=lender_count,y=loan_amount))+geom_point()+geom_smooth()
```



11. Learning Outcomes Achieved:

1. We understood the sources of larger data sets.
2. We understood how the larger data-sets are maintained and managed.
3. We understood the basic usages of ggplot2 graphics package.

12. Conclusion:

We understood the exploratory data analysis and the methods required to do it in R. Also, we have done operations on larger data sets and performed GGplot of the data set to analyze the relativity of the data.

13. Experiment/Assignment Evaluation

<table>
<tr><td colspan="6"><b>Experiment/Assignment Evaluation:</b></td></tr>
<tr><td><b>Sr. No.</b></td><td colspan="3"><b>Parameters</b></td><td><b>Marks obtained</b></td><td><b>Out of</b></td></tr>
<tr><td><b>1</b></td><td colspan="3">Technical Understanding (Assessment may be done based on Q & A <b><u>or</u></b> any other relevant method.) Teacher should mention the other method used -</td><td></td><td>6</td></tr>
<tr><td><b>2</b></td><td colspan="3">Neatness/presentation</td><td></td><td>2</td></tr>
<tr><td><b>3</b></td><td colspan="3">Punctuality</td><td></td><td>2</td></tr>
<tr><td colspan="2"><b>Date of performance (DOP)</b></td><td></td><td><b>Total marks obtained</b></td><td></td><td><b>10</b></td></tr>
<tr><td colspan="2"><b>Date of checking (DOC)</b></td><td></td><td colspan="3"><b>Signature of teacher</b></td></tr>
</table>

References:

1. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What are different ways to store larger data-set?
2. What are names of packages required to extract data from data-set stored in standard spreadsheet.
3. What are various plotting functions in ggplot2?

Subject:	**R Programming Lab. (ITL804)**		
Class:	**BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20**		
Name of Student:	**Kazi Jawwad A Rahim**		
Roll No:	**28**	**Date of performance (DOP) :**	
Assignment/Experiment No:	**07**	**Date of checking (DOC) :**	
Title: Program to demonstrate regression and correlation in tabular data including categorical data.			
Marks:		**Teacher's Signature:**	

1. **Aim**: To understand the exploratory data analysis and the methods required to do it in R.

2. **Prerequisites**:
   1. Working with larger data-sets.

3. **Hardware Requirements**:
   1. PC with minimum 2GB RAM

4. **Software Requirements:**
   1. Windows / Linux OS.
   2. R version 3.6 or higher

5. **Learning Objectives:**
   1. To understand the basic elements of larger data-sets.
   2. To understand numerical and categorical variables in larger data-sets.
   3. To understand how to apply regression to design decision model on the larger data-sets.

6. **Learning Objectives Applicable: LO 5, LO 6**

7. **Program Outcomes Applicable: PO 4, PO 5**

8. **Program Education Objectives Applicable: PEO 4, PEO 6**

10. Results:

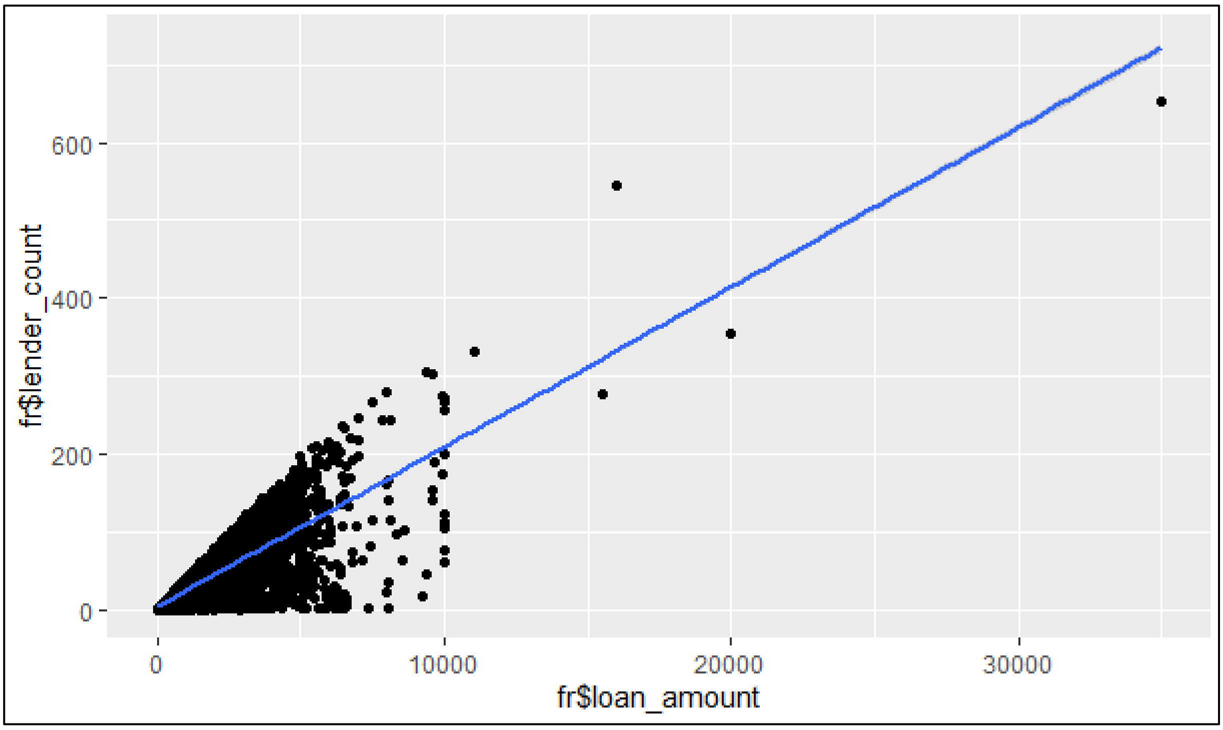
Here we have considered a large data set *"lendingdata.csv"* of 15 columns and 27518 rows.

