DDS 6306 - Case Study #2

David Loveday

4/11/2021

Youtube Presentation Link: https://www.youtube.com/watch?v=wSlwNIm8y2M (https://www.youtube.com/watch?v=wSlwNIm8y2M)

GitHub Webpage Link: https://dloveday.github.io/ (https://dloveday.github.io/)

Executive Summary

DDSAnalytics has concluded an initial analysis of employee attribute data which demonstrates the ability to predict, using a naïve Bayes classification model, an individual employee's voluntary attrition potential, as well as their monthly income using a multiple linear regression model. Many of these explanatory attributes may already exist in employee files while the others could be easily, and cost-effectively, be collected.

The analysis found these independent variables to be most impactful and their specific demographic values to be most at-risk for attrition:

1. Does the employee work overtime? Most At-Risk: "Yes"
2. Employee's total years with the Company Most At-Risk: 0 – 10
3. Employee's martial status Most At-Risk: "Single"
4. Employee's tenure in their current role Most At-Risk: 0 - 4
5. Monthly Income Most At-Risk: $0 - $5,811
6. Department in which employee works Most At-Risk: "Sales"
7. Role held by the employee Most At-Risk: "Sales Rep"
8. Age of employee Most At-Risk: 18 – 28
9. Employee's tenure with their current manager Most At-Risk: 0 - 4

DDSAnalytics has also found that a relatively simple multiple linear regression (MLR) model can effectively describe, and predict, an employee's monthly income.

The MLR model below achieves a statistically significant (p-value << 0.05) solution with an Adjusted R2 = 91%.

$$\boldsymbol{MonthlyIncome} = \boldsymbol{\beta 0 + \beta 1 Distance + \beta 2 JobLevel + \beta 3 PercSalaryHike + \beta 4 TotalWorkingYears + \beta 5 YearsWithCurrentManager}$$

Residuals: Min 1Q Median 3Q Max -5759 -872 16 740 4035

Coefficients: Estimate Std. Error t value Pr(>|t|)
$\boldsymbol{\beta 0}$ (Intercept) -1707.30 227.30 -7.51 1.5e-13 $\boldsymbol{\beta 1}$ DistanceFromHome -15.57 5.74 -2.71 0.0068 $\boldsymbol{\beta 2}$ JobLevel 3723.77 68.43 54.41 < 2e-16 $\boldsymbol{\beta 3}$ PercentSalaryHike 9.57 12.72 0.75 0.4519
$\boldsymbol{\beta 4}$ TotalWorkingYears 68.12 10.41 6.54 1.0e-10 $\boldsymbol{\beta 5}$ YearsWithCurrManager -60.04 14.70 -4.09 4.8e-05

Residual standard error: 1370 on 864 degrees of freedom Multiple R-squared: 0.911, Adjusted R-squared: 0.911
F-statistic: 1.78e+03 on 5 and 864 DF, p-value: <2e-16

Project Desciption

Talent management is defined as the iterative process of developing and retaining employees. It may include workforce planning, employee training programs, identifying high-potential employees and reducing/preventing voluntary employee turnover (attrition). To gain a competitive edge over its competition, DDSAnalytics is planning to leverage data science for talent management. The executive leadership has identified predicting employee turnover as its first application of data science for talent management. Before the business green lights the project, they have tasked your data science team to conduct an analysis of existing employee data.

Project Deliverables

1. Identify the top factors which contribute tot turnover. Clearly document and defend the analysis.
2. Discuss any other material insights, trends, or observations gleaned from the dataset.
3. Construct models which predict attrition and monthly income

Dataset Description

The dataset provided to the project team captures general demographic information for each employee (example: age, gender, marital status) company-specific information (example: total years with the company, department, role), along with employee-reported satisfaction scores for several areas (example: Work Life Balance, Environment Satisfaction).

- Total rows: 870
- Total gross/net columns: 36/32 – ID/EmployeeCount/Over18/StandardHours [Columns with singular values removed from the analysis]
- Character/Categorical Fields: 8
- Numeric Fields: 24 – 10 integer fields behave more as categorical variables

Mercifully, the dataset did not contain any missing or null values

WORKFLOW & OUTPUTS

Initialize libraries

```
library(ggplot2)
library(tidyverse)
library(dplyr)
library(tidyr)
library(maps)
library(ggthemes)
library(plotly)
library(reshape)
library(githubinstall)
library(envirofacts)
library(eia)
library(stringr)
library(mapproj)
library(countrycode)
library(WDI)
library(stringr)
library(jsonlite)
library(plyr)
library(Rmisc)
library(class)
library(caret)
library(e1071)
library(scales)
library(RCurl) #getURL
library(rvest) #html_table, html_node
library(maps)
library(fiftystater)
library(mapproj)
library(GGally)
library(fpp)
library(shiny)
library(psych)
library(Hmisc)
library(corrplot)
library(caTools)
library(rpart)
library(data.table)
library(DT)
library(gridExtra)
library(Metrics)
library(randomForest)
library(pROC)
library(car)
library(asbio)
library(boot)
library(DAAG)
library(olsrr)
```

Import Data / Format & Condition

```
path.employee.dB <- "C:\\Users\\dloveday\\Dropbox\\Family\\School\\SMU\\Courses\\Spring 2021\\DS
6306 - Doing Data Science\\Lecture Notes\\Unit 14 and 15 Case Study 2\\CaseStudy2-data.csv"

make.eda.plots <- "Yes"
  column.histograms <- "Yes"
  demographics.plots <- "Yes"
  correlation.matrix.plots <- "Yes"
#
######################## FORMAT & CONDITION dB ########################
#
employee.dB <- read.csv(path.employee.dB)
raw.employe.dB <- read.csv(path.employee.dB)
```

Data Diagnostics

```
count.duplicates <- sum(duplicated(employee.dB))
summary.employee.dB <- summary(employee.dB)
```

Variable Conversions, Numericizations & Factorizations

```
# Make corresponding numeric field for "Attrition", if 'Yes' (= 1), if 'No' (= 0)
employee.dB$Attrition.Numeric <- ifelse(employee.dB$Attrition == "Yes", 1, 0)

# Make corresponding numeric field for "BusinessTravel"
employee.dB$BusinessTravel.Non_Travel <- ifelse(employee.dB$BusinessTravel == "Non-Travel", 1, 0
)
employee.dB$BusinessTravel.Travel_Rarely <- ifelse(employee.dB$BusinessTravel == "Travel_Rarely"
, 1, 0)
employee.dB$BusinessTravel.Travel_Frequently <- ifelse(employee.dB$BusinessTravel == "Travel_Fre
quently", 1, 0)

# Make corresponding numeric field for "Department"
employee.dB$Department.Sales <- ifelse(employee.dB$Department == "Sales", 1, 0)
employee.dB$Department.Research_Development <- ifelse(employee.dB$Department == "Research & Deve
lopment", 1, 0)
employee.dB$Department.Human_Resources <- ifelse(employee.dB$Department == "Human Resources", 1,
0)

# Make corresponding numeric field for "EducationalField"
employee.dB$EducationField.LifeSciences <- ifelse(employee.dB$EducationField == "Life Sciences",
1, 0)
employee.dB$EducationField.Medical <- ifelse(employee.dB$EducationField == "Medical", 1, 0)
employee.dB$EducationField.Marketing <- ifelse(employee.dB$EducationField == "Marketing", 1, 0)
employee.dB$EducationField.TechnicalDegree <- ifelse(employee.dB$EducationField == "Technical De
gree", 1, 0)
employee.dB$EducationField.Other <- ifelse(employee.dB$EducationField == "Other", 1, 0)
employee.dB$EducationField.HumanResources <- ifelse(employee.dB$EducationField == "Human Resourc
es", 1, 0)

# Make corresponding numeric field for "Gender"
for (i in 1:nrow(employee.dB)){
  if(employee.dB$Gender[i] == "Male") (employee.dB$Gender.Numeric[i] <- 0)
  if(employee.dB$Gender[i] == "Female") (employee.dB$Gender.Numeric[i] <- 1)
}

# Make corresponding numeric field for "JobRole"
employee.dB$JobRole.SalesExecutive <- ifelse(employee.dB$JobRole == "Sales Executive", 1, 0)
employee.dB$JobRole.ResearchDirector <- ifelse(employee.dB$JobRole == "Research Director", 1, 0)
employee.dB$JobRole.ManufacturingDirector <- ifelse(employee.dB$JobRole == "Manufacturing Direct
or", 1, 0)
employee.dB$JobRole.ResearchScientist <- ifelse(employee.dB$JobRole == "Research Scientist", 1,
0)
employee.dB$JobRole.SalesRepresentative <- ifelse(employee.dB$JobRole == "Sales Representative",
1, 0)
employee.dB$JobRole.HealthcareRepresentative <- ifelse(employee.dB$JobRole == "Healthcare Repres
entative", 1, 0)
employee.dB$JobRole.Manager <- ifelse(employee.dB$JobRole == "Manager", 1, 0)
employee.dB$JobRole.HumanResources <- ifelse(employee.dB$JobRole == "Human Resources", 1, 0)
employee.dB$JobRole.LaboratoryTechnician <- ifelse(employee.dB$JobRole == "Laboratory Technicia
n", 1, 0)

# Make corresponding numeric field for "MaritalStatus"
employee.dB$MaritalStatus.Divorced <- ifelse(employee.dB$MaritalStatus == "Divorced", 1, 0)
```

```
employee.dB$MaritalStatus.Single <- ifelse(employee.dB$MaritalStatus == "Single", 1, 0)
employee.dB$MaritalStatus.Married <- ifelse(employee.dB$MaritalStatus == "Married", 1, 0)

# Make corresponding numeric field for "OverTime"
for (i in 1:nrow(employee.dB)){
  if(employee.dB$OverTime[i] == "No") (employee.dB$OverTime.Numeric[i] <- 0)
  if(employee.dB$OverTime[i] == "Yes") (employee.dB$OverTime.Numeric[i] <- 1)
}



categorical.string.variables <- c("BusinessTravel", "Department", "EducationField", "Gender", "J
obRole", "MaritalStatus", "OverTime")

categorical.string.binary.variables <- c("BusinessTravel.Non_Travel", "BusinessTravel.Travel_Rar
ely", "BusinessTravel.Travel_Frequently",
                                         "Department.Sales", "Department.Research_Development",
"Department.Human_Resources",
                                         "EducationField.LifeSciences", "EducationField.Medical"
, "EducationField.Marketing", "EducationField.TechnicalDegree", "EducationField.Other", "Educati
onField.HumanResources",
                                         "Gender.Numeric",
                                         "JobRole.SalesExecutive", "JobRole.ResearchDirector",
"JobRole.ManufacturingDirector", "JobRole.ResearchScientist", "JobRole.SalesRepresentative", "Jo
bRole.HealthcareRepresentative", "JobRole.Manager", "JobRole.HumanResources", "JobRole.Laborator
yTechnician",
                                         "MaritalStatus.Divorced", "MaritalStatus.Single", "Mari
talStatus.Married",
                                         "OverTime.Numeric")

categorical.numeric.variables <- c("Education", "EnvironmentSatisfaction", "JobInvolvement", "Jo
bLevel", "JobSatisfaction", "NumCompaniesWorked", "PercentSalaryHike", "TrainingTimesLastYear",
"WorkLifeBalance", "YearsAtCompany",  "YearsInCurrentRole", "YearsSinceLastPromotion", "YearsWit
hCurrManager")

continious.numeric.variables <- c("Age", "DailyRate", "DistanceFromHome", "EmployeeNumber", "Hou
rlyRate", "MonthlyIncome", "MonthlyRate")

all.variables <- c(categorical.string.variables, categorical.numeric.variables, continious.numer
ic.variables)

all.evaluation.variables <- c(categorical.string.variables,
                              c("Education.Quartiles", "EnvironmentSatisfaction.Quartiles", "Job
Involvement.Quartiles", "JobLevel.Quartiles", "JobSatisfaction.Quartiles", "NumCompaniesWorked.Q
uartiles", "PercentSalaryHike.Quartiles", "TrainingTimesLastYear.Quartiles", "WorkLifeBalance.Qu
artiles", "YearsAtCompany.Quartiles",  "YearsInCurrentRole.Quartiles", "YearsSinceLastPromotion.
Quartiles", "YearsWithCurrManager.Quartiles"),
                              c("Age.Quartiles", "DailyRate.Quartiles", "DistanceFromHome.Quarti
les", "EmployeeNumber.Quartiles", "HourlyRate.Quartiles", "MonthlyIncome.Quartiles", "MonthlyRat
e.Quartiles")
                              )


# Calculate and Record Quartiles for All Numeric Variables
```

```
for (i in 1:length(categorical.numeric.variables)) {
  employee.dB[ ,paste0(categorical.numeric.variables[i],".Quartiles") ] <- cut(employee.dB[ ,cat
egorical.numeric.variables[i]], 4, include.lowest = TRUE, labels = c("0-25%", "25-50%", "50-75%"
, "75-100%"))
}

for (i in 1:length(continious.numeric.variables)) {
  employee.dB[ ,paste0(continious.numeric.variables[i],".Quartiles") ] <- cut(employee.dB[ ,cont
inious.numeric.variables[i]], 4, include.lowest = TRUE, labels = c("0-25%", "25-50%", "50-75%",
"75-100%"))
}
```

