

Project title: IBM HR data Analysis

Name: Aman Sharma

E-Mail: aman.rajendra1998@gmail.com

College: Birla Institute of Technology Mesra Ranchi

```
setwd("C:/Users/AMAN/Desktop/iiml/project")
ds <- read.csv(paste("ibm.csv", sep=""))
attach(ds)
View(ds)
```

The dimensions of the dataset is

```
dim(ds)

## [1] 1470   35
#1470 rows and 35 columns
```

Describing and summarizing the dataset

```
library(psych)
str(ds)

## 'data.frame': 1470 obs. of 35 variables:
## $ i..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 ...
## $ 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 2 4 ...
## $ EmployeeCount : int 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 ...
## $ Overtime                 : Factor w/ 2 levels "No","Yes": 2 1 2 2 1 1 2 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 ...

summary(ds)

##      i..Age      Attrition          BusinessTravel DailyRate
## Min.    :18.00  No :1233  Non-Travel       : 150  Min.    : 102.0
## 1st Qu.:30.00  Yes: 237   Travel_Frequently: 277  1st Qu.: 465.0
## Median  :36.00                           Travel_Rarely   :1043  Median   : 802.0
## Mean    :36.92                           Mean          : 802.5
## 3rd Qu.:43.00                           3rd Qu.:1157.0
## Max.    :60.00                           Max.       :1499.0
##
##              Department DistanceFromHome Education
## Human Resources   : 63   Min.    : 1.000  Min.    :1.000
## Research & Development:961  1st Qu.: 2.000  1st Qu.:2.000
## Sales             :446   Median   : 7.000  Median   :3.000
##                               Mean     : 9.193  Mean     :2.913
##                               3rd Qu.:14.000 3rd Qu.:4.000
##                               Max.    :29.000  Max.    :5.000
##
##      EducationField EmployeeCount EmployeeNumber
## Human Resources : 27   Min.    :1       Min.    : 1.0
## Life Sciences   :606   1st Qu.:1       1st Qu.: 491.2
## Marketing       :159   Median   :1       Median   :1020.5
## Medical          :464   Mean     :1       Mean     :1024.9
## Other            : 82   3rd Qu.:1       3rd Qu.:1555.8
## Technical Degree:132  Max.    :1       Max.    :2068.0
##
##      EnvironmentSatisfaction Gender HourlyRate JobInvolvement
## Min.    :1.000  Female:588  Min.    : 30.00  Min.    :1.00
## 1st Qu.:2.000          Male :882  1st Qu.: 48.00  1st Qu.:2.00
## Median  :3.000                           Median   : 66.00  Median   :3.00
## Mean    :2.722                           Mean     : 65.89  Mean     :2.73
## 3rd Qu.:4.000                           3rd Qu.: 83.75 3rd Qu.:3.00
## Max.    :4.000                           Max.    :100.00  Max.    :4.00
##
##      JobLevel          JobRole JobSatisfaction
## Min.    :1.000  Sales Executive   :326  Min.    :1.000
## 1st Qu.:1.000  Research Scientist:292  1st Qu.:2.000
## Median  :2.000  Laboratory Technician:259  Median :3.000
## Mean    :2.064  Manufacturing Director:145  Mean    :2.729

```

```

## 3rd Qu.:3.000  Healthcare Representative:131  3rd Qu.:4.000
## Max. :5.000   Manager                  :102   Max. :4.000
##          (Other)                 :215
## MaritalStatus MonthlyIncome   MonthlyRate NumCompaniesWorked
## Divorced:327   Min.    : 1009   Min.    : 2094   Min.    :0.000
## Married :673   1st Qu.: 2911   1st Qu.: 8047   1st Qu.:1.000
## Single  :470   Median  : 4919   Median  :14236   Median  :2.000
##           Mean    : 6503   Mean    :14313   Mean    :2.693
##           3rd Qu.: 8379   3rd Qu.:20462   3rd Qu.:4.000
##           Max.   :19999   Max.   :26999   Max.   :9.000
##
## Over18   Overtime   PercentSalaryHike PerformanceRating
## Y:1470   No :1054   Min.    :11.00   Min.    :3.000
##          Yes: 416   1st Qu.:12.00   1st Qu.:3.000
##           Median :14.00   Median :3.000
##           Mean   :15.21   Mean   :3.154
##           3rd Qu.:18.00   3rd Qu.:3.000
##           Max.   :25.00   Max.   :4.000
##
## RelationshipSatisfaction StandardHours StockOptionLevel TotalWorkingYears
## Min.    :1.000            Min.    :80     Min.    :0.0000   Min.    : 0.00
## 1st Qu.:2.000            1st Qu.:80     1st Qu.:0.0000   1st Qu.: 6.00
## Median  :3.000            Median :80     Median :1.0000   Median :10.00
## Mean    :2.712            Mean   :80     Mean   :0.7939   Mean   :11.28
## 3rd Qu.:4.000            3rd Qu.:80     3rd Qu.:1.0000   3rd Qu.:15.00
## Max.   :4.000            Max.   :80     Max.   :3.0000   Max.   :40.00
##
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany  YearsInCurrentRole
## Min.    :0.000            Min.    :1.000   Min.    : 0.000   Min.    : 0.000
## 1st Qu.:2.000            1st Qu.:2.000   1st Qu.: 3.000   1st Qu.: 2.000
## Median  :3.000            Median :3.000   Median : 5.000   Median : 3.000
## Mean    :2.799            Mean   :2.761   Mean   : 7.008   Mean   : 4.229
## 3rd Qu.:3.000            3rd Qu.:3.000   3rd Qu.: 9.000   3rd Qu.: 7.000
## Max.   :6.000            Max.   :4.000   Max.   :40.000   Max.   :18.000
##
## YearsSinceLastPromotion YearsWithCurrManager
## Min.    : 0.000            Min.    : 0.000
## 1st Qu.: 0.000            1st Qu.: 2.000
## Median  : 1.000            Median : 3.000
## Mean    : 2.188            Mean   : 4.123
## 3rd Qu.: 3.000            3rd Qu.: 7.000
## Max.   :15.000            Max.   :17.000
##

```

```
describe(ds)
```

	vars	n	mean	sd	median	trimmed
## i..Age		1 1470	36.92	9.14	36.0	36.47
## Attrition*		2 1470	1.16	0.37	1.0	1.08
## BusinessTravel*		3 1470	2.61	0.67	3.0	2.76
## DailyRate		4 1470	802.49	403.51	802.0	803.83
## Department*		5 1470	2.26	0.53	2.0	2.25
## DistanceFromHome		6 1470	9.19	8.11	7.0	8.08
## Education		7 1470	2.91	1.02	3.0	2.98
## EducationField*		8 1470	3.25	1.33	3.0	3.10

## EmployeeCount	9	1470	1.00	0.00	1.0	1.00			
## EmployeeNumber	10	1470	1024.87	602.02	1020.5	1023.40			
## EnvironmentSatisfaction	11	1470	2.72	1.09	3.0	2.78			
## Gender*	12	1470	1.60	0.49	2.0	1.62			
## HourlyRate	13	1470	65.89	20.33	66.0	66.02			
## JobInvolvement	14	1470	2.73	0.71	3.0	2.74			
## JobLevel	15	1470	2.06	1.11	2.0	1.90			
## JobRole*	16	1470	5.46	2.46	6.0	5.61			
## JobSatisfaction	17	1470	2.73	1.10	3.0	2.79			
## MaritalStatus*	18	1470	2.10	0.73	2.0	2.12			
## MonthlyIncome	19	1470	6502.93	4707.96	4919.0	5667.24			
## MonthlyRate	20	1470	14313.10	7117.79	14235.5	14286.48			
## NumCompaniesWorked	21	1470	2.69	2.50	2.0	2.36			
## Over18*	22	1470	1.00	0.00	1.0	1.00			
## OverTime*	23	1470	1.28	0.45	1.0	1.23			
## PercentSalaryHike	24	1470	15.21	3.66	14.0	14.80			
## PerformanceRating	25	1470	3.15	0.36	3.0	3.07			
## RelationshipSatisfaction	26	1470	2.71	1.08	3.0	2.77			
## StandardHours	27	1470	80.00	0.00	80.0	80.00			
## StockOptionLevel	28	1470	0.79	0.85	1.0	0.67			
## TotalWorkingYears	29	1470	11.28	7.78	10.0	10.37			
## TrainingTimesLastYear	30	1470	2.80	1.29	3.0	2.72			
## WorkLifeBalance	31	1470	2.76	0.71	3.0	2.77			
## YearsAtCompany	32	1470	7.01	6.13	5.0	5.99			
## YearsInCurrentRole	33	1470	4.23	3.62	3.0	3.85			
## YearsSinceLastPromotion	34	1470	2.19	3.22	1.0	1.48			
## YearsWithCurrManager	35	1470	4.12	3.57	3.0	3.77			
			mad	min	max	range	skew	kurtosis	se
## i..Age			8.90	18	60	42	0.41	-0.41	0.24
## Attrition*			0.00	1	2	1	1.84	1.39	0.01
## BusinessTravel*			0.00	1	3	2	-1.44	0.69	0.02
## DailyRate			510.01	102	1499	1397	0.00	-1.21	10.52
## Department*			0.00	1	3	2	0.17	-0.40	0.01
## DistanceFromHome			7.41	1	29	28	0.96	-0.23	0.21
## Education			1.48	1	5	4	-0.29	-0.56	0.03
## EducationField*			1.48	1	6	5	0.55	-0.69	0.03
## EmployeeCount			0.00	1	1	0	NaN	NaN	0.00
## EmployeeNumber			790.97	1	2068	2067	0.02	-1.23	15.70
## EnvironmentSatisfaction			1.48	1	4	3	-0.32	-1.20	0.03
## Gender*			0.00	1	2	1	-0.41	-1.83	0.01
## HourlyRate			26.69	30	100	70	-0.03	-1.20	0.53
## JobInvolvement			0.00	1	4	3	-0.50	0.26	0.02
## JobLevel			1.48	1	5	4	1.02	0.39	0.03
## JobRole*			2.97	1	9	8	-0.36	-1.20	0.06
## JobSatisfaction			1.48	1	4	3	-0.33	-1.22	0.03
## MaritalStatus*			1.48	1	3	2	-0.15	-1.12	0.02
## MonthlyIncome			3260.24	1009	19999	18990	1.37	0.99	122.79
## MonthlyRate			9201.76	2094	26999	24905	0.02	-1.22	185.65
## NumCompaniesWorked			1.48	0	9	9	1.02	0.00	0.07
## Over18*			0.00	1	1	0	NaN	NaN	0.00
## OverTime*			0.00	1	2	1	0.96	-1.07	0.01
## PercentSalaryHike			2.97	11	25	14	0.82	-0.31	0.10
## PerformanceRating			0.00	3	4	1	1.92	1.68	0.01
## RelationshipSatisfaction			1.48	1	4	3	-0.30	-1.19	0.03

```

## StandardHours          0.00   80    80     0    NaN    NaN   0.00
## StockOptionLevel      1.48    0     3     3    0.97   0.35   0.02
## TotalWorkingYears     5.93    0    40    40    1.11   0.91   0.20
## TrainingTimesLastYear 1.48    0     6     6    0.55   0.48   0.03
## WorkLifeBalance       0.00    1     4     3   -0.55   0.41   0.02
## YearsAtCompany        4.45    0    40    40    1.76   3.91   0.16
## YearsInCurrentRole    4.45    0    18    18    0.92   0.47   0.09
## YearsSinceLastPromotion 1.48    0    15    15    1.98   3.59   0.08
## YearsWithCurrManager   4.45    0    17    17    0.83   0.16   0.09

```

The categorical variables of the data set are as follows Attrition,Business Travel,Department,Education Field,Gender,Job Role,Marital Status,Over 18,Overtime,Education,Environment Satisfaction,Job Involvement,Job Level,Job Satisfaction,Performance Rating,Relationship Satisfaction,Stock Option Level,Work Life Balance.

One way contingency tables

Attrition

```



```

Business Travel

```



```

Department

```



```

Education Field

```



```

Gender

```
table(ds$Gender)

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

Job Role

```
table(ds$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
```

Marital Status

```
table(ds$MaritalStatus)

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

Over 18

```
table(ds$Over18)

##
##      Y
## 1470
```

Overtime

```
table(ds$OverTime)

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

Education

```
table(ds$Education)

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

Environment Satisfaction

```
table(ds$EnvironmentSatisfaction)

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

Job Involvement

```
table(ds$JobInvolvement)

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

Job Level

```
table(ds$JobLevel)

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

Job Satisfaction

```
table(ds$JobSatisfaction)

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

Performance Rating

```
table(ds$PerformanceRating)

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

Relationship Satisfaction

```
table(ds$RelationshipSatisfaction)

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

Stock Option Level

```
table(ds$StockOptionLevel)

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

Work Life Balance

```
table(ds$WorkLifeBalance)

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

Two way contingency tables

Job Satisfaction and Performance Rating

```
xtabs(~JobSatisfaction+PerformanceRating)

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

Job Involvement and Environment Satisfaction

```
xtabs(~JobInvolvement+EnvironmentSatisfaction)

##          EnvironmentSatisfaction
## JobInvolvement 1   2   3   4
##                 1 17 23 22 21
##                 2 74 60 124 117
##                 3 164 165 266 273
##                 4 29 39 41 35
```

Job Role and Department

```
xtabs(~JobRole+Department)
```

```
##                                     Department
## JobRole                               Human Resources Research & Development Sales
## Healthcare Representative           0                  131      0
## Human Resources                     52                  0      0
## Laboratory Technician               0                  259      0
## Manager                            11                  54     37
## Manufacturing Director             0                  145      0
## Research Director                  0                  80      0
## Research Scientist                 0                  292      0
## Sales Executive                   0                  0     326
## Sales Representative                0                  0     83
```

Education Field and Department

```
xtabs(~EducationField+Department)
```

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

Overtime and Work life balance

```
xtabs(~OverTime+WorkLifeBalance)
```

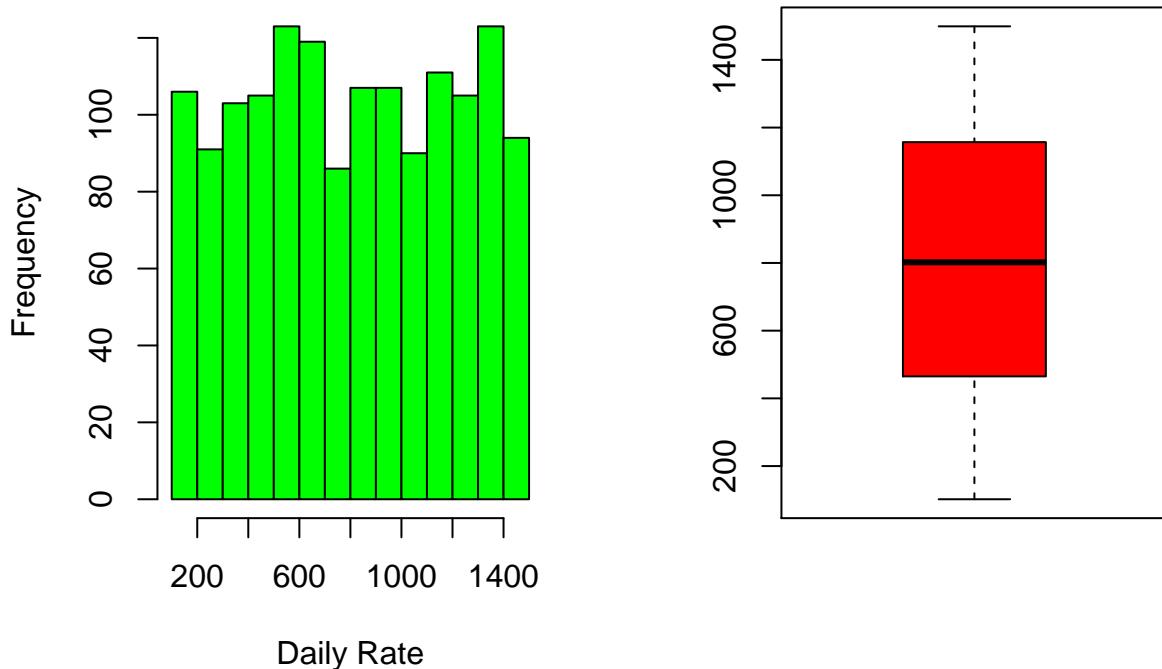
```
##           WorkLifeBalance
## OverTime   1   2   3   4
##       No    58  240  639 117
##       Yes   22  104  254  36
```

Plots of different variables of dataset

```
par(mfrow=c(1,2))
hist(ds$DailyRate,col = "green",xlab = "Daily Rate"
     ,main = "Histogram of Daily Rate")

boxplot(DailyRate,data=ds,col = "red")
```

Histogram of Daily Rate

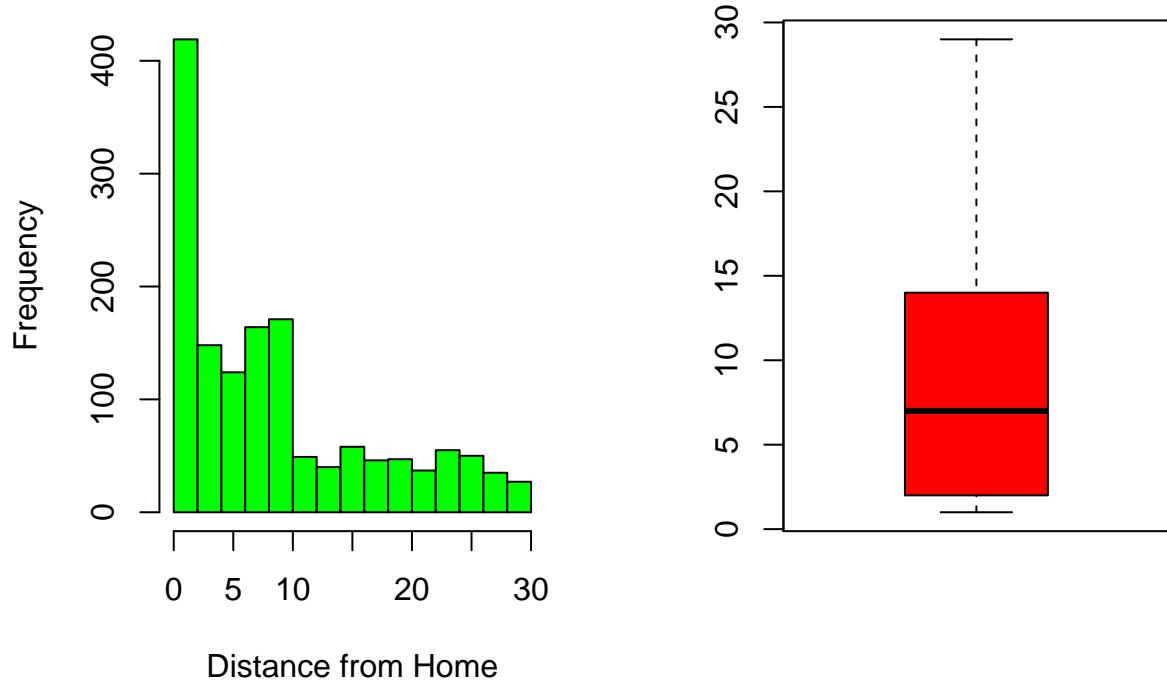


```
hist(ds$DistanceFromHome,col = "green",xlab = "Distance from Home"
```

```
,main = "Histogram of Distance from Home")
```

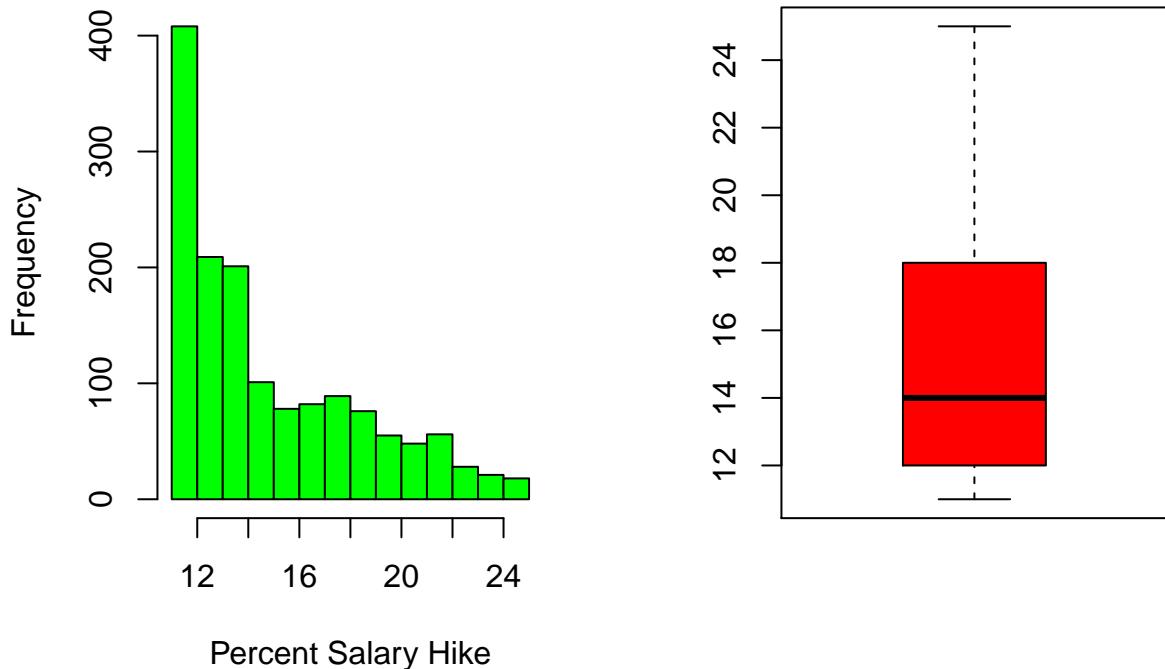
```
boxplot(DistanceFromHome,data=ds,col = "red")
```

Histogram of Distance from Hom



```
hist(ds$PercentSalaryHike,col = "green",xlab = "Percent Salary Hike"  
,main = "Histogram of Percent salary Hike")  
  
boxplot(PercentSalaryHike,data=ds,col = "red")
```

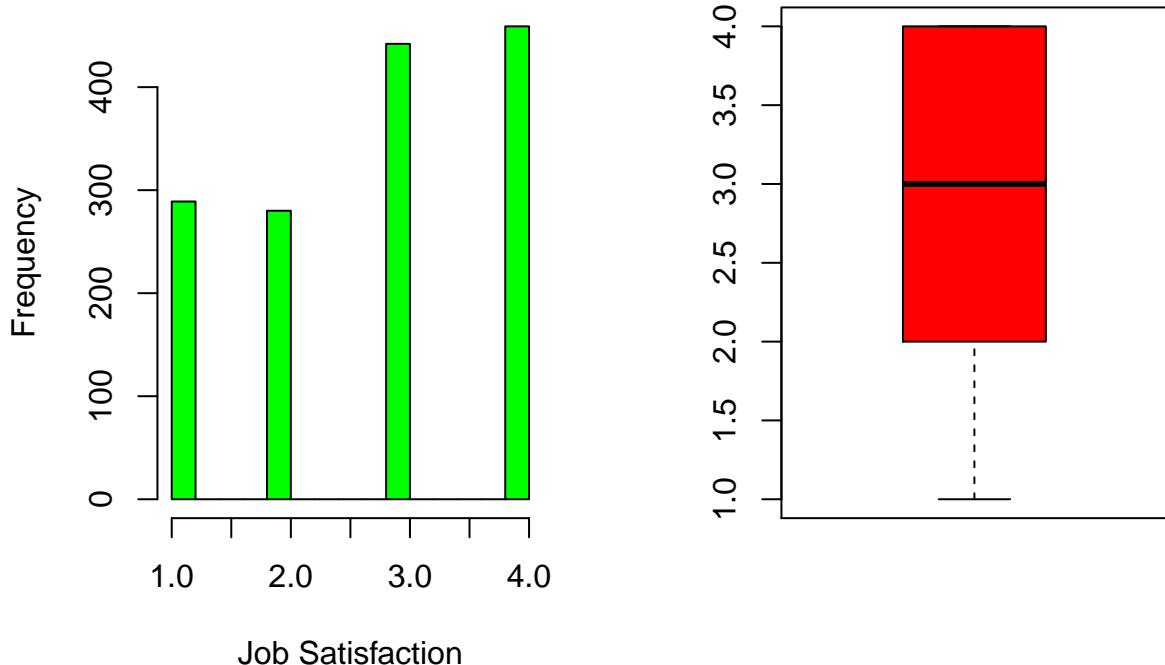
Histogram of Percent salary Hik



```
hist(ds$JobSatisfaction,col = "green",xlab = "Job Satisfaction"  
,main = "Histogram of job satisfaction")
```

```
boxplot(JobSatisfaction,data=ds,col = "red")
```

Histogram of job satisfaction

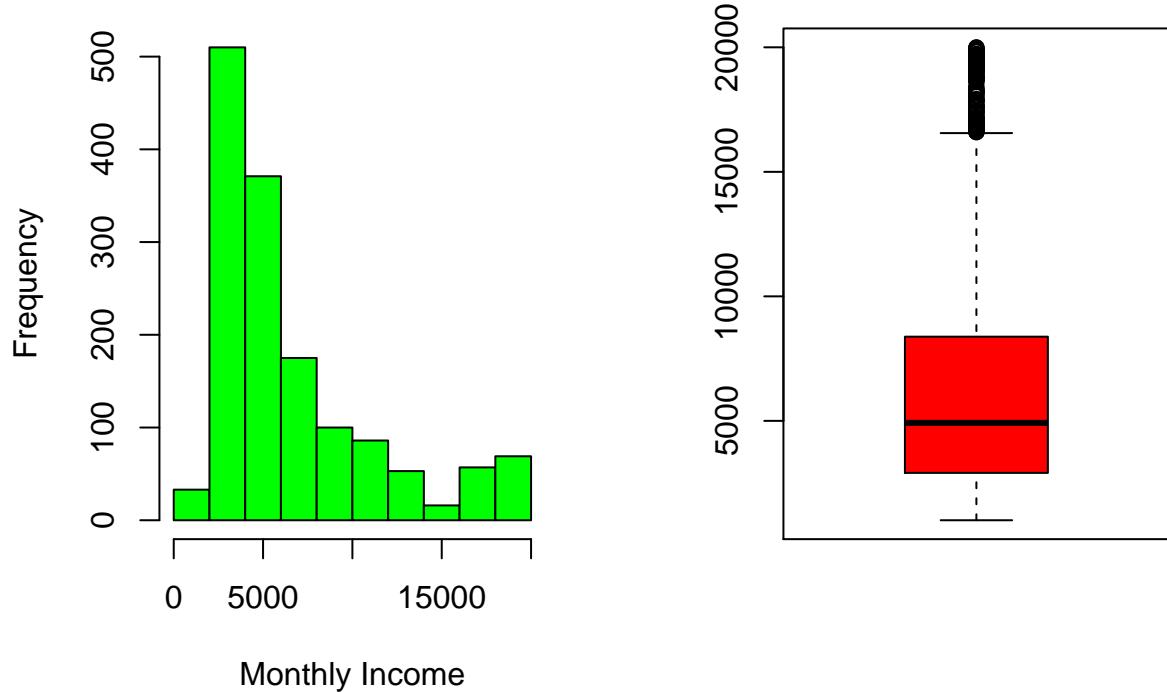


