Reading the dataset and visualizing its length and breadth

```
employees.df <-read.csv(paste("WA_Fn-UseC_-HR-Employee-
Attrition.csv",sep=""))
View(employees.df)
attach(employees.df)
nrow(employees.df)

## [1] 1470

ncol(employees.df)

## [1] 35
```

Descriptive statistics for each variable

```
library(psych)
describe(employees.df)[,c(1:5)]

##                            vars     n       mean       sd   median
## ï..Age                        1 1470      36.92     9.14     36.0
## Attrition*                    2 1470       1.16     0.37      1.0
## BusinessTravel*               3 1470       2.61     0.67      3.0
## DailyRate                     4 1470     802.49   403.51    802.0
## Department*                   5 1470       2.26     0.53      2.0
## DistanceFromHome              6 1470       9.19     8.11      7.0
## Education                     7 1470       2.91     1.02      3.0
## EducationField*               8 1470       3.25     1.33      3.0
## EmployeeCount                 9 1470       1.00     0.00      1.0
## EmployeeNumber               10 1470    1024.87   602.02   1020.5
## EnvironmentSatisfaction      11 1470       2.72     1.09      3.0
## Gender*                      12 1470       1.60     0.49      2.0
## HourlyRate                   13 1470      65.89    20.33     66.0
## JobInvolvement               14 1470       2.73     0.71      3.0
## JobLevel                     15 1470       2.06     1.11      2.0
## JobRole*                     16 1470       5.46     2.46      6.0
## JobSatisfaction              17 1470       2.73     1.10      3.0
## MaritalStatus*               18 1470       2.10     0.73      2.0
## MonthlyIncome                19 1470    6502.93  4707.96   4919.0
## MonthlyRate                  20 1470   14313.10  7117.79  14235.5
## NumCompaniesWorked           21 1470       2.69     2.50      2.0
## Over18*                      22 1470       1.00     0.00      1.0
## OverTime*                    23 1470       1.28     0.45      1.0
## PercentSalaryHike            24 1470      15.21     3.66     14.0
## PerformanceRating            25 1470       3.15     0.36      3.0
## RelationshipSatisfaction     26 1470       2.71     1.08      3.0
## StandardHours                27 1470      80.00     0.00     80.0
## StockOptionLevel             28 1470       0.79     0.85      1.0
## TotalWorkingYears            29 1470      11.28     7.78     10.0
## TrainingTimesLastYear        30 1470       2.80     1.29      3.0
## WorkLifeBalance              31 1470       2.76     0.71      3.0
## YearsAtCompany               32 1470       7.01     6.13      5.0
```

```
## YearsInCurrentRole        33 1470    4.23    3.62    3.0
## YearsSinceLastPromotion   34 1470    2.19    3.22    1.0
## YearsWithCurrManager      35 1470    4.12    3.57    3.0
```

Inspect the datatypes

```
str(employees.df)
```

```
## 'data.frame':    1470 obs. of  35 variables:
##  $ ï..Age                  : int  41 49 37 33 27 32 59 30 38 36 ...
##  $ Attrition               : Factor w/ 2 levels "No","Yes": 2 1 2 1 1 1 1
1 1 1 ...
##  $ BusinessTravel          : Factor w/ 3 levels "Non-
Travel","Travel_Frequently",..: 3 2 3 2 3 2 3 3 2 3 ...
##  $ DailyRate               : int  1102 279 1373 1392 591 1005 1324 1358
216 1299 ...
##  $ Department              : Factor w/ 3 levels "Human Resources",..: 3 2
2 2 2 2 2 2 2 2 ...
##  $ DistanceFromHome        : int  1 8 2 3 2 2 3 24 23 27 ...
##  $ Education               : int  2 1 2 4 1 2 3 1 3 3 ...
##  $ EducationField          : Factor w/ 6 levels "Human Resources",..: 2 2
5 2 4 2 4 2 4 ...
##  $ EmployeeCount           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ EmployeeNumber          : int  1 2 4 5 7 8 10 11 12 13 ...
##  $ EnvironmentSatisfaction : int  2 3 4 4 1 4 3 4 4 3 ...
##  $ Gender                  : Factor w/ 2 levels "Female","Male": 1 2 2 1 2
2 1 2 2 2 ...
##  $ HourlyRate              : int  94 61 92 56 40 79 81 67 44 94 ...
##  $ JobInvolvement          : int  3 2 2 3 3 3 4 3 2 3 ...
##  $ JobLevel                : int  2 2 1 1 1 1 1 1 3 2 ...
##  $ JobRole                 : Factor w/ 9 levels "Healthcare
Representative",..: 8 7 3 7 3 3 3 3 5 1 ...
##  $ JobSatisfaction         : int  4 2 3 3 2 4 1 3 3 3 ...
##  $ MaritalStatus           : Factor w/ 3 levels "Divorced","Married",..: 3
2 3 2 2 3 2 1 3 2 ...
##  $ MonthlyIncome           : int  5993 5130 2090 2909 3468 3068 2670 2693
9526 5237 ...
##  $ MonthlyRate             : int  19479 24907 2396 23159 16632 11864 9964
13335 8787 16577 ...
##  $ NumCompaniesWorked      : int  8 1 6 1 9 0 4 1 0 6 ...
##  $ Over18                  : Factor w/ 1 level "Y": 1 1 1 1 1 1 1 1 1 1
...
##  $ OverTime                : Factor w/ 2 levels "No","Yes": 2 1 2 2 1 1 2
1 1 1 ...
##  $ PercentSalaryHike       : int  11 23 15 11 12 13 20 22 21 13 ...
##  $ PerformanceRating       : int  3 4 3 3 3 3 4 4 4 3 ...
##  $ RelationshipSatisfaction: int  1 4 2 3 4 3 1 2 2 2 ...
##  $ StandardHours           : int  80 80 80 80 80 80 80 80 80 80 ...
##  $ StockOptionLevel        : int  0 1 0 0 1 0 3 1 0 2 ...
##  $ TotalWorkingYears       : int  8 10 7 8 6 8 12 1 10 17 ...
##  $ TrainingTimesLastYear   : int  0 3 3 3 3 2 3 2 2 3 ...
##  $ WorkLifeBalance         : int  1 3 3 3 3 2 2 3 3 2 ...
```

```
##  $ YearsAtCompany        : int  6 10 0 8 2 7 1 1 9 7 ...
##  $ YearsInCurrentRole     : int  4 7 0 7 2 7 0 0 7 7 ...
##  $ YearsSinceLastPromotion : int  0 1 0 3 2 3 0 0 1 7 ...
##  $ YearsWithCurrManager   : int  5 7 0 0 2 6 0 0 8 7 ...
```

One way contigency tables for categorical variables :

Attrition

```
table(Attrition)
```

```
## Attrition
##   No  Yes
## 1233  237
```

Business Travel

```
table(BusinessTravel)
```

```
## BusinessTravel
##      Non-Travel Travel_Frequently      Travel_Rarely
##             150               277               1043
```

Department

```
table(Department)
```

```
## Department
##       Human Resources Research & Development                  Sales
##                    63                    961                    446
```

Education

```
table(Education)
```

```
## Education
##   1   2   3   4   5
## 170 282 572 398  48
```

Eduaction Field

```
table(EducationField)
```

```
## EducationField
##  Human Resources    Life Sciences         Marketing         Medical
##             27             606               159             464
##          Other Technical Degree
##             82             132
```

Enviornment Satisfaction

```
table(EnvironmentSatisfaction)
```

```
## EnvironmentSatisfaction
##   1   2   3   4
## 284 287 453 446
```

Gender

```
table(Gender)
```

```
## Gender
## Female   Male
##    588    882
```

Job Involvement

```
table(JobInvolvement)
```

```
## JobInvolvement
##   1   2   3   4
##  83 375 868 144
```

Job Level

```
table(JobLevel)
```

```
## JobLevel
##   1   2   3   4   5
## 543 534 218 106  69
```