Calculate Attribute densities & correlations for total, retained employees, lost employee populations

```
# Calculate Attrition Rate by Category - For All & Retained & Lost
categories.dB <- data.frame()

for (i in 1:length(all.evaluation.variables)) {
  #
  unique.vector.temp   <-   unique(employee.dB[, c(paste0(all.evaluation.variables[i]))])
  category.vector.temp <-   rep(c(paste0(all.evaluation.variables[i])), times = length(unique.ve
ctor.temp))
  category.dF.temp <- data.frame(Category = category.vector.temp, Variable = unique.vector.temp)
  #
  categories.dB <- rbind(categories.dB, category.dF.temp)
}

categories.dB$demographic <- paste0(categories.dB$Category," - ", categories.dB$Variable)

retained.employee.dB <- filter(employee.dB, employee.dB$Attrition == "No")
lost.employee.dB <- filter(employee.dB, employee.dB$Attrition == "Yes")

for (i in 1:nrow(categories.dB)) {
  #
  categories.dB$all.demographic.count[i] <- sum( ldply( employee.dB[, categories.dB$Category
[i]], function(c) sum(c==categories.dB$Variable[i]) ) )

  categories.dB$retained.demographic.count[i] <- sum( ldply( retained.employee.dB[, categories.d
B$Category[i]], function(c) sum(c==categories.dB$Variable[i]) ) )
  categories.dB$retained.demographic.density[i] <- categories.dB$retained.demographic.count[i] /
nrow(retained.employee.dB)

  categories.dB$lost.demographic.count[i] <- sum( ldply( lost.employee.dB[, categories.dB$Catego
ry[i]], function(c) sum(c==categories.dB$Variable[i]) ) )
  categories.dB$lost.demographic.density[i] <- categories.dB$lost.demographic.count[i] / nrow(lo
st.employee.dB)
  #
}
categories.dB$demographic.density.contrast.delta <- categories.dB$lost.demographic.density - cat
egories.dB$retained.demographic.density
categories.dB$demographic.density.contrast.perc <- (categories.dB$demographic.density.contrast.d
elta / categories.dB$retained.demographic.density) * 100
categories.dB$retention.rate <- categories.dB$retained.demographic.count/categories.dB$all.demog
raphic.count
categories.dB$attrition.rate <- categories.dB$lost.demographic.count/categories.dB$all.demograph
ic.count
#
```

Calculate Pearson’s correlations & Chi-Square Test

```
#
#### Define Numeric Fields to be Used for Correlation Matrix
correlation.matrix.vector <- c(categorical.numeric.variables, continious.numeric.variables, cate
gorical.string.binary.variables)

correlation.mtx.employee.dB <- rcorr(   as.matrix(employee.dB[, c("Attrition.Numeric" ,correlati
on.matrix.vector)])   )
correlation.mtx.employee.dB.coeff <- correlation.mtx.employee.dB$r
correlation.mtx.employee.dB.p <- correlation.mtx.employee.dB$P

positive.over.correlation.mtx <- data.frame(matrix(nrow = 1, ncol = 3))
colnames(positive.over.correlation.mtx) <- c("Var1", "Var2", "correl.coeff")

negative.over.correlation.mtx <- data.frame(matrix(nrow = 1, ncol = 3))
colnames(negative.over.correlation.mtx) <- c("Var1", "Var2", "correl.coeff")


correlation.threshold <- 0.75

for (i in 1:ncol(correlation.mtx.employee.dB.coeff)) {
  for (j in 1:nrow(correlation.mtx.employee.dB.coeff)) {
    if (correlation.mtx.employee.dB.coeff[j,i] > correlation.threshold) {
      positive.over.correlation.mtx <- rbind(positive.over.correlation.mtx, c((colnames(correlat
ion.mtx.employee.dB.coeff))[i], (rownames(correlation.mtx.employee.dB.coeff))[j],  correlation.m
tx.employee.dB.coeff[j,i]              ))
    }
    if (correlation.mtx.employee.dB.coeff[j,i] < correlation.threshold*-1) {
      negative.over.correlation.mtx <- rbind(negative.over.correlation.mtx, c((colnames(correlat
ion.mtx.employee.dB.coeff))[i], (rownames(correlation.mtx.employee.dB.coeff))[j],  correlation.m
tx.employee.dB.coeff[j,i]              ))
    }
  }
}
positive.over.correlation.mtx <- filter(positive.over.correlation.mtx, positive.over.correlatio
n.mtx$correl.coeff != 1 & !is.na(positive.over.correlation.mtx))
negative.over.correlation.mtx <- filter(negative.over.correlation.mtx, negative.over.correlatio
n.mtx$correl.coeff != 1 & !is.na(negative.over.correlation.mtx))

#########################################################################################################
#####################################
##################################################### CHI-SQUARED TEST (ATTRITION) ###############
#####################################
#########################################################################################################
#####################################
#
employee.dB.chi.sq <- data.frame(Variable = correlation.matrix.vector)

for (i in 1:nrow(employee.dB.chi.sq)) {
  employee.dB.chi.sq$p.value[i] <- chisq.test(employee.dB$Attrition, employee.dB[,correlation.ma
trix.vector[i]])$p.value
}
influential.variables.chi.sq <- (filter(employee.dB.chi.sq, employee.dB.chi.sq$p.value < 0.05))
$Variable
```

```
employee.dB.Categorical.Binary.chi.sq <- data.frame(Variable = categorical.string.binary.variabl
es)

for (i in 1:nrow(employee.dB.Categorical.Binary.chi.sq)) {
  employee.dB.Categorical.Binary.chi.sq$p.value[i] <- chisq.test(employee.dB$Attrition, employe
e.dB[, categorical.string.binary.variables[i]])$p.value
}
#
########################################################################################################
#####################################
################################################## CHI-SQUARED TEST (MONTHLYINCOME) ###########
########################################
########################################################################################################
#####################################
#
employee.dB.chi.sq.MI <- data.frame(Variable = correlation.matrix.vector)

for (i in 1:nrow(employee.dB.chi.sq.MI)) {
  employee.dB.chi.sq.MI$p.value[i] <- chisq.test(employee.dB$MonthlyIncome, employee.dB[,correla
tion.matrix.vector[i]])$p.value
}
influential.variables.chi.sq.MI <- (filter(employee.dB.chi.sq.MI, employee.dB.chi.sq.MI$p.value
 < 0.05))$Variable

employee.dB.Categorical.Binary.chi.sq.MI <- data.frame(Variable = categorical.string.binary.vari
ables)

for (i in 1:nrow(employee.dB.Categorical.Binary.chi.sq.MI)) {
  employee.dB.Categorical.Binary.chi.sq.MI$p.value[i] <- chisq.test(employee.dB$Attrition, emplo
yee.dB[, categorical.string.binary.variables[i]])$p.value
}
```

STEP 1A - EDA - DEFINE PLOT FUNCTIONS

```
continious.histogram.plot.single.variable <- function(x_name, fill_source) {
  #dev.new()
  employee.dB.plot <- employee.dB[, c(paste0(x_name), paste0(fill_source))]
  employee.dB.plot %>% ggplot(aes(x = employee.dB.plot[,1], y = ..density..))+
    theme_bw()+
    geom_histogram(alpha = 0.35, color = "grey80")+
    geom_density(aes(color = employee.dB.plot[,1]), alpha = 0.01, size = 0.85, show.legend = FAL
SE)+
    labs(x = x_name, fill = "Attrition")
}

continious.histogram.plot <- function(x_name, fill_source) {
  #dev.new()
  employee.dB.plot <- employee.dB[, c(paste0(x_name), paste0(fill_source))]
  employee.dB.plot %>% ggplot(aes(x = employee.dB.plot[,1], y = ..density.., fill = employee.dB.
plot[,2]))+
    theme_bw()+
    geom_histogram(position = "dodge", alpha = 0.35, color = "grey80")+
    geom_density(aes(color = employee.dB.plot[,2]), alpha = 0.01, size = 0.85, show.legend = FAL
SE)+
    labs(x = x_name, fill = "Attrition")
}

categorical.barchart.plot.single.variable <- function(x_name, fill_source) {
  #dev.new()
  employee.dB.plot <- employee.dB[, c(paste0(x_name), paste0(fill_source))]

  employee.dB.plot.table.total <- data.frame(table(employee.dB.plot[,2]))
  var1.name <- employee.dB.plot.table.total[1,1]
  var1.freq <- employee.dB.plot.table.total[1,2]
  var2.name <- employee.dB.plot.table.total[2,1]
  var2.freq <- employee.dB.plot.table.total[2,2]


  employee.dB.plot.table <- table(employee.dB.plot)

  employee.dB.plot.table <- data.frame(table(employee.dB.plot))
  for (i in 1:nrow(employee.dB.plot.table)) {
    if (employee.dB.plot.table[i,2] == var1.name) { employee.dB.plot.table$Density[i] <- employe
e.dB.plot.table$Freq[i] / var1.freq }
    if (employee.dB.plot.table[i,2] == var2.name) { employee.dB.plot.table$Density[i] <- employe
e.dB.plot.table$Freq[i] / var2.freq }
  }

  employee.dB.plot.table %>% ggplot(aes(x = employee.dB.plot.table[,1], y = Density))+
    theme_bw()+
    geom_bar(stat = "identity", alpha = 0.35, color = "grey80")+
    labs(x = x_name)
}



categorical.barchart.plot <- function(x_name, fill_source) {
```

```
  #dev.new()
  employee.dB.plot <- employee.dB[, c(paste0(x_name), paste0(fill_source))]

  employee.dB.plot.table.total <- data.frame(table(employee.dB.plot[,2]))
  var1.name <- employee.dB.plot.table.total[1,1]
  var1.freq <- employee.dB.plot.table.total[1,2]
  var2.name <- employee.dB.plot.table.total[2,1]
  var2.freq <- employee.dB.plot.table.total[2,2]


  employee.dB.plot.table <- table(employee.dB.plot)

  employee.dB.plot.table <- data.frame(table(employee.dB.plot))
  for (i in 1:nrow(employee.dB.plot.table)) {
    if (employee.dB.plot.table[i,2] == var1.name) { employee.dB.plot.table$Density[i] <- employe
e.dB.plot.table$Freq[i] / var1.freq }
    if (employee.dB.plot.table[i,2] == var2.name) { employee.dB.plot.table$Density[i] <- employe
e.dB.plot.table$Freq[i] / var2.freq }
  }


  employee.dB.plot.table %>% ggplot(aes(x = employee.dB.plot.table[,1], y = Density, fill = empl
oyee.dB.plot.table[,2]))+
    theme_bw()+
    geom_bar(stat = "identity", position = "dodge", alpha = 0.35, color = "grey80")+
    labs(x = x_name, fill = fill_source)
}

continious.scatter.plot <- function(x_name, y_name, fill_source) {
  #dev.new()
  employee.dB.plot <- employee.dB[, c(paste0(x_name), paste0(y_name), paste0(fill_source))]
  employee.dB.plot %>% ggplot(aes(x = employee.dB.plot[,1], y = employee.dB.plot[,2], color = em
ployee.dB.plot[,3]))+
    theme_bw()+
    geom_point(position = "jitter")+
    labs(x = x_name, y = "Attrition", color = "Attrition")
}
```

