

## Experiment-2

### Computing Summary Statistics /plotting and visualizing data using Tabulation and Graphical Representations

#### Aim:

The purpose of this experiment is to learn the different alignment of data set and various graphical representations in R

#### Procedure:

Step by step procedure to conduct the required experiment –

1. Arrangement of data using various R functions
2. Visualize the data set using various R functions

#### Code and Results:

```
#creating a vector empid
empid=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)
empid

## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

# creating a vector age
age=c(30,37,45,32,50,60,35,32,34,43,32,30,43,50,60)
age

## [1] 30 37 45 32 50 60 35 32 34 43 32 30 43 50 60

# creating a vector gender
gender=c(0,1,0,1,1,1,0,0,1,0,0,1,1,0,0)
gender

## [1] 0 1 0 1 1 1 0 0 1 0 0 1 1 0 0

# creating a vector status
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

# reating a data frame (Combining vectors)
empinfo=data.frame(empid,age,gender,status)
empinfo

##      empid age gender 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

# labeling character to numeric
empinfo$gender=factor(empinfo$gender,labels=c("male","female"))
empinfo$gender

## [1] male   female male   female female female male   male   female male
## [11] male   female female male    male
## Levels: male female

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

## [1] staff   staff   faculty faculty staff   staff   staff   faculty
faculty
## [10] staff   faculty staff   faculty staff   faculty
## Levels: staff faculty

empinfo

##      empid age gender  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

```

```
# Extract male data
```

```
male=subset(empinfo,empinfo$gender=="male")
```

```
male
```

```
##      empid age gender  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
```

```
# Extract female data
```

```
female=subset(empinfo, empinfo$gender=='female')
```

```
female
```

```
##      empid age gender  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
```

```
# summary statistics for empinfo data
```

```
summary(empinfo)
```

```
##      empid          age          gender          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 of male,female and age
```

```
summary(male)
```

```
##      empid          age          gender          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(female)
```

```
##      empid      age      gender      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(age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      30.00  32.00   37.00   40.87  47.50   60.00
```

```
# creating table (one-way)
```

```
table1=table(empinfo$gender)
table1
```

```
##
##      male female
##         8       7
```

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

```
##
##      staff faculty
##         8         7
```

```
# creating table (two-way)
```

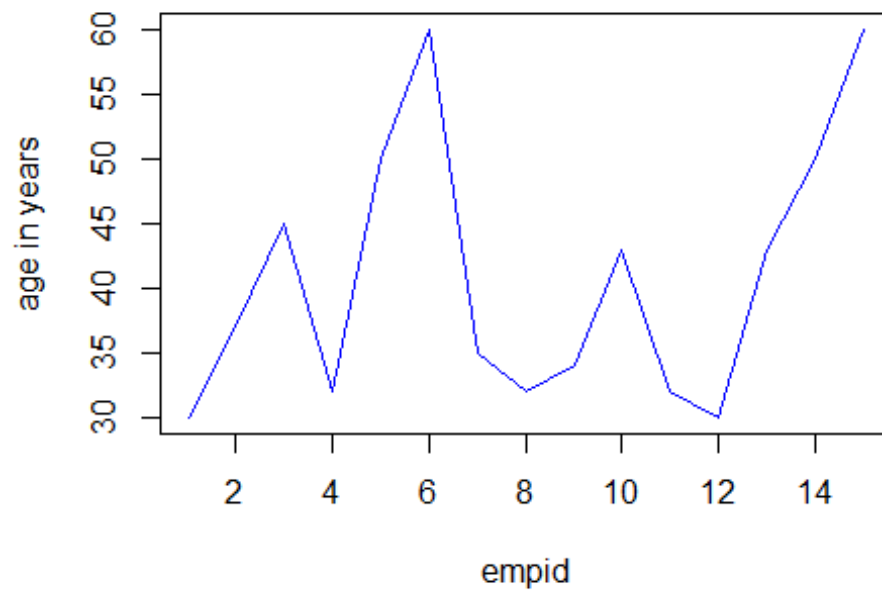
```
table3=table(empinfo$gender, empinfo$status)
table3
```

```
##
##           staff faculty
##      male         4         4
##      female        4         3
```

