Experiment Lab-II

Computation of tables and graphs-summary statistics

Aim: To represent the various types of data using tabulation and graphical representation

Question No1:-Computation of tables and graphs-summary statistics for employee data

Creating vector:-

```
>empid=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)        #creating  a vector empid
> empid
      [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
> age=c(30,37,45,32,50,60,35,32,34,43,32,30,43,50,60)
                                                    # creating a  vector age
> age
       [1] 30 37 45 32 50 60 35 32 34 43 32 30 43 50 60
> sex=c(0,1,0,1,1,1,0,0,1,0,0,1,1,0,0)
> sex
 [1] 0 1 0 1 1 1 0 0 1 0 0 1 1 0 0
> status=c(1,1,2,2,1,1,1,2,2,1,2,1,2,1,2)
> status
       [1] 1 1 2 2 1 1 1 2 2 1 2 1 2 1 2
```

Creating a data frame (Combining vectors):

```
> empinfo=data.frame(empid,age,sex,status)
```

> empinfo

	empid	age	sex	status
1	1	30	0	1
2	2	37	1	1
3	3	45	0	2
4	4	32	1	2
5	5	50	1	1
6	6	60	1	1
7	7	35	0	1
8	8	32	0	2
9	9	34	1	2
10	10	43	0	1
11	11	32	0	2
12	12	30	1	1
13	13	43	1	2
14	14	50	0	1
15	15	60	0	2

> empinfo$sex=factor(empinfo$sex,labels=c("male","female"))

> empinfo$status=factor(empinfo$status,labels=c("staff","faculty"))

>empinfo

	empid	age	sex	status
1	1	30	male	staff
2	2	37	female	staff
3	3	45	male	faculty
4	4	32	female	faculty
5	5	50	female	staff
6	6	60	female	staff
7	7	35	male	staff
8	8	32	male	faculty
9	9	34	female	faculty
10	10	43	male	staff
11	11	32	male	faculty
12	12	30	female	staff
13	13	43	female	faculty
14	14	50	male	staff
15	15	60	male	faculty

#The following command shows male data only

```
> sexm=subset(empinfo,empinfo$sex=='male')
> sexm    #it shows Male data only
   empid age  sex  status
1      1  30 male   staff
3      3  45 male faculty
7      7  35 male   staff
8      8  32 male faculty
10    10  43 male   staff
11    11  32 male faculty
14    14  50 male   staff
15    15  60 male faculty
```

#The following command shows female data only

```
> sexf=subset(empinfo,empinfo$sex=='female')
> sexf
   empid age    sex  status
2      2  37 female   staff
4      4  32 female faculty
5      5  50 female   staff
6      6  60 female   staff
9      9  34 female faculty
12    12  30 female   staff
13    13  43 female faculty
```

? Similarly create staff data set and faculty dataset

- Summary statistics for empinfo data

```
> summary(empinfo)
     empid           age             sex        status
 Min.   : 1.0   Min.   :30.00   male  :8   staff  :8
 1st Qu.: 4.5   1st Qu.:32.00   female:7   faculty:7
 Median : 8.0   Median :37.00
 Mean   : 8.0   Mean   :40.87
 3rd Qu.:11.5   3rd Qu.:47.50
 Max.   :15.0   Max.   :60.00
```

- Summary statistics for male and female employees data

```
> summary(sexf)
     empid            age            sex        status
 Min.   : 2.000   Min.   :30.00   male  :0   staff  :4
 1st Qu.: 4.500   1st Qu.:33.00   female:7   faculty:3
 Median : 6.000   Median :37.00
 Mean   : 7.286   Mean   :40.86
 3rd Qu.:10.500   3rd Qu.:46.50
 Max.   :13.000   Max.   :60.00
> summary(sexm)
     empid            age            sex        status
 Min.   : 1.000   Min.   :30.00   male  :8   staff  :4
 1st Qu.: 6.000   1st Qu.:32.00   female:0   faculty:4
 Median : 9.000   Median :39.00
 Mean   : 8.625   Mean   :40.88
 3rd Qu.:11.750   3rd Qu.:46.25
 Max.   :15.000   Max.   :60.00
```

- Summary statistics for age

> summary(empinfo$age)

Min. 1st Qu. Median Mean 3rd Qu. Max.

