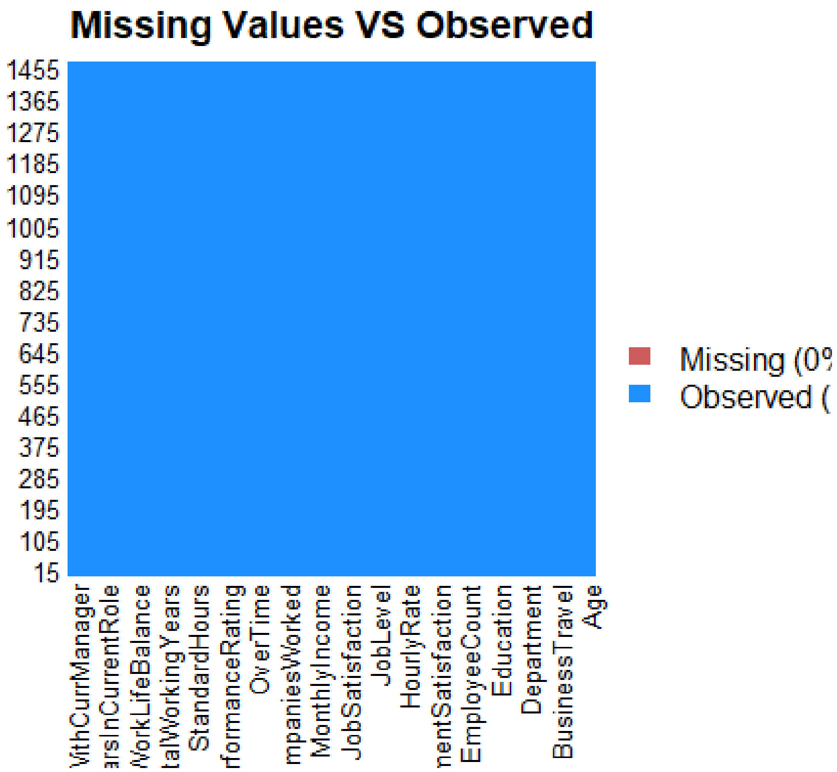
```
> ##Importing Data and inital analyses
> #Importing csv file from a location
> attr<- read.csv(file="MVA/Attrition Dataset.csv", header=TRUE, sep=",")
> attr <- as.data.frame(attr)
> glimpse(attr)

Observations: 1,470
Variables: 35
$ Age                      <int> 41, 49, 37, 33, 27, 32, 59, 30, 38, 36, 35,
29, 31, 34, 28, 29, 32, 22, 5...
$ Attrition                <fct> Yes, No, Yes, No, No, No, No, No, No, No, No
, No, No, No, Yes, No, No, No...
$ BusinessTravel           <fct> Travel_Rarely, Travel_Frequently, Travel_Rar
ely, Travel_Frequently, Trave...
$ DailyRate                <int> 1102, 279, 1373, 1392, 591, 1005, 1324, 1358
, 216, 1299, 809, 153, 670, 1...
$ Department               <fct> Sales, Research & Development, Research & De
velopment, Research & Develop...
$ DistanceFromHome         <int> 1, 8, 2, 3, 2, 2, 3, 24, 23, 27, 16, 15, 26,
19, 24, 21, 5, 16, 2, 2, 11,...
$ Education                <int> 2, 1, 2, 4, 1, 2, 3, 1, 3, 3, 3, 2, 1, 2, 3,
4, 2, 2, 4, 3, 2, 4, 4, 2, 1...
$ EducationField           <fct> Life Sciences, Life Sciences, Other, Life Sc
iences, Medical, Life Science...
$ EmployeeCount            <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
$ EmployeeNumber           <int> 1, 2, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16
, 18, 19, 20, 21, 22, 23, 24,...
$ EnvironmentSatisfaction  <int> 2, 3, 4, 4, 1, 4, 3, 4, 4, 3, 1, 4, 1, 2, 3,
2, 1, 4, 1, 4, 1, 3, 1, 3, 2...
$ Gender                   <fct> Female, Male, Male, Female, Male, Male, Fema
le, Male, Male, Male, Male, F...
$ HourlyRate               <int> 94, 61, 92, 56, 40, 79, 81, 67, 44, 94, 84,
49, 31, 93, 50, 51, 80, 96, 7...
$ JobInvolvement           <int> 3, 2, 2, 3, 3, 3, 4, 3, 2, 3, 4, 2, 3, 3, 2,
4, 4, 4, 2, 3, 4, 2, 3, 3, 3...
$ JobLevel                 <int> 2, 2, 1, 1, 1, 1, 1, 1, 3, 2, 1, 2, 1, 1, 1,
3, 1, 1, 4, 1, 2, 1, 3, 1, 1...
$ JobRole                  <fct> Sales Executive, Research Scientist, Laborat
ory Technician, Research Scie...
$ JobSatisfaction          <int> 4, 2, 3, 3, 2, 4, 1, 3, 3, 3, 2, 3, 3, 4, 3,
1, 2, 4, 4, 4, 3, 1, 2, 4, 1...
$ MaritalStatus            <fct> Single, Married, Single, Married, Married, S
ingle, Married, Divorced, Sin...
$ MonthlyIncome            <int> 5993, 5130, 2090, 2909, 3468, 3068, 2670, 26
93, 9526, 5237, 2426, 4193, 2...
$ MonthlyRate              <int> 19479, 24907, 2396, 23159, 16632, 11864, 996
4, 13335, 8787, 16577, 16479,...
$ NumCompaniesWorked       <int> 8, 1, 6, 1, 9, 0, 4, 1, 0, 6, 0, 0, 1, 0, 5,
1, 0, 1, 2, 5, 0, 7, 0, 1, 2...
$ Over18                   <fct> Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y,
Y, Y, Y, Y, Y, Y, Y, Y, Y, Y...
$ OverTime                 <fct> Yes, No, Yes, Yes, No, No, Yes, No, No, No,
No, Yes, No, No, Yes, No, Yes...
$ PercentSalaryHike        <int> 11, 23, 15, 11, 12, 13, 20, 22, 21, 13, 13,
12, 17, 11, 14, 11, 12, 13, 1...
$ PerformanceRating        <int> 3, 4, 3, 3, 3, 3, 4, 4, 4, 3, 3, 3, 3, 3, 3,
3, 3, 3, 3, 3, 3, 4, 3, 3, 3...
$ RelationshipSatisfaction <int> 1, 4, 2, 3, 4, 3, 1, 2, 2, 2, 3, 4, 4, 3, 2,
3, 4, 2, 3, 3, 4, 2, 3, 4, 3...
$ StandardHours            <int> 80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 80,
80, 80, 80, 80, 80, 80, 80, 8...
$ StockOptionLevel         <int> 0, 1, 0, 0, 1, 0, 3, 1, 0, 2, 1, 0, 1, 1, 0,
1, 2, 2, 0, 0, 1, 0, 0, 0, 0...
```

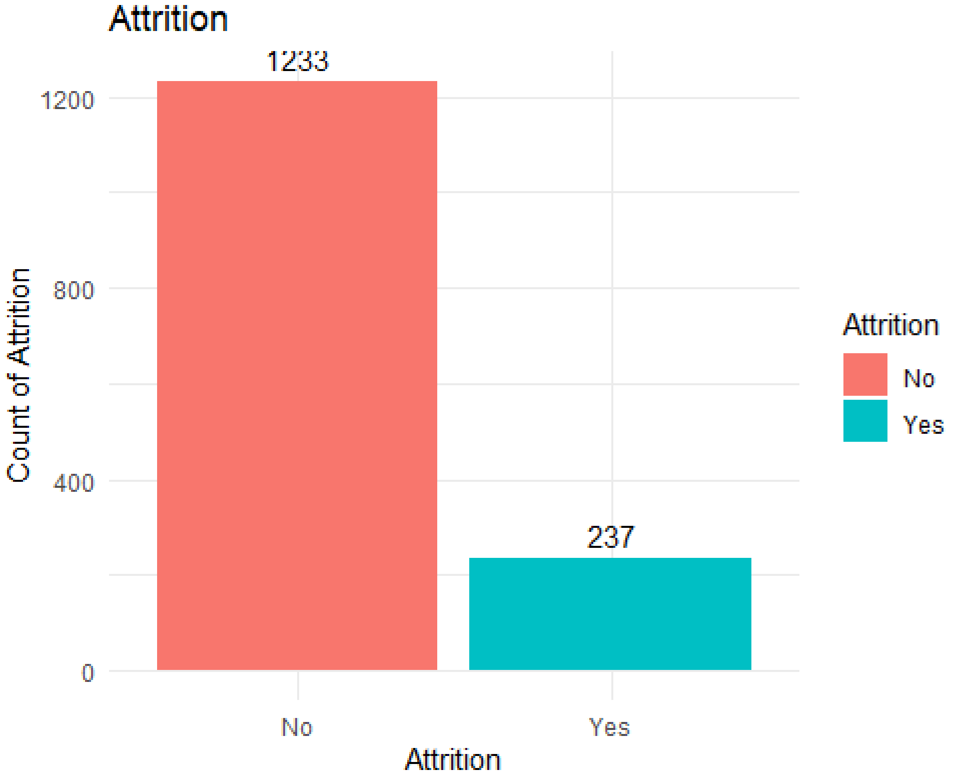
```
$ TotalWorkingYears        <int> 8, 10, 7, 8, 6, 8, 12, 1, 10, 17, 6, 10, 5, 
3, 6, 10, 7, 1, 31, 6, 5, 10,...
$ TrainingTimesLastYear    <int> 0, 3, 3, 3, 3, 2, 3, 2, 2, 3, 5, 3, 1, 2, 4, 
1, 5, 2, 3, 3, 5, 4, 4, 6, 2...
$ WorkLifeBalance          <int> 1, 3, 3, 3, 3, 2, 2, 3, 3, 2, 3, 3, 2, 3, 3, 
3, 2, 2, 3, 3, 2, 3, 3, 3, 3...
$ YearsAtCompany           <int> 6, 10, 0, 8, 2, 7, 1, 1, 9, 7, 5, 9, 5, 2, 4
, 10, 6, 1, 25, 3, 4, 5, 12, ...
$ YearsInCurrentRole       <int> 4, 7, 0, 7, 2, 7, 0, 0, 7, 7, 4, 5, 2, 2, 2, 
9, 2, 0, 8, 2, 2, 3, 6, 0, 2...
$ YearsSinceLastPromotion  <int> 0, 1, 0, 3, 2, 3, 0, 0, 1, 7, 0, 0, 4, 1, 0, 
8, 0, 0, 3, 1, 1, 0, 2, 0, 1...
$ YearsWithCurrManager     <int> 5, 7, 0, 0, 2, 6, 0, 0, 8, 7, 3, 8, 3, 2, 3, 
8, 5, 0, 7, 2, 3, 3, 11, 0, ...


> #Dimension of the dataset
> dim(attr)
[1] 1470   35

> #View the first 5 rows of the dataset
> head(attr)
  Age Attrition    BusinessTravel DailyRate             Department DistanceFr
omHome Education EducationField
1  41       Yes     Travel_Rarely      1102                  Sales
1         2  Life Sciences
2  49        No Travel_Frequently       279 Research & Development
8         1  Life Sciences
3  37       Yes     Travel_Rarely      1373 Research & Development
2         2          Other
4  33        No Travel_Frequently      1392 Research & Development
3         4  Life Sciences
5  27        No     Travel_Rarely       591 Research & Development
2         1        Medical
6  32        No Travel_Frequently      1005 Research & Development
2         2  Life Sciences
  EmployeeCount EmployeeNumber EnvironmentSatisfaction Gender HourlyRate JobI
nvolvement JobLevel
1             1              1                       2 Female         94
3        2
2             1              2                       3   Male         61
2        2
3             1              4                       4   Male         92
2        1
4             1              5                       4 Female         56
3        1
5             1              7                       1   Male         40
3        1
6             1              8                       4   Male         79
3        1
                JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRa
te NumCompaniesWorked Over18
1       Sales Executive               4        Single          5993       194
79                  8      Y
2    Research Scientist               2       Married          5130       249
07                  1      Y
3 Laboratory Technician               3        Single          2090        23
96                  6      Y
4    Research Scientist               3       Married          2909       231
59                  1      Y
5 Laboratory Technician               2       Married          3468       166
32                  9      Y
6 Laboratory Technician               4        Single          3068       118
64                  0      Y
```

```
  OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction Stand
ardHours StockOptionLevel
1      Yes                11                 3                        1
80                0
2       No                23                 4                        4
80                1
3      Yes                15                 3                        2
80                0
4      Yes                11                 3                        3
80                0
5       No                12                 3                        4
80                1
6       No                13                 3                        3
80                0
  TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany Year
sInCurrentRole
1                 8                     0               1              6
4
2                10                     3               3             10
7
3                 7                     3               3              0
0
4                 8                     3               3              8
7
5                 6                     3               3              2
2
6                 8                     2               2              7
7
  YearsSinceLastPromotion YearsWithCurrManager
1                       0                    5
2                       1                    7
3                       0                    0
4                       3                    0
5                       2                    2
6                       3                    6
> summary(attr)
      Age        Attrition            BusinessTravel   DailyRate
Department
 Min.   :18.00   No :1233   Non-Travel       : 150   Min.   : 102.0   Human R
esources       : 63
 1st Qu.:30.00   Yes: 237   Travel_Frequently: 277   1st Qu.: 465.0   Researc
h & Development:961
 Median :36.00              Travel_Rarely    :1043   Median : 802.0   Sales
:446
 Mean   :36.92                                       Mean   : 802.5
 3rd Qu.:43.00                                       3rd Qu.:1157.0
 Max.   :60.00                                       Max.   :1499.0

 DistanceFromHome   Education             EducationField EmployeeCount Emplo
yeeNumber
 Min.   : 1.000   Min.   :1.000   Human Resources : 27    Min.   :1     Min.
:   1.0
 1st Qu.: 2.000   1st Qu.:2.000   Life Sciences   :606    1st Qu.:1     1st Q
u.: 491.2
 Median : 7.000   Median :3.000   Marketing       :159    Median :1     Media
n :1020.5
 Mean   : 9.193   Mean   :2.913   Medical         :464    Mean   :1     Mean
:1024.9
 3rd Qu.:14.000   3rd Qu.:4.000   Other           : 82    3rd Qu.:1     3rd Q
u.:1555.8
 Max.   :29.000   Max.   :5.000   Technical Degree:132    Max.   :1     Max.
:2068.0
```

