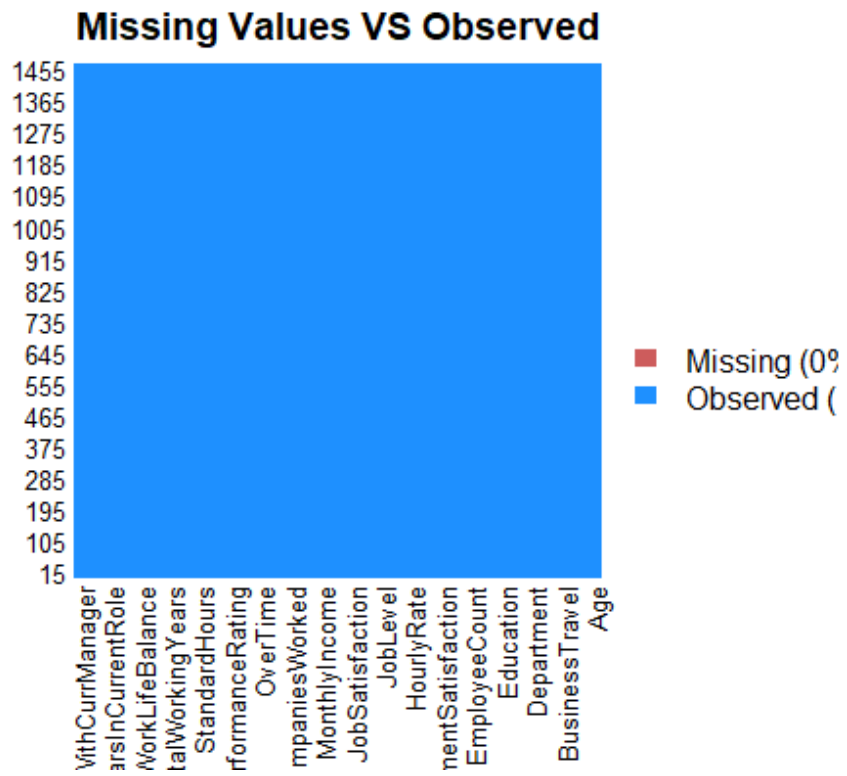
```
> ##Importing Data and inital analyses
> #Importing csv file from a location
> attr<- read.csv(file="MVA/Attrition Dataset.csv", header=TRUE, sep=",")
> attr <- as.data.frame(attr)
> glimpse(attr)

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
```

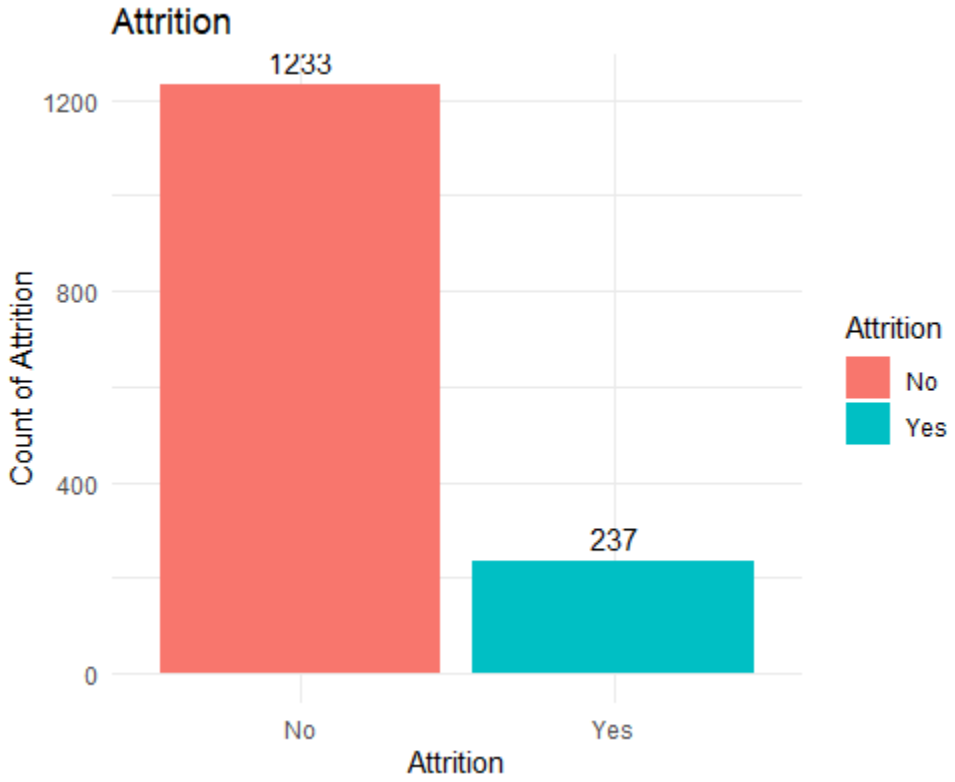
```
 EnvironmentSatisfaction    Gender      HourlyRate     JobInvolvement    JobL
evel
 Min.   :1.000           Female:588   Min.   : 30.00   Min.   :1.00   Min.
:1.000
 1st Qu.:2.000           Male  :882   1st Qu.: 48.00   1st Qu.:2.00   1st Qu.
:1.000
 Median :3.000                        Median : 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")
```

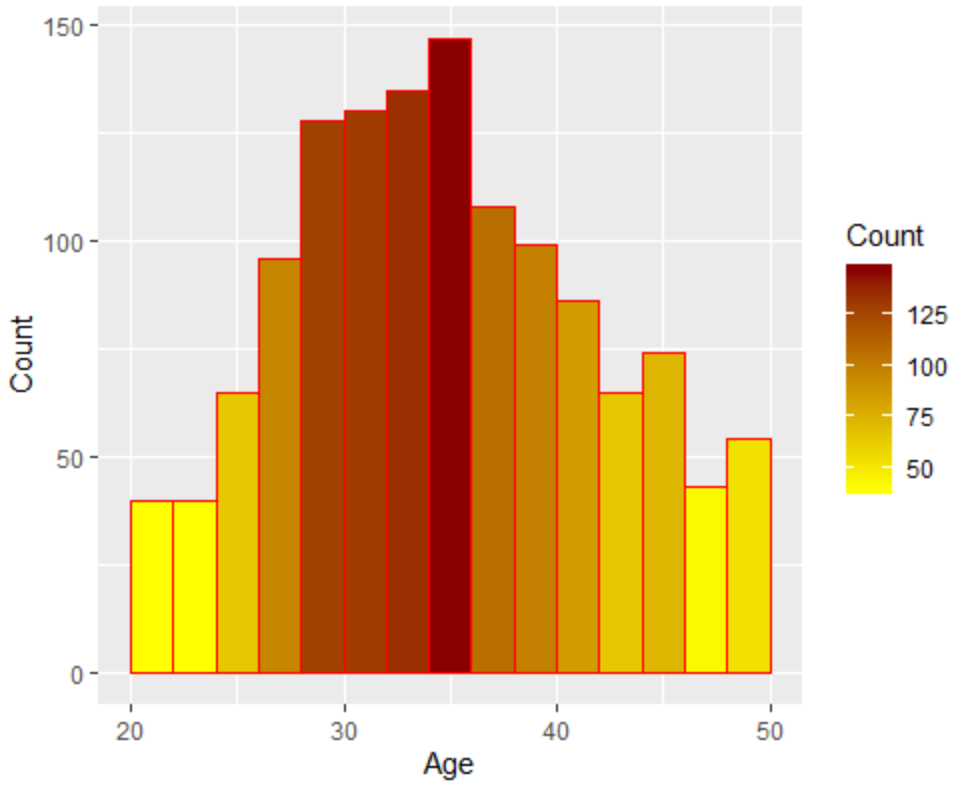



```
> #Removing redundant columns
> attr$EmployeeNumber<- NULL
> attr$StandardHours <- NULL
> attr$Over18 <- NULL
> attr$EmployeeCount <- NULL
> #Converting data type of categorical column
> attr$Education <- factor(attr$Education)
> attr$EnvironmentSatisfaction <- factor(attr$EnvironmentSatisfaction)
> 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) %>%
```

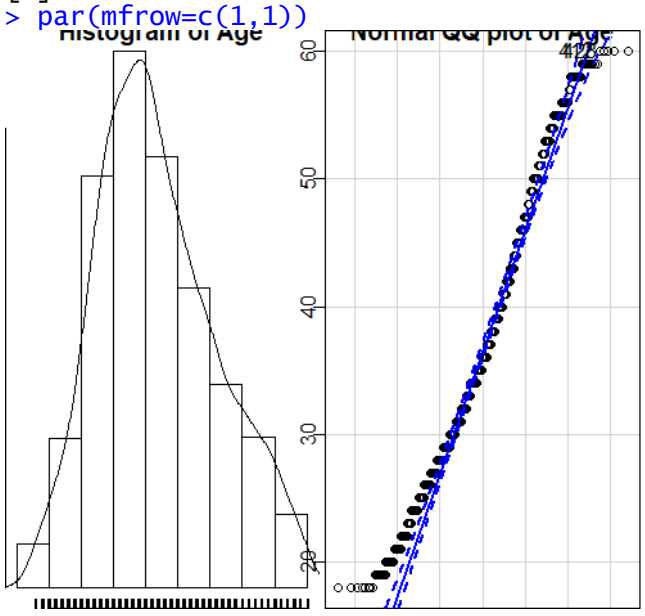
```
+    tally() %>%
+    ggplot(aes(x =Attrition,y = n,fill=Attrition)) +
+    geom_bar(stat = "identity") +
+    theme_minimal()+
+    labs(x="Attrition", y="Count of Attrition")+
+    ggtitle("Attrition")+
+    geom_text(aes(label = n), vjust = -0.5, position = position_dodge(0.9))
```



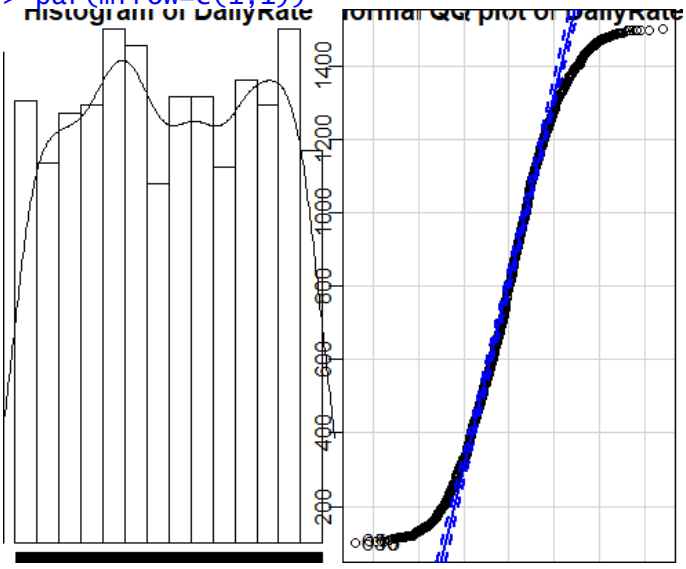

```
#Influence of features on Attrition
> ggplot(data=attr, aes(attr$Age)) +
+   geom_histogram(breaks=seq(20, 50, by=2),
+                  col="red",
+                  aes(fill=..count..))+
+   labs(x="Age", y="Count")+
+   scale_fill_gradient("Count", low="yellow", high="dark red")
```



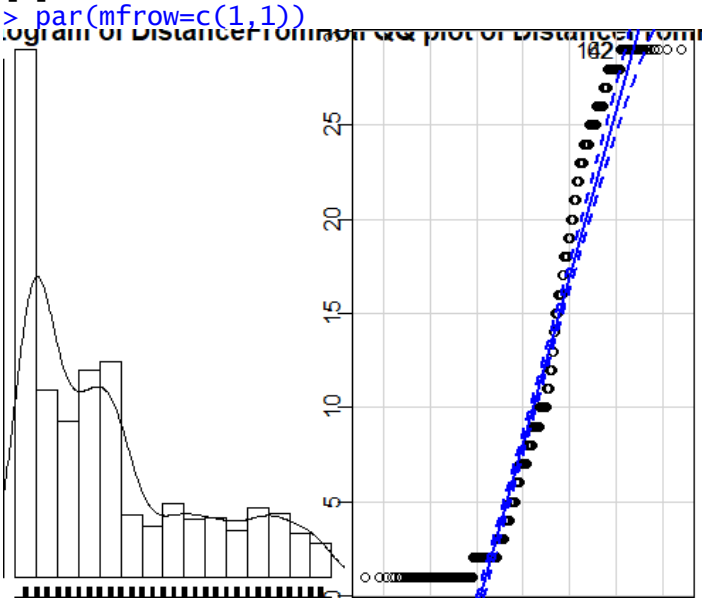

```
> #Checking for distributions in numerical columns
> #The qqPlot show a few extreme outliers which break the assumption of 95% c
onfidence
> #normal distribution
> par(mfrow = c(1,2))
> hist(attr$Age,xlab='',main = 'Histogram of Age',freq = FALSE)
> lines(density(attr$Age,na.rm = T))
> rug(jitter(attr$Age))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

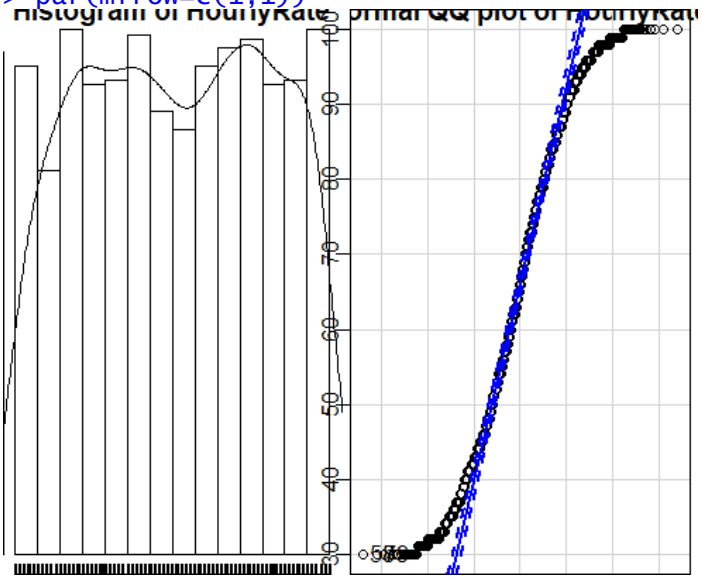



