Statistical Analysis
of
University Questionnaire Survey

Submitted to:

Dr. Manas Patra

Submitted by:

Abhishek Chand [2018MSBDA008]

Ankit Singh [2018MSBDA003]

Atul Joshi [2018MSBDA005]

Introduction

This questionnaire is designed to take a survey of the institutional feedback. It consists of five sections comprising 21 questions. The major sections are

1. Academics (Curriculum/Syllabus Evaluation Process)
2. Instructor
3. Hostel and Mess
4. Academic facilities (Labs/E-resources/Workshops etc)
5. Others (General Awareness about University)

The objectives of the feedback system are as follows:

- To help the faculty modification and improvement in the teaching methodology at the University
- To develop a communication bridge between the management and students.
- To maintain the functioning of the University system in the field of hostel & mess facilities, teaching, library, sports, administration, etc at its level best and further improve them.

The feedback was conducted at level of UG, PG and PhD students, and around 400 responses were received. The feedback was collected on the requirements regarding provided facilities, curriculum, teaching quality & services.

Methodology

Respondents included the population of Central University of at level of both the UG and PG and PhD students

A questionnaire in the form of Online Google Form and Offline Feedback Form was used to gather data.
The questionnaire was divided into five-sections that required students to rate their level of satisfaction from a scale of [1 to 5] for 17 questions and [Yes, No, Maybe] for 4 questions.

Input Data

Manual Cleaning Process

A total of 467 data was collected which contained duplicate and spurious information. In order to get relevant and valid data, cleaning was performed at two levels.

Level 1

The data entries where the Enrolment Number was not valid was excluded by manual cleaning using Excel.

Level 2

In case of duplicate entries based on Enrolment number, all entries after first entry were discarded by manual cleaning using Excel.

Separating Input Data

```
> library(readr)
```

```
> df <- read_csv("data.csv")
```

The data is read using read_csv() into a data frame.

The dataframe is also then divided into 3 sections – UG, PG and PhD

```
> bachelors=df[substr(df$`ENROLLMENT NUMBER`,5,5)=='i',]
> bachelors=rbind(bachelors,df[substr(df$`ENROLLMENT
NUMBER`,5,5)=='I',])
> bachelors=rbind(bachelors,df[substr(df$`ENROLLMENT
NUMBER`,5,5)=='b',])
```

```
> bachelors=rbind(bachelors,df[substr(df$`ENROLLMENT
NUMBER`,5,5)=='B',])
```

```
> masters=df[substr(df$`ENROLLMENT NUMBER`,5,5)=='m',]
> masters=rbind(bachelors,df[substr(df$`ENROLLMENT
NUMBER`,5,5)=='M',])
```

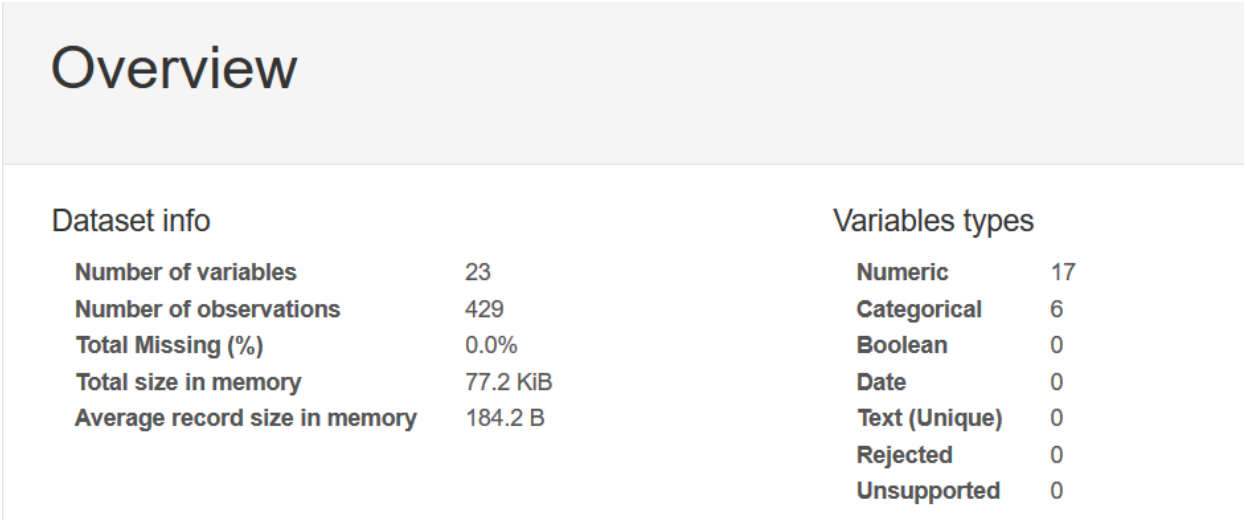
```
> phd=df[substr(df$`ENROLLMENT NUMBER`,5,5)=='p',]
> phd=rbind(bachelors,df[substr(df$`ENROLLMENT
NUMBER`,5,5)=='P',])
```

Profiling

```
from pandas import read_csv
import pandas_profiling as pp

data=read_csv('file:///C:/Users/abhic/Documents/data.csv')
profile=pp.ProfileReport(data)
profile.to_file(outputfile='out_pro.html')
```

Overview



4. DRINKING WATER QUALITY

Numeric

Distinct count	5	**Mean**	1.9068
Unique (%)	1.2%	**Minimum**	1
Missing (%)	0.0%	**Maximum**	5
Missing (n)	0	**Zeros (%)**	0.0%
Infinite (%)	0.0%		
Infinite (n)	0		