```
fr = read.csv("lendingdata.csv")
```

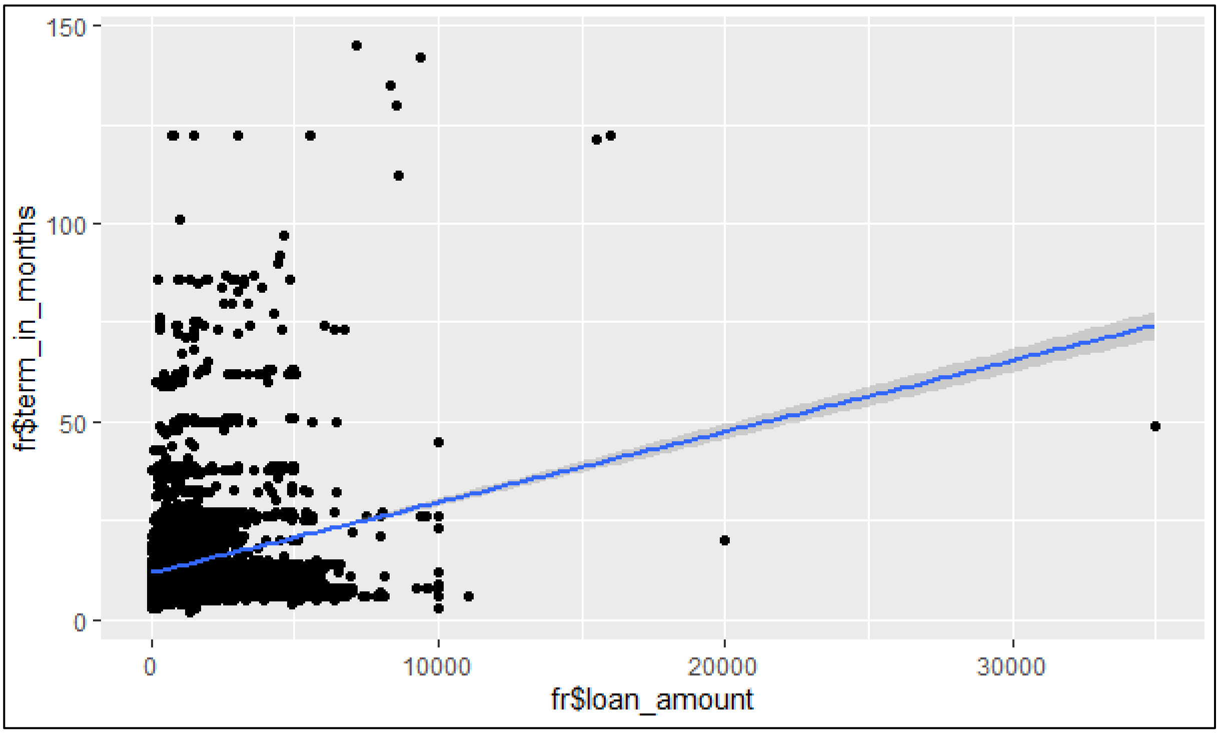
We are now considering three columns namely *loan_amount*, *lender_count* and *term_in_months*.

We will now plot regression line for above mentioned columns in pair of any two columns.

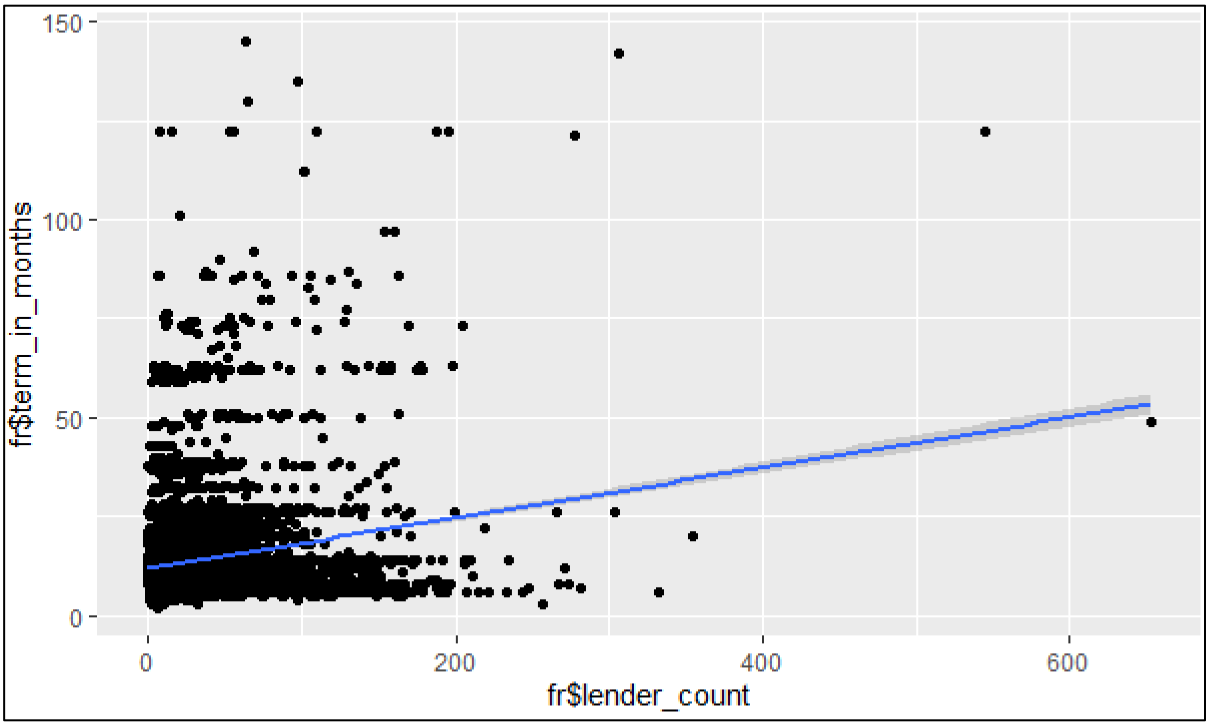
```
ggplot(fr,aes(x=fr$loan_amount,y=fr$lender_count))+geom_point()+geom_smooth(method=l
m,formula=y~x)
```