30.00 32.00 37.00 40.87 47.50 60.00

- Creating one-way table

1. For sex

```
> table1=table(empinfo$sex)
> table1

  male female
     8      7
```

2. For status

```
> table2=table(empinfo$status)
> table2

  staff faculty
      8       7
```

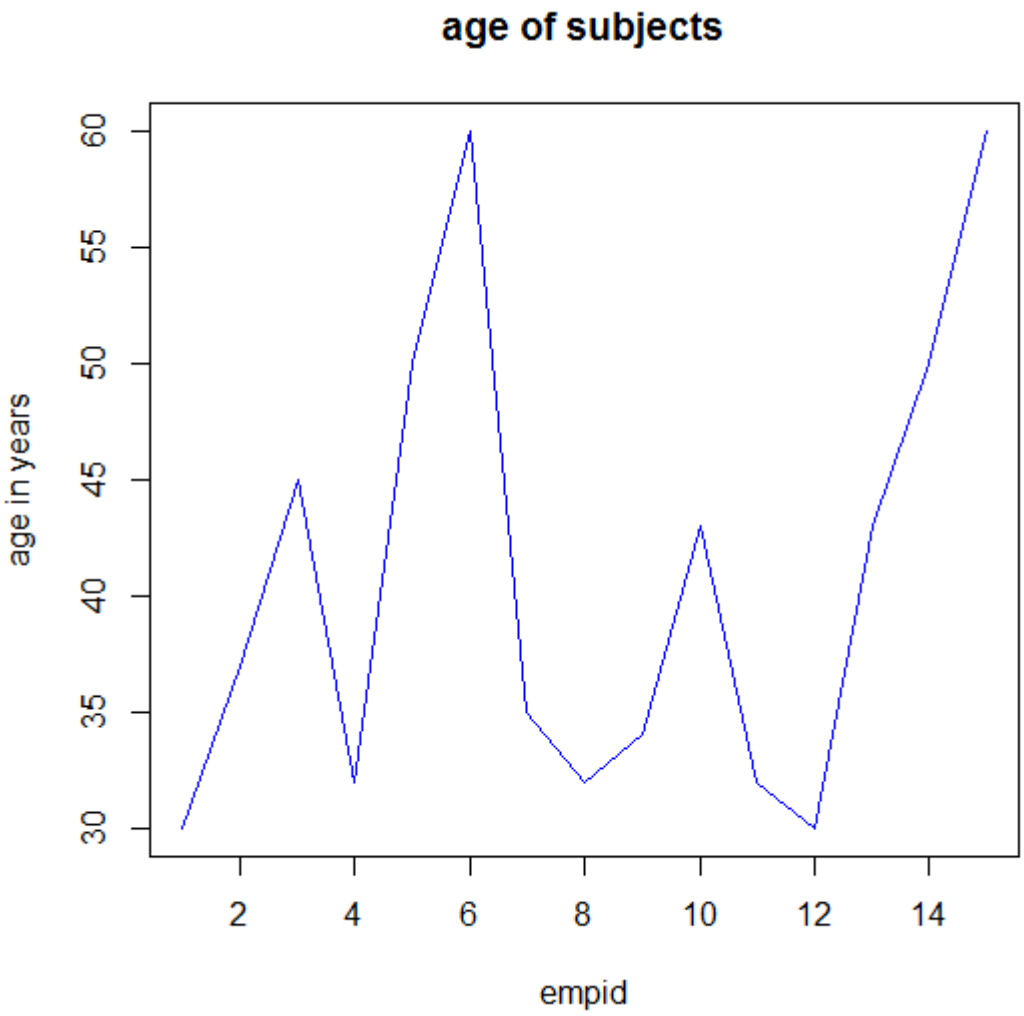
- Creating two-way table

```
> table3=table(empinfo$sex,empinfo$status)
> table3

        staff faculty
male        4       4
female      4       3
```