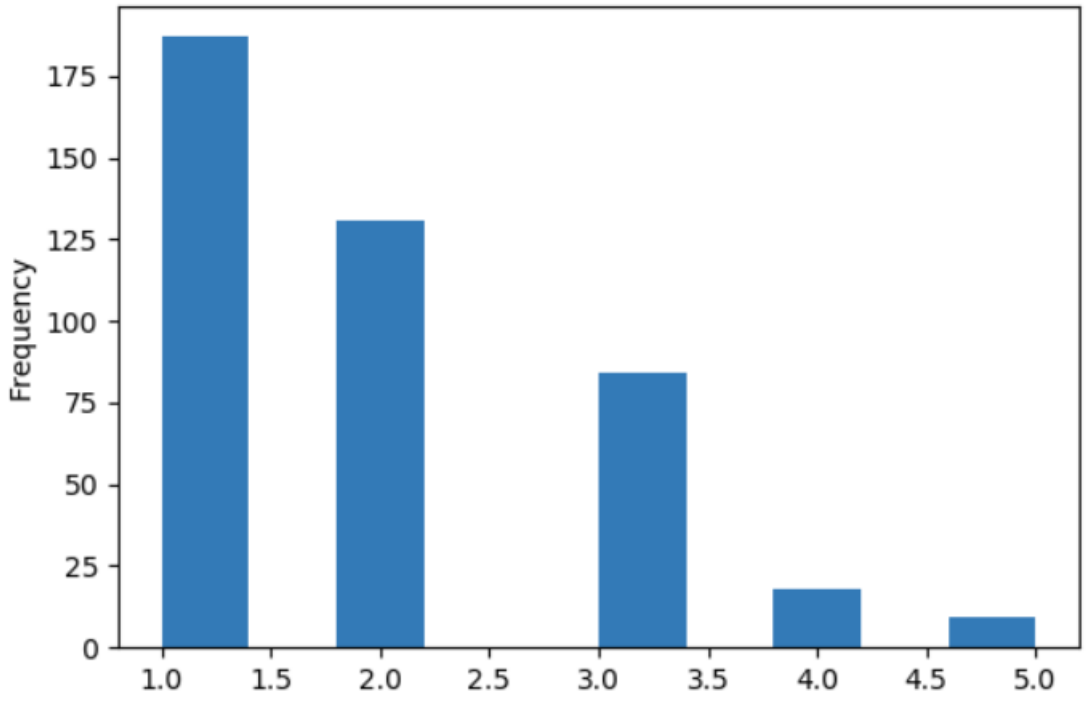
Statistics | Histogram | Common Values | Extreme Values

Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
Median	2
Q3	3
95-th percentile	4
Maximum	5
Range	4
Interquartile range	2

Descriptive statistics

Standard deviation	0.99093
Coef of variation	0.51969
Kurtosis	0.45838
Mean	1.9068
MAD	0.79051
Skewness	0.96958
Sum	818
Variance	0.98194
Memory size	3.4 KiB



Correlation

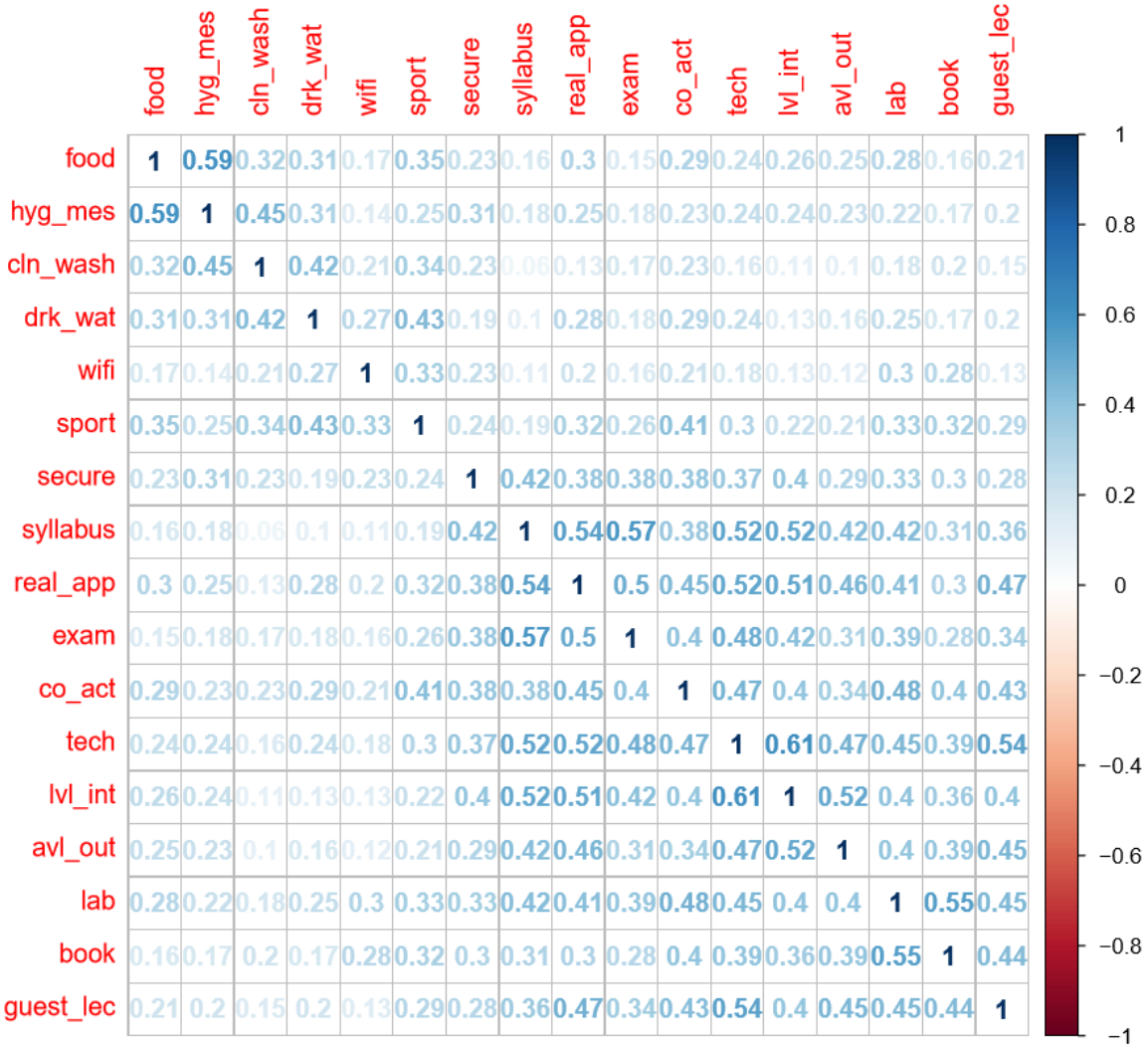
- a quantity measuring the extent of interdependence of variable quantities.