```
ggplot(fr,aes(x=fr$loan_amount,y=fr$term_in_months))+geom_point()+geom_smooth(method
=lm,formula=y~x)
```

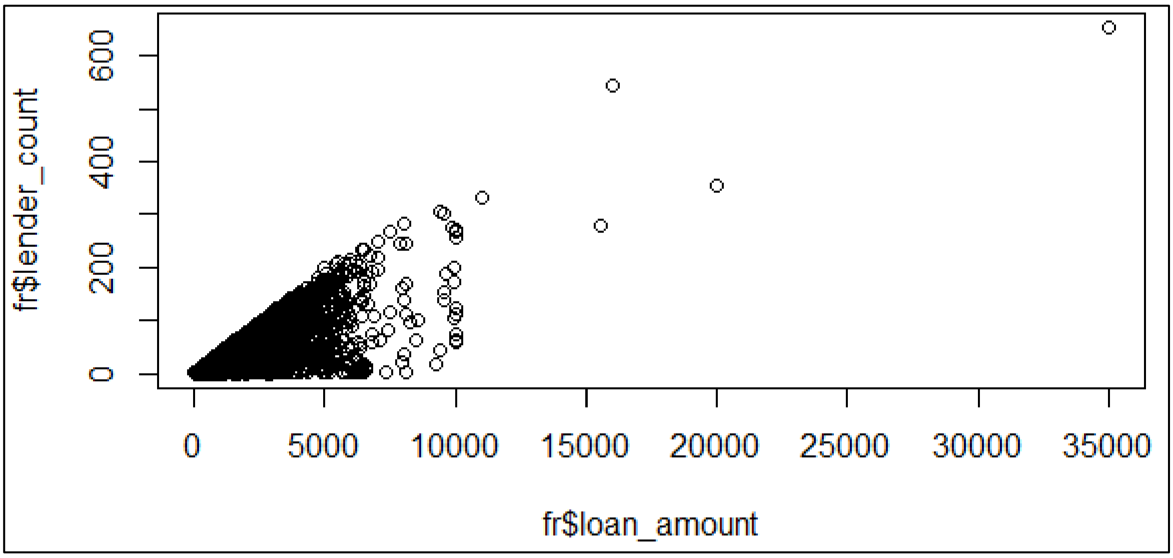


```
ggplot(fr,aes(x=fr$lender_count,y=fr$term_in_months))+geom_point()+geom_smooth(method
=lm,formula=y~x)
```

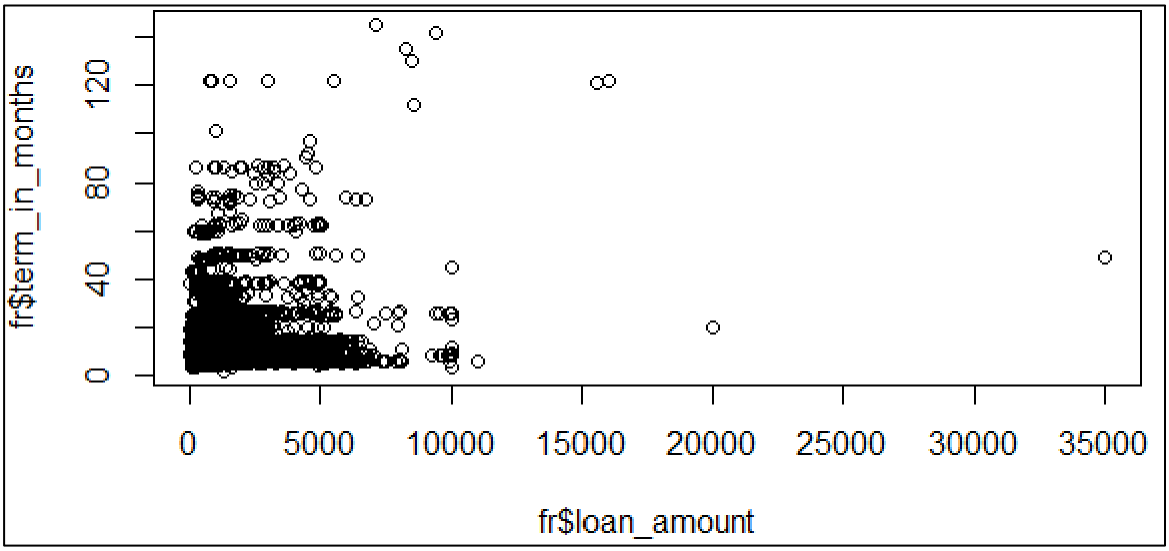


Following are the correlations and their visualization.