Job Role

```
table(JobRole)
```

```
## JobRole
## Healthcare Representative           Human Resources
##                       131                        52
##     Laboratory Technician                   Manager
##                       259                       102
##     Manufacturing Director         Research Director
##                       145                        80
##        Research Scientist           Sales Executive
##                       292                       326
##       Sales Representative
##                        83
```

Job Satisfaction

```
table(JobSatisfaction)
```

```
## JobSatisfaction
##   1   2   3   4
## 289 280 442 459
```

Marital Status

```
table(MaritalStatus)
```

```
## MaritalStatus
## Divorced  Married   Single
##      327      673      470
```

Overtime

```
table(OverTime)
```

```
## OverTime
##   No  Yes
## 1054  416
```

Performance Rating

```
table(PerformanceRating)
```

```
## PerformanceRating
##    3    4
## 1244  226
```

Relationship Satisfaction

```
table(RelationshipSatisfaction)
```

```
## RelationshipSatisfaction
##   1   2   3   4
## 276 303 459 432
```

Stock Option Level

```
table(StockOptionLevel)
```

```
## StockOptionLevel
##   0   1   2   3
## 631 596 158  85
```

Work-life balance

```
table(WorkLifeBalance)
```

```
## WorkLifeBalance
##   1   2   3   4
##  80 344 893 153
```

Two way contigency tables:

Since there are 17 categorical variables in the dataset, it is humanly impossible to write the code for every 17C2 = 136 combinations. So I will try and list the most important ones from the 136 combinations.

Education and Education Field

```
xtabs(~Education+EducationField, data=employees.df)
```

```
##           EducationField
## Education Human Resources Life Sciences Marketing Medical Other
##         1              2             67        14      63     5
##         2              2            116        24      99    19
##         3             16            233        59     183    24
##         4              5            173        52     104    33
##         5              2             17        10      15     1
##           EducationField
## Education Technical Degree
##         1               19
##         2               22
##         3               57
##         4               31
##         5                3
```

Department and Education Field

```
xtabs(~Department+EducationField, data=employees.df)
```

```
##                          EducationField
## Department                Human Resources Life Sciences Marketing Medical
##    Human Resources                     27            16         0      13
##    Research & Development               0           440         0     363
##    Sales                                0           150       159      88
##                          EducationField
## Department                Other Technical Degree
##    Human Resources           3               4
##    Research & Development    64              94
##    Sales                     15              34
```

Department and Gender

```
xtabs(~Gender+Department, data=employees.df)
```

```
##         Department
## Gender   Human Resources Research & Development Sales
##    Female             20                    379   189
##    Male               43                    582   257
```

Job level and job involvement

```
xtabs(~JobLevel+JobInvolvement, data=employees.df)
```

```
##          JobInvolvement
## JobLevel   1   2   3   4
##        1  30 137 318  58
##        2  35 128 317  54
##        3  10  66 128  14
##        4   3  27  62  14
##        5   5  17  43   4
```

Work life balance and enviornment satisfaction

```
xtabs(~WorkLifeBalance+EnvironmentSatisfaction, data=employees.df)

##                 EnvironmentSatisfaction
## WorkLifeBalance   1   2   3   4
##               1  20  13  22  25
##               2  67  78 100  99
##               3 167 171 278 277
##               4  30  25  53  45
```

Job Satisfaction and Performance Rating

```
xtabs(~JobSatisfaction+PerformanceRating, data=employees.df)

##                 PerformanceRating
## JobSatisfaction   3   4
##               1 241  48
##               2 237  43
##               3 386  56
##               4 380  79
```
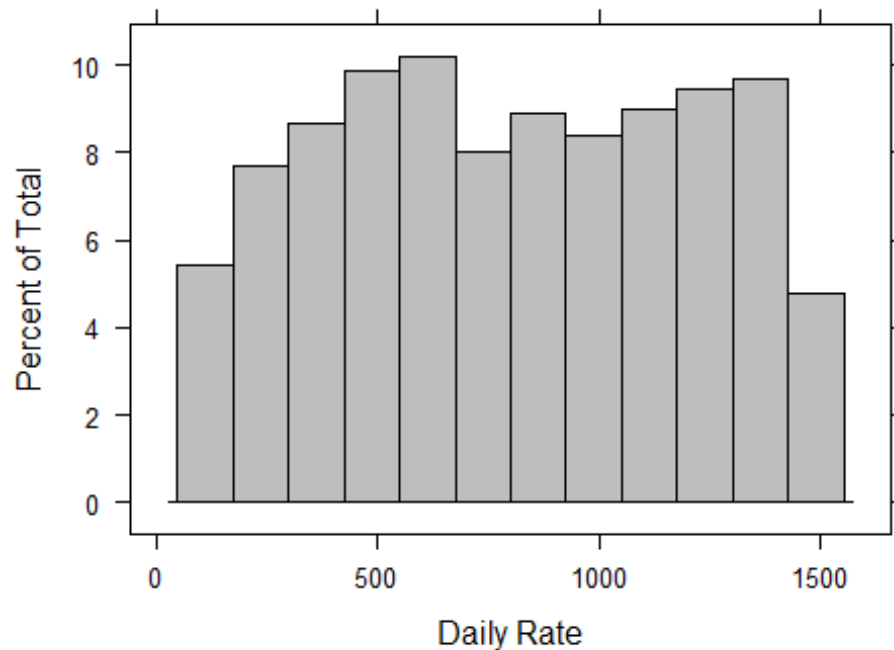
Overtime and Performance Rating

```
xtabs(~OverTime+PerformanceRating, data=employees.df)

##          PerformanceRating
## OverTime   3   4
##      No  893 161
##      Yes 351  65
```

Review the distribution of Daily Rate

```
library(lattice)
histogram(~DailyRate, data = employees.df,
main = "Distribution of Daily Rate", xlab="Daily Rate", col='grey' )
```

Distribution of Daily Rate

Comparison of Daily Rate with the given variables

```r
aggregate(cbind(DailyRate,YearsAtCompany,MonthlyIncome) ~ Gender,
data = employees.df, mean)

##    Gender DailyRate YearsAtCompany MonthlyIncome
## 1 Female   808.2738       7.231293      6686.566
## 2   Male   798.6270       6.859410      6380.508

library(car)

##
## Attaching package: 'car'

## The following object is masked from 'package:psych':
##
##     logit

#Distribution of Monthly Income and Years at Company
scatterplot(MonthlyIncome~YearsAtCompany,data=employees.df,main="Distribution
of monthly income with work experience",ylab="Monthly Income",xlab = "Years
at company")
```

Distribution of monthly income with work experience
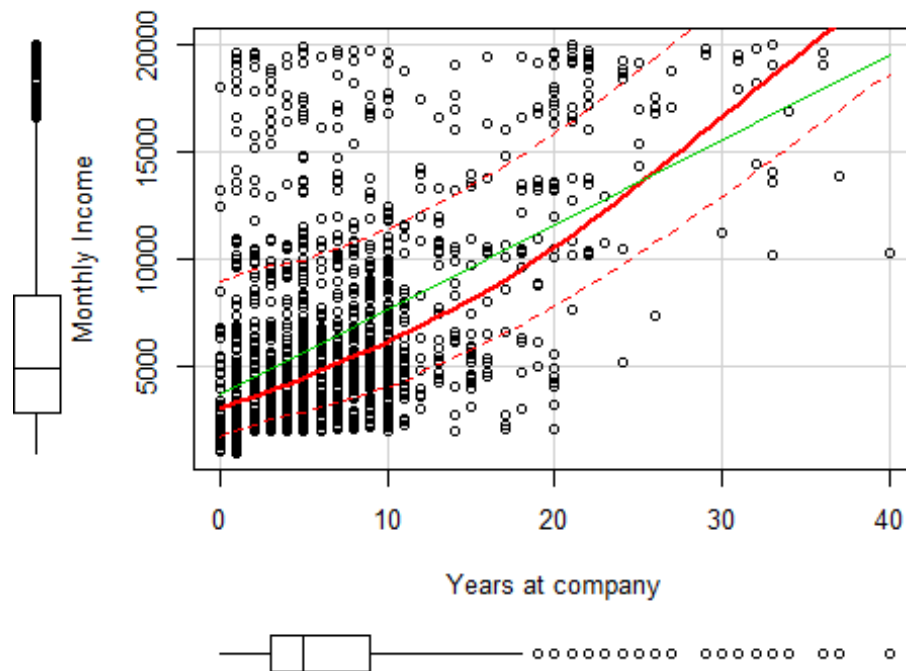

```
#Distribution of Monthly Income and Years at Company
boxplot(MonthlyIncome~YearsAtCompany,data=employees.df,main="Distribution of
monthly income with work experience",ylab="Monthly Income",xlab = "Years at
company")
```