The **corrplot** package is a graphical display of a correlation matrix, confidence interval. It also contains some algorithms to do matrix reordering.

```
> library(corrplot)
```

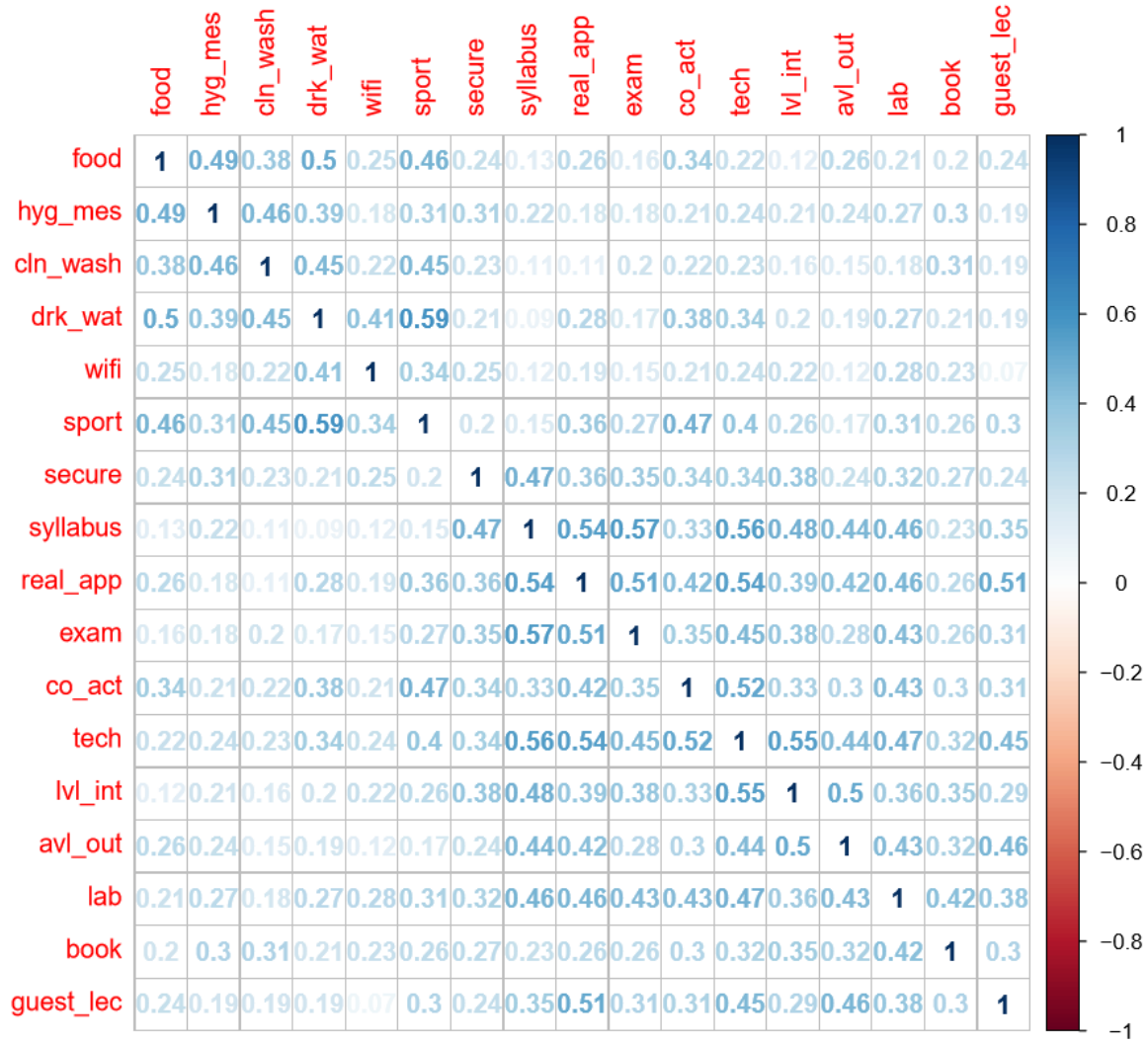
Total

```
> t=df[,3:19]
> colnames(t) <- c("food", "hyg_mes", "cln_wash",
"drk_wat", "wifi", "sport", "secure", "syllabus",
"real_app", "exam", "co_act", "tech", "lvl_int",
"avl_out", "lab", "book", "guest_lec")
> M<-cor(t)
> corrplot(M, method="number")
```



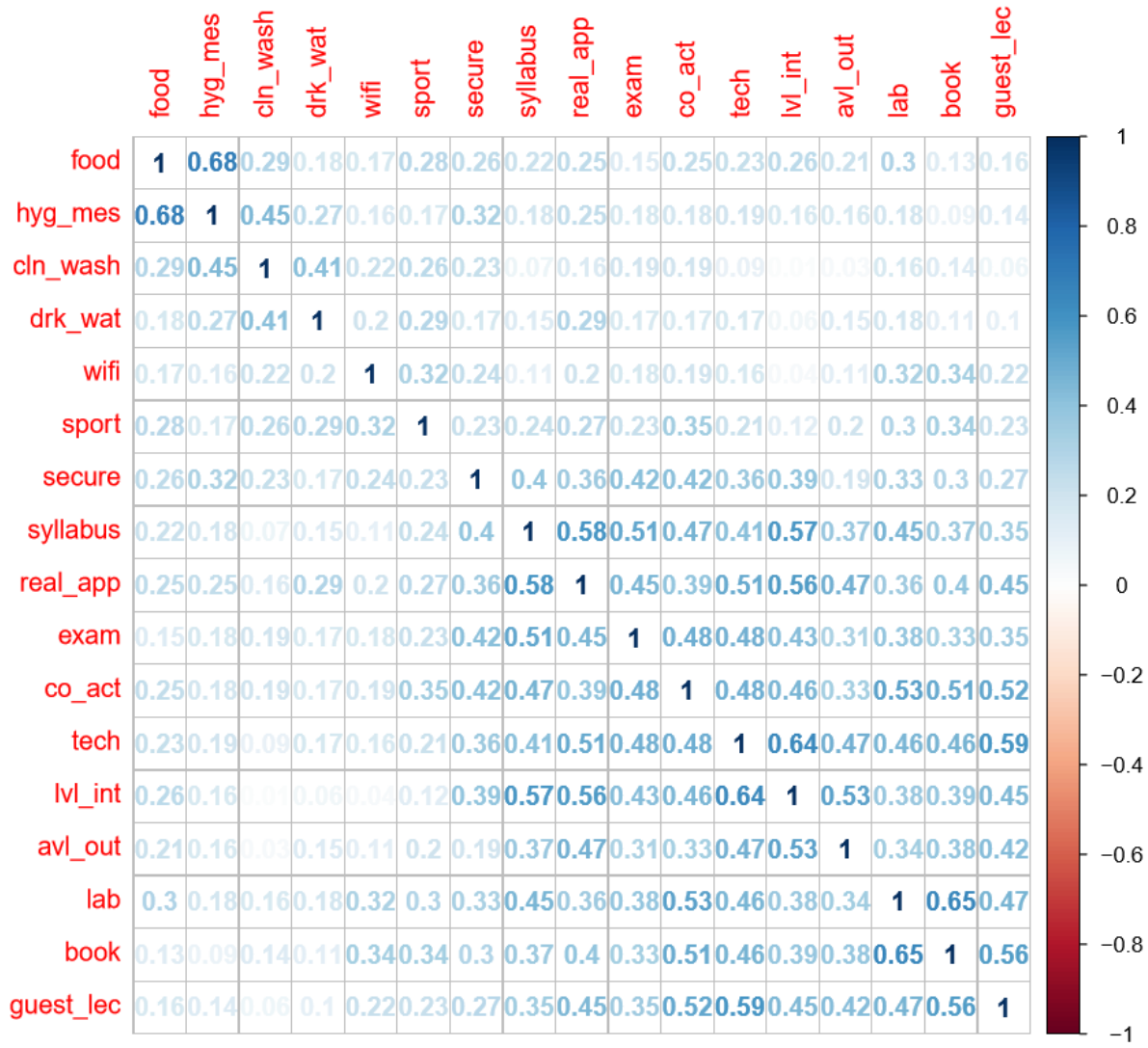
UG

```
> b=bachelors[,3:19]
> colnames(b) <- c("food", "hyg_mes", "cln_wash",
"drk_wat", "wifi", "sport", "secure", "syllabus",
"real_app", "exam", "co_act", "tech", "lvl_int",
"avl_out", "lab", "book", "guest_lec")
> Mb<-cor(b)
> corrplot(Mb, method="number")
```