```
> par(mfrow = c(1,2))
> hist(attr$DistanceFromHome,xlab='',main = 'Histogram of DistanceFromHome',f
req = FALSE)
> lines(density(attr$DistanceFromHome,na.rm = T))
> rug(jitter(attr$DistanceFromHome))
> qqPlot(attr$DistanceFromHome,main='Normal QQ plot of DistanceFromHome')
[1]  62 142
> par(mfrow=c(1,1))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

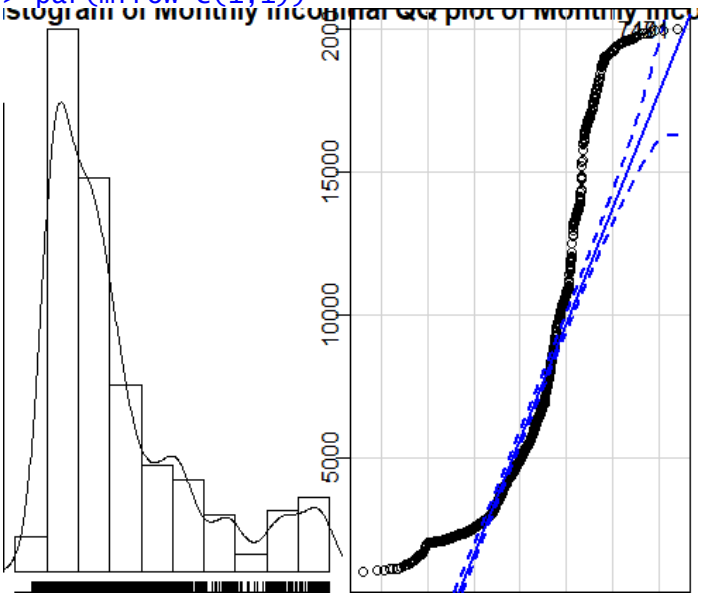



```
> par(mfrow = c(1,2))
> hist(attr$HourlyRate,xlab='',main = 'Histogram of HourlyRate',freq = FALSE)
> lines(density(attr$HourlyRate,na.rm = T))
```

```
> rug(jitter(attr$HourlyRate))
> qqPlot(attr$HourlyRate,main='Normal QQ plot of HourlyRate')
[1] 58 79
> par(mfrow=c(1,1))
```

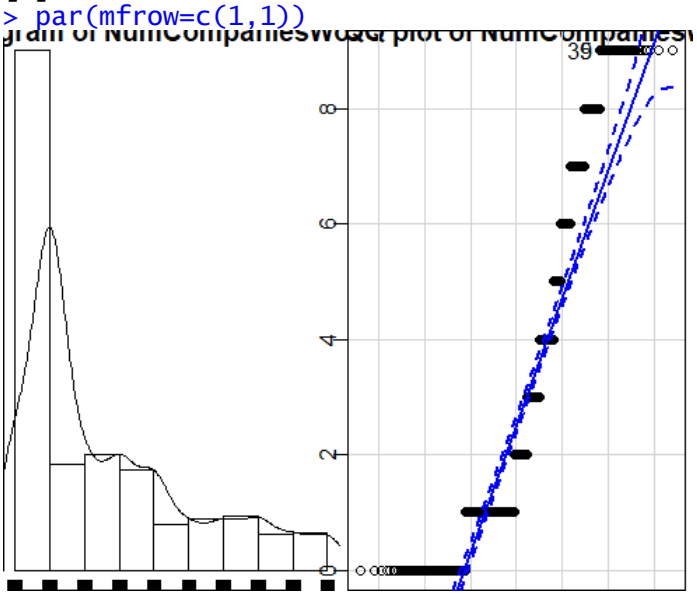


```
> par(mfrow = c(1,2))
> hist(attr$MonthlyIncome,xlab='',main = 'Histogram of Monthly Income',freq =
FALSE)
> lines(density(attr$MonthlyIncome,na.rm = T))
> rug(jitter(attr$MonthlyIncome))
> qqPlot(attr$MonthlyIncome,main='Normal QQ plot of Monthly Income')
[1] 191 747
> par(mfrow=c(1,1))
```

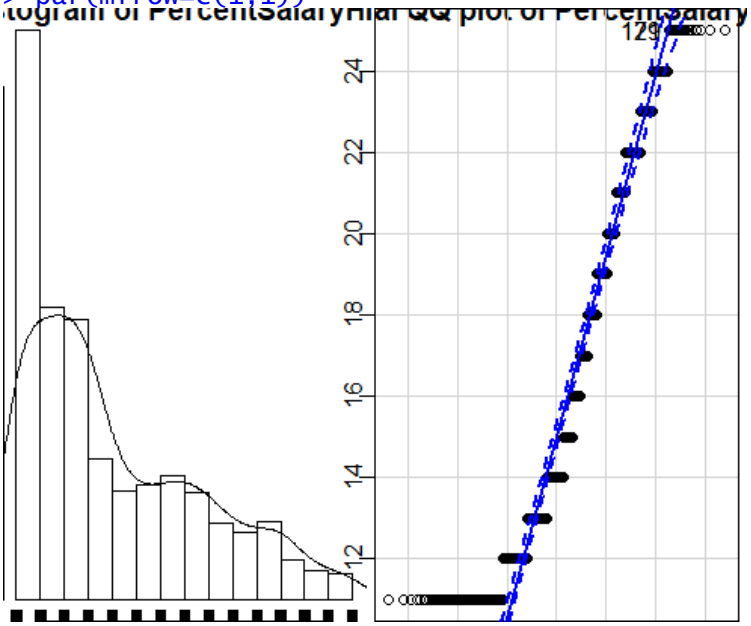


```
> par(mfrow = c(1,2))
> hist(attr$NumCompaniesWorked,xlab='',main = 'Histogram of NumCompaniesWorke
d',freq = FALSE)
> lines(density(attr$NumCompaniesWorked,na.rm = T))
> rug(jitter(attr$NumCompaniesWorked))
```

```
> qqPlot(attr$NumCompaniesWorked,main='Normal QQ plot of NumCompaniesWorked')
[1]  5 39
> par(mfrow=c(1,1))
```

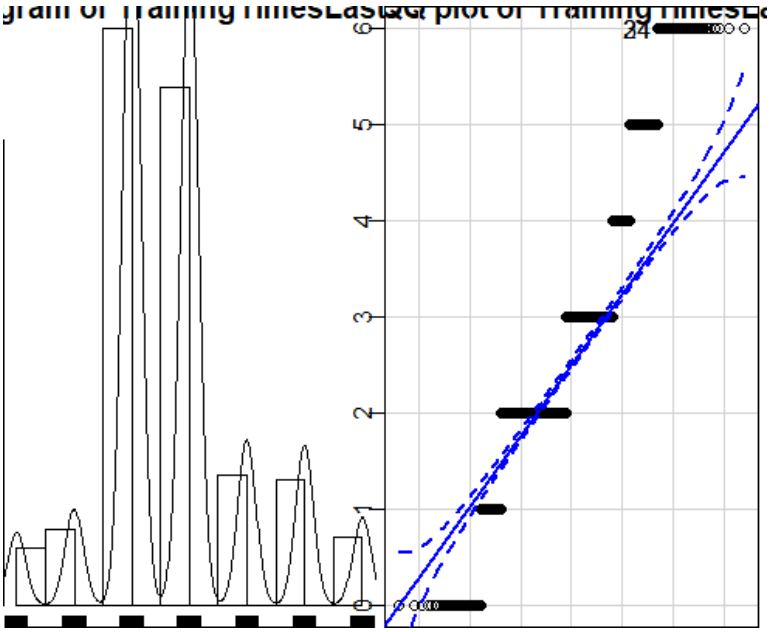



```
> par(mfrow = c(1,2))
> hist(attr$PercentSalaryHike,xlab='',main = 'Histogram of PercentSalaryHike'
,freq = FALSE)
> lines(density(attr$PercentSalaryHike,na.rm = T))
> rug(jitter(attr$PercentSalaryHike))
> qqPlot(attr$PercentSalaryHike,main='Normal QQ plot of PercentSalaryHike')
[1] 121 179
> par(mfrow=c(1,1))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

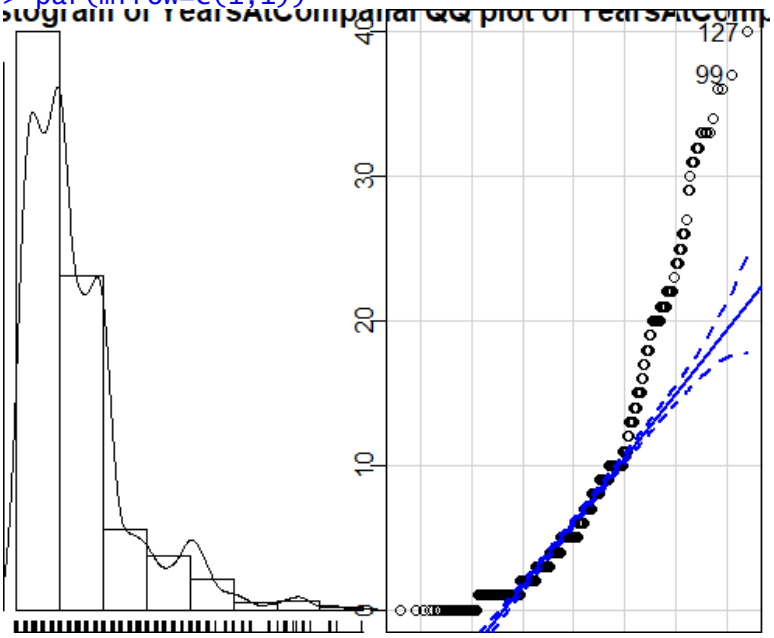



```
> par(mfrow = c(1,2))
> hist(attr$TrainingTimesLastYear,xlab='',main = 'Histogram of TrainingTimesL
astYear',freq = FALSE)
> lines(density(attr$TrainingTimesLastYear,na.rm = T))
> rug(jitter(attr$TrainingTimesLastYear))
```

```
> qqPlot(attr$TrainingTimesLastYear,main='Normal QQ plot of TrainingTimesLast
Year')
[1] 24 34
> par(mfrow=c(1,1))
```

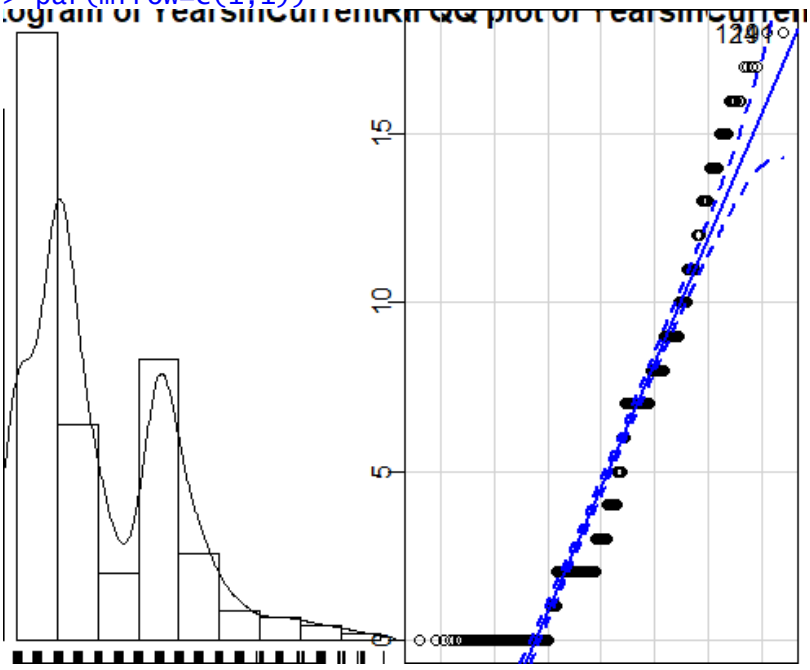