STEP 1B - EDA - CREATE VISUALS

```
if (make.eda.plots == "Yes") {

  if (column.histograms == "Yes") {
  #
  ###################
  ### CONTINIOUS VARIABLES ###
  ###################
  #
  ## AGE ##
  #continious.histogram.plot.single.variable("Age", "Attrition")
  #continious.histogram.plot("Age", "Attrition")

  ## BUSINESS TRAVEL ##
  continious.histogram.plot.single.variable("DailyRate", "Attrition")
  continious.histogram.plot("DailyRate", "Attrition")

  ## DISTANCE FROM HOME ##
  continious.histogram.plot.single.variable("DistanceFromHome", "Attrition")
  continious.histogram.plot("DistanceFromHome", "Attrition")

  ## EDUCATION ##
  continious.histogram.plot.single.variable("Education", "Attrition")
  continious.histogram.plot("Education", "Attrition")

  ## HOURLY RATE ##
  continious.histogram.plot.single.variable("HourlyRate", "Attrition")
  continious.histogram.plot("HourlyRate", "Attrition")

  ## MONTHLY INCOME ##
  continious.histogram.plot.single.variable("MonthlyIncome", "Attrition")
  continious.histogram.plot("MonthlyIncome", "Attrition")

  ## MONTHLY RATE ##
  continious.histogram.plot.single.variable("MonthlyRate", "Attrition")
  continious.histogram.plot("MonthlyRate", "Attrition")

  ## PERCENTAGE SALARY HIKE ##
  continious.histogram.plot.single.variable("PercentSalaryHike", "Attrition")
  continious.histogram.plot("PercentSalaryHike", "Attrition")

  ## TOTAL WORKING YEARS ##
  continious.histogram.plot.single.variable("TotalWorkingYears", "Attrition")
  continious.histogram.plot("TotalWorkingYears", "Attrition")

  ## YEARS AT COMPANY ##
  continious.histogram.plot.single.variable("YearsAtCompany", "Attrition")
  continious.histogram.plot("YearsAtCompany", "Attrition")

  ## YEARS IN CURRENT ROLE ##
  continious.histogram.plot.single.variable("YearsInCurrentRole", "Attrition")
  continious.histogram.plot("YearsInCurrentRole", "Attrition")
```

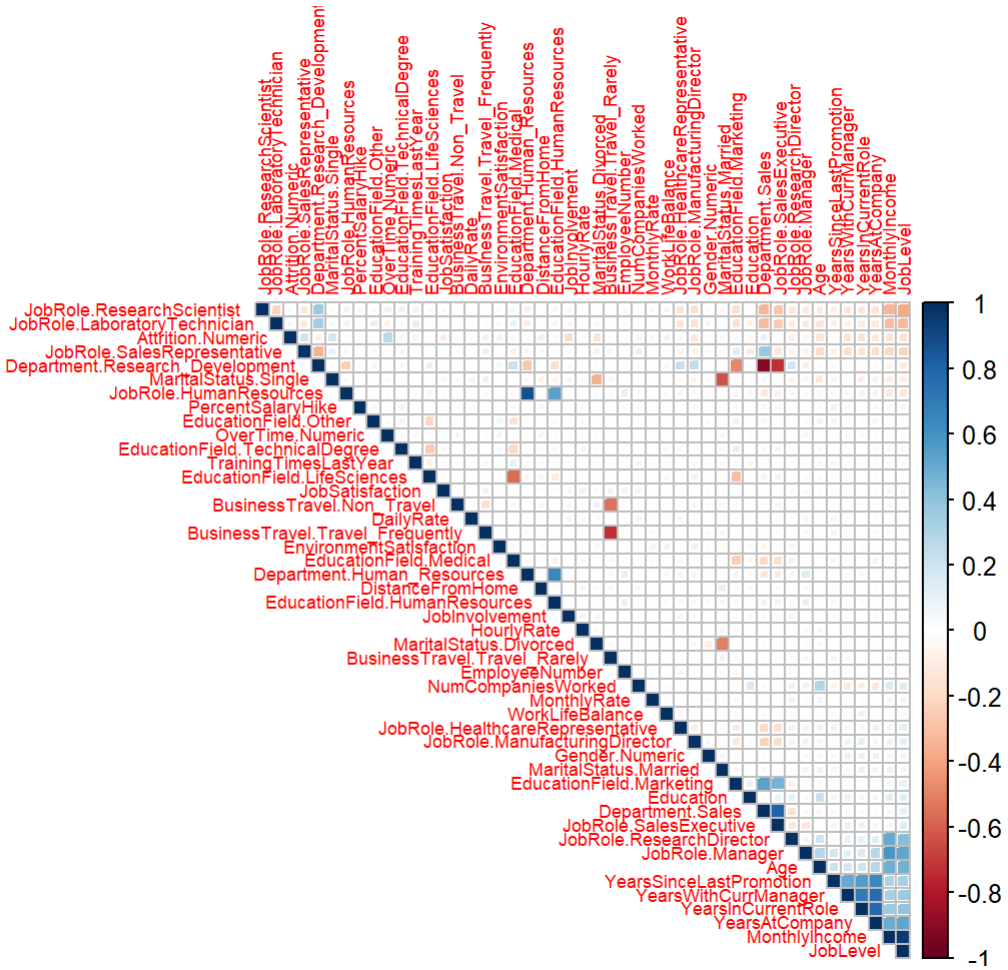
```
####################
### CATEGORICAL VARIABLES ###
####################
#
## RAW ATTRITION ##
#dev.new()
employee.dB %>% ggplot(aes(x = Attrition, fill = Attrition))+
  theme_bw()+
  geom_bar(stat = "count", alpha = 0.35, color = "grey80", show.legend = FALSE)+
  labs(x = "Attrition", y = "Count")

## BUSINESS TRAVEL ##
categorical.barchart.plot.single.variable("BusinessTravel", "Attrition")
categorical.barchart.plot("BusinessTravel", "Attrition")

## DEPARTMENT ##
categorical.barchart.plot.single.variable("Department", "Attrition")
categorical.barchart.plot("Department", "Attrition")

#dev.new()
employee.dB %>% ggplot(aes(x = employee.dB$Department))+
  theme_bw()+
  geom_bar(stat = "count", alpha = 0.35, color = "grey80")+
  labs(x = "Department")


## ENVIRONMENTAL SATISFACTION ##
categorical.barchart.plot.single.variable("EnvironmentSatisfaction", "Attrition")
categorical.barchart.plot("EnvironmentSatisfaction", "Attrition")

## EDUCATION ##
categorical.barchart.plot.single.variable("Education", "Attrition")
categorical.barchart.plot("Education", "Attrition")

## EDUCATION FIELD ##
categorical.barchart.plot.single.variable("EducationField", "Attrition")
categorical.barchart.plot("EducationField", "Attrition")

## ENVIRONMENTAL SATISFACTION  ##
categorical.barchart.plot.single.variable("EnvironmentSatisfaction", "Attrition")
categorical.barchart.plot("EnvironmentSatisfaction", "Attrition")

## GENDER ##
categorical.barchart.plot.single.variable("Gender", "Attrition")
categorical.barchart.plot("Gender", "Attrition")

## JOB INVOLVEMENT ##
categorical.barchart.plot.single.variable("JobInvolvement", "Attrition")
categorical.barchart.plot("JobInvolvement", "Attrition")

## JOB LEVEL ##
categorical.barchart.plot.single.variable("JobLevel", "Attrition")
categorical.barchart.plot("JobLevel", "Attrition")
```

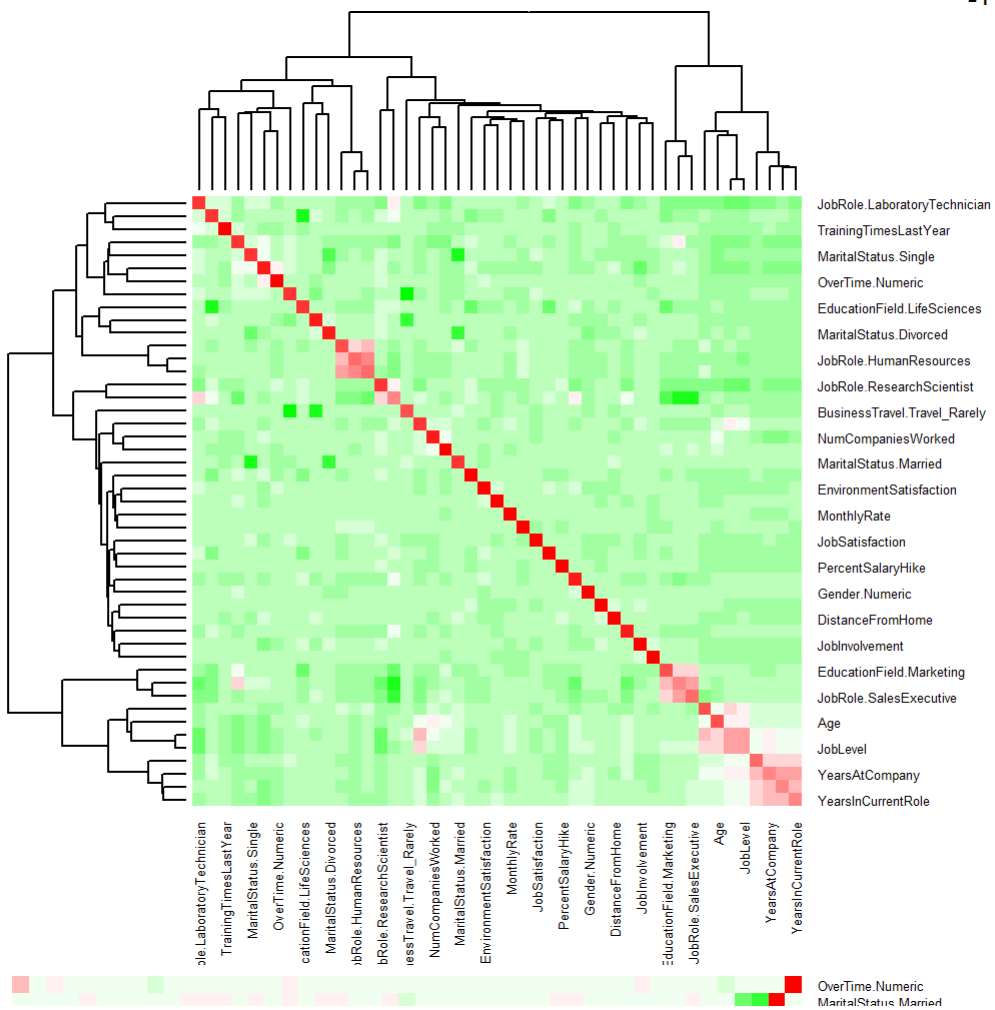
```
## JOB ROLE ##
categorical.barchart.plot.single.variable("JobRole", "Attrition")
categorical.barchart.plot("JobRole", "Attrition")

## JOB SATISFACTION ##
categorical.barchart.plot.single.variable("JobSatisfaction", "Attrition")
categorical.barchart.plot("JobSatisfaction", "Attrition")

## MARITAL STATUS ##
categorical.barchart.plot.single.variable("MaritalStatus", "Attrition")
categorical.barchart.plot("MaritalStatus", "Attrition")

## NUMBER OF COMPANIES WORKED ##
categorical.barchart.plot.single.variable("NumCompaniesWorked", "Attrition")
categorical.barchart.plot("NumCompaniesWorked", "Attrition")

## OVERTIME ##
categorical.barchart.plot.single.variable("OverTime", "Attrition")
categorical.barchart.plot("OverTime", "Attrition")

## PERFORMANCE RATING ##
categorical.barchart.plot.single.variable("PerformanceRating", "Attrition")
categorical.barchart.plot("PerformanceRating", "Attrition")

## RELATIONSHIP SATISFACTION ##
categorical.barchart.plot.single.variable("RelationshipSatisfaction", "Attrition")
categorical.barchart.plot("RelationshipSatisfaction", "Attrition")

## STOCK OPTION LEVEL ##
categorical.barchart.plot.single.variable("StockOptionLevel", "Attrition")
categorical.barchart.plot("StockOptionLevel", "Attrition")

## TRAINING TIMES LAST YEAR ##
categorical.barchart.plot.single.variable("TrainingTimesLastYear", "Attrition")
categorical.barchart.plot("TrainingTimesLastYear", "Attrition")

## WORKLIFE BALANCE ##
categorical.barchart.plot.single.variable("WorkLifeBalance", "Attrition")
categorical.barchart.plot("WorkLifeBalance", "Attrition")

## YEARS AT COMPANY ##
categorical.barchart.plot.single.variable("YearsAtCompany", "Attrition")
categorical.barchart.plot("YearsAtCompany", "Attrition")

## YEARS IN CURRENT ROLE  ##
categorical.barchart.plot.single.variable("YearsInCurrentRole", "Attrition")
categorical.barchart.plot("YearsInCurrentRole", "Attrition")

## YEARS SINCE LAST PROMOTION  ##
categorical.barchart.plot.single.variable("YearsSinceLastPromotion", "Attrition")
categorical.barchart.plot("YearsSinceLastPromotion", "Attrition")

## YEARS WITH CURRENT MANAGER  ##
```