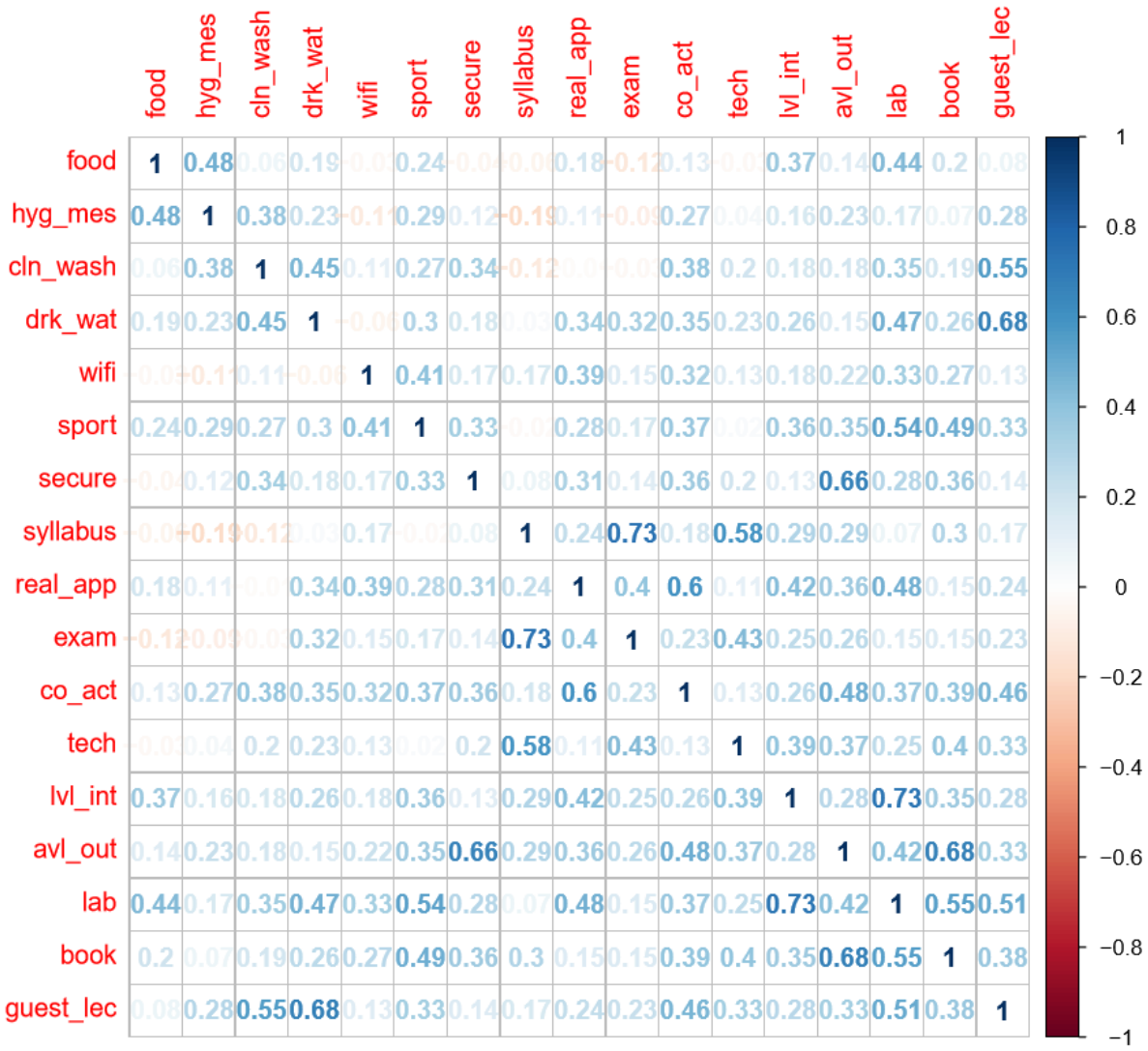
PG

```
> m=masters[,3:19]
> colnames(m) <- c("food", "hyg_mes", "cln_wash",
"drk_wat", "wifi", "sport", "secure", "syllabus",
"real_app", "exam", "co_act", "tech", "lvl_int",
"avl_out", "lab", "book", "guest_lec")
> Mm<-cor(m)
> corrplot(Mm, method="number")
```



PhD

```
> p=phd[,3:19]
> colnames(p) <- c("food", "hyg_mes", "cln_wash",
"drk_wat", "wifi", "sport", "secure", "syllabus",
"real_app", "exam", "co_act", "tech", "lvl_int",
"avl_out", "lab", "book", "guest_lec")
> Mp<-cor(p)
> corrplot(Mp, method="number")
```