```
hist(ds$MonthlyIncome,col = "green",xlab="Monthly Income"
```

```
,main = "Histogram of Monthly Income")
```

```
boxplot(MonthlyIncome,data=ds,col = "red")
```

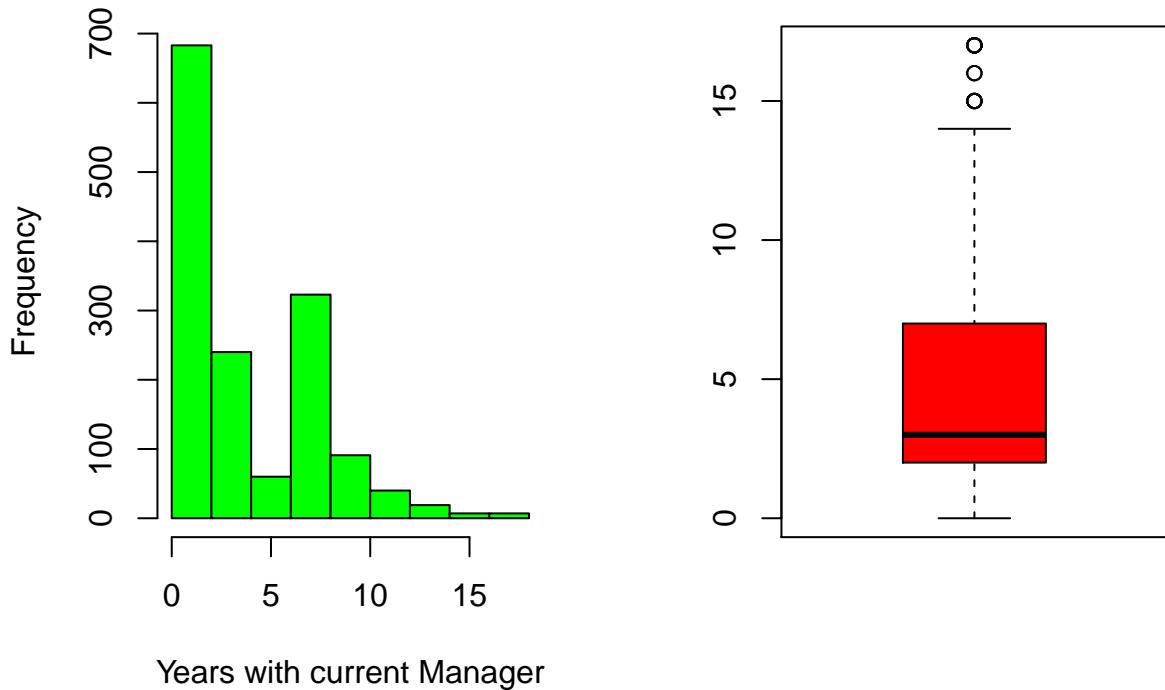
Histogram of Monthly Income



```
hist(ds$YearsWithCurrManager,col = "green",xlab = "Years with current Manager"  
,main = "Histogram of Years with Current Manager")
```

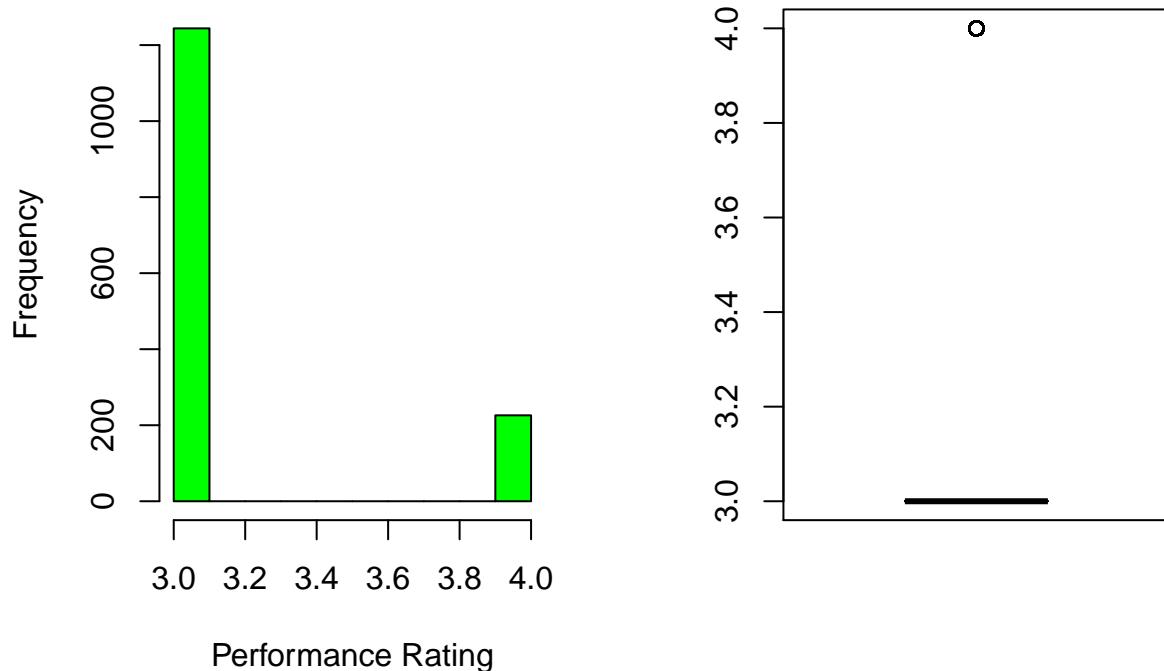
```
boxplot(YearsWithCurrManager,data=ds,col = "red")
```

Histogram of Years with Current Manager



```
hist(ds$PerformanceRating,col = "green",xlab = "Performance Rating"  
,main="Histogram of Performance Rating")  
  
boxplot(PerformanceRating,data=ds,col = "red")
```

Histogram of Performance Rating



Scatter Plots

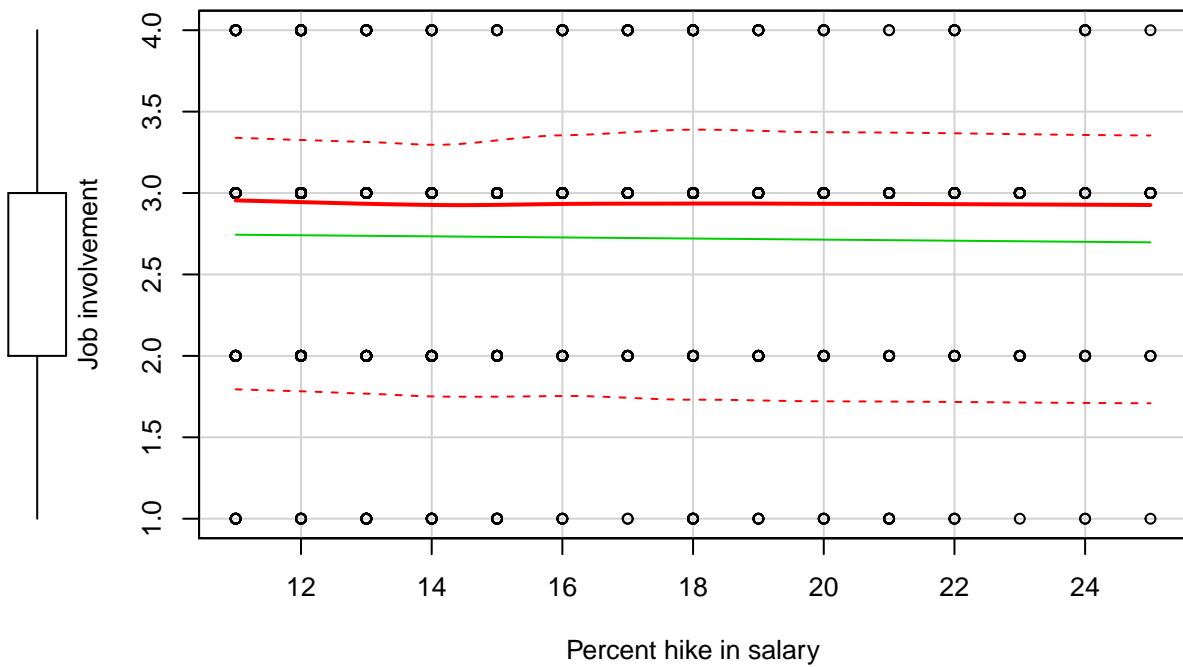
```
library(car)

##
## Attaching package: 'car'
## The following object is masked from 'package:psych':
##      logit
#Monthly income with years at company
scatterplot(MonthlyIncome~YearsAtCompany,data=ds,main="Distribution of monthly
income with work experience",ylab="Monthly Income"
,xlab = "Years at company")
```



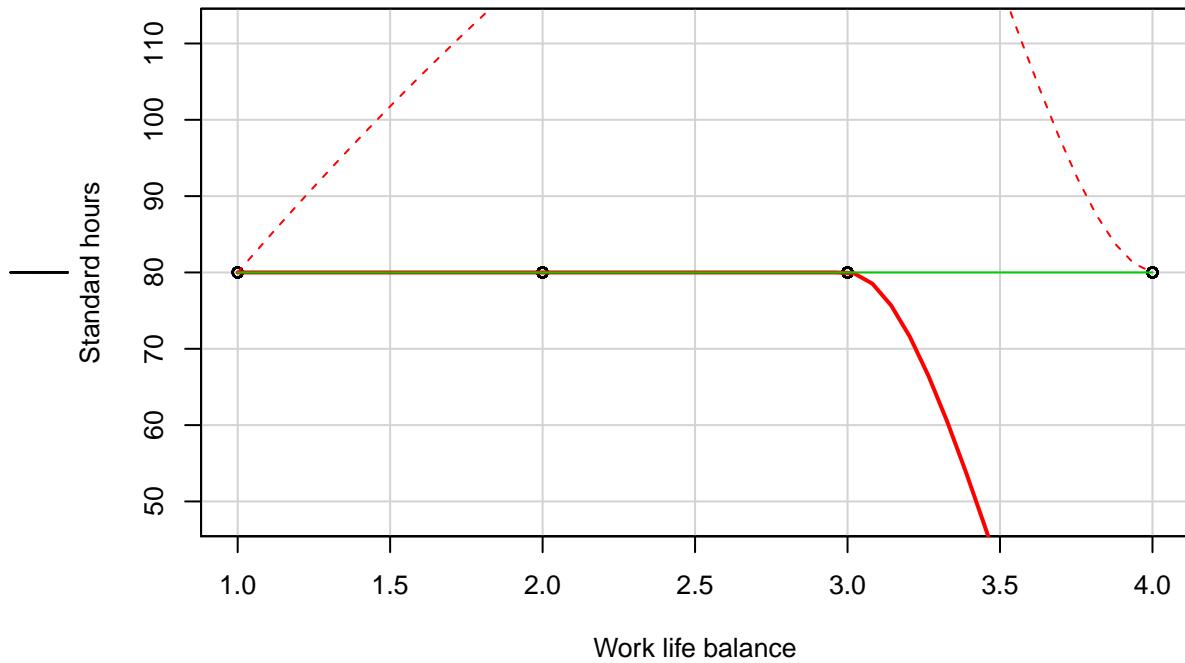
```
#Job Involvement with Percent Salary Hike
scatterplot(JobInvolvement~PercentSalaryHike,data=ds,main="Distribution of
job involvement with percent hike in salary"
,ylab="Job involvement",xlab = "Percent hike in salary")
```

Distribution of job involvement with percent hike in salary



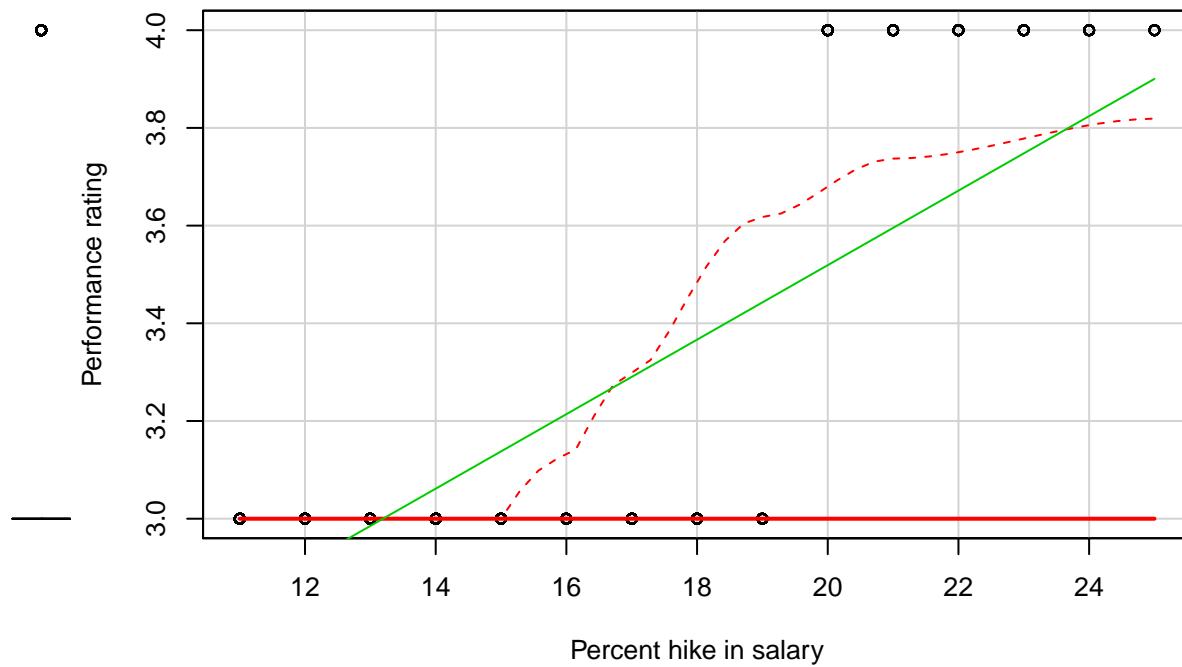
```
#Standard Hours with Work life Balance
scatterplot(StandardHours~WorkLifeBalance,data=ds,main="Distribution of Standard hours
with work life balance",ylab="Standard hours",xlab = "Work life balance")
```

Distribution of Standard hours with work life balance



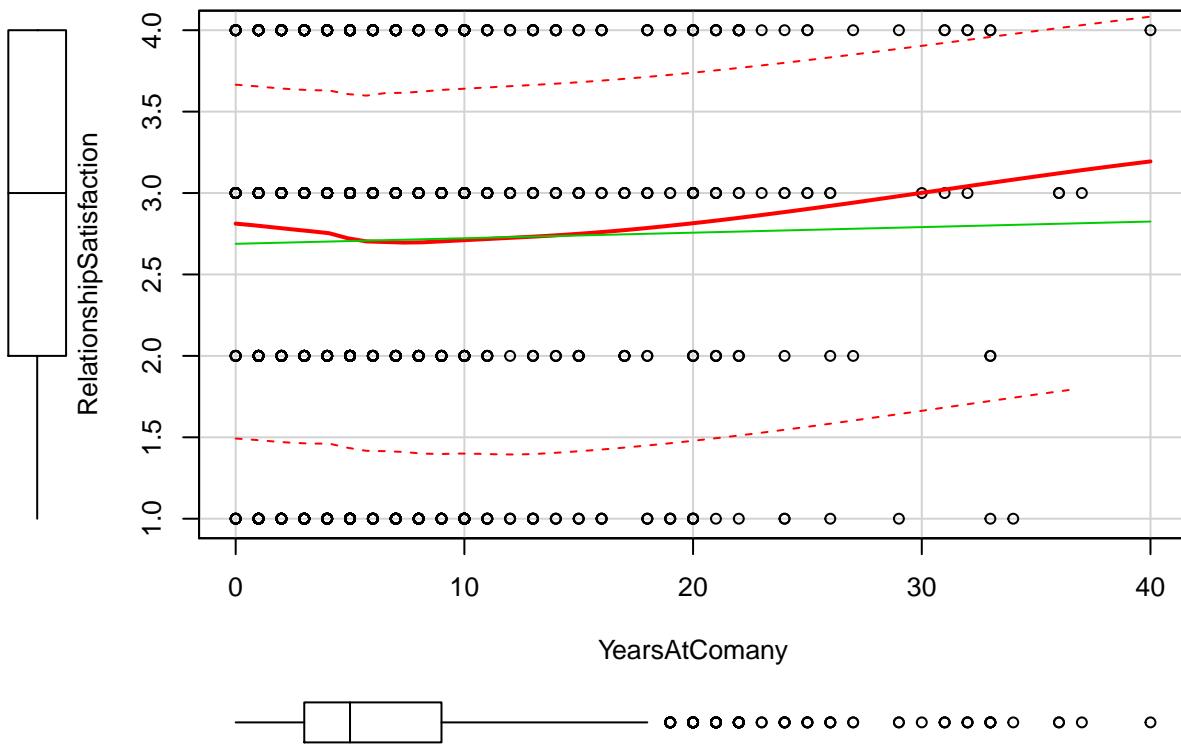
```
#Performance Rating with Percent hike in salary
scatterplot(PerformanceRating~PercentSalaryHike,data=ds,main="Distribution of
    performance rating with percent hike in salary",ylab="Performance rating"
    ,xlab = "Percent hike in salary")
```

Distribution of performance rating with percent hike in salary



```
#Relationship Satisfaction with Years at company
scatterplot(RelationshipSatisfaction~YearsAtCompany,data=ds,main="Distribution of
relationship satisfaction with years at company",
ylab="RelationshipSatisfaction",xlab = "YearsAtComany")
```

Distribution of relationship satisfaction with years at company



Dependency of Attrition on different variables

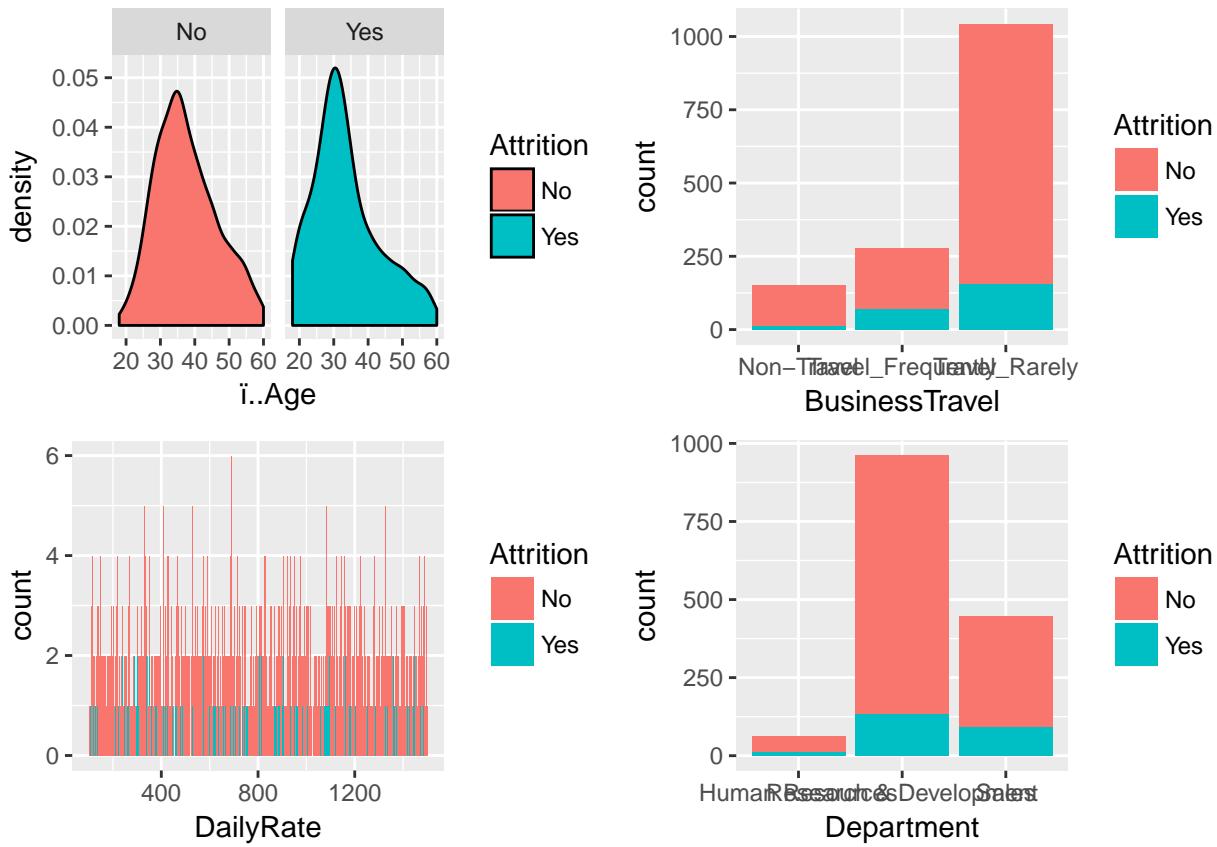
```

library(ggplot2)

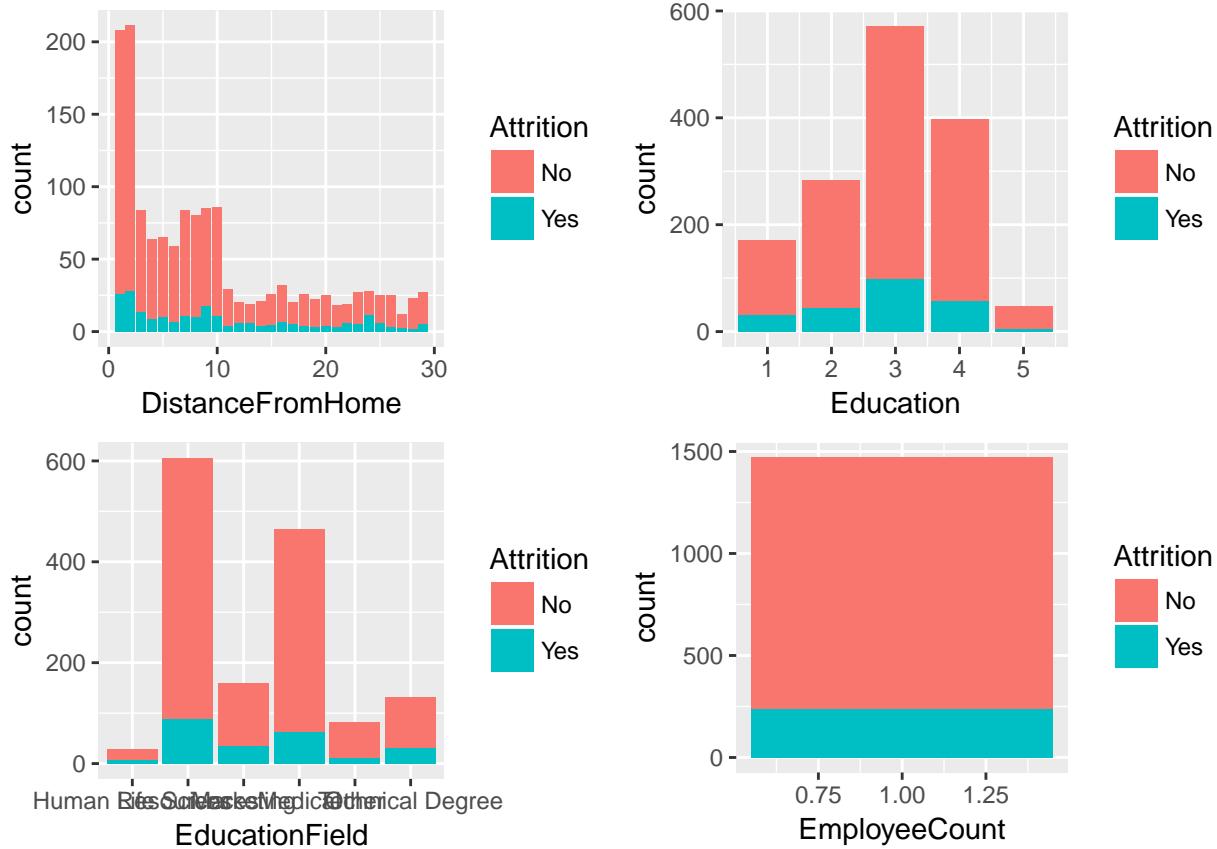
##
## Attaching package: 'ggplot2'
## The following objects are masked from 'package:psych':
##     %+%, alpha
library(grid)
library(gridExtra)
a<-ggplot(ds,aes(i..Age,fill=Attrition))+geom_density()+facet_grid(~Attrition)
b<-ggplot(ds,aes(BusinessTravel,fill=Attrition))+geom_bar()
c<-ggplot(ds,aes(DailyRate,fill=Attrition))+geom_bar()
d<-ggplot(ds,aes(Department,fill=Attrition))+geom_bar()
grid.arrange(a,b,c,d,ncol=2)

## Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x$label)): font
## metrics unknown for character 0x11

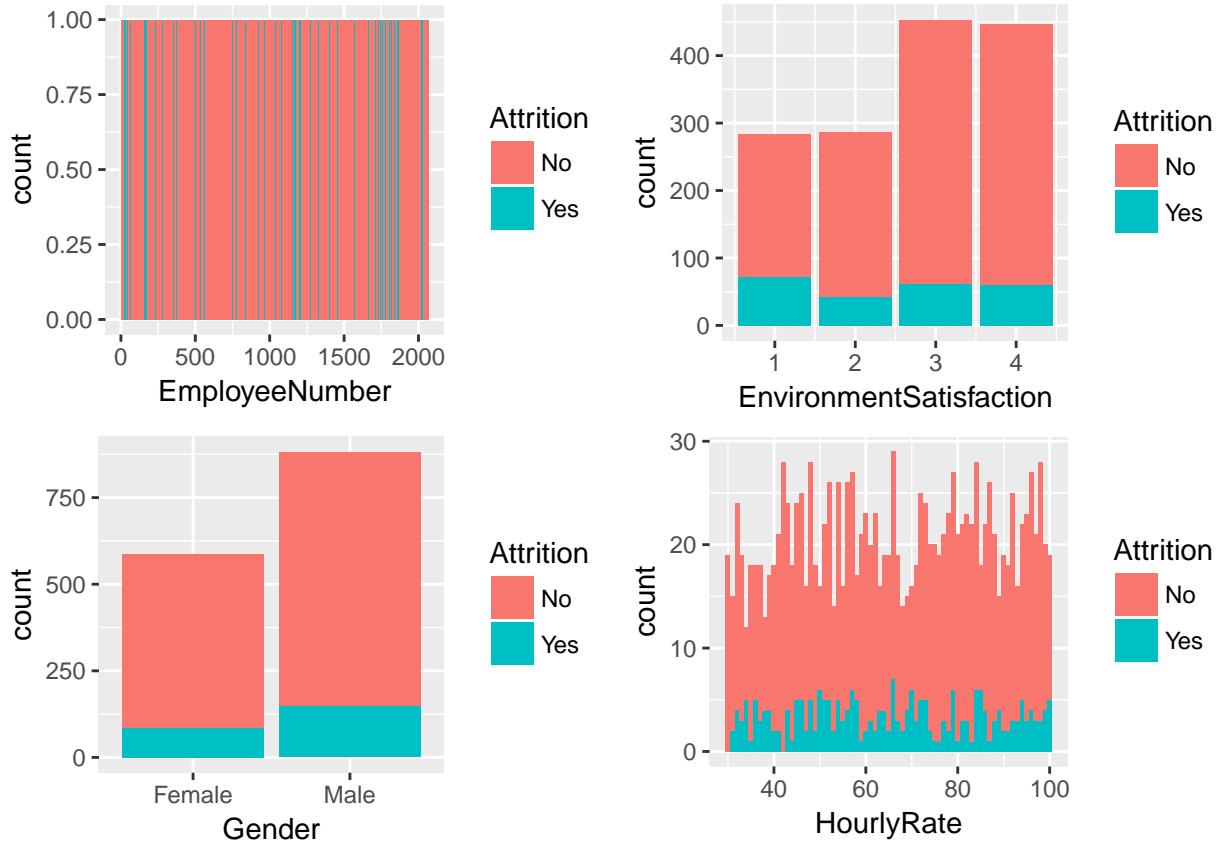
```



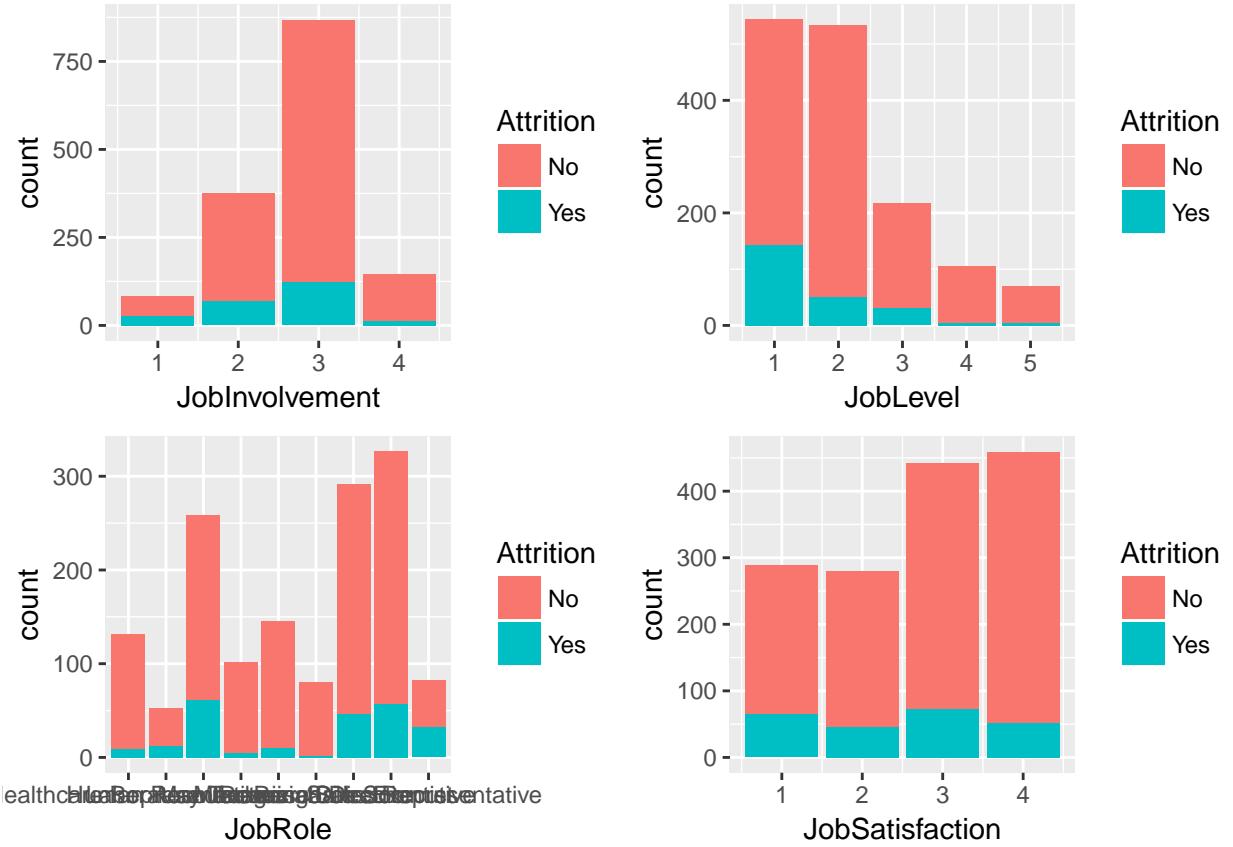
```
e<-ggplot(ds,aes(DistanceFromHome,fill=Attrition))+geom_bar()
f<-ggplot(ds,aes(Education,fill=Attrition))+geom_bar()
g<-ggplot(ds,aes(EducationField,fill=Attrition))+geom_bar()
h<-ggplot(ds,aes(EmployeeCount,fill=Attrition))+geom_bar()
grid.arrange(e,f,g,h,ncol=2)
```