```
> par(mfrow = c(1,2))
> hist(attr$YearsAtCompany,xlab='',main = 'Histogram of YearsAtCompany',freq
= FALSE)
> lines(density(attr$YearsAtCompany,na.rm = T))
> rug(jitter(attr$YearsAtCompany))
> qqPlot(attr$YearsAtCompany,main='Normal QQ plot of YearsAtCompany')
[1] 127  99
> par(mfrow=c(1,1))
```

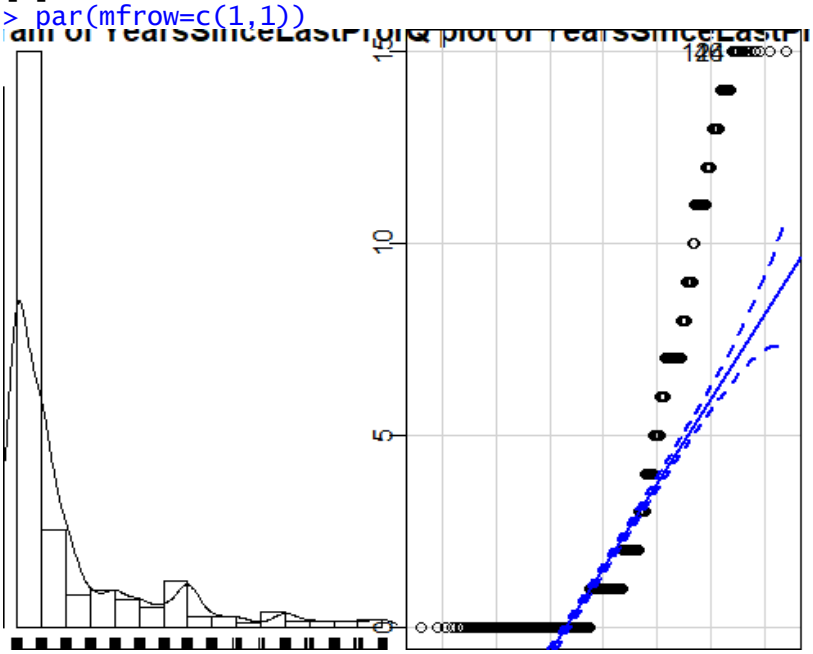



```
> par(mfrow = c(1,2))
```

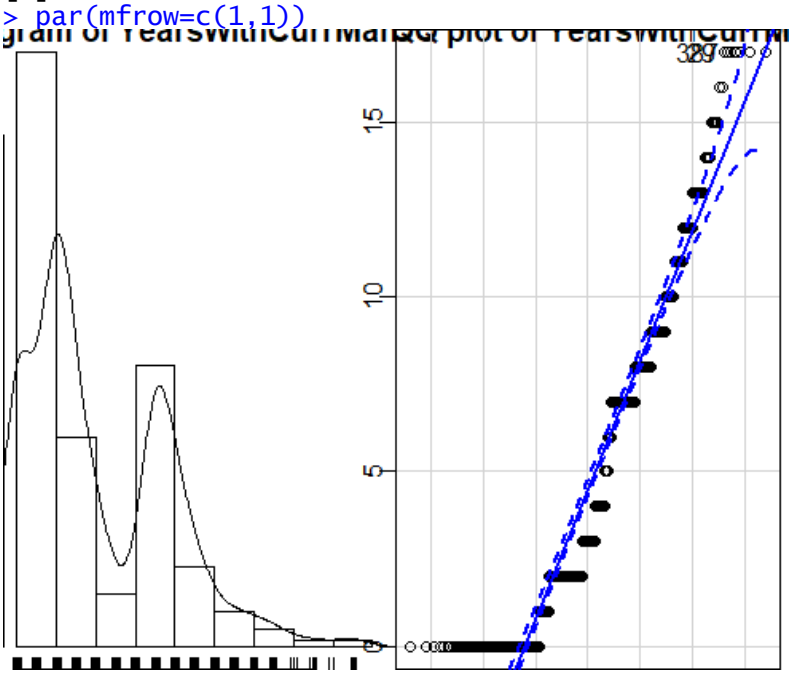
```
> hist(attr$YearsInCurrentRole,xlab='',main = 'Histogram of YearsInCurrentRol
e',freq = FALSE)
> lines(density(attr$YearsInCurrentRole,na.rm = T))
> rug(jitter(attr$YearsInCurrentRole))
> qqPlot(attr$YearsInCurrentRole,main='Normal QQ plot of YearsInCurrentRole')
[1] 124 191
> par(mfrow=c(1,1))
```



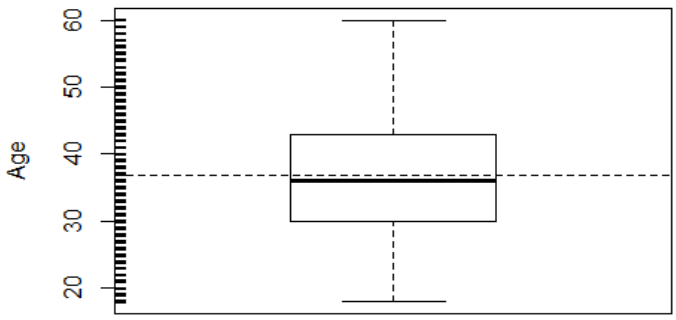

```
> par(mfrow = c(1,2))
> hist(attr$YearsSinceLastPromotion,xlab='',main = 'Histogram of YearsSinceLa
stPromotion',freq = FALSE)
> lines(density(attr$YearsSinceLastPromotion,na.rm = T))
> rug(jitter(attr$YearsSinceLastPromotion))
> qqPlot(attr$YearsSinceLastPromotion,main='Normal QQ plot of YearsSinceLastP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



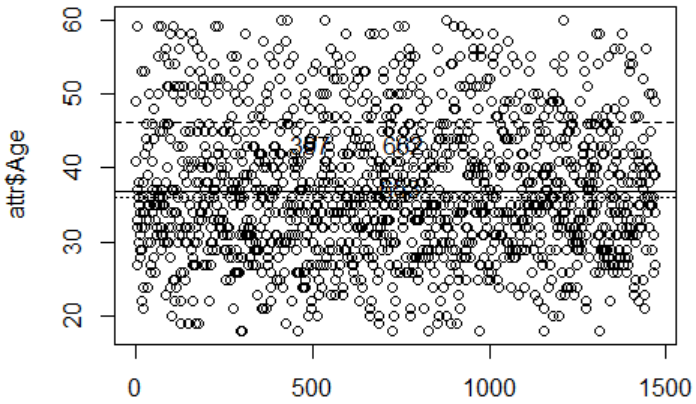
```
> par(mfrow = c(1,2))
> hist(attr$YearsWithCurrManager,xlab='',main = 'Histogram of YearsWithCurrMa
nager',freq = FALSE)
> lines(density(attr$YearsWithCurrManager,na.rm = T))
> rug(jitter(attr$YearsWithCurrManager))
> qqPlot(attr$YearsWithCurrManager,main='Normal QQ plot of YearsWithCurrManag
er')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



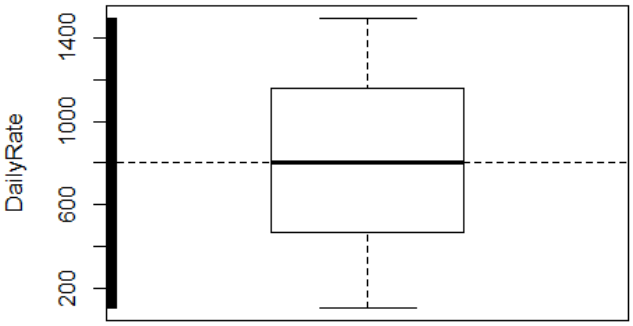

```
> #Boxplot distributions for our numeric columns
> #The dashed line shows the mean and the dark center line shows the median
> #Difference between these two lines depict the deviation from the central l
imit theorem
> #Boxplot distributions for  Age
> boxplot(attr$Age, ylab = "Age")
> rug(jitter(attr$Age),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



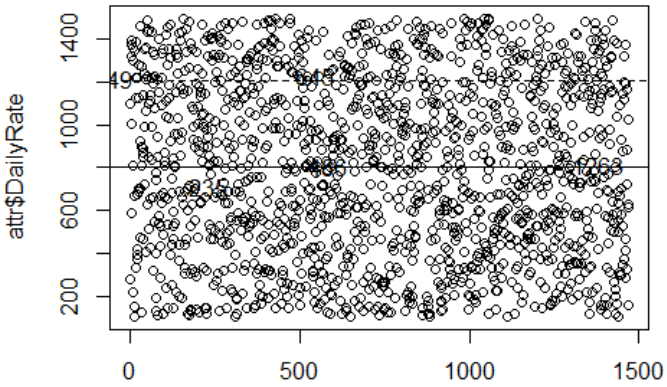
```
> #Plotting the Age  with 3 lines for mean, median and mean+std
> plot(attr$Age, xlab = "")
> abline(h = mean(attr$Age, na.rm = T), lty = 1)
> abline(h = mean(attr$Age, na.rm = T) + sd(attr$Age, na.rm = T),lty = 2)
> abline(h = median(attr$Age, na.rm = T), lty = 3)
> identify(attr$Age)
[1]  286 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



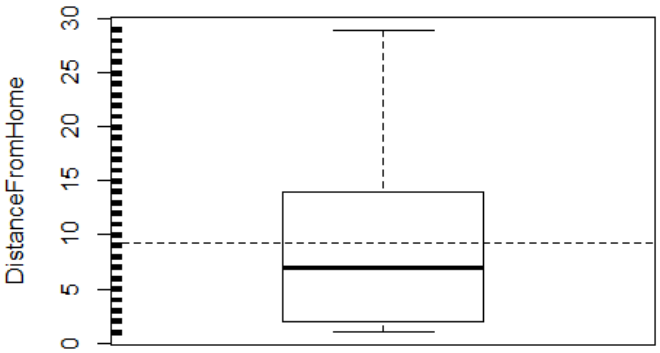

```
> #Boxplot distributions for Daily rate
> boxplot(attr$DailyRate, ylab = "DailyRate",outline = TRUE)
> rug(jitter(attr$DailyRate), side = 2)
> abline(h = mean(attr$DailyRate, na.rm = T), lty = 2)
```



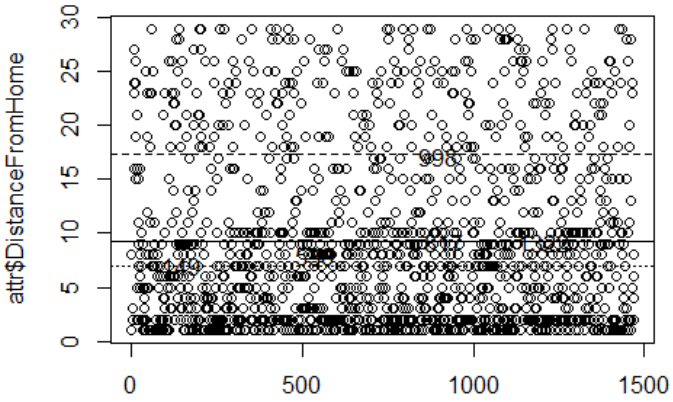

```
> #Plotting the DailyRate  with 3 lines for mean, median and mean+std
> plot(attr$DailyRate, xlab = "")
> abline(h = mean(attr$DailyRate, na.rm = T), lty = 1)
> abline(h = mean(attr$DailyRate, na.rm = T) + sd(attr$DailyRate, na.rm = T),
lty = 2)
> abline(h = median(attr$DailyRate, na.rm = T), lty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



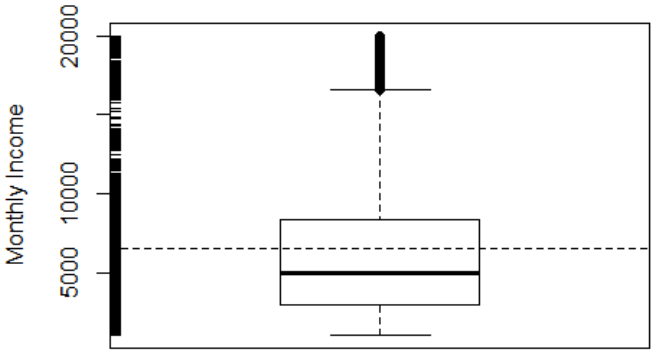

```
> #Boxplot distributions for Distance from home
> boxplot(attr$DistanceFromHome, ylab = "DistanceFromHome",outline = TRUE)
> rug(jitter(attr$DistanceFromHome), side = 2)
> abline(h = mean(attr$DistanceFromHome, na.rm = T), lty = 2)
```