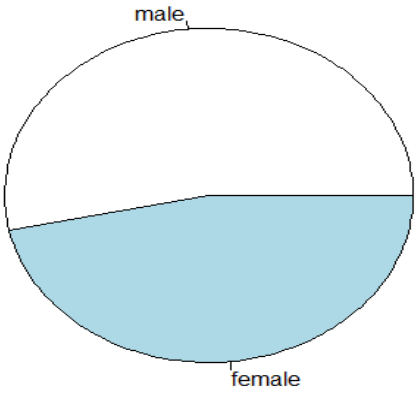
Gaphical reperesentation in R

>plot(empinfo$age,type="l",main="age of subjects",xlab="empid",ylab="age in years",col="blue")



Pie Chart:-

> table4<-table(empinfo$sex)
> pie(table4)

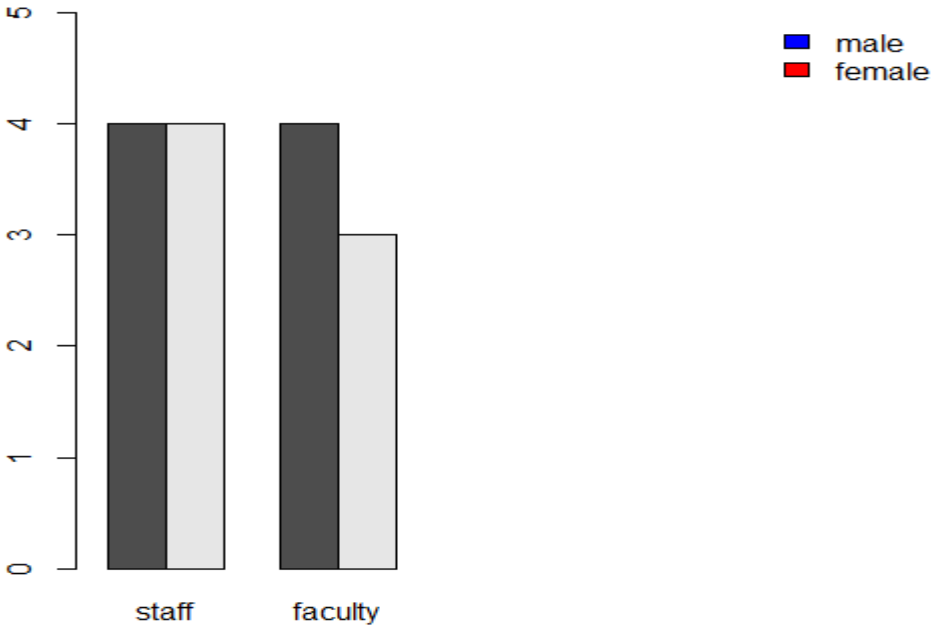


> table5=table(empinfo$sex,empinfo$status)

> barplot(table5,beside=T,xlim=c(1,15),ylim=c(0,5))

