



CENTRAL UNIVERSITY OF RAJASTHAN  
**DEPARTMENT OF DATA SCIENCE**

*DECEMBER, 2018*

**UNIVERSITY FEEDBACK SURVEY**

UNDER THE GUIDANCE OF-  
**Dr. MANAS PATRA**

**Submitted by**

Priya Sinha (18)  
Aishwarya Verma (02)  
Antara Khan (14)

**CANDIDATE'S DECLARATION**

We here declare that the work embodied in this project entitled “survey project on UNIVERSITY FEEDBACK” towards partial fulfilment of the requirement for the award of degree of M.SC IN BIG DATA ANALYTICS, is a bonafide piece of work carried out by as under the extreme supervision of Dr. Manas Patra, Department of Data science, Central university of Rajasthan. The work presented in this project has not been submitted by us for the award of any other degree of this or any other university. We have taken care in all respects to honour the intellectual property right and we have acknowledged the contribution(s) of others for using them for this academic purpose.

**DATE:**

**AISHWARYA VERMA (02)**

**ANTARA KHAN(14)**

**PLACE :**

**PRIYA SINHA(18)**

## **CERTIFICATE**

This is to certify that Ms. Aishwarya Verma, Ms. Antara Khan and Ms. Priya Sinha had completed their “survey project” under my supervision and the above statement made by them is correct and true to the best of my knowledge.

**Dr. Manas Patra**

HEAD OF DEPARTMENT

Data Science Department

Central UNIVERSITY OF RAJASTHAN.

**PLACE:** Ajmer, Rajasthan

**DATE**

# ACKNOWLEDGEMENT

I feel much honored in presenting this dissertation report in such an authenticable form of sheer endurance and continual efforts of inspiring excellence from various coordinating factor of cooperation and sincere efforts drawn from all sources of knowledge. I express my sincere central university of Rajasthan, for his valuable guidance and infilling support for the completion of this project work. We would also lie to thanks Dr. Manas Patra department of data science for this support and valuable guidance.

**Date:** December 2018

**Place:**

## **ABSTRACT**

This project we have undertaken is “survey project on University Feedback”. This represents the total 467 students’ responses of university feedback forms. First, we have cleaned the data of 467 responses and got 422 cleaned data responses. We have analyzed the whole survey data through some test and graphically by the help of scatter plots, bar plots and also by making contingency table and applied Chi-square test which are shown the correlation between different columns on our feedback data.

The main purpose of this project to analyze feedback data and read the trends between the quantity we have chosen.

# **TABLE OF CONTENTS**

<b>UNIVERSITY FEEDBACK FORM.....</b>	<b>6-9</b>
<b>Data collection.....</b>	<b>10</b>
<b>Data cleaning (manually).....</b>	<b>11</b>
<b>Data cleaning (in R) .....</b>	<b>12</b>
<b>Histogram.....</b>	<b>13</b>
<b>Density Plot.....</b>	<b>14</b>
<b>Bar plot.....</b>	<b>15</b>
<b>Contingency Table.....</b>	<b>16-21</b>
• Factor().	
• Table()	
• Chi-square test	
• Scatter Plot	
<b>Summary of data in R.....</b>	<b>22-23</b>
<b>Conclusion .....</b>	<b>24</b>
<b>References.....</b>	<b>25</b>

# UNIVERSITY FEEDBACK FORM

12/11/2018

UNIVERSITY FEEDBACK FORM

## UNIVERSITY FEEDBACK FORM

Please fill this form only once .

\* Required

1, ENROLLMENT NUMBER \*

\_\_\_\_\_

## RATING PARAMETERS

RATING :- 5 : EXCELLENT , 4 : VERY GOOD , 3 : GOOD , 2 : AVERAGE , 1 : BELOW AVERAGE

2, GENDER \*

Mark only one oval.

☐ Female

☐ Male

☐ Other

## HOSTEL AND MESS

3, 1. FOOD QUALITY AND VARIETY \*

Mark only one oval.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4, 2. HYGIENE & MAINTENANCE OF MESS \*

Mark only one oval.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5, 3. CLEANLINESS OF WASHROOMS \*

Mark only one oval.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**6. 4. DRINKING WATER QUALITY \****Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**7. 5. WI-FI AND INTERNET SERVICES \****Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**8. 6. SPORTS AND GYM \****Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**9. 7. HOW SECURE DO YOU FEEL IN CAMPUS? \****Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**ACADEMICS****10. 1.ORGANISATION AND COVERAGE OF SYLLABUS \****Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**11. 2.RELEVANCE AND REAL WORLD APPLICATION \****Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**12. 3.EXAMINATION PATTERN AND GRADING SYSTEM***Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**13. 4. CO- CURRICULAR ACTIVITIES \***

*Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## **ABOUT THE INSTRUCTOR**

---

**14. 1. TECHNICAL CONTENT IN TEACHING \***

*Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**15. 2. LEVEL OF INTERACTION DURING LECTURES \***

*Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**16. 3. AVAILABILITY OUTSIDE CLASS \***

*Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## **ACADEMIC FACILITIES**

---

**17. 1. LAB FACILITIES (EX: EQUIPMENT) \***

*Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**18. 2. AVAILABILITY OF BOOKS, E-RESOURCES, STUDY MATERIALS, ETC. \***

*Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



**19. 3.QUALITY OF GUEST LECTURES, WORKSHOPS, SEMINARS, ETC \***

*Mark only one oval.*

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**OTHERS**

**20. 1. ARE YOU AWARE ABOUT SPARSH FOUNDATION AND ITS ACTIVITIES? \***

*Mark only one oval.*

- ☐ Yes  
☐ No  
☐ Maybe

**21. 2. DO YOU FEEL GENDER DISCRIMINATION EXISTS IN CURAJ? \***

*Mark only one oval.*

- ☐ Yes  
☐ No  
☐ Maybe

**22. 3. HAVE YOU EXPERIENCED HARASSMENT IN ANY FORM (EX: PHYSICAL, VERBAL, ETC) IN CURAJ? \***

*Mark only one oval.*

- ☐ Yes  
☐ No  
☐ Maybe

**23. 4. ARE YOU SATISFIED WITH THE CCTV MONITORING IN CURAJ? \***

*Mark only one oval.*

- ☐ Yes  
☐ No  
☐ Maybe

# DATA COLLECTION

We have collected the data from the students regarding university's feedback. We have total 23 variables in it and the responses given in 5 categories (1,2,3,4,5) in some questions and for some it has 3 categories(Yes,No,Maybe).