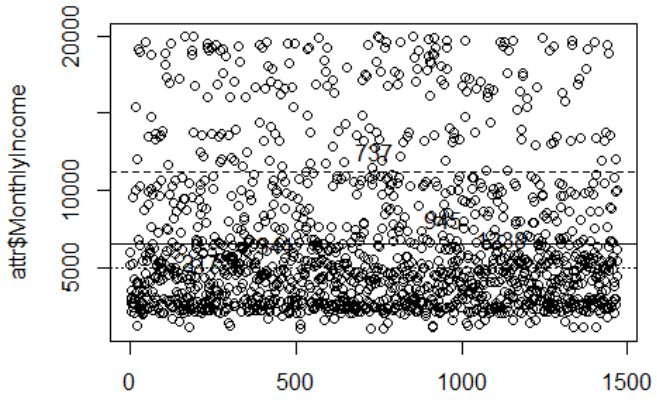

```
> plot(attr$DistanceFromHome, xlab = "")
> abline(h = mean(attr$DistanceFromHome, na.rm = T), lty = 1)
> abline(h = mean(attr$DistanceFromHome, na.rm = T) + sd(attr$DistanceFromHom
e, na.rm = T),lty = 2)
> abline(h = median(attr$DistanceFromHome, na.rm = T), lty = 3)
> identify(attr$DistanceFromHome)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



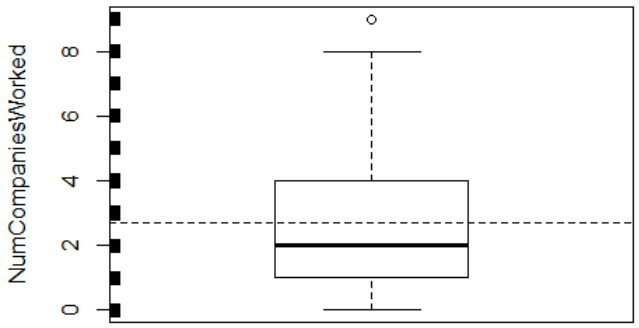

```
> #Boxplot distributions for Monthly Income
> boxplot(attr$MonthlyIncome, ylab = "Monthly Income")
> rug(jitter(attr$MonthlyIncome), side = 2)
> abline(h = mean(attr$MonthlyIncome, na.rm = T), lty = 2)
```



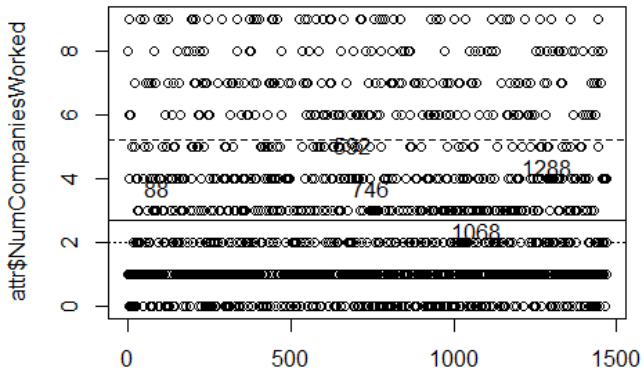

```
> #Plotting the Monthly Income and Age  with 3 lines for mean, median and mea
n+std
> plot(attr$MonthlyIncome, xlab = "")
> abline(h = mean(attr$MonthlyIncome, na.rm = T), lty = 1)
> abline(h = mean(attr$MonthlyIncome, na.rm = T) + sd(attr$MonthlyIncome, na.
rm = T),lty = 2)
> abline(h = median(attr$MonthlyIncome,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



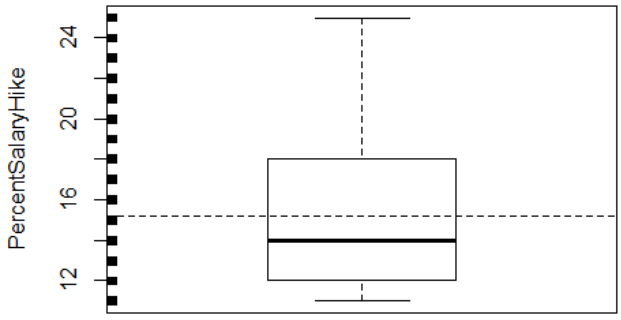

```
> #Boxplot distributions for  NumCompaniesWorked
> boxplot(attr$NumCompaniesWorked, ylab = "NumCompaniesWorked")
> rug(jitter(attr$NumCompaniesWorked), side = 2)
> abline(h = mean(attr$NumCompaniesWorked, na.rm = T), lty = 2)
```



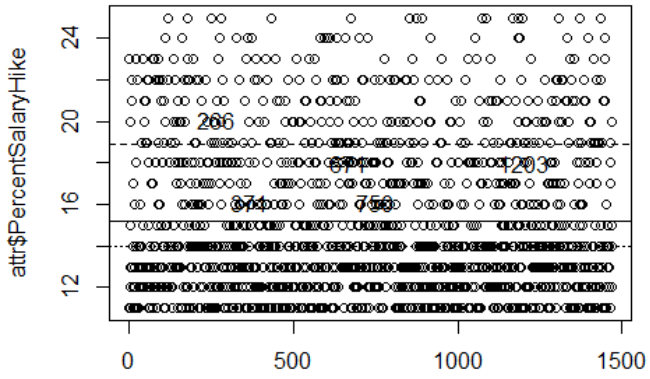

```
> #Plotting the NumCompaniesWorked  with 3 lines for mean, median and mean+st
d
> plot(attr$NumCompaniesWorked, xlab = "")
> abline(h = mean(attr$NumCompaniesWorked, na.rm = T), lty = 1)
> abline(h = mean(attr$NumCompaniesWorked, na.rm = T) + sd(attr$NumCompaniesW
orked, na.rm = T),lty = 2)
> abline(h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



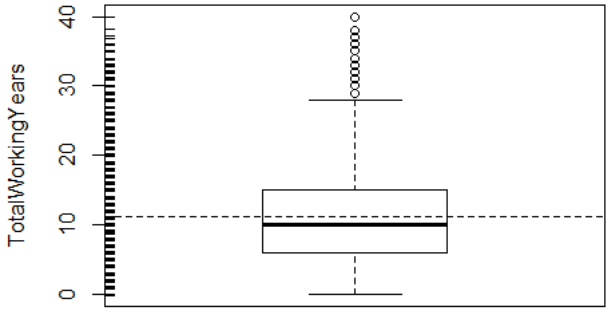

```
> #Boxplot distributions for  PercentSalaryHike
> boxplot(attr$PercentSalaryHike, ylab = "PercentSalaryHike")
> rug(jitter(attr$PercentSalaryHike), side = 2)
> abline(h = mean(attr$PercentSalaryHike, na.rm = T), lty = 2)
```



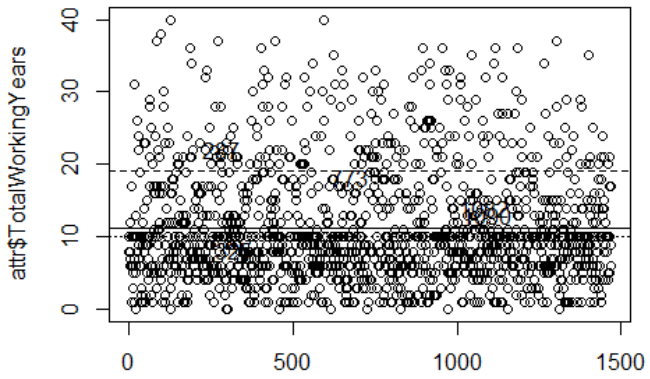

```
> #Plotting the PercentSalaryHike  with 3 lines for mean, median and mean+std
> plot(attr$PercentSalaryHike, xlab = "")
> abline(h = mean(attr$PercentSalaryHike, na.rm = T), lty = 1)
> abline(h = mean(attr$PercentSalaryHike, na.rm = T) + sd(attr$PercentSalaryH
ike, na.rm = T),lty = 2)
> abline(h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



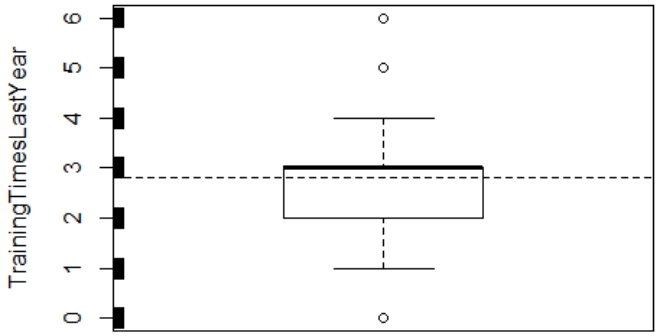

```
> #Boxplot distributions for  TotalWorkingYears
> boxplot(attr$TotalWorkingYears, ylab = "TotalWorkingYears")
> rug(jitter(attr$TotalWorkingYears), side = 2)
> abline(h = mean(attr$TotalWorkingYears, na.rm = T), lty = 2)
```



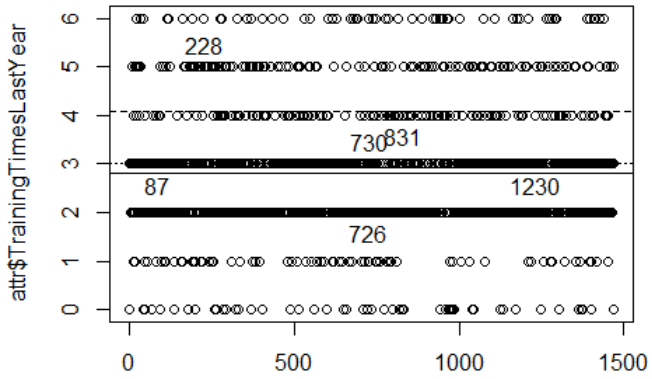

```
> #Plotting the TotalWorkingYears  with 3 lines for mean, median and mean+std
> plot(attr$TotalWorkingYears, xlab = "")
> abline(h = mean(attr$TotalWorkingYears, na.rm = T), lty = 1)
> abline(h = mean(attr$TotalWorkingYears, na.rm = T) + sd(attr$TotalWorkingYe
ars, na.rm = T),lty = 2)
> abline(h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



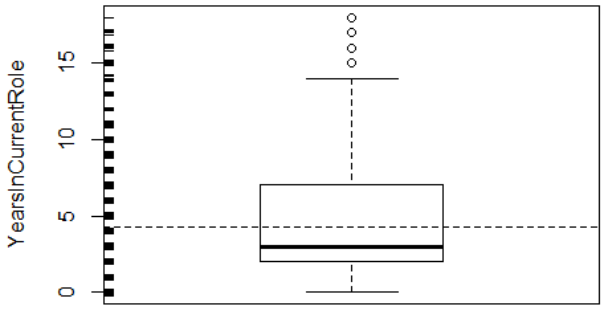

```
> #Boxplot distributions for  TrainingTimesLastYear
> boxplot(attr$TrainingTimesLastYear, ylab = "TrainingTimesLastYear")
> rug(jitter(attr$TrainingTimesLastYear), side = 2)
> abline(h = mean(attr$TrainingTimesLastYear, na.rm = T), lty = 2)
```



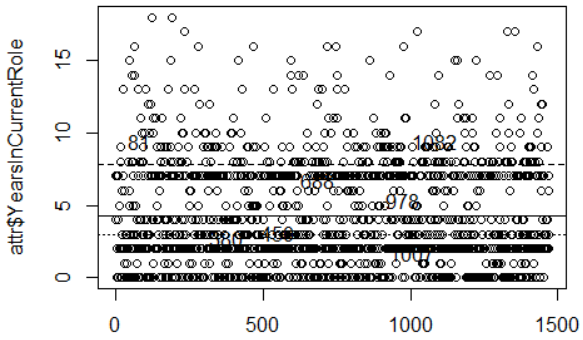

```
> #Plotting the TrainingTimesLastYear  with 3 lines for mean, median and mean
+std
> plot(attr$TrainingTimesLastYear, xlab = "")
> abline(h = mean(attr$TrainingTimesLastYear, na.rm = T), lty = 1)
> abline(h = mean(attr$TrainingTimesLastYear, na.rm = T) + sd(attr$TrainingTi
mesLastYear, na.rm = T),lty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



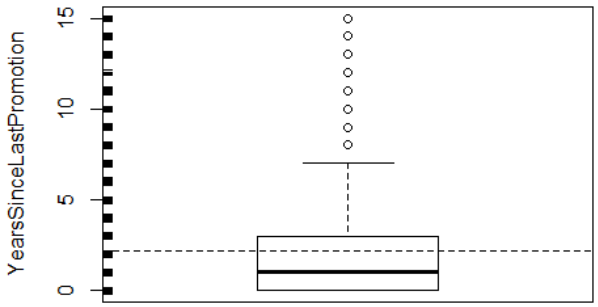

```
> #Boxplot distributions for  YearsInCurrentRole
> boxplot(attr$YearsInCurrentRole, ylab = "YearsInCurrentRole")
> rug(jitter(attr$YearsInCurrentRole), side = 2)
> abline(h = mean(attr$YearsInCurrentRole, na.rm = T), lty = 2)
```