```
cor(fr$loan_amount,fr$lender_count)

>>>0.8151209

plot(fr$loan_amount,fr$lender_count)
```

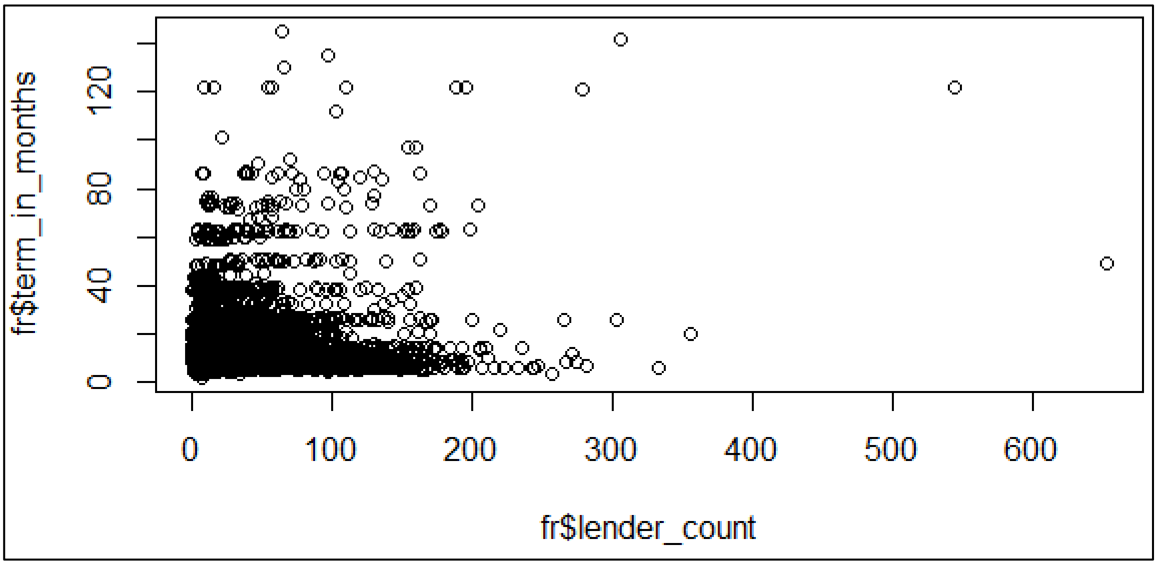


```
cor(fr$loan_amount,fr$term_in_months)

>>>0.2063649

plot(fr$loan_amount,fr$term_in_months)
```



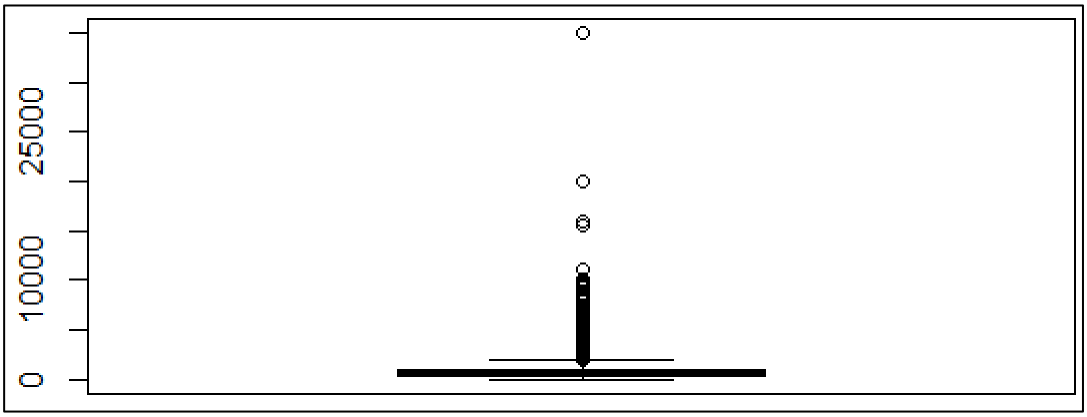
```
cor(fr$lender_count,fr$term_in_months)