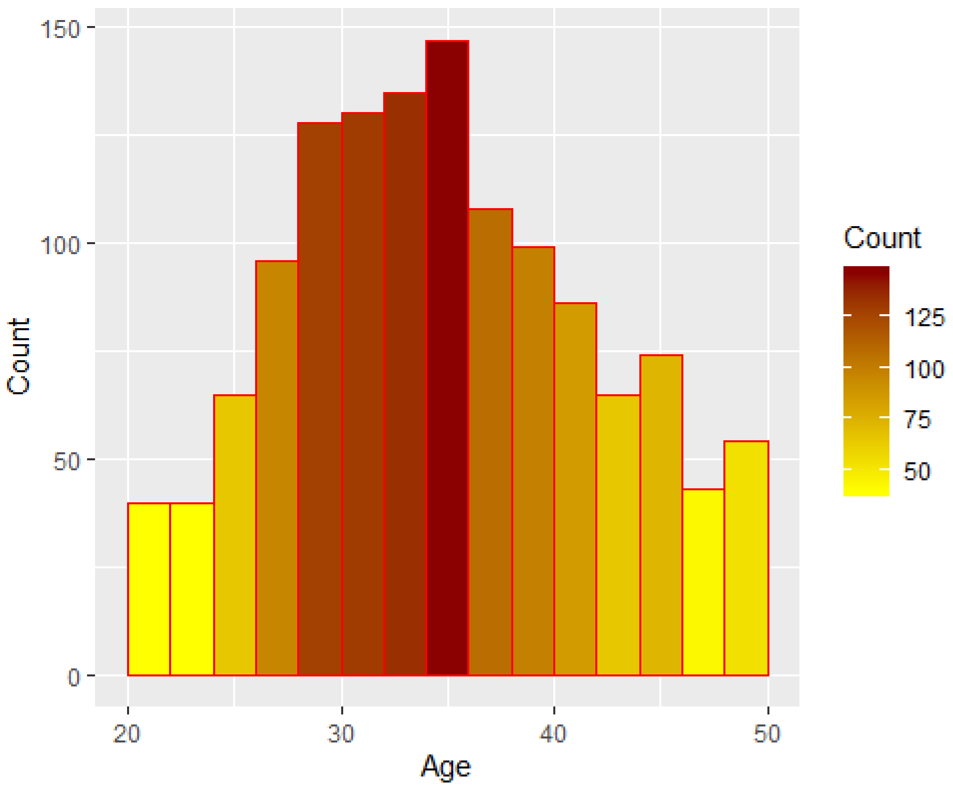
```
 EnvironmentSatisfaction    Gender      HourlyRate     JobInvolvement    JobL
evel
 Min.   :1.000           Female:588   Min.   : 30.00   Min.   :1.00   Min.   
:1.000  
 1st Qu.:2.000           Male  :882   1st Qu.: 48.00   1st Qu.:2.00   1st Qu.
:1.000  
 Median :3.000                        Median : 66.00   Median :3.00   Median 
:2.000  
 Mean   :2.722                        Mean   : 65.89   Mean   :2.73   Mean   
:2.064  
 3rd Qu.:4.000                        3rd Qu.: 83.75   3rd Qu.:3.00   3rd Qu.
:3.000  
 Max.   :4.000                        Max.   :100.00   Max.   :4.00   Max.   
:5.000  

                      JobRole    JobSatisfaction  MaritalStatus MonthlyIncome
MonthlyRate   
 Sales Executive          :326   Min.   :1.000   Divorced:327   Min.   : 1009
Min.   : 2094  
 Research Scientist       :292   1st Qu.:2.000   Married :673   1st Qu.: 2911
1st Qu.: 8047  
 Laboratory Technician    :259   Median :3.000   Single  :470   Median : 4919
Median :14236  
 Manufacturing Director   :145   Mean   :2.729                  Mean   : 6503
Mean   :14313  
 Healthcare Representative:131   3rd Qu.:4.000                  3rd Qu.: 8379
3rd Qu.:20462  
 Manager                  :102   Max.   :4.000                  Max.   :19999
Max.   :26999  
 (Other)                  :215                                               
 NumCompaniesWorked Over18   OverTime   PercentSalaryHike PerformanceRating R
elationshipSatisfaction
 Min.   :0.000      Y:1470   No :1054   Min.   :11.00     Min.   :3.000     M
in.   :1.000  
 1st Qu.:1.000               Yes: 416   1st Qu.:12.00     1st Qu.:3.000     1
st Qu.:2.000  
 Median :2.000                          Median :14.00     Median :3.000     M
edian :3.000  
 Mean   :2.693                          Mean   :15.21     Mean   :3.154     M
ean   :2.712  
 3rd Qu.:4.000                          3rd Qu.:18.00     3rd Qu.:3.000     3
rd Qu.:4.000  
 Max.   :9.000                          Max.   :25.00     Max.   :4.000     M
ax.   :4.000  

 StandardHours StockOptionLevel TotalWorkingYears TrainingTimesLastYear WorkL
ifeBalance YearsAtCompany  
 Min.   :80    Min.   :0.0000   Min.   : 0.00     Min.   :0.000         Min.   
:1.000   Min.   : 0.000  
 1st Qu.:80    1st Qu.:0.0000   1st Qu.: 6.00     1st Qu.:2.000         1st Q
u.:2.000   1st Qu.: 3.000  
 Median :80    Median :1.0000   Median :10.00     Median :3.000         Media
n :3.000   Median : 5.000  
 Mean   :80    Mean   :0.7939   Mean   :11.28     Mean   :2.799         Mean   
:2.761   Mean   : 7.008  
 3rd Qu.:80    3rd Qu.:1.0000   3rd Qu.:15.00     3rd Qu.:3.000         3rd Q
u.:3.000   3rd Qu.: 9.000  
 Max.   :80    Max.   :3.0000   Max.   :40.00     Max.   :6.000         Max.   
:4.000   Max.   :40.000  

 YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager
 Min.   : 0.000     Min.   : 0.000          Min.   : 0.000      
 1st Qu.: 2.000     1st Qu.: 0.000          1st Qu.: 2.000      
 Median : 3.000     Median : 1.000          Median : 3.000      
 Mean   : 4.229     Mean   : 2.188          Mean   : 4.123
 3rd Qu.: 7.000     3rd Qu.: 3.000          3rd Qu.: 7.000
 Max.   :18.000     Max.   :15.000          Max.   :17.000


> #Rename the Age column
> colnames(attr)[1] <- "Age"
> #Calculating the number of null values in each of the columns
> colSums(sapply(attr,is.na))
                      Age                Attrition           BusinessTravel
DailyRate
                        0                        0                        0
0
               Department         DistanceFromHome                Education
EducationField
                        0                        0                        0
0
            EmployeeCount           EmployeeNumber  EnvironmentSatisfaction
Gender
                        0                        0                        0
0
               HourlyRate           JobInvolvement                 JobLevel
JobRole
                        0                        0                        0
0
          JobSatisfaction            MaritalStatus            MonthlyIncome
MonthlyRate
                        0                        0                        0
0
       NumCompaniesWorked                   Over18                 OverTime
PercentSalaryHike
                        0                        0                        0
0
        PerformanceRating RelationshipSatisfaction            StandardHours
StockOptionLevel
                        0                        0                        0
0
        TotalWorkingYears    TrainingTimesLastYear          WorkLifeBalance
YearsAtCompany
                        0                        0                        0
0
       YearsInCurrentRole  YearsSinceLastPromotion     YearsWithCurrManager
                        0                        0                        0
> missmap(attr,main="Missing Values VS Observed")
```



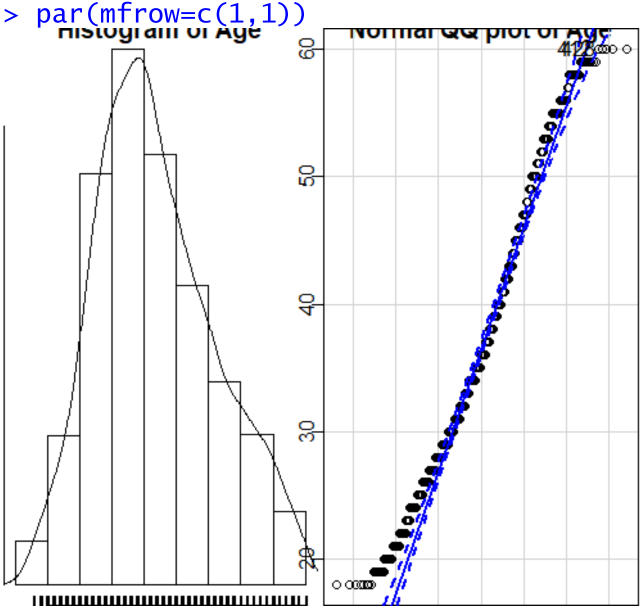
```
> #Removing redundant columns
> attr$EmployeeNumber<- NULL
> attr$StandardHours <- NULL
> attr$Over18 <- NULL
> attr$EmployeeCount <- NULL
> #Converting data type of categorical column
> attr$Education <- factor(attr$Education)
> attr$EnvironmentSatisfaction <- factor(attr$EnvironmentSatisfaction)
> attr$JobInvolvement <- factor(attr$JobInvolvement)
> attr$JobLevel <- factor(attr$JobLevel)
> attr$JobSatisfaction <- factor(attr$JobSatisfaction)
> attr$PerformanceRating <- factor(attr$PerformanceRating)
> attr$RelationshipSatisfaction <- factor(attr$RelationshipSatisfaction)
> attr$StockOptionLevel <- factor(attr$StockOptionLevel)
> attr$WorkLifeBalance <- factor(attr$WorkLifeBalance)
> #Assigning categorical and numerical variable to temporary variable
> catvar<-c('BusinessTravel','Department','Education','EducationField','Envir
onmentSatisfaction','Gender',
+            'JobRole','JobInvolvement','JobLevel','JobSatisfaction',
+            'MaritalStatus','PerformanceRating','RelationshipSatisfaction','S
tockOptionLevel','WorkLifeBalance')
> numvar<-c('Age','DailyRate','DistanceFromHome','HourlyRate',
+            'MonthlyIncome','MonthlyRate','NumCompaniesWorked','PercentSalary
Hike','TotalWorkingYears',
+            'TrainingTimesLastYear','YearsAtCompany',
+            'YearsInCurrentRole','YearsSinceLastPromotion','YearsWithCurrMana
ger')


> ##Exploratory Data Analysis
> 
> #Vizualization of Attrition
> attr %>%
+   group_by(Attrition) %>%
+   tally() %>%
+   ggplot(aes(x =Attrition,y = n,fill=Attrition)) +
+   geom_bar(stat = "identity") +
+   theme_minimal()+
+   labs(x="Attrition", y="Count of Attrition")+
+   ggtitle("Attrition")+
+   geom_text(aes(label = n), vjust = -0.5, position = position_dodge(0.9))
```



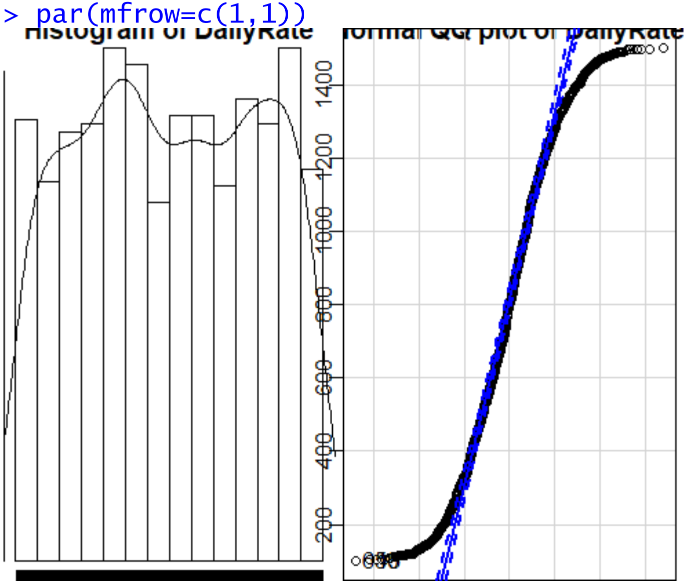
```
#Influence of features on Attrition
> ggplot(data=attr, aes(attr$Age)) +
+   geom_histogram(breaks=seq(20, 50, by=2),
+                  col="red",
+                  aes(fill=..count..))+
+   labs(x="Age", y="Count")+
+   scale_fill_gradient("Count", low="yellow", high="dark red")
```



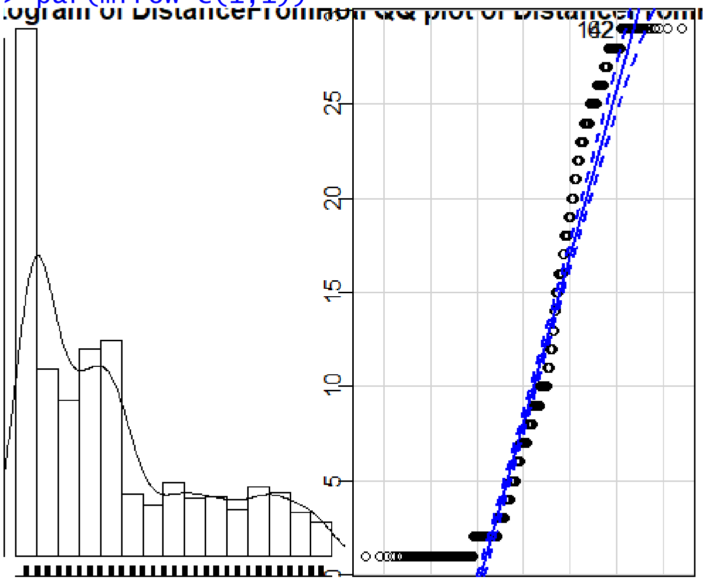
```
> #Checking for distributions in numerical columns
> #The qqPlot show a few extreme outliers which break the assumption of 95% c
onfidence
> #normal distribution
> par(mfrow = c(1,2))
> hist(attr$Age,xlab='',main = 'Histogram of Age',freq = FALSE)
> lines(density(attr$Age,na.rm = T))
> rug(jitter(attr$Age))
> qqPlot(attr$Age,main='Normal QQ plot of Age')
[1] 412 428
> par(mfrow=c(1,1))
```



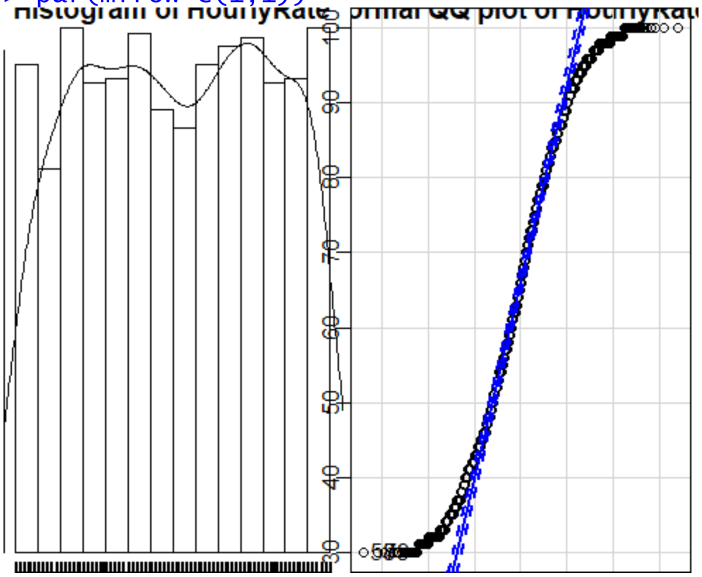
```
> par(mfrow = c(1,2))
> hist(attr$DailyRate,xlab='',main = 'Histogram of DailyRate',freq = FALSE)
> lines(density(attr$DailyRate,na.rm = T))
> rug(jitter(attr$DailyRate))
> qqPlot(attr$DailyRate,main='Normal QQ plot of DailyRate')
[1] 650  15
> par(mfrow=c(1,1))
```