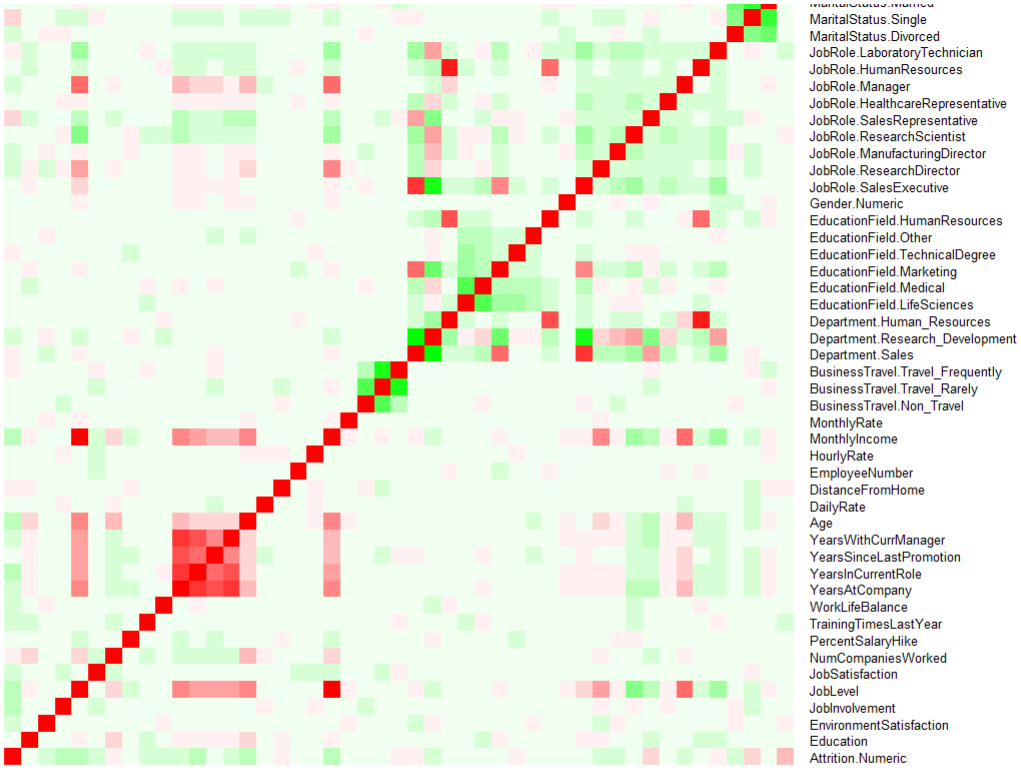
```
  categorical.barchart.plot.single.variable("YearsWithCurrManager", "Attrition")
  categorical.barchart.plot("YearsWithCurrManager", "Attrition")

  #
  ####################
  ### TEMPORAL ###
  ####################
  #
  ## EMPLOYEE NUMBER ##
  continious.scatter.plot("EmployeeNumber", "Attrition.Numeric", "Attrition")
  }

  #
  ####################
  ### DEMOGRAPHICS ###
  ####################
  #
  if (demographics.plots == "Yes") {

    ## RETENTION DELTA FOR EACH DEMOGRAPHIC (BARCHART)
    #dev.new()
    categories.dB %>% arrange(desc(categories.dB$demographic.density.contrast.delta)) %>%
      ggplot(aes(x = reorder(demographic,-demographic.density.contrast.delta), y = demographic.d
ensity.contrast.delta, fill = Category ))+
      theme_bw()+
      theme(legend.key.width = unit(0.1,"cm"))+
      theme(legend.key.height = unit(0.2,"cm"))+
      theme(axis.text.x = element_text(size = 0.5))+
      theme(axis.text.x = element_text(angle = 45))+
      theme(legend.title = element_text(size = 1))+
      theme(legend.text = element_text(size = 5))+
      geom_bar(stat = "identity")+
      guides(fill = guide_legend(ncol = 1))+
      labs(x = "Demographic", y = "Population Denisty Delta")

    ## RETENTION DELTA FOR EACH DEMOGRAPHIC (BARCHART) - TOP 10
    #dev.new()
    categories.dB %>% arrange(desc(categories.dB$demographic.density.contrast.delta)) %>% head(1
0) %>%
      ggplot(aes(x = reorder(demographic,-demographic.density.contrast.delta), y = demographic.d
ensity.contrast.delta, fill = Category ))+
      theme_bw()+
      theme(legend.key.width = unit(0.1,"cm"))+
      theme(legend.key.height = unit(0.2,"cm"))+
      theme(axis.text.x = element_text(size = 10))+
      #theme(axis.text.x = element_text(angle = 45))+
      theme(legend.title = element_text(size = 1))+
      theme(legend.text = element_text(size = 1))+
      geom_bar(stat = "identity")+
      guides(fill = guide_legend(ncol = 1))+
      labs(x = "Demographic", y = "Demographic Denisty Contrast")

    ## RETENTION DELTA FOR EACH DEMOGRAPHIC (BARCHART) - BOTTOM 10
    #dev.new()
```

```
    categories.dB %>% arrange(desc(categories.dB$demographic.density.contrast.delta)) %>% tail(1
0) %>%
      ggplot(aes(x = reorder(demographic,-demographic.density.contrast.delta), y = demographic.d
ensity.contrast.delta, fill = Category ))+
      theme_bw()+
      theme(legend.key.width = unit(0.1,"cm"))+
      theme(legend.key.height = unit(0.2,"cm"))+
      theme(axis.text.x = element_text(size = 10))+
      #theme(axis.text.x = element_text(angle = 45))+
      theme(legend.title = element_text(size = 1))+
      theme(legend.text = element_text(size = 1))+
      geom_bar(stat = "identity")+
      guides(fill = guide_legend(ncol = 1))+
      labs(x = "Demographic", y = "Demographic Denisty Contrast")


    ##############################################################################################

    ## DELTA PERCENTAGE FOR EACH DEMOGRAPHIC (BARCHART)
    #dev.new()
    categories.dB %>% arrange(desc(categories.dB$demographic.density.contrast.perc)) %>%
      ggplot(aes(x = reorder(demographic,-demographic.density.contrast.perc), y = demographic.de
nsity.contrast.perc, fill = Category ))+
      theme_bw()+
      theme(legend.key.width = unit(0.1,"cm"))+
      theme(legend.key.height = unit(0.2,"cm"))+
      theme(axis.text.x = element_text(size = 0.5))+
      theme(axis.text.x = element_text(angle = 45))+
      theme(legend.title = element_text(size = 1))+
      theme(legend.text = element_text(size = 5))+
      geom_bar(stat = "identity")+
      guides(fill = guide_legend(ncol = 1))+
      labs(x = "Demographic", y = "Population Denisty Delta Percentage (%)")

  } #DEMOGRAPHICS CLOSING BRACKETS

  #
  ####################
  ### CORRELATION MATRIX ###
  ####################
  #
  if (correlation.matrix.plots == "Yes") {
    #dev.new()
    corrplot(correlation.mtx.employee.dB.coeff, method = "square", type = "upper", order = "FPC"
, tl.cex = 0.55)

    #dev.new()
    palette = colorRampPalette(c("green", "white", "red")) (20)
    heatmap(x = correlation.mtx.employee.dB.coeff, col = palette, symm = TRUE, cexRow = 0.5, cex
Col = 0.5, scale = "column")

    #dev.new()
    palette = colorRampPalette(c("green", "white", "red")) (20)
    heatmap(x = correlation.mtx.employee.dB.coeff, col = palette, symm = TRUE, Colv = NA, Rowv =
```

```
NA, cexRow = 0.5, cexCol = 0.5)

    #dev.new()
    correlation.mtx.plot <- ggcorr(correlation.mtx.employee.dB.coeff, size = 1.5, label = TRUE,
 label_size = 2, label_color = "black", nbreaks = 8, label_alpha = TRUE, color = "grey5", layou
t.exp = 1)
    multiplot(correlation.mtx.plot)

    }
#
} # ALL EDA CLOSING BRACKET
```


JobRole.ResearchScientist
JobRole.LaboratoryTechnician
Attrition.Numeric
JobRole.SalesRepresentative
Department.Research_Development
MaritalStatus.Single
JobRole.HumanResources
PercentSalaryHike
EducationField.Other
OverTime.Numeric
EducationField.TechnicalDegree
TrainingTimesLastYear
EducationField.LifeSciences
JobSatisfaction
BusinessTravel.Non_Travel
DailyRate
BusinessTravel.Travel_Frequently
EnvironmentSatisfaction
EducationField.Medical
Department.Human_Resources
DistanceFromHome
EducationField.HumanResources
JobInvolvement
HourlyRate
MaritalStatus.Divorced
BusinessTravel.Travel_Rarely
EmployeeNumber
NumCompaniesWorked
MonthlyRate
WorkLifeBalance
JobRole.HealthcareRepresentative
JobRole.ManufacturingDirector
Gender.Numeric
MaritalStatus.Married
EducationField.Marketing
Education
Department.Sales
JobRole.SalesExecutive
JobRole.ResearchDirector
JobRole.Manager
Age
YearsSinceLastPromotion
YearsWithCurrManager
YearsInCurrentRole
YearsAtCompany
MonthlyIncome
JobLevel
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