Distribution of monthly income with work experien
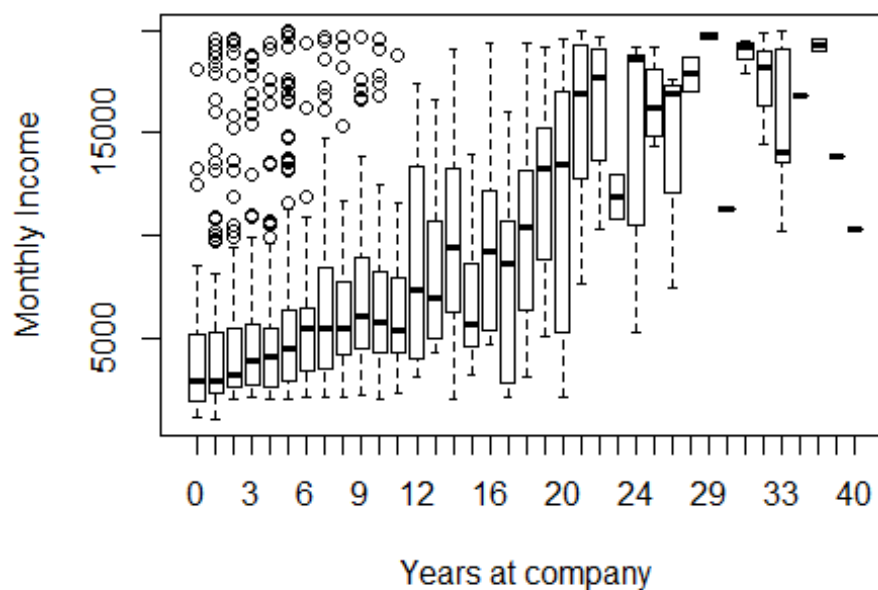
Comparison of Monthly Income and Job Satisfaction

```
#Distribution of Monthly Income and Job Satisfaction
scatterplot(MonthlyIncome~JobSatisfaction,data=employees.df,main="Distributio
n of monthly income with job satisfaction",ylab="Monthly Income",xlab = "Job
Satisfaction")
```


Distribution of monthly income with job satisfaction
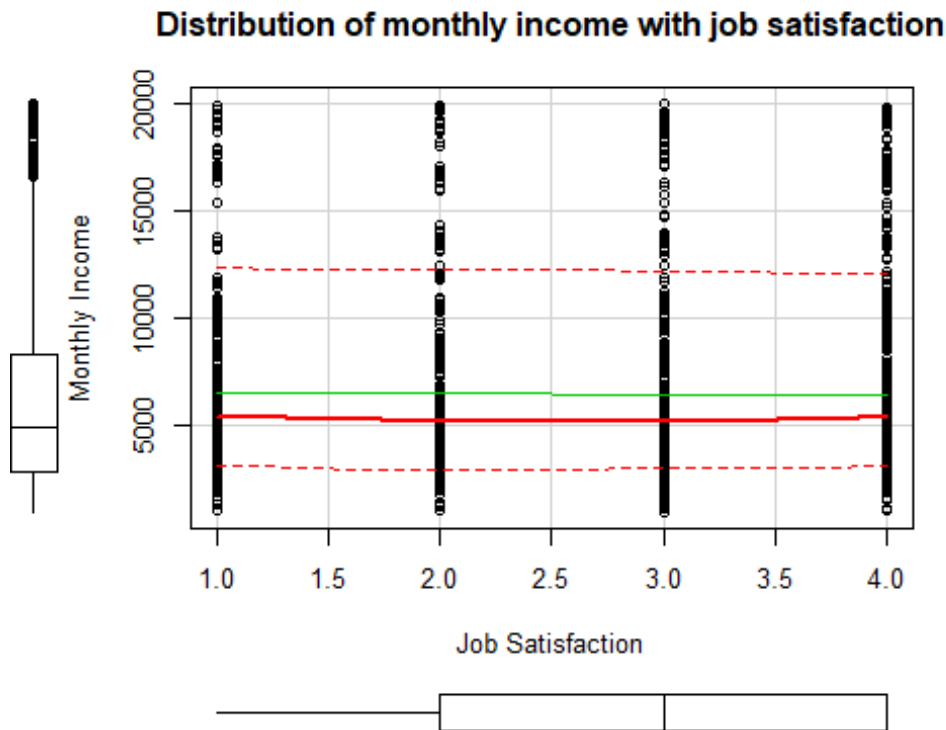
```
#Distribution of Monthly Income and Job Satisfaction
boxplot(MonthlyIncome~JobSatisfaction,data=employees.df,main="Distribution of
monthly income with job satisfaction",ylab="Monthly Income",xlab = "Job
Satisfaction")
```

Distribution of monthly income with job satisfactio
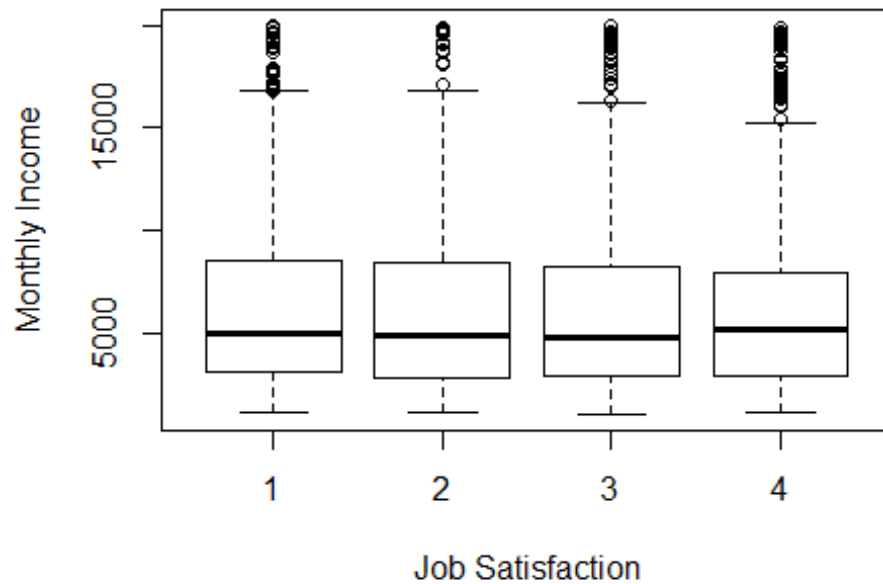
Comparison of Percent Salary Hike and Job Satisfaction

```
#Distribution of percent sa;ary hike and job satisfaction
scatterplot(PercentSalaryHike~JobSatisfaction,data=employees.df,main="Distrib
ution of percent salary hike with job satisfaction",ylab="Percent Salary
Hike",xlab = "Job Satisfaction")
```

Distribution of percent salary hike with job satisfaction


```
#Distribution of percent sa;ary hike and job satisfaction
boxplot(PercentSalaryHike~JobSatisfaction,data=employees.df,main="Distributio
n of percent salary hike with job satisfaction",ylab="Percent Salary
Hike",xlab = "Job Satisfaction")
```