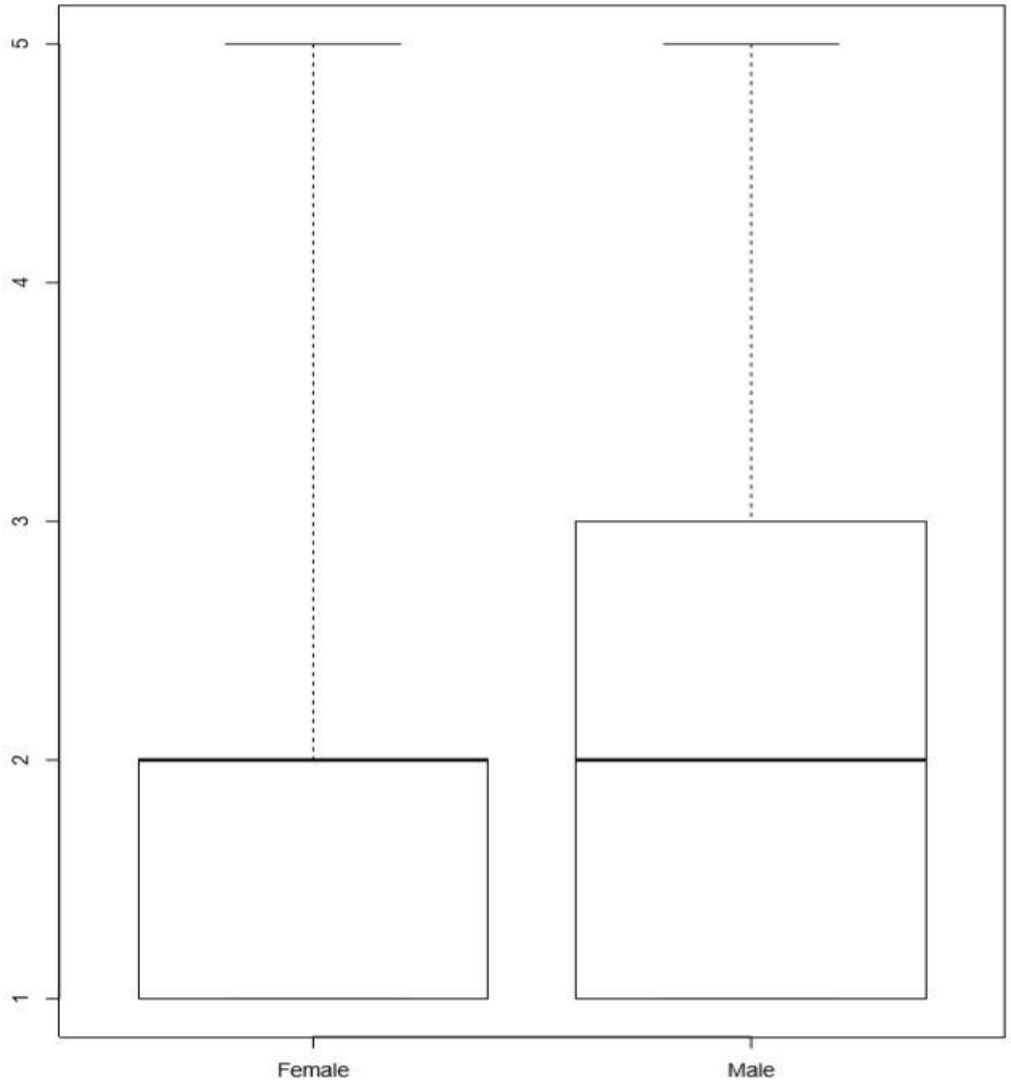
Box Plots

- Produce box-and-whisker plot(s) of the given (grouped) values.
-

Box Plots of Drinking Water Quality grouped by Gender

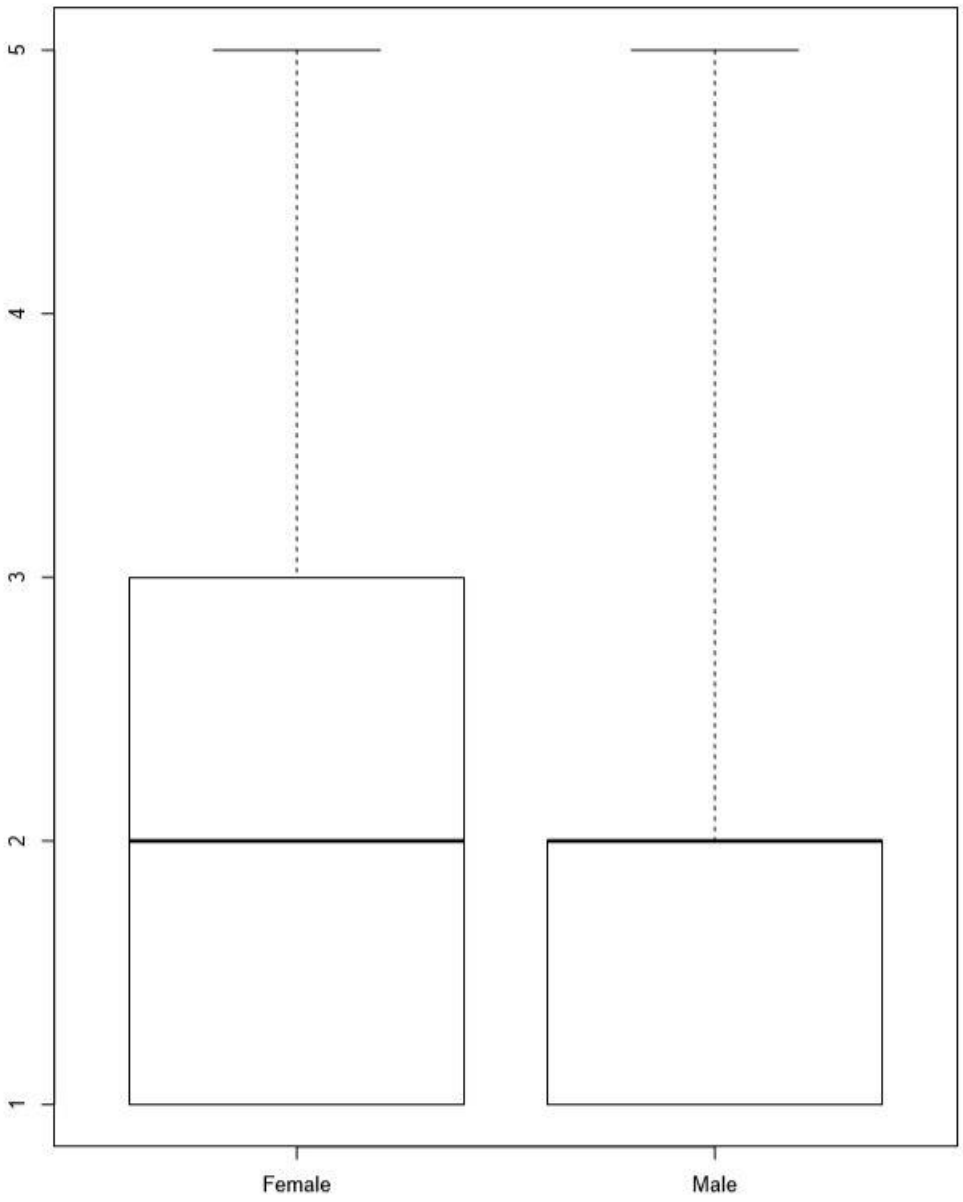
Total

```
> boxplot(ndf$drk_wat ~ ndf$gender, data = ndf, range = 0)
```