	A	B	C	D	E	F	G
	Timestamp	ENROLLMENT NUMBER	GENDER	1. FOOD QUALITY AND VARIETY	3. CLEANLINESS OF WASHROOMS	2. HYGIENE & MAINTENANCE OF MESS	4. DRINKING WA
2	10/18/2018 12:35:42	2018msbda007	Male	3	3	3	3
3	10/21/2018 18:14:49	2018MSBDA001	Male	3	4	3	3
4	10/21/2018 18:15:08	2018msbda015	Male	3	3	4	4
5	10/21/2018 18:15:43	2018MSBDA006	Male	3	3	4	4
6	10/21/2018 18:16:16	2018msbda003	Male	2	3	3	3
7	10/21/2018 18:16:52	2018MSBDA009	Female	3	1	3	3
8	10/21/2018 18:18:01	2018MSSES010	Female	3	2	3	3
9	10/21/2018 18:18:09	2018MSBDA005	Male	4	3	4	4
10	10/21/2018 18:20:32	2015IMSCS023	Male	3	3	2	2
11	10/21/2018 18:20:37	2018IMSCS003	Male	3	3	3	3
12	10/21/2018 18:21:21	2018imscs023	Male	2	3	2	2
13	10/21/2018 18:21:53	2014IMSCS003	Male	2	2	3	3
14	10/21/2018 18:30:26	2018MSBDA012	Male	2	3	3	3
15	10/21/2018 18:32:20	2018msbda023	Male	2	4	2	2
16	10/21/2018 18:32:54	2018MSSB001	Female	4	3	4	4
17	10/21/2018 18:35:38	2018msbda018	Female	3	2	2	2
18	10/21/2018 18:37:57	2016imss006	Female	2	2	2	2
19	10/21/2018 18:42:20	2017IMSST008	Female	1	1	1	1
20	10/21/2018 18:43:38	2015imsch014	Female	2	2	2	2
21	10/21/2018 18:44:58	2018imsp013	Male	3	1	3	3
22	10/21/2018 18:47:07	2018mtcse004	Male	1	4	2	2
23	10/21/2018 18:48:14	2018IMSCH026	Female	2	2	2	2
24	10/21/2018 18:48:43	2017IMSMB007	Female	3	3	4	4
25	10/21/2018 18:52:16	2017mscs001	Male	3	2	4	4
26	10/21/2018 18:58:11	2016imsec006	Male	1	3	3	3

	A	B	C	D	E	F	G
	Timestamp	ENROLLMENT NUMBER	GENDER	1. FOOD QUALITY AND VARIETY	3. CLEANLINESS OF WASHROOMS	2. HYGIENE & MAINTENANCE OF MESS	4. DRINKING WA
444	11/2/2018 14:31:20	2018imsmb012	Male	2	2	2	2
445	11/2/2018 14:31:27	2018IMSBMT015	Female	3	1	2	2
446	11/2/2018 14:32:25	2018imsbmt006	Female	1	1	1	1
447	11/2/2018 14:33:00	2017-008	Female	1	1	1	1
448	11/2/2018 14:33:06	2018MSSN002	Female	3	2	3	3
449	11/2/2018 14:33:46	2017macms007	Female	2	1	3	3
450	11/2/2018 14:34:10	2017msta005	Female	2	2	2	2
451	11/2/2018 14:34:14	2018MSYT007	Female	3	3	3	3
452	11/2/2018 14:35:04	2017msch001	Female	2	2	3	3
453	11/2/2018 14:35:23	2017msta002	Female	3	3	3	3
454	11/2/2018 14:35:41	2018MSBYT004	Female	3	2	4	4
455	11/2/2018 14:37:07	2018IMSBEC004	Female	3	1	2	2
456	11/2/2018 14:37:13	2017msm003	Female	2	3	3	3
457	11/2/2018 14:37:30	2017msw007	Female	3	3	3	3
458	11/2/2018 14:38:30	2017msw001	Female	2	2	3	3
459	11/2/2018 14:38:45	2018IMSBCH002	Female	4	3	3	3
460	11/2/2018 14:39:41	2018imsbch007	Female	2	2	2	2
461	11/2/2018 14:39:52	2017msm001	Female	3	3	2	2
462	11/2/2018 14:39:57	2018MAEN001	Female	2	1	1	1
463	11/2/2018 14:40:50	2018msta018	Female	4	3	3	3
464	11/2/2018 14:41:20	2017msats003	Female	4	2	3	3
465	11/2/2018 17:57:10	2018phdpfarm001	Male	4	2	1	1
466	11/2/2018 18:00:10	2015phdpfarm001	Male	4	2	4	4
467	11/3/2018 13:03:39	2014phdpfarm08	Male	4	3	4	4

First, we delete the timestamp and delete the unwanted rows.

feedback - Microsoft Excel

Home Insert Page Layout Formulas Data Review View

Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A A

Cut Copy Paste Format Painter

Wrap Text Merge & Center

General \$ % + =

Conditional Formatting Format as Table Normal Bad Good Neutral Calculation Check Cell