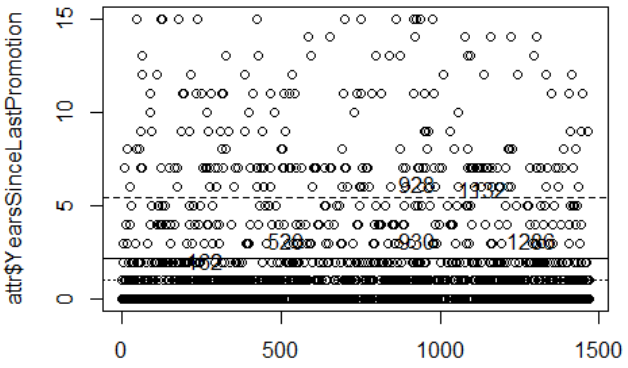

```
> #Plotting the YearsInCurrentRole  with 3 lines for mean, median and mean+st
d
> plot(attr$YearsInCurrentRole, xlab = "")
> abline(h = mean(attr$YearsInCurrentRole, na.rm = T), lty = 1)
> abline(h = mean(attr$YearsInCurrentRole, na.rm = T) + sd(attr$YearsInCurren
tRole, na.rm = T),lty = 2)
> abline(h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



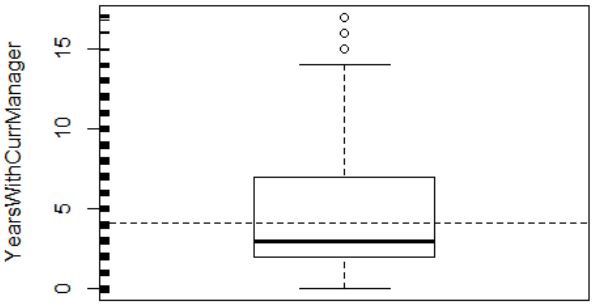

```
> #Boxplot distributions for  YearsSinceLastPromotion
> boxplot(attr$YearsSinceLastPromotion, ylab = "YearsSinceLastPromotion")
> rug(jitter(attr$YearsSinceLastPromotion), side = 2)
> abline(h = mean(attr$YearsSinceLastPromotion, na.rm = T), lty = 2)
```



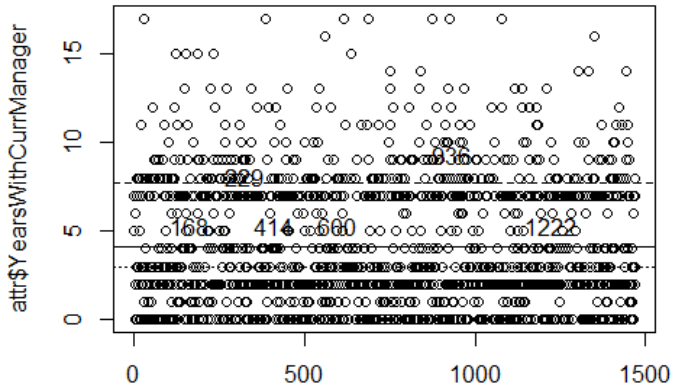

```
> #Plotting the YearsSinceLastPromotion  with 3 lines for mean, median and me
an+std
> plot(attr$YearsSinceLastPromotion, xlab = "")
> abline(h = mean(attr$YearsSinceLastPromotion, na.rm = T), lty = 1)
> abline(h = mean(attr$YearsSinceLastPromotion, na.rm = T) + sd(attr$YearsSin
ceLastPromotion, 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
```



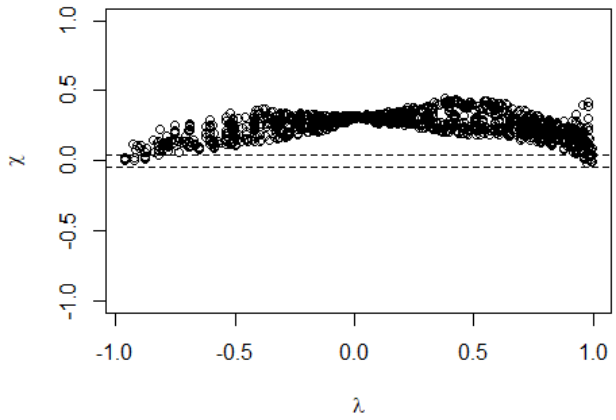

```
> #Boxplot distributions for  YearsWithCurrManager
> boxplot(attr$YearsWithCurrManager, ylab = "YearsWithCurrManager")
> rug(jitter(attr$YearsWithCurrManager), side = 2)
> abline(h = mean(attr$YearsWithCurrManager, na.rm = T), lty = 2)
```



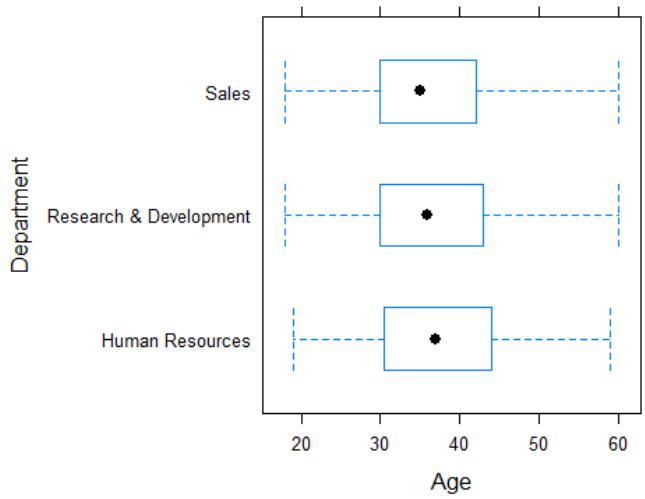

```
> #Boxplot distributions for  YearsWithCurrManager
> plot(attr$YearsWithCurrManager, xlab = "")
> abline(h = mean(attr$YearsWithCurrManager, na.rm = T), lty = 1)
> abline(h = mean(attr$YearsWithCurrManager, na.rm = T) + sd(attr$YearsWithCu
rrManager, na.rm = T),lty = 2)
> abline(h = median(attr$YearsWithCurrManager,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



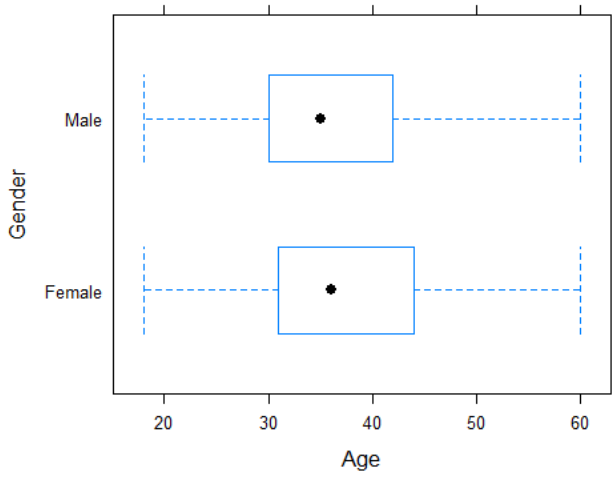

```
> #Chi Plot for inspecting the independence
> chi.plot(attr$MonthlyIncome,attr$Age)
```



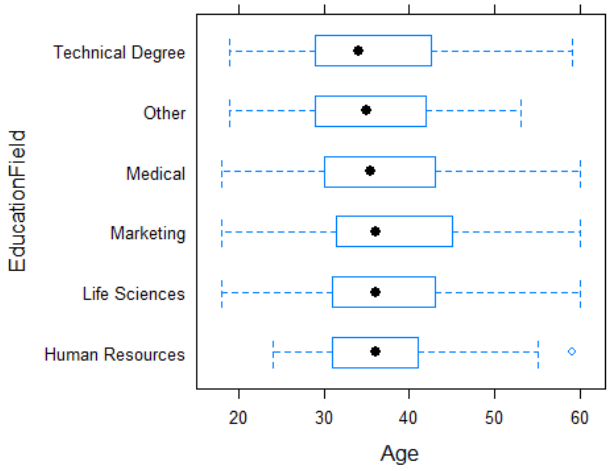

```
> #Plotting joint boxplots for various categories wrt Age
> bwplot(attr$Department ~ attr$Age, data=attr, ylab='Department',xlab='Age')
```



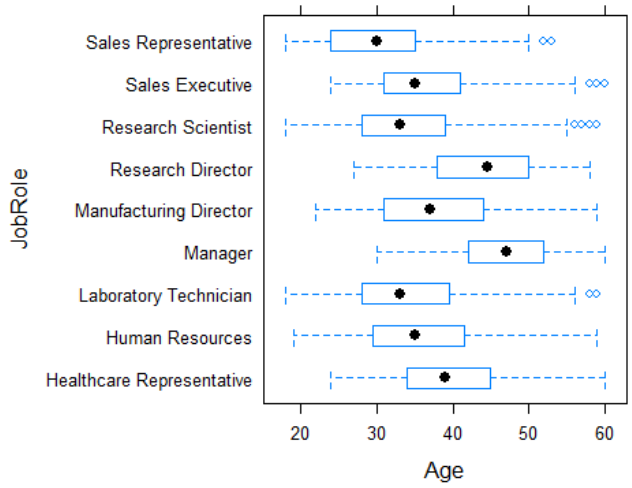

```
> bwplot(attr$Gender ~ attr$Age, data=attr, ylab='Gender',xlab='Age')
```



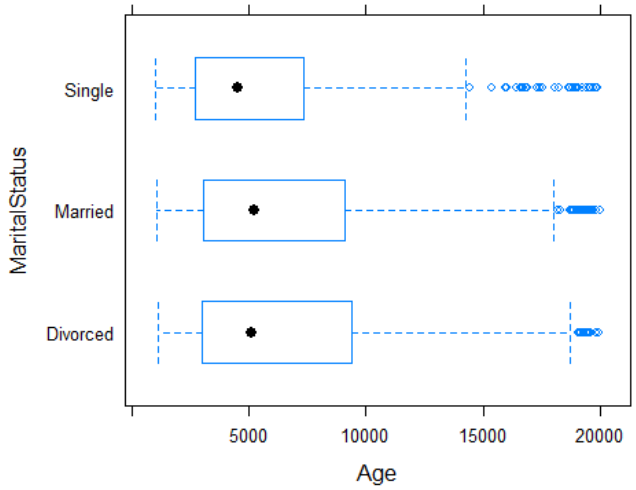

```
> bwplot(attr$EducationField ~ attr$Age, data=attr, ylab='EducationField',xla
b='Age')
```



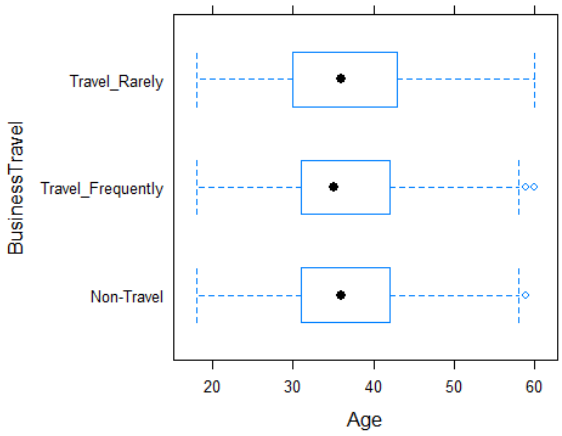
```
> bwplot(attr$JobRole ~ attr$Age, data=attr, ylab='JobRole',xlab='Age')
```



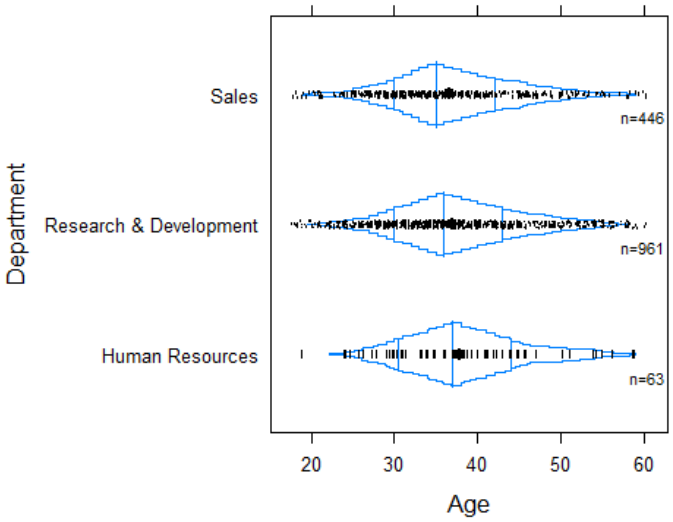

```
> bwplot(attr$MaritalStatus ~ attr$MonthlyIncome, data=attr, ylab='MaritalSta
tus',xlab='Age')
```



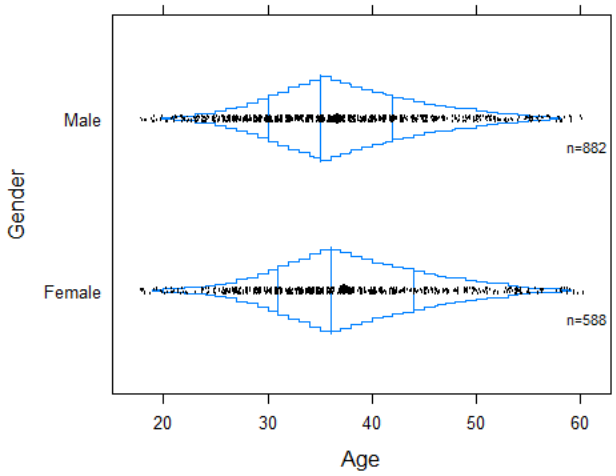

```
> bwplot(attr$BusinessTravel ~ attr$Age, data=attr, ylab='BusinessTravel',xla
b='Age')
```



```
#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')
```