>>>0.1846157
```

plot(fr$lender_count,fr$term_in_months)

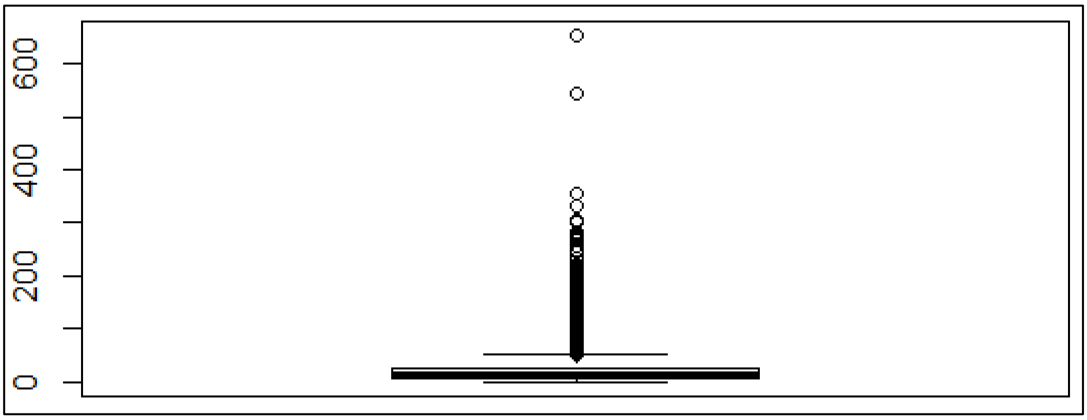


Now we will visualize correlation of categorical variable with a numeric variable using Boxplot for above mentioned three columns.

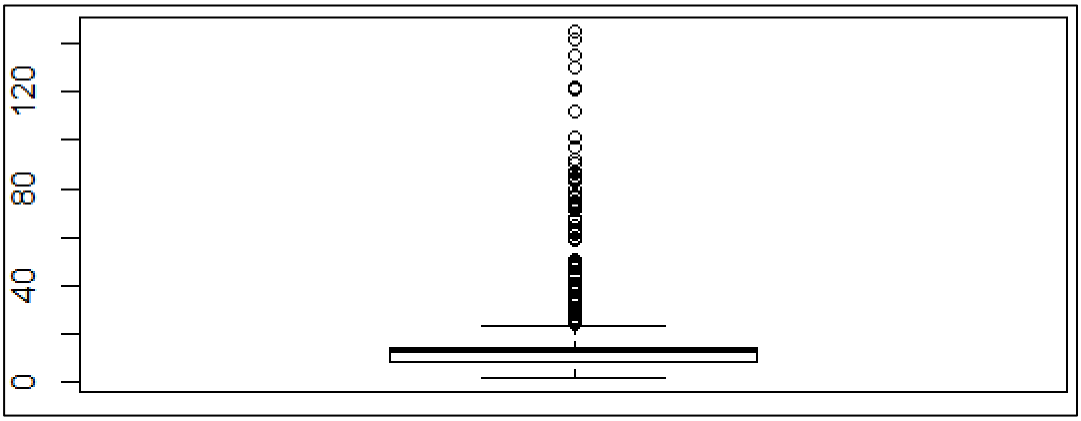
boxplot(fr$loan_amount)



boxplot(fr$lender_count)

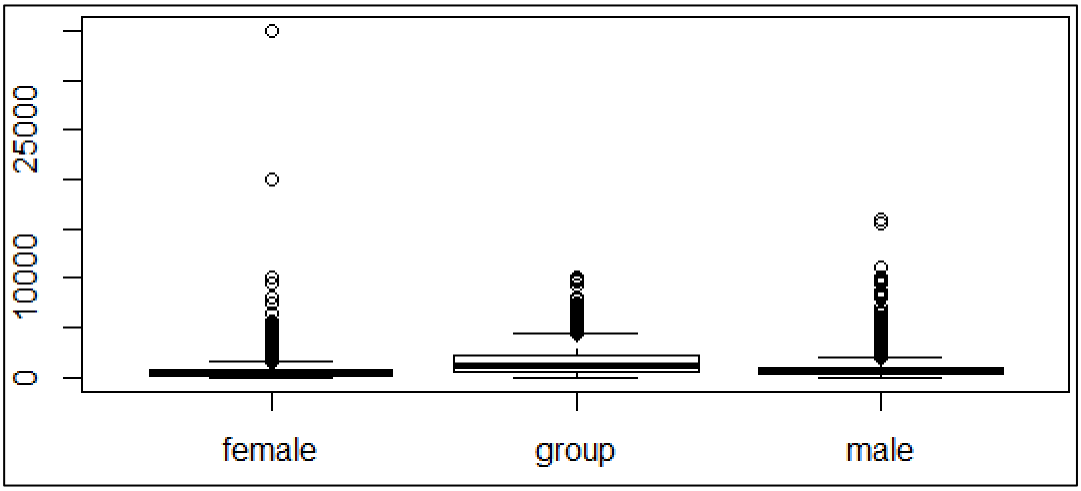