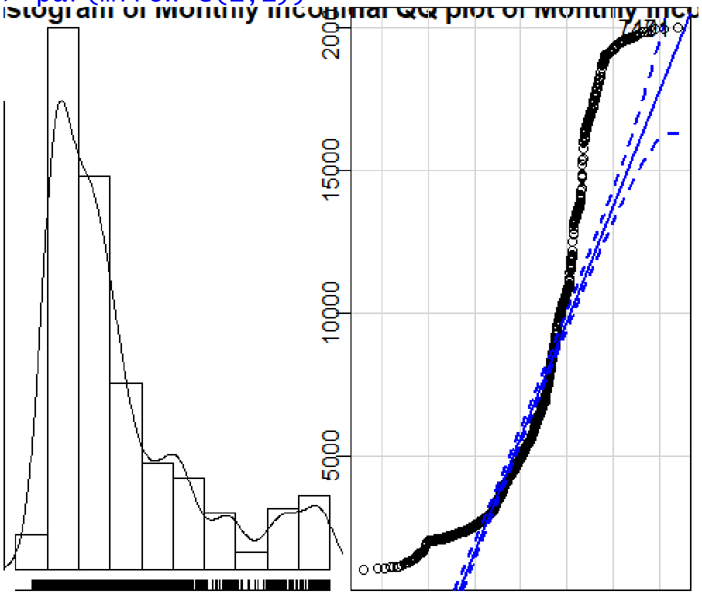
```
> par(mfrow = c(1,2))
> hist(attr$DistanceFromHome,xlab='',main = 'Histogram of DistanceFromHome',f
req = FALSE)
> lines(density(attr$DistanceFromHome,na.rm = T))
> rug(jitter(attr$DistanceFromHome))
> qqPlot(attr$DistanceFromHome,main='Normal QQ plot of DistanceFromHome')
[1]  62 142
> par(mfrow=c(1,1))
```



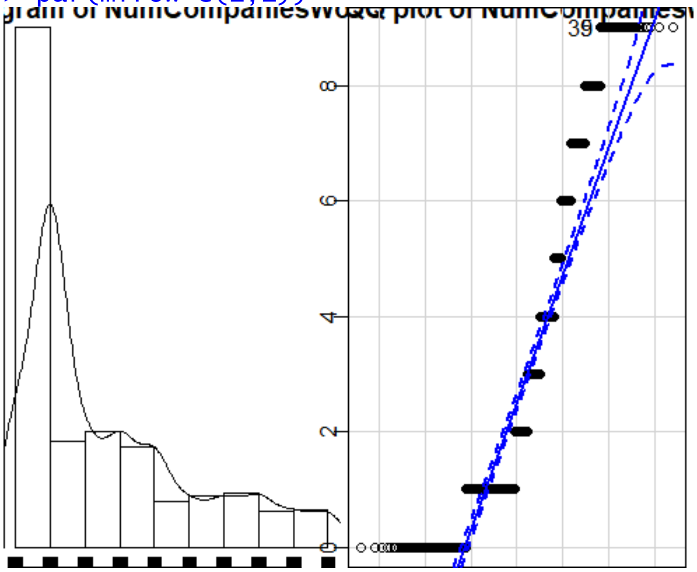
```
> par(mfrow = c(1,2))
> hist(attr$HourlyRate,xlab='',main = 'Histogram of HourlyRate',freq = FALSE)
> lines(density(attr$HourlyRate,na.rm = T))
> rug(jitter(attr$HourlyRate))
> qqPlot(attr$HourlyRate,main='Normal QQ plot of HourlyRate')
[1] 58 79
> par(mfrow=c(1,1))
```



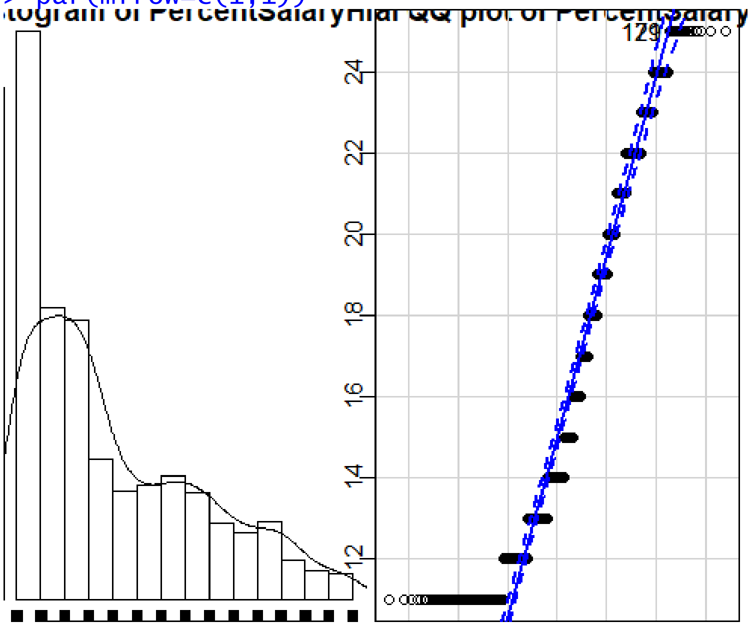
```
> par(mfrow = c(1,2))
> hist(attr$MonthlyIncome,xlab='',main = 'Histogram of Monthly Income',freq =
FALSE)
> lines(density(attr$MonthlyIncome,na.rm = T))
> rug(jitter(attr$MonthlyIncome))
> qqPlot(attr$MonthlyIncome,main='Normal QQ plot of Monthly Income')
[1] 191 747
> par(mfrow=c(1,1))
```



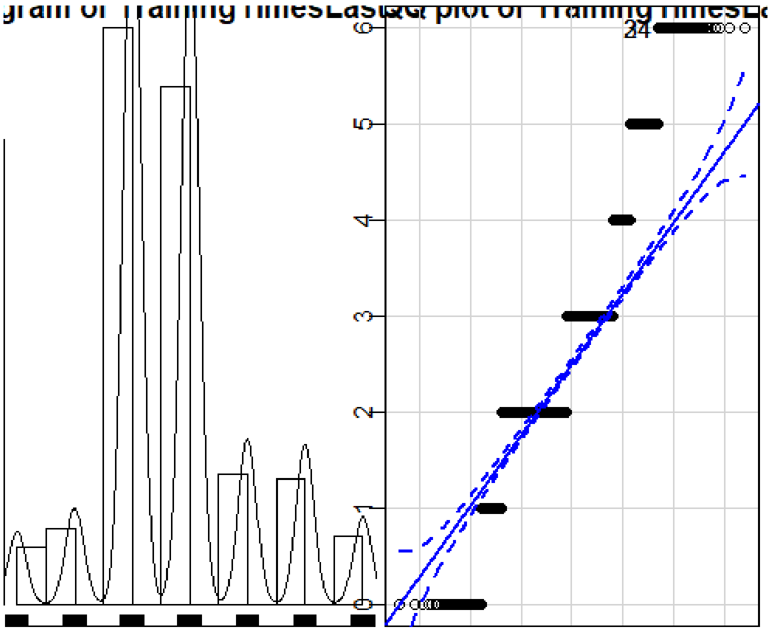
```
> par(mfrow = c(1,2))
> hist(attr$NumCompaniesWorked,xlab='',main = 'Histogram of NumCompaniesWorke
d',freq = FALSE)
> lines(density(attr$NumCompaniesWorked,na.rm = T))
> rug(jitter(attr$NumCompaniesWorked))
> qqPlot(attr$NumCompaniesWorked,main='Normal QQ plot of NumCompaniesWorked')
[1]  5 39
> par(mfrow=c(1,1))
```



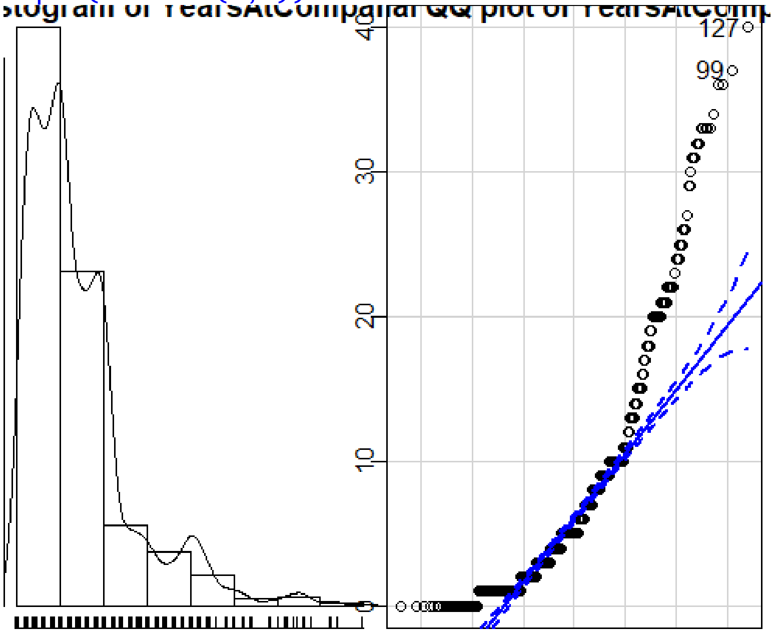
```
> par(mfrow = c(1,2))
> hist(attr$PercentSalaryHike,xlab='',main = 'Histogram of PercentSalaryHike'
,freq = FALSE)
> lines(density(attr$PercentSalaryHike,na.rm = T))
> rug(jitter(attr$PercentSalaryHike))
> qqPlot(attr$PercentSalaryHike,main='Normal QQ plot of PercentSalaryHike')
[1] 121 179
> par(mfrow=c(1,1))
```



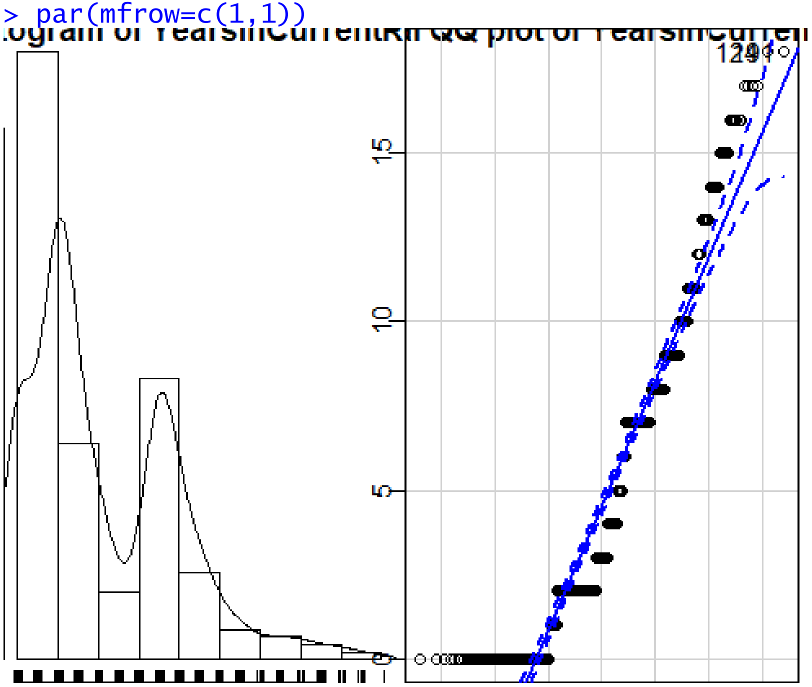
```
> par(mfrow = c(1,2))
> hist(attr$TrainingTimesLastYear,xlab='',main = 'Histogram of TrainingTimesL
astYear',freq = FALSE)
> lines(density(attr$TrainingTimesLastYear,na.rm = T))
> rug(jitter(attr$TrainingTimesLastYear))
> qqPlot(attr$TrainingTimesLastYear,main='Normal QQ plot of TrainingTimesLast
Year')
[1] 24 34
> par(mfrow=c(1,1))
```



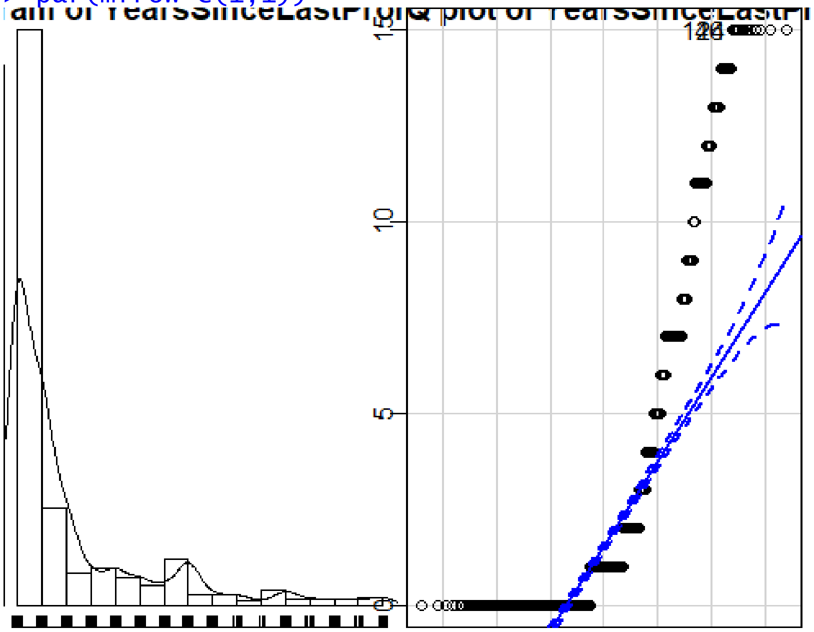
```
> par(mfrow = c(1,2))
> hist(attr$YearsAtCompany,xlab='',main = 'Histogram of YearsAtCompany',freq
= FALSE)
> lines(density(attr$YearsAtCompany,na.rm = T))
> rug(jitter(attr$YearsAtCompany))
> qqPlot(attr$YearsAtCompany,main='Normal QQ plot of YearsAtCompany')
[1] 127  99
> par(mfrow=c(1,1))
```