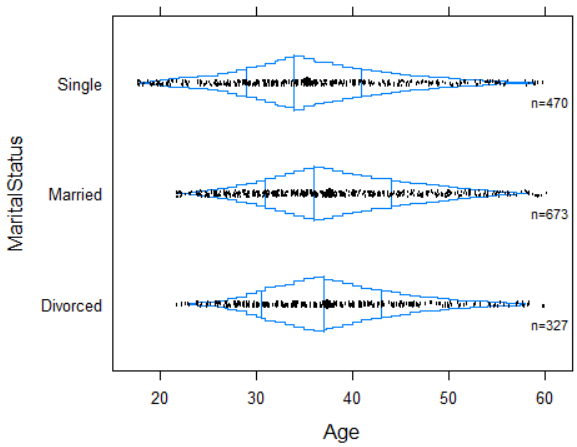

```
> bwplot(attr$Gender ~ attr$Age, data=attr,panel=panel.bpplot,
+        probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='Gender',xlab='Age')
```



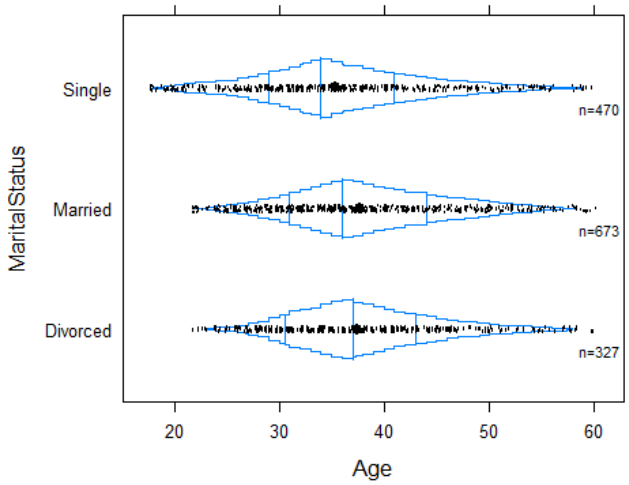

```
> bwplot(attr$EducationField ~ attr$Age, data=attr,panel=panel.bpplot,
+        probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='EducationField',xlab='Age')
```


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

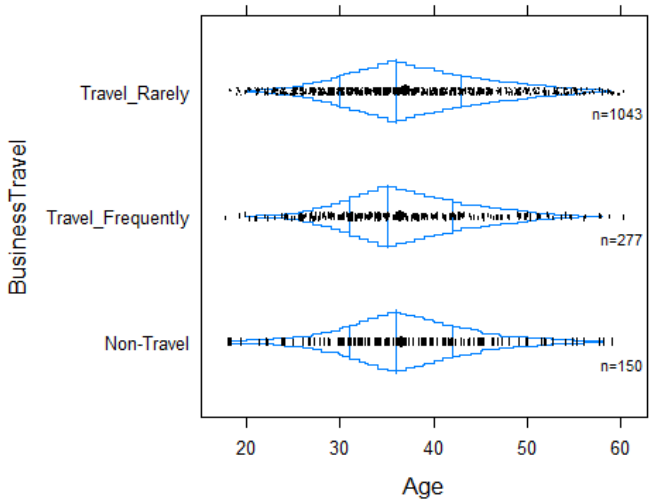

```
> bwplot(attr$MartialStatus ~ attr$Age, data=attr,panel=panel.bpplot,
+        probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='MartialStatus',xlab='Age')
```



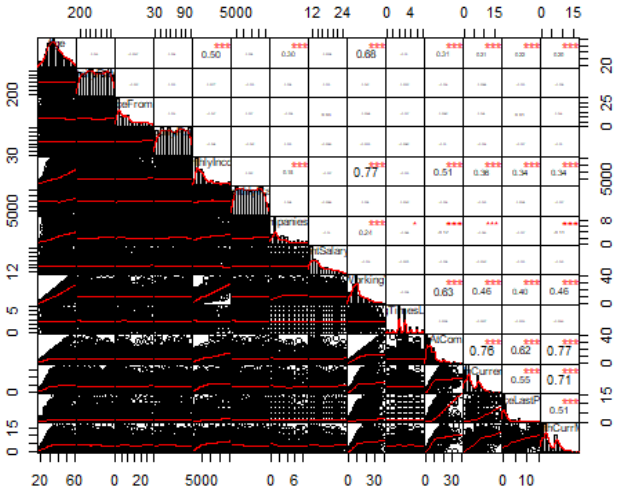
```
> bwplot(attr$MaritalStatus ~ attr$Age, data=attr,panel=panel.bpplot,
+        probs=seq(.01,.49,by=.01), datadensity=TRUE, ylab='MaritalStatus',xlab='Age')
```




```
> 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)
```

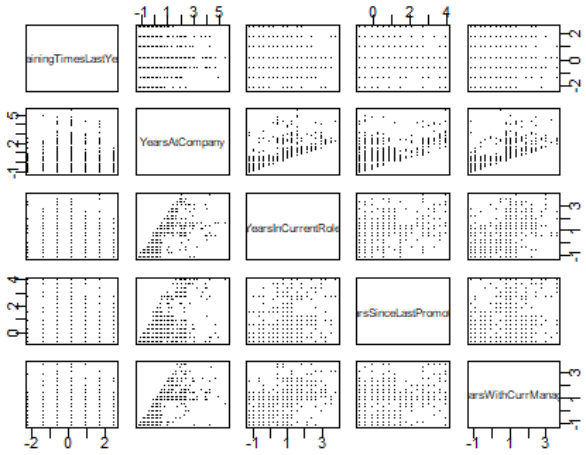


```
#-------------------------------------------------------------------------------------------
> ##Creating Temporary Variables
> #-------------------------------------------------------------------------------------------
>
> #Converting double/int columns to numeric
> numeric_col <- c("Age","DailyRate","DistanceFromHome","HourlyRate",
+                  "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, 0.155654
$ 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,
```

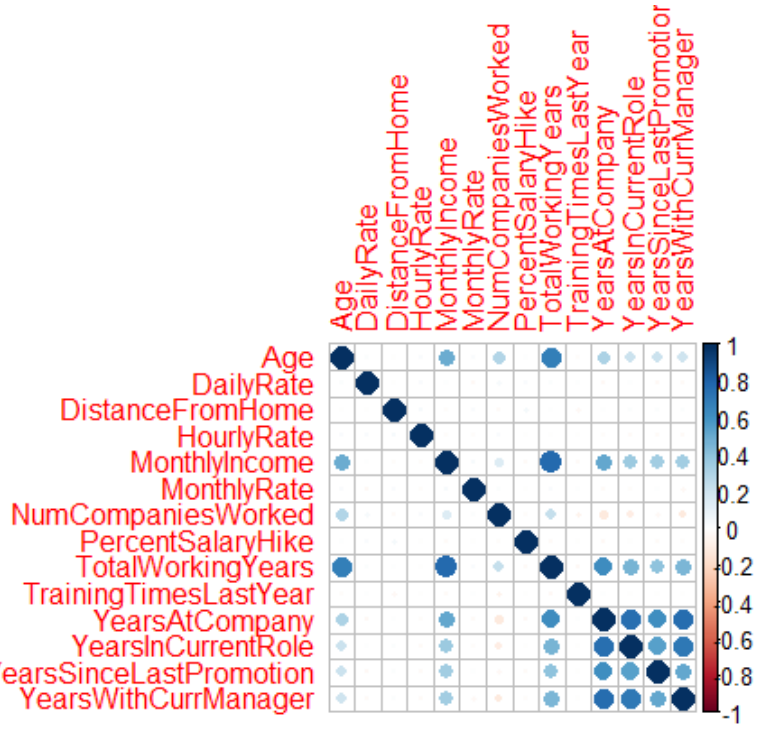
```
##Matrix Plots, Covariance and Corelations Plots
#ScatterPlot matrix
pairs(attr_final[,10:14],pch=".",cex=1.5)
```




```
##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

> #PCA
> #plot.new(); dev.off()
> #Considering the numeric columns that will help to get variance in data
> attr_pca <- attr[,numvar]
> #Plotting correlation plot to understand the how feature are related to each other
> correplot<-cor(attr_pca)
> corrplot(correplot,method="circle")
```




```
> #Finding the principal components of data
> attr_pca <- prcomp(attr_pca,scale=TRUE)
> attr_pca
Standard deviations (1, .., p=14):
 [1] 2.0041891 1.2846031 1.0336848 1.0257290 1.0046944 0.9962218 0.9763963 0.9578606 0.84
[11] 0.6854527 0.5321872 0.4397545 0.3742537

Rotation (n x k) = (14 x 14):
                                PC1          PC2          PC3          PC4          PC5
Age                       0.280157344 -0.472170158  0.003362193  0.004488409 -0.039563410
DailyRate                -0.006815197 -0.077962430 -0.207301367 -0.609569867 -0.211568990
DistanceFromHome          0.004812032  0.041564987 -0.664884791  0.306131593  0.048941659
HourlyRate               -0.011288550 -0.062668026 -0.352147686 -0.255816205  0.602292088
MonthlyIncome             0.360622909 -0.290395305  0.052415102  0.025332267 -0.034941693
MonthlyRate               0.001123298 -0.086158010  0.020312197  0.664085954 -0.101166486
NumCompaniesWorked        0.030991906 -0.560133264  0.005628265 -0.041875610  0.017785645
PercentSalaryHike        -0.015351368  0.004618486 -0.465841883 -0.055689609 -0.698726672
TotalWorkingYears         0.415285665 -0.318115831  0.009368263  0.007027664 -0.024159198
TrainingTimesLastYear    -0.010993402  0.092457674  0.409028173 -0.138279489 -0.293982017
YearsAtCompany            0.443443529  0.213079968  0.002115638 -0.010571214  0.024921329
YearsInCurrentRole        0.391353065  0.279423881 -0.048111956 -0.038785223 -0.004927194
YearsSinceLastPromotion   0.344322397  0.198658357  0.003993040  0.027659809  0.019935007
YearsWithCurrManager      0.386171187  0.295138965 -0.031745944 -0.034459502  0.021898300
                                PC7          PC8          PC9          PC10         PC11
Age                       -0.098196914 -0.05927715 -0.183114693  0.005033984 -0.74367068 -0
DailyRate                  0.715405171 -0.02770642 -0.028707475  0.040304455 -0.01980752  0
DistanceFromHome           0.031447533 -0.65217193  0.037737577  0.002338630  0.02927699  0
HourlyRate                -0.221010405  0.40142111 -0.004675476  0.018009772  0.03725997 -0
MonthlyIncome             -0.012272736 -0.03685912 -0.377381332  0.104651321  0.61775910 -0
MonthlyRate                0.482943083  0.40448871  0.056690883 -0.044889268 -0.01681584  0
NumCompaniesWorked        -0.032989593 -0.03355765  0.775796629 -0.129586743  0.19687866  0
PercentSalaryHike         -0.376210309  0.38335261  0.012190972  0.019568502  0.04082482  0
TotalWorkingYears         -0.029511945 -0.04398227 -0.196663458 -0.038585533  0.08301594  0
TrainingTimesLastYear     -0.217564575 -0.29622601  0.130785998 -0.017811234  0.02954240 -0
YearsAtCompany             0.005335572  0.01862614 -0.001551392 -0.104225054  0.03762072  0
YearsInCurrentRole         0.062086964  0.05420752  0.201595025 -0.271683842 -0.04458982 -0
YearsSinceLastPromotion    0.022129234  0.03850513  0.306725567  0.845951303 -0.08053456  0
```