JobRole.LaboratoryTechnician
TrainingTimesLastYear
MaritalStatus.Single
OverTime.Numeric
EducationField.LifeSciences
MaritalStatus.Divorced
JobRole.HumanResources
JobRole.ResearchScientist
BusinessTravel.Travel_Rarely
NumCompaniesWorked
MaritalStatus.Married
EnvironmentSatisfaction
MonthlyRate
JobSatisfaction
PercentSalaryHike
Gender.Numeric
DistanceFromHome
JobInvolvement
EducationField.Marketing
JobRole.SalesExecutive
Age
JobLevel
YearsAtCompany
YearsInCurrentRole
OverTime.Numeric
MaritalStatus.Married


MaritalStatus.Single
MaritalStatus.Divorced
JobRole.LaboratoryTechnician
JobRole.HumanResources
JobRole.Manager
JobRole.HealthcareRepresentative
JobRole.SalesRepresentative
JobRole.ResearchScientist
JobRole.ManufacturingDirector
JobRole.ResearchDirector
JobRole.SalesExecutive
Gender.Numeric
EducationField.HumanResources
EducationField.Other
EducationField.TechnicalDegree
EducationField.Marketing
EducationField.Medical
EducationField.LifeSciences
Department.Human_Resources
Department.Research_Development
Department.Sales
BusinessTravel.Travel_Frequently
BusinessTravel.Travel_Rarely
BusinessTravel.Non_Travel
MonthlyRate
MonthlyIncome
HourlyRate
EmployeeNumber
DistanceFromHome
DailyRate
Age
YearsWithCurrManager
YearsSinceLastPromotion
YearsInCurrentRole
YearsAtCompany
WorkLifeBalance
TrainingTimesLastYear
PercentSalaryHike
NumCompaniesWorked
JobSatisfaction
JobLevel
JobInvolvement
EnvironmentSatisfaction
Education
Attrition.Numeric
Attrition.Numeric
Education
EnvironmentSatisfaction
JobInvolvement
JobLevel
JobSatisfaction
NumCompaniesWorked
PercentSalaryHike
TrainingTimesLastYear
WorkLifeBalance
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrManager
Age
DailyRate
DistanceFromHome
EmployeeNumber
HourlyRate
MonthlyIncome
MonthlyRate
BusinessTravel.Non_Travel
BusinessTravel.Travel_Rarely
BusinessTravel.Travel_Frequently
Department.Sales
Department.Research_Development
Department.Human_Resources
EducationField.LifeSciences
EducationField.Medical
EducationField.Marketing
EducationField.TechnicalDegree
EducationField.Other
EducationField.HumanResources
Gender.Numeric
JobRole.SalesExecutive
JobRole.ResearchDirector
JobRole.ManufacturingDirector
JobRole.ResearchScientist
JobRole.SalesRepresentative
JobRole.HealthcareRepresentative
JobRole.Manager
JobRole.HumanResources
JobRole.LaboratoryTechnician
MaritalStatus.Divorced
MaritalStatus.Single
MaritalStatus.Married
OverTime.Numeric
[-1,-0.75]
(-0.75,-0.5]
(-0.5,-0.25]
(-0.25,0]
(0,0.25]
(0.25,0.5]
(0.5,0.75]
(0.75,1]

STEP 2 - CLASSIFICATION ANALYSIS

```
alt.employee.dB <- read.csv(path.employee.dB)

attach(alt.employee.dB)

# Explore the raw data set
str(alt.employee.dB)
```

```
## 'data.frame':    870 obs. of  36 variables:
##  $ ID                      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Age                     : int  32 40 35 32 24 27 41 37 34 34 ...
##  $ Attrition               : chr  "No" "No" "No" "No" ...
##  $ BusinessTravel          : chr  "Travel_Rarely" "Travel_Rarely" "Travel_Frequently" "Travel
_Rarely" ...
##  $ DailyRate               : int  117 1308 200 801 567 294 1283 309 1333 653 ...
##  $ Department              : chr  "Sales" "Research & Development" "Research & Development"
"Sales" ...
##  $ DistanceFromHome        : int  13 14 18 1 2 10 5 10 10 10 ...
##  $ Education               : int  4 3 2 4 1 2 5 4 4 4 ...
##  $ EducationField          : chr  "Life Sciences" "Medical" "Life Sciences" "Marketing" ...
##  $ EmployeeCount           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ EmployeeNumber          : int  859 1128 1412 2016 1646 733 1448 1105 1055 1597 ...
##  $ EnvironmentSatisfaction : int  2 3 3 3 1 4 2 4 3 4 ...
##  $ Gender                  : chr  "Male" "Male" "Male" "Female" ...
##  $ HourlyRate              : int  73 44 60 48 32 32 90 88 87 92 ...
##  $ JobInvolvement          : int  3 2 3 3 3 3 4 2 3 2 ...
##  $ JobLevel                : int  2 5 3 3 1 3 1 2 1 2 ...
##  $ JobRole                 : chr  "Sales Executive" "Research Director" "Manufacturing Direct
or" "Sales Executive" ...
##  $ JobSatisfaction         : int  4 3 4 4 4 1 3 4 3 3 ...
##  $ MaritalStatus           : chr  "Divorced" "Single" "Single" "Married" ...
##  $ MonthlyIncome           : int  4403 19626 9362 10422 3760 8793 2127 6694 2220 5063 ...
##  $ MonthlyRate             : int  9250 17544 19944 24032 17218 4809 5561 24223 18410 15332
...
##  $ NumCompaniesWorked      : int  2 1 2 1 1 1 2 2 1 1 ...
##  $ Over18                  : chr  "Y" "Y" "Y" "Y" ...
##  $ OverTime                : chr  "No" "No" "No" "No" ...
##  $ PercentSalaryHike       : int  11 14 11 19 13 21 12 14 19 14 ...
##  $ PerformanceRating       : int  3 3 3 3 3 4 3 3 3 3 ...
##  $ RelationshipSatisfaction: int  3 1 3 3 3 3 1 3 4 2 ...
##  $ StandardHours           : int  80 80 80 80 80 80 80 80 80 80 ...
##  $ StockOptionLevel        : int  1 0 0 2 0 2 0 3 1 1 ...
##  $ TotalWorkingYears       : int  8 21 10 14 6 9 7 8 1 8 ...
##  $ TrainingTimesLastYear   : int  3 2 2 3 2 4 5 5 2 3 ...
##  $ WorkLifeBalance         : int  2 4 3 3 3 2 2 3 3 2 ...
##  $ YearsAtCompany          : int  5 20 2 14 6 9 4 1 1 8 ...
##  $ YearsInCurrentRole      : int  2 7 2 10 3 7 2 0 1 2 ...
##  $ YearsSinceLastPromotion : int  0 4 2 5 1 1 0 0 0 7 ...
##  $ YearsWithCurrManager    : int  3 9 2 7 3 7 3 0 0 7 ...
```

```
#Convert characters to factors
alt.employee.dB$Attrition <- as.factor(alt.employee.dB$Attrition)
alt.employee.dB$BusinessTravel <- as.factor(alt.employee.dB$BusinessTravel)
alt.employee.dB$Department <- as.factor(alt.employee.dB$Department)
alt.employee.dB$EducationField <- as.factor(alt.employee.dB$EducationField)
alt.employee.dB$Gender <- as.factor(alt.employee.dB$Gender)
alt.employee.dB$JobRole <- as.factor(alt.employee.dB$JobRole)
alt.employee.dB$MaritalStatus <- as.factor(alt.employee.dB$MaritalStatus)
alt.employee.dB$Over18 <- as.factor(alt.employee.dB$Over18)
alt.employee.dB$OverTime <- as.factor(alt.employee.dB$OverTime)

# Coerce the integer (more or less categorical) variables in to factors
categorical.variables <- c('RelationshipSatisfaction', 'PerformanceRating', 'WorkLifeBalance',
'JobInvolvement', 'JobSatisfaction', 'JobLevel', 'StockOptionLevel')
alt.employee.dB[,categorical.variables] <- lapply(alt.employee.dB[,categorical.variables] , fact
or, ordered = TRUE)
str(alt.employee.dB)
```

```
## 'data.frame':    870 obs. of  36 variables:
##  $ ID                      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Age                     : int  32 40 35 32 24 27 41 37 34 34 ...
##  $ Attrition               : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ BusinessTravel          : Factor w/ 3 levels "Non-Travel","Travel_Frequently",..: 3 3 2 3
2 2 3 3 3 2 ...
##  $ DailyRate               : int  117 1308 200 801 567 294 1283 309 1333 653 ...
##  $ Department              : Factor w/ 3 levels "Human Resources",..: 3 2 2 3 2 2 2 3 3 2 ...
##  $ DistanceFromHome        : int  13 14 18 1 2 10 5 10 10 10 ...
##  $ Education               : int  4 3 2 4 1 2 5 4 4 4 ...
##  $ EducationField          : Factor w/ 6 levels "Human Resources",..: 2 4 2 3 6 2 4 2 2 6 ...
##  $ EmployeeCount           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ EmployeeNumber          : int  859 1128 1412 2016 1646 733 1448 1105 1055 1597 ...
##  $ EnvironmentSatisfaction : int  2 3 3 3 1 4 2 4 3 4 ...
##  $ Gender                  : Factor w/ 2 levels "Female","Male": 2 2 2 1 1 2 2 1 1 2 ...
##  $ HourlyRate              : int  73 44 60 48 32 32 90 88 87 92 ...
##  $ JobInvolvement          : Ord.factor w/ 4 levels "1"<"2"<"3"<"4": 3 2 3 3 3 3 4 2 3 2 ...
##  $ JobLevel                : Ord.factor w/ 5 levels "1"<"2"<"3"<"4"<..: 2 5 3 3 1 3 1 2 1 2
...
##  $ JobRole                 : Factor w/ 9 levels "Healthcare Representative",..: 8 6 5 8 7 5 7
8 9 1 ...
##  $ JobSatisfaction         : Ord.factor w/ 4 levels "1"<"2"<"3"<"4": 4 3 4 4 4 1 3 4 3 3 ...
##  $ MaritalStatus           : Factor w/ 3 levels "Divorced","Married",..: 1 3 3 2 3 1 2 1 2 2
...
##  $ MonthlyIncome           : int  4403 19626 9362 10422 3760 8793 2127 6694 2220 5063 ...
##  $ MonthlyRate             : int  9250 17544 19944 24032 17218 4809 5561 24223 18410 15332
...
##  $ NumCompaniesWorked      : int  2 1 2 1 1 1 2 2 1 1 ...
##  $ Over18                  : Factor w/ 1 level "Y": 1 1 1 1 1 1 1 1 1 1 ...
##  $ OverTime                : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 2 2 2 1 ...
##  $ PercentSalaryHike       : int  11 14 11 19 13 21 12 14 19 14 ...
##  $ PerformanceRating       : Ord.factor w/ 2 levels "3"<"4": 1 1 1 1 1 2 1 1 1 1 ...
##  $ RelationshipSatisfaction: Ord.factor w/ 4 levels "1"<"2"<"3"<"4": 3 1 3 3 3 3 1 3 4 2 ...
##  $ StandardHours           : int  80 80 80 80 80 80 80 80 80 80 ...
##  $ StockOptionLevel        : Ord.factor w/ 4 levels "0"<"1"<"2"<"3": 2 1 1 3 1 3 1 4 2 2 ...
##  $ TotalWorkingYears       : int  8 21 10 14 6 9 7 8 1 8 ...
##  $ TrainingTimesLastYear   : int  3 2 2 3 2 4 5 5 2 3 ...
##  $ WorkLifeBalance         : Ord.factor w/ 4 levels "1"<"2"<"3"<"4": 2 4 3 3 3 2 2 3 3 2 ...
##  $ YearsAtCompany          : int  5 20 2 14 6 9 4 1 1 8 ...
##  $ YearsInCurrentRole      : int  2 7 2 10 3 7 2 0 1 2 ...
##  $ YearsSinceLastPromotion : int  0 4 2 5 1 1 0 0 0 7 ...
##  $ YearsWithCurrManager    : int  3 9 2 7 3 7 3 0 0 7 ...
```

```
# Calculate the Time an Employee spend at their last job
alt.employee.dB$TimeAtLastJob <- ifelse(alt.employee.dB$NumCompaniesWorked!=0, alt.employee.dB$T
otalWorkingYears-alt.employee.dB$YearsAtCompany/alt.employee.dB$NumCompaniesWorked,0)
alt.employee.dB$AgeGroup <- as.factor(
                        ifelse(alt.employee.dB$Age<=30,"Early", ifelse(
                          alt.employee.dB$Age<=43,"Mid","Late"
                        ))
                      )
#convert MonthlyIncome
alt.employee.dB$MonthlyIncomeGroup <- as.factor(
  ifelse(alt.employee.dB$MonthlyIncome <= 2911,"1st.Quartile", ifelse(
    alt.employee.dB$MonthlyIncome <= 6503,"2nd.Quartile", ifelse(
      alt.employee.dB$MonthlyIncome <= 8379,"3rd.Quartile","4th.Quartile"
    )))
)
#convert  YearsAtCompany
alt.employee.dB$YrsAtCompanyGroup <- as.factor(
  ifelse(alt.employee.dB$YearsAtCompany <= 3,"1st.Quartile", ifelse(
    alt.employee.dB$YearsAtCompany <= 7,"2nd.Quartile", ifelse(
      alt.employee.dB$YearsAtCompany <= 10,"3rd.Quartile","4th.Quartile"
    )))
)
#convert YearsWithCurrManager
alt.employee.dB$YrsWtCurrManagerGroup <- as.factor(
  ifelse(alt.employee.dB$YearsWithCurrManager <= 2,"1st.Quartile", ifelse(
    alt.employee.dB$YearsWithCurrManager <= 4,"2nd.Quartile", ifelse(
      alt.employee.dB$YearsWithCurrManager <= 7,"3rd.Quartile","4th.Quartile"
    )))
)
#convert YearsInCurrentRole
alt.employee.dB$YrsInCurrentRoleGroup <- as.factor(
  ifelse(alt.employee.dB$YearsInCurrentRole <= 2,"LessTh2", ifelse(
    alt.employee.dB$YearsInCurrentRole <= 4,"2nd.Quartile", ifelse(
      alt.employee.dB$YearsInCurrentRole <= 7,"3rd.Quartile","4th.Quartile"
    )))
)

#create a new data frame with factors
alt.employee.dB_cat = alt.employee.dB[, c(2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19
, 20, 21, 22, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36,  37, 38)]

#check for missing values
colSums(is.na(alt.employee.dB_cat))
```

```
##                     Age               Attrition          BusinessTravel
##                       0                       0                       0
##               DailyRate              Department        DistanceFromHome
##                       0                       0                       0
##               Education           EducationField  EnvironmentSatisfaction
##                       0                       0                       0
##                  Gender              HourlyRate          JobInvolvement
##                       0                       0                       0
##                JobLevel                 JobRole         JobSatisfaction
##                       0                       0                       0
##           MaritalStatus           MonthlyIncome             MonthlyRate
##                       0                       0                       0
##       NumCompaniesWorked                OverTime       PercentSalaryHike
##                       0                       0                       0
##        PerformanceRating RelationshipSatisfaction        StockOptionLevel
##                       0                       0                       0
##        TotalWorkingYears    TrainingTimesLastYear         WorkLifeBalance
##                       0                       0                       0
##           YearsAtCompany       YearsInCurrentRole  YearsSinceLastPromotion
##                       0                       0                       0
##     YearsWithCurrManager            TimeAtLastJob                AgeGroup
##                       0                       0                       0
```

```
#Split the Data
set.seed(1234)
alt.employee.dB_cat_split <- sample.split(alt.employee.dB_cat$Attrition, SplitRatio = 0.70)
alt.employee.dB_cat_train <- subset(alt.employee.dB_cat,alt.employee.dB_cat_split == T)
alt.employee.dB_cat_test <- subset(alt.employee.dB_cat,alt.employee.dB_cat_split == F)

# compare the dimention of splitted dataset
dim(alt.employee.dB_cat)
```

```
## [1] 870  33
```