Distribution of percent salary hike with job satisfact
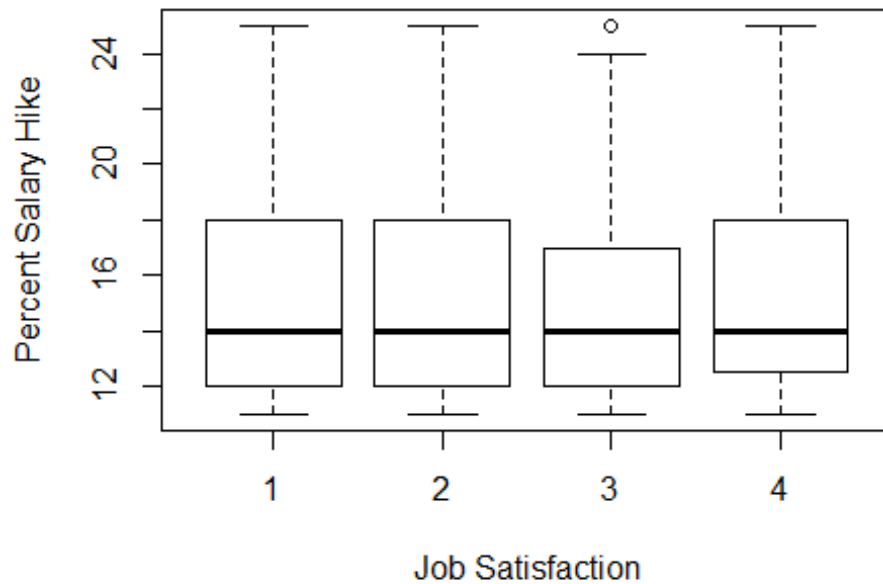
Review the distribution of percent salary hike

```
histogram(~PercentSalaryHike, data = employees.df,
main = "Distribution of Daily Rate", xlab="Daily Rate", col='grey' )
```

Distribution of Daily Rate
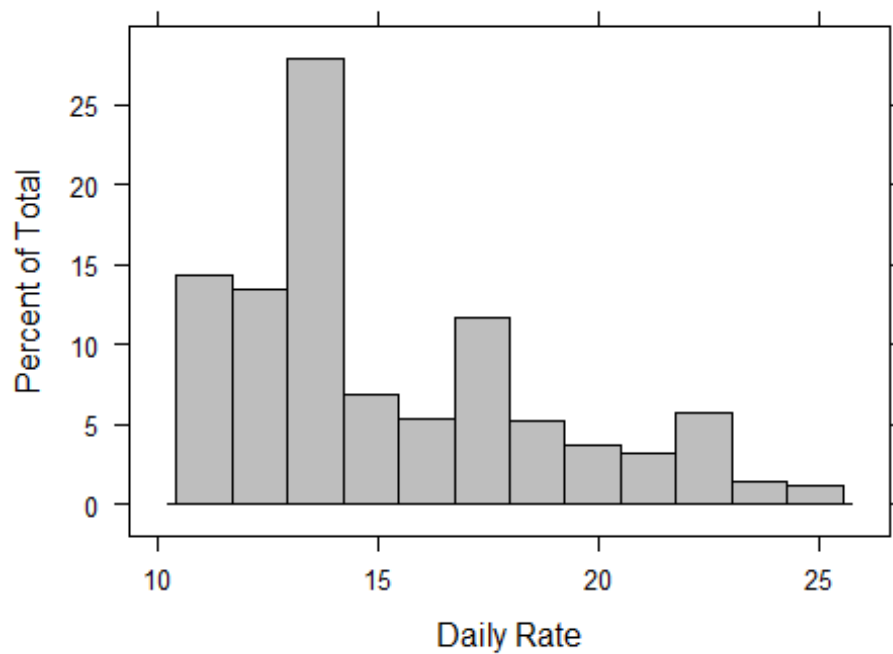
Comparison of Distance From Home and Enviornment Satisfaction

```
#Distribution of Distance From Home and Enviornment Satisfaction
scatterplot(DistanceFromHome~EnvironmentSatisfaction,data=employees.df,main="
Distribution of distance from home with enviornment
satisfaction",ylab="Distance from home",xlab = "Enviornment Satisfaction")
```


Distribution of distance from home with enviornment satisfac
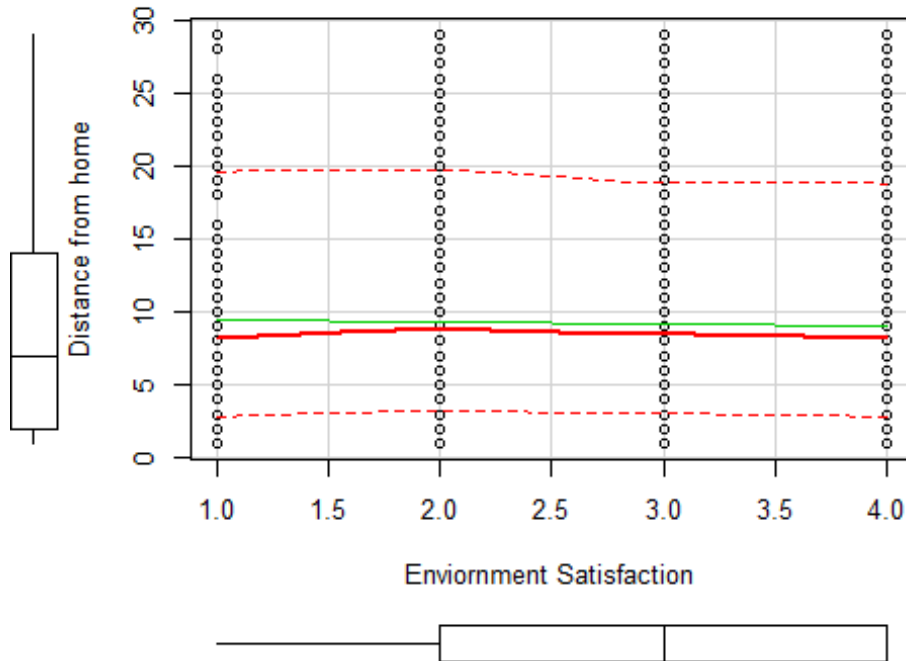
```
#Distribution of Distance From Home and Enviornment Satisfaction
boxplot(DistanceFromHome~EnvironmentSatisfaction,data=employees.df,main="Dist
ribution of distance from home with enviornment satisfaction",ylab="Distance
from home",xlab = "Enviornment Satisfaction")
```

ribution of distance from home with enviornment sati
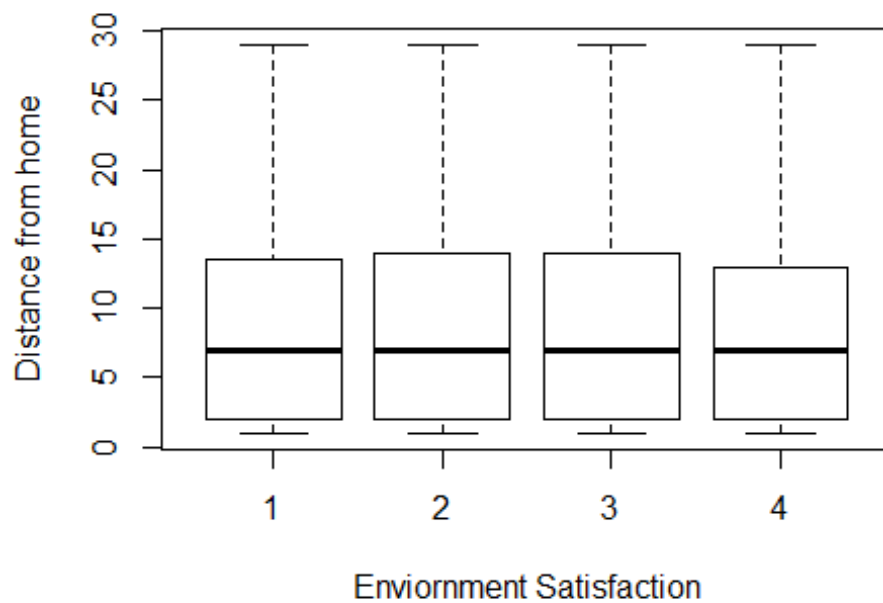
Review the distribution of Distance From Home

```
histogram(~DistanceFromHome, data = employees.df,
main = "Distribution of Distance From Home", xlab="Distance From Home",
col='grey' )
```