AutoSum Fill Clear Sort & Filter Find & Select

ENROLLMENT NUMBER

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
437	2018mae012	Female	4	3	4	1	2	1	5	4	4	4	4	4	5	4	1	4	4	Maybe	No	No	Yes	
438	2017mscm009	Female	2	1	2	2	3	3	4	4	5	3	5	5	5	5	4	5	3	Yes	Yes	No	Yes	
439	2018imses015	Male	2	3	3	1	1	1	4	4	3	3	4	3	3	3	2	4	4	No	Maybe	No	Yes	
440	2017IMSMT012	Male	2	1	3	1	1	1	4	3	2	3	2	1	2	1	2	3	2	No	Yes	No	No	
441	2018msm007	Female	1	1	1	1	3	4	4	4	2	3	1	2	2	1	1	2	3	Maybe	Yes	Yes	Maybe	
442	2018MSESO06	Female	3	3	3	1	1	2	3	5	4	4	3	4	4	3	4	3	Yes	Yes	No	No		
443	2018imscs026	Male	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	3	1	No	Maybe	No	No	
444	2018imsmb012	Male	2	2	2	3	3	3	4	4	4	4	3	3	4	3	4	2	1	No	No	No	Yes	
445	2018IMSMBMT015	Female	3	1	2	2	2	3	4	4	3	4	2	3	1	1	3	1	4	Yes	No	Maybe	No	
446	2018imsbmt006	Female	1	1	1	1	1	1	2	4	2	3	1	2	3	2	2	1	1	Maybe	Yes	Yes	Maybe	
447		Female	1	1	1	3	3	4	4	3	3	4	2	3	3	4	5	4	4	No	Yes	No	Maybe	
448	2018MSSN002	Female	3	2	3	2	3	1	5	5	3	4	4	4	5	5	4	4	4	No	No	No	Yes	
449	2017macms007	Female	2	1	3	1	5	4	5	4	3	4	5	3	3	4	5	5	3	Yes	Yes	No	No	
450	2017msta005	Female	2	2	2	1	3	2	3	3	1	3	2	2	3	1	2	3	2	Maybe	Yes	No	No	
451	2018MSYT007	Female	3	3	4	5	4	3	4	3	3	3	3	4	3	4	5	5	4	No	Yes	No	Yes	
452	2017msch001	Female	2	2	3	1	3	2	4	5	3	3	4	3	3	3	3	3	3	Yes	No	No	No	
453	2017msa002	Female	3	3	3	1	4	2	3	4	3	3	3	3	3	3	3	3	3	No	Yes	No	Maybe	
454	2018MSBYT004	Female	3	2	4	1	3	2	4	2	2	4	2	4	3	2	1	1	2	Yes	No	No	Yes	
455	2018IMSBECo04	Female	1	2	2	2	3	3	5	3	3	3	5	4	3	4	3	4	2	No	No	No	Yes	
456	2017msm003	Female	2	3	3	2	3	4	4	4	3	3	3	3	3	3	3	3	3	4	Maybe	No	No	Yes
457	2017msw007	Female	3	3	3	2	3	1	4	2	2	3	1	3	2	1	1	2	3	No	No	Yes	Maybe	
458	2017msw001	Female	2	2	3	1	2	2	3	3	1	3	1	3	2	1	3	3	3	No	No	No	Yes	
459	2018IMSBCH002	Female	4	3	3	4	3	3	5	4	4	4	5	4	5	5	4	4	3	Yes	No	No	Yes	
460	2018imsbch007	Female	2	2	2	1	2	3	4	3	3	3	5	3	4	4	4	3	4	No	No	No	Yes	
461	2017msm001	Female	3	3	2	2	4	4	4	4	3	4	4	4	5	4	3	4	3	Maybe	No	No	Yes	
462	2018MAEN001	Female	2	1	1	3	2	4	4	4	4	3	3	3	4	5	3	2	4	Yes	Yes	Yes	Maybe	
463	2018msta018	Female	4	3	3	3	4	3	5	4	3	4	4	4	4	3	5	4	4	No	No	No	Yes	
464	2017msats003	Female	4	2	3	1	4	1	4	4	4	4	3	2	4	4	3	3	4	Yes	Yes	No	Yes	
465	2018phdpharm001	Male	4	2	1	1	3	1	3	5	3	4	1	5	5	1	3	1	1	Yes	No	No	No	
466	2015phdpharm001	Male	4	2	4	1	4	1	3	3	3	4	1	3	2	2	2	1	1	Yes	No	No	No	
467	2014phdpharm08	Male	4	3	4	1	4	1	2	5	3	4	1	4	4	2	3	2	2	Yes	No	No	No	

feedback

10:33 PM

12/10/2018

feedback - Microsoft Excel

ENROLLMENT NUMBER																									
1	ENROLLMENT NU	GENDER	1. FOOD Q	3. CLEANL	2. HYGIENE	4. DRINKI	5. WI-FI	AT G	SPORTS	7. HOW SE	1. ORGANI	2. RELEVAN	3. EXAMIN	4. CO-CUT	1. TECHN	2. LEVEL O	3. AVAILAB	1. LAB FAC	2. AVAILAB	3. QUALITY	1. ARE YOU	2. DO YOU	3. HAVE Y	4. ARE	
2	2018msbda007	Male	3	3	3	3	2	4	2	4	3	4	3	4	4	4	4	4	4	4	4	Yes	No	No	No
3	2018MSBDA001	Male	3	3	4	3	3	4	1	5	5	5	5	4	5	5	5	5	4	4	4	Yes	Maybe	No	Maybe
4	2018msbda015	Male	3	3	4	3	2	1	4	3	2	4	2	3	2	3	2	4	2	4	Maybe	No	No	Maybe	
5	2018MSBDA006	Male	3	3	4	3	2	2	2	2	3	3	3	3	3	4	4	4	4	3	2	No	No	Yes	
6	2018msbda003	Male	2	3	3	3	1	1	5	4	3	3	3	3	3	4	4	4	3	3	4	No	Maybe	No	Yes
7	2018MSBDA009	Female	3	1	3	2	3	2	4	3	2	4	3	3	4	4	4	4	4	4	3	Yes	Yes	No	Yes
8	2018MSES010	Female	3	2	3	3	2	4	5	4	5	5	5	5	4	4	4	5	5	5	5	Yes	Yes	No	Yes
9	2018MSBDA005	Male	4	3	4	3	3	3	3	4	5	5	4	4	5	5	5	5	5	3	4	No	Maybe	No	Yes
10	2015IMSCS023	Male	3	3	2	2	2	3	3	4	4	3	3	3	3	2	1	4	5	2	Yes	Yes	No	No	
11	2018IMSCS003	Male	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Yes	No	No	Yes
12	2018imscs023	Male	2	3	2	3	4	2	4	5	4	5	3	5	5	5	5	4	3	3	3	Yes	Yes	No	Yes
13	2014IMSCS003	Male	2	2	3	4	4	3	4	3	4	3	2	3	3	2	1	2	2	No	Yes	Maybe	No	Yes	
14	2018MSBDA012	Male	2	3	2	5	3	5	3	5	3	4	2	3	3	3	4	4	3	Maybe	Maybe	No	No	Yes	
15	2018msbda023	Male	2	4	2	3	5	3	5	3	2	4	2	4	3	3	5	5	5	2	Maybe	Maybe	No	Yes	
16	2018MSSB001	Female	4	3	4	1	2	1	5	5	4	5	5	5	5	5	5	5	5	5	No	Yes	No	Yes	
17	2018msbda018	Female	3	2	2	3	4	3	5	3	3	4	4	3	3	4	4	4	4	3	No	No	No	Yes	
18	2016imst006	Female	2	2	2	1	2	2	3	4	3	3	1	2	2	2	3	2	1	Yes	Maybe	No	No	Yes	
19	2017IMSST008	Female	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Yes	Yes	Yes	No	
20	2015imsch014	Female	2	2	2	2	2	2	2	3	4	4	5	2	4	4	4	4	5	3	Yes	Yes	No	No	
21	2018imspH013	Male	3	1	3	1	1	1	1	5	5	3	1	5	3	5	5	5	4	3	1	Yes	No	No	Yes
22	2018mtcse004	Male	1	4	2	3	4	4	4	4	4	4	4	2	4	4	4	4	1	5	2	No	Yes	No	Maybe
23	2018MSCHO26	Female	2	2	2	1	1	2	5	4	3	1	4	3	5	3	5	5	5	3	No	No	No	Maybe	
24	2017IMSMB007	Female	3	3	4	1	4	3	3	3	2	3	3	3	3	3	4	2	4	1	Yes	Maybe	No	Yes	
25	2017mcs001	Male	3	2	4	1	2	3	2	3	2	3	1	2	2	2	3	4	2	3	3	Yes	Yes	Yes	No
26	2016imsec006	Male	1	3	3	2	4	3	5	3	3	4	3	2	3	2	3	1	3	2	1	Maybe	Yes	No	No
27	2015imses012	Female	1	1	1	1	2	1	2	1	2	1	2	1	1	1	1	1	1	1	2	No	No	Yes	No
28		Female	1	2	3	3	4	3	4	4	4	3	2	3	3	3	4	3	4	2	3	No	Yes	Maybe	No
29		Female	3	4	3	2	3	3	3	5	3	4	2	2	4	3	4	4	3	4	Yes	No	No	Yes	
30	2017mcom007	Male	4	2	3	2	4	3	4	5	3	4	4	4	4	5	3	5	4	4	Yes	Maybe	No	No	
31	2018msn001	Male	2	1	2	2	1	2	3	4	3	3	3	3	3	4	4	3	2	3	2	Yes	No	No	Maybe
32	2018msn001	Male	2	4	2	1	3	2	5	2	2	2	2	2	2	2	2	1	1	2	No	Yes	No	No	