```
boxplot(fr$term_in_months)
```

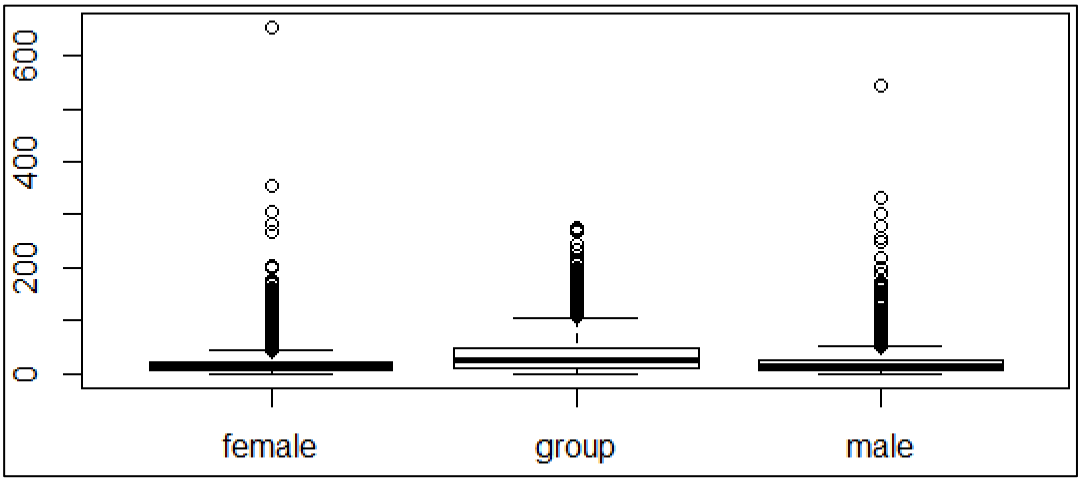


Now let's apply a function by splitting the *loan_amount*, *lender_count* and *term_in_months* as per the genders each, it will display multiple boxplots for different possible genders.

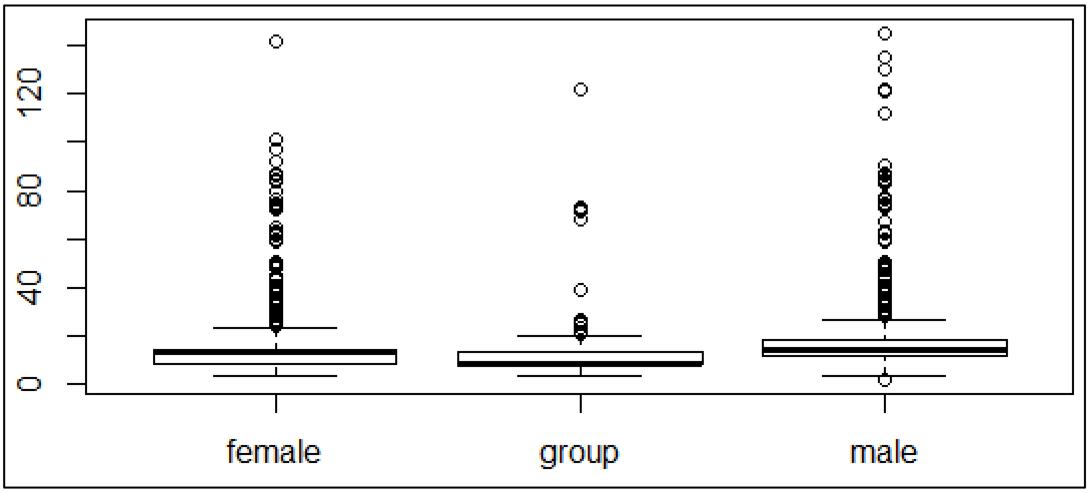
```
boxplot(split(fr$loan_amount,fr$borrower_genders))
```



```
boxplot(split(fr$lender_count,fr$borrower_genders))
```

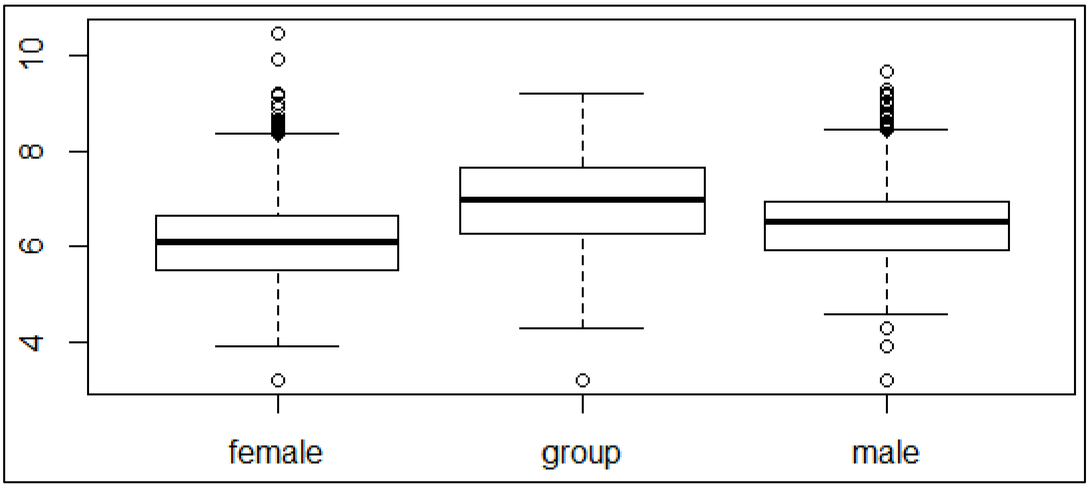