Distribution of Distance From Home
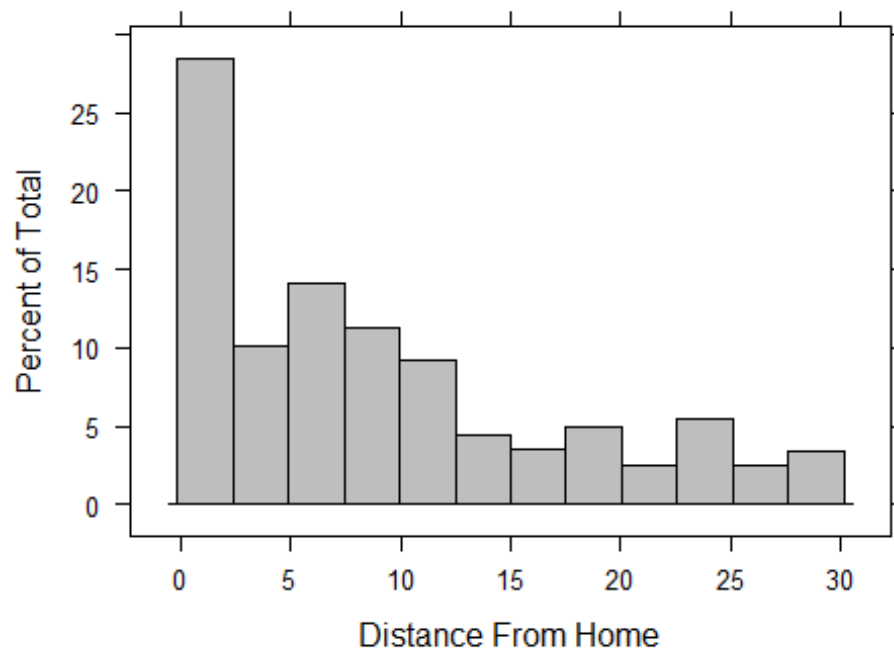
Corrogram and Correlation Matrix :

```r
library(corrplot)

## corrplot 0.84 loaded

colnames(employees.df)

##  [1] "ï..Age"                  "Attrition"
##  [3] "BusinessTravel"          "DailyRate"
##  [5] "Department"              "DistanceFromHome"
##  [7] "Education"               "EducationField"
##  [9] "EmployeeCount"           "EmployeeNumber"
## [11] "EnvironmentSatisfaction" "Gender"
## [13] "HourlyRate"              "JobInvolvement"
## [15] "JobLevel"                "JobRole"
## [17] "JobSatisfaction"         "MaritalStatus"
## [19] "MonthlyIncome"           "MonthlyRate"
## [21] "NumCompaniesWorked"      "Over18"
## [23] "OverTime"                "PercentSalaryHike"
## [25] "PerformanceRating"       "RelationshipSatisfaction"
## [27] "StandardHours"           "StockOptionLevel"
## [29] "TotalWorkingYears"       "TrainingTimesLastYear"
## [31] "WorkLifeBalance"         "YearsAtCompany"
## [33] "YearsInCurrentRole"      "YearsSinceLastPromotion"
## [35] "YearsWithCurrManager"

dataColumns <-
employees.df[,c("ï..Age","DailyRate","DistanceFromHome","Education","Employee
```

```
Count","EnvironmentSatisfaction","HourlyRate","JobInvolvement","JobLevel","Jo
bSatisfaction","MonthlyIncome","MonthlyRate","NumCompaniesWorked","PercentSal
aryHike","PerformanceRating","RelationshipSatisfaction","StandardHours","Stoc
kOptionLevel","TotalWorkingYears","TrainingTimesLastYear","WorkLifeBalance","
YearsAtCompany","YearsInCurrentRole","YearsSinceLastPromotion","YearsWithCurr
Manager")]
```