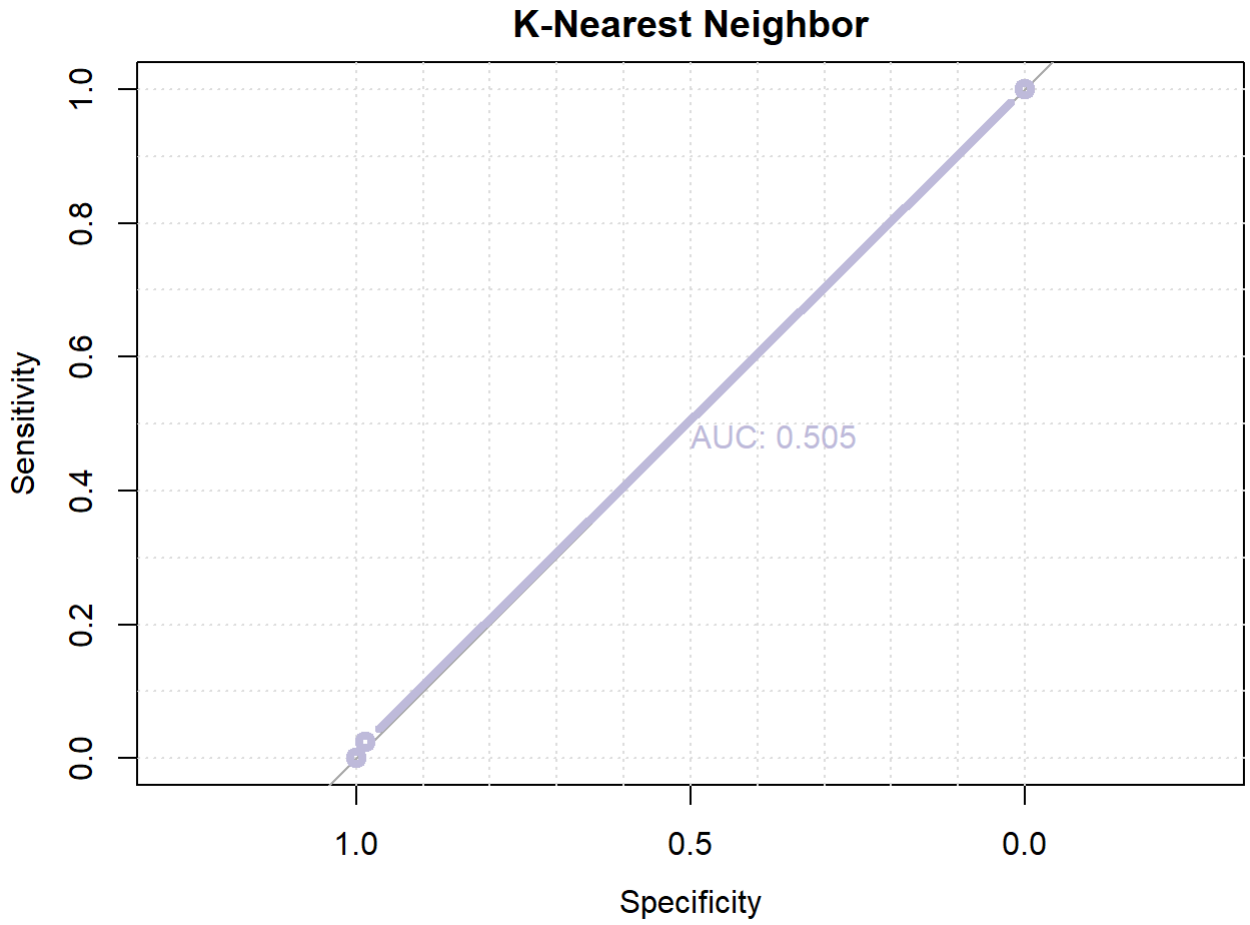
```
dim(alt.employee.dB_cat_test)
```

```
## [1] 261  33
```

```
dim(alt.employee.dB_cat_train)
```

```
## [1] 609  33
```

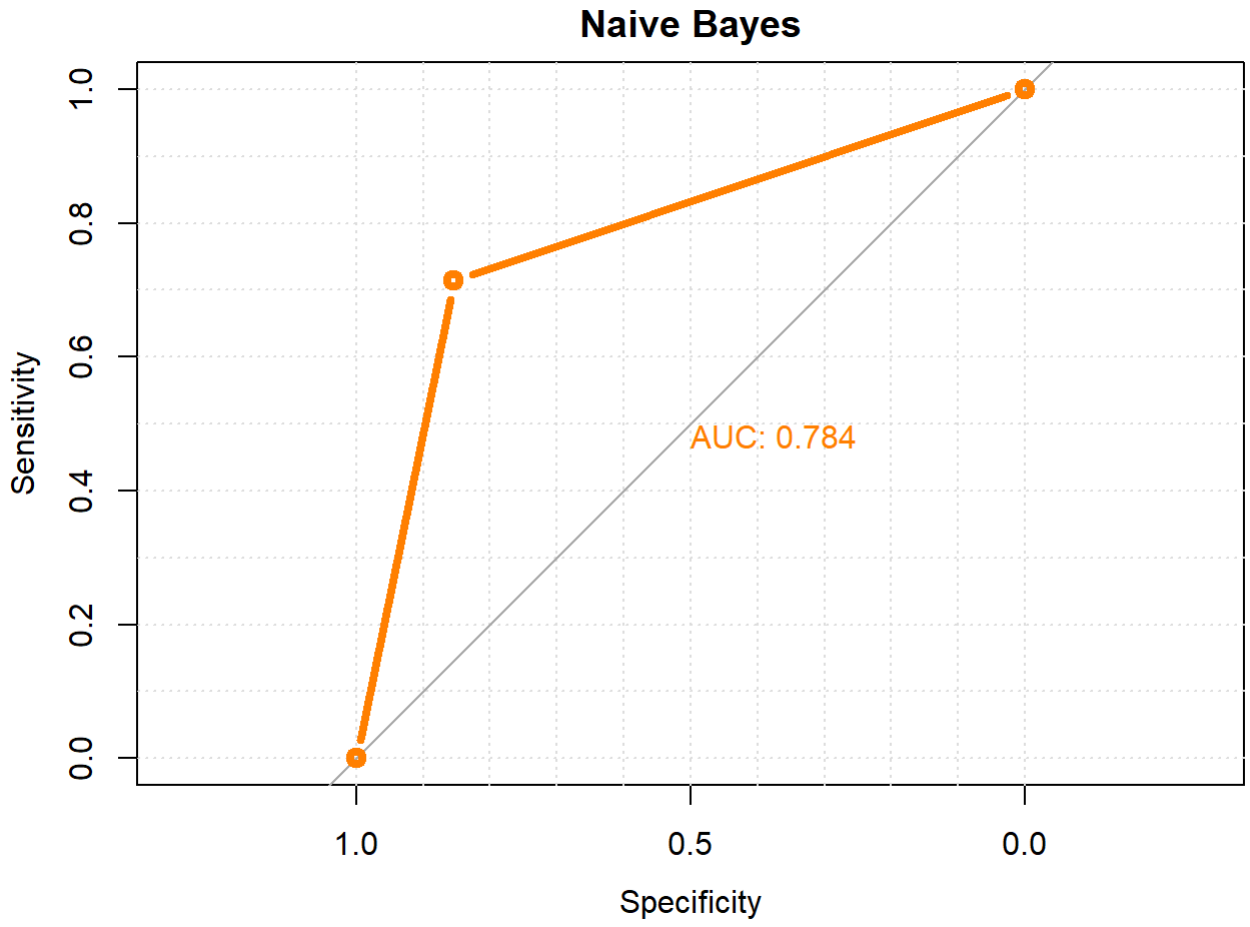
```
################ kNN Classifier ################

#Fit the model on train data
KnnModel <- train(Attrition~.,alt.employee.dB_cat_train,method = 'knn',trControl = trainControl
(method = 'repeatedcv',number = 3))
#Predict with test set
Knn_pred <- predict(KnnModel,alt.employee.dB_cat_test)
#Print confusion matrix
Knn_CM <- confusionMatrix(alt.employee.dB_cat_test$Attrition, Knn_pred)
#Plot the curve
plot.roc(as.numeric(alt.employee.dB_cat_test$Attrition), as.numeric(Knn_pred),lwd=4, type="b",gr
id.lty=3, grid=TRUE, print.auc=TRUE,print.auc.col= "#BEBADA", col ="#BEBADA", main = "K-Nearest
 Neighbor")
```




```
################ Naive Bayes Classifier ################

# Fit the model on train data
NBModel <- naiveBayes(Attrition~., data=alt.employee.dB_cat_train)

# Validate on test set
NBM_pred <- predict(NBModel, newdata=alt.employee.dB_cat_test)

# Print confusion matrix
NBM_CM <- confusionMatrix(NBM_pred, alt.employee.dB_cat_test$Attrition)
NBM_CM
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No Yes
##        No  187  12
##        Yes  32  30
##
##                Accuracy : 0.8314
##                  95% CI : (0.7804, 0.8748)
##     No Information Rate : 0.8391
##     P-Value [Acc > NIR] : 0.668819
##
##                   Kappa : 0.4765
##
##  Mcnemar's Test P-Value : 0.004179
##
##             Sensitivity : 0.8539
##             Specificity : 0.7143
##          Pos Pred Value : 0.9397
##          Neg Pred Value : 0.4839
##              Prevalence : 0.8391
##          Detection Rate : 0.7165
##    Detection Prevalence : 0.7625
##       Balanced Accuracy : 0.7841
##
##        'Positive' Class : No
##
```