```
i<-ggplot(ds,aes(EmployeeNumber,fill=Attrition))+geom_bar()
j<-ggplot(ds,aes(EnvironmentSatisfaction,fill=Attrition))+geom_bar()
k<-ggplot(ds,aes(Gender,fill=Attrition))+geom_bar()
l<-ggplot(ds,aes(HourlyRate,fill=Attrition))+geom_bar()
grid.arrange(i,j,k,l,ncol=2)
```



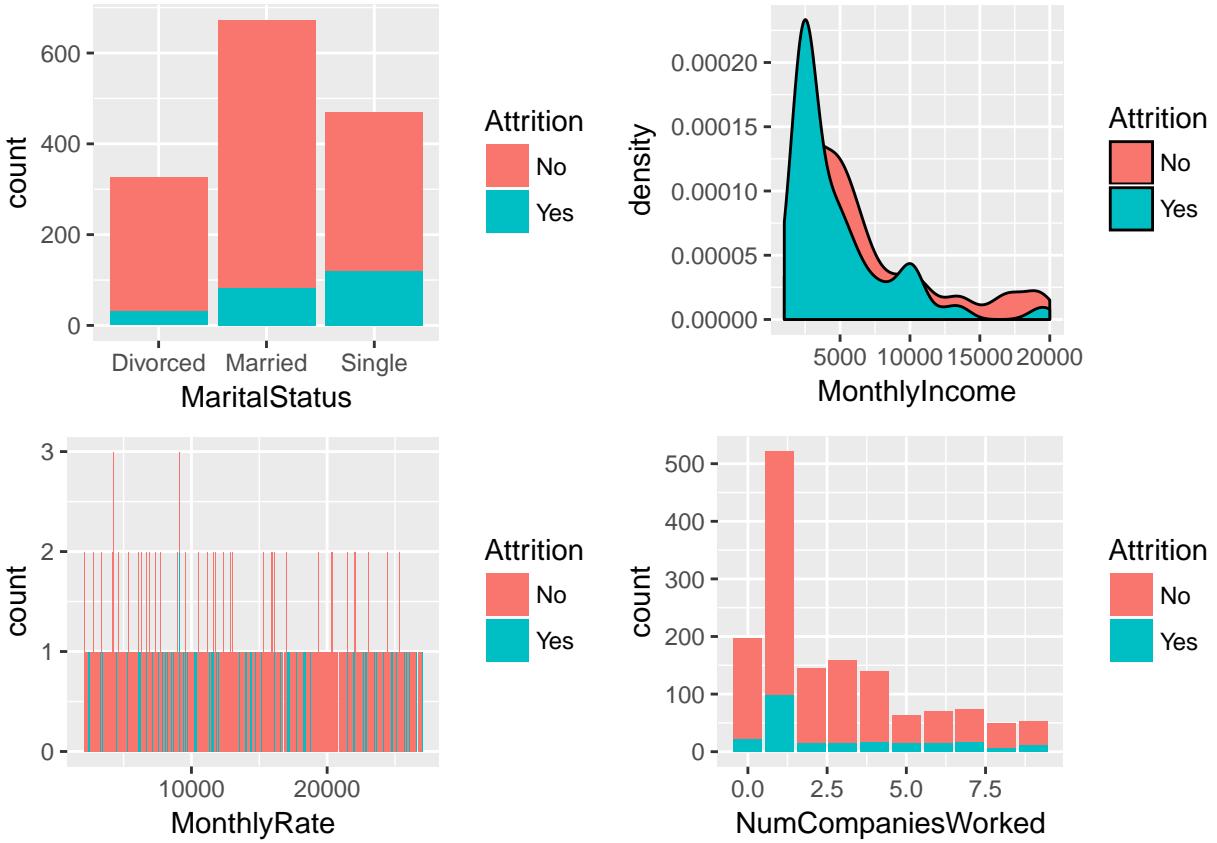
```
m<-ggplot(ds,aes(JobInvolvement,fill=Attrition))+geom_bar()
n<-ggplot(ds,aes(JobLevel,fill=Attrition))+geom_bar()
o<-ggplot(ds,aes(JobRole,fill=Attrition))+geom_bar()
p<-ggplot(ds,aes(JobSatisfaction,fill=Attrition))+geom_bar()
grid.arrange(m,n,o,p,ncol=2)
```



```

q<-ggplot(ds,aes(MaritalStatus,fill=Attrition))+geom_bar()
r<-ggplot(ds,aes(MonthlyIncome,fill=Attrition))+geom_density()
s<-ggplot(ds,aes(MonthlyRate,fill=Attrition))+geom_bar()
t<-ggplot(ds,aes(NumCompaniesWorked,fill=Attrition))+geom_bar()
grid.arrange(q,r,s,t,ncol=2)

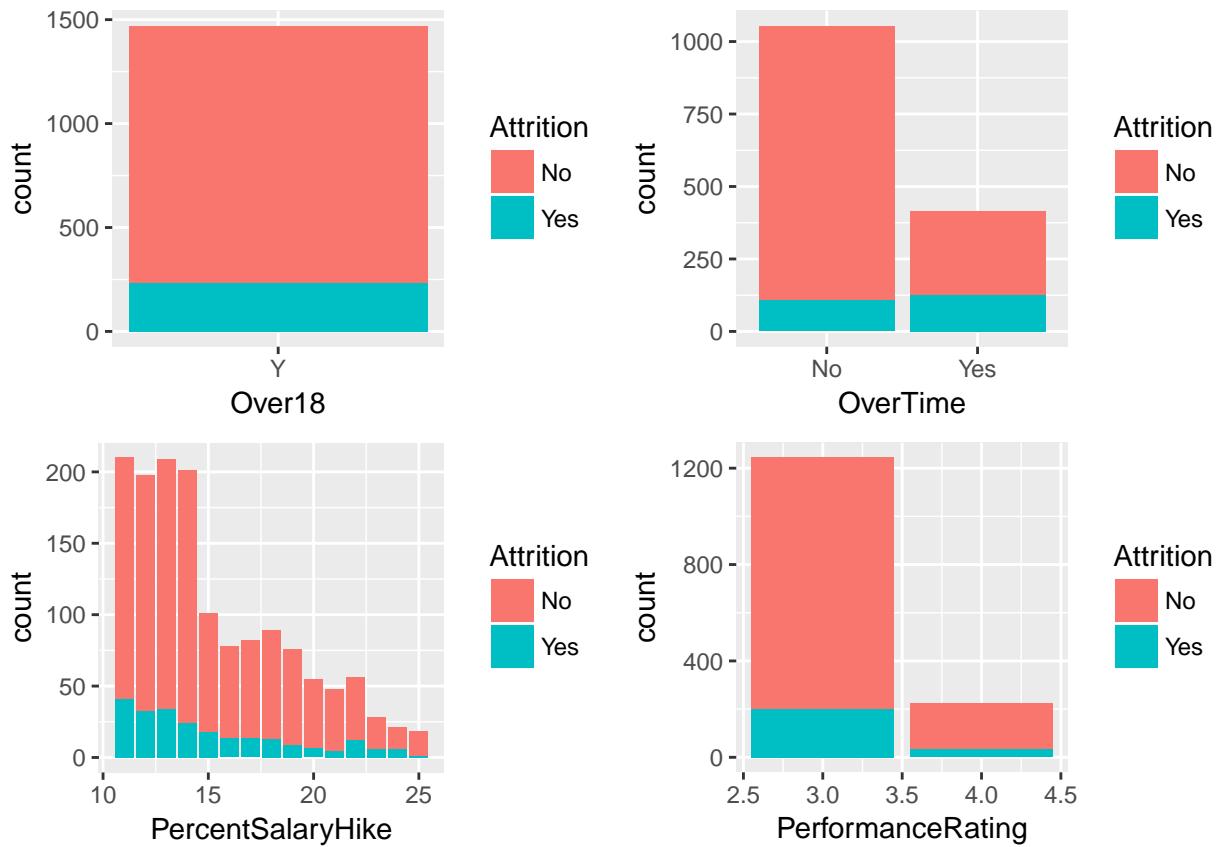
```



```

u<-ggplot(ds,aes(Over18,fill=Attrition))+geom_bar()
v<-ggplot(ds,aes(OverTime,fill=Attrition))+geom_bar()
w<-ggplot(ds,aes(PercentSalaryHike,fill=Attrition))+geom_bar()
x<-ggplot(ds,aes(PerformanceRating,fill=Attrition))+geom_bar()
grid.arrange(u,v,w,x,ncol=2)

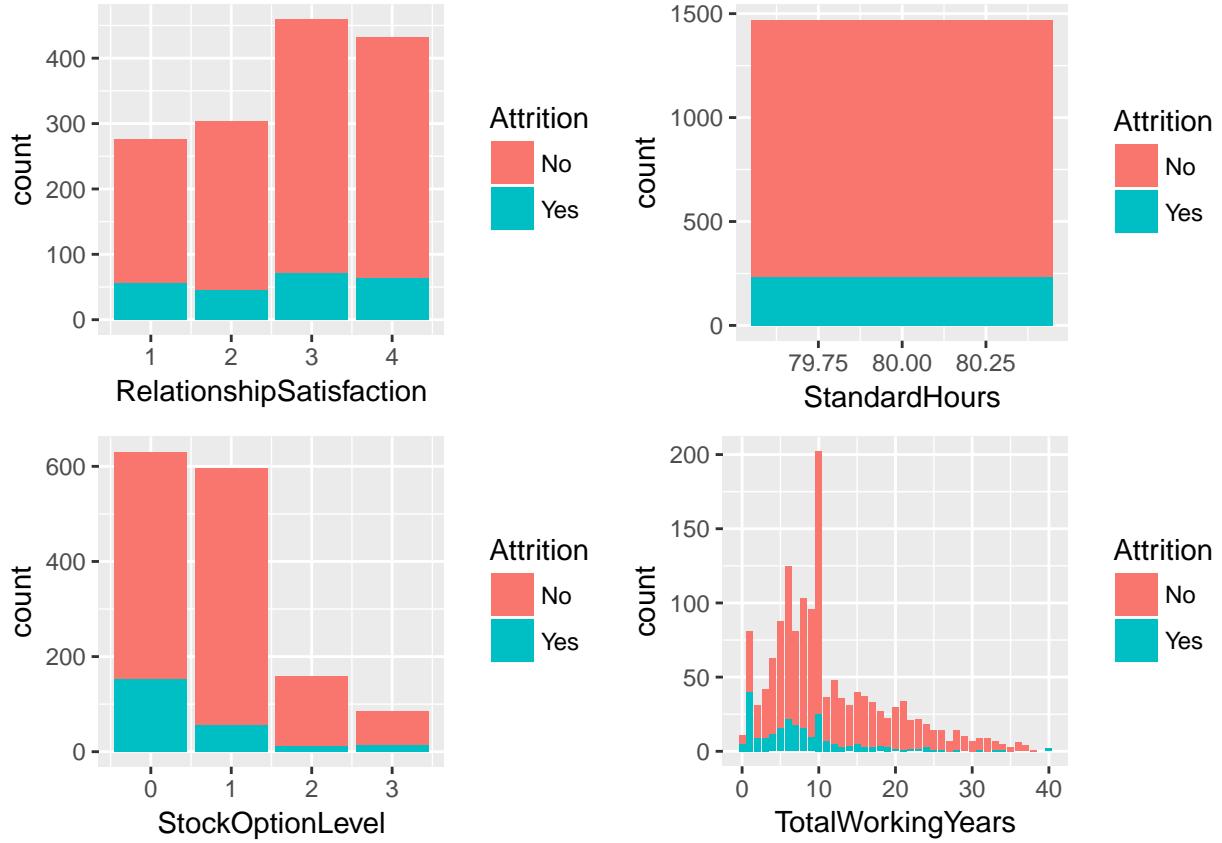
```



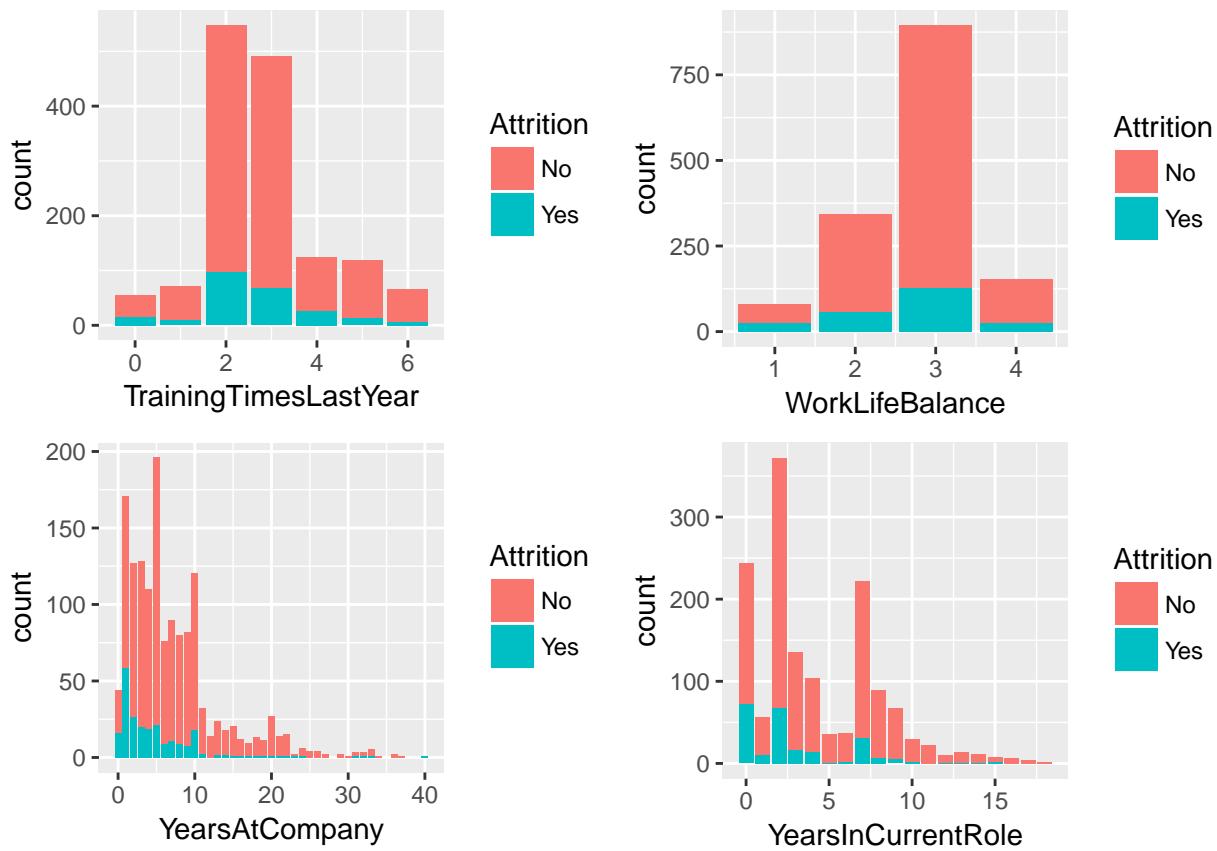
```

y<-ggplot(ds,aes(RelationshipSatisfaction,fill=Attrition))+geom_bar()
z<-ggplot(ds,aes(StandardHours,fill=Attrition))+geom_bar()
a1<-ggplot(ds,aes(StockOptionLevel,fill=Attrition))+geom_bar()
a2<-ggplot(ds,aes(TotalWorkingYears,fill=Attrition))+geom_bar()
grid.arrange(y,z,a1,a2,ncol=2)

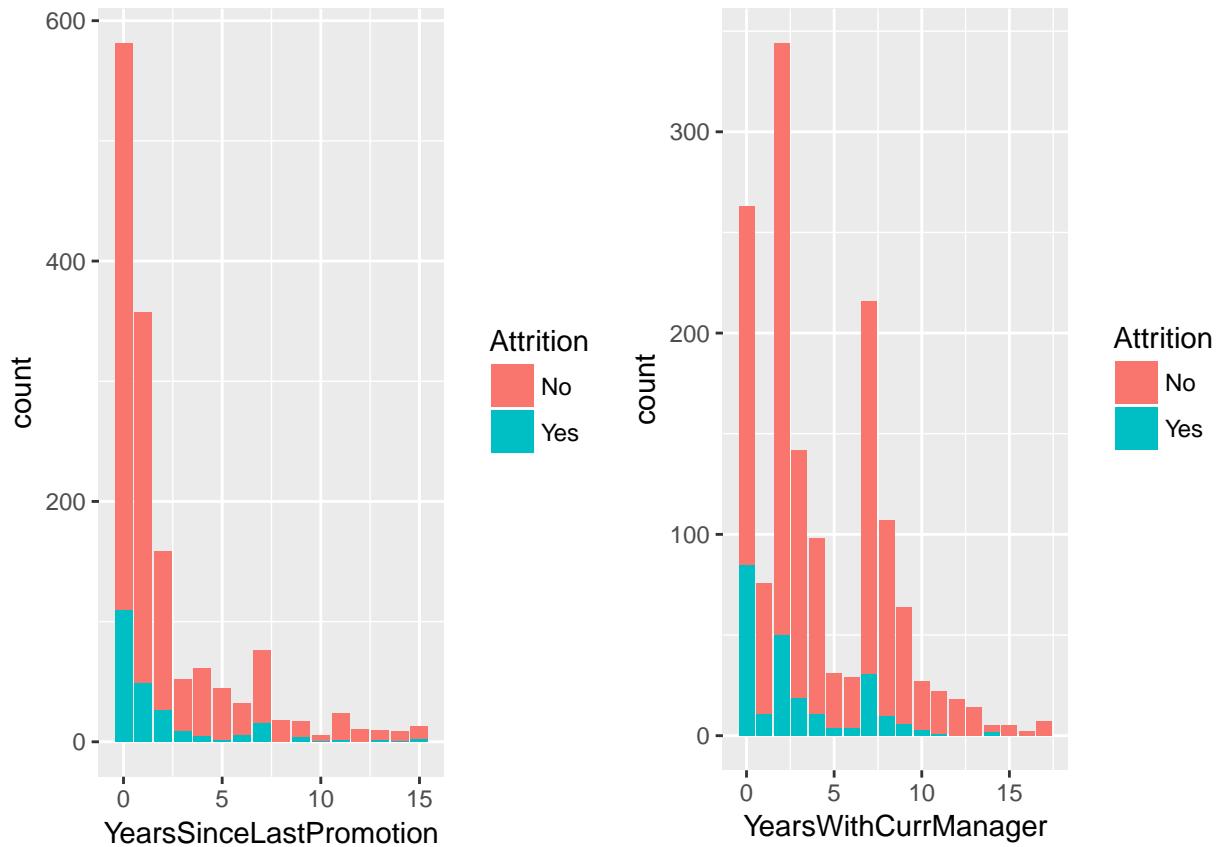
```



```
a3<-ggplot(ds,aes(TrainingTimesLastYear,fill=Attrition))+geom_bar()
a4<-ggplot(ds,aes(WorkLifeBalance,fill=Attrition))+geom_bar()
a5<-ggplot(ds,aes(YearsAtCompany,fill=Attrition))+geom_bar()
a6<-ggplot(ds,aes(YearsInCurrentRole,fill=Attrition))+geom_bar()
grid.arrange(a3,a4,a5,a6,ncol=2)
```



```
a7<-ggplot(ds,aes(YearsSinceLastPromotion,fill=Attrition))+geom_bar()
a8<-ggplot(ds,aes(YearsWithCurrManager,fill=Attrition))+geom_bar()
grid.arrange(a7,a8,ncol=2)
```



From the above plots some points which can be inferred are as follows:

1. Employee with age around 30 leave most.
2. Employee who travels the most leave the most.
3. Less number of people attrited from HR department.
4. Employees who have left the organization are near to the Office.
5. As the Job Level increases the number of people quitting decreases.
6. Lower Job Satisfaction levels higher attrition.
7. Attrition is on higher side for Single.
8. High level of attrition among employees having low income.
9. Larger Proportion of Overtime Employees are quitting.
10. Larger proportions of 1 & 2 rating are quitting.
11. Larger proportions of levels 1 & 2 quit.
12. Larger proportion of new comers are quitting the organization.
13. New Manager is a big cause for quitting.

Correlation Matrix

```
library(corrplot)

## corrplot 0.84 loaded

frame <- ds[,c("i..Age", "DailyRate", "DistanceFromHome", "Education", "EmployeeCount",
             "EnvironmentSatisfaction", "HourlyRate", "JobInvolvement", "JobLevel",
             "JobSatisfaction", "MonthlyIncome", "MonthlyRate", "NumCompaniesWorked",
             "PercentSalaryHike", "PerformanceRating", "RelationshipSatisfaction",
             "StandardHours", "StockOptionLevel", "TotalWorkingYears",
             "TrainingTimesLastYear", "WorkLifeBalance", "YearsAtCompany",
             "YearsInCurrentRole", "YearsSinceLastPromotion", "YearsWithCurrManager")]

round(cor(frame), 2)

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

##          i..Age DailyRate DistanceFromHome Education
## i..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
## i..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
## i..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
## i..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		
## i..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		
## i..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
## i..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
## i..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
## i..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
## i..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

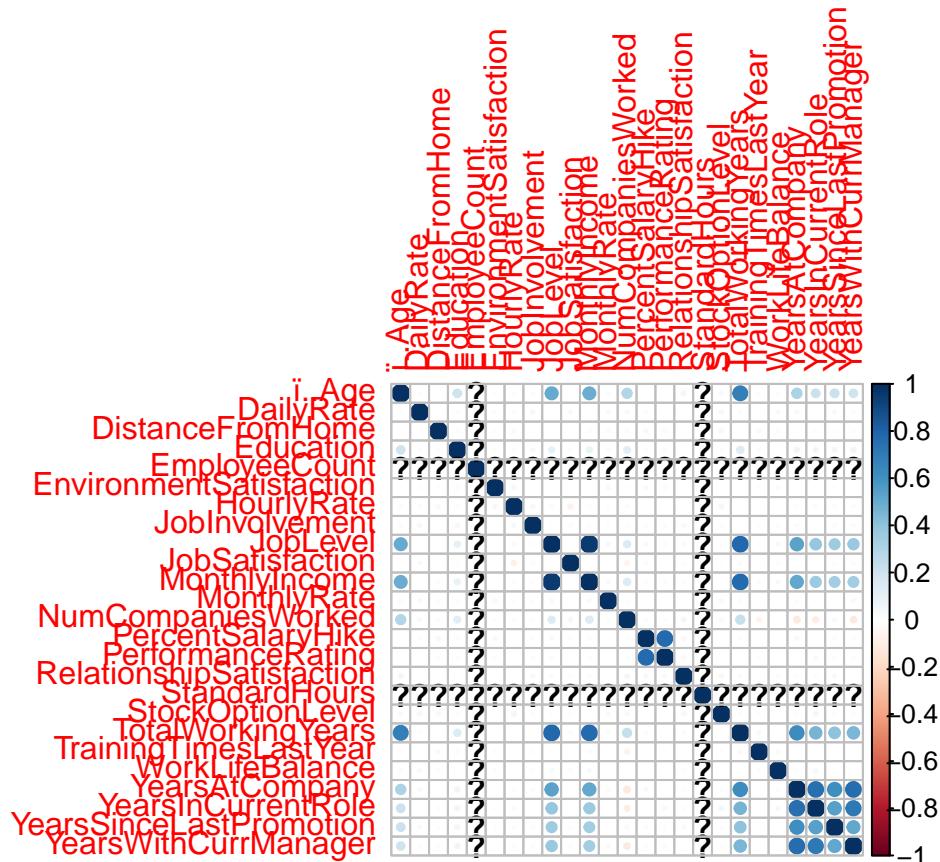
```

## TrainingTimesLastYear          0.00
## WorkLifeBalance               0.01
## YearsAtCompany                0.62
## YearsInCurrentRole            0.55
## YearsSinceLastPromotion       1.00
## YearsWithCurrManager          0.51

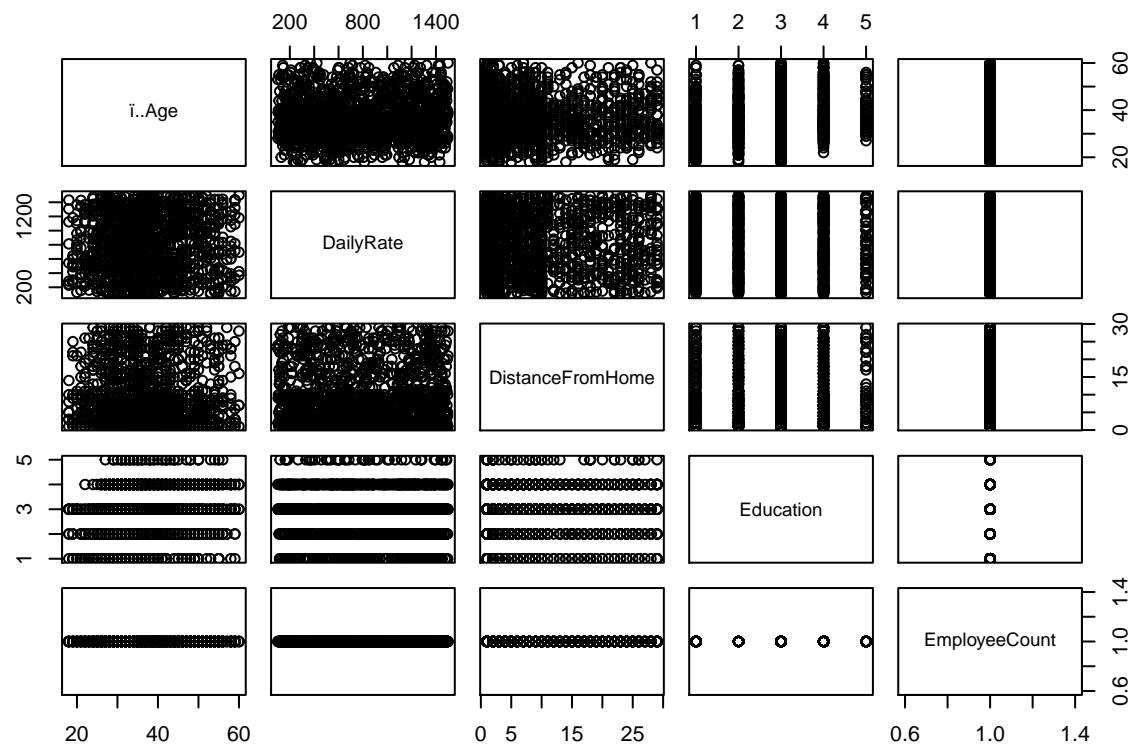
corrplot(cor(frame))