Correlation Matrix :

```
res <- cor(dataColumns)

## Warning in cor(dataColumns): the standard deviation is zero

round(res, 2)

##                            ï..Age DailyRate DistanceFromHome Education
## ï..Age                       1.00      0.01             0.00      0.21
## DailyRate                    0.01      1.00             0.00     -0.02
## DistanceFromHome             0.00      0.00             1.00      0.02
## Education                    0.21     -0.02             0.02      1.00
## EmployeeCount                  NA        NA               NA        NA
## EnvironmentSatisfaction      0.01      0.02            -0.02     -0.03
## HourlyRate                   0.02      0.02             0.03      0.02
## JobInvolvement               0.03      0.05             0.01      0.04
## JobLevel                     0.51      0.00             0.01      0.10
## JobSatisfaction              0.00      0.03             0.00     -0.01
## MonthlyIncome                0.50      0.01            -0.02      0.09
## MonthlyRate                  0.03     -0.03             0.03     -0.03
## NumCompaniesWorked           0.30      0.04            -0.03      0.13
## PercentSalaryHike            0.00      0.02             0.04     -0.01
## PerformanceRating            0.00      0.00             0.03     -0.02
## RelationshipSatisfaction     0.05      0.01             0.01     -0.01
## StandardHours                  NA        NA               NA        NA
## StockOptionLevel             0.04      0.04             0.04      0.02
## TotalWorkingYears            0.68      0.01             0.00      0.15
## TrainingTimesLastYear       -0.02      0.00            -0.04     -0.03
## WorkLifeBalance             -0.02     -0.04            -0.03      0.01
## YearsAtCompany               0.31     -0.03             0.01      0.07
## YearsInCurrentRole           0.21      0.01             0.02      0.06
## YearsSinceLastPromotion      0.22     -0.03             0.01      0.05
## YearsWithCurrManager         0.20     -0.03             0.01      0.07
##                            EmployeeCount EnvironmentSatisfaction HourlyRate
## ï..Age                                NA                    0.01       0.02
## DailyRate                             NA                    0.02       0.02
## DistanceFromHome                      NA                   -0.02       0.03
## Education                             NA                   -0.03       0.02
## EmployeeCount                          1                      NA         NA
## EnvironmentSatisfaction               NA                    1.00      -0.05
## HourlyRate                            NA                   -0.05       1.00
## JobInvolvement                        NA                   -0.01       0.04
## JobLevel                              NA                    0.00      -0.03
## JobSatisfaction                       NA                   -0.01      -0.07
```

```
## MonthlyIncome                         NA                  -0.01          -0.02
## MonthlyRate                           NA                   0.04          -0.02
## NumCompaniesWorked                    NA                   0.01           0.02
## PercentSalaryHike                     NA                  -0.03          -0.01
## PerformanceRating                     NA                  -0.03           0.00
## RelationshipSatisfaction             NA                   0.01           0.00
## StandardHours                         NA                     NA             NA
## StockOptionLevel                      NA                   0.00           0.05
## TotalWorkingYears                     NA                   0.00           0.00
## TrainingTimesLastYear                 NA                  -0.02          -0.01
## WorkLifeBalance                       NA                   0.03           0.00
## YearsAtCompany                        NA                   0.00          -0.02
## YearsInCurrentRole                    NA                   0.02          -0.02
## YearsSinceLastPromotion               NA                   0.02          -0.03
## YearsWithCurrManager                  NA                   0.00          -0.02
##                         JobInvolvement JobLevel JobSatisfaction
## ï..Age                            0.03     0.51            0.00
## DailyRate                         0.05     0.00            0.03
## DistanceFromHome                  0.01     0.01            0.00
## Education                         0.04     0.10           -0.01
## EmployeeCount                       NA       NA              NA
## EnvironmentSatisfaction          -0.01     0.00           -0.01
## HourlyRate                        0.04    -0.03           -0.07
## JobInvolvement                    1.00    -0.01           -0.02
## JobLevel                         -0.01     1.00            0.00
## JobSatisfaction                  -0.02     0.00            1.00
## MonthlyIncome                    -0.02     0.95           -0.01
## MonthlyRate                      -0.02     0.04            0.00
## NumCompaniesWorked                0.02     0.14           -0.06
## PercentSalaryHike                -0.02    -0.03            0.02
## PerformanceRating                -0.03    -0.02            0.00
## RelationshipSatisfaction          0.03     0.02           -0.01
## StandardHours                       NA       NA              NA
## StockOptionLevel                  0.02     0.01            0.01
## TotalWorkingYears                -0.01     0.78           -0.02
## TrainingTimesLastYear            -0.02    -0.02           -0.01
## WorkLifeBalance                  -0.01     0.04           -0.02
## YearsAtCompany                   -0.02     0.53            0.00
## YearsInCurrentRole                0.01     0.39            0.00
## YearsSinceLastPromotion          -0.02     0.35           -0.02
## YearsWithCurrManager              0.03     0.38           -0.03
##                         MonthlyIncome MonthlyRate NumCompaniesWorked
## ï..Age                           0.50        0.03               0.30
## DailyRate                        0.01       -0.03               0.04
## DistanceFromHome                -0.02        0.03              -0.03
## Education                        0.09       -0.03               0.13
## EmployeeCount                      NA          NA                 NA
## EnvironmentSatisfaction         -0.01        0.04               0.01
## HourlyRate                      -0.02       -0.02               0.02
## JobInvolvement                  -0.02       -0.02               0.02
## JobLevel                         0.95        0.04               0.14
```

```
## JobSatisfaction                    -0.01             0.00                 -0.06
## MonthlyIncome                        1.00             0.03                  0.15
## MonthlyRate                          0.03             1.00                  0.02
## NumCompaniesWorked                   0.15             0.02                  1.00
## PercentSalaryHike                   -0.03            -0.01                 -0.01
## PerformanceRating                   -0.02            -0.01                 -0.01
## RelationshipSatisfaction            0.03             0.00                  0.05
## StandardHours                          NA               NA                    NA
## StockOptionLevel                     0.01            -0.03                  0.03
## TotalWorkingYears                    0.77             0.03                  0.24
## TrainingTimesLastYear               -0.02             0.00                 -0.07
## WorkLifeBalance                      0.03             0.01                 -0.01
## YearsAtCompany                       0.51            -0.02                 -0.12
## YearsInCurrentRole                   0.36            -0.01                 -0.09
## YearsSinceLastPromotion              0.34             0.00                 -0.04
## YearsWithCurrManager                 0.34            -0.04                 -0.11
##                           PercentSalaryHike PerformanceRating
## ï..Age                                 0.00              0.00
## DailyRate                              0.02              0.00
## DistanceFromHome                       0.04              0.03
## Education                             -0.01             -0.02
## EmployeeCount                            NA                NA
## EnvironmentSatisfaction               -0.03             -0.03
## HourlyRate                            -0.01              0.00
## JobInvolvement                        -0.02             -0.03
## JobLevel                              -0.03             -0.02
## JobSatisfaction                        0.02              0.00
## MonthlyIncome                         -0.03             -0.02
## MonthlyRate                           -0.01             -0.01
## NumCompaniesWorked                    -0.01             -0.01
## PercentSalaryHike                      1.00              0.77
## PerformanceRating                      0.77              1.00
## RelationshipSatisfaction              -0.04             -0.03
## StandardHours                            NA                NA
## StockOptionLevel                       0.01              0.00
## TotalWorkingYears                     -0.02              0.01
## TrainingTimesLastYear                 -0.01             -0.02
## WorkLifeBalance                        0.00              0.00
## YearsAtCompany                        -0.04              0.00
## YearsInCurrentRole                     0.00              0.03
## YearsSinceLastPromotion               -0.02              0.02
## YearsWithCurrManager                  -0.01              0.02
##                           RelationshipSatisfaction StandardHours
## ï..Age                                        0.05            NA
## DailyRate                                     0.01            NA
## DistanceFromHome                              0.01            NA
## Education                                    -0.01            NA
## EmployeeCount                                   NA            NA
## EnvironmentSatisfaction                       0.01            NA
## HourlyRate                                    0.00            NA
## JobInvolvement                                0.03            NA
```

```
## JobLevel                                          0.02                  NA
## JobSatisfaction                                   -0.01                 NA
## MonthlyIncome                                      0.03                 NA
## MonthlyRate                                        0.00                 NA
## NumCompaniesWorked                                 0.05                 NA
## PercentSalaryHike                                 -0.04                 NA
## PerformanceRating                                 -0.03                 NA
## RelationshipSatisfaction                           1.00                 NA
## StandardHours                                        NA                  1
## StockOptionLevel                                  -0.05                 NA
## TotalWorkingYears                                  0.02                 NA
## TrainingTimesLastYear                              0.00                 NA
## WorkLifeBalance                                    0.02                 NA
## YearsAtCompany                                     0.02                 NA
## YearsInCurrentRole                                -0.02                 NA
## YearsSinceLastPromotion                            0.03                 NA
## YearsWithCurrManager                               0.00                 NA
##                             StockOptionLevel TotalWorkingYears
## ï..Age                                  0.04              0.68
## DailyRate                               0.04              0.01
## DistanceFromHome                        0.04              0.00
## Education                               0.02              0.15
## EmployeeCount                             NA                NA
## EnvironmentSatisfaction                 0.00              0.00
## HourlyRate                              0.05              0.00
## JobInvolvement                          0.02             -0.01
## JobLevel                                0.01              0.78
## JobSatisfaction                         0.01             -0.02
## MonthlyIncome                           0.01              0.77
## MonthlyRate                            -0.03              0.03
## NumCompaniesWorked                      0.03              0.24
## PercentSalaryHike                       0.01             -0.02
## PerformanceRating                       0.00              0.01
## RelationshipSatisfaction               -0.05              0.02
## StandardHours                             NA                NA
## StockOptionLevel                        1.00              0.01
## TotalWorkingYears                       0.01              1.00
## TrainingTimesLastYear                   0.01             -0.04
## WorkLifeBalance                         0.00              0.00
## YearsAtCompany                          0.02              0.63
## YearsInCurrentRole                      0.05              0.46
## YearsSinceLastPromotion                 0.01              0.40
## YearsWithCurrManager                    0.02              0.46
##                             TrainingTimesLastYear WorkLifeBalance
## ï..Age                                      -0.02           -0.02
## DailyRate                                    0.00           -0.04
## DistanceFromHome                            -0.04           -0.03
## Education                                   -0.03            0.01
## EmployeeCount                                  NA              NA
## EnvironmentSatisfaction                     -0.02            0.03
## HourlyRate                                  -0.01            0.00
```

```
## JobInvolvement                         -0.02              -0.01
## JobLevel                               -0.02               0.04
## JobSatisfaction                        -0.01              -0.02
## MonthlyIncome                          -0.02               0.03
## MonthlyRate                             0.00               0.01
## NumCompaniesWorked                     -0.07              -0.01
## PercentSalaryHike                      -0.01               0.00
## PerformanceRating                      -0.02               0.00
## RelationshipSatisfaction               0.00               0.02
## StandardHours                            NA                 NA
## StockOptionLevel                        0.01               0.00
## TotalWorkingYears                      -0.04               0.00
## TrainingTimesLastYear                   1.00               0.03
## WorkLifeBalance                         0.03               1.00
## YearsAtCompany                          0.00               0.01
## YearsInCurrentRole                     -0.01               0.05
## YearsSinceLastPromotion                 0.00               0.01
## YearsWithCurrManager                    0.00               0.00
##                           YearsAtCompany YearsInCurrentRole
## ï..Age                          0.31               0.21
## DailyRate                      -0.03               0.01
## DistanceFromHome                0.01               0.02
## Education                       0.07               0.06
## EmployeeCount                    NA                 NA
## EnvironmentSatisfaction         0.00               0.02
## HourlyRate                     -0.02              -0.02
## JobInvolvement                 -0.02               0.01
## JobLevel                        0.53               0.39
## JobSatisfaction                 0.00               0.00
## MonthlyIncome                   0.51               0.36
## MonthlyRate                    -0.02              -0.01
## NumCompaniesWorked             -0.12              -0.09
## PercentSalaryHike              -0.04               0.00
## PerformanceRating               0.00               0.03
## RelationshipSatisfaction        0.02              -0.02
## StandardHours                    NA                 NA
## StockOptionLevel                0.02               0.05
## TotalWorkingYears               0.63               0.46
## TrainingTimesLastYear           0.00              -0.01
## WorkLifeBalance                 0.01               0.05
## YearsAtCompany                  1.00               0.76
## YearsInCurrentRole              0.76               1.00
## YearsSinceLastPromotion         0.62               0.55
## YearsWithCurrManager            0.77               0.71
##                           YearsSinceLastPromotion YearsWithCurrManager
## ï..Age                               0.22                     0.20
## DailyRate                           -0.03                    -0.03
## DistanceFromHome                     0.01                     0.01
## Education                            0.05                     0.07
## EmployeeCount                         NA                       NA
## EnvironmentSatisfaction              0.02                     0.00
```

```
## HourlyRate                    -0.03        -0.02
## JobInvolvement                -0.02         0.03
## JobLevel                       0.35         0.38
## JobSatisfaction               -0.02        -0.03
## MonthlyIncome                  0.34         0.34
## MonthlyRate                    0.00        -0.04
## NumCompaniesWorked            -0.04        -0.11
## PercentSalaryHike             -0.02        -0.01
## PerformanceRating              0.02         0.02
## RelationshipSatisfaction       0.03         0.00
## StandardHours                    NA           NA
## StockOptionLevel               0.01         0.02
## TotalWorkingYears              0.40         0.46
## TrainingTimesLastYear          0.00         0.00
## WorkLifeBalance                0.01         0.00
## YearsAtCompany                 0.62         0.77
## YearsInCurrentRole             0.55         0.71
## YearsSinceLastPromotion        1.00         0.51
## YearsWithCurrManager           0.51         1.00
```