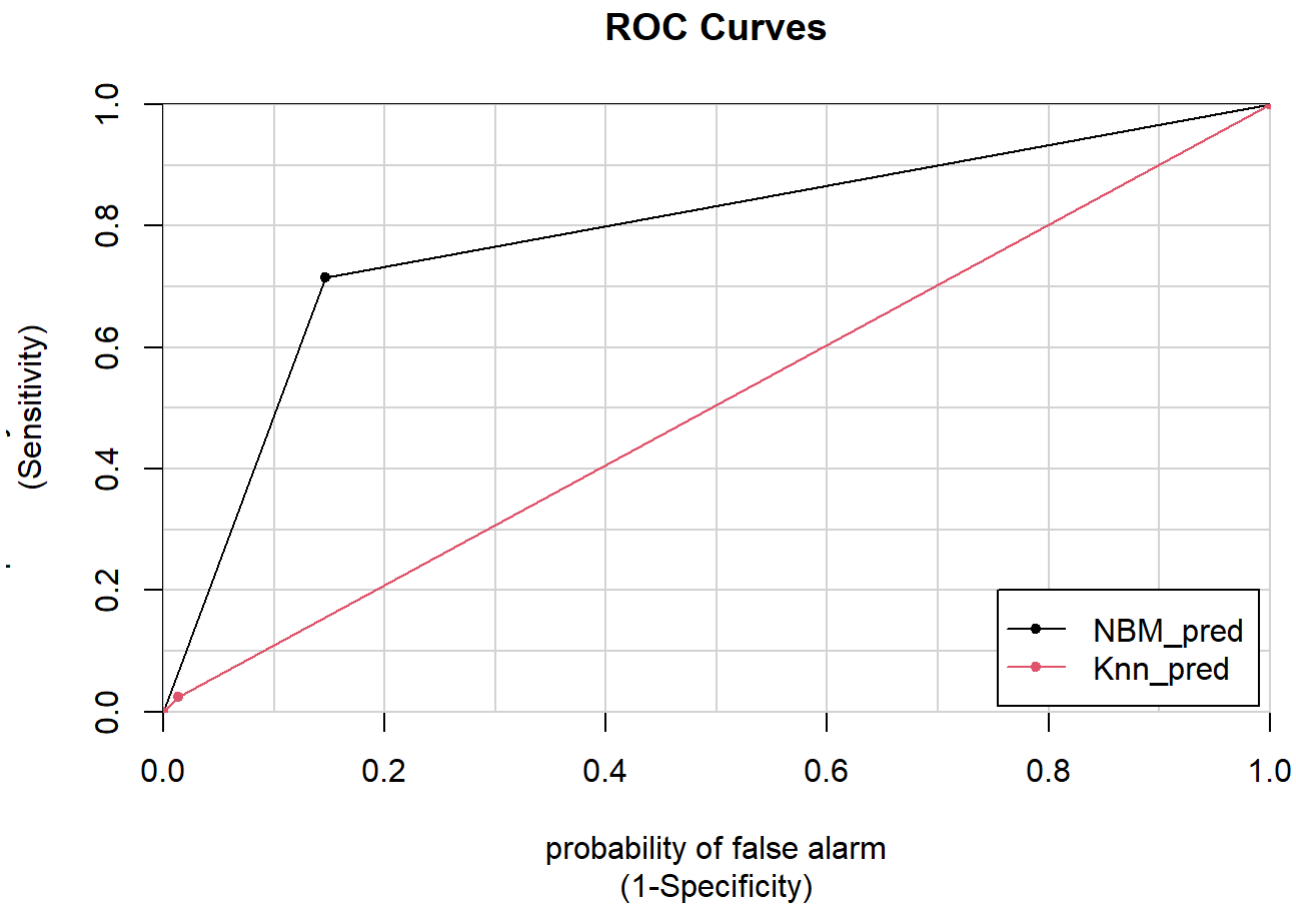
```
#Plot
NBM_ROC <- plot.roc(as.numeric(alt.employee.dB_cat_test$Attrition), as.numeric(NBM_pred),lwd=4,
 type="b",grid.lty=3, grid=TRUE, print.auc=TRUE,print.auc.col= "#FF7F00", col ="#FF7F00", main =
"Naive Bayes")
```




```
NBM_ROC
```

```
##
## Call:
## plot.roc.default(x = as.numeric(alt.employee.dB_cat_test$Attrition),     predictor = as.numer
ic(NBM_pred), lwd = 4, type = "b", grid.lty = 3,     grid = TRUE, print.auc = TRUE, print.auc.co
l = "#FF7F00",     col = "#FF7F00", main = "Naive Bayes")
##
## Data: as.numeric(NBM_pred) in 219 controls (as.numeric(alt.employee.dB_cat_test$Attrition) 1)
< 42 cases (as.numeric(alt.employee.dB_cat_test$Attrition) 2).
## Area under the curve: 0.7841
```

```
################ ROC Comaprison ################

#dev.new()
colAUC(cbind(NBM_pred, Knn_pred), alt.employee.dB_cat_test$Attrition, plotROC = TRUE)
```



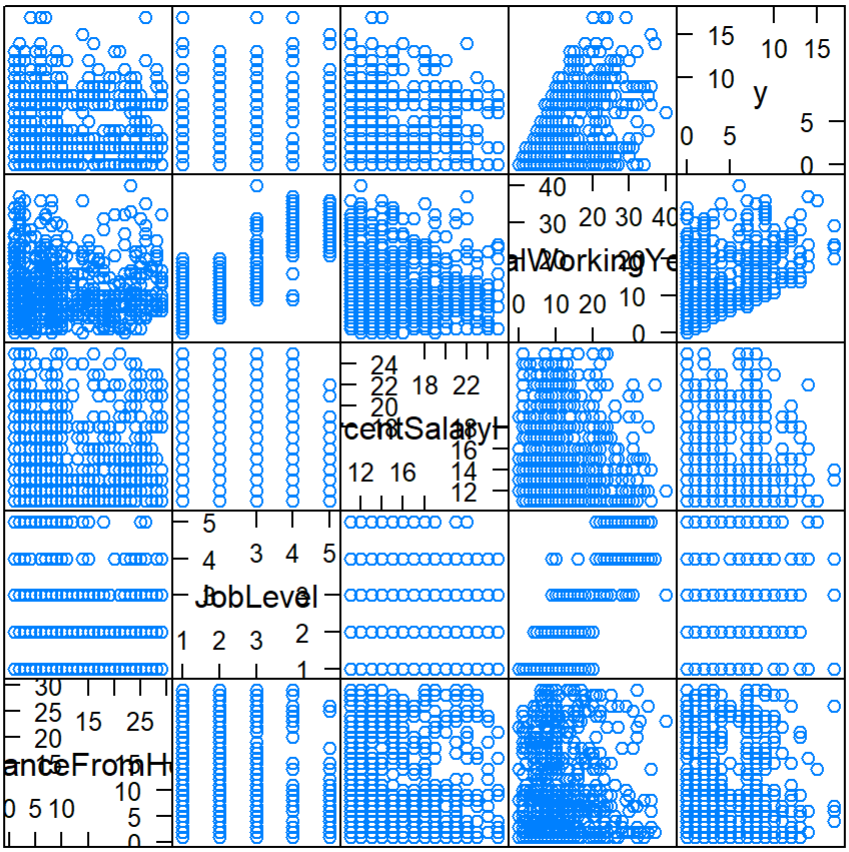

```
##             NBM_pred  Knn_pred
## No vs. Yes 0.7840835 0.5050554
```

```
#########################################################################
################ Generate Outputs for Grading Submission ###############

# Import & Condition Data
path.class.prediction.dB <- "C:\\Users\\dloveday\\Dropbox\\Family\\School\\SMU\\Courses\\Spring
 2021\\DS 6306 - Doing Data Science\\Lecture Notes\\Unit 14 and 15 Case Study 2\\CaseStudy2CompS
et No Attrition.csv"
class.prediction.dB <- read.csv(path.class.prediction.dB)

categorical.variables <- c('RelationshipSatisfaction', 'PerformanceRating', 'WorkLifeBalance',
'JobInvolvement', 'JobSatisfaction', 'JobLevel', 'StockOptionLevel')
class.prediction.dB[,categorical.variables] <- lapply(class.prediction.dB[,categorical.variable
s] , factor, ordered = TRUE)

class.prediction.dB$BusinessTravel <- as.factor(class.prediction.dB$BusinessTravel)
class.prediction.dB$Department <- as.factor(class.prediction.dB$Department)
class.prediction.dB$EducationField <- as.factor(class.prediction.dB$EducationField)
class.prediction.dB$Gender <- as.factor(class.prediction.dB$Gender)
class.prediction.dB$JobRole <- as.factor(class.prediction.dB$JobRole)
class.prediction.dB$MaritalStatus <- as.factor(class.prediction.dB$MaritalStatus)
class.prediction.dB$Over18 <- as.factor(class.prediction.dB$Over18)
class.prediction.dB$OverTime <- as.factor(class.prediction.dB$OverTime)

class.prediction.dB$TimeAtLastJob <- ifelse(class.prediction.dB$NumCompaniesWorked!=0, class.pre
diction.dB$TotalWorkingYears-class.prediction.dB$YearsAtCompany/class.prediction.dB$NumCompanies
Worked,0)

class.prediction.dB$TimeAtLastJob <- ifelse(class.prediction.dB$NumCompaniesWorked!=0, class.pre
diction.dB$TotalWorkingYears-class.prediction.dB$YearsAtCompany/class.prediction.dB$NumCompanies
Worked,0)
class.prediction.dB$AgeGroup <- as.factor(
  ifelse(class.prediction.dB$Age<=30,"Early", ifelse(
    class.prediction.dB$Age<=43,"Mid","Late"
  ))
)
#convert MonthlyIncome
class.prediction.dB$MonthlyIncomeGroup <- as.factor(
  ifelse(class.prediction.dB$MonthlyIncome <= 2911,"1st.Quartile", ifelse(
    class.prediction.dB$MonthlyIncome <= 6503,"2nd.Quartile", ifelse(
      class.prediction.dB$MonthlyIncome <= 8379,"3rd.Quartile","4th.Quartile"
    )))
)
#convert  YearsAtCompany
class.prediction.dB$YrsAtCompanyGroup <- as.factor(
  ifelse(class.prediction.dB$YearsAtCompany <= 3,"1st.Quartile", ifelse(
    class.prediction.dB$YearsAtCompany <= 7,"2nd.Quartile", ifelse(
      class.prediction.dB$YearsAtCompany <= 10,"3rd.Quartile","4th.Quartile"
    )))
)
#convert YearsWithCurrManager
class.prediction.dB$YrsWtCurrManagerGroup <- as.factor(
  ifelse(class.prediction.dB$YearsWithCurrManager <= 2,"1st.Quartile", ifelse(
    class.prediction.dB$YearsWithCurrManager <= 4,"2nd.Quartile", ifelse(
```

```
      class.prediction.dB$YearsWithCurrManager <= 7,"3rd.Quartile","4th.Quartile"
    )))
)
#convert YearsInCurrentRole
class.prediction.dB$YrsInCurrentRoleGroup <- as.factor(
  ifelse(class.prediction.dB$YearsInCurrentRole <= 2,"LessTh2", ifelse(
    class.prediction.dB$YearsInCurrentRole <= 4,"2nd.Quartile", ifelse(
      class.prediction.dB$YearsInCurrentRole <= 7,"3rd.Quartile","4th.Quartile"
    )))
)


# Make predictions using Test Model
class.prediction.dB$Attrition <- predict(NBModel, newdata=class.prediction.dB)

# Generate Outputs
write.csv(class.prediction.dB[, c("ID", "Attrition")], file.path("C:\\Users\\dloveday\\Dropbox
\\Family\\School\\SMU\\Courses\\Spring 2021\\DS 6306 - Doing Data Science\\Lecture Notes\\Unit 1
4 and 15 Case Study 2\\DL Work\\Outputs for Submission\\","Case2PredictionsLoveday Attrition.cs
v"), row.names = FALSE)
```