```

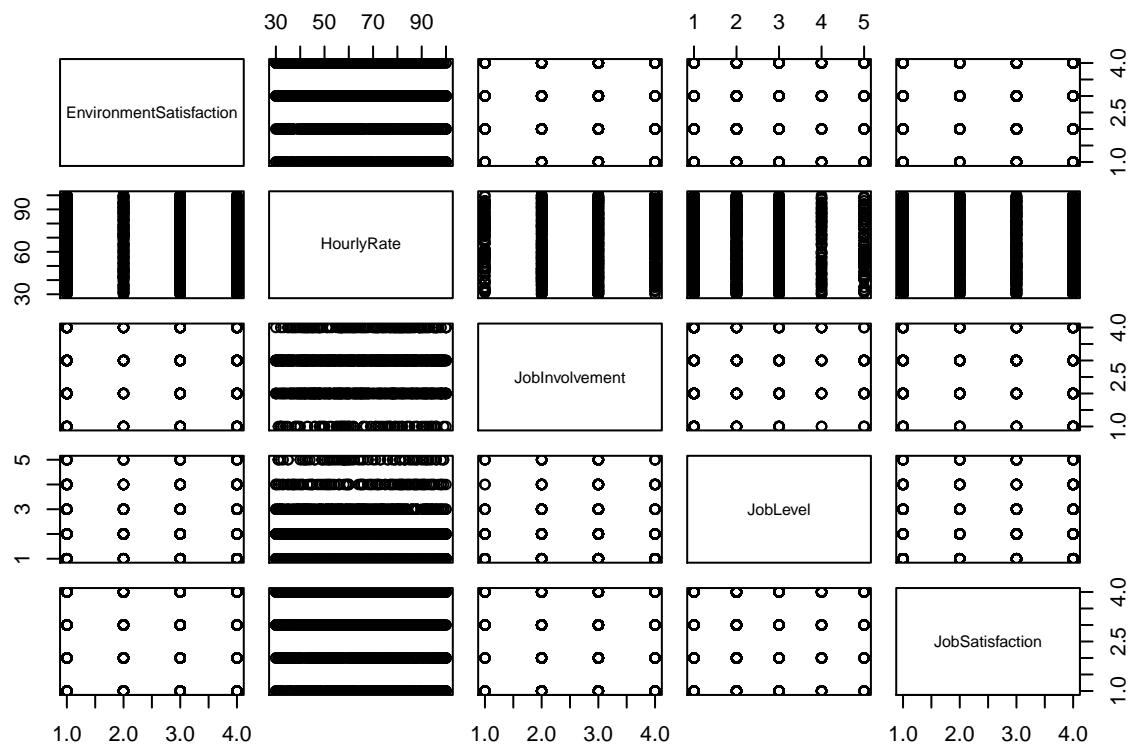
Warning in cor(frame): the standard deviation is zero



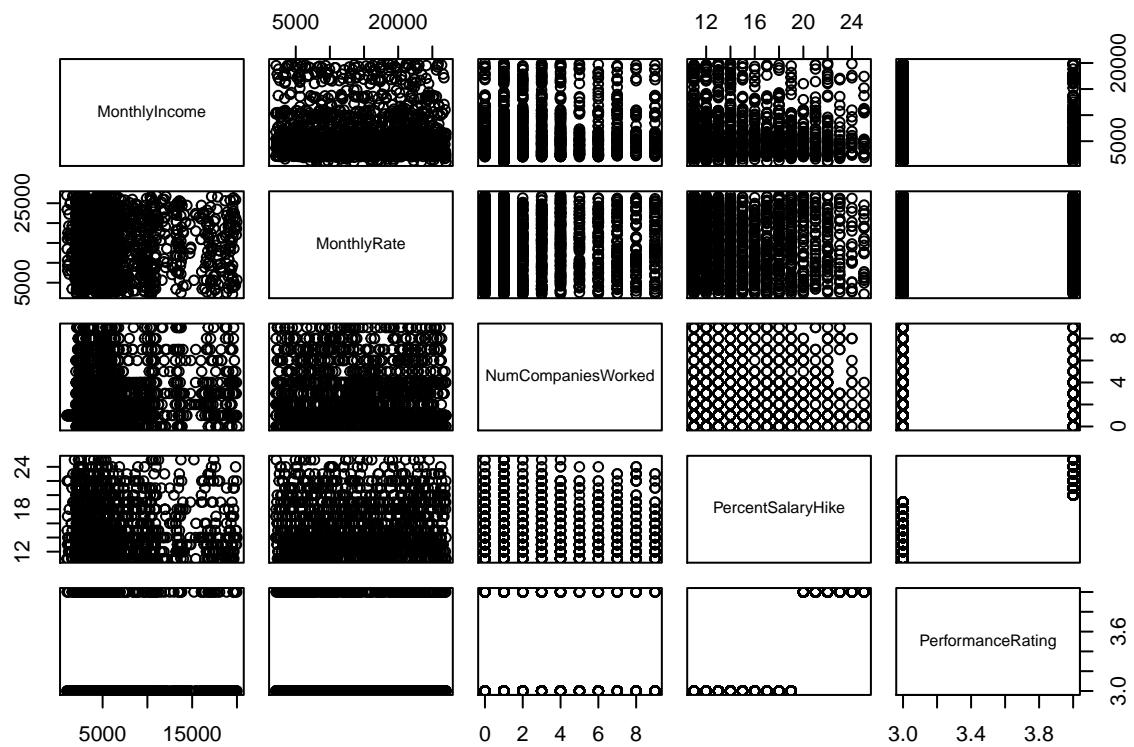
```
pairs(frame[1:5])
```



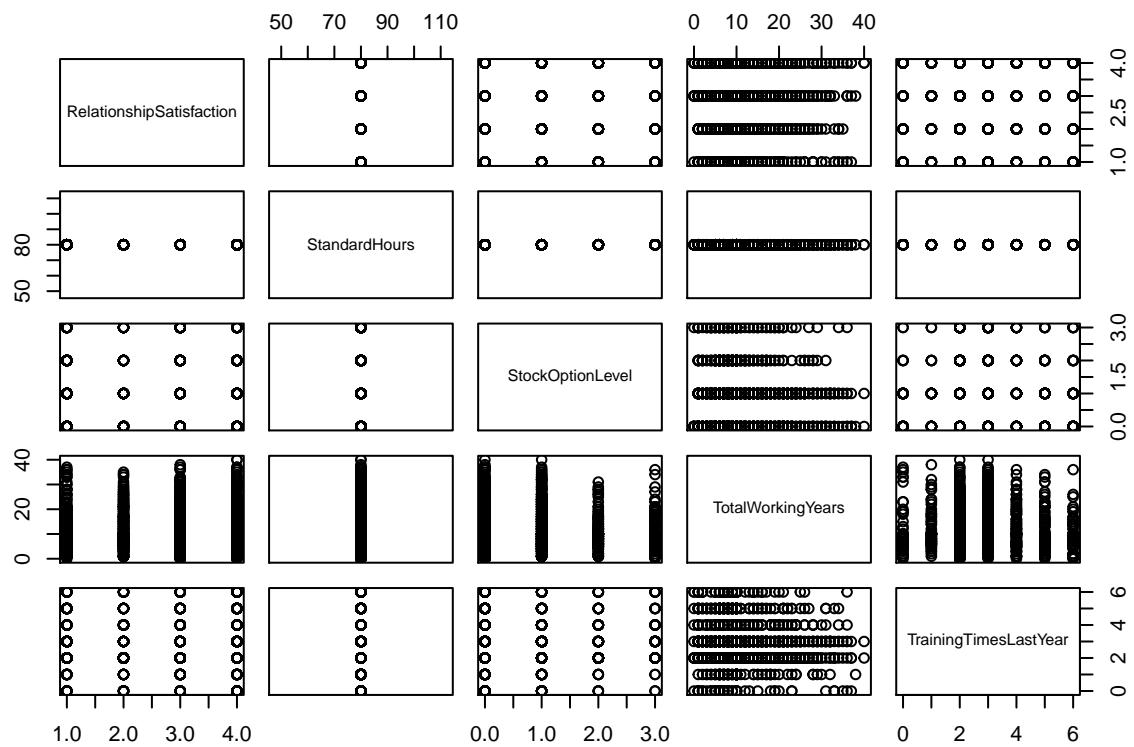
```
pairs(frame[6:10])
```



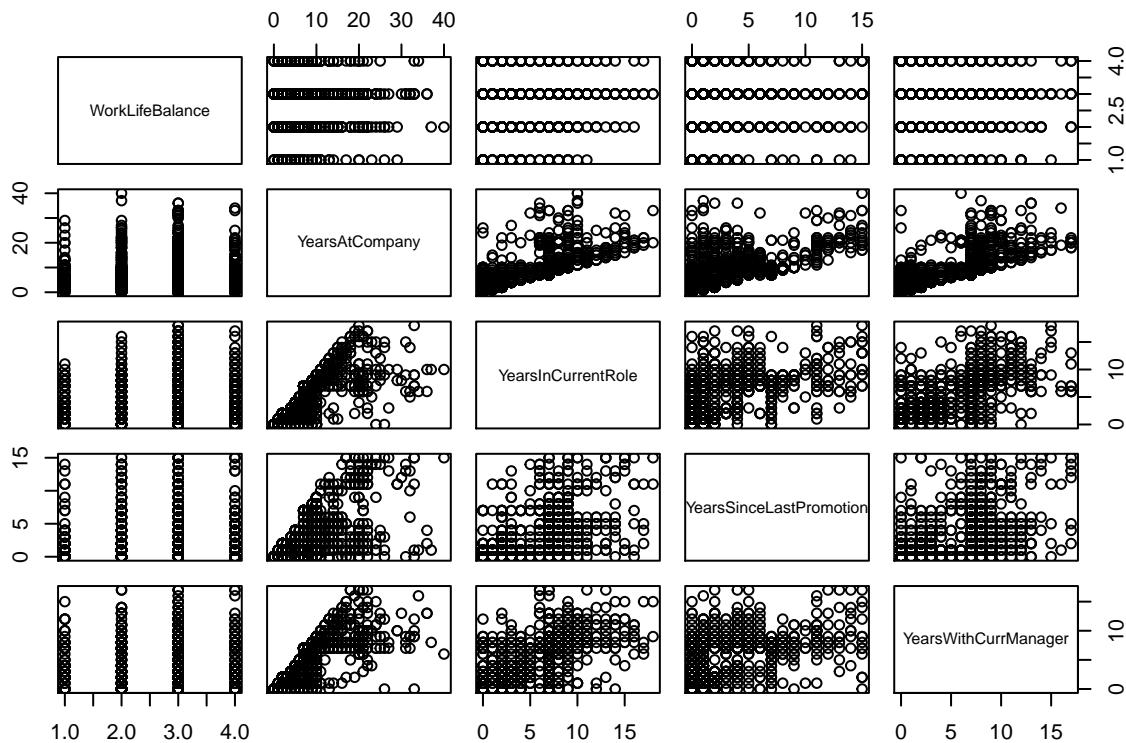
```
pairs(frame[11:15])
```



```
pairs(frame[16:20])
```



```
pairs(frame[21:25])
```



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

T test and Chi Square test

Hypothesis 1: Let there is no significant effect of manager on attrition

```
t.test(YearsWithCurrManager~Attrition,data=ds)

##
##  Welch Two Sample t-test
##
## data: YearsWithCurrManager by Attrition
## t = 6.6334, df = 365.1, p-value = 1.185e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.065929 1.964223
## sample estimates:
## mean in group No mean in group Yes
##          4.367397           2.852321
```

From the above t test we can justify by looking at the p value that there is a significant relationship between attrition and number of working years with the current manager. So we can reject our Hypothesis 1.

Hypothesis 2:The monthly income is not dependent on the performance rating of the employee.

```
t.test(MonthlyIncome~PerformanceRating,data=ds)

##
##  Welch Two Sample t-test
##
## data: MonthlyIncome by PerformanceRating
## t = 0.66007, df = 313.96, p-value = 0.5097
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -442.4770 889.2376
## sample estimates:
## mean in group 3 mean in group 4
##       6537.274      6313.894
```

From the above t test we can justify that there is no significant dependency between monthly income and performance rating. As the p value is more than 0.05 we can't reject our hypothesis 2.

```
t.test(PerformanceRating~Attrition,data=ds)
```

```
##
##  Welch Two Sample t-test
##
## data: PerformanceRating by Attrition
## t = -0.10999, df = 331.22, p-value = 0.9125
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.05350780 0.04784086
## sample estimates:
## mean in group No mean in group Yes
##       3.153285      3.156118
```

```
t.test(JobSatisfaction~Attrition,data=ds)
```

```
##
##  Welch Two Sample t-test
##
## data: JobSatisfaction by Attrition
## t = 3.9261, df = 328.59, p-value = 0.0001052
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.1547890 0.4656797
## sample estimates:
## mean in group No mean in group Yes
##       2.778589      2.468354
```

```
t.test(MonthlyIncome~Attrition,data=ds)
```

```
##
##  Welch Two Sample t-test
##
## data: MonthlyIncome by Attrition
## t = 7.4826, df = 412.74, p-value = 4.434e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```

## 1508.244 2583.050
## sample estimates:
## mean in group No mean in group Yes
## 6832.740 4787.093
t.test(YearsAtCompany~Attrition,data=ds)

##
## Welch Two Sample t-test
##
## data: YearsAtCompany by Attrition
## t = 5.2826, df = 338.21, p-value = 2.286e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 1.404805 3.071629
## sample estimates:
## mean in group No mean in group Yes
## 7.369019 5.130802
chisq.test(ds$JobRole,ds$Attrition)

##
## Pearson's Chi-squared test
##
## data: ds$JobRole and ds$Attrition
## X-squared = 86.19, df = 8, p-value = 2.752e-15
chisq.test(ds$RelationshipSatisfaction,ds$Attrition)

##
## Pearson's Chi-squared test
##
## data: ds$RelationshipSatisfaction and ds$Attrition
## X-squared = 5.2411, df = 3, p-value = 0.155
chisq.test(ds$YearsSinceLastPromotion,ds$Attrition)

## Warning in chisq.test(ds$YearsSinceLastPromotion, ds$Attrition): Chi-
## squared approximation may be incorrect
##
## Pearson's Chi-squared test
##
## data: ds$YearsSinceLastPromotion and ds$Attrition
## X-squared = 21.845, df = 15, p-value = 0.1119
chisq.test(ds$Education,ds$Attrition)

##
## Pearson's Chi-squared test
##
## data: ds$Education and ds$Attrition
## X-squared = 3.074, df = 4, p-value = 0.5455

```

From the above test we can also infer that:

1. Attrition is dependent on Job Satisfaction.
2. Attrition is dependent on Montly income.

3. Attrition is dependent on number of years an employee has worked in the company.
4. Attrition is dependent on the Job Role.
5. Attrition does not depend upon Performance Rating.
6. Attrition does not depend upon Relationship Satisfaction.
7. Attrition does not depend upon number of years since last promotion.
8. Attrition does not depend upon the level of education.

Linear Regression

MODEL 1

```

library(car)
ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
m1 <- Attrition ~ i..Age + BusinessTravel + DailyRate + Department + DistanceFromHome + Education +
  EducationField + EmployeeNumber + EnvironmentSatisfaction + Gender + HourlyRate + JobInvolvement +
  JobLevel + JobRole + JobSatisfaction + MaritalStatus + MonthlyIncome + MonthlyRate + NumCompaniesWorked +
  OverTime + PercentSalaryHike + PerformanceRating + RelationshipSatisfaction + StockOptionLevel +
  TotalWorkingYears + TrainingTimesLastYear + WorkLifeBalance +
  YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion + YearsWithCurrManager

f1 <- lm(m1, data=ds)
summary(f1)

##
## Call:
## lm(formula = m1, data = ds)
##
## Residuals:
##      Min        1Q        Median        3Q       Max
## -0.55266 -0.20551 -0.08396  0.08281  1.14588
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.563e+00  1.779e-01   8.784 < 2e-16
## i..Age                   -3.504e-03  1.327e-03  -2.640 0.008370
## BusinessTravelTravel_Frequently 1.523e-01  3.305e-02   4.609 4.41e-06
## BusinessTravelTravel_Rarely    6.561e-02  2.853e-02   2.300 0.021586
## DailyRate                 -2.698e-05  2.120e-05  -1.272 0.203414
## DepartmentResearch & Development 1.293e-01  1.171e-01   1.104 0.269643
## DepartmentSales              1.053e-01  1.211e-01   0.869 0.384814
## DistanceFromHome             3.624e-03  1.048e-03   3.457 0.000562
## Education                  1.909e-03  8.543e-03   0.223 0.823252
## EducationFieldLife Sciences -1.225e-01  8.376e-02  -1.462 0.143969
## EducationFieldMarketing     -8.209e-02  8.923e-02  -0.920 0.357706
## EducationFieldMedical       -1.344e-01  8.409e-02  -1.598 0.110168
## EducationFieldOther          -1.443e-01  8.995e-02  -1.604 0.108977
## EducationFieldTechnical Degree -2.674e-02  8.748e-02  -0.306 0.759905
## EmployeeNumber               -7.553e-06  1.420e-05  -0.532 0.594843
## EnvironmentSatisfaction     -4.040e-02  7.800e-03  -5.179 2.55e-07

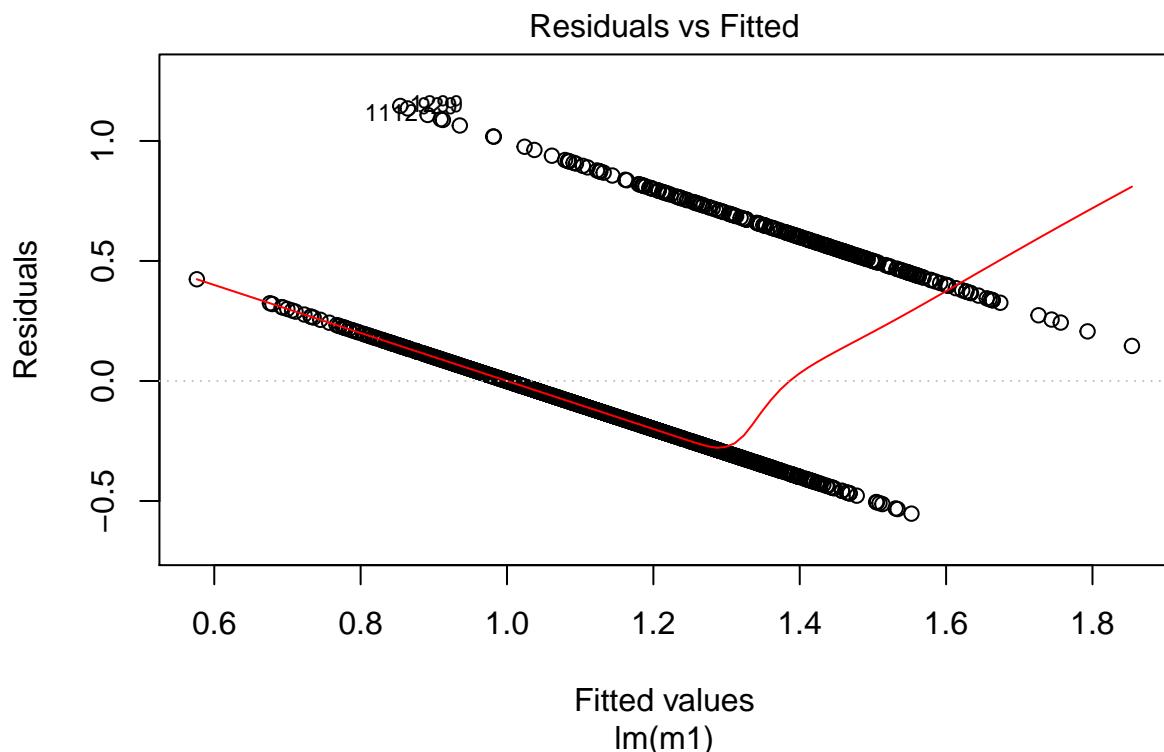
```

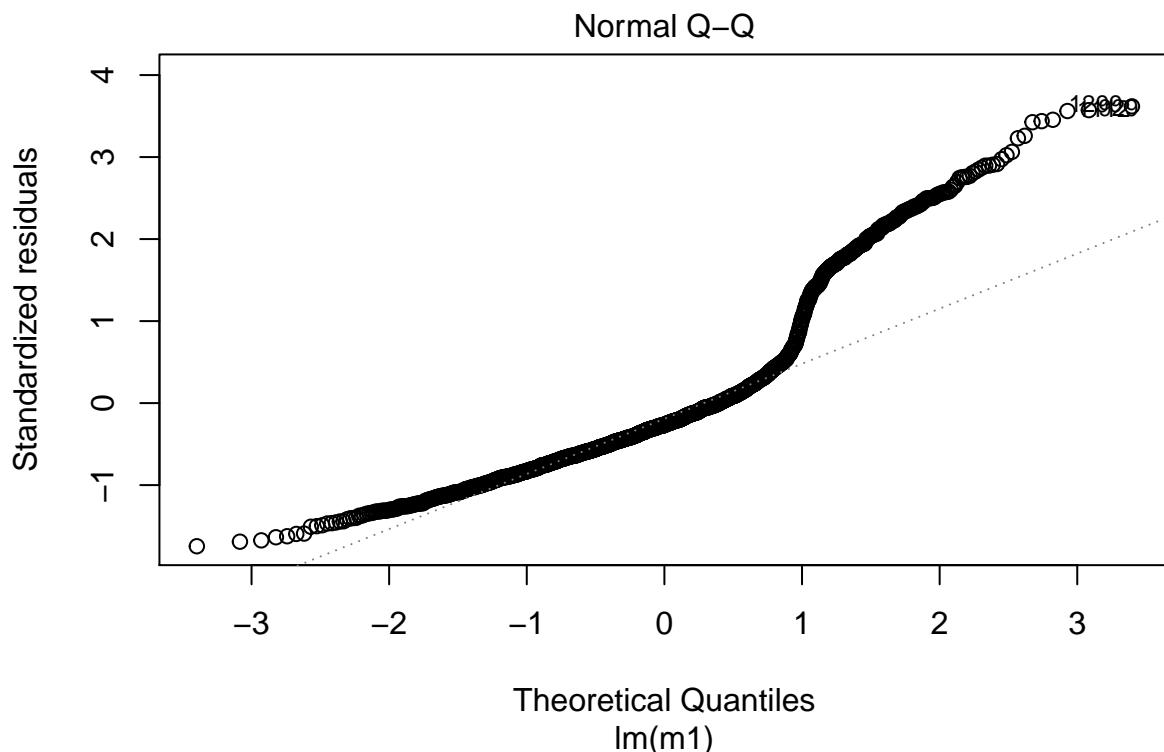
```

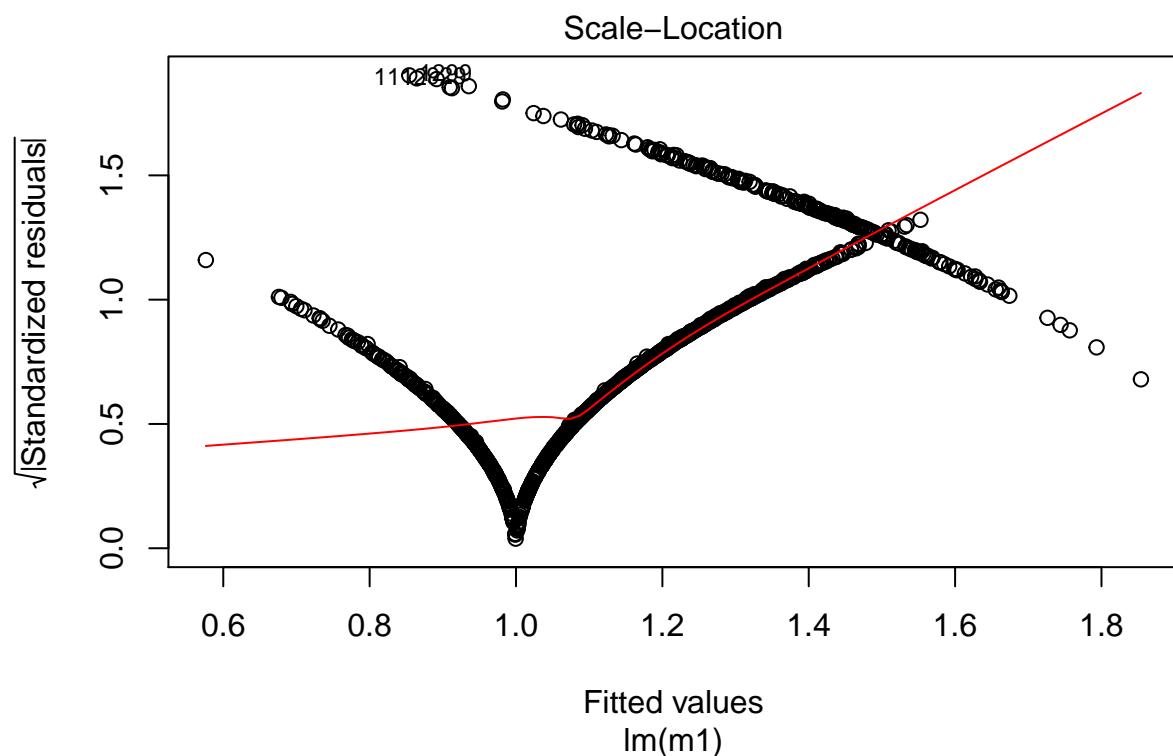
## GenderMale           3.527e-02  1.742e-02  2.025  0.043058
## HourlyRate          -1.688e-04 4.188e-04 -0.403  0.686901
## JobInvolvement      -5.800e-02 1.199e-02 -4.836  1.47e-06
## JobLevel             -5.416e-03 2.855e-02 -0.190  0.849544
## JobRoleHuman Resources 2.163e-01 1.224e-01  1.767  0.077495
## JobRoleLaboratory Technician 1.369e-01 4.001e-02  3.421  0.000642
## JobRoleManager        5.061e-02 6.793e-02  0.745  0.456363
## JobRoleManufacturing Director 1.466e-02 3.921e-02  0.374  0.708604
## JobRoleResearch Director -3.382e-03 6.056e-02 -0.056  0.955470
## JobRoleResearch Scientist 3.858e-02 3.960e-02  0.974  0.330155
## JobRoleSales Executive   1.017e-01 7.748e-02  1.313  0.189440
## JobRoleSales Representative 2.553e-01 8.608e-02  2.965  0.003073
## JobSatisfaction       -3.735e-02 7.718e-03 -4.839  1.45e-06
## MaritalStatusMarried   1.323e-02 2.299e-02  0.575  0.565056
## MaritalStatusSingle    1.102e-01 3.145e-02  3.503  0.000475
## MonthlyIncome          1.460e-06 7.600e-06  0.192  0.847726
## MonthlyRate            4.697e-07 1.193e-06  0.394  0.693790
## NumCompaniesWorked     1.720e-02 3.807e-03  4.519  6.72e-06
## OverTimeYes            2.105e-01 1.896e-02 11.102 < 2e-16
## PercentSalaryHike      -2.181e-03 3.675e-03 -0.594  0.552852
## PerformanceRating      1.826e-02 3.717e-02  0.491  0.623347
## RelationshipSatisfaction -2.330e-02 7.892e-03 -2.953  0.003202
## StockOptionLevel        -1.654e-02 1.367e-02 -1.210  0.226380
## TotalWorkingYears       -3.715e-03 2.417e-03 -1.537  0.124436
## TrainingTimesLastYear   -1.341e-02 6.635e-03 -2.021  0.043491
## WorkLifeBalance         -3.137e-02 1.206e-02 -2.601  0.009384
## YearsAtCompany          5.499e-03 2.989e-03  1.840  0.065995
## YearsInCurrentRole      -9.218e-03 3.876e-03 -2.378  0.017517
## YearsSinceLastPromotion 1.081e-02 3.416e-03  3.164  0.001588
## YearsWithCurrManager    -9.565e-03 3.971e-03 -2.408  0.016150
##
## (Intercept)               ***
## ii..Age                  **
## BusinessTravelTravel_Frequently *** 
## BusinessTravelTravel_Rarely *
## DailyRate                 ***
## DepartmentResearch & Development
## DepartmentSales
## DistanceFromHome          ***
## Education
## EducationFieldLife Sciences
## EducationFieldMarketing
## EducationFieldMedical
## EducationFieldOther
## EducationFieldTechnical Degree
## EmployeeNumber
## EnvironmentSatisfaction   ***
## GenderMale                *
## HourlyRate
## JobInvolvement            ***
## JobLevel
## JobRoleHuman Resources    .
## JobRoleLaboratory Technician *** 
## JobRoleManager

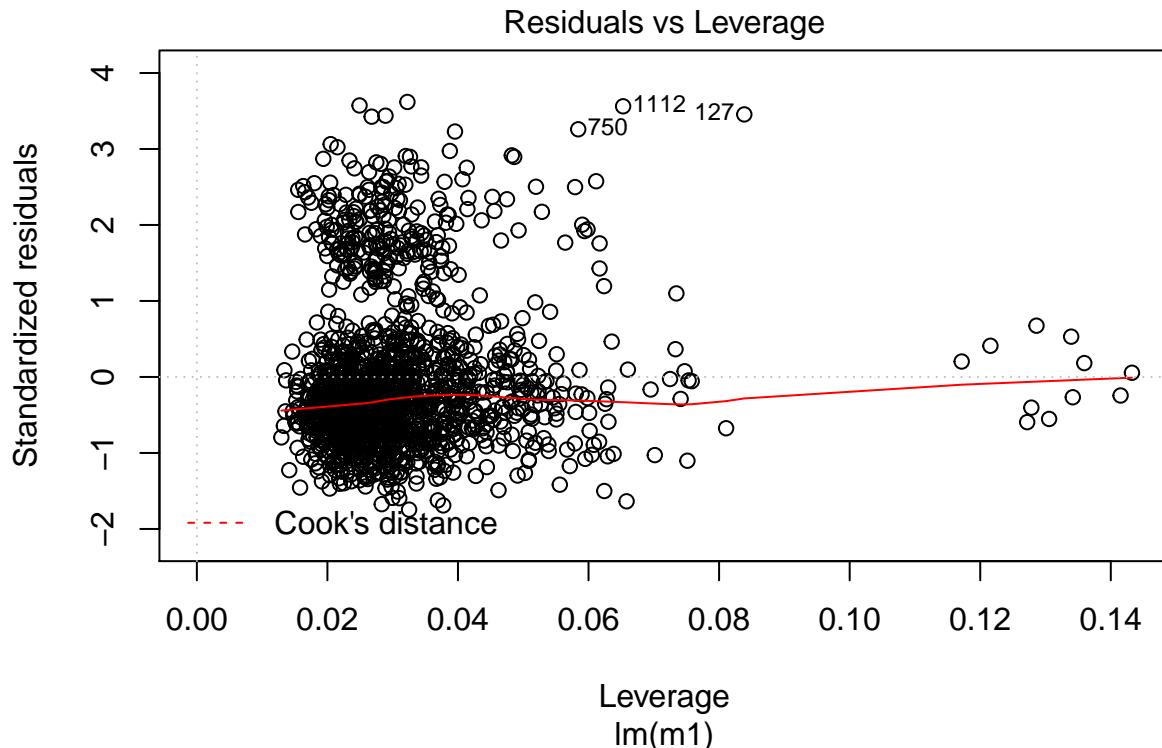
```

```
## JobRoleManufacturing Director
## JobRoleResearch Director
## JobRoleResearch Scientist
## JobRoleSales Executive
## JobRoleSales Representative      **
## JobSatisfaction                  ***
## MaritalStatusMarried
## MaritalStatusSingle             ***
## MonthlyIncome
## MonthlyRate
## NumCompaniesWorked              ***
## OverTimeYes                      ***
## PercentSalaryHike
## PerformanceRating
## RelationshipSatisfaction        **
## StockOptionLevel
## TotalWorkingYears
## TrainingTimesLastYear           *
## WorkLifeBalance                 **
## YearsAtCompany                  .
## YearsInCurrentRole              *
## YearsSinceLastPromotion         **
## YearsWithCurrManager            *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3219 on 1424 degrees of freedom
## Multiple R-squared:  0.2578, Adjusted R-squared:  0.2343
## F-statistic: 10.99 on 45 and 1424 DF,  p-value: < 2.2e-16
plot(f1)
```