Corrogram :

```
N <- cor(dataColumns)

## Warning in cor(dataColumns): the standard deviation is zero

corrplot(N, method="circle")
```
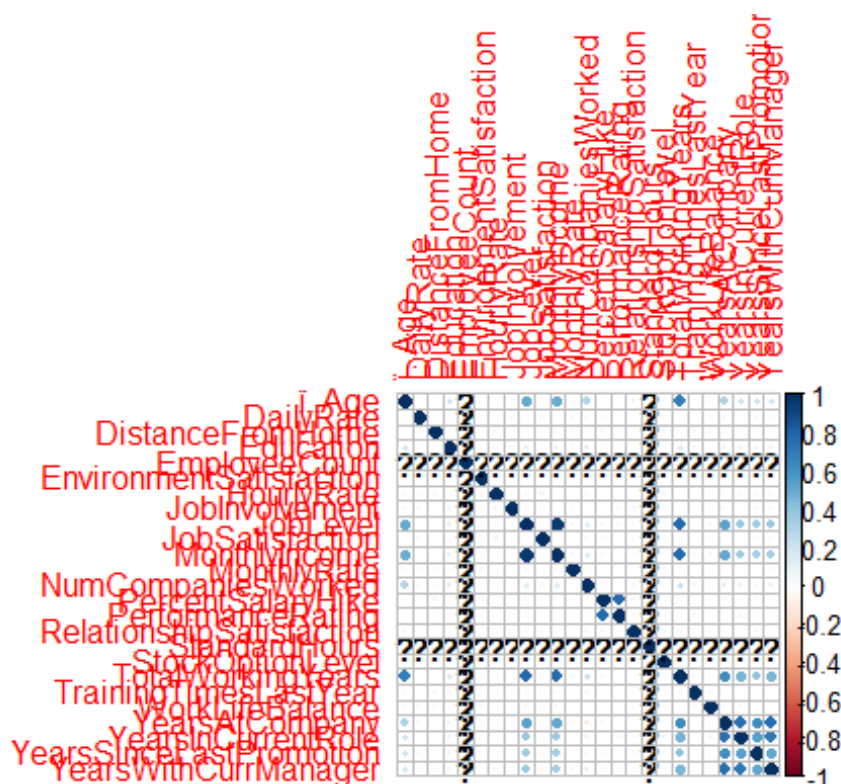

Scatterplot Matrix :

```
pairs(dataColumns[1:6])
```
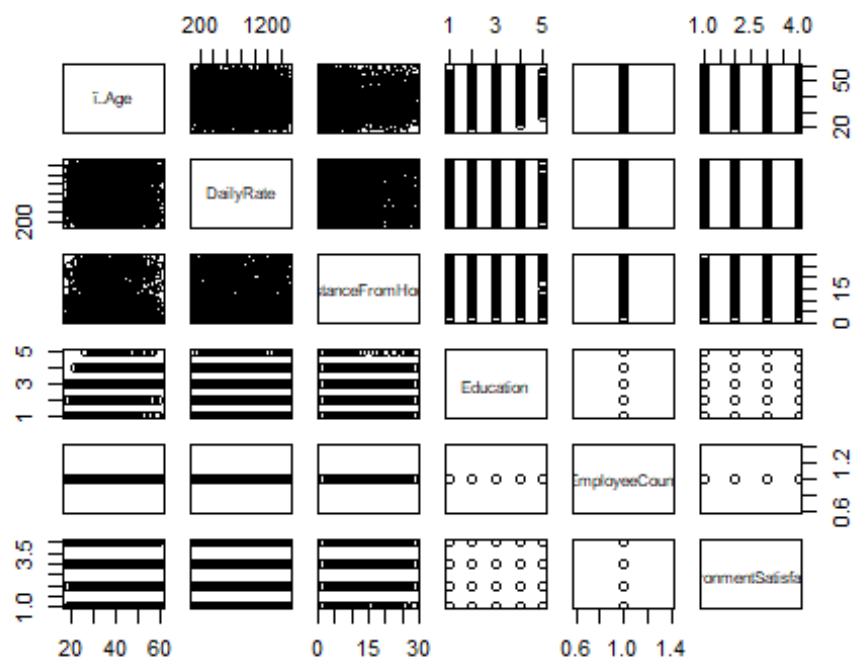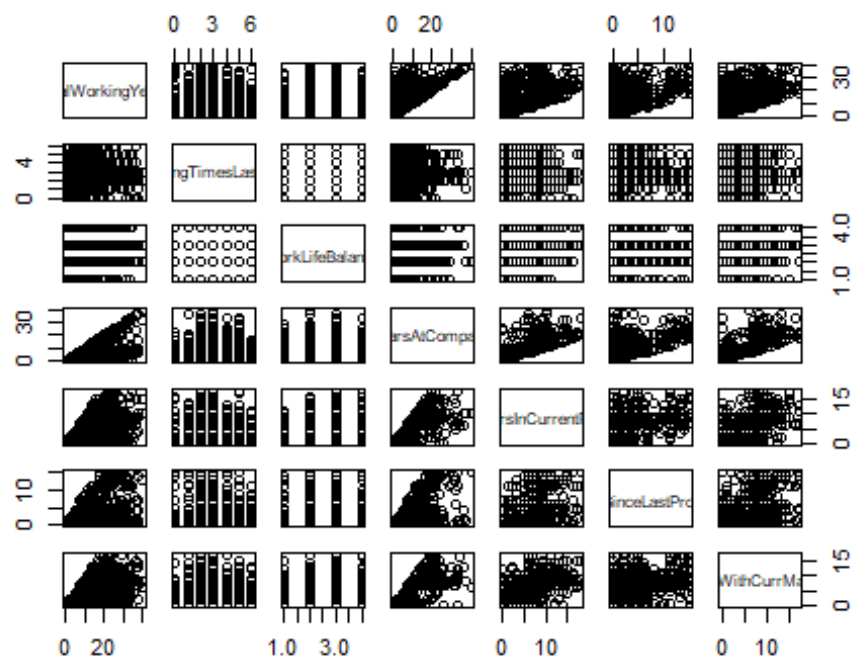
```
pairs(dataColumns[7:12])
```
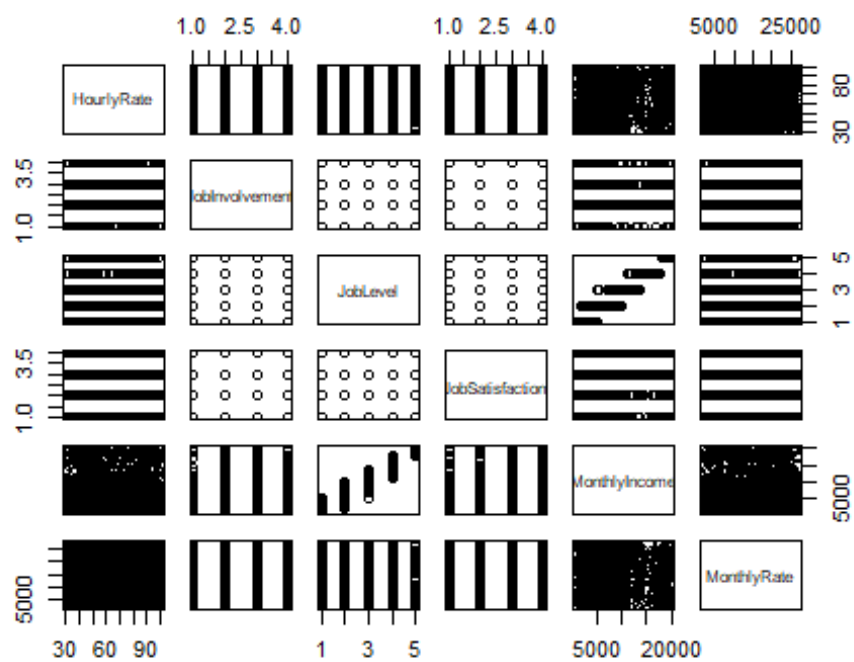


```
pairs(dataColumns[13:18])
```
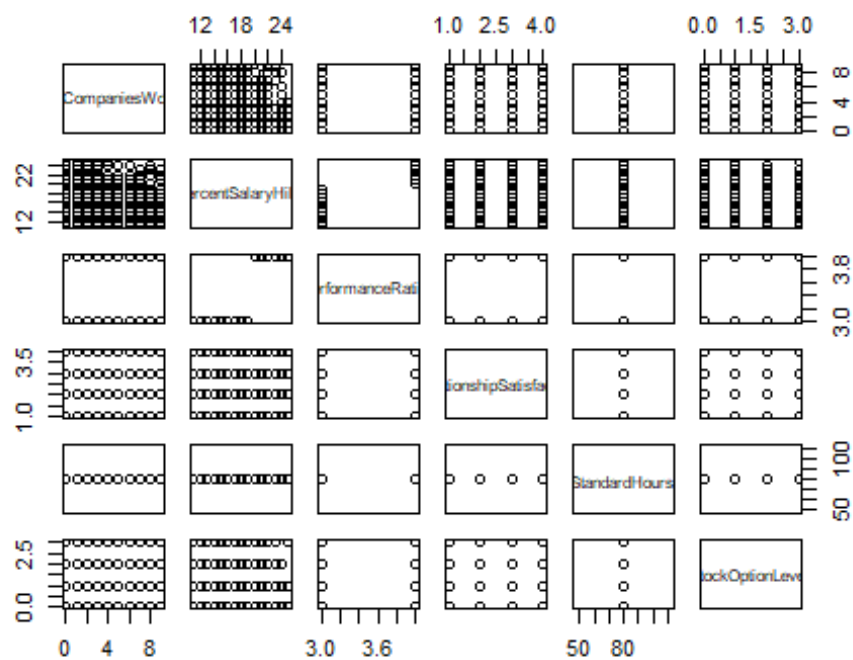
```
pairs(dataColumns[19:25])
```