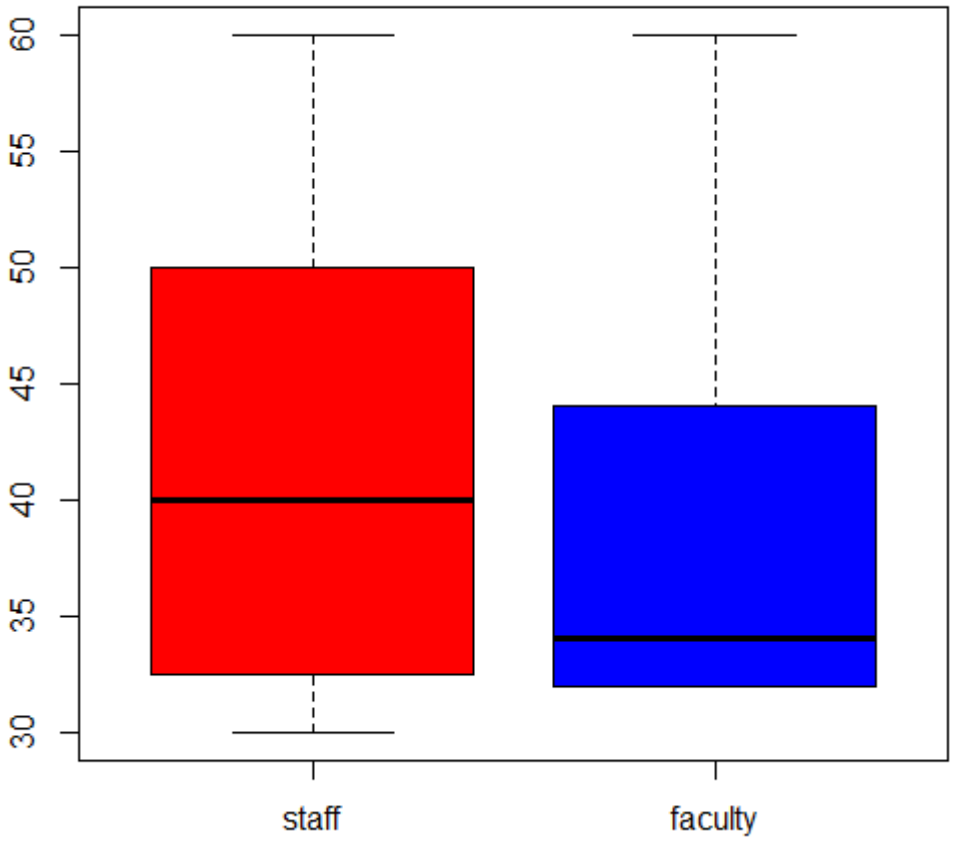
>legend("topright",legend=rownames(table5),fill=c('blue','red'),bty="n")

`bty`	the type of box to be drawn around the legend. The allowed values are `"o"` (the default) and `"n"`.



BOXPLOT:-

- **boxplot(empinfo$age~empinfo$status,col=c("red","blue"))**



Question2 :-(Life satisfaction data)

#downloading the Raw data into R

- Life_Data<-read.csv("C:\\Users\\aadmin\\Desktop\\mokesh\\data1.csv")

read.

```
# Printing the data
>Life__Data
#checking the summary statistics of the continues variables
>summary(Life_Data$IncomeC)
>summary(Life_Data$Income)
# creating labels
>Life_Data$Gender=factor(Life_Data$Gender,labels=c("male","female"))
>Life_Data$Married=factor(Life_Data$Married,labels=c("no","yes"))
>Life_Data$Smoke=factor(Life_Data$Smoke,labels=c("no","yes"))
>Life_Data$Finish=factor(Life_Data$Finish,labels=c("no","yes"))
#checking the Frequency of the discrete variables
table(Life_Data$Age)
table(Life_Data$Gender)
table(Life_Data$Married)
table(Life_Data$Smoke)
table(Life_Data$Finish)

#generating cross tabulations
table(Life_Data$Gender,Life_Data$Married)
table(Life_Data$Gender,Life_Data$Finish)
table(Life_Data$Gender,Life_Data$Finish,Life_Data$Married)
table(Life_Data$LifeSat,Life_Data$Smoke)
```

#generating Graphs

```
counts <- table(Life_Data$LifeSatC )

barplot(counts, main="Lift Satisfaction Distribution",xlab="Score on Life
Satisfaction",)
```

Question 3: Challenging Experiment

Create your own (Student Record) dataset and do the summary statistics and graphs with interpretation. Use atleast 50 observations with five variables.