10:33 PM

12/10/2018

# DATA CLEANING(IN R)

In order to do data cleaning in R, we import the data through this command.

```
> o_data<-read.csv("C:\\Users\\DELL\\Desktop\\o_data.csv")
```

So as to check whether our data consist some NA values we have a inbuilt function in R. If it shows "TRUE" then our data must contain NA values otherwise not.

```
> any(is.na(o_data))
[1] TRUE
```

Now to sum the NA values, we have another inbuilt function which adds up all the count of NA values.

```
> sum(is.na(o_data))
[1] 444
```

So to remove these ,we use another R command:

```
c1_data<-na.omit(o_data)
```

Now we again check for NA values in our data(c1\_data):

```
> any(is.na(c1_data))
[1] FALSE
```

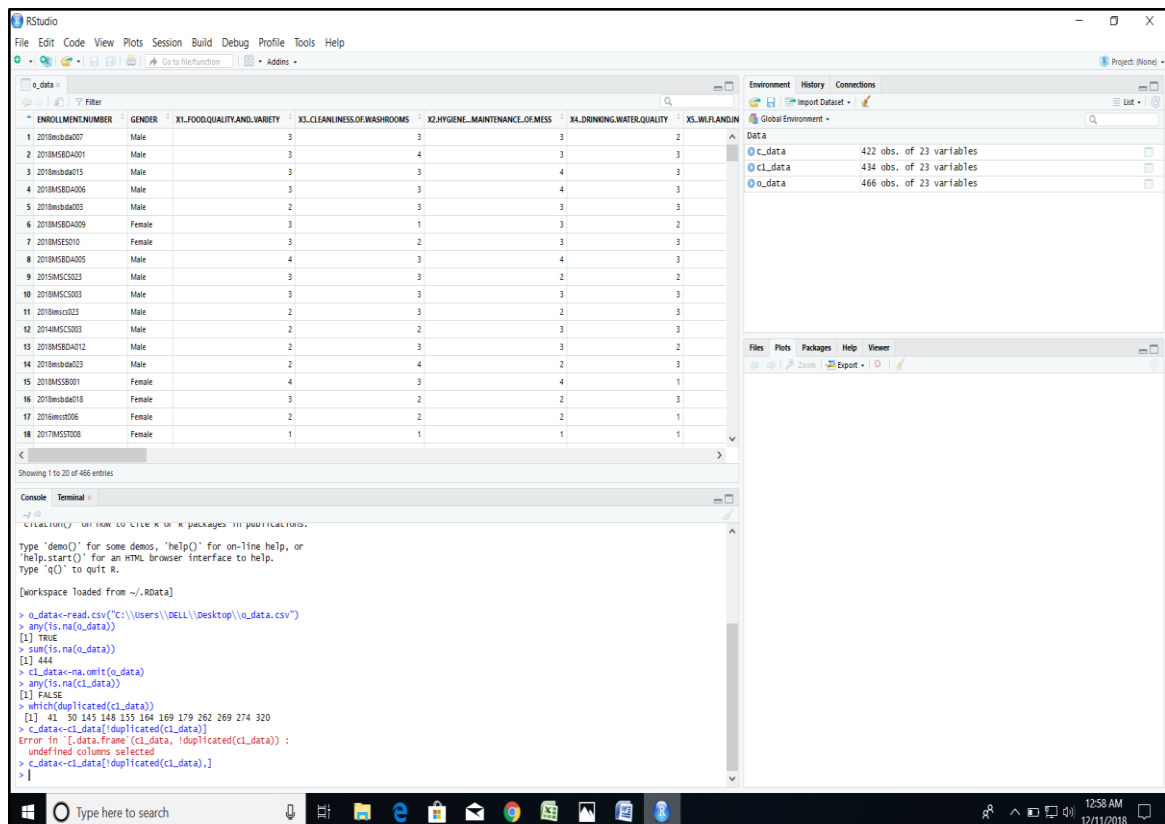
Now we check for any duplicate values via another R inbuilt function so it gives the number of duplicated rows as:

```
> which(duplicated(c1_data))
[1] 41 50 145 148 155 164 169 179 262 269 274 320
```

So as to remove these duplicated rows command is as follows:-

```
c_data<-c1_data[!duplicated(c1_data),]
```

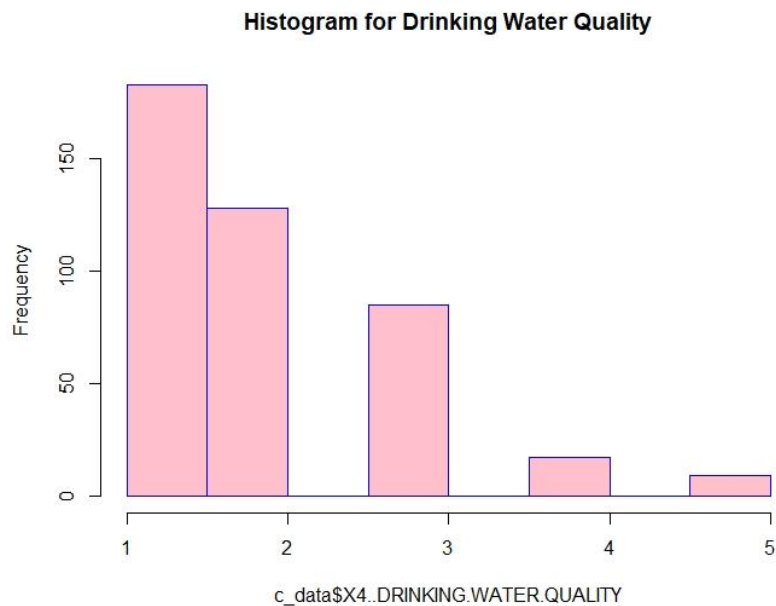
Hence we left with 422 rows with 23 variables as shown in figure:



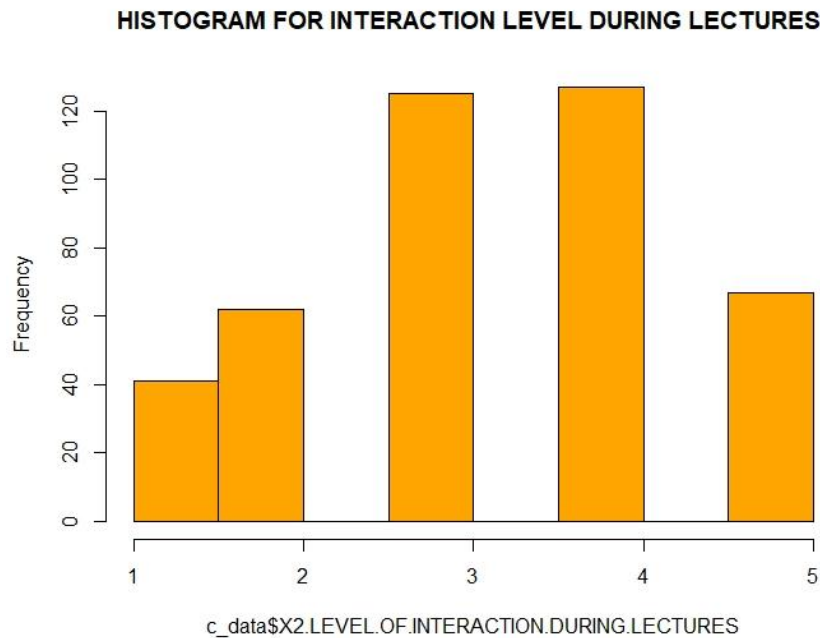
# HISTOGRAM

A histogram is a plot to show the frequency of continuous variable but we have all the discrete values in our data. So in order to plot these we have:

```
> hist(c_data$X4..DRINKING.WATER.QUALITY,main="Histogram for Drinking water  
Quality",border="blue",col="pink")
```



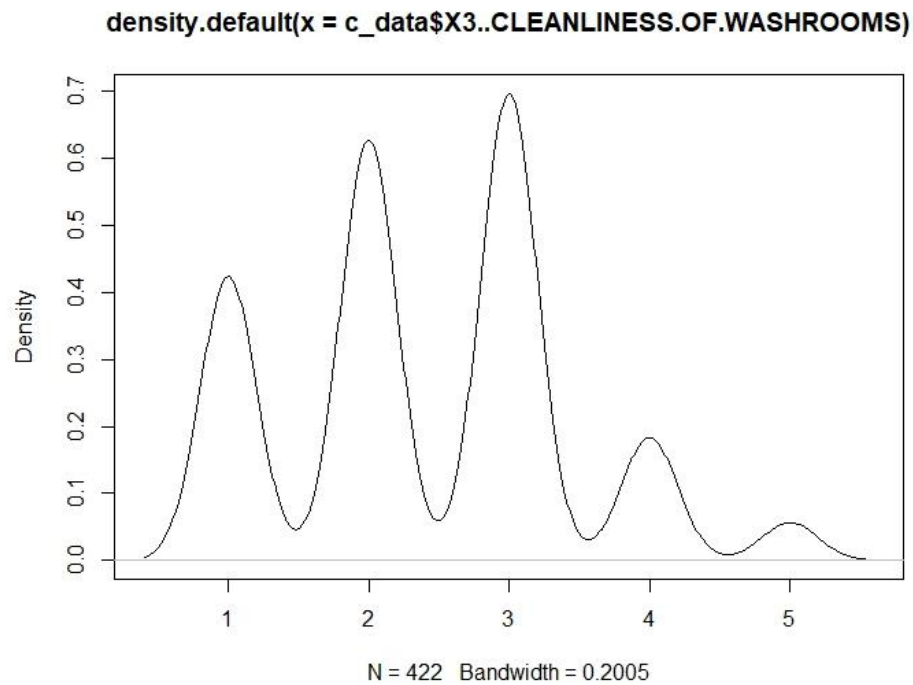
```
> hist(c_data$X2.LEVEL.OF.INTERACTION.DURING.LECTURES,main="HISTOGRAM FOR  
INTERACTION LEVEL DURING LECTURES ",border="black",col="orange")
```



# DENSITY PLOTS

It is a much more effective way of showing the distribution of a variable.

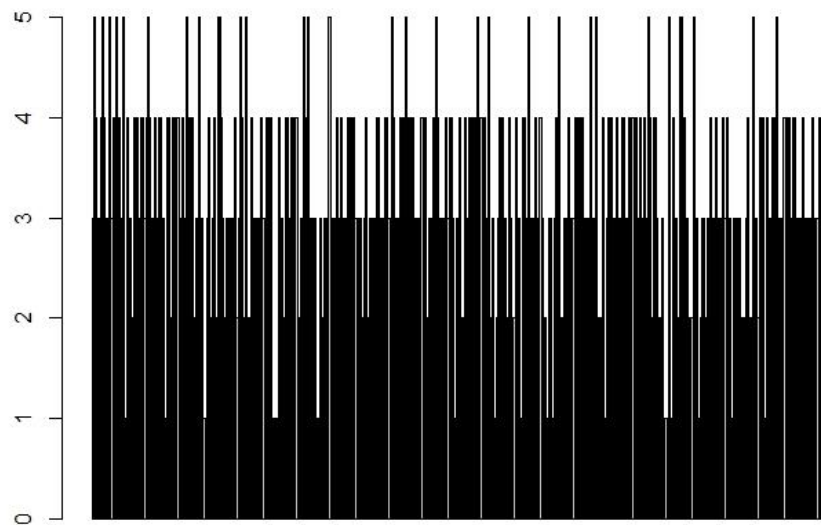
```
> plot(density(c_data$X3..CLEANLINESS.OF.WASHROOMS))
```



## BAR PLOTS

The above two plots discussed is for continuous variable,so in order to plot graph for categorical variable we use barplots.

```
> barplot(c_data$X3.EXAMINATION.PATTERN.AND.GRADING.SYSTEM)
```



## CONTINGENCY TABLE

Contingency tables (also called crosstabs or two-way tables) are used in statistics to summarize the relationship between several categorical variables. A contingency table is a special type of frequency distribution table, where two variables are shown simultaneously.

So here we are taking only two columns from our data in order to define a relationship between two variables.