MODEL 2

```

ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
m2 <- Attrition ~ i..Age + BusinessTravel + DailyRate + Department + DistanceFromHome + Education +
    EducationField + EmployeeNumber + EnvironmentSatisfaction + Gender

f2 <- lm(m2, data = ds)
summary(f2)

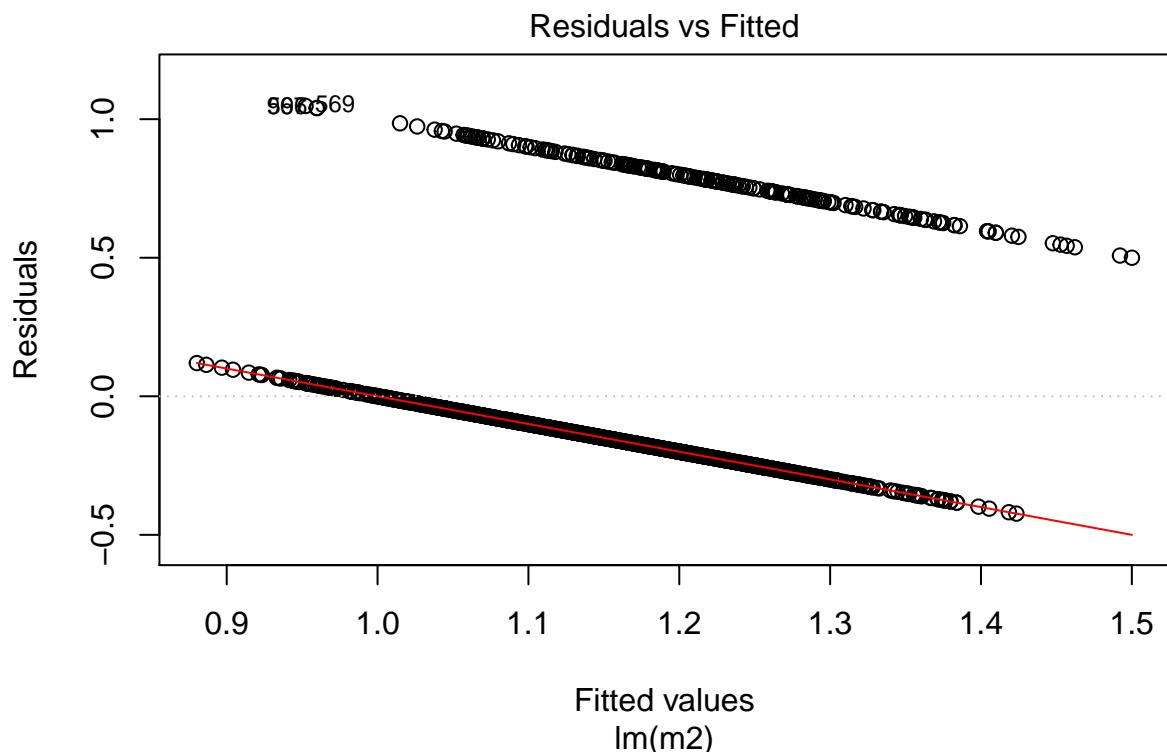
## 
## Call:
## lm(formula = m2, data = ds)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.42349 -0.20326 -0.12648 -0.02365  1.04772 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               1.488e+00  9.338e-02 15.940 < 2e-16  
## i..Age                  -6.098e-03  1.044e-03 -5.842 6.36e-09  
## BusinessTravelTravel_Frequently 1.710e-01  3.615e-02  4.730 2.47e-06  
## BusinessTravelTravel_Rarely    7.566e-02  3.118e-02  2.427 0.015356  
## DailyRate                -4.510e-05  2.314e-05 -1.949 0.051510  

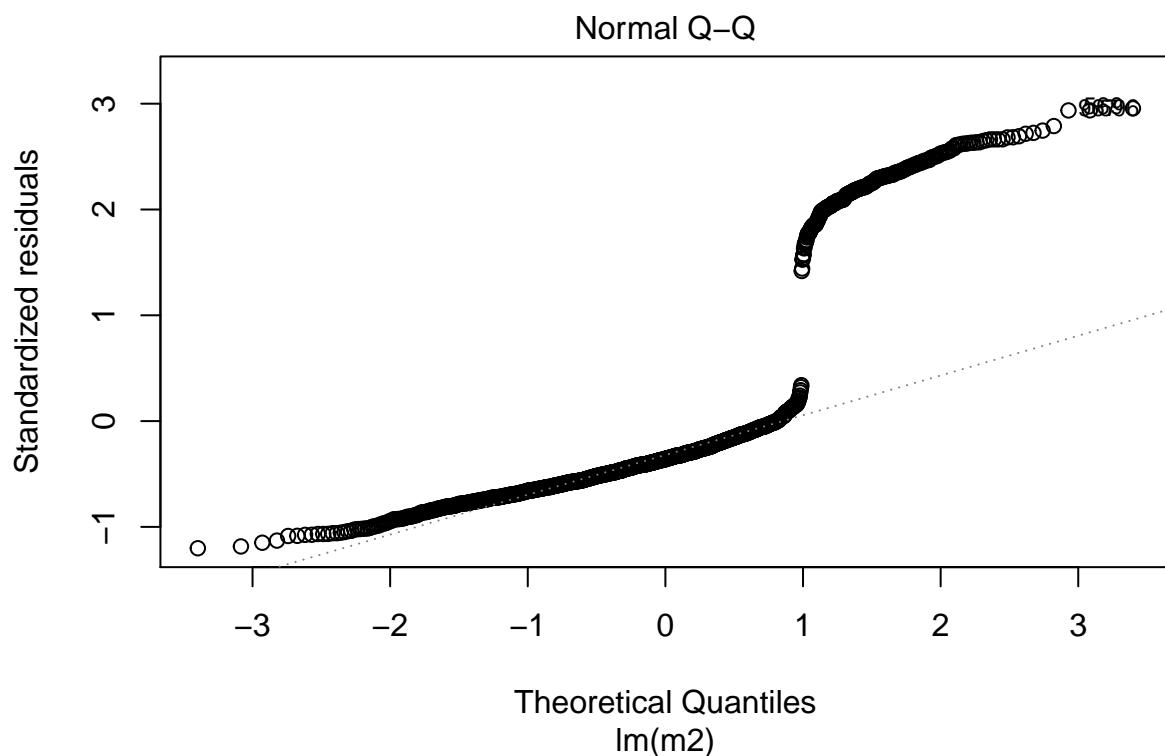
```

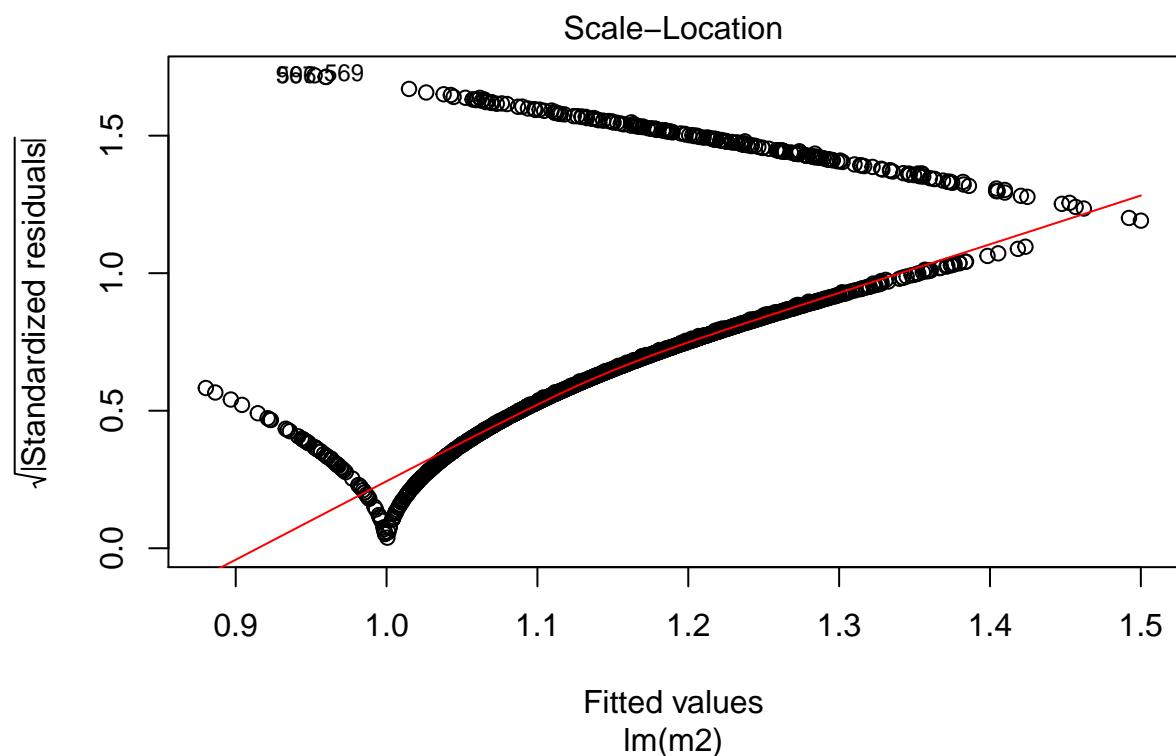
```

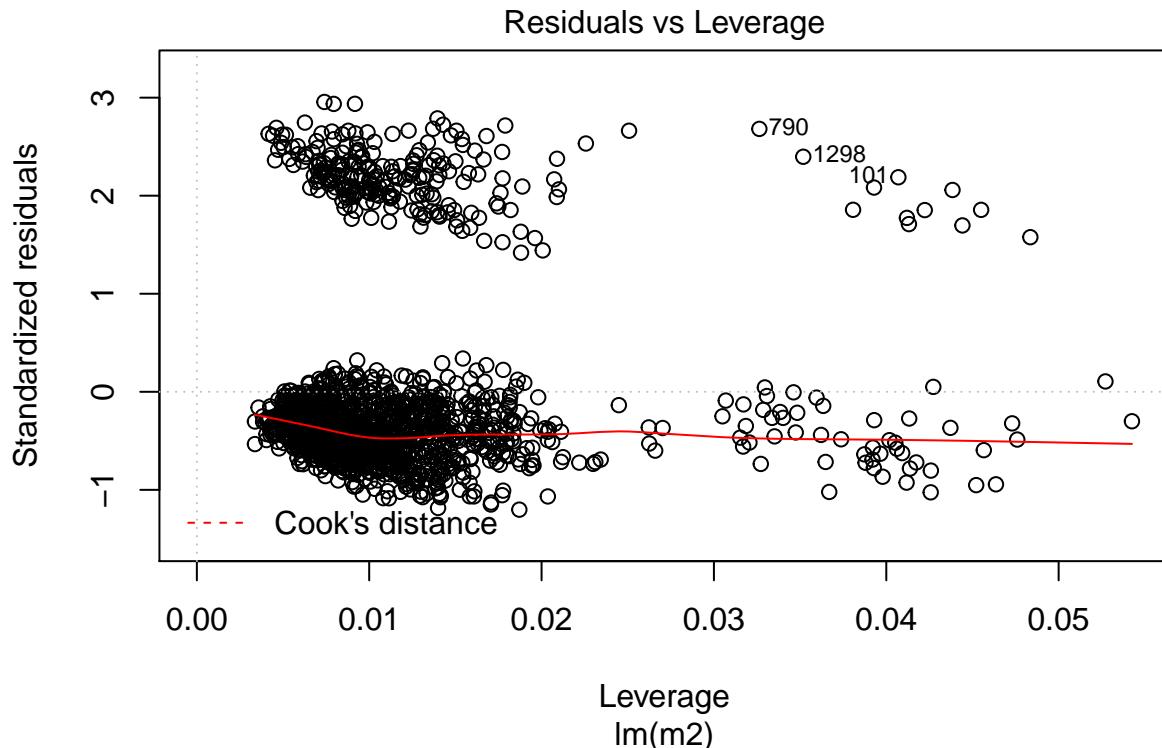
## DepartmentResearch & Development -1.038e-02 6.057e-02 -0.171 0.863998
## DepartmentSales 4.023e-02 6.313e-02 0.637 0.524068
## DistanceFromHome 3.487e-03 1.148e-03 3.038 0.002421
## Education -1.934e-03 9.336e-03 -0.207 0.835936
## EducationFieldLife Sciences -1.087e-01 9.141e-02 -1.189 0.234680
## EducationFieldMarketing -7.260e-02 9.740e-02 -0.745 0.456175
## EducationFieldMedical -1.172e-01 9.175e-02 -1.277 0.201779
## EducationFieldOther -1.172e-01 9.811e-02 -1.195 0.232319
## EducationFieldTechnical Degree -9.022e-03 9.536e-02 -0.095 0.924638
## EmployeeNumber -1.008e-05 1.552e-05 -0.650 0.516067
## EnvironmentSatisfaction -3.304e-02 8.529e-03 -3.873 0.000112
## GenderMale 2.275e-02 1.901e-02 1.197 0.231564
##
## (Intercept) ***
## i..Age ***
## BusinessTravelTravel_Frequently ***
## BusinessTravelTravel_Rarely *
## DailyRate .
## DepartmentResearch & Development
## DepartmentSales
## DistanceFromHome **
## Education
## EducationFieldLife Sciences
## EducationFieldMarketing
## EducationFieldMedical
## EducationFieldOther
## EducationFieldTechnical Degree
## EmployeeNumber
## EnvironmentSatisfaction ***
## GenderMale
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3558 on 1453 degrees of freedom
## Multiple R-squared: 0.07492, Adjusted R-squared: 0.06474
## F-statistic: 7.355 on 16 and 1453 DF, p-value: < 2.2e-16
plot(f2)

```









MODEL 3

```

ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
m3 <- Attrition~HourlyRate+JobInvolvement+JobLevel+JobRole+JobSatisfaction+MaritalStatus+
    MonthlyIncome+MonthlyRate+NumCompaniesWorked+OverTime+PercentSalaryHike

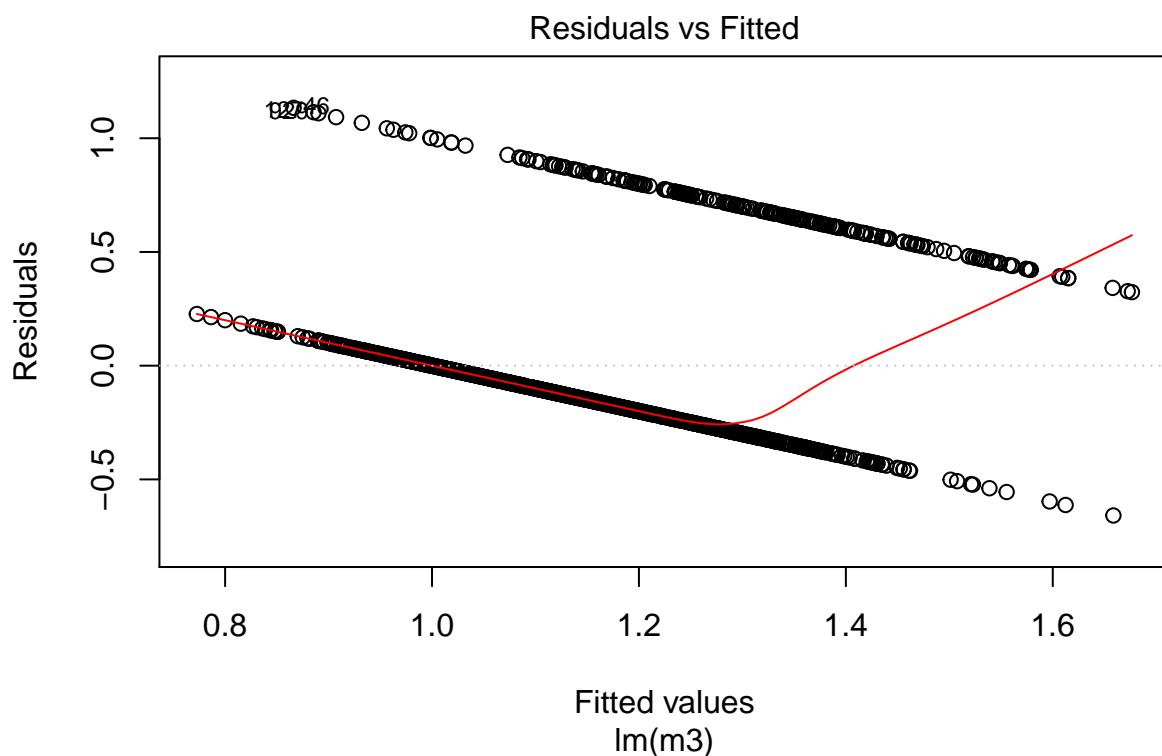
f3<-lm(m3,data=ds)
summary(f3)

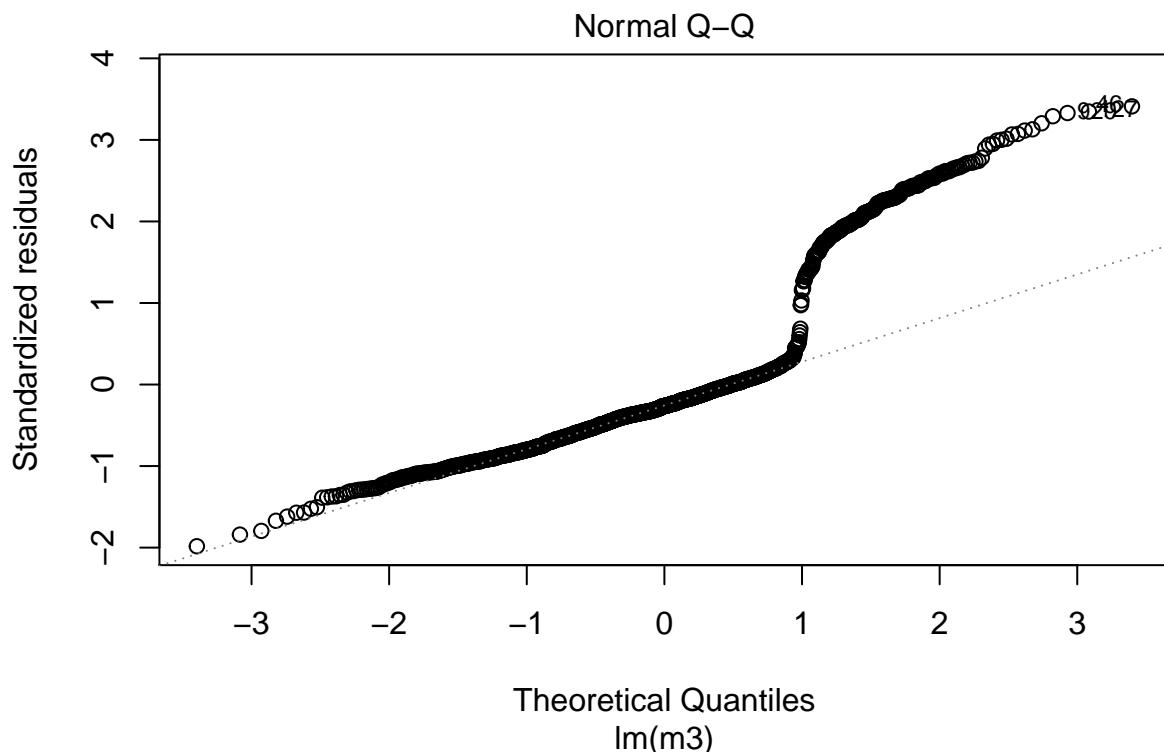
##
## Call:
## lm(formula = m3, data = ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.65860 -0.20574 -0.08826  0.03467  1.13379 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               1.285e+00  8.570e-02 14.991 < 2e-16 ***
## HourlyRate                -1.008e-04  4.333e-04 -0.233  0.81603  
## JobInvolvement             -6.186e-02  1.238e-02 -4.995 6.59e-07 ***
## JobLevel                  -3.183e-02  2.843e-02 -1.119  0.26323  
## JobRoleHuman Resources    1.402e-01  5.743e-02  2.440  0.01479 *  
## 
```