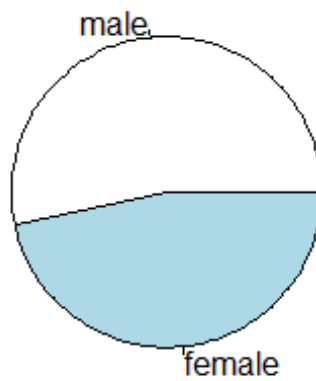
```
# Graphical representation (scatterplot)
```

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

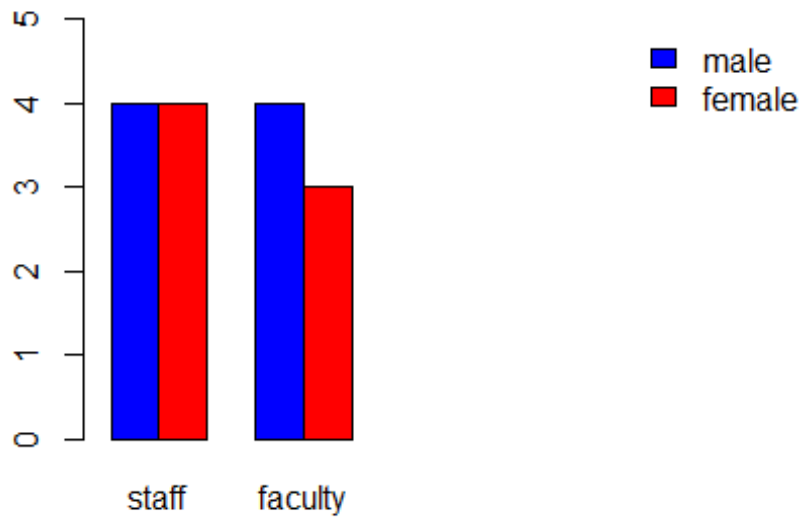
### Age of employees



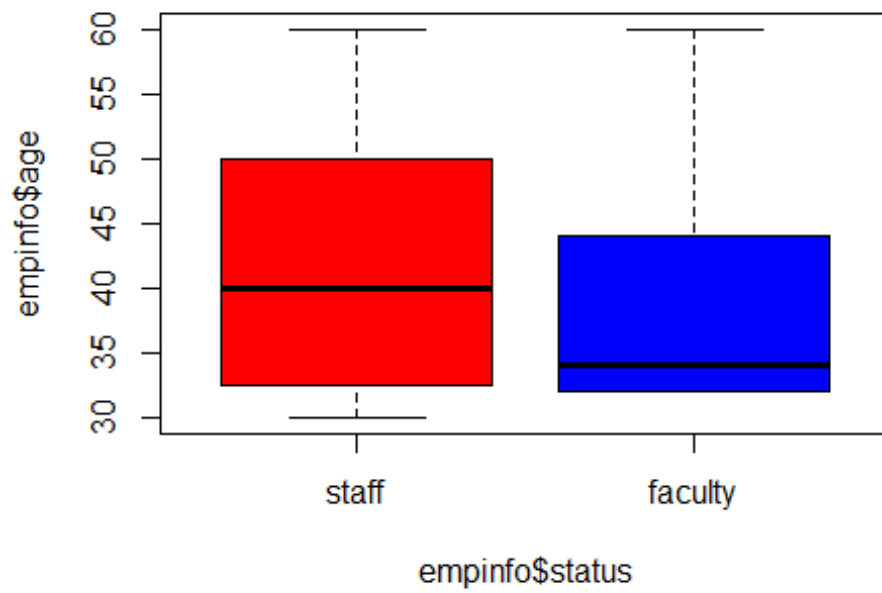
```
# Graphical representation (Pie chart)  
pie(table1)
```



```
# Graphical representation (Bar plot)
barplot(table3,beside=T,xlim=c(1,15),ylim=c(0,5),col=c("blue", "red"))
legend("topright",legend=rownames(table3),fill=c('blue','red'),bty="n")
```



```
# Graphical representation (Box plot)
boxplot(empinfo$age~empinfo$status,col=c('red','blue'))
```



### Conclusion:

Different alignment of data set and various graphical representations in R have been explored and executed.