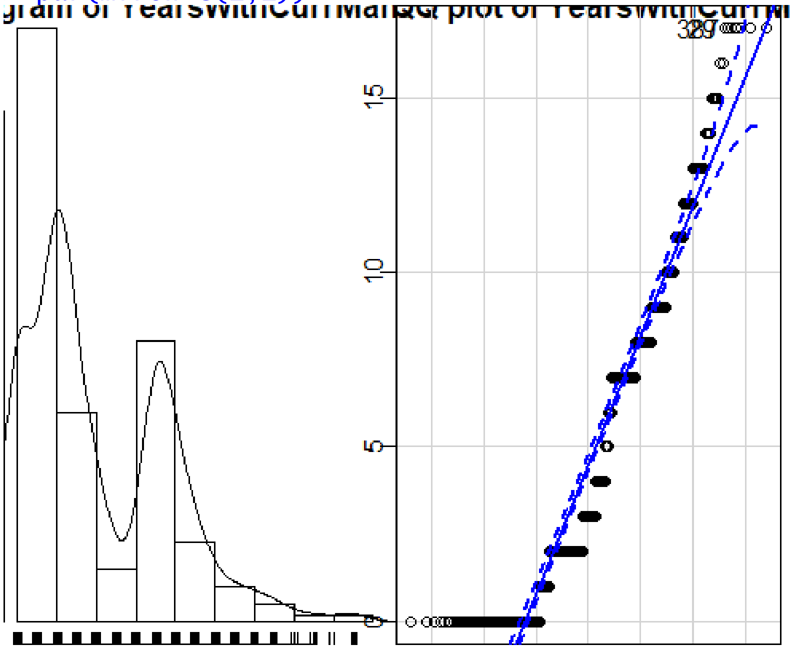
```
> par(mfrow = c(1,2))
> hist(attr$YearsInCurrentRole,xlab='',main = 'Histogram of YearsInCurrentRol
e',freq = FALSE)
> lines(density(attr$YearsInCurrentRole,na.rm = T))
> rug(jitter(attr$YearsInCurrentRole))
> qqPlot(attr$YearsInCurrentRole,main='Normal QQ plot of YearsInCurrentRole')
[1] 124 191
> par(mfrow=c(1,1))
```



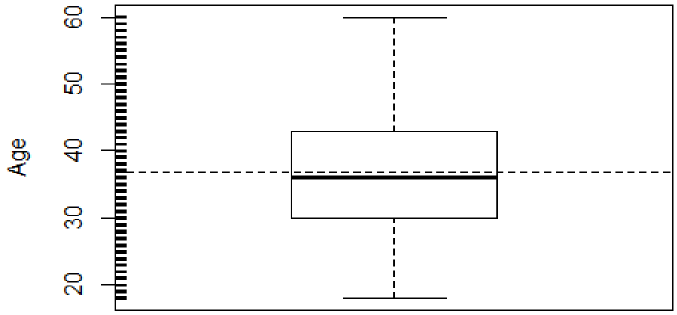
```
> par(mfrow = c(1,2))
> hist(attr$YearsSinceLastPromotion,xlab='',main = 'Histogram of YearsSinceLa
stPromotion',freq = FALSE)
> lines(density(attr$YearsSinceLastPromotion,na.rm = T))
> rug(jitter(attr$YearsSinceLastPromotion))
> qqPlot(attr$YearsSinceLastPromotion,main='Normal QQ plot of YearsSinceLastP
romotion')
[1]  46 124
> par(mfrow=c(1,1))
```



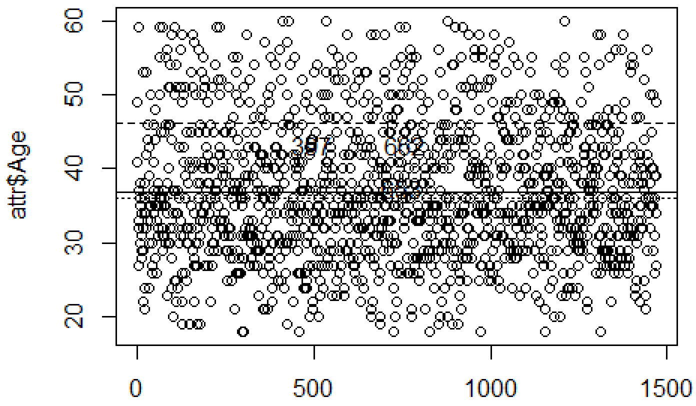
```
> par(mfrow = c(1,2))
> hist(attr$YearsWithCurrManager,xlab='',main = 'Histogram of YearsWithCurrMa
nager',freq = FALSE)
> lines(density(attr$YearsWithCurrManager,na.rm = T))
> rug(jitter(attr$YearsWithCurrManager))
> qqPlot(attr$YearsWithCurrManager,main='Normal QQ plot of YearsWithCurrManag
er')
[1]  29 387
> par(mfrow=c(1,1))
```



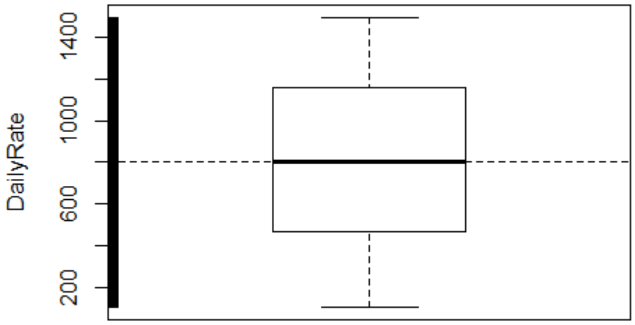
```
> #Boxplot distributions for our numeric columns
> #The dashed line shows the mean and the dark center line shows the median
> #Difference between these two lines depict the deviation from the central l
imit theorem
> #Boxplot distributions for  Age
> boxplot(attr$Age, ylab = "Age")
> rug(jitter(attr$Age), side = 2)
> abline(h = mean(attr$Age, na.rm = T), lty = 2)
```



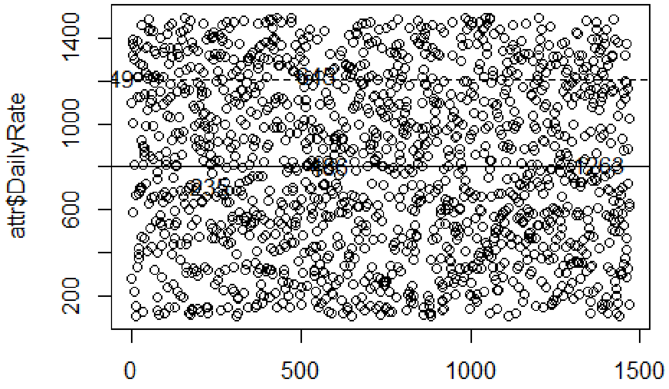
```
> #Plotting the Age  with 3 lines for mean, median and mean+std
> plot(attr$Age, xlab = "")
> abline(h = mean(attr$Age, na.rm = T), lty = 1)
> abline(h = mean(attr$Age, na.rm = T) + sd(attr$Age, na.rm = T),lty = 2)
> abline(h = median(attr$Age, na.rm = T), lty = 3)
> identify(attr$Age)
[1]  286  696  709  720 1174 1323
```



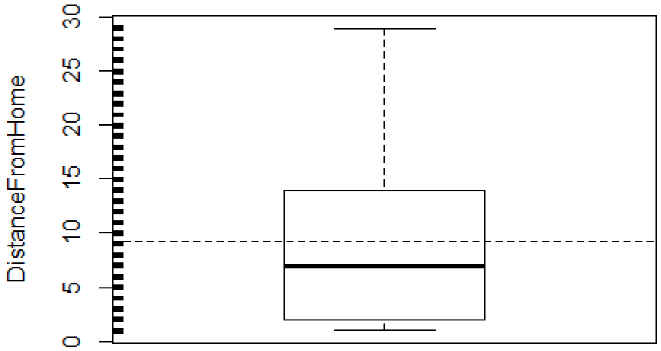
```
> #Boxplot distributions for Daily rate
> boxplot(attr$DailyRate, ylab = "DailyRate",outline = TRUE)
> rug(jitter(attr$DailyRate), side = 2)
> abline(h = mean(attr$DailyRate, na.rm = T), lty = 2)
```



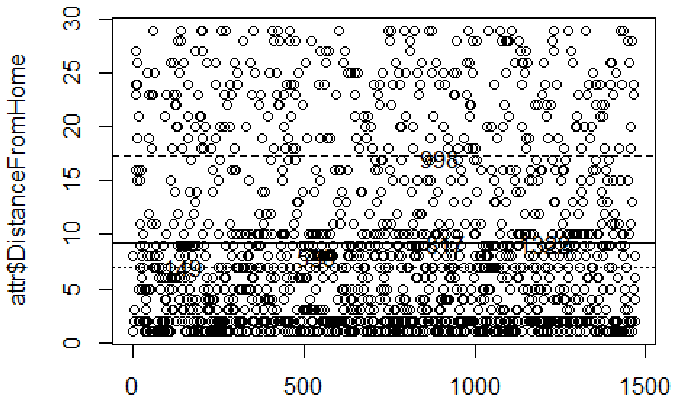
```
> #Plotting the DailyRate  with 3 lines for mean, median and mean+std
> plot(attr$DailyRate, xlab = "")
> abline(h = mean(attr$DailyRate, na.rm = T), lty = 1)
> abline(h = mean(attr$DailyRate, na.rm = T) + sd(attr$DailyRate, na.rm = T),
lty = 2)
> abline(h = median(attr$DailyRate, na.rm = T), lty = 3)
> identify(attr$DailyRate)
[1]   49  235  486  645 1263
```



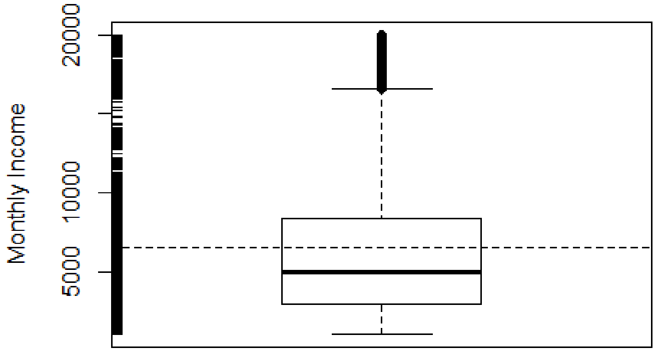
```
> #Boxplot distributions for Distance from home
> boxplot(attr$DistanceFromHome, ylab = "DistanceFromHome",outline = TRUE)
> rug(jitter(attr$DistanceFromHome), side = 2)
> abline(h = mean(attr$DistanceFromHome, na.rm = T), lty = 2)
```



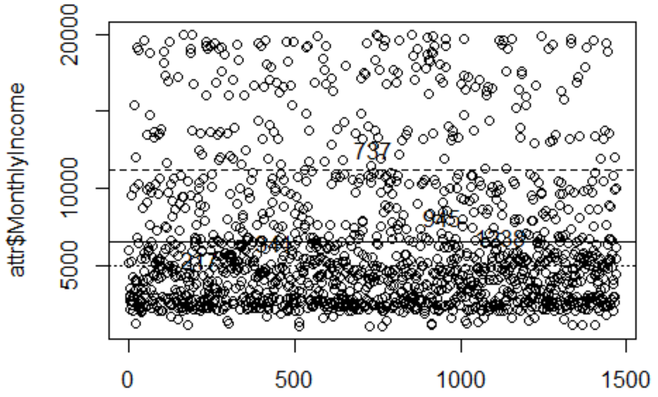
```
> plot(attr$DistanceFromHome, xlab = "")
> abline(h = mean(attr$DistanceFromHome, na.rm = T), lty = 1)
> abline(h = mean(attr$DistanceFromHome, na.rm = T) + sd(attr$DistanceFromHom
e, na.rm = T),lty = 2)
> abline(h = median(attr$DistanceFromHome, na.rm = T), lty = 3)
> identify(attr$DistanceFromHome)
[1]  149  538  817  998 1322
```



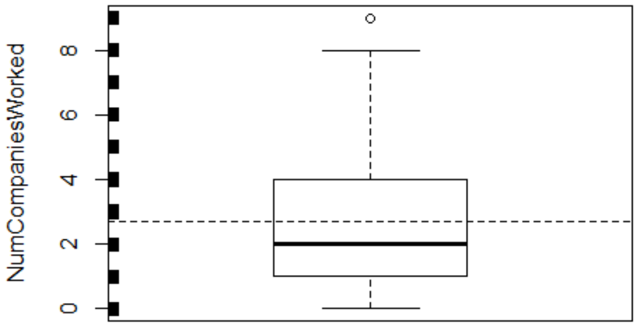
```
> #Boxplot distributions for Monthly Income
> boxplot(attr$MonthlyIncome, ylab = "Monthly Income")
> rug(jitter(attr$MonthlyIncome), side = 2)
> abline(h = mean(attr$MonthlyIncome, na.rm = T), lty = 2)
```



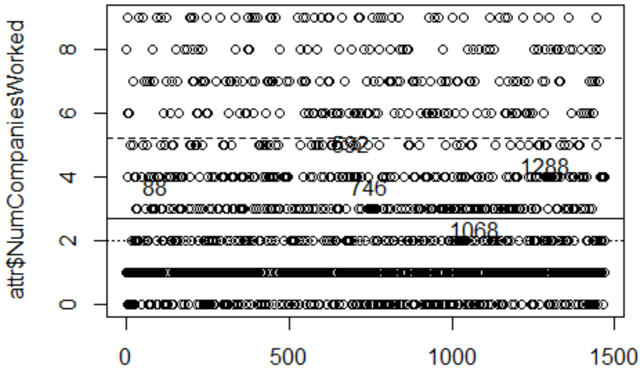
```
> #Plotting the Monthly Income and Age  with 3 lines for mean, median and mea
n+std
> plot(attr$MonthlyIncome, xlab = "")
> abline(h = mean(attr$MonthlyIncome, na.rm = T), lty = 1)
> abline(h = mean(attr$MonthlyIncome, na.rm = T) + sd(attr$MonthlyIncome, na.
rm = T),lty = 2)
> abline(h = median(attr$MonthlyIncome, na.rm = T), lty = 3)
> identify(attr$MonthlyIncome)
[1]  217  341  737  945 1238
```



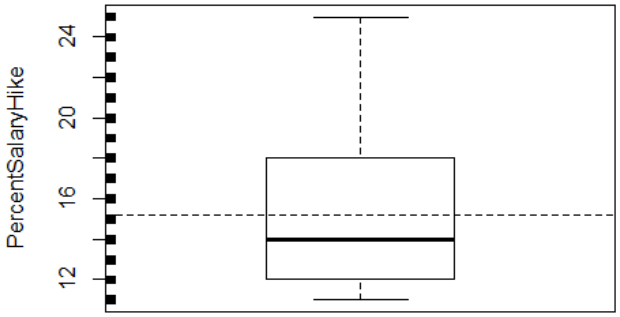
```
> #Boxplot distributions for  NumCompaniesWorked
> boxplot(attr$NumCompaniesWorked, ylab = "NumCompaniesWorked")
> rug(jitter(attr$NumCompaniesWorked), side = 2)
> abline(h = mean(attr$NumCompaniesWorked, na.rm = T), lty = 2)
```



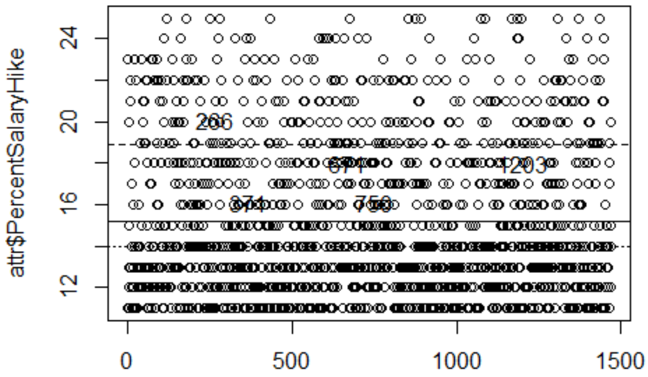
```
> #Plotting the NumCompaniesWorked  with 3 lines for mean, median and mean+st
d
> plot(attr$NumCompaniesWorked, xlab = "")
> abline(h = mean(attr$NumCompaniesWorked, na.rm = T), lty = 1)
> abline(h = mean(attr$NumCompaniesWorked, na.rm = T) + sd(attr$NumCompaniesW
orked, na.rm = T),lty = 2)
> abline(h = median(attr$NumCompaniesWorked, na.rm = T), lty = 3)
> identify(attr$NumCompaniesWorked)
[1]   88  592  746 1068 1288
```