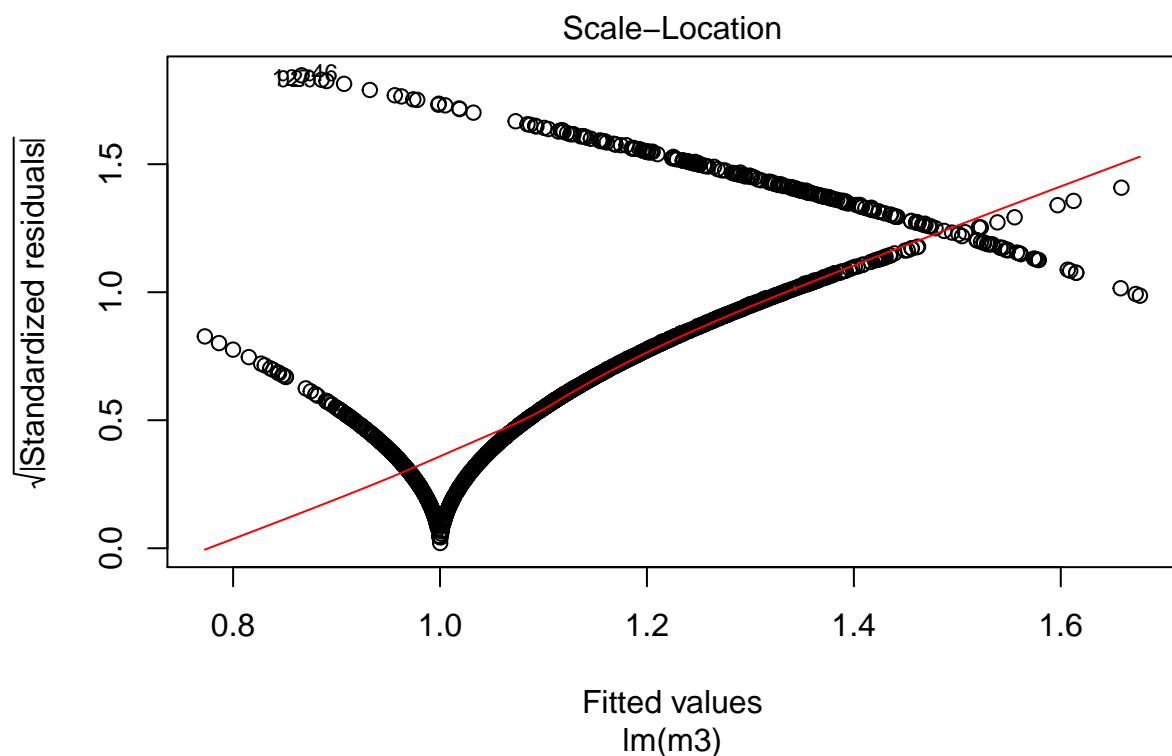
```

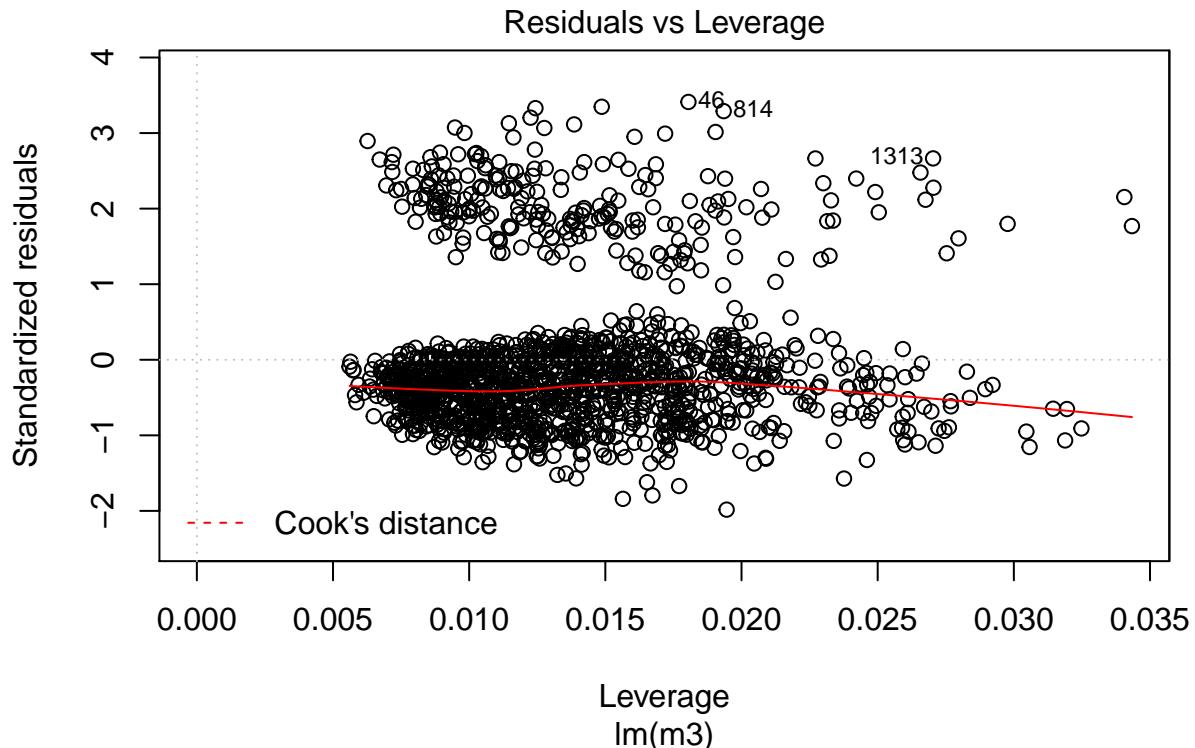
## JobRoleLaboratory Technician  1.350e-01  4.133e-02   3.267  0.00111 **
## JobRoleManager               3.141e-02  6.192e-02   0.507  0.61208
## JobRoleManufacturing Director -4.662e-03  4.047e-02  -0.115  0.90832
## JobRoleResearch Director    -1.427e-02  6.143e-02  -0.232  0.81629
## JobRoleResearch Scientist    4.390e-02  4.092e-02   1.073  0.28355
## JobRoleSales Executive      9.321e-02  3.485e-02   2.674  0.00758 **
## JobRoleSales Representative  2.738e-01  5.254e-02   5.211  2.14e-07 ***
## JobSatisfaction            -3.679e-02  7.981e-03  -4.609  4.40e-06 ***
## MaritalStatusMarried        2.628e-02  2.267e-02   1.159  0.24657
## MaritalStatusSingle         1.395e-01  2.435e-02   5.730  1.22e-08 ***
## MonthlyIncome              1.292e-06  7.728e-06   0.167  0.86720
## MonthlyRate                4.186e-07  1.234e-06   0.339  0.73447
## NumCompaniesWorked          1.183e-02  3.574e-03   3.312  0.00095 ***
## OverTimeYes                 2.082e-01  1.950e-02  10.677 < 2e-16 ***
## PercentSalaryHike           -1.207e-03  2.402e-03  -0.503  0.61531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3354 on 1450 degrees of freedom
## Multiple R-squared:  0.1796, Adjusted R-squared:  0.1688
## F-statistic:  16.7 on 19 and 1450 DF,  p-value: < 2.2e-16
plot(f3)

```









MODEL 4

```

ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
m4 <- Attrition~PerformanceRating+RelationshipSatisfaction+StockOptionLevel+TotalWorkingYears+
      TrainingTimesLastYear+WorkLifeBalance+YearsAtCompany+YearsInCurrentRole+YearsSinceLastPromotion+
      YearsWithCurrManager

## [1] 5 7 0 0 2 6 0 0 8 7 3 8 3 2 3 8 5 0 7 2 3 3 11
## [24] 0 3 8 7 2 17 1 0 3 8 0 0 4 2 2 0 3 0 0 1 7 7 8
## [47] 7 0 8 1 0 2 3 3 0 12 8 2 7 7 8 9 8 9 8 2 3 0 2
## [70] 0 3 4 0 9 0 8 8 0 2 2 0 4 0 9 3 2 0 3 8 7 11 0
## [93] 7 5 7 3 2 4 8 2 2 0 0 10 0 2 7 3 1 0 8 6 4 2 0
## [116] 3 8 8 0 12 9 7 3 15 7 3 6 0 1 11 2 1 2 7 7 1 3 9
## [139] 2 7 4 3 2 4 7 2 2 7 4 0 13 6 5 15 7 9 0 0 11 2 2
## [162] 2 2 4 1 4 8 4 7 4 1 0 2 10 3 2 2 0 7 1 2 2 2 2
## [185] 3 6 9 15 9 0 9 3 1 2 8 7 2 2 0 7 0 7 0 7 0 9 2
## [208] 7 2 3 8 8 7 7 4 0 1 5 8 5 7 1 7 13 4 0 8 1 8 1
## [231] 2 15 2 6 4 9 7 12 2 2 2 0 7 7 2 2 9 2 2 0 8 5
## [254] 2 2 2 8 9 0 4 3 5 4 7 0 0 8 5 10 11 13 9 4 1 0 2
## [277] 2 7 4 6 3 6 9 8 3 7 0 2 2 2 2 7 2 4 2 8 0 9 0
## [300] 2 4 0 7 7 7 8 8 1 2 3 0 11 2 8 10 7 7 7 2 9 4 7
## [323] 7 2 9 9 3 0 1 1 0 4 3 0 8 2 7 2 2 7 4 9 10 7 13
## [346] 2 2 0 0 2 0 7 0 7 2 7 0 2 0 12 2 7 2 0 3 0 7 3
## [369] 0 2 0 2 0 4 7 0 7 2 4 2 3 0 4 2 0 0 17 3 2 1 9

```

```

## [392] 0 2 2 12 2 2 3 7 2 6 7 4 7 8 2 4 2 4 2 2 10 8 4
## [415] 4 0 0 12 2 4 7 1 0 8 0 7 7 11 4 1 3 0 7 2 8 7 3
## [438] 2 0 0 2 0 7 2 7 7 10 9 13 7 5 8 7 0 2 5 1 0 1 7
## [461] 0 3 7 0 2 2 8 7 9 2 1 0 3 9 4 4 0 7 6 2 0 3 2
## [484] 2 7 3 0 0 2 0 0 7 2 2 6 2 2 1 2 2 5 0 10 8 0 2
## [507] 8 2 8 7 1 2 1 1 7 0 2 2 5 8 2 4 0 8 7 0 10 8 3
## [530] 8 7 7 7 12 3 6 2 7 10 2 7 5 0 4 13 7 0 2 2 3 2 6
## [553] 1 3 7 2 2 0 7 1 0 16 8 4 1 2 3 4 4 9 2 0 1 2 3
## [576] 4 1 3 12 3 2 2 7 2 9 0 0 4 3 0 7 2 0 5 7 8 2 2
## [599] 1 4 8 0 0 0 8 2 3 7 3 0 8 6 5 2 7 0 17 3 1 8 1
## [622] 11 3 3 4 2 3 0 0 4 2 2 2 3 3 15 8 2 3 2 4 3 2 3
## [645] 2 2 2 8 0 8 2 7 9 7 4 3 1 3 1 2 0 0 2 0 11 2 2
## [668] 0 2 2 0 0 3 2 4 2 7 7 0 0 1 11 2 0 0 7 17 9 0 0
## [691] 5 0 3 7 4 7 8 2 0 8 0 2 7 3 8 7 0 7 7 2 0 0 0 2
## [714] 2 3 2 11 2 8 7 1 5 2 7 2 2 3 0 8 1 7 0 2 3 1 2
## [737] 8 1 8 2 2 0 0 3 0 7 10 0 2 9 11 13 14 8 2 1 7 9 4
## [760] 4 12 0 2 0 0 2 7 1 5 3 2 0 3 8 2 0 2 0 7 7 9 3
## [783] 2 8 9 1 2 2 7 6 2 7 8 2 4 0 6 0 2 7 0 0 2 0 1
## [806] 12 0 7 7 7 9 0 1 5 9 2 3 9 2 4 2 3 2 1 0 2 0 2
## [829] 0 3 0 2 3 2 3 4 10 13 14 5 7 2 0 7 4 0 4 9 2 3 0
## [852] 2 8 0 2 11 2 2 7 3 0 9 0 4 0 3 0 2 3 3 7 0 1 7
## [875] 9 17 2 4 3 9 2 7 7 10 3 1 7 0 12 8 4 7 0 2 9 1 8
## [898] 5 7 0 7 0 3 0 0 7 0 10 8 1 1 0 2 11 0 2 8 2 10 4
## [921] 10 2 13 8 2 0 17 12 5 2 7 7 2 1 8 0 2 2 1 3 7 7
## [944] 7 5 2 0 7 8 7 6 9 2 2 10 11 7 0 6 8 7 1 10 7 6 2
## [967] 9 3 13 8 0 3 0 0 0 8 9 0 8 7 0 2 3 11 4 7 4 2 3
## [990] 3 2 2 5 2 2 8 4 7 1 9 4 2 2 2 7 9 3 7 7 5 0 1
## [1013] 0 2 0 3 0 4 8 2 2 2 3 0 6 3 3 2 3 7 9 3 0 7 7
## [1036] 2 2 2 2 2 8 3 3 8 2 2 3 3 12 2 8 0 0 8 7 9 2 0
## [1059] 10 0 0 0 2 4 3 2 2 2 0 0 3 2 2 7 0 0 12 7 17 2 1
## [1082] 3 2 0 6 7 9 7 0 7 7 2 3 9 1 7 7 0 2 7 2 5 2 7
## [1105] 0 2 7 2 2 3 0 9 4 2 6 0 13 1 0 7 2 0 5 4 6 3 0
## [1128] 2 2 1 8 7 2 2 1 12 0 3 13 2 10 1 4 5 3 1 2 4 7 7
## [1151] 10 0 2 0 0 7 12 9 2 8 9 2 0 7 3 2 2 2 3 2 2 3 4
## [1174] 2 7 0 2 11 2 9 3 12 2 4 7 11 7 9 9 4 6 2 2 3 2 0
## [1197] 2 2 2 2 2 2 0 2 7 0 7 2 13 0 13 3 7 2 7 4 8 4 7
## [1220] 2 7 4 0 10 2 6 3 3 2 0 2 9 7 8 0 3 2 0 2 4 8 2
## [1243] 8 3 7 1 2 2 4 2 2 7 2 7 4 0 5 0 0 7 7 0 0 4 0
## [1266] 7 4 7 2 8 2 0 4 0 7 2 3 2 7 3 7 7 6 3 2 2 2 0
## [1289] 7 8 7 8 2 2 2 5 2 4 7 3 8 14 3 10 7 7 2 2 3 2 2
## [1312] 0 0 0 7 0 7 2 10 0 3 0 7 2 6 0 2 8 8 0 14 9 0 7
## [1335] 3 2 4 0 0 0 7 2 0 2 4 2 8 8 16 0 7 7 2 0 0 3 4
## [1358] 7 4 1 2 7 3 9 7 0 2 2 11 2 7 0 9 7 0 3 3 3 0 0
## [1381] 4 4 3 2 1 1 6 2 2 7 3 0 9 7 0 8 2 3 8 9 2 4 0
## [1404] 10 7 8 4 2 4 4 10 2 7 2 11 0 8 2 7 2 0 12 2 2 8 1
## [1427] 1 0 2 7 11 7 8 4 0 2 2 3 0 4 2 9 2 14 8 10 7 11 4
## [1450] 2 7 9 1 0 1 2 2 2 1 3 4 0 6 7 0 3 7 3 8 2

```

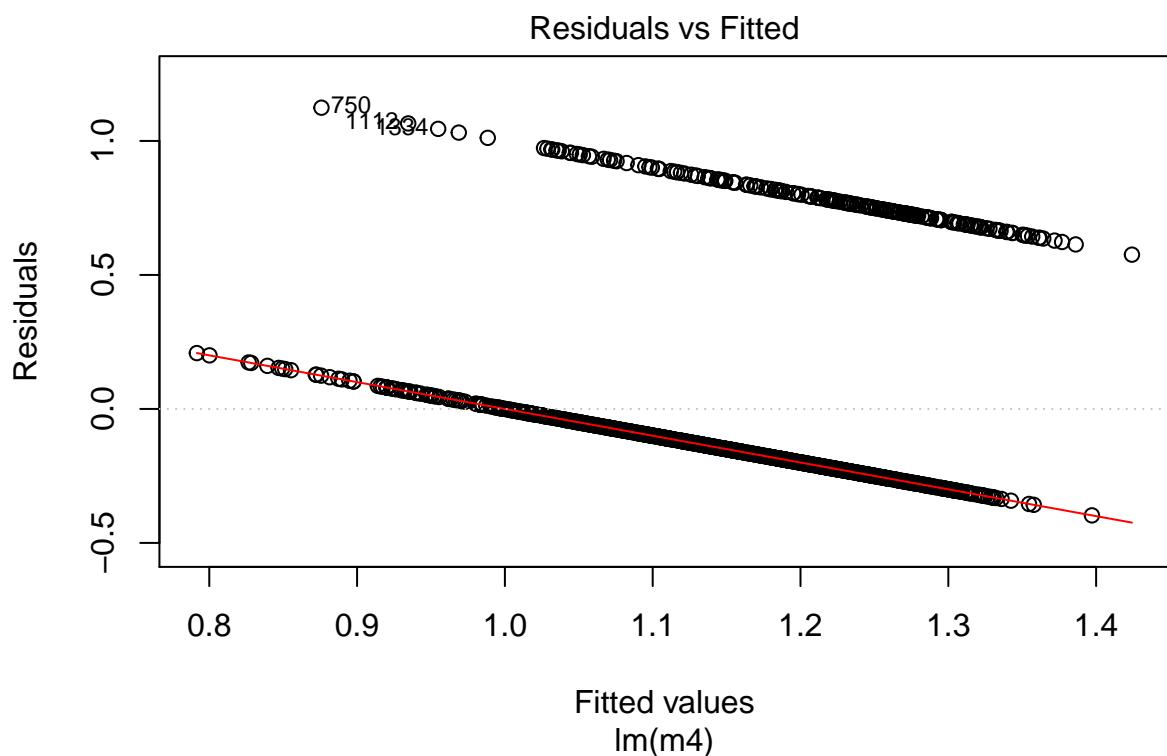
```
f4<-lm(m4,data=ds)
summary(f4)
```

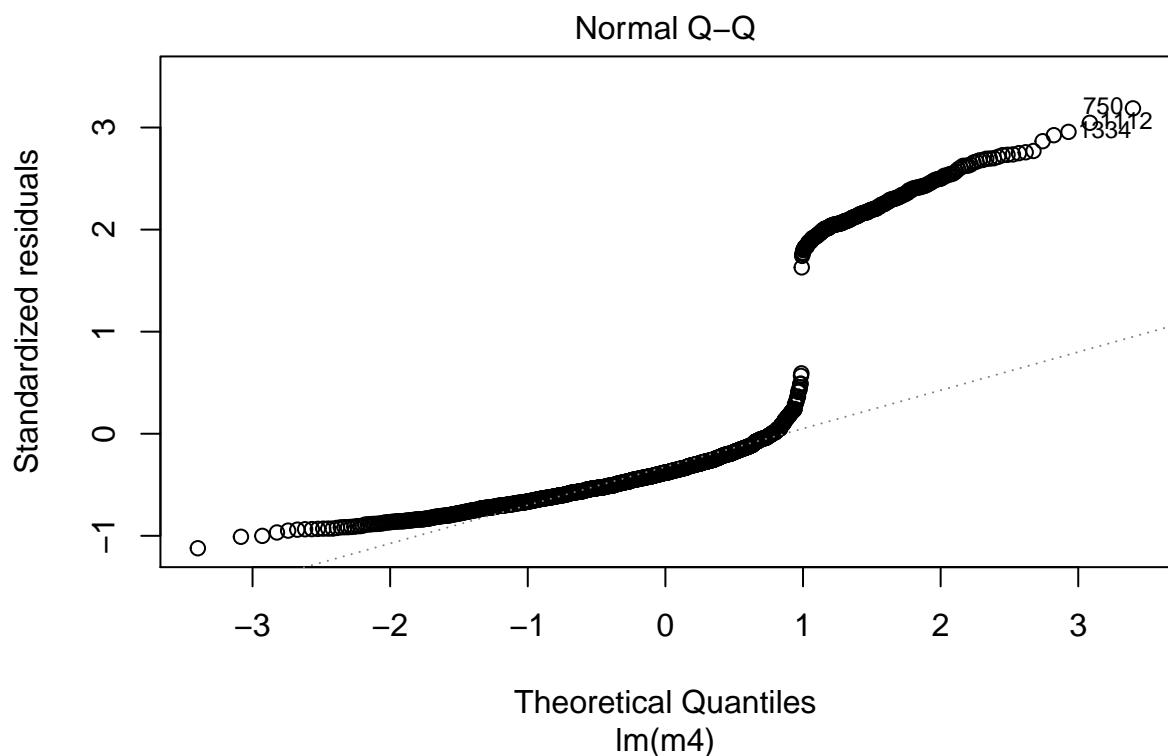
```
##
## Call:
## lm(formula = m4, data = ds)
##
```

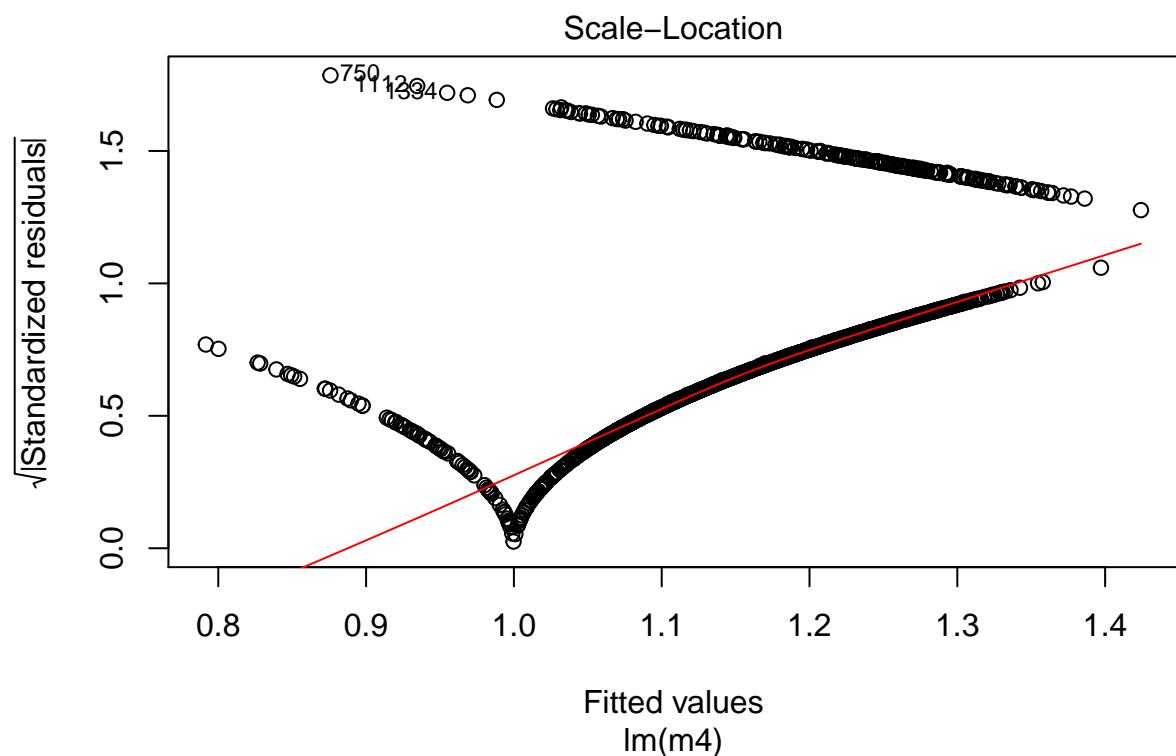
```

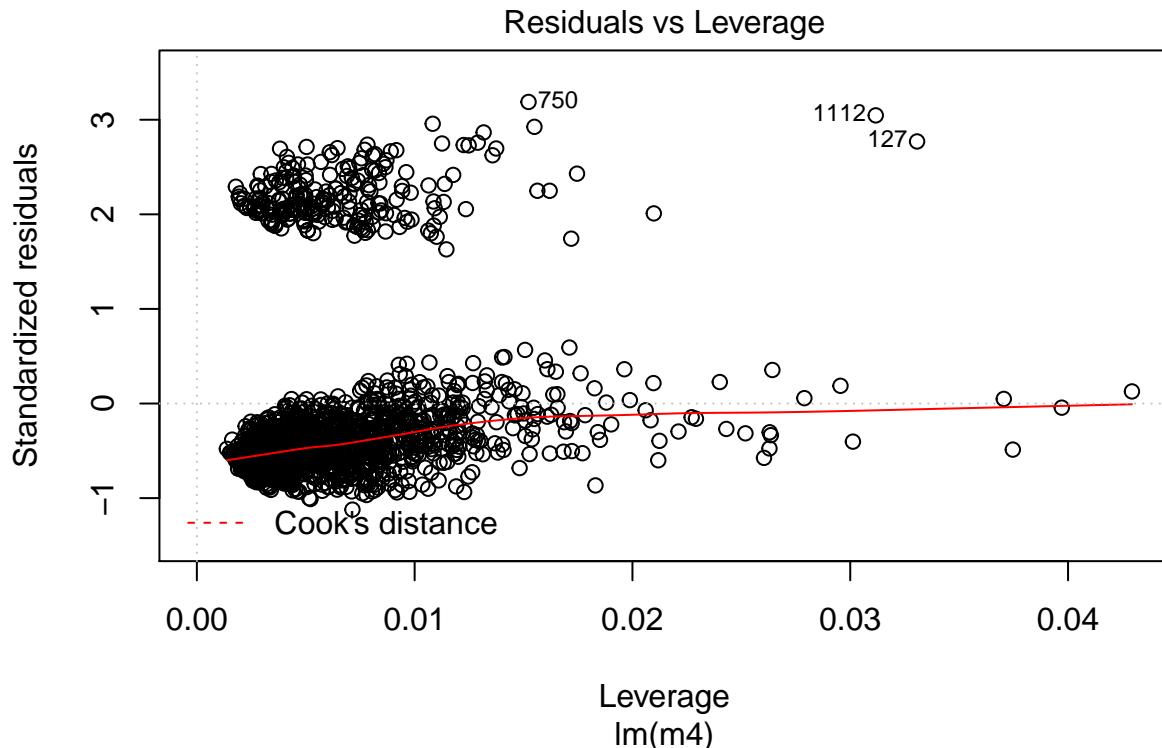
## Residuals:
##      Min       1Q    Median       3Q      Max
## -0.39717 -0.20498 -0.13527 -0.02517  1.12410
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.4798556  0.0962004 15.383 < 2e-16 ***
## PerformanceRating     0.0053350  0.0257424  0.207 0.835846
## RelationshipSatisfaction -0.0180317 0.0086053 -2.095 0.036306 *
## StockOptionLevel      -0.0566255  0.0109128 -5.189 2.41e-07 ***
## TotalWorkingYears     -0.0072541  0.0015349 -4.726 2.51e-06 ***
## TrainingTimesLastYear -0.0178166  0.0072044 -2.473 0.013511 *
## WorkLifeBalance        -0.0281344  0.0131588 -2.138 0.032676 *
## YearsAtCompany         0.0008911  0.0028035  0.318 0.750654
## YearsInCurrentRole    -0.0154146  0.0040029 -3.851 0.000123 ***
## YearsSinceLastPromotion 0.0122202  0.0037098  3.294 0.001011 **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3553 on 1460 degrees of freedom
## Multiple R-squared:  0.0727, Adjusted R-squared:  0.06698
## F-statistic: 12.72 on 9 and 1460 DF,  p-value: < 2.2e-16
plot(f4)

```









MODEL 5

```

ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
m5 <- Attrition~StockOptionLevel+TotalWorkingYears+YearsInCurrentRole+YearsSinceLastPromotion
+YearsWithCurrManager

## [1] 5 7 0 0 2 6 0 0 8 7 3 8 3 2 3 8 5 0 7 2 3 3 11
## [24] 0 3 8 7 2 17 1 0 3 8 0 0 4 2 2 0 3 0 0 1 7 7 8
## [47] 7 0 8 1 0 2 3 3 0 12 8 2 7 7 8 9 8 9 8 2 3 0 2
## [70] 0 3 4 0 9 0 8 8 0 2 2 0 4 0 9 3 2 0 3 8 7 11 0
## [93] 7 5 7 3 2 4 8 2 2 0 0 10 0 2 7 3 1 0 8 6 4 2 0
## [116] 3 8 8 0 12 9 7 3 15 7 3 6 0 1 11 2 1 2 7 7 1 3 9
## [139] 2 7 4 3 2 4 7 2 2 7 4 0 13 6 5 15 7 9 0 0 11 2 2
## [162] 2 2 4 1 4 8 4 7 4 1 0 2 10 3 2 2 0 7 1 2 2 2 2
## [185] 3 6 9 15 9 0 9 3 1 2 8 7 2 2 0 7 0 7 0 7 0 9 2
## [208] 7 2 3 8 8 7 7 4 0 1 5 8 5 7 1 7 13 4 0 8 1 8 1
## [231] 2 15 2 6 4 9 7 12 2 2 2 2 0 7 7 2 2 9 2 2 0 8 5
## [254] 2 2 2 8 9 0 4 3 5 4 7 0 0 8 5 10 11 13 9 4 1 0 2
## [277] 2 7 4 6 3 6 9 8 3 7 0 2 2 2 7 2 4 2 8 0 9 0
## [300] 2 4 0 7 7 7 8 8 1 2 3 0 11 2 8 10 7 7 2 9 4 7
## [323] 7 2 9 9 3 0 1 1 0 4 3 0 8 2 7 2 2 7 4 9 10 7 13
## [346] 2 2 0 0 2 0 7 0 7 2 7 0 2 0 12 2 7 2 0 3 0 7 3
## [369] 0 2 0 2 0 4 7 0 7 2 4 2 3 0 4 2 0 0 17 3 2 1 9
## [392] 0 2 2 12 2 2 3 7 2 6 7 4 7 8 2 4 2 4 2 2 10 8 4