**Question 1:** Is there a significant relationship between “DRINKING WATER QUALITY” and “HYGIENE MAINTENANCE OF MESS”.

To take only two columns:

```
t1<-c_data[,5:6]
```

Now since we have very less responses in 1 and 5 categories so we merge 1 with 2 and 4 with 5 and make only 3 categories:

```
> t1$X2.HYGIENE...MAINTENANCE..OF.MESS[t1$X2.HYGIENE...MAINTENANCE..OF.MESS==2]<-1
> t1$X2.HYGIENE...MAINTENANCE..OF.MESS[t1$X2.HYGIENE...MAINTENANCE..OF.MESS==4]<-5
> t1$X2.HYGIENE...MAINTENANCE..OF.MESS[t1$X2.HYGIENE...MAINTENANCE..OF.MESS==3]<-2
> t1$X2.HYGIENE...MAINTENANCE..OF.MESS[t1$X2.HYGIENE...MAINTENANCE..OF.MESS==5]<-3

> t1$X4..DRINKING.WATER.QUALITY[t1$X4..DRINKING.WATER.QUALITY==2]<-1
> t1$X4..DRINKING.WATER.QUALITY[t1$X4..DRINKING.WATER.QUALITY==4]<-5
> t1$X4..DRINKING.WATER.QUALITY[t1$X4..DRINKING.WATER.QUALITY==3]<-2
> t1$X4..DRINKING.WATER.QUALITY[t1$X4..DRINKING.WATER.QUALITY==5]<-3
```

Now convert the data into table through table() function in R:

```
> addmargins(table(t1))
```

	X4..DRINKING.WATER.QUALITY			
X2.HYGIENE...MAINTENANCE..OF.MESS	1	2	3	Sum
1	140	25	1	166
2	121	29	15	165
3	50	31	10	91
Sum	311	85	26	422

```
> A<-table(t1)
> A
```

	X4..DRINKING.WATER.QUALITY			
X2.HYGIENE...MAINTENANCE..OF.MESS	1	2	3	
1	140	25	1	
2	121	29	15	
3	50	31	10	

Now we apply chi-square test to check the relationship:-

```
> chisq.test(A)
```

Pearson's Chi-squared test

```
data: A
X-squared = 32.389, df = 4, p-value = 1.593e-06
```

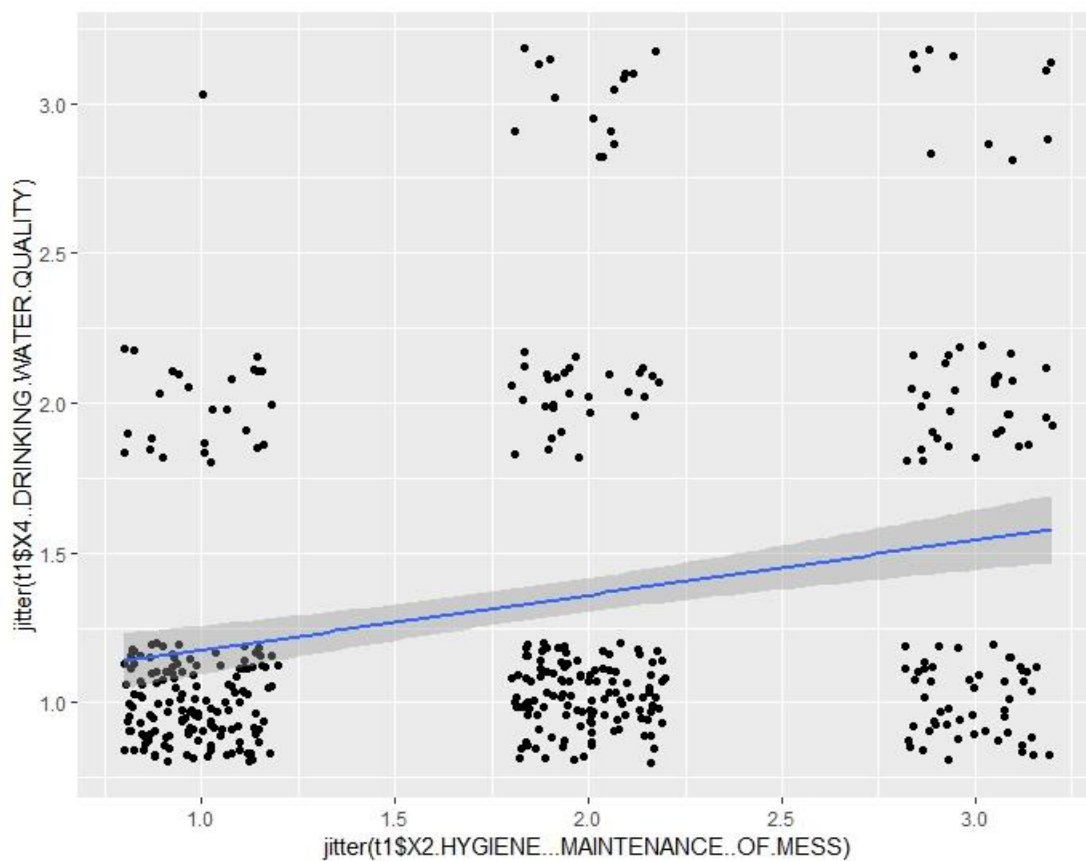


Since we get  $p < 0.05$  Hence these two variables are in fact dependent and the correlation between two variables is given by:-

```
cor(t1$X2.HYGIENE...MAINTENANCE..OF.MESS,t1$X4..DRINKING.WATER.QUALITY)[1]  
0.2576383
```

Now in order to make scatter plot we use the following code: `> g1<-`

```
ggplot(t1,aes(jitter(t1$X2.HYGIENE...MAINTENANCE..OF.MESS),jitter(t1$X  
4..DRINKING.WATER.QUALITY)))+geom_point()+geom_smooth(method="lm")  
> g1
```



**Question 2:** Is there a dependency between Gender and the Awareness about the sparsh foundation?

We follow the same procedure for this also as we have seen in previous question.

To take only two columns:

```
> t2<-c_data[,c(2,20)]
```

Since the data in the data is in non\_numeric form, we convert it into numeric form by using factor() function in R:

```
>t2$GENDER<-factor(t2$GENDER)
```

```
>t2$ X1..ARE.YOU.AWARE.ABOUT.SPASH.FOUNDATION.AND.ITS.ACTIVITIES.<-  
factor(t2$ X1..ARE.YOU.AWARE.ABOUT.SPASH.FOUNDATION.AND.ITS.ACTIVITIES.)
```

Now convert the data into table form through table() function in R:

```
> addmargins(table(t2))  
X1..ARE.YOU.AWARE.ABOUT.SPASH.FOUNDATION.AND.ITS.ACTIVITIES.  
GENDER  Maybe  No  Yes  Sum  
Female   30   42   92  164  
Male    31   95  132  258  
Sum      61  137  224  422
```

```
> B<-table(t2)
```

```
> B
```

```
X1..ARE.YOU.AWARE.ABOUT.SPASH.FOUNDATION.AND.ITS.ACTIVITIES.  
GENDER  Maybe  No  Yes  
Female   30   42   92  
Male    31   95  132
```