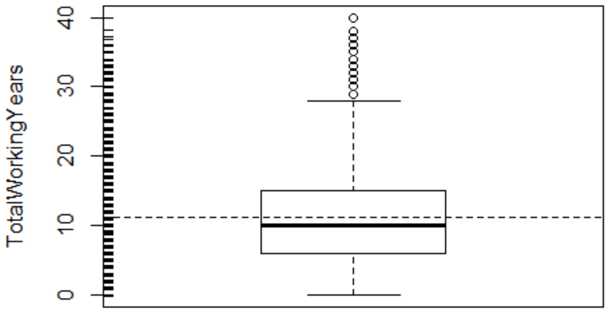
```
> #Boxplot distributions for  PercentSalaryHike
> boxplot(attr$PercentSalaryHike, ylab = "PercentSalaryHike")
> rug(jitter(attr$PercentSalaryHike), side = 2)
> abline(h = mean(attr$PercentSalaryHike, na.rm = T), lty = 2)
```



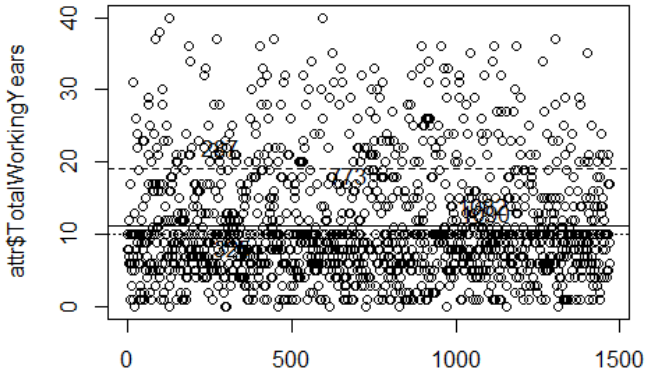
```
> #Plotting the PercentSalaryHike  with 3 lines for mean, median and mean+std
> plot(attr$PercentSalaryHike, xlab = "")
> abline(h = mean(attr$PercentSalaryHike, na.rm = T), lty = 1)
> abline(h = mean(attr$PercentSalaryHike, na.rm = T) + sd(attr$PercentSalaryH
ike, na.rm = T),lty = 2)
> abline(h = median(attr$PercentSalaryHike, na.rm = T), lty = 3)
> identify(attr$PercentSalaryHike)
[1]  266  371  671  750 1203
```



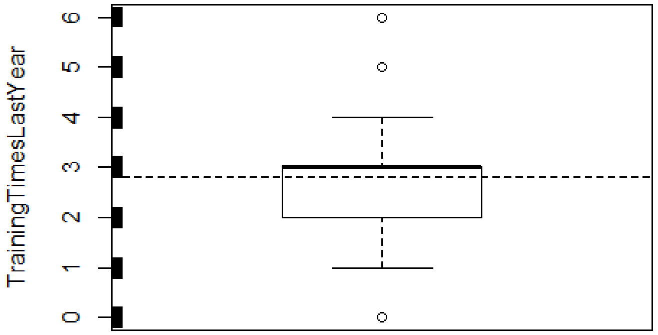
```
> #Boxplot distributions for  TotalWorkingYears
> boxplot(attr$TotalWorkingYears, ylab = "TotalWorkingYears")
> rug(jitter(attr$TotalWorkingYears), side = 2)
> abline(h = mean(attr$TotalWorkingYears, na.rm = T), lty = 2)
```



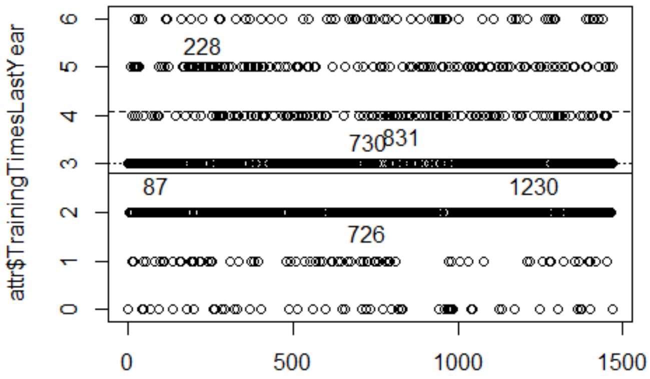
```
> #Plotting the TotalWorkingYears  with 3 lines for mean, median and mean+std
> plot(attr$TotalWorkingYears, xlab = "")
> abline(h = mean(attr$TotalWorkingYears, na.rm = T), lty = 1)
> abline(h = mean(attr$TotalWorkingYears, na.rm = T) + sd(attr$TotalWorkingYe
ars, na.rm = T),lty = 2)
> abline(h = median(attr$TotalWorkingYears, na.rm = T), lty = 3)
> identify(attr$TotalWorkingYears)
[1]  287  325  773 1082 1090
```



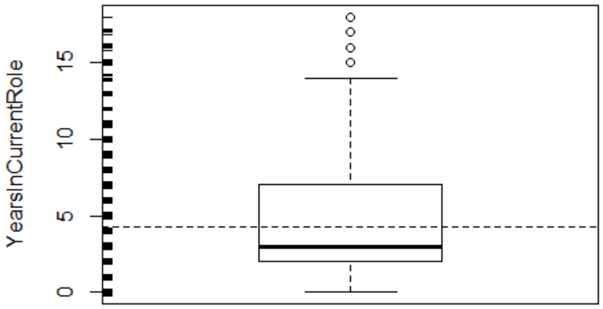
```
> #Boxplot distributions for  TrainingTimesLastYear
> boxplot(attr$TrainingTimesLastYear, ylab = "TrainingTimesLastYear")
> rug(jitter(attr$TrainingTimesLastYear), side = 2)
> abline(h = mean(attr$TrainingTimesLastYear, na.rm = T), lty = 2)
```



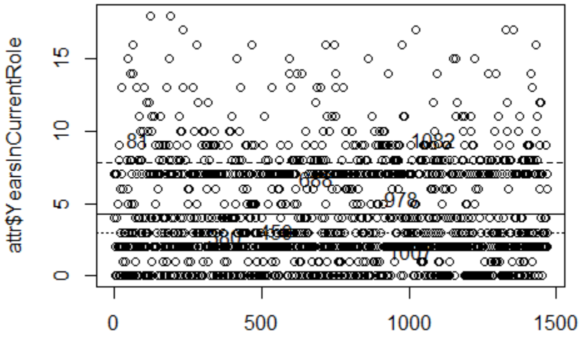
```
> #Plotting the TrainingTimesLastYear  with 3 lines for mean, median and mean
+std
> plot(attr$TrainingTimesLastYear, xlab = "")
> abline(h = mean(attr$TrainingTimesLastYear, na.rm = T), lty = 1)
> abline(h = mean(attr$TrainingTimesLastYear, na.rm = T) + sd(attr$TrainingTi
mesLastYear, na.rm = T),lty = 2)
> abline(h = median(attr$TrainingTimesLastYear, na.rm = T), lty = 3)
> identify(attr$TrainingTimesLastYear)
[1]   87  228  726  730  831 1230
```



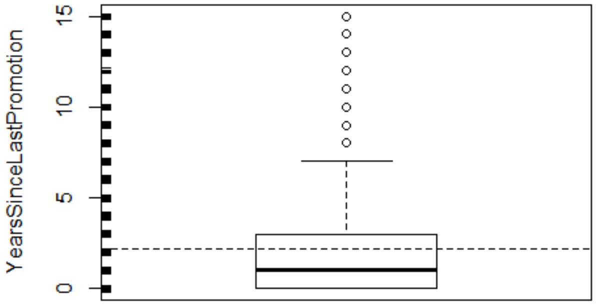
```
> #Boxplot distributions for  YearsInCurrentRole
> boxplot(attr$YearsInCurrentRole, ylab = "YearsInCurrentRole")
> rug(jitter(attr$YearsInCurrentRole), side = 2)
> abline(h = mean(attr$YearsInCurrentRole, na.rm = T), lty = 2)
```



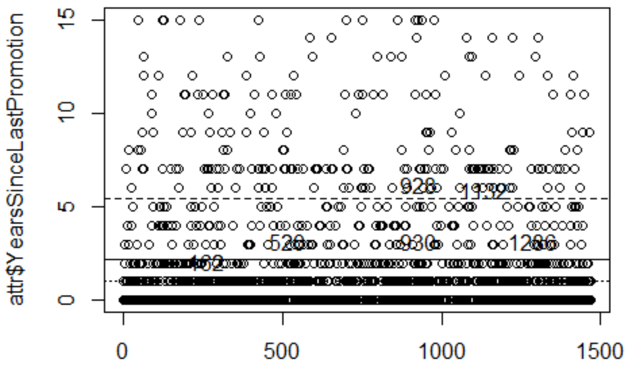
```
> #Plotting the YearsInCurrentRole  with 3 lines for mean, median and mean+st
d
> plot(attr$YearsInCurrentRole, xlab = "")
> abline(h = mean(attr$YearsInCurrentRole, na.rm = T), lty = 1)
> abline(h = mean(attr$YearsInCurrentRole, na.rm = T) + sd(attr$YearsInCurren
tRole, na.rm = T),lty = 2)
> abline(h = median(attr$YearsInCurrentRole, na.rm = T), lty = 3)
> identify(attr$YearsInCurrentRole)
[1]   81  380  450  688  978 1007 1082
```



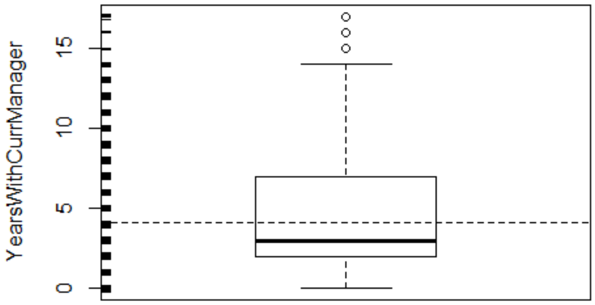
```
> #Boxplot distributions for  YearsSinceLastPromotion
> boxplot(attr$YearsSinceLastPromotion, ylab = "YearsSinceLastPromotion")
> rug(jitter(attr$YearsSinceLastPromotion), side = 2)
> abline(h = mean(attr$YearsSinceLastPromotion, na.rm = T), lty = 2)
```



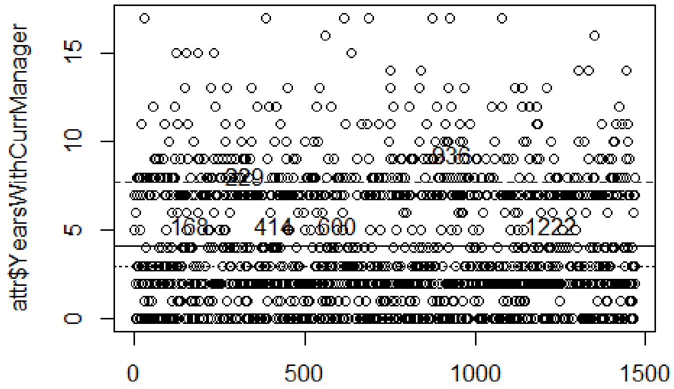
```
> #Plotting the YearsSinceLastPromotion  with 3 lines for mean, median and me
an+std
> plot(attr$YearsSinceLastPromotion, xlab = "")
> abline(h = mean(attr$YearsSinceLastPromotion, na.rm = T), lty = 1)
> abline(h = mean(attr$YearsSinceLastPromotion, na.rm = T) + sd(attr$YearsSin
ceLastPromotion, na.rm = T),lty = 2)
> abline(h = median(attr$YearsSinceLastPromotion, na.rm = T), lty = 3)
> identify(attr$YearsSinceLastPromotion)
[1]  162  520  928  930 1132 1286
```



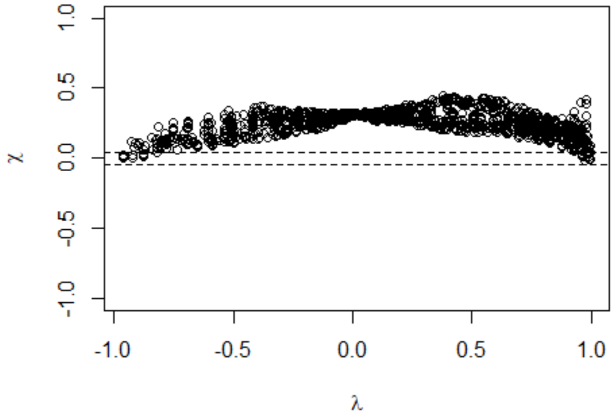
```
> #Boxplot distributions for  YearsWithCurrManager
> boxplot(attr$YearsWithCurrManager, ylab = "YearsWithCurrManager")
> rug(jitter(attr$YearsWithCurrManager), side = 2)
> abline(h = mean(attr$YearsWithCurrManager, na.rm = T), lty = 2)
```



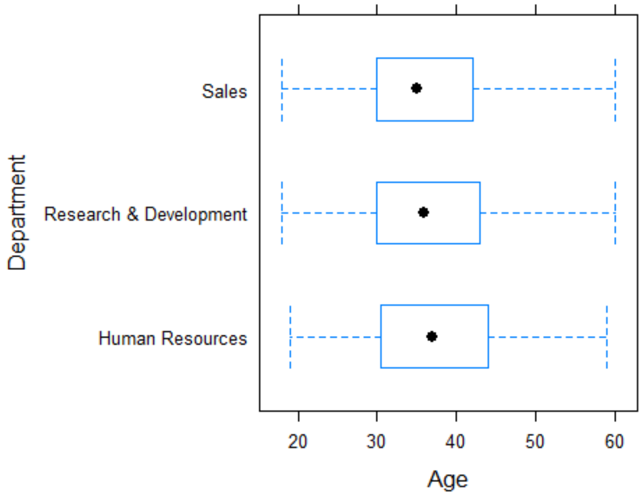
```
> #Boxplot distributions for  YearsWithCurrManager
> plot(attr$YearsWithCurrManager, xlab = "")
> abline(h = mean(attr$YearsWithCurrManager, na.rm = T), lty = 1)
> abline(h = mean(attr$YearsWithCurrManager, na.rm = T) + sd(attr$YearsWithCu
rrManager, na.rm = T),lty = 2)
> abline(h = median(attr$YearsWithCurrManager, na.rm = T), lty = 3)
> identify(attr$YearsWithCurrManager)
[1]  168  229  414  600  936 1222
```