What variables does attrition depend on :

```
##Attrition VS How frequently an employee goes to business travel
library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha

ggplot(employees.df,aes(BusinessTravel,fill=Attrition))+geom_bar(position=pos
ition_dodge())+labs(x="Travel Frequency",y="Count",title="Attrition Vs
Business Travel")
```


Attrition Vs Business Travel
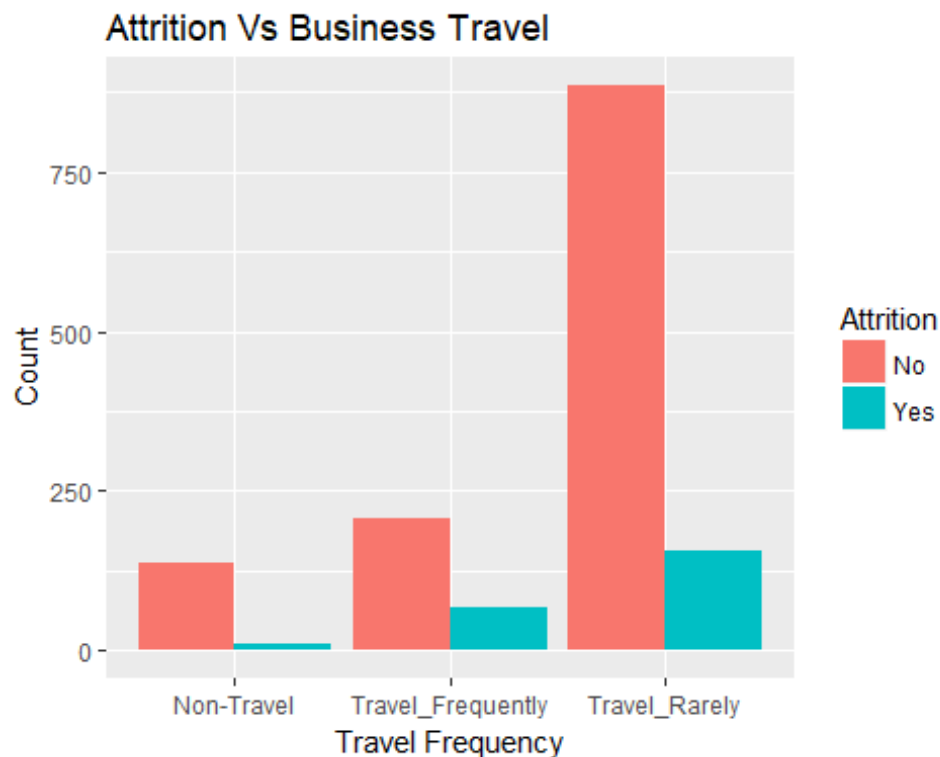
```
table_travel<-table(employees.df$BusinessTravel, employees.df$Attrition)
chisq.test(table_travel)

##
##  Pearson's Chi-squared test
##
## data:  table_travel
## X-squared = 24.182, df = 2, p-value = 5.609e-06
```

#The barplot shows that employees who travel rarely do not frequently quit
the job. Thus, attrition is dependent on business travel, and the chi square
test proves this.

#Attrition VS Distance from home
`t.test(DistanceFromHome~Attrition, data=employees.df)`

```
##
##   Welch Two Sample t-test
##
## data:  DistanceFromHome by Attrition
## t = -2.8882, df = 322.72, p-value = 0.004137
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -2.8870025 -0.5475146
## sample estimates:
##   mean in group No mean in group Yes
##           8.915653          10.632911
```

#As the p-value is less than alpha(0.05), the attrition of employee depends
on distance from home.

#Attrition VS Job Level
t.test(JobLevel~Attrition, data=employees.df)

```
##
##   Welch Two Sample t-test
##
## data:  JobLevel by Attrition
## t = 7.3859, df = 376.25, p-value = 9.845e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.3733861 0.6443231
## sample estimates:
##   mean in group No mean in group Yes
##           2.145985          1.637131
```

#As p-value is less than alpha, attrition depends on job level.

#Attrition VS Job Role
table_role<-table(employees.df$JobRole, employees.df$Attrition)
chisq.test(table_role)

```
##
##   Pearson's Chi-squared test
##
## data:  table_role
## X-squared = 86.19, df = 8, p-value = 2.752e-15
```

#As p-value is less than alpha, attrition depends on job role.

#Attrition VS Job Satisfaction
table_job_sat<-table(employees.df$JobSatisfaction, employees.df$Attrition)
chisq.test(table_job_sat)

```
##
##   Pearson's Chi-squared test
##
## data:  table_job_sat
## X-squared = 17.505, df = 3, p-value = 0.0005563
```

```
#As p-value is less than alpha, attrition depends on job satisfaction of
employees.
```

So Attrition depends on : Business Travel Distance From Home Job Level Job Role Job Satisfaction