```
boxplot(split(fr$term_in_months,fr$borrower_genders))
```

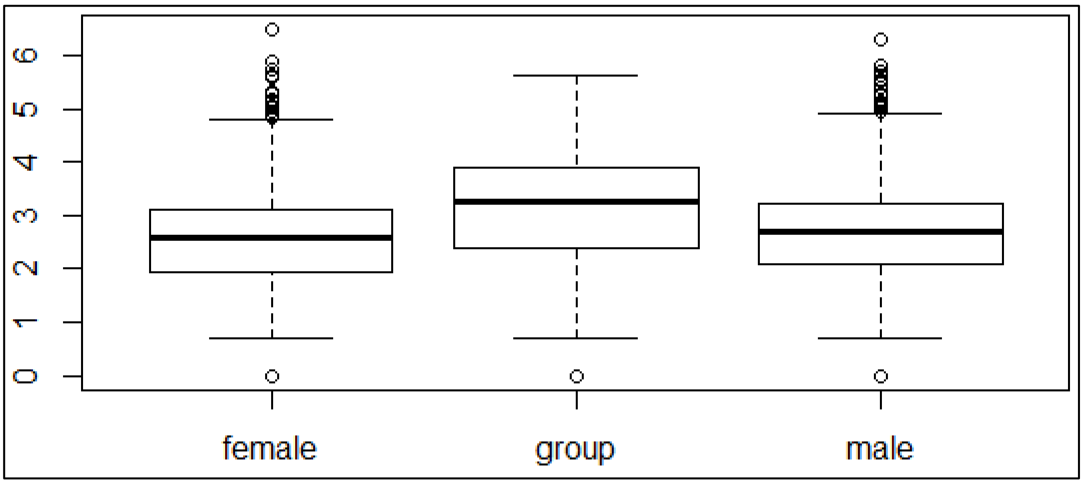


We can take log of *loan_amount, lender_count* and *term_in_months* each to have a broader view.

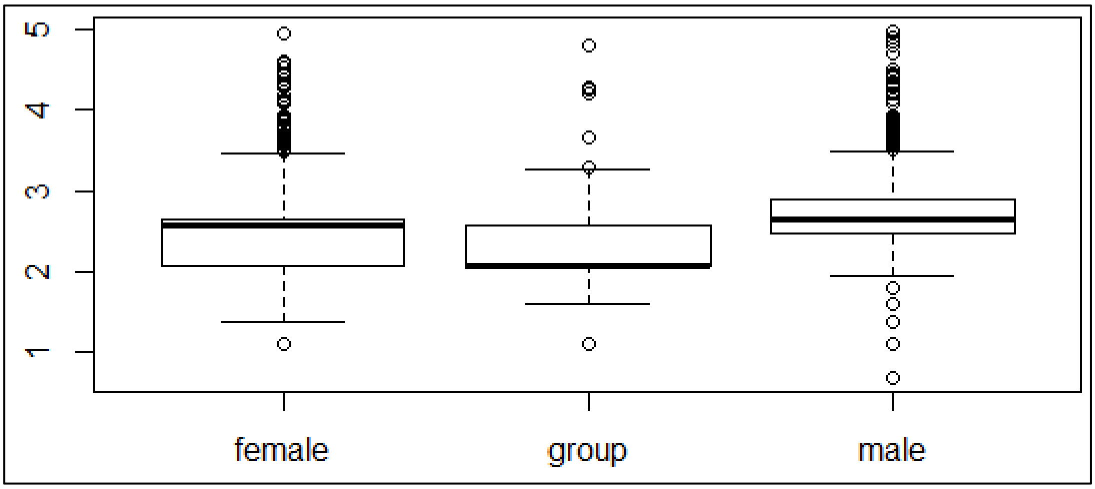
```
boxplot(split(log(fr$loan_amount),fr$borrower_genders))
```



```
boxplot(split(log(fr$lender_count),fr$borrower_genders))
```



```
boxplot(split(log(fr$term_in_months),fr$borrower_genders))
```



11. Learning Outcomes Achieved:

1. We understood the basic elements of larger data-sets.
2. We understood numerical and categorical variables in larger data-sets.
3. We understood how to apply regression to design decision model on the larger data-sets.

12. Conclusion:

We have successfully demonstrated the exploratory data analysis and the methods required to do it in R. Also, we have plotted the regression line, correlations between columns and boxplots.

13. Experiment/Assignment Evaluation

Experiment/Assignment Evaluation:				
Sr. No.	**Parameters**		**Marks obtained**	**Out of**
1	Technical Understanding (Assessment may be done based on Q & A **or** any other relevant method.) Teacher should mention the other method used -			6
2	Neatness/presentation			2
3	Punctuality			2
Date of performance (DOP)		**Total marks obtained**		**10**
Date of checking (DOC)		**Signature of teacher**		

References:

1. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What does it mean by categorical variables in data-sets?
2. What does it mean by regression?
3. What is correlation and how is it useful in data-science?

Department of Information Technology, FAMT Ratnagiri

ASSIGNMENT-I

BE(IT), Sem-VIII(CBCGS) Sub.- R Programming Lab (ITL804)

Extracting data from larger data set and performing exploratory analysis

		Module	Level	CO
1)	List different spreadsheet file formats used for storing the larger data-set. Explain any two in short.	M3	R	LO3
2)	Give package and libraries required for- extracting the data from CSV and excel file format. Also, give a sample codes.	M1	U	LO1
3)	What is exploratory analysis of the data? Explain with a sample code.	M6	U	LO6

.....................x.......................

Assignment-I

1. List different spreadsheet file formats used for storing the larger data-set. Explain any two in short.

ANS. Following is a list of common spreadsheet file formats along with their file extensions.

File Extension	File Format
CSV	Comma Separated Values File
DIF	Microsoft Data Interchange Format
ODS	OpenDocument Spreadsheet
OTS	OpenDocument Spreadsheet Template
TSV	Tab Separated Values File
XLM	Microsoft Excel Macro File
XLS	Microsoft Excel Binary File Format
XLSB	Microsoft Excel Binary Spreadsheet File
XLSM	Microsoft Excel Open XML Macro-Enabled Spreadsheet
XLSX	Microsoft Excel Open XML Spreadsheet
XLT	Microsoft Excel Template File
XLTM	Microsoft Excel Open XML Macro-Enabled Spreadsheet Template
XLTX	Microsoft Excel Open XML Spreadsheet Template

1. CSV: Files with CSV (Comma Separated Values) extension represent plain text files that contain records of data with comma separated values. Each line in a CSV file is a new record from the set of records contained in the file. Such files are generated when data transfer is intended from one storage system to another. Since all applications can recognize records separated by comma, import of such data files to database is done very conveniently. Almost all spreadsheet applications such as Microsoft Excel or OpenOffice Calc can import CSV without much effort. Data imported from such files is arranged in cells of a spreadsheet for representation to user.

2. XLSX: XLSX is well-known format for Microsoft Excel documents that was introduced by Microsoft with the release of Microsoft Office 2007. Based on structure organized according to the Open Packaging Conventions as of the OOXML standard ECMA-376, the new format is a zip package that contains a number of XML files. The underlying structure and files can be examined by simply unzipping the .xlsx file.

2. Give package and libraries required for- extracting the data from CSV and excel file format. Also, give a sample code.

ANS. Steps to loading and extracting data from csv and excel file format is as follows.

1. Save the CSV/Excel file in the same location as of the script.
2. For CSV file use