Now we apply chi square test to check the dependency.

```
> chisq.test(B)
```

Pearson's Chi-squared test

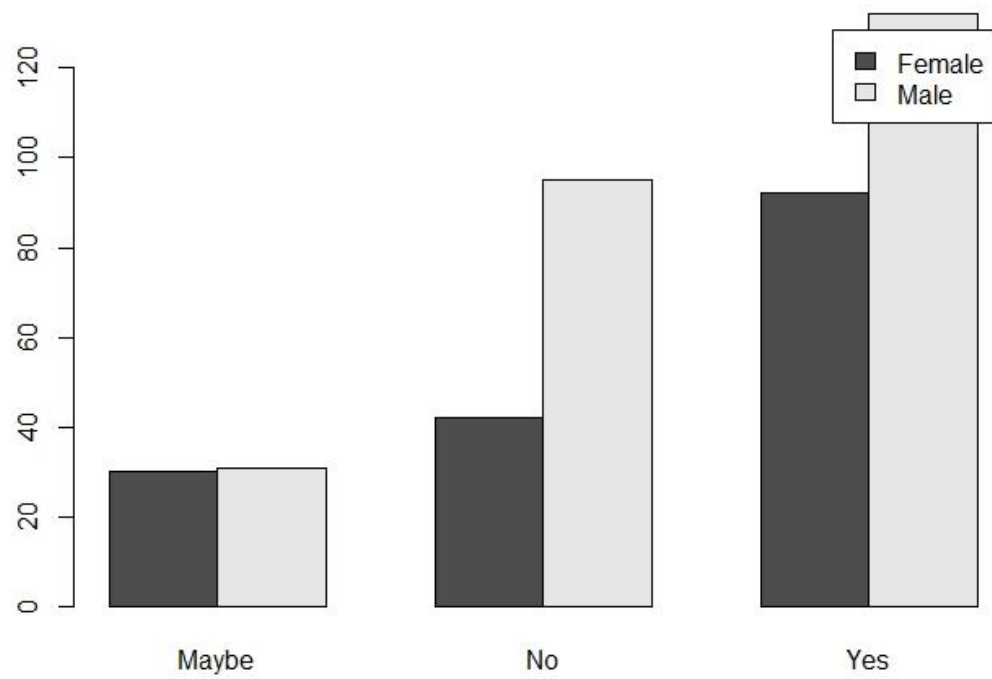
```
data: B  
X-squared = 7.0756, df = 2, p-value = 0.02908
```

Since the p-value<0.05 ,hence we conclude that these two variables are dependent on each other and the correlation is given by following R command:

```
>  
cor(rank(t2$GENDER),rank(t2$X1..ARE.YOU.AWARE.ABOUT.SPASH.FOUNDATION.AND.ITS  
.ACTIVITIES.))  
[1] -0.0119484
```

Here we know that both variables are non-numeric so to find out the correlation between these two variables we use rank().  
We draw barplot for non-numeric data.

```
> barplot(B,beside = TRUE,legend=TRUE,)
```



**Question 3:** We are going to check relation between Gender and cleanliness of washroom. We are taking null hypothesis as both female and male thinks that washrooms are clean in CURAJ campus.

We apply the same procedure :

```
> t3<-c_data[,c(2,9)]
```

Since the data is in non-numeric form so first we convert it into numeric form first:

```
> t3$GENDER<-factor(t3$GENDER)
t3$X7..HOW.SECURE.DO.YOU.FEEL.IN.CAMPUS.<factor(t3$X7..HOW.SECURE.DO.YOU.FEEL.IN.CAMPUS.)
```

Now reduce the 5 categories into 3:

```
> t3$X3..CLEANLINESS.OF.WASHROOMS[t3$X3..CLEANLINESS.OF.WASHROOMS==2]<-1
> t3$X3..CLEANLINESS.OF.WASHROOMS[t3$X3..CLEANLINESS.OF.WASHROOMS==4]<-5
> t3$X3..CLEANLINESS.OF.WASHROOMS[t3$X3..CLEANLINESS.OF.WASHROOMS==3]<-2
> t3$X3..CLEANLINESS.OF.WASHROOMS[t3$X3..CLEANLINESS.OF.WASHROOMS==5]<-3
```

Convert the following data into table:

```
> addmargins(table(t3))
```

	X3..CLEANLINESS.OF.WASHROOMS			
GENDER	1	2	3	Sum
Female	92	54	18	164
Male	131	94	33	258
Sum	223	148	51	422

```
> C<-table(t3)
```

```
> C
```

	X3..CLEANLINESS.OF.WASHROOMS		
GENDER	1	2	3
Female	92	54	18
Male	131	94	33

Now we apply chi-square test on it:

```
> chisq.test(C)
```

Pearson's Chi-squared test

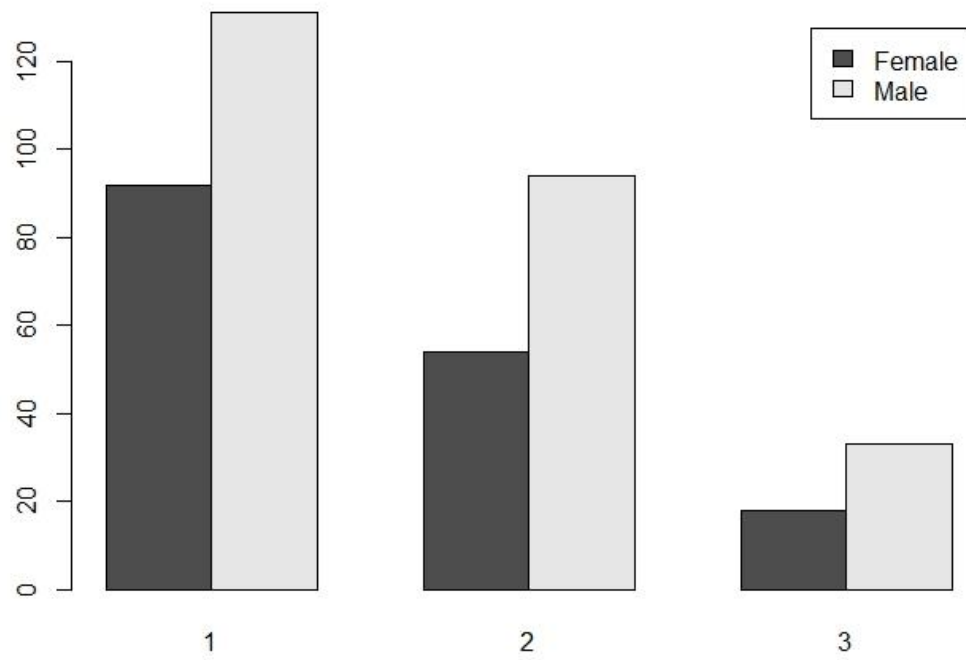
data: C  
X-squared = 1.1625, df = 2, p-value = 0.5592