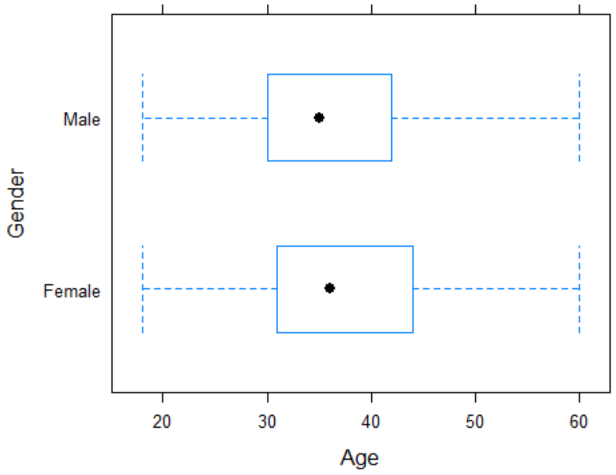
```
> #Chi Plot for inspecting the independence
> chi.plot(attr$MonthlyIncome,attr$Age)
```



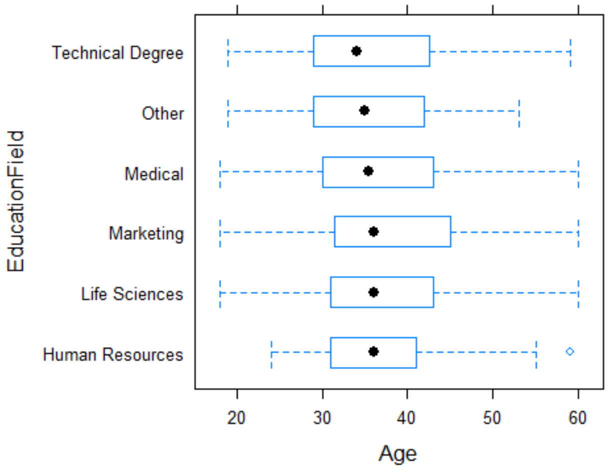
```
> #Plotting joint boxplots for various categories wrt Age
> bwplot(attr$Department ~ attr$Age, data=attr, ylab='Department',xlab='Age')
```



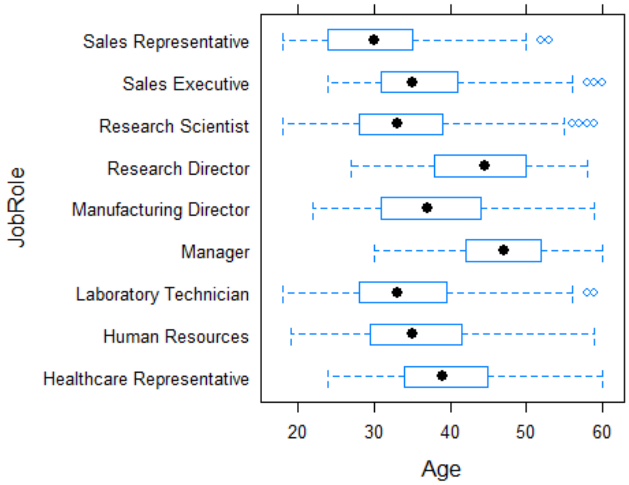
```
> bwplot(attr$Gender ~ attr$Age, data=attr, ylab='Gender',xlab='Age')
```



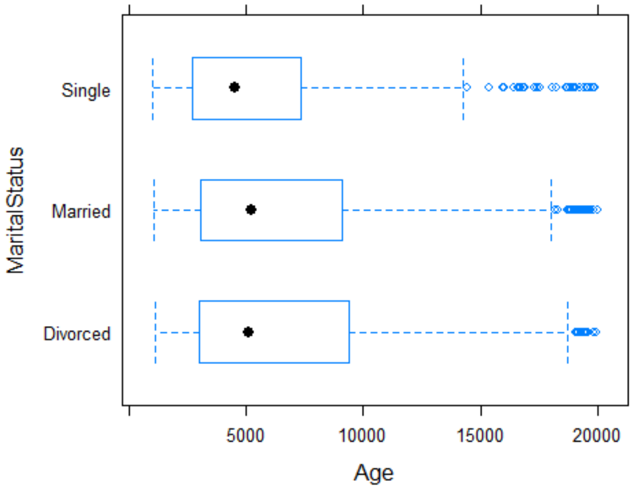
```
> bwplot(attr$EducationField ~ attr$Age, data=attr, ylab='EducationField',xla
b='Age')
```



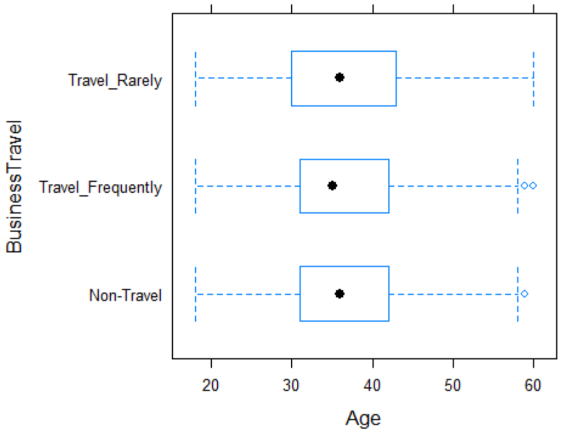
```
> bwplot(attr$JobRole ~ attr$Age, data=attr, ylab='JobRole',xlab='Age')
```



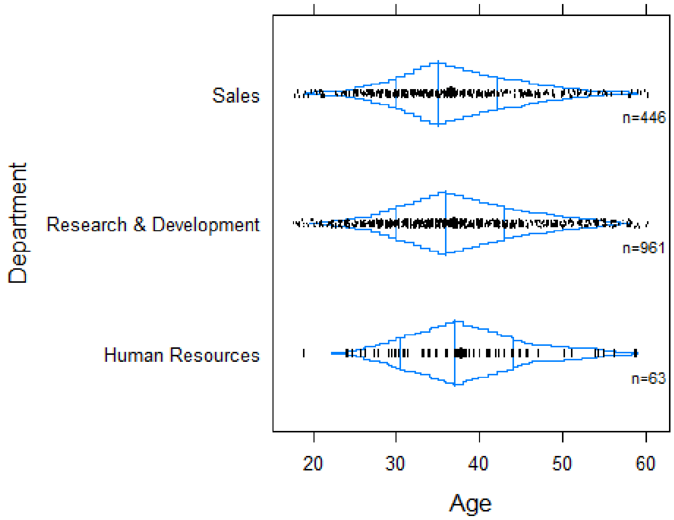
```
> bwplot(attr$MaritalStatus ~ attr$MonthlyIncome, data=attr, ylab='MaritalSta
tus',xlab='Age')
```



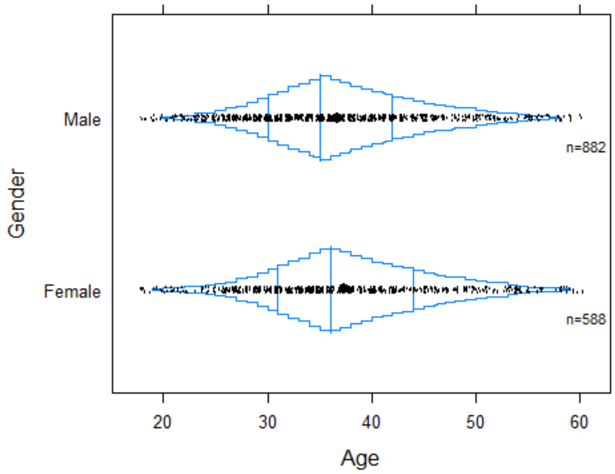
```
> bwplot(attr$BusinessTravel ~ attr$Age, data=attr, ylab='BusinessTravel',xla
b='Age')
```



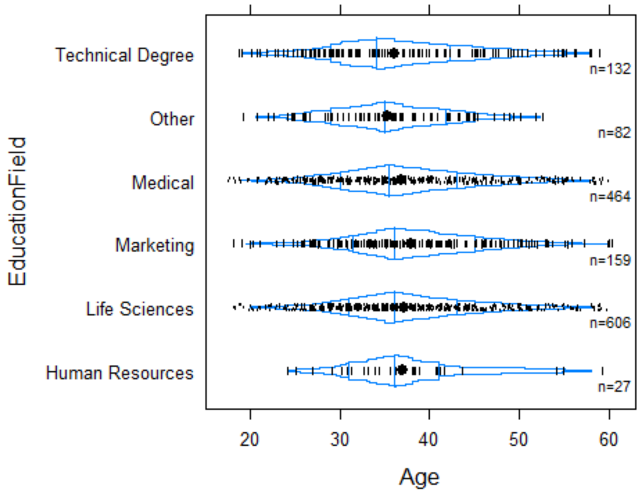
```
#Plotting stripplots for various categories wrt numerical column TotalCharges
> bwplot(attr$Department ~ attr$Age, data=attr,panel=panel.bpplot,
+         probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='Department',xlab=
'Age')
```



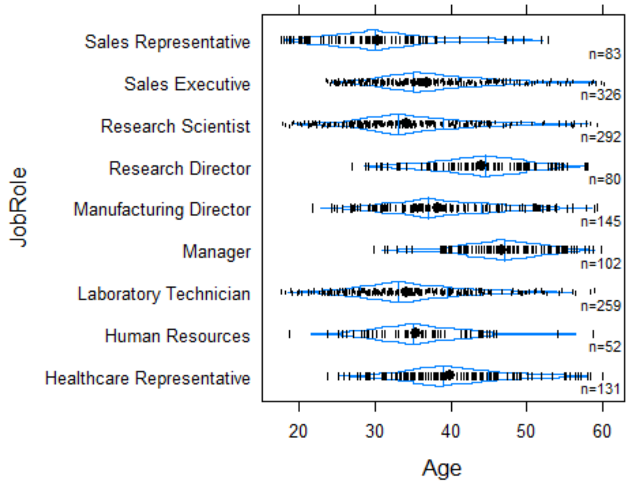
```
> bwplot(attr$Gender ~ attr$Age, data=attr,panel=panel.bpplot,
+         probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='Gender',xlab='Age')
```



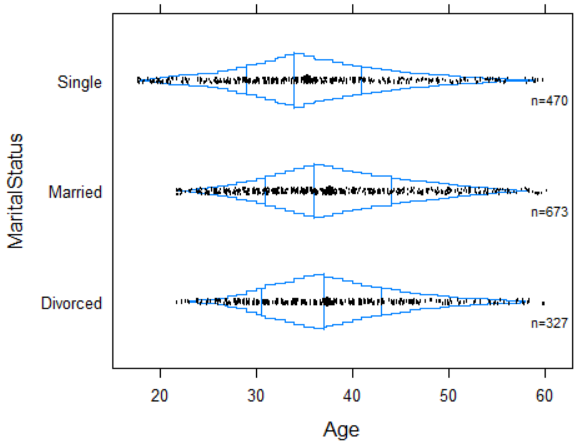
```
> bwplot(attr$EducationField ~ attr$Age, data=attr,panel=panel.bpplot,
+         probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='EducationField',xlab='Age')
```



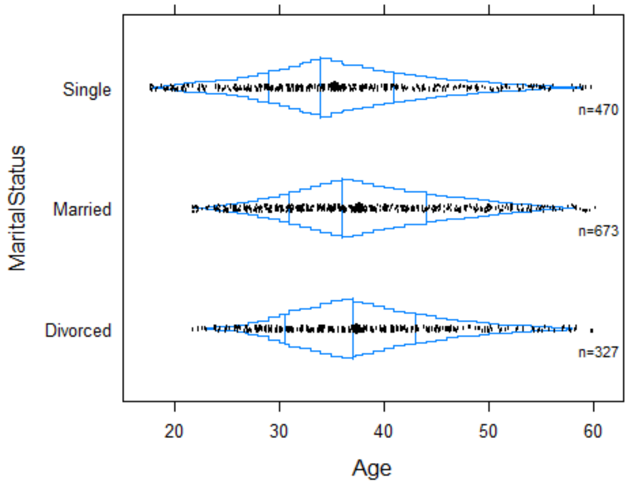
```
> bwplot(attr$JobRole ~ attr$Age, data=attr,panel=panel.bpplot,
+         probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='JobRole',xlab='Age')
```



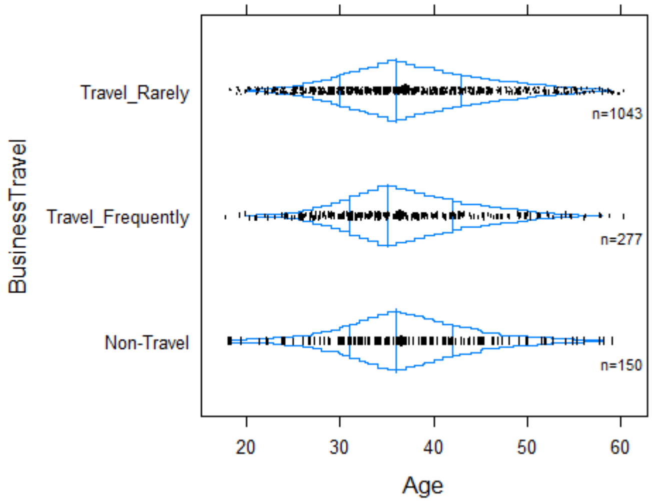
```
> bwplot(attr$MartialStatus ~ attr$Age, data=attr,panel=panel.bpplot,
+         probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='MartialStatus',xlab='Age')
```



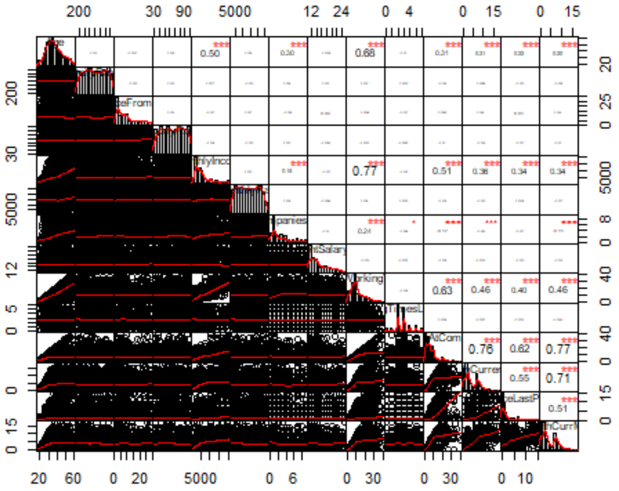
```
> bwplot(attr$MaritalStatus ~ attr$Age, data=attr,panel=panel.bpplot,
+        probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='MaritalStatus',xlab='Age')
```



```
> bwplot(attr$BusinessTravel ~ attr$Age, data=attr,panel=panel.bpplot,
+        probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='BusinessTravel',xlab='Age')
```