```

```

## [415] 4 0 0 12 2 4 7 1 0 8 0 7 7 11 4 1 3 0 7 2 8 7 3
## [438] 2 0 0 2 0 7 2 7 7 10 9 13 7 5 8 7 0 2 5 1 0 1 7
## [461] 0 3 7 0 2 2 8 7 9 2 1 0 3 9 4 4 0 7 6 2 0 3 2
## [484] 2 7 3 0 0 2 0 0 7 2 2 6 2 2 1 2 2 5 0 10 8 0 2
## [507] 8 2 8 7 1 2 1 1 7 0 2 2 5 8 2 4 0 8 7 0 10 8 3
## [530] 8 7 7 7 12 3 6 2 7 10 2 7 5 0 4 13 7 0 2 2 3 2 6
## [553] 1 3 7 2 2 0 7 1 0 16 8 4 1 2 3 4 4 9 2 0 1 2 3
## [576] 4 1 3 12 3 2 2 7 2 9 0 0 4 3 0 7 2 0 5 7 8 2 2
## [599] 1 4 8 0 0 0 8 2 3 7 3 0 8 6 5 2 7 0 17 3 1 8 1
## [622] 11 3 3 4 2 3 0 0 4 2 2 2 3 3 15 8 2 3 2 4 3 2 3
## [645] 2 2 2 8 0 8 2 7 9 7 4 3 1 3 1 2 0 0 2 0 11 2 2
## [668] 0 2 2 0 0 3 2 4 2 7 7 0 0 1 11 2 0 0 7 17 9 0 0
## [691] 5 0 3 7 4 7 8 2 0 8 0 2 7 3 8 7 0 7 7 2 0 0 0 2
## [714] 2 3 2 11 2 8 7 1 5 2 7 2 2 3 0 8 1 7 0 2 3 1 2
## [737] 8 1 8 2 2 0 0 3 0 7 10 0 2 9 11 13 14 8 2 1 7 9 4
## [760] 4 12 0 2 0 0 2 7 1 5 3 2 0 3 8 2 0 2 0 7 7 9 3
## [783] 2 8 9 1 2 2 7 6 2 7 8 2 4 0 6 0 2 7 0 0 2 0 1
## [806] 12 0 7 7 7 9 0 1 5 9 2 3 9 2 4 2 3 2 1 0 2 0 2
## [829] 0 3 0 2 3 2 3 4 10 13 14 5 7 2 0 7 4 0 4 9 2 3 0
## [852] 2 8 0 2 11 2 2 7 3 0 9 0 4 0 3 0 2 3 3 7 0 1 7
## [875] 9 17 2 4 3 9 2 7 7 10 3 1 7 0 12 8 4 7 0 2 9 1 8
## [898] 5 7 0 7 0 3 0 0 7 0 10 8 1 1 0 2 11 0 2 8 2 10 4
## [921] 10 2 13 8 2 0 17 12 5 2 7 7 7 2 1 8 0 2 2 1 3 7 7
## [944] 7 5 2 0 7 8 7 6 9 2 2 10 11 7 0 6 8 7 1 10 7 6 2
## [967] 9 3 13 8 0 3 0 0 0 8 9 0 8 7 0 2 3 11 4 7 4 2 3
## [990] 3 2 2 5 2 2 8 4 7 1 9 4 2 2 2 7 9 3 7 7 5 0 1
## [1013] 0 2 0 3 0 4 8 2 2 2 3 0 6 3 3 2 3 7 9 3 0 7 7
## [1036] 2 2 2 2 2 8 3 3 8 2 2 3 3 12 2 8 0 0 8 7 9 2 0
## [1059] 10 0 0 0 2 4 3 2 2 2 0 0 3 2 2 7 0 0 12 7 17 2 1
## [1082] 3 2 0 6 7 9 7 0 7 7 2 3 9 1 7 7 0 2 7 2 5 2 7
## [1105] 0 2 7 2 2 3 0 9 4 2 6 0 13 1 0 7 2 0 5 4 6 3 0
## [1128] 2 2 1 8 7 2 2 1 12 0 3 13 2 10 1 4 5 3 1 2 4 7 7
## [1151] 10 0 2 0 0 7 12 9 2 8 9 2 0 7 3 2 2 2 3 2 2 3 4
## [1174] 2 7 0 2 11 2 9 3 12 2 4 7 11 7 9 9 4 6 2 2 3 2 0
## [1197] 2 2 2 2 2 0 2 7 0 7 2 13 0 13 3 7 2 7 4 8 4 7
## [1220] 2 7 4 0 10 2 6 3 3 2 0 2 9 7 8 0 3 2 0 2 4 8 2
## [1243] 8 3 7 1 2 2 4 2 2 7 2 7 4 0 5 0 0 7 7 0 0 4 0
## [1266] 7 4 7 2 8 2 0 4 0 7 2 3 2 7 3 7 7 6 3 2 2 2 0
## [1289] 7 8 7 8 2 2 2 5 2 4 7 3 8 14 3 10 7 7 2 2 3 2 2
## [1312] 0 0 0 7 0 7 2 10 0 3 0 7 2 6 0 2 8 8 0 14 9 0 7
## [1335] 3 2 4 0 0 0 7 2 0 2 4 2 8 8 16 0 7 7 2 0 0 3 4
## [1358] 7 4 1 2 7 3 9 7 0 2 2 11 2 7 0 9 7 0 3 3 3 0 0
## [1381] 4 4 3 2 1 1 6 2 2 7 3 0 9 7 0 8 2 3 8 9 2 4 0
## [1404] 10 7 8 4 2 4 4 10 2 7 2 11 0 8 2 7 2 0 12 2 2 8 1
## [1427] 1 0 2 7 11 7 8 4 0 2 2 3 0 4 2 9 2 14 8 10 7 11 4
## [1450] 2 7 9 1 0 1 2 2 2 1 3 4 0 6 7 0 3 7 3 8 2

```

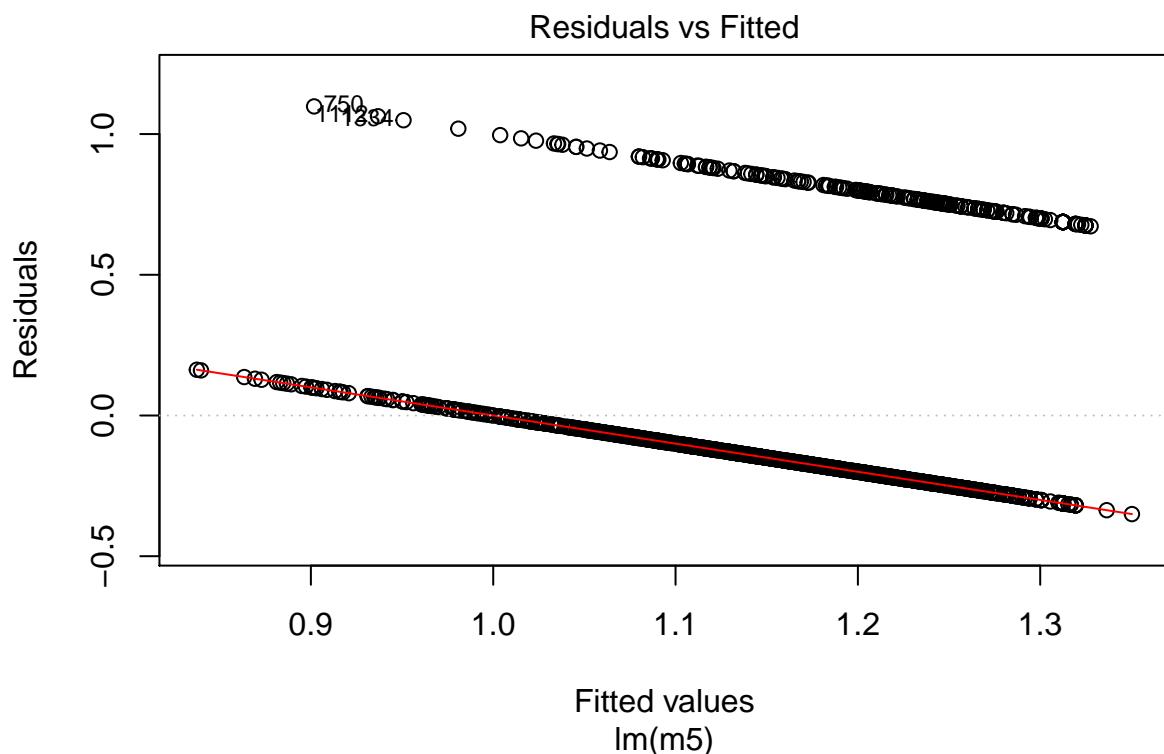
```
f5<-lm(m5,data=ds)
summary(f5)
```

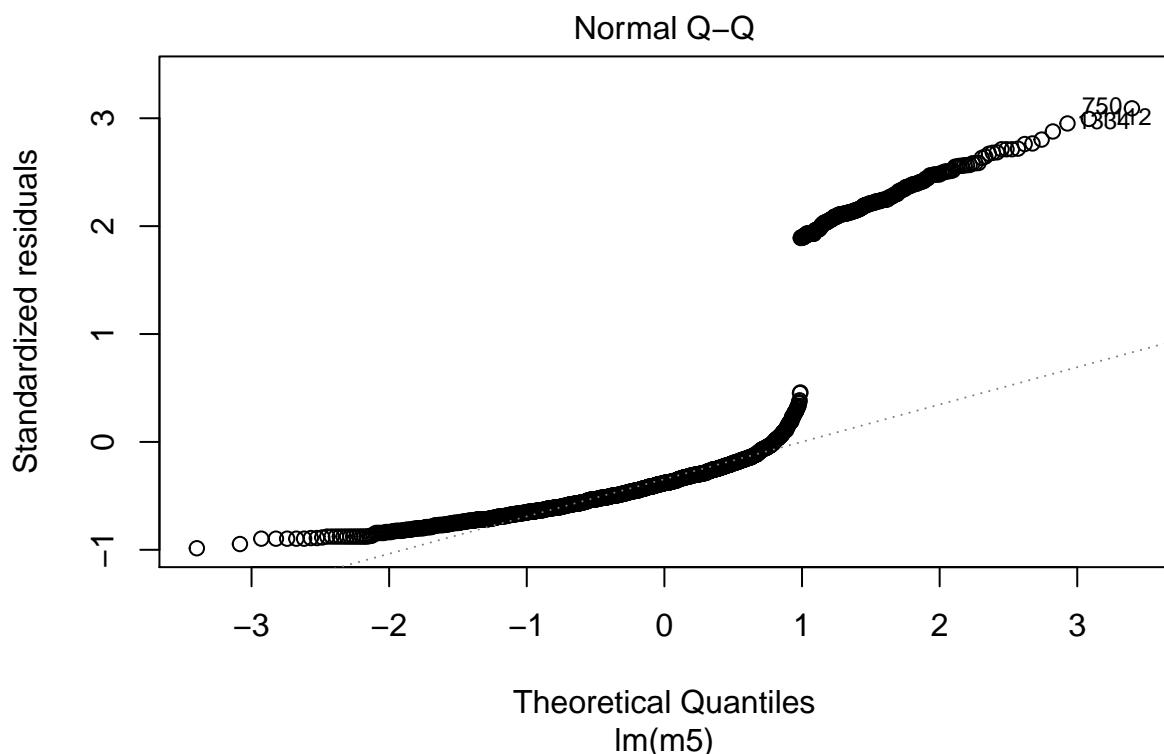
```
##
## Call:
## lm(formula = m5, data = ds)
##
## Residuals:
```

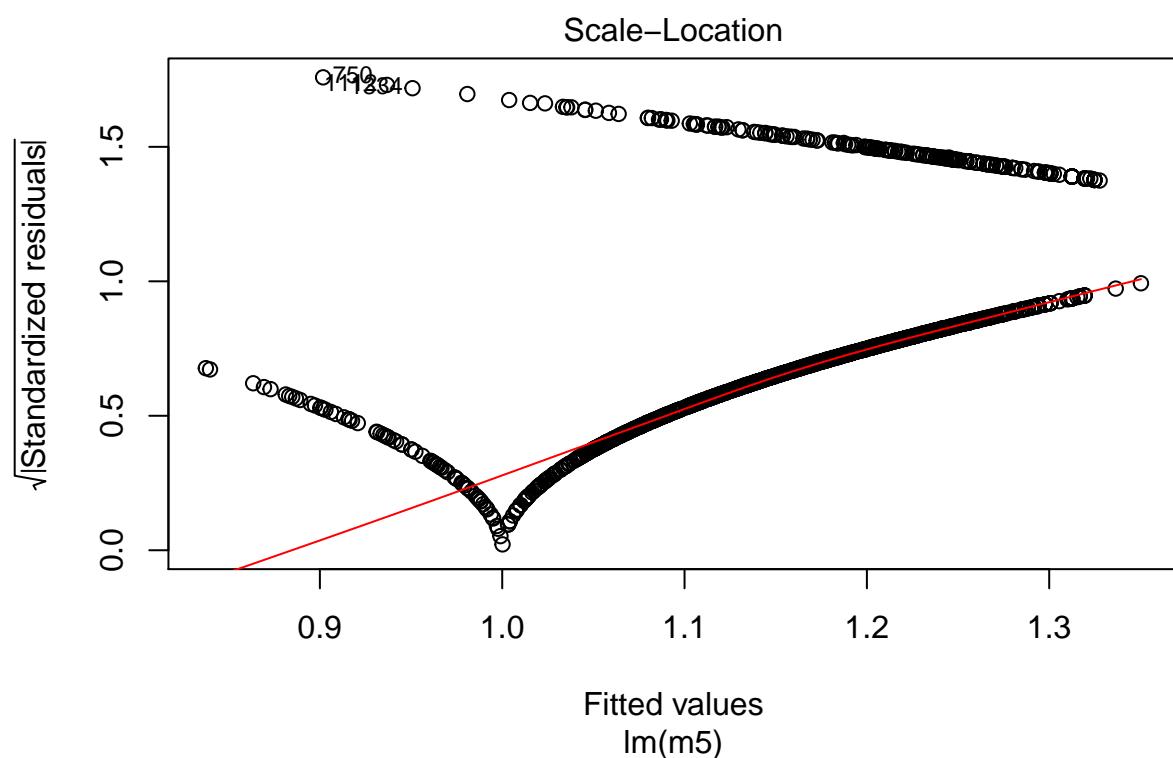
```

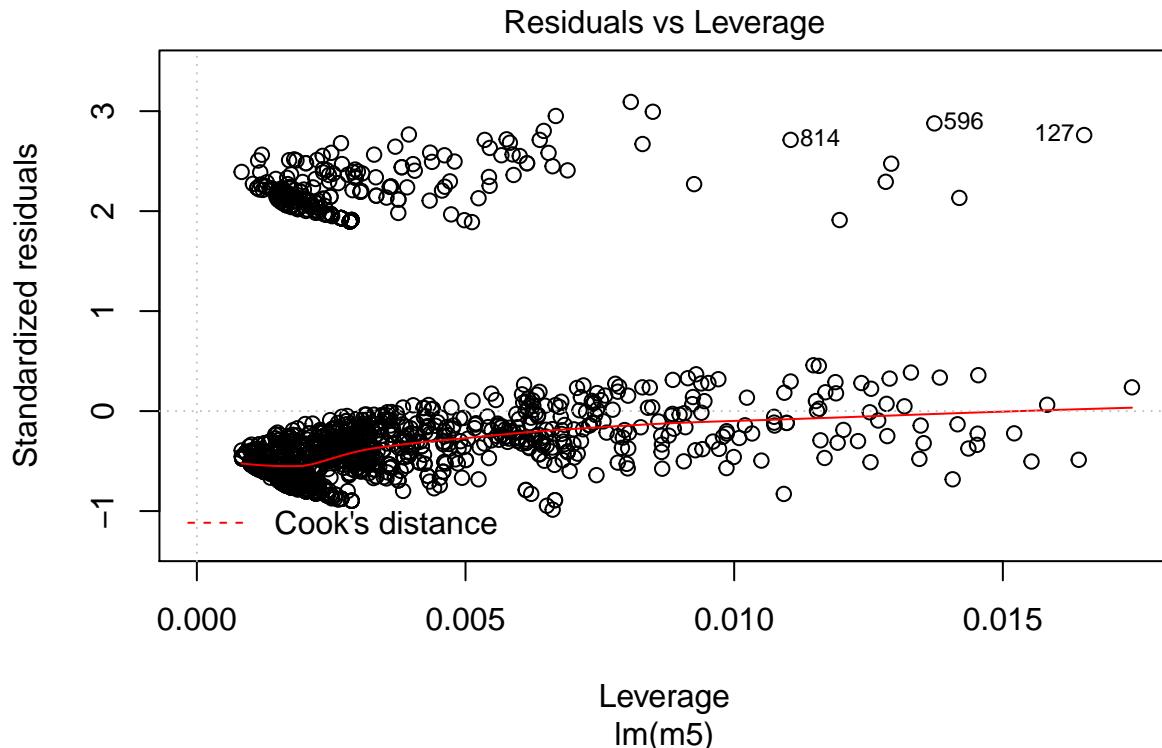
##      Min     1Q   Median     3Q    Max
## -0.35031 -0.20607 -0.13577 -0.03972 1.09823
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.319384  0.019143 68.921 < 2e-16 ***
## StockOptionLevel     -0.056040  0.010937 -5.124 3.39e-07 ***
## TotalWorkingYears    -0.006920  0.001377 -5.027 5.60e-07 ***
## YearsInCurrentRole   -0.014798  0.003236 -4.573 5.22e-06 ***
## YearsSinceLastPromotion 0.012326  0.003528  3.494  0.00049 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3566 on 1465 degrees of freedom
## Multiple R-squared:  0.06267, Adjusted R-squared:  0.06011
## F-statistic: 24.49 on 4 and 1465 DF, p-value: < 2.2e-16
plot(f5)

```









MODEL 6

```

ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
m6 <- Attrition ~ i..Age + DistanceFromHome + EnvironmentSatisfaction + JobInvolvement + JobSatisfaction +
  NumCompaniesWorked + RelationshipSatisfaction + TrainingTimesLastYear + WorkLifeBalance +
  YearsInCurrentRole + YearsSinceLastPromotion + YearsWithCurrManager

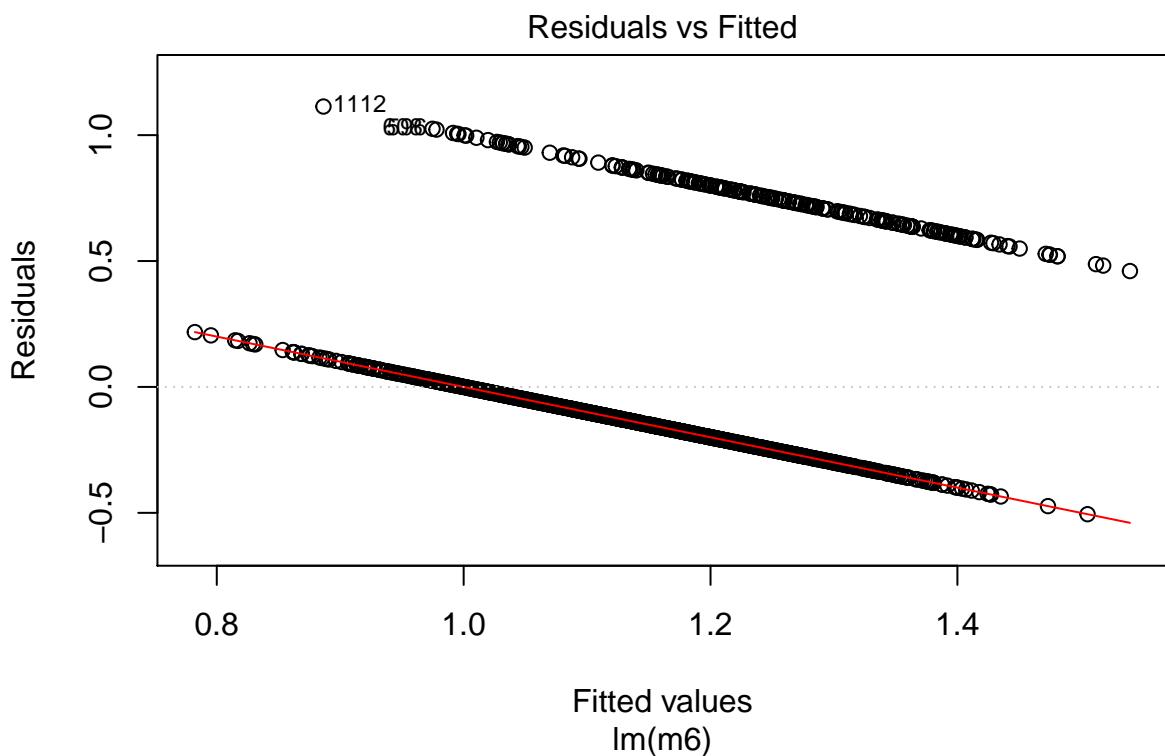
f6<-lm(m6,data=ds)
summary(f6)

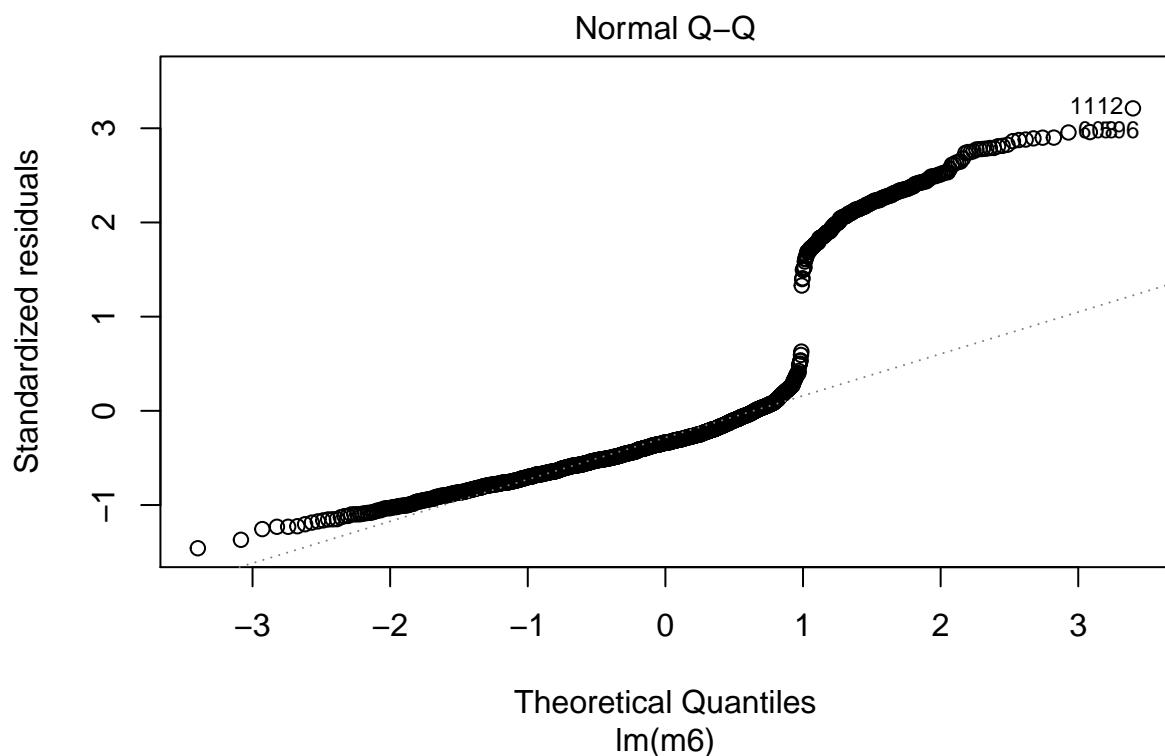
##
## Call:
## lm(formula = m6, data = ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.50542 -0.20300 -0.11714  0.00538  1.11358 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             1.931118   0.077068 25.057 < 2e-16 ***
## i..Age                  -0.006350   0.001092 -5.818 7.32e-09 ***
## DistanceFromHome        0.003526   0.001124  3.138 0.001735 ** 
## EnvironmentSatisfaction -0.034553   0.008329 -4.149 3.54e-05 ***
```

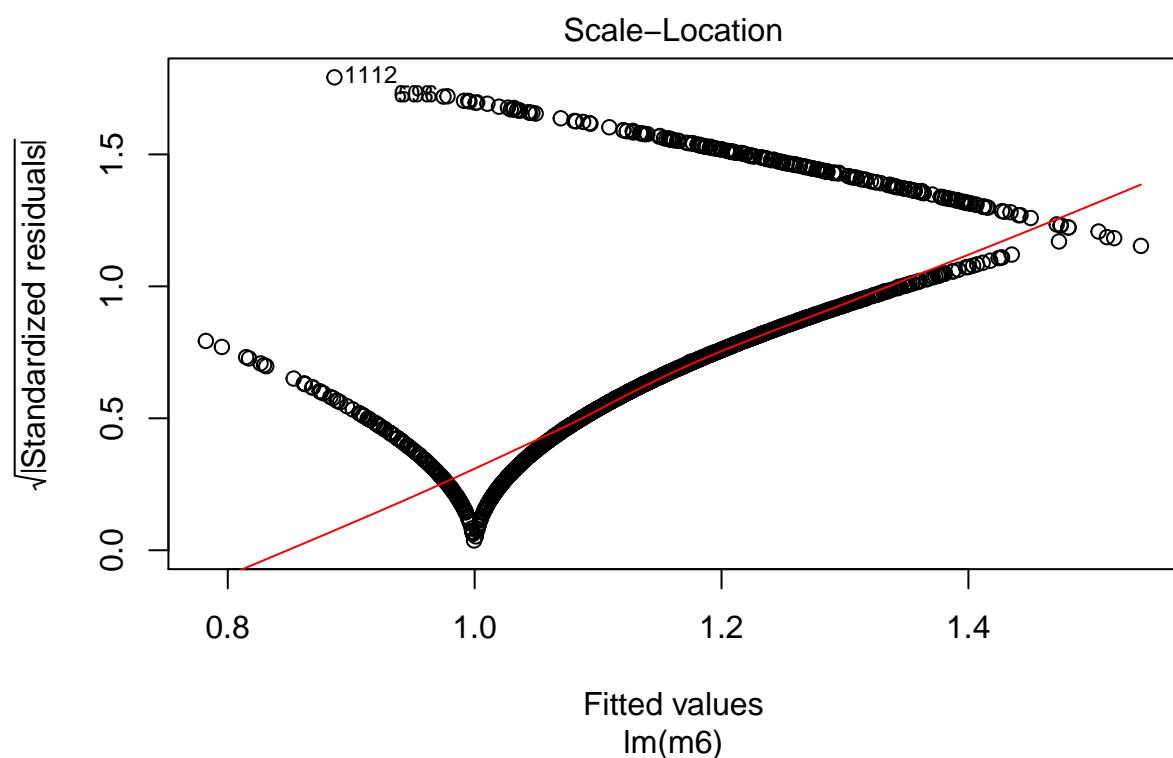
```

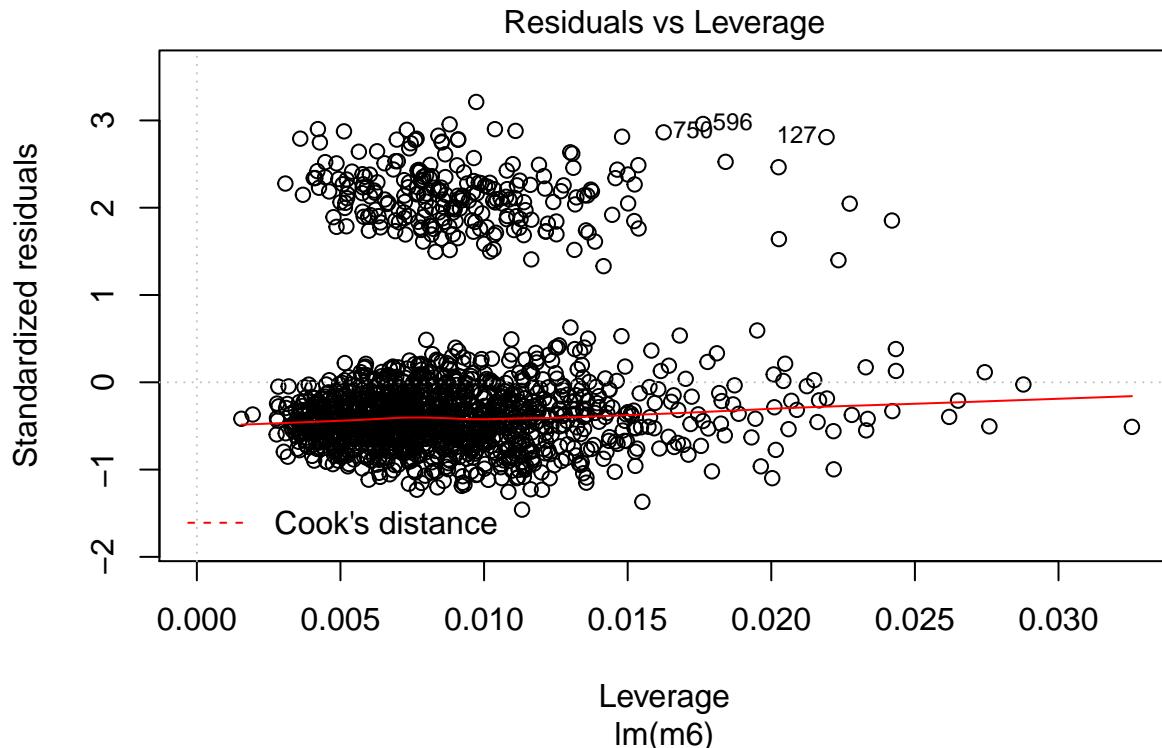
## JobInvolvement      -0.064307  0.012813 -5.019 5.84e-07 ***
## JobSatisfaction    -0.035441  0.008270 -4.285 1.94e-05 ***
## NumCompaniesWorked  0.010555  0.003906  2.702 0.006968 **
## RelationshipSatisfaction -0.014332 0.008445 -1.697 0.089895 .
## TrainingTimesLastYear -0.016697 0.007078 -2.359 0.018453 *
## WorkLifeBalance     -0.030479  0.012927 -2.358 0.018512 *
## YearsInCurrentRole  -0.011063  0.003792 -2.917 0.003583 **
## YearsSinceLastPromotion  0.012603  0.003473  3.629 0.000295 ***
## YearsWithCurrManager -0.009936  0.003748 -2.651 0.008119 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3484 on 1457 degrees of freedom
## Multiple R-squared:  0.1101, Adjusted R-squared:  0.1028
## F-statistic: 15.02 on 12 and 1457 DF,  p-value: < 2.2e-16
plot(f6)