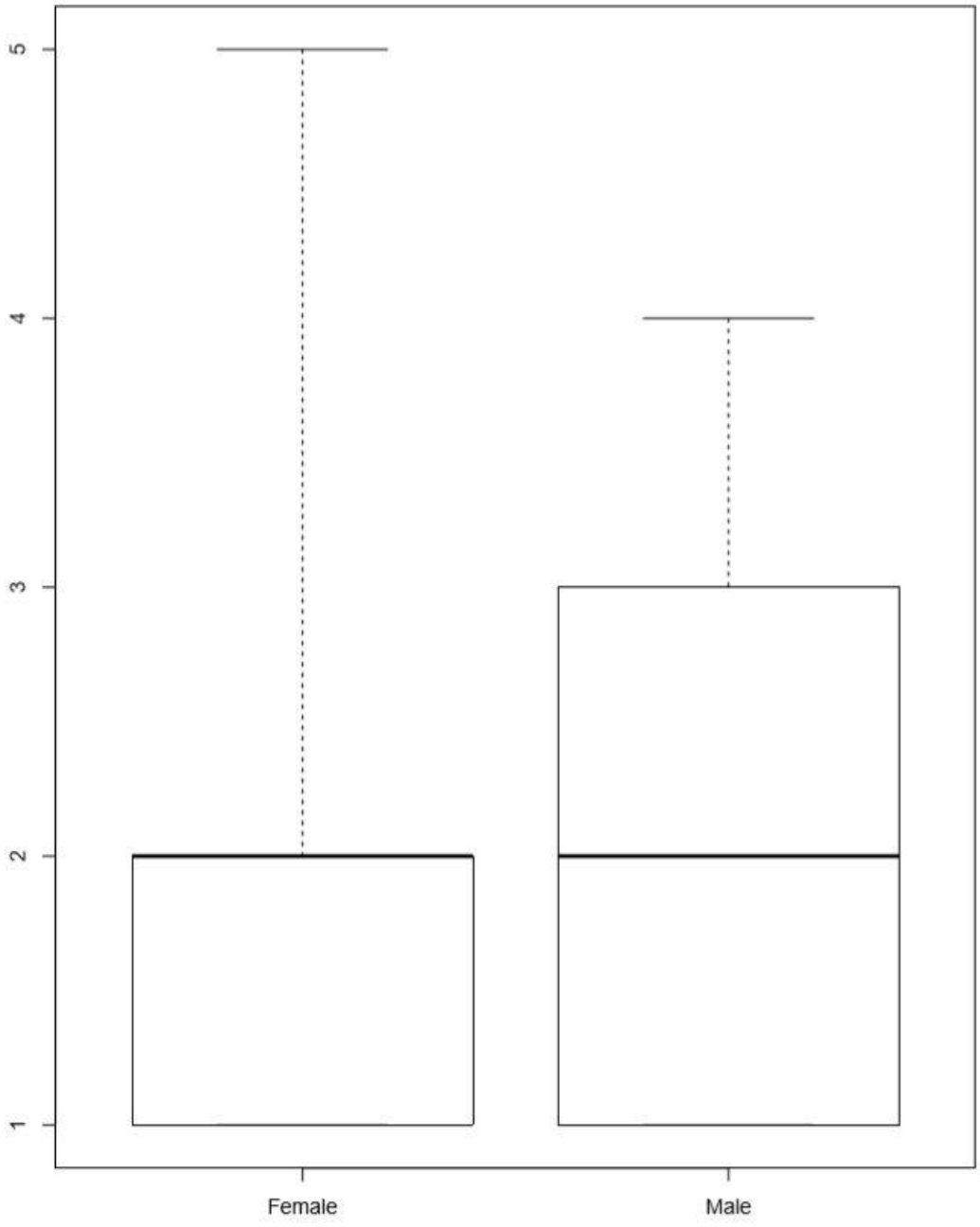
Bachelors

```
> boxplot(nb$drk_wat ~ nb$gender, data = nb, range = 0)
```



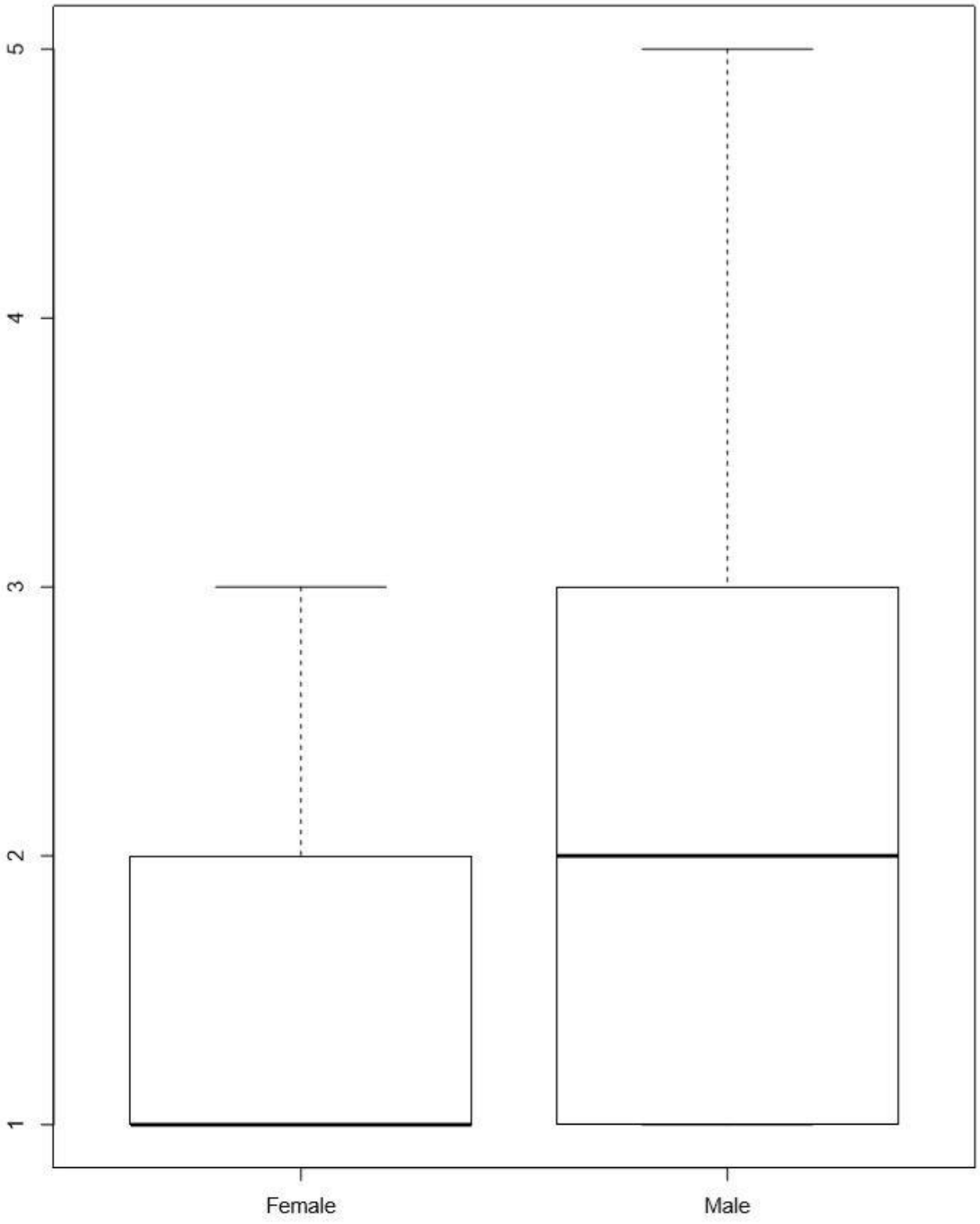
Masters

```
> boxplot(nm$drk_wat ~ nm$gender, data = nm, range = 0)
```