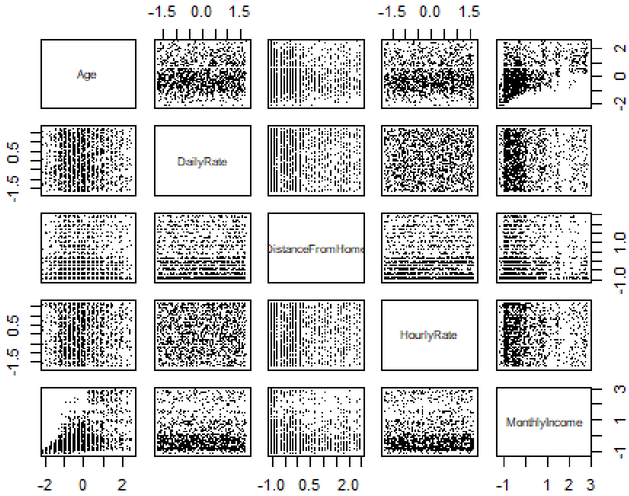
```
> data<-attr[,c('Age','DailyRate','DistanceFromHome','HourlyRate',
+               'MonthlyIncome','MonthlyRate','NumCompaniesWorked','PercentSalaryHike','T
+               'TrainingTimesLastYear','YearsAtCompany',
+               'YearsInCurrentRole','YearsSinceLastPromotion','YearsWithCurrManager')]
> chart.Correlation(data,histogram = TRUE,pch=19)
```



```
#---------------------------------------------------------------------------------
> ##Creating Temporary Variables
> #---------------------------------------------------------------------------------
>
> #Converting double/int columns to numeric
> numeric_col <- c("Age","DailyRate","DistanceFromHome","HourlyRate",
+                  "MonthlyIncome","MonthlyRate","NumCompaniesWorked","PercentSalaryHike"
+                  "TrainingTimesLastYear","YearsAtCompany",
+                  "YearsInCurrentRole","YearsSinceLastPromotion","YearsWithCurrManager")
> attr[numeric_col] <- sapply(attr[numeric_col], as.numeric)

e out the numeric columns from categorical columns and storing them as a seperate datafra
> attr_i <- attr[,c("Age","DailyRate","DistanceFromHome","HourlyRate",
+                   "MonthlyIncome","MonthlyRate","NumCompaniesWorked","PercentSalaryHike
+                   "TrainingTimesLastYear","YearsAtCompany",
+                   "YearsInCurrentRole","YearsSinceLastPromotion","YearsWithCurrManager"
> attr_i <- data.frame(scale(attr_i))

> #Creating temporary variables for the categorical data
> attr_c <- attr[,-c(2,3,5,8,10,11,12,13,14,15,19,21,22,23)]
> temporary<- data.frame(sapply(attr_c,function(x) data.frame(model.matrix(~x-1,data =att
> head(temporary)
  Education.x2 Education.x3 Education.x4 Education.x5 EnvironmentSatisfaction.x2 Environm
1            1            0            0            0                          1
2            0            0            0            0                          0
3            1            0            0            0                          0
4            0            0            1            0                          0
5            0            0            0            0                          0
6            1            0            0            0                          0
  EnvironmentSatisfaction.x4 MaritalStatus.xMarried MaritalStatus.xSingle OverTime StockO
1                          0                      0                     1        1
2                          0                      1                     0        0
3                          1                      0                     1        1
4                          1                      1                     0        1
5                          0                      1                     0        0
6                          1                      0                     1        0
  StockOptionLevel.x2 StockOptionLevel.x3 WorkLifeBalance.x2 WorkLifeBalance.x3 WorkLifeB
1                   0                   0                  0                  0
2                   0                   0                  0                  1
3                   0                   0                  0                  1
4                   0                   0                  0                  1
5                   0                   0                  0                  1
6                   0                   0                  1                  0
```

```
> View(attr)

> #Combining the temporary and the numeric columns and create the final dataset
> attr_final <- cbind(attr_i,temporary)
> head(attr_final)
          Age  DailyRate DistanceFromHome HourlyRate MonthlyIncome MonthlyRate NumCompani
1  0.44619856  0.7422739       -1.0105654  1.3826677    -0.1083127   0.7257730          2
2  1.32191535 -1.2973331       -0.1470997 -0.2405949    -0.2916193   1.4883696         -0
3  0.00834016  1.4138821       -0.8872132  1.2842882    -0.9373347  -1.6742711          1
4 -0.42951824  1.4609690       -0.7638609 -0.4865438    -0.7633739   1.2427877         -0
5 -1.08630583 -0.5241163       -0.8872132 -1.2735802    -0.6446387   0.3257890          2
6 -0.53898284  0.5018828       -0.8872132  0.6448211    -0.7296013  -0.3440822         -1
  PercentSalaryHike TotalWorkingYears TrainingTimesLastYear YearsAtCompany YearsInCurrent
1       -1.15016269        -0.4214990            -2.1712429   -0.164557109        -0.0632
2        2.12858163        -0.1644554             0.1556541    0.488341541         0.7647
3       -0.05724792        -0.5500208             0.1556541   -1.143905083        -1.1672
4       -1.15016269        -0.4214990             0.1556541    0.161892216         0.7647
5       -0.87693400        -0.6785426             0.1556541   -0.817455758        -0.6152
6       -0.60370530        -0.4214990            -0.6199782   -0.001332446         0.7647
  YearsSinceLastPromotion YearsWithCurrManager Education.x2 Education.x3 Education.x4 Edu
1             -0.67891464            0.2457504            1            0            0
2             -0.36858985            0.8062671            0            0            0
3             -0.67891464           -1.1555415            1            0            0
4              0.25205973           -1.1555415            0            0            1
5             -0.05826506           -0.5950247            0            0            0
6              0.25205973            0.5260087            1            0            0
  EnvironmentSatisfaction.x2 EnvironmentSatisfaction.x3 EnvironmentSatisfaction.x4 Marita
1                          1                          0                          0
2                          0                          1                          0
3                          0                          0                          1
4                          0                          0                          1
5                          0                          0                          0
6                          0                          0                          1
  MaritalStatus.xSingle OverTime StockOptionLevel.x1 StockOptionLevel.x2 StockOptionLevel
1                     1        1                   0                   0
2                     0        0                   1                   0
3                     1        1                   0                   0
4                     0        1                   0                   0
5                     0        0                   1                   0
6                     1        0                   0                   0
  WorkLifeBalance.x2 WorkLifeBalance.x3 WorkLifeBalance.x4
1                  0                  0                  0
2                  0                  1                  0
3                  0                  1                  0
4                  0                  1                  0
5                  0                  1                  0
6                  1                  0                  0
> glimpse(attr_final)
Observations: 1,470
Variables: 30
$ Age                        <dbl> 0.44619856, 1.32191535, 0.00834016, -0.42951824, -1.08
$ DailyRate                  <dbl> 0.74227393, -1.29733311, 1.41388208, 1.46096900, -0.52
$ DistanceFromHome           <dbl> -1.01056544, -0.14709966, -0.88721318, -0.76386093, -0
$ HourlyRate                 <dbl> 1.38266773, -0.24059489, 1.28428818, -0.48654378, -1.2
$ MonthlyIncome              <dbl> -0.108312654, -0.291619349, -0.937334707, -0.763373892
$ MonthlyRate                <dbl> 0.7257730, 1.4883696, -1.6742711, 1.2427877, 0.3257890
$ NumCompaniesWorked         <dbl> 2.1244130, -0.6778187, 1.3237753, -0.6778187, 2.524731
$ PercentSalaryHike          <dbl> -1.15016269, 2.12858163, -0.05724792, -1.15016269, -0.
$ TotalWorkingYears          <dbl> -0.42149902, -0.16445544, -0.55002081, -0.42149902, -0
$ TrainingTimesLastYear      <dbl> -2.1712429, 0.1556541, 0.1556541, 0.1556541, 0.1556541
$ YearsAtCompany             <dbl> -0.164557109, 0.488341541, -1.143905083, 0.161892216,
$ YearsInCurrentRole         <dbl> -0.06327437, 0.76473737, -1.16729002, 0.76473737, -0.6
$ YearsSinceLastPromotion    <dbl> -0.67891464, -0.36858985, -0.67891464, 0.25205973, -0.
$ YearsWithCurrManager       <dbl> 0.2457504, 0.8062671, -1.1555415, -1.1555415, -0.59502
$ Education.x2                <dbl> 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1,
$ Education.x3                <dbl> 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0,
$ Education.x4                <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
$ Education.x5                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
$ EnvironmentSatisfaction.x2  <dbl> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0,
$ EnvironmentSatisfaction.x3  <dbl> 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0,
$ EnvironmentSatisfaction.x4  <dbl> 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1,
$ MaritalStatus.xMarried      <dbl> 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
$ MaritalStatus.xSingle       <dbl> 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0,
$ OverTime                    <dbl> 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1,
$ StockOptionLevel.x1         <dbl> 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0,
$ StockOptionLevel.x2         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1,
$ StockOptionLevel.x3         <dbl> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
$ WorkLifeBalance.x2          <dbl> 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1,
$ WorkLifeBalance.x3          <dbl> 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0,
$ WorkLifeBalance.x4          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

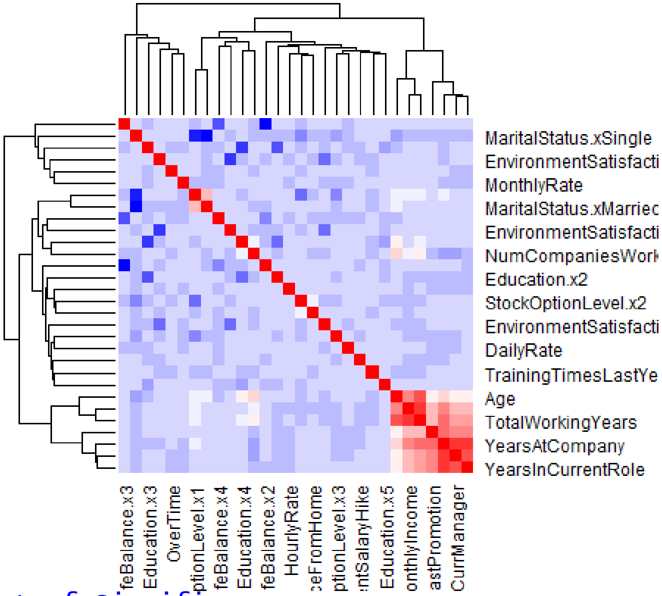
```
##Matrix Plots, Covariance and Corelations Plots
#ScatterPlot matrix
pairs(attr_final[,1:5],pch=".",cex=1.5)
```



```
> #CorrelationMatrix
> cormatrix <- round(cor(attr_final),4)
> cormatrix
                            Age DailyRate DistanceFromHome HourlyRate MonthlyIncome Mo
Age                      1.0000    0.0107          -0.0017     0.0243        0.4979
DailyRate                0.0107    1.0000          -0.0050     0.0234        0.0077
DistanceFromHome        -0.0017   -0.0050           1.0000     0.0311       -0.0170
HourlyRate               0.0243    0.0234           0.0311     1.0000       -0.0158
MonthlyIncome            0.4979    0.0077          -0.0170    -0.0158        1.0000
MonthlyRate              0.0281   -0.0322           0.0275    -0.0153        0.0348
NumCompaniesWorked       0.2996    0.0382          -0.0293     0.0222        0.1495
PercentSalaryHike        0.0036    0.0227           0.0402    -0.0091       -0.0273
TotalWorkingYears        0.6804    0.0145           0.0046    -0.0023        0.7729
TrainingTimesLastYear   -0.0196    0.0025          -0.0369    -0.0085       -0.0217
YearsAtCompany           0.3113   -0.0341           0.0095    -0.0196        0.5143
YearsInCurrentRole       0.2129    0.0099           0.0188    -0.0241        0.3638
YearsSinceLastPromotion  0.2165   -0.0332           0.0100    -0.0267        0.3450
YearsWithCurrManager     0.2021   -0.0264           0.0144    -0.0201        0.3441
Education.x2            -0.0033    0.0237           0.0008     0.0080       -0.0286
Education.x3            -0.0389   -0.0409           0.0050    -0.0097        0.0024
Education.x4             0.1573    0.0141          -0.0035     0.0054        0.0427
Education.x5             0.0598   -0.0077           0.0296     0.0230        0.0693
```