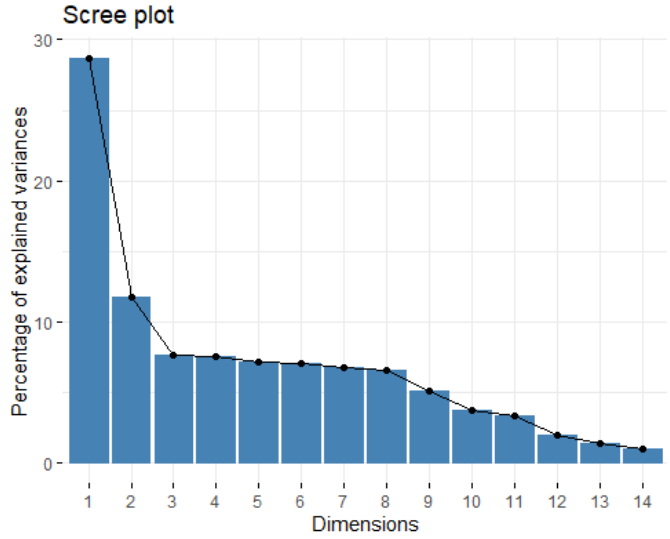
```
YearsWithCurrManager     0.011525930  0.04176204  0.161413516 -0.407140185 -0.06576770  0
                               PC13          PC14
Age                       0.1893016403  0.237072230
DailyRate                -0.0099081253  0.018837870
DistanceFromHome          0.0057709912  0.011991386
HourlyRate                0.0049037331 -0.003314012
MonthlyIncome             0.4041044054  0.279006032
MonthlyRate              -0.0083381972  0.009367738
NumCompaniesWorked       -0.0369554196  0.107234673
PercentSalaryHike        -0.0183445446  0.010585240
TotalWorkingYears        -0.4158888971 -0.705989494
TrainingTimesLastYear     0.0008431661 -0.012085362
YearsAtCompany           -0.6494086309  0.562584645
YearsInCurrentRole        0.1948014145 -0.130643692
YearsSinceLastPromotion   0.0977011821 -0.083447864
YearsWithCurrManager      0.3959144832 -0.121010943
> names(attr_pca)
[1] "sdev"     "rotation" "center"   "scale"    "x"
> head(attr_pca)
$sdev
 [1] 2.0041891 1.2846031 1.0336848 1.0257290 1.0046944 0.9962218 0.9763963 0.9578606 0.84
[11] 0.6854527 0.5321872 0.4397545 0.3742537

$rotation
                                 PC1           PC2           PC3           PC4           PC5
Age                       0.280157344 -0.472170158  0.003362193  0.004488409 -0.039563410
DailyRate                -0.006815197 -0.077962430 -0.207301367 -0.609569867 -0.211568990
DistanceFromHome          0.004812032  0.041564987 -0.664884791  0.306131593  0.048941659
HourlyRate               -0.011288550 -0.062668026 -0.352147686 -0.255816205  0.602292088
MonthlyIncome             0.360622909 -0.290395305  0.052415102  0.025332267 -0.034941693
MonthlyRate               0.001123298 -0.086158010  0.020312197  0.664085954 -0.101166486
NumCompaniesWorked        0.030991906 -0.560133264  0.005628265 -0.041875610  0.017785645
PercentSalaryHike        -0.015351368  0.004618486 -0.465841883 -0.055689609 -0.698726672
TotalWorkingYears         0.415285665 -0.318115831  0.009368263  0.007027664 -0.024159198
TrainingTimesLastYear    -0.010993402  0.092457674  0.409028173 -0.138279489 -0.293982017
YearsAtCompany            0.443443529  0.213079968  0.002115638 -0.010571214  0.024921329
YearsInCurrentRole        0.391353065  0.279423881 -0.048111956 -0.038785223 -0.004927194
YearsSinceLastPromotion   0.344322397  0.198658357  0.003993040  0.027659809  0.019935007
YearsWithCurrManager      0.386171187  0.295138965 -0.031745944 -0.034459502  0.021898300
                                 PC7           PC8           PC9          PC10          PC11
Age                      -0.098196914 -0.05927715 -0.183114693  0.005033984 -0.74367068 -0
DailyRate                 0.715405171 -0.02770642 -0.028707475  0.040304455 -0.01980752  0
DistanceFromHome          0.031447533 -0.65217193  0.037737577  0.002338630  0.02927699  0
HourlyRate               -0.221010405  0.40142111 -0.004675476  0.018009772  0.03725997 -0
MonthlyIncome            -0.012272736 -0.03685912 -0.377381332  0.104651321  0.61775910 -0
MonthlyRate               0.482943083  0.40448871  0.056690883 -0.044889268 -0.01681584  0
NumCompaniesWorked       -0.032989593 -0.03355765  0.775796629 -0.129586743  0.19687866  0
PercentSalaryHike        -0.376210309  0.38335261  0.012190972  0.019568502  0.04082482  0
TotalWorkingYears        -0.029511945 -0.04398227 -0.196663458 -0.038585533  0.08301594  0
TrainingTimesLastYear    -0.217564575 -0.29622601  0.130785998 -0.017811234  0.02954240 -0
YearsAtCompany            0.005335572  0.01862614 -0.001551392 -0.104225054  0.03762072  0
YearsInCurrentRole        0.062086964  0.05420752  0.201595025 -0.271683842 -0.04458982 -0
YearsSinceLastPromotion   0.022129234  0.03850513  0.306725567  0.845951303 -0.08053456  0
YearsWithCurrManager      0.011525930  0.04176204  0.161413516 -0.407140185 -0.06576770  0
                               PC13          PC14
Age                       0.1893016403  0.237072230
DailyRate                -0.0099081253  0.018837870
DistanceFromHome          0.0057709912  0.011991386
HourlyRate                0.0049037331 -0.003314012
MonthlyIncome             0.4041044054  0.279006032
MonthlyRate              -0.0083381972  0.009367738
NumCompaniesWorked       -0.0369554196  0.107234673
PercentSalaryHike        -0.0183445446  0.010585240
TotalWorkingYears        -0.4158888971 -0.705989494
```

```
TrainingTimesLastYear    0.0008431661 -0.012085362
YearsAtCompany          -0.6494086309  0.562584645
YearsInCurrentRole       0.1948014145 -0.130643692
YearsSinceLastPromotion  0.0977011821 -0.083447864
YearsWithCurrManager     0.3959144832 -0.121010943

$center
                     Age               DailyRate        DistanceFromHome                Hour
               36.923810              802.485714                9.192517                65.
           MonthlyIncome             MonthlyRate      NumCompaniesWorked       PercentSala
             6502.931293           14313.103401                2.693197                15.
      TotalWorkingYears   TrainingTimesLastYear         YearsAtCompany      YearsInCurre
              11.279592                2.799320                7.008163                 4.
YearsSinceLastPromotion    YearsWithCurrManager
               2.187755                4.123129

$scale
                     Age               DailyRate        DistanceFromHome                Hour
               9.135373              403.509100                8.106864                20.
           MonthlyIncome             MonthlyRate      NumCompaniesWorked       PercentSala
             4707.956783           7117.786044                2.498009                 3.
      TotalWorkingYears   TrainingTimesLastYear         YearsAtCompany      YearsInCurre
               7.780782                1.289271                6.126525                 3.
YearsSinceLastPromotion    YearsWithCurrManager
               3.222430                3.568136

$x
                PC1          PC2           PC3          PC4           PC5           PC6
   [1,] -0.243044079 -1.805296680 -0.3104169553 -0.384994275  2.0169575215  0.984793983
   [2,]  0.854288163  0.381073595 -0.5252390003  1.608645799 -1.5927436292 -0.481865695 -
   [3,] -2.198275394 -1.413301606 -0.0582282745 -2.596801032  0.6022422032 -0.023230640 -
   [4,] -0.570845554  0.682294015  0.9530781591 -0.112517223  0.0144209522 -0.415235118
   [5,] -1.567685771 -0.965438549  1.6427717726  0.537047071 -0.0703157168  0.940560119
   [6,] -0.033228256  1.476192828  0.1768758528 -0.871158691  0.9282306482  0.367849216
   [7,] -1.176981984 -2.344142438 -0.5229233979 -1.666338771 -0.8793271137 -0.417767837 -
   [8,] -2.668563284  0.310022541 -2.6128166708 -0.330963952 -1.2486460778 -0.052250851
   [9,]  0.910894519  1.213005861 -1.5039201417  1.144947290 -1.1238829367  1.049881443 -
  [10,]  1.339778839 -0.208584655 -1.9076328247 -0.281672326  1.0866328829 -1.252674274
  [11,] -1.227202504  0.961999888  0.0613545738  0.034499533  0.4709461858 -2.043427871 -
  [12,] -0.073116546  1.760186112  0.5344916830  1.247906142  0.5189044912  0.442256861 -
  [13,] -1.087474516  1.023377864 -1.5147071841  1.573775407 -0.7475549471  1.511464717
  [14,] -1.814073316  0.601874393 -1.3014136722 -1.101977678  1.6522331947 -0.334506872
  [15,] -1.659410211  0.292702788 -0.0672556843  1.527211158  0.0369416431 -0.176183833 -
  [16,]  1.744251934  1.686663324 -1.1177364895 -0.381693162  0.6609099755  1.223658522
  [17,] -1.118546968  1.132058901  1.4189672811  0.273681898  0.7687430566 -1.460717606 -
  [18,] -2.875446285  0.684977918 -1.2101773533 -1.069482747  1.5219403902 -0.105065952
  [19,]  4.306398521 -1.060188116  0.2102166478 -0.374474127 -0.3205736368 -1.006842601
  [20,] -1.274793286 -0.518252552  1.7607337154 -0.263605442  0.4775025923  1.296215086 -
  [21,] -1.707674413  1.416074395 -0.2866612249 -0.911443288  0.0740619531 -1.694982203 -
  [22,] -0.973935456 -1.013413533 -1.1120212000 -1.850388185 -1.3640205357 -0.666890243 -
  [23,]  1.686994016  1.312303531  1.5096245660  1.224455487  0.2429059015 -0.652215366
  [24,] -3.186960204  1.127496005  0.4028852195  0.661084708  0.6427677466 -2.782189999 -
  [25,] -1.236420346  0.070998887  0.2998566707  0.248336464  1.5088956715  0.030328006
  [26,]  4.335210193 -1.187855044  0.8439064948 -1.135518002  0.1502712115  0.089980426
  [27,]  0.202667449  1.487707345 -1.0496066009 -1.510499974 -1.5194444667 -1.172856717 -
  [28,]  0.522611662  0.517960123  0.7496484428  1.189509235  0.3725504816  0.467749272
  [29,]  4.167497779  1.068632778  1.1917707595 -0.655936499 -0.3603395503  0.680646739 -
  [30,]  0.855668649 -2.594944348  0.7337329455  0.725830150  0.9881939586 -0.389610046
  [31,] -2.101453288 -0.598109160  0.9218830885 -1.261397525  1.1285699242  0.185406411 -
  [32,] -0.628686779 -0.506577343  1.0557480273 -0.359363863 -1.2767722294 -1.259019980
  [33,] -0.585760317  1.290594877  1.0175278792  0.759970743  0.8434796204 -1.537202603 -
  [34,] -1.453000790 -0.878346751  1.6553451372 -1.466979255 -0.7652620288 -1.050427892 -
  [35,] -2.295426038  0.497391245  0.3961407175 -1.173667681  0.0495687553  1.396660954 -
  [36,] -0.832994065  0.042990012  0.7006144001 -0.308495224  0.3417332881 -0.722274641
```

```
[37,] -1.354693336 -0.382172652 -0.0113330121 -1.423881781  1.0314915947  0.517495020 -
[38,] -1.788459375  0.520002398  0.3186189447 -1.168483914  1.3010327250 -0.499080505 -
[39,] -1.940888185 -1.947119717  0.3014065269 -0.616302576  0.6580712905 -0.112112477 -
[40,] -0.826827035  0.269317986  0.4683770185 -1.617065750 -1.5424619460  1.078602496 -
[41,] -2.540242339  0.231335668  0.9249250633 -0.019213294  1.0357898965  0.023182442 -
[42,] -2.775064596  0.699871648  2.2688244419 -0.255818018 -1.5958498425 -1.351223462
[43,] -2.660977369  0.709039810 -1.1151323620 -0.173311857  0.1700305471  0.580484544
[44,]  0.494241988  1.355663630 -0.4558314906 -0.773576913 -0.0893201999  2.889936395
[45,]  0.755465347  1.607572626 -0.5109637132 -0.506888091 -1.4233075836  1.057147451 -
[46,]  5.804951893  1.078091201 -0.6975281151  0.051444706  0.4337848886  1.614106696
[47,]  0.806658069  1.653442819 -2.3068106160 -0.295882847 -0.4722918379  0.134169049 -
[48,] -1.995881882 -0.926896408 -1.9960316143  0.633599363 -0.1729888148  1.021151201 -
[49,]  0.844245935 -0.616849916 -0.8194206960 -0.921810229 -0.7717068286 -1.818536061 -
[50,] -2.418330657  0.297326189 -0.3513459999 -1.101751982 -1.5156578641  1.499053343
[51,] -0.617002088 -3.634989196  0.3562096372  0.118216390  1.4389894618 -0.220229442 -
[52,] -1.874898008  0.830881805  0.6248015543 -1.142134976 -0.3600199657  0.329946720