PhD

```
> boxplot(np$drk_wat ~ np$gender, data = np, range = 0)
```



Box Plot Conclusion

From above, we can interpret that:

- Median=2 for all cases except PhD females where it is 1
- Distribution of values is not symmetric
- Distribution is skewed towards lower values

Conclusion:

- Males are more satisfied with Drinking Water Quality as compared to females.
- PhD females are extremely unsatisfied with Drinking Water Quality.
- Rating above 3 are extremely rare. This means that the water quality should be improved.

Chi-Squared Test

All the hypotheses below are tested at 5% level of significance

1. H_0: Gender Discrimination is independent of gender

H_A: Gender Discrimination is gender biased

Total

```
> C=with(ndf,table(gender,gend_discm))
> C
        gend_discm
gender   Maybe  No Yes
  Female    33  60  72
  Male      43 120 101
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 3.5183, df = 2, p-value = 0.1722
```

p-value> 0.05

- ⇨ can't reject the null hypothesis
- ⇨ Gender discrimination is independent of gender bias for the whole population
- ⇨ Gender doesn't determine whether a person feels gender discriminated

UG

```
> C=with(nb,table(gender,gend_discm))
> C
        gend_discm
gender   Maybe No Yes
  Female    14 18  19
  Male      23 62  71
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 4.2719, df = 2, p-value = 0.1181
```

p-value> 0.05

- ⇨ can't reject the null hypothesis
- ⇨ Gender discrimination is independent of gender bias for the UG population
- ⇨ Gender doesn't determine whether a person feels gender discriminated

PG

```
> C=with(nm,table(gender,gend_discm))
> C
        gend_discm
gender   Maybe No Yes
  Female    15 36  49
  Male      17 45  24
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 8.6821, df = 2, p-value = 0.01302
```

p-value< 0.05

- ⇨ rejecting the null hypothesis
- ⇨ Gender discrimination is gender biased for the PG population
- ⇨ Gender determines whether a person feels gender discriminated for PG population.

PhD

```
> C=with(np,table(gender,gend_discm))
> C
        gend_discm
gender   Maybe No Yes
  Female     4  5   4
  Male       3 13   6
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 1.9105, df = 2, p-value = 0.3847
```

p-value> 0.05

- ⇨ can’t reject the null hypothesis
- ⇨ Gender discrimination is independent of gender bias for the PhD population
- ⇨ Gender doesn’t determine whether a person feels gender discriminated

2 H_0: Technical content in teaching is independent of real-world applications in course

H_A: Technical content in teaching is dependent on real world applications in course

Total

```
> C=with(ndf,table(tech,real_app))
> C
    real_app
tech  1  2  3  4  5
   1 24 15  5  0  6
   2 18 47 26  6  0
   3 10 34 81 19  0
   4  5 12 36 38  7
   5  1  2 12 13 12
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 225.2, df = 16, p-value < 2.2e-16
```

p-value< 0.05

- ⇨ rejecting the null hypothesis
- ⇨ technical content in teaching and real-world applications of course are not independent for whole population

UG

```
> C=with(nb,table(tech,real_app))
> C
    real_app
tech  1  2  3  4  5
   1 20  9  2  0  1
   2 13 33 18  3  0
   3  6 16 39 10  0
   4  3  5 11  8  2
   5  0  0  3  3  2
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 107.05, df = 16, p-value = 1.626e-15
```

p-value< 0.05

- ⇨ rejecting the null hypothesis
- ⇨ technical content in teaching and real-world applications of course are not independent for UG population.
- ⇨

PG

```
> C=with(nm,table(tech,real_app))
> C
    real_app
tech  1  2  3  4  5
   1  2  5  2  0  1
   2  4 14  6  3  0
   3  4 15 31  8  0
   4  2  7 22 29  3
   5  0  2  8 10  8
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 81.312, df = 16, p-value = 9.645e-11
```

p-value< 0.05

- ⇨ rejecting the null hypothesis
- ⇨ technical content in teaching and real-world applications of course are not independent for PG population.

PhD

```
> C=with(np,table(tech,real_app))
> C
    real_app
tech  1  2  3  4  5
   1  2  1  1  0  4
   2  1  0  2  0  0
   3  0  3 10  1  0
   4  0  0  3  1  2
   5  1  0  1  0  2
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 22.718, df = 16, p-value = 0.1215
```

p-value> 0.05

- ⇨ can’t reject the null hypothesis
- ⇨ technical content in teaching and real-world applications of course are independent for PhD population.