```
EnvironmentSatisfaction.x2 -0.0224   -0.0133           0.0247     0.0254       -0.0229
EnvironmentSatisfaction.x3 -0.0110    0.0029          -0.0013     0.0158       -0.0029
EnvironmentSatisfaction.x4  0.0219    0.0164          -0.0190    -0.0574        0.0036
MaritalStatus.xMarried      0.0839    0.0400           0.0302     0.0364        0.0568
MaritalStatus.xSingle      -0.1192   -0.0758          -0.0274    -0.0334       -0.0894
OverTime                    0.0281    0.0091           0.0255    -0.0078        0.0061
StockOptionLevel.x1         0.1072    0.0211          -0.0227    -0.0064        0.0907
StockOptionLevel.x2        -0.0281   -0.0092           0.0872     0.0638       -0.0244
StockOptionLevel.x3        -0.0046    0.0446          -0.0066     0.0092       -0.0355
WorkLifeBalance.x2          0.0160    0.0342           0.0091     0.0166       -0.0048
WorkLifeBalance.x3         -0.0101   -0.0126           0.0131     0.0122        0.0077
WorkLifeBalance.x4         -0.0133   -0.0316          -0.0386    -0.0243        0.0176
                           NumCompaniesWorked PercentSalaryHike TotalWorkingYears Trainin
Age                                    0.2996            0.0036            0.6804
DailyRate                              0.0382            0.0227            0.0145
DistanceFromHome                      -0.0293            0.0402            0.0046
HourlyRate                             0.0222           -0.0091           -0.0023
MonthlyIncome                          0.1495           -0.0273            0.7729
MonthlyRate                            0.0175           -0.0064            0.0264
NumCompaniesWorked                     1.0000           -0.0102            0.2376
PercentSalaryHike                     -0.0102            1.0000           -0.0206
TotalWorkingYears                      0.2376           -0.0206            1.0000
TrainingTimesLastYear                 -0.0661           -0.0052           -0.0357
YearsAtCompany                        -0.1184           -0.0360            0.6281
YearsInCurrentRole                    -0.0908           -0.0015            0.4604
YearsSinceLastPromotion               -0.0368           -0.0222            0.4049
YearsWithCurrManager                  -0.1103           -0.0120            0.4592
Education.x2                          -0.0211           -0.0029           -0.0355
Education.x3                           0.0014           -0.0171           -0.0020
Education.x4                           0.0951           -0.0069            0.0905
Education.x5                           0.0134            0.0219            0.0662
EnvironmentSatisfaction.x2            -0.0199            0.0023           -0.0263
EnvironmentSatisfaction.x3            -0.0242            0.0234           -0.0124
EnvironmentSatisfaction.x4             0.0319           -0.0414            0.0138
MaritalStatus.xMarried                -0.0161            0.0209            0.0535
MaritalStatus.xSingle                 -0.0192           -0.0014           -0.0895
OverTime                              -0.0208           -0.0054            0.0128
StockOptionLevel.x1                    0.0060            0.0508            0.0968
StockOptionLevel.x2                   -0.0084           -0.0085           -0.0438
StockOptionLevel.x3                    0.0398           -0.0190           -0.0168
WorkLifeBalance.x2                    -0.0048           -0.0347            0.0192
WorkLifeBalance.x3                    -0.0374            0.0327           -0.0087
WorkLifeBalance.x4                     0.0356           -0.0213            0.0012
                           YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion Year
Age                                0.3113             0.2129                  0.2165
DailyRate                         -0.0341             0.0099                 -0.0332
DistanceFromHome                   0.0095             0.0188                  0.0100
HourlyRate                        -0.0196            -0.0241                 -0.0267
MonthlyIncome                      0.5143             0.3638                  0.3450
MonthlyRate                       -0.0237            -0.0128                  0.0016
NumCompaniesWorked                -0.1184            -0.0908                 -0.0368
PercentSalaryHike                 -0.0360            -0.0015                 -0.0222
TotalWorkingYears                  0.6281             0.4604                  0.4049
TrainingTimesLastYear              0.0036            -0.0057                 -0.0021
YearsAtCompany                     1.0000             0.7588                  0.6184
YearsInCurrentRole                 0.7588             1.0000                  0.5481
YearsSinceLastPromotion            0.6184             0.5481                  1.0000
YearsWithCurrManager               0.7692             0.7144                  0.5102
Education.x2                      -0.0314            -0.0423                 -0.0289
Education.x3                      -0.0193             0.0142                  0.0046
Education.x4                       0.0549             0.0320                  0.0310
Education.x5                       0.0404             0.0306                  0.0297
EnvironmentSatisfaction.x2        -0.0012             0.0138                  0.0187
EnvironmentSatisfaction.x3         0.0212             0.0241                  0.0110
EnvironmentSatisfaction.x4        -0.0127            -0.0058                  0.0001
MaritalStatus.xMarried             0.0449             0.0655                  0.0541
MaritalStatus.xSingle             -0.0709            -0.0865                 -0.0531
OverTime                          -0.0117            -0.0298                 -0.0122
StockOptionLevel.x1                0.0828             0.0606                  0.0435
StockOptionLevel.x2               -0.0123             0.0241                  0.0036
StockOptionLevel.x3               -0.0289            -0.0020                 -0.0162
WorkLifeBalance.x2                 0.0079            -0.0203                  0.0187
WorkLifeBalance.x3                 0.0047             0.0340                 -0.0076
WorkLifeBalance.x4                 0.0006             0.0116                  0.0064
                           Education.x2 Education.x3 Education.x4 Education.x5 Environmen
Age                             -0.0033      -0.0389       0.1573       0.0598
DailyRate                        0.0237      -0.0409       0.0141      -0.0077
DistanceFromHome                 0.0008       0.0050      -0.0035       0.0296
HourlyRate                       0.0080      -0.0097       0.0054       0.0230
MonthlyIncome                   -0.0286       0.0024       0.0427       0.0693
MonthlyRate                     -0.0043      -0.0258      -0.0027       0.0053
NumCompaniesWorked              -0.0211       0.0014       0.0951       0.0134
PercentSalaryHike               -0.0029      -0.0171      -0.0069       0.0219
TotalWorkingYears               -0.0355      -0.0020       0.0905       0.0662
TrainingTimesLastYear            0.0182      -0.0024      -0.0382       0.0286
YearsAtCompany                  -0.0314      -0.0193       0.0549       0.0404
YearsInCurrentRole              -0.0423       0.0142       0.0320       0.0306
YearsSinceLastPromotion         -0.0289       0.0046       0.0310       0.0297
YearsWithCurrManager            -0.0173      -0.0068       0.0476       0.0291
Education.x2                     1.0000      -0.3888      -0.2969      -0.0895
Education.x3                    -0.3888       1.0000      -0.4863      -0.1466
Education.x4                    -0.2969      -0.4863       1.0000      -0.1119
Education.x5                    -0.0895      -0.1466      -0.1119       1.0000
EnvironmentSatisfaction.x2       0.0215      -0.0517       0.0011       0.0447
EnvironmentSatisfaction.x3      -0.0371       0.0445      -0.0121      -0.0397
EnvironmentSatisfaction.x4       0.0167       0.0166      -0.0258       0.0036
MaritalStatus.xMarried           0.0031      -0.0249      -0.0007       0.0309
MaritalStatus.xSingle           -0.0377       0.0333      -0.0041      -0.0111
OverTime                         0.0238      -0.0492       0.0114       0.0035
StockOptionLevel.x1              0.0551       0.0145      -0.0354       0.0120
StockOptionLevel.x2              0.0094      -0.0427       0.0555      -0.0267
StockOptionLevel.x3             -0.0245       0.0055      -0.0001       0.0201
WorkLifeBalance.x2              -0.0040       0.0268      -0.0258      -0.0292
WorkLifeBalance.x3              -0.0046      -0.0185       0.0038       0.0458
WorkLifeBalance.x4               0.0093      -0.0207       0.0280      -0.0250
                           EnvironmentSatisfaction.x3 EnvironmentSatisfaction.x4 MaritalS
Age                                           -0.0110                     0.0219
DailyRate                                      0.0029                     0.0164
DistanceFromHome                              -0.0013                    -0.0190
HourlyRate                                     0.0158                    -0.0574
MonthlyIncome                                 -0.0029                     0.0036
MonthlyRate                                    0.0312                     0.0207
NumCompaniesWorked                            -0.0242                     0.0319
PercentSalaryHike                              0.0234                    -0.0414
TotalWorkingYears                             -0.0124                     0.0138
TrainingTimesLastYear                          0.0216                    -0.0327
YearsAtCompany                                 0.0212                    -0.0127
YearsInCurrentRole                             0.0241                    -0.0058
YearsSinceLastPromotion                        0.0110                     0.0001
YearsWithCurrManager                           0.0022                    -0.0078
Education.x2                                  -0.0371                     0.0167
Education.x3                                   0.0445                     0.0166
Education.x4                                  -0.0121                    -0.0258
Education.x5                                  -0.0397                     0.0036
EnvironmentSatisfaction.x2                    -0.3287                    -0.3251
EnvironmentSatisfaction.x3                     1.0000                    -0.4405
EnvironmentSatisfaction.x4                    -0.4405                     1.0000
MaritalStatus.xMarried                         0.0373                    -0.0540
MaritalStatus.xSingle                     -0.0342                    0.0362
OverTime                                   0.0157                    0.0453
StockOptionLevel.x1                        0.0430                   -0.0055
StockOptionLevel.x2                       -0.0223                    0.0194
StockOptionLevel.x3                        0.0051                   -0.0240
WorkLifeBalance.x2                        -0.0209                   -0.0188
WorkLifeBalance.x3                         0.0085                    0.0184
WorkLifeBalance.x4                         0.0282                   -0.0069
                           MaritalStatus.xSingle OverTime StockOptionLevel.x1 StockOption
Age                                      -0.1192   0.0281              0.1072
DailyRate                                -0.0758   0.0091              0.0211
DistanceFromHome                         -0.0274   0.0255             -0.0227
HourlyRate                               -0.0334  -0.0078             -0.0064
MonthlyIncome                            -0.0894   0.0061              0.0907
MonthlyRate                               0.0373   0.0214             -0.0354
NumCompaniesWorked                       -0.0192  -0.0208              0.0060
PercentSalaryHike                        -0.0014  -0.0054              0.0508
TotalWorkingYears                        -0.0895   0.0128              0.0968
TrainingTimesLastYear                     0.0241  -0.0791             -0.0252
YearsAtCompany                           -0.0709  -0.0117              0.0828
YearsInCurrentRole                       -0.0865  -0.0298              0.0606
YearsSinceLastPromotion                  -0.0531  -0.0122              0.0435
YearsWithCurrManager                     -0.0478  -0.0416              0.0492
Education.x2                             -0.0377   0.0238              0.0551
Education.x3                              0.0333  -0.0492              0.0145
Education.x4                             -0.0041   0.0114             -0.0354
Education.x5                             -0.0111   0.0035              0.0120
EnvironmentSatisfaction.x2               -0.0212  -0.0008              0.0092
EnvironmentSatisfaction.x3               -0.0342   0.0157              0.0430
EnvironmentSatisfaction.x4                0.0362   0.0453             -0.0055
MaritalStatus.xMarried                   -0.6300  -0.0135              0.3564
MaritalStatus.xSingle                     1.0000  -0.0065             -0.5661
OverTime                                 -0.0065   1.0000             -0.0051
StockOptionLevel.x1                      -0.5661  -0.0051              1.0000
StockOptionLevel.x2                      -0.2379  -0.0327             -0.2866
StockOptionLevel.x3                      -0.1698   0.0320             -0.2046
WorkLifeBalance.x2                       -0.0241   0.0237              0.0246
WorkLifeBalance.x3                        0.0253   0.0040             -0.0285
WorkLifeBalance.x4                       -0.0044  -0.0361             -0.0092
                           StockOptionLevel.x3 WorkLifeBalance.x2 WorkLifeBalance.x3 Work
Age                                    -0.0046             0.0160            -0.0101
DailyRate                               0.0446             0.0342            -0.0126
DistanceFromHome                       -0.0066             0.0091             0.0131
HourlyRate                              0.0092             0.0166             0.0122
MonthlyIncome                          -0.0355            -0.0048             0.0077
MonthlyRate                            -0.0227            -0.0049             0.0243
NumCompaniesWorked                      0.0398            -0.0048            -0.0374
PercentSalaryHike                      -0.0190            -0.0347             0.0327
TotalWorkingYears                      -0.0168             0.0192            -0.0087
TrainingTimesLastYear                   0.0137            -0.0124             0.0197
YearsAtCompany                         -0.0289             0.0079             0.0047
YearsInCurrentRole                     -0.0020            -0.0203             0.0340
YearsSinceLastPromotion                -0.0162             0.0187            -0.0076
YearsWithCurrManager                   -0.0249            -0.0015             0.0125
Education.x2                           -0.0245            -0.0040            -0.0046
Education.x3                            0.0055             0.0268            -0.0185
Education.x4                           -0.0001            -0.0258             0.0038
Education.x5                            0.0201            -0.0292             0.0458
EnvironmentSatisfaction.x2              0.0250             0.0439            -0.0118
EnvironmentSatisfaction.x3              0.0051            -0.0209             0.0085
EnvironmentSatisfaction.x4             -0.0240            -0.0188             0.0184
MaritalStatus.xMarried                 -0.0171            -0.0145            -0.0107
MaritalStatus.xSingle                  -0.1698            -0.0241             0.0253
OverTime                                0.0320             0.0237             0.0040
```

```
StockOptionLevel.x1                      -0.2046           0.0246            -0.0285
StockOptionLevel.x2                      -0.0860           0.0261            -0.0179
StockOptionLevel.x3                       1.0000           0.0076             0.0320
WorkLifeBalance.x2                        0.0076           1.0000            -0.6876
WorkLifeBalance.x3                        0.0320          -0.6876             1.0000
WorkLifeBalance.x4                       -0.0176          -0.1884            -0.4240

> #Heatmap for correlation matrix
> #Negative correlations are shown in blue and positive in red
> col<- colorRampPalette(c("blue", "white", "red"))(20)
> heatmap(cormatrix, col=col, symm=TRUE)
```



```
##Test of Significance

#T-Test
#Null Hypothesis - The two means are equal
#Alternate Hypothesis - Difference in the two means is not zero
#pvalue >= 0.05, accept null hypothesis
#Or
#else accept the alternate hypothesis

#Univariate mean comparison using t test
> #Monthly Income and Attrition
> with(data=attr,t.test(attr$MonthlyIncome[attr$Attrition=="Yes"],attr$MonthlyIncome[attr
al=TRUE))

        Two Sample t-test

data:  attr$MonthlyIncome[attr$Attrition == "Yes"] and attr$MonthlyIncome[attr$Attrition
t = -6.2039, df = 1468, p-value = 7.147e-10
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2692.446 -1398.847
sample estimates:
mean of x mean of y
 4787.093  6832.740

> #HourlyRate and Attrition
> with(data=attr,t.test(attr$HourlyRate[attr$Attrition=="Yes"],attr$HourlyRate[attr$Attri
E))

        Two Sample t-test
```

```
data:  attr$HourlyRate[attr$Attrition == "Yes"] and attr$HourlyRate[attr$Attrition == "No
t = -0.26229, df = 1468, p-value = 0.7931
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.207565  2.450946
sample estimates:
mean of x mean of y
 65.57384  65.95215

> #Daily Rate and Attrition
> with(data=attr,t.test(attr$DailyRate[attr$Attrition=="Yes"],attr$DailyRate[attr$Attriti
)

	Two Sample t-test

data:  attr$DailyRate[attr$Attrition == "Yes"] and attr$DailyRate[attr$Attrition == "No"]
t = -2.1741, df = 1468, p-value = 0.02986
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -118.209251   -6.073932
sample estimates:
mean of x mean of y
 750.3629  812.5045

> #Age and Attrition
> with(data=attr,t.test(attr$Age[attr$Attrition=="Yes"],attr$Age[attr$Attrition=="No"],va

	Two Sample t-test

data:  attr$Age[attr$Attrition == "Yes"] and attr$Age[attr$Attrition == "No"]
t = -6.1787, df = 1468, p-value = 8.356e-10
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.208825 -2.698450
sample estimates:
mean of x mean of y
 33.60759  37.56123

> #DistanceFromHome and Attrition
> with(data = attr,t.test(attr$DistanceFromHome[attr$Attrition=="Yes"],attr$Age[attr$Attr
TRUE))

	Two Sample t-test

data:  attr$DistanceFromHome[attr$Attrition == "Yes"] and attr$Age[attr$Attrition == "No"
t = -43.048, df = 1468, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -28.15538 -25.70126
sample estimates:
mean of x mean of y
 10.63291  37.56123

> #Monthly Income and Gender
> with(data = attr,t.test(attr$MonthlyIncome[attr$Gender=="Male"],attr$MonthlyIncome[attr
ual = TRUE))

	Two Sample t-test

data:  attr$MonthlyIncome[attr$Gender == "Male"] and attr$MonthlyIncome[attr$Gender == "F
t = -1.2213, df = 1468, p-value = 0.2222
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -797.6470  185.5303
```

```
sample estimates:
mean of x mean of y 
 6380.508  6686.566 

> #DistanceFromHome and Gender
> with(data = attr,t.test(attr$DistanceFromHome[attr$Gender=="Male"],attr$DistanceFromHom
var.equal = TRUE))

	Two Sample t-test

data:  attr$DistanceFromHome[attr$Gender == "Male"] and attr$DistanceFromHome[attr$Gender
t = -0.070902, df = 1468, p-value = 0.9435
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8775316  0.8163071
sample estimates:
mean of x mean of y 
 9.180272  9.210884 

> #Monthly Income and gender
> t2testgender <- hotelling.test(attr$MonthlyIncome + attr$DistanceFromHome ~ attr$Gender
> cat("T2 statistic =",t2testgender$stat[[1]],"\n")
T2 statistic = 1.499903 
> print(t2testgender)
Test stat:  0.74944 
Numerator df:  2 
Denominator df:  1467 
P-value:  0.4728 

> #Monthly Income and Attrition
> t2testattr <- hotelling.test(attr$MonthlyIncome + attr$DistanceFromHome ~ attr$Attritio
> cat("T2 statistic =",t2testattr$stat[[1]],"\n")
T2 statistic = 47.28597 
> print(t2testattr)
Test stat:  23.627 
Numerator df:  2 
Denominator df:  1467 
P-value:  7.957e-11 
```