[53,] -0.595964470 -1.218466650 -0.8408839570 -2.447215614 -0.9399446833  0.774272087 -
[54,] -0.697653611 -0.440311397 -0.1001609790 -1.002941201  0.6741005136 -0.305412825
[55,] -2.079810230 -0.955185144 -1.8161177942  0.441813227 -1.3960822450  0.253432093
[56,]  3.708466615  1.814109184 -0.0097541922 -0.036909451  2.1090226075  0.552998695
[57,]  0.809848967  1.103897503 -1.7108414088 -0.291327045 -1.1137226784 -0.318923581 -
[58,] -1.465681948  0.058005177 -0.5799431432  0.624721643 -1.1676958094  0.275086845
[59,]  0.172118882  0.857390902 -0.3222266947 -0.286150650 -1.6974104391  0.722818482 -
[60,] -0.134488687  0.774824505  0.0742047436 -1.668328387 -0.8320217703  1.458220908 -
[61,]  1.154466224  1.922306493  0.5340354541  0.208456672 -2.0449281318  1.172737829 -
[62,]  1.015549604  1.685679343 -1.0836803069  0.522264077  0.9425346951  0.960676995
[63,]  5.230825960 -1.151655660  0.8119758029  0.364485427 -0.0405314789  0.826552609
[64,]  5.030282900 -0.755218265 -1.7678975054 -2.044562287  1.6478909242 -0.737221001
[65,]  3.788984539  1.946298043 -0.0945560272 -1.067010987  0.3184679516  0.860439229
[66,]  0.660742882 -2.427444222  0.7464731633  1.051907583 -0.8666771968  0.700339962
[67,] -0.689313482  0.194433528 -0.4992692899 -0.091816631  0.9077818943 -1.467101142
[68,] -0.360119052 -2.235551626 -0.3033093916 -0.208683127 -0.8858009134  0.057772762
[69,] -1.590660053 -0.244778958  0.5118655703 -0.944308747  1.1711060680  0.909011154 -
[70,] -2.349000231 -0.024496709 -1.0212757577  1.625278907  0.8108621115  0.977055716 -
[71,]  0.537061435 -3.049510452  0.9612799574  0.240109595  0.1951466351  0.179765350
                PC8           PC9           PC10          PC11          PC12           PC
 [1,]  1.5917738507  1.1863471823 -0.8721216003 -0.0964053395  0.1652689774  2.697995e-
 [2,]  1.4396138690 -0.2915207036 -0.9133153760 -1.2704956060 -0.1223848585  3.003787e-
 [3,]  0.1645450520  0.7037177946  0.2334286037 -0.2063109991  0.0757180104 -2.034191e-
 [4,]  0.3788485663 -0.0232142896  0.4483748666 -0.4255970757 -1.1905425766 -5.888785e-
 [5,] -0.2110090034  2.3132854086 -0.0092829889  0.7911383832  0.0932147601 -1.001822e-
 [6,]  0.8232618244 -0.2219554949 -0.0872069258 -0.4352547853 -0.2311142456  2.074828e-
 [7,]  0.6783978017 -0.4439152914  0.2850426885 -1.9955214056  0.0476998780 -7.608564e-
 [8,] -0.3684730529 -0.4875080936  0.4800518063  0.0543954145  0.1020736714 -4.659415e-
 [9,] -0.9351740522 -0.8467583073 -0.7791175369  0.1173708729 -0.0207647030  7.178969e-
[10,] -0.9887551265  1.8168832792  0.5496258442  0.0893646974  0.1719743118  1.482873e-
[11,] -0.7428933771 -0.3487121205 -0.3578297611 -0.5029560805 -0.2422407358 -3.914638e-
[12,] -1.0813690872 -0.3767620037 -1.1058405989  0.0965394361  0.4399899746 -2.594555e-
[13,] -1.2884619372 -0.0363538480  0.8308304610 -0.2190499489  0.3017206595 -5.138678e-
[14,] -0.8195524226 -0.7310455341  0.3812739185 -0.4949317648 -0.0110627441  2.587105e-
[15,] -1.9271549894  1.2343495139 -0.5198251249  0.4250203594  0.1772497877 -2.805626e-
[16,] -1.2962378063  0.1672283742  0.9190505173  0.6052322133 -0.1396419825  7.886224e-
[17,] -0.0866130578 -0.4359922773 -0.4908235512 -0.1712840455  0.4964863948 -6.197885e-
[18,] -0.3703968876 -0.4515944960  0.4707221281  0.6870011577  0.0480464742 -1.323742e-
[19,]  1.1419762947 -1.3201148540 -0.5553725566 -0.0202145544  0.0106513546 -1.330532e+
[20,] -0.9326540622  0.6299363640 -0.0098247146 -0.2880597580 -0.0739027515  1.175321e-
[21,]  0.0042898132 -0.3112305780  0.1801517860  0.6691640181  0.1912063602 -6.152712e-
[22,]  0.3446067836  1.3648622584 -0.4792426584  0.2130812121  0.0911878469 -3.605121e-
[23,] -0.2496196395 -0.6875524464 -0.9335162824  0.6027692112  0.8094891808  7.005028e-
[24,] -0.3520700093  0.2822430507  0.2555611279  0.6350192662  0.0179911182 -1.389628e-
[25,]  0.5077089246 -0.1612550579 -0.0186130089 -0.3041024579  0.2026623054 -1.190615e-
[26,] -0.6259760526 -0.5375940379 -0.5583140528  0.3347066400 -1.0951792994  8.320370e-
[27,] -0.6537874675  0.3570074959  0.9290619928 -0.0189461399  1.1560600116 -2.554831e-
[28,] -0.0648482848 -0.7354981588  0.5503138531 -0.7380133792 -0.9496643546  8.389053e-
```

```
 [29,] -1.2648397307  0.3155536872 -1.1110140575 -0.1549320265  2.1008653026 -1.743973e-
 [30,]  0.9457428880 -1.6915126800  0.7155359778  1.0274567321 -0.2257845320  7.444228e-
 [31,] -0.1895100955  0.2414887412  0.1464112473 -0.0495438258 -0.1690167656 -2.514129e-
 [32,] -1.1586344712 -0.3666531671  0.0562237019 -0.6611355034  0.2298295473  2.991824e-
 [33,] -0.1491699406  0.0895757307 -0.5992509197 -0.0332054293  1.5600371817 -5.381913e-
 [34,] -1.5329294741 -0.2065832996  0.1820633909 -0.4482305144  0.1077454163 -8.542825e-
 [35,]  0.2134201607 -0.1299260602  0.8657952064  0.4687972163  0.1202434785 -4.949487e-
 [36,]  0.7203396398 -0.3895527430 -0.1574012260 -1.2429413892  0.1954059923  1.912162e-
 [37,]  0.2688390818 -0.9067448813  0.0479551510 -1.6866316116 -0.1441330064  4.304295e-
 [38,]  0.6935995880 -0.2003120841  0.5271573093 -0.6067308937 -0.0279386513  2.824748e-
 [39,]  0.2429085978  1.7646246196 -0.1129573297  0.2507919469 -0.1314289782 -1.482093e-
 [40,] -0.1179524881 -0.3580190226  0.0556865761  0.1396045303  0.0353093528 -1.389953e-
 [41,]  0.0181823818 -0.5107722810  0.3263458397 -0.5465772678 -0.0602546854  5.658773e-
 [42,] -0.7388494377 -0.0587506994  0.2787045616  0.0926434197  0.0901210156 -1.196858e-
 [43,] -1.9884712967 -0.3763602810  0.3022550599  0.1729522027  0.2552767122 -3.327794e-
 [44,] -0.3462124514 -0.6086221030 -0.7113741266  0.7779144715 -0.3180020334  4.068577e-
 [45,]  1.5058488466  0.0966839023 -0.4001732785  0.0586933619 -0.1521141805 -3.945470e-
 [46,]  0.0542470485 -0.2605203348  2.2309102556  0.7861699813 -1.1179707031  4.072934e-
 [47,] -0.3676769926  0.0803763634  1.3065788470 -0.3439303856  0.5396748825  1.926447e-
 [48,]  0.0379981462  0.0139823408  0.2379323987 -0.1482978143  0.0345722067 -2.626249e-
 [49,]  1.5880220441  0.4128727770 -1.2321509722 -0.6278155609  0.3533099316  1.678164e-
 [50,] -0.4960323101 -0.6439180110  0.3515209761 -0.5484838822  0.2054958400  1.408071e-
 [51,]  1.1704670085  0.8200564211 -0.1123759892 -0.2977384011  0.1285674151 -6.899389e-
 [52,] -0.3941110717 -0.1764307635  0.5591763859  0.0448168714  0.0596852002  2.356005e-
 [53,]  0.8857772453  0.2519075061  0.0130817786 -0.4850733698  0.0278447233  1.985134e-
 [54,] -0.6111756017 -0.7690695559  0.6284517406  0.6011938058  0.6113545988  3.056229e-
 [55,] -0.5984018598  1.3970676765 -0.0643320469  0.9721652618 -0.1949743264 -3.069309e-
 [56,]  1.7422401316  0.1408058227 -0.0199461188  0.7410413537 -0.5297862525  1.093202e+
 [57,] -0.0108146406 -0.3106679951 -0.8493487541  0.3863867848 -0.1458192162  6.892529e-
 [58,] -1.8215348214  0.3621471351  0.4006851921 -0.2172714122  0.0608335537  2.928318e-
 [59,]  0.3335120551  0.4740138563 -0.8411057977  0.4163000283 -0.0858083206  2.931242e-
 [60,]  0.2284006406 -0.6348604010 -0.7475512841 -0.2920548376  0.2496513450  4.888034e-
 [61,]  0.7408606053  0.5122781140  0.5793496320 -0.0105572994 -0.0042066309  5.488545e-
 [62,] -2.4733087118  0.5540510535  1.2979592812 -0.9935329529  1.2766076994  1.734229e-
 [63,] -0.4001018830  0.0656254162  2.1963468297  0.5895255433  1.4470277217 -1.086096e+
 [64,] -1.7992385929  1.6436332545 -0.5355851584 -1.3269684512 -1.2755865234 -5.269145e-
 [65,] -0.1566109309  0.5395449026  1.4430413025  0.0325762440 -1.0121301749  2.073019e-
 [66,] -0.5389467786 -2.0662901268  0.1446763110 -0.3168046990  0.3722514197  2.468134e-
 [67,]  0.6267828319 -0.5362398552 -0.3750836045 -0.0545365082 -0.4224485744  2.452230e-
 [68,]  0.3270971145 -1.6969908177  0.3987065997 -0.0530068030  0.1509487886 -4.247686e-
 [69,]  0.3792991801  0.4293974391  0.0802128266 -0.3229283966 -0.0234897322 -3.881229e-
 [70,]  1.5255487349 -1.1502222404  0.3958031962 -0.5136583440 -0.0367616655  1.303524e-
 [71,]  0.4884620531  0.4469007175 -0.3625035208 -1.6091056814  0.0695982865 -7.853735e-
 [ reached getOption("max.print") -- omitted 1399 rows ]

> summary(attr_pca)
Importance of components:
                          PC1    PC2     PC3     PC4    PC5     PC6    PC7     PC8    PC9
Standard deviation     2.0042 1.2846 1.03368 1.02573 1.0047 0.99622 0.9764 0.95786 0.8500
Proportion of Variance 0.2869 0.1179 0.07632 0.07515 0.0721 0.07089 0.0681 0.06554 0.0516
Cumulative Proportion  0.2869 0.4048 0.48111 0.55626 0.6284 0.69925 0.7673 0.83288 0.8845
                          PC12    PC13   PC14
Standard deviation     0.53219 0.43975 0.3743
Proportion of Variance 0.02023 0.01381 0.0100
Cumulative Proportion  0.97618 0.99000 1.0000
> #Extract variance against features
> eigenvalues<-attr_pca$sdev^2
> eigenvalues
 [1] 4.0167738 1.6502052 1.0685042 1.0521201 1.0094108 0.9924579 0.9533497 0.9174969 0.72
[11] 0.4698454 0.2832233 0.1933840 0.1400658
> sum(eigenvalues)
[1] 14
> names(eigenvalues) <- paste("PC",1:14,sep="")
> eigenvalues
```

```
      PC1       PC2       PC3       PC4       PC5       PC6       PC7       PC8       PC9
4.0167738 1.6502052 1.0685042 1.0521201 1.0094108 0.9924579 0.9533497 0.9174969 0.7224654
     PC12      PC13      PC14
0.2832233 0.1933840 0.1400658
> sumoflambdas <- sum(eigenvalues)
> sumoflambdas
[1] 14
> #Variance %
> pctvar<- (eigenvalues/sumoflambdas)*100
> pctvar
      PC1       PC2       PC3       PC4       PC5       PC6       PC7       PC8       PC9
28.691241 11.787180  7.632173  7.515143  7.210077  7.088985  6.809641  6.553550  5.160467
     PC12      PC13      PC14
 2.023023  1.381314  1.000470
> #Calculate cumulative of variance
> cumvar <- cumsum(pctvar)
> cumvar
      PC1       PC2       PC3       PC4       PC5       PC6       PC7       PC8       PC9
 28.69124  40.47842  48.11059  55.62574  62.83581  69.92480  76.73444  83.28799  88.44846
     PC12      PC13      PC14
 97.61822  98.99953 100.00000

> #Visualize PCA using Scree plot
> fviz_screeplot(attr_pca, ncp=14)
```




```
> summary(attr_pca)
Importance of components:
                          PC1    PC2     PC3     PC4    PC5     PC6    PC7     PC8    PC9
Standard deviation     2.0042 1.2846 1.03368 1.02573 1.0047 0.99622 0.9764 0.95786 0.8500
Proportion of Variance 0.2869 0.1179 0.07632 0.07515 0.0721 0.07089 0.0681 0.06554 0.0516
Cumulative Proportion  0.2869 0.4048 0.48111 0.55626 0.6284 0.69925 0.7673 0.83288 0.8845
                          PC12    PC13   PC14
Standard deviation     0.53219 0.43975 0.3743
Proportion of Variance 0.02023 0.01381 0.0100
Cumulative Proportion  0.97618 0.99000 1.0000
```