Now here p-value>0.05 Hence we don't reject the null hypothesis i.e., these two variable are independent of each other. Correlation is given by:

```
> cor(rank(t3$GENDER),rank(t3$X3..CLEANLINESS.OF.WASHROOMS))
[1] 0.0520023
```

Barplot for this:-

```
> barplot(C,legend=TRUE,beside=TRUE)
```



# SUMMARY OF DATA IN R

> summary(c\_data)

```

      ENROLLMENT.NUMBER  GENDER  X1..FOOD.QUALITY.AND..VARIETY
X3..CLEANLINESS.OF.WASHROOMS X2.HYGIENE...MAINTENANCE..OF.MESS
2016imsch012 : 2      : 0  Min.   :1.000      Min.
:1.000      Min.   :1.000
2017imsmb008 : 2      Female:164  1st Qu.:2.000      1st
Qu.:2.000      1st Qu.:2.000
2017MSATS006 : 2      Male  :258  Median :2.000      Median
:2.000      Median :3.000
2017mscs001  : 2      Mean   :2.481      Mean
:2.408      Mean   :2.749
2018IMSBMT012: 2      3rd Qu.:3.000      3rd
Qu.:3.000      3rd Qu.:3.000
2018IMSCH011 : 2      Max.    :5.000      Max.
:5.000      Max.    :5.000
(Other)      :410
X4..DRINKING.WATER.QUALITY X5..WI.FI.AND.INTERNET.SERVICES
X6..SPORTS.AND.GYM X7..HOW.SECURE.DO.YOU.FEEL.IN.CAMPUS.
Min.   :1.000      Min.   :1.000      Min.   :1.000
Min.   :1.000      1st Qu.:2.000      1st Qu.:1.000
1st Qu.:1.000      1st Qu.:3.000      Median :2.000
1st Qu.:3.000      Median :4.000      Median :2.000
Median :2.000      Mean   :1.912      Mean   :2.284
Median :4.000      Mean   :3.486      3rd Qu.:3.000
Mean   :1.912      3rd Qu.:4.000      3rd Qu.:3.000
Mean   :3.486      Max.    :5.000      Max.    :5.000
3rd Qu.:3.000      Max.    :5.000
3rd Qu.:4.000
Max.    :5.000
Max.    :5.000

X1.ORGANISATION.AND.COVERAGE.OF.SYLLABUS
X2.RELEVANCE.AND.REAL.WORLD.APPLICATION
X3.EXAMINATION.PATTERN.AND.GRADING.SYSTEM
Min.   :1.00      Min.   :1.000
Min.   :1.00      1st Qu.:2.000
1st Qu.:3.00      Median :3.000
1st Qu.:3.00      Mean   :2.765
Median :3.00      3rd Qu.:3.000
Median :3.00      Max.    :5.000
Mean   :3.41
Mean   :3.14
3rd Qu.:4.00
3rd Qu.:4.00
Max.    :5.00
Max.    :5.00

X4..CO..CURRICULAR..ACTIVITIES X1.technical.CONTENT.IN.TEACHING
X2.LEVEL.OF.INTERACTION.DURING.LECTURES X3.AVAILABILITY.OUTSIDE.CLASS
Min.   :1.000      Min.   :1.000      Min.
:1.000      Min.   :1.000
1st Qu.:2.000      1st Qu.:2.000      1st
Qu.:3.000      1st Qu.:2.000
Median :3.000      Median :3.000      Median
:3.000      Median :3.000
Mean   :2.687      Mean   :2.953      Mean
:3.277      Mean   :2.882

```

3rd Qu.:4.000  
 Qu.:4.000  
 Max. :5.000  
 :5.000

3rd Qu.:4.000  
 3rd Qu.:4.000  
 Max. :5.000  
 Max. :5.000

3rd  
 Max.

X1.LAB.FACILITIES..EX..EQUIPMENT.  
 X2.AVAILABILITY.OF.BOOKS..E.RESOURCE..STUDY.MATERIALS..ETC.  
 Min. :1.000 Min. :1.000  
 1st Qu.:2.000 1st Qu.:2.000  
 Median :3.000 Median :3.000  
 Mean :2.976 Mean :3.021  
 3rd Qu.:4.000 3rd Qu.:4.000  
 Max. :5.000 Max. :5.000

X3.QUALITY.OF.GUEST.LECTURES..WORKSHOPS..SEMINARS..ETC  
 X1..ARE.YOU.AWARE.ABOUT.SPASH.FOUNDATION.AND.ITS.ACTIVITIES.  
 Min. :1.000 : 0  
 1st Qu.:2.000 Maybe: 61  
 Median :3.000 No :137  
 Mean :2.682 Yes :224  
 3rd Qu.:3.000  
 Max. :5.000

X2..DO.YOU.FEEL.GENDER.DISCRIMINATION.EXISTS.IN.CURAJ.  
 X3..HAVE.YOU.EXPERIENCED.HARASSMENT.IN.ANY.FORM..EX..PHYSICAL..VERBAL..ETC..I  
 N.CURAJ.  
 : 0 : 0  
 Maybe: 75 Maybe: 30  
 No :175 No :337  
 Yes :172 Yes : 55

X4..ARE.YOU.SATISFIED.WITH.THE.CCTV.MONITORING.IN..CURAJ.  
 : 0  
 Maybe: 73  
 No :180  
 Yes :169

# CONCLUSION

By analyzing the survey we have discovered the following conclusions:

>From the histogram of drinking water quality , most of the students rate it for 1 or 2 out of 5. So, the drinking water quality is very bad and students are not satisfied in this concern.

>From the histogram of level of interaction during lecture, most of the students rate in 3-4 out of 5. So, we can think level of interaction during lecture is good and students are satisfied in this concern. It shows beneficial for all the students.

>From the density graph of cleanliness of washroom, vote an average remarks here.

>We have taken three null hypotheses:

1. Drinking water quality and hygiene maintenance of mess are independent of each other.

2. Gender and awareness of sparse foundation is not correlated to each other.

3. Both male and female students think that washrooms are clean in curaj.

>Applying chi-square test, it is proved that drinking water quality and hygiene maintenance of mess are not independent. They are dependent on each other. We have shown also the correlation between them to prove it.

>By chi-square test, it is proved that gender and awareness of sparse foundation are correlated to each other.

>By checking relation between gender and cleanliness of washroom, we accept that both male and female students think that washrooms are clean in curaj. We have tested it by chi-square test and accept this hypothesis.



# REFERENCES

- GOOGLE.COM
- [https://www.tutorialspoint.com/r/r\\_chi\\_square\\_tests.htm](https://www.tutorialspoint.com/r/r_chi_square_tests.htm)
- Statistical Method by Freedman
- <https://stackoverflow.com/>
- <https://www.r-project.org/help.html>