```
fr=read.csv("file.csv")
```

where fr is data frame object
file.csv is the dataset

3. For Excel file use

```
fr=read.xls("file.xlsx")
```

where fr is data frame object
file.xlsx is the data set

e.g.

```
df=read.csv("https://s3.amazonaws.com/assets.datacamp.com/blog_assets/scores_timed.csv")
print(df)
```

OUTPUT:

```
  X1.6.12.01.03.0       X50.WORST
1 2;16;07:42:51;0         32;BEST
2 3;19;12:01:29;0              50
3 4;13;03:22:50;0 14; INTERMEDIATE
4  5;8;09:30:03;0        40;WORST
```

3. What is exploratory analysis of the data? Explain with a sample code.

ANS. EDA: In statistics, exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task. Exploratory data analysis was promoted by John Tukey to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. EDA is different from initial data analysis (IDA), which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA.

SAMPLE CODE:

```
basic_eda <- function(data)
{
  glimpse(data)
  df_status(data)
  freq(data)
  profiling_num(data)
  plot_num(data)
  describe(data)
}
```

Glimpse gives the no of observations (rows) and variables and a head of the first cases.
freq function runs for all factor or character variables automatically
profiling_num runs for all numerical/integer variables automatically
Describe is useful to have a quick picture for all the variables.

Department of Information Technology, FAMT Ratnagiri

ASSIGNMENT-II

BE(IT), Sem-VIII(CBCGS)   Sub.- R Programming Lab (ITL804)

Applying data mining algorithm on the data extracted in Assignment-1, and visualization and interpretation of results.

		Module	Level	CO
1)	Write a code to read the larger data-set contains in the file at *http://famt.ac.in/eResource/it/lendingdata.csv.*	M5	C	LO5
2)	What is data cleaning? Explain in detail.	M6	U	LO6
3)	List various regression models used in statistics for estimating the result.	M6	U	LO6

......................x........................

Assignment-II

1. Write a code to read the larger data-set contains in the file at http://famt.ac.in/eResource/it/lendingdata.csv.

ANS. Now, I’m using a large data set **“lendingdata.csv”** of about 15 columns and 27518 rows.

```
fr = read.csv("lendingdata.csv")
```

OUTPUT:

```
ncol(fr)
[1] 15

nrow(fr)
[1] 27518
```

Now, I’m listing one of the columns data as follows

```
fr$country
  [1] Cambodia                  Philippines
  [3] Peru                    Tajikistan
  [5] Uganda                    Jordan
  [7] Tajikistan                Cambodia
  [9] Nicaragua                  Nigeria
 [11] Colombia                   Nicaragua
 [13] Colombia                   Philippines
 [15] Ecuador                   Colombia
```

And so on

2. What is data cleaning? Explain in detail.

ANS. Data Cleaning is the process of transforming raw data into consistent data that can be analyzed. It is aimed at improving the content of statistical statements based on the data as well as their reliability. Data cleaning may profoundly influence the statistical statements based on the data. R has a set of comprehensive tools that are specifically designed to clean data in an effective and comprehensive manner. It mainly has three steps as follows-

STEP 1: Initial Exploratory Analysis: The first step to the overall data cleaning process involves an initial exploration of the data frame that you have just imported into R. It is very important to understand how you can import data into R and save it as a data frame.

STEP 2: Visual Exploratory Analysis: There are 2 types of plots that you should use during your cleaning process –The Histogram and the BoxPlot

Histogram: The histogram is very useful in visualizing the overall distribution of a numeric column. We can determine if the distribution of data is normal or bi-modal or unimodal or any other kind of distribution of interest. We can also use Histograms to figure out if there are outliers in the particular numerical column under study.

BoxPlot: Boxplots are super useful because it shows you the median, along with the first, second and third quartiles. BoxPlots are the best way of spotting outliers in your data frame.

STEP 3: Correcting the errors: This step focuses on the methods that you can use to correct all the errors that you have seen.

3. List various regression models used in statistics for estimating the result.

ANS. List of various regression models-

1. Histogram
2. Boxplot
3. Correlation
4. Plot
5. GGPlot