```









MODEL 7

```

ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
m7 <- Attrition~BusinessTravel+DistanceFromHome+EnvironmentSatisfaction+JobInvolvement+JobSatisfaction+
  MonthlyIncome+MonthlyRate+NumCompaniesWorked+OverTime+PercentSalaryHike+PerformanceRating+
  RelationshipSatisfaction+WorkLifeBalance+YearsInCurrentRole+YearsSinceLastPromotion+YearsWithCurrManager

f7<-lm(m7,data=ds)
summary(f7)

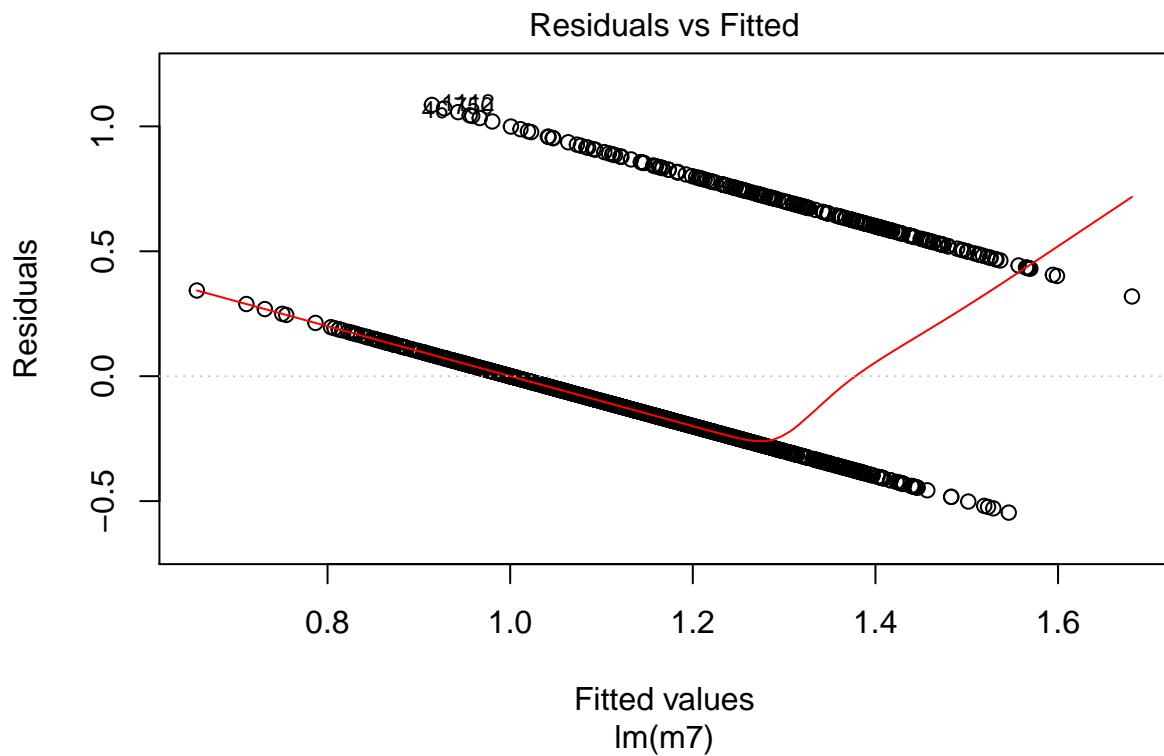
##
## Call:
## lm(formula = m7, data = ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.54600 -0.20668 -0.10262  0.04462  1.08574 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               1.600e+00  1.122e-01 14.261 < 2e-16 ***
## BusinessTravel_Travel_Frequently 1.591e-01  3.407e-02  4.670 3.29e-06 ***
## BusinessTravel_Travel_Rarely     6.615e-02  2.938e-02  2.251  0.02453 *  
## DistanceFromHome            3.273e-03  1.082e-03  3.026  0.00252 ** 
## 
```

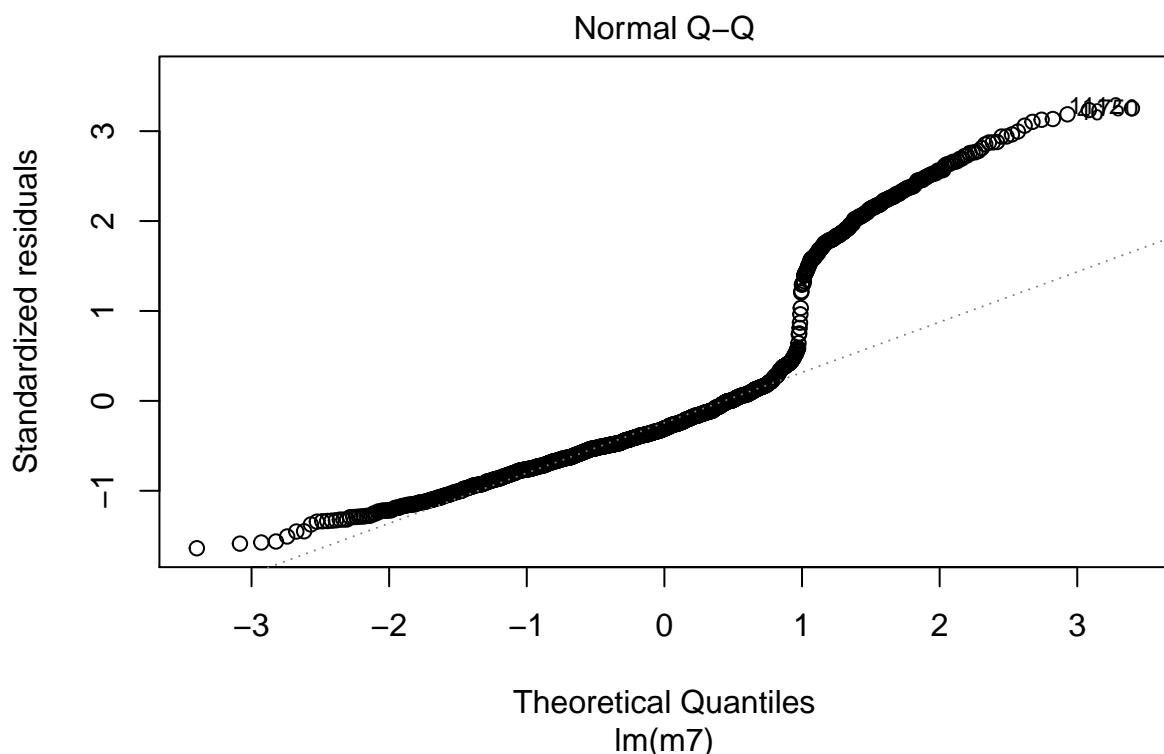
```

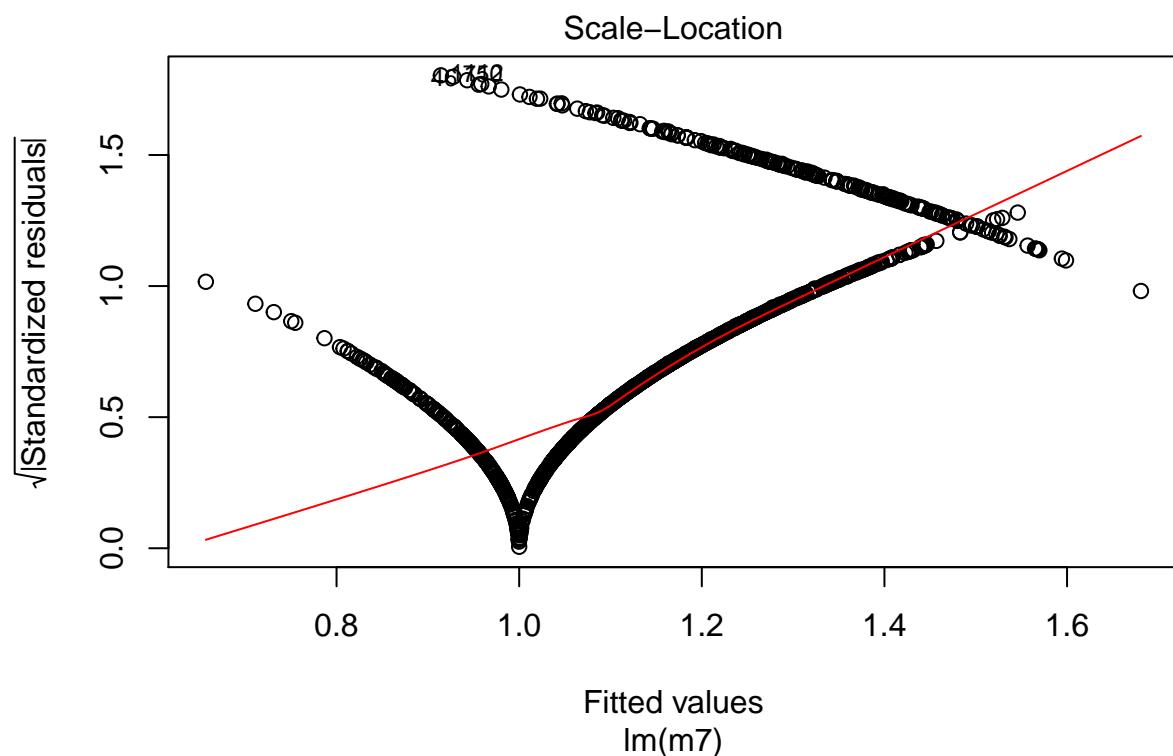
## EnvironmentSatisfaction -4.088e-02 8.038e-03 -5.086 4.14e-07 ***
## JobInvolvement -6.833e-02 1.234e-02 -5.539 3.60e-08 ***
## JobSatisfaction -3.786e-02 7.960e-03 -4.756 2.17e-06 ***
## MonthlyIncome -1.135e-05 2.091e-06 -5.426 6.76e-08 ***
## MonthlyRate 5.610e-07 1.232e-06 0.455 0.64904
## NumCompaniesWorked 9.264e-03 3.616e-03 2.562 0.01050 *
## OverTimeYes 2.031e-01 1.955e-02 10.390 < 2e-16 ***
## PercentSalaryHike -3.624e-03 3.784e-03 -0.958 0.33839
## PerformanceRating 2.022e-02 3.836e-02 0.527 0.59818
## RelationshipSatisfaction -2.066e-02 8.137e-03 -2.539 0.01122 *
## WorkLifeBalance -2.439e-02 1.243e-02 -1.962 0.04994 *
## YearsInCurrentRole -9.762e-03 3.668e-03 -2.661 0.00787 **
## YearsSinceLastPromotion 1.275e-02 3.369e-03 3.786 0.00016 ***
## YearsWithCurrManager -8.183e-03 3.623e-03 -2.259 0.02405 *

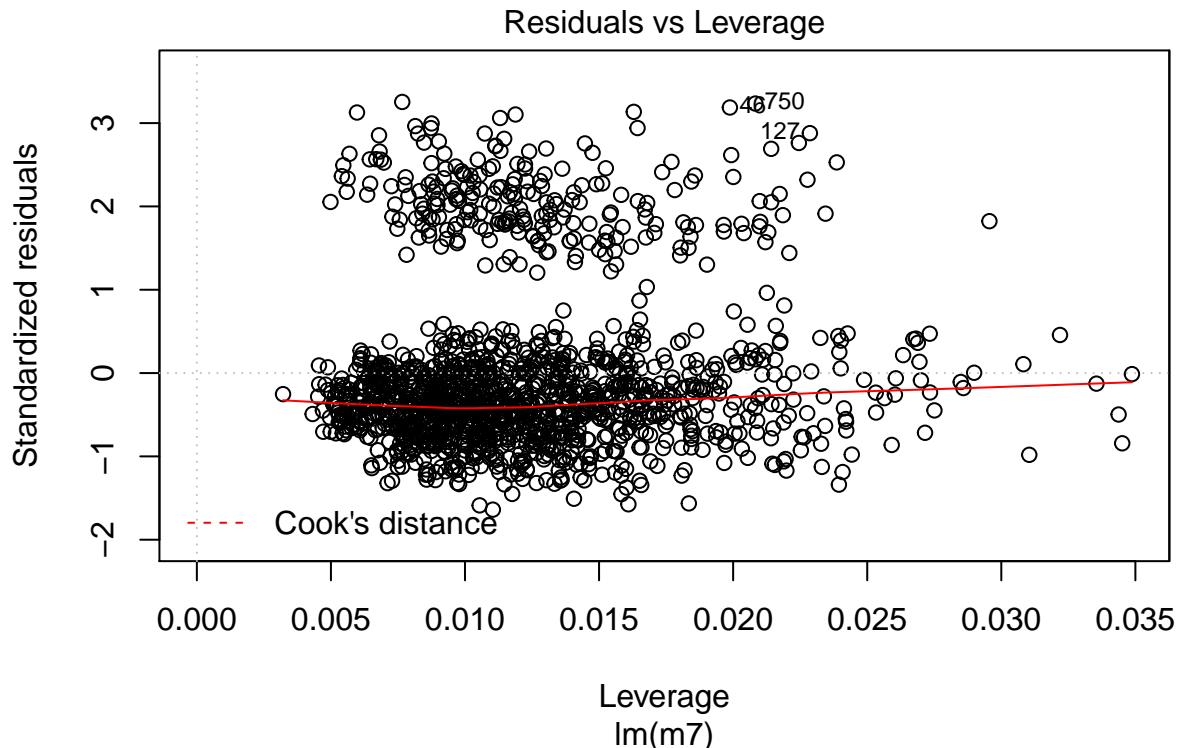
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.335 on 1452 degrees of freedom
## Multiple R-squared: 0.1803, Adjusted R-squared: 0.1707
## F-statistic: 18.79 on 17 and 1452 DF, p-value: < 2.2e-16
plot(f7)

```









Prediction using Model 1

```

ds[, c(2)] <- sapply(ds[, c(2)], as.numeric)
set.seed(2)
library(caTools)

split<-sample.split(ds$Attrition, SplitRatio = 0.7)
train_data<-subset(ds,split=="TRUE")
test_data<-subset(ds,split=="FALSE")

library(car)

m1 <- Attrition ~ i..Age+BusinessTravel+DailyRate+Department+DistanceFromHome+Education+EducationField+
EmployeeNumber+EnvironmentSatisfaction+Gender+HourlyRate+JobInvolvement+JobLevel+JobRole+
JobSatisfaction+MaritalStatus+MonthlyIncome+MonthlyRate+NumCompaniesWorked+OverTime+
PercentSalaryHike+PerformanceRating+RelationshipSatisfaction+StockOptionLevel+
TotalWorkingYears+TrainingTimesLastYear+WorkLifeBalance+YearsAtCompany+
YearsInCurrentRole+YearsSinceLastPromotion+YearsWithCurrManager

model<-lm(m1,data=ds)
vif(model)

##                                     GVIF Df GVIF^(1/(2*Df))
## i..Age                           2.083404  1      1.443400

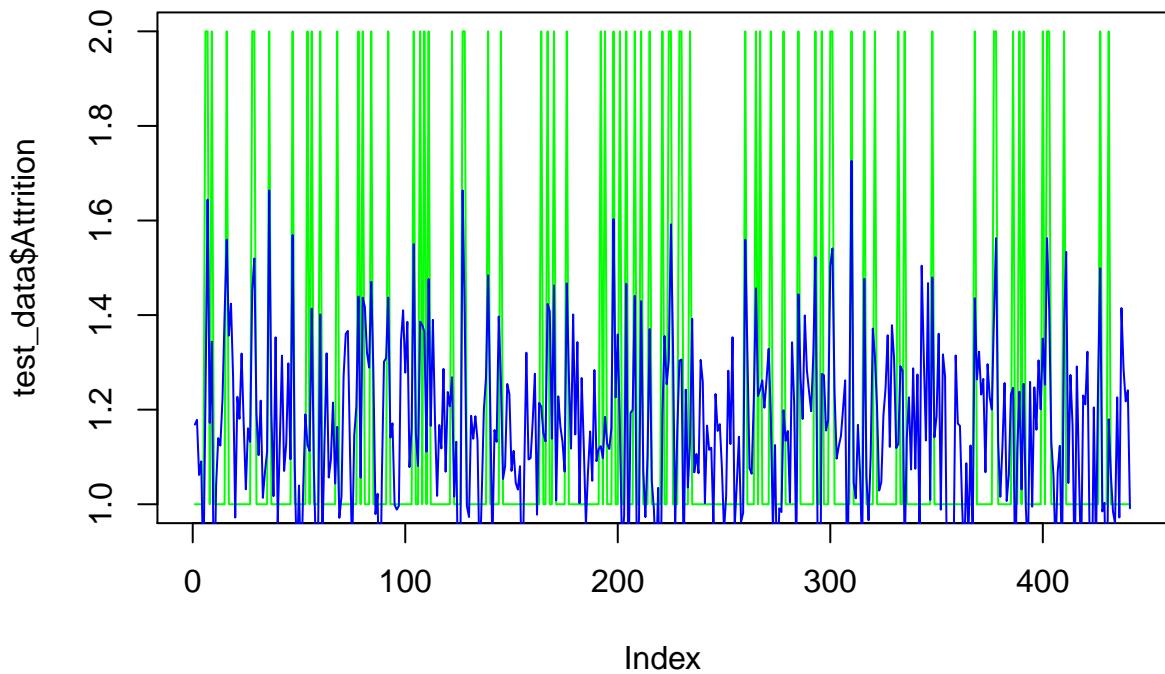
```

```

## BusinessTravel          1.050939  2      1.012498
## DailyRate                1.037997  1      1.018821
## Department            110.107406  2      3.239322
## DistanceFromHome        1.023764  1      1.011812
## Education               1.085391  1      1.041821
## EducationField           2.740818  5      1.106084
## EmployeeNumber           1.035944  1      1.017813
## EnvironmentSatisfaction 1.030723  1      1.015245
## Gender                   1.032778  1      1.016257
## HourlyRate               1.027692  1      1.013751
## JobInvolvement           1.032529  1      1.016134
## JobLevel                 14.157614  1      3.762660
## JobRole                  753.422074  8      1.512916
## JobSatisfaction          1.027094  1      1.013457
## MaritalStatus             2.002689  2      1.189607
## MonthlyIncome             18.152179  1      4.260537
## MonthlyRate               1.021823  1      1.010853
## NumCompaniesWorked        1.282293  1      1.132384
## OverTime                  1.035040  1      1.017369
## PercentSalaryHike         2.564495  1      1.601404
## PerformanceRating         2.550755  1      1.597108
## RelationshipSatisfaction 1.032391  1      1.016067
## StockOptionLevel          1.922337  1      1.386484
## TotalWorkingYears          5.012936  1      2.238959
## TrainingTimesLastYear     1.037466  1      1.018561
## WorkLifeBalance            1.029394  1      1.014591
## YearsAtCompany             4.753637  1      2.180284
## YearsInCurrentRole         2.795648  1      1.672019
## YearsSinceLastPromotion   1.717851  1      1.310668
## YearsWithCurrManager       2.847138  1      1.687347

predic<-predict(model,test_data)
plot(test_data$Attrition,type="l",lty=1.8,col="green")
lines(predic,type = "l",col="blue")

```



From the above linear regression analysis we can conclude that Model 1 is the best fit model as it's R squared value is the highest.

Summary of linear regression

1. Employees who travel the most quit the most.
2. Employees who overtime quit the job.
3. Employees who have not been promoted for long quit the job.
4. Employees who are not satisfied with the job quit.
5. Employees who are not satisfied with the working environment are not much involved in the job and quit the most.
6. It can be noticed from models that mostly single, Lab Technician and Sales Representative quit the job.
7. Employees who are not having a good relationship with their colleague and manager quit the job.
8. Employees who are not having a stable work life balance quit the job.
9. Employees who leave near to the company strangely quit the job.

Logistic Regression

```
ds1 <- read.csv(paste("ibm.csv", sep=""))
ds1$Attrition<-ifelse(ds1$Attrition=="Yes",1,0)
set.seed(2)
library(caTools)
split<-sample.split(ds1$Attrition,SplitRatio = 0.7)
train_data<-subset(ds1,split=="TRUE")
test_data<-subset(ds1,split=="FALSE")
library(car)
m1 <-Attrition~ i..Age+BusinessTravel+DailyRate+Department+DistanceFromHome+Education+
  EducationField+EmployeeNumber+EnvironmentSatisfaction+Gender+HourlyRate+
  JobInvolvement+JobLevel+JobRole+JobSatisfaction+MaritalStatus+MonthlyIncome+
  MonthlyRate+NumCompaniesWorked+OverTime+PercentSalaryHike+PerformanceRating+
  RelationshipSatisfaction+StockOptionLevel+TotalWorkingYears+TrainingTimesLastYear+
  WorkLifeBalance+YearsAtCompany+YearsInCurrentRole+YearsSinceLastPromotion+YearsWithCurrManager

model<-glm(m1,train_data,family = "binomial")
summary(model)

##
## Call:
## glm(formula = m1, family = "binomial", data = train_data)
##
## Deviance Residuals:
##      Min        1Q     Median        3Q       Max
## -1.8346   -0.4845   -0.2543   -0.0996    3.3348
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -1.140e+01  4.669e+02  -0.024 0.980519
## i..Age                   -2.319e-02  1.573e-02  -1.474 0.140347
## BusinessTravelTravel_Frequently  1.727e+00  4.709e-01   3.668 0.000245
## BusinessTravelTravel_Rarely    7.505e-01  4.282e-01   1.753 0.079666
## DailyRate                  -2.587e-04  2.650e-04  -0.976 0.328862
## DepartmentResearch & Development 1.280e+01  4.669e+02   0.027 0.978119
## DepartmentSales             1.314e+01  4.669e+02   0.028 0.977551
## DistanceFromHome            4.384e-02  1.283e-02   3.418 0.000631
## Education                  3.103e-02  1.047e-01   0.296 0.766978
## EducationFieldLife Sciences -4.080e-01  1.035e+00  -0.394 0.693457
## EducationFieldMarketing     -1.318e-01  1.090e+00  -0.121 0.903778
## EducationFieldMedical       -4.946e-01  1.036e+00  -0.477 0.633067
## EducationFieldOther          -7.027e-01  1.146e+00  -0.613 0.539797
## EducationFieldTechnical Degree  3.076e-01  1.080e+00   0.285 0.775798
## EmployeeNumber              -1.627e-04  1.828e-04  -0.890 0.373445
## EnvironmentSatisfaction     -3.728e-01  9.939e-02  -3.751 0.000176
## GenderMale                  3.790e-01  2.240e-01   1.692 0.090662
## HourlyRate                  5.999e-04  5.406e-03   0.111 0.911642
## JobInvolvement               -5.728e-01  1.450e-01  -3.951 7.78e-05
## JobLevel                     4.367e-01  3.843e-01   1.136 0.255787
## JobRoleHuman Resources       1.421e+01  4.669e+02   0.030 0.975714
## JobRoleLaboratory Technician 1.286e+00  5.481e-01   2.346 0.018959
## JobRoleManager                4.050e-01  9.724e-01   0.417 0.677004
```

```

## JobRoleManufacturing Director      -3.012e-01  6.405e-01  -0.470  0.638212
## JobRoleResearch Director         -6.899e-01  1.083e+00  -0.637  0.523926
## JobRoleResearch Scientist        3.295e-01  5.636e-01   0.585  0.558850
## JobRoleSales Executive          5.468e-01  1.174e+00   0.466  0.641402
## JobRoleSales Representative     1.369e+00  1.241e+00   1.103  0.269884
## JobSatisfaction                  -4.117e-01  9.887e-02  -4.164  3.13e-05
## MaritalStatusMarried            7.830e-01  3.365e-01   2.327  0.019978
## MaritalStatusSingle             1.525e+00  4.265e-01   3.576  0.000349
## MonthlyIncome                   -8.019e-05  9.617e-05  -0.834  0.404415
## MonthlyRate                     1.105e-05  1.540e-05   0.718  0.473027
## NumCompaniesWorked              2.092e-01  4.665e-02   4.484  7.32e-06
## OverTimeYes                     1.958e+00  2.342e-01   8.361  < 2e-16
## PercentSalaryHike               -7.075e-03  4.685e-02  -0.151  0.879978
## PerformanceRating                1.989e-02  4.879e-01   0.041  0.967475
## RelationshipSatisfaction       -3.365e-01  9.845e-02  -3.418  0.000630
## StockOptionLevel                 7.905e-02  1.828e-01  -0.432  0.665431
## TotalWorkingYears                9.197e-02  3.587e-02  -2.564  0.010346
## TrainingTimesLastYear           1.901e-01  8.970e-02  -2.120  0.034026
## WorkLifeBalance                 2.854e-01  1.518e-01  -1.880  0.060093
## YearsAtCompany                  1.002e-01  4.661e-02   2.151  0.031507
## YearsInCurrentRole              1.320e-01  5.374e-02  -2.457  0.014012
## YearsSinceLastPromotion        1.485e-01  5.002e-02   2.969  0.002985
## YearsWithCurrManager            1.254e-01  5.352e-02  -2.343  0.019134
##
## (Intercept)
## i..Age
## BusinessTravelTravel_Frequently *** 
## BusinessTravelTravel_Rarely    .
## DailyRate
## DepartmentResearch & Development
## DepartmentSales
## DistanceFromHome                ***
## Education
## EducationFieldLife Sciences
## EducationFieldMarketing
## EducationFieldMedical
## EducationFieldOther
## EducationFieldTechnical Degree
## EmployeeNumber
## EnvironmentSatisfaction      ***
## GenderMale
## HourlyRate
## JobInvolvement                  ***
## JobLevel
## JobRoleHuman Resources
## JobRoleLaboratory Technician    *
## JobRoleManager
## JobRoleManufacturing Director
## JobRoleResearch Director
## JobRoleResearch Scientist
## JobRoleSales Executive
## JobRoleSales Representative
## JobSatisfaction                  ***
## MaritalStatusMarried            *

```

```

## MaritalStatusSingle           ***
## MonthlyIncome
## MonthlyRate
## NumCompaniesWorked          ***
## OverTimeYes                  ***
## PercentSalaryHike
## PerformanceRating
## RelationshipSatisfaction    ***
## StockOptionLevel
## TotalWorkingYears            *
## TrainingTimesLastYear        *
## WorkLifeBalance              .
## YearsAtCompany               *
## YearsInCurrentRole           *
## YearsSinceLastPromotion     **
## YearsWithCurrManager         *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 909.34  on 1028  degrees of freedom
## Residual deviance: 604.72  on  983  degrees of freedom
## AIC: 696.72
##
## Number of Fisher Scoring iterations: 14
res<-predict(model,train_data,type="response")
(table(ActualValue=train_data$Attrition,PredictedValue=res>0.5))

##             PredictedValue
## ActualValue FALSE TRUE
##      0     840   23
##      1     89    77

library(ROCR)

## Loading required package: gplots

##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
## 
##      lowess

ROCRPred<-prediction(res,train_data$Attrition)
ROCRPerf<-performance(ROCRPred,"tpr","fpr")
plot(ROCRPerf,colorize=TRUE,print.cutoffs.at=seq(0.1,by=0.1))

```