3. H_0: Cases of harassment are independent of gender

H_A: Cases of harassment is gender biased

Total

```
> C=with(ndf,table(gender,harass))
> C
        harass
gender   Maybe  No Yes
  Female    12 137  16
  Male      19 207  38
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 2.0509, df = 2, p-value = 0.3586
```

p-value> 0.05

- ⇨ can't reject the null hypothesis
- ⇨ Cases of harassment are independent of gender

UG

```
> C=with(nb,table(gender,harass))
> C
        harass
gender   Maybe  No Yes
  Female     5  38   8
  Male      10 122  24
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 0.68102, df = 2, p-value = 0.7114
```

p-value> 0.05

- ⇨ can't reject the null hypothesis
- ⇨ Cases of harassment are independent of gender for UG population

PG

```
> C=with(nm,table(gender,harass))
> C
        harass
gender   Maybe No Yes
  Female     5 89   6
  Male       8 70   8
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 2.2072, df = 2, p-value = 0.3317
```

p-value> 0.05

- ⇨ can’t reject the null hypothesis
- ⇨ Cases of harassment are independent of gender for PG population

PhD

```
> C=with(np,table(gender,harass))
> C
        harass
gender   Maybe No Yes
  Female     2  9   2
  Male       1 15   6
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 1.6266, df = 2, p-value = 0.4434
```

p-value> 0.05

- ⇨ can’t reject the null hypothesis
- ⇨ Cases of harassment are independent of gender for PhD population

4. H_0: Satisfaction with Exam Pattern and Grading System is independent of Syllabus Quality (Organization and Coverage of Syllabus)

H_A: Satisfaction with Exam Pattern and Grading System is not independent of Syllabus Quality (Organization and Coverage of Syllabus)

Total

```
> C=with(ndf,table(exam,syllabus))
> C
    syllabus
exam  1  2  3  4  5
   1 22  6  6  3  1
   2  2 18 24 12  1
   3  4 16 79 56 16
   4  3  6 31 56 36
   5  0  1  3  6 21
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 287.6, df = 16, p-value < 2.2e-16
```

p-value< 0.05

- ⇨ rejecting the null hypothesis
- ⇨ Satisfaction with Exam Pattern and Grading System is not independent of Syllabus Quality (Organization and Coverage of Syllabus)

UG

```
> C=with(nb,table(exam,syllabus))
> C
    syllabus
exam  1  2  3  4  5
   1 14  5  4  1  1
   2  1 11 14  6  1
   3  3  7 46 25  5
   4  1  2 16 16 18
   5  0  1  1  4  4
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 132.13, df = 16, p-value < 2.2e-16
```

p-value< 0.05

- ⇨ rejecting the null hypothesis

- ⇨ Satisfaction with Exam Pattern and Grading System is not independent of Syllabus Quality (Organization and Coverage of Syllabus) for UG population

PG

```
> C=with(nm,table(exam,syllabus))
> C
    syllabus
exam  1  2  3  4  5
   1  4  1  2  2  0
   2  1  5  9  5  0
   3  1  8 25 29 11
   4  1  4 11 35 15
   5  0  0  1  2 14
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 104.54, df = 16, p-value = 4.842e-15
```

p-value< 0.05

- ⇨ rejecting the null hypothesis
- ⇨ Satisfaction with Exam Pattern and Grading System is not independent of Syllabus Quality (Organization and Coverage of Syllabus) for PG population

PhD

```
> C=with(np,table(exam,syllabus))
> C
    syllabus
exam 1 2 3 4 5
   1 4 0 0 0 0
   2 0 2 1 1 0
   3 0 1 8 2 0
   4 1 0 3 5 3
   5 0 0 1 0 3
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 54.64, df = 16, p-value = 4.075e-06
```

p-value< 0.05

- ⇨ rejecting the null hypothesis
- ⇨ Satisfaction with Exam Pattern and Grading System is not independent of Syllabus Quality (Organization and Coverage of Syllabus) for PhD population

5. H_0: Gender discrimination is independent of awareness towards Sparsh Foundation

H_A: Gender discrimination is not independent of awareness towards Sparsh Foundation

Total

```
> C=with(ndf,table(gend_discm,sparsh))
> C
           sparsh
gend_discm Maybe No Yes
     Maybe    18 23  35
     No       28 57  95
     Yes      16 61  96
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 9.3056, df = 4, p-value = 0.0539
```

p-value> 0.05

- ⇨ can't reject the null hypothesis
- ⇨ Gender discrimination is not independent of awareness towards Sparsh Foundation

UG

```
> C=with(nb,table(gend_discm,sparsh))
> C
           sparsh
gend_discm Maybe No Yes
     Maybe     6 13  18
     No        9 29  42
     Yes       7 35  48
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 2.048, df = 4, p-value = 0.7269
```

p-value> 0.05

- ⇨ can't reject the null hypothesis
- ⇨ Gender discrimination is not independent of awareness towards Sparsh Foundation for UG population

PG

```
> C=with(nm,table(gend_discm,sparsh))
> C
          sparsh
gend_discm Maybe No Yes
     Maybe    10 10  12
     No       16 26  39
     Yes       8 26  39
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 6.542, df = 4, p-value = 0.1622
```

p-value> 0.05

- ⇨ can’t reject the null hypothesis
- ⇨ Gender discrimination is not independent of awareness towards Sparsh Foundation for PG population.

PhD

```
> C=with(np,table(gend_discm,sparsh))
> C
          sparsh
gend_discm Maybe No Yes
     Maybe     2  0   5
     No        3  2  13
     Yes       1  0   9
> chisq.test(C)

        Pearson's Chi-squared test

data:  C
X-squared = 3.0226, df = 4, p-value = 0.554
```

p-value> 0.05

- ⇨ can’t reject the null hypothesis
- ⇨ Gender discrimination is not independent of awareness towards Sparsh Foundation for PhD population.

Conclusion

- 42.9 % students are not satisfied by CCTV MONITORING IN CURAJ. This could be improved to make the students in campus feel more secure.
- 43.6 % students rate DRINKING WATER QUALITY as 1. This shows that most students are extremely dissatisfied by water quality, and it could be improved.
- 35.4 % responses for FOOD QUALITY AND VARIETY are 2, this means that Food is barely average in the University and some steps could be taken to improve it.
- 32.6 % students rated the SECURE FEELING IN CAMPUS as 4. This shows that the campus is secure but further improvement is welcome.
- 30.8 % students rated SPORTS & GYM as 1. This show that the facility provided in Gym is extremely poor and also equipment should be improved.

References

1. Statistics: David Freedman, Pobert Pisani & Roger Purves, W. W. Norten & Co. 4th Edition 2007.
2. https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html
3. https://www.rdocumentation.org/packages/graphics/versions/3.5.1/topics/boxplot
4. https://www.rdocumentation.org/packages/stats/versions/3.5.1/topics/chisq.test