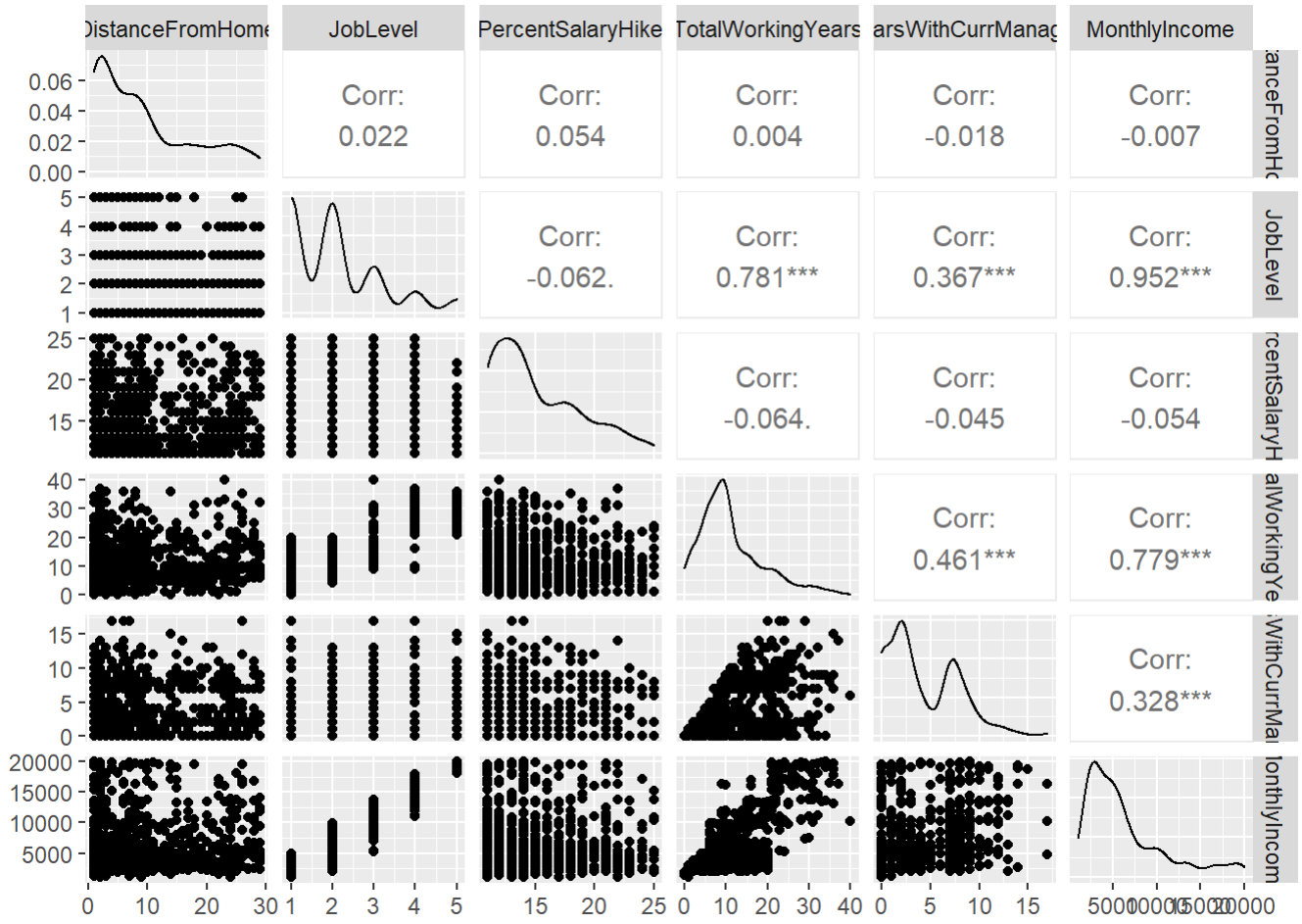
STEP 3 - MLR ANALYSIS FOR MonthlyIncome

```
# All Integer Variables in Dataset for MLR
employee.dB.integers <- employee.dB %>% select_if(is.integer)
employee.dB.integers <- employee.dB.integers[,c(2:ncol(employee.dB.integers))]

# Statistically Significant (p-value <0.05) Coefficient from Total Integer MLR Output
stat.signif.variables <- c("DistanceFromHome", "JobLevel", "PercentSalaryHike", "TotalWorkingYea
rs", "YearsWithCurrManager", "MonthlyIncome")

subset.variables <- stat.signif.variables

employee.dB.MI <- as.data.frame(employee.dB[, c(subset.variables)])
employee.dB.MI.log <- log(employee.dB.MI)
employee.dB.MI.sqrt <- sqrt(employee.dB.MI)

#dev.new()
featurePlot(x=employee.dB.MI[,1:4], y=employee.dB.MI[,5], plot="pairs", auto.key=list(columns=3
))
```

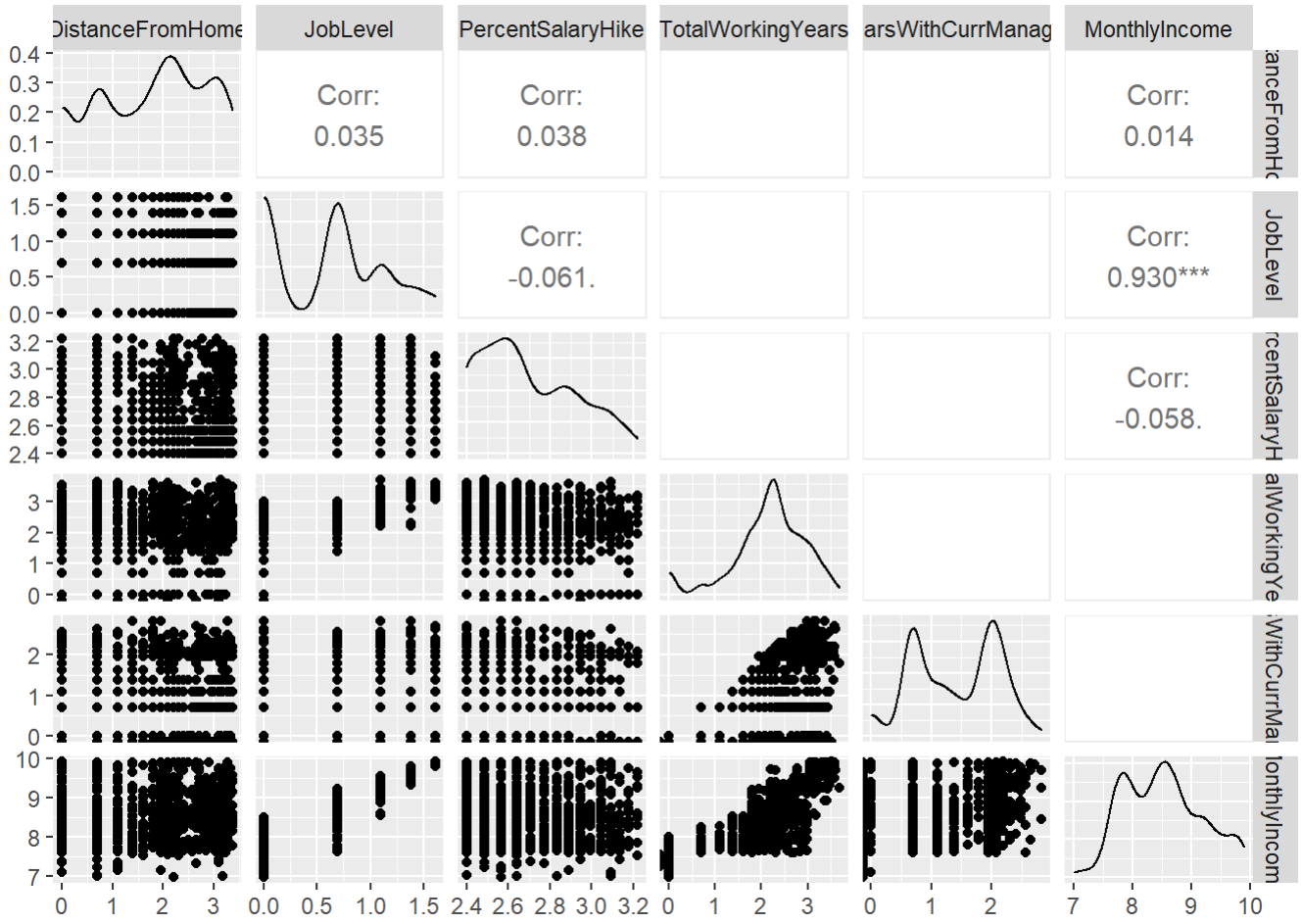


Scatter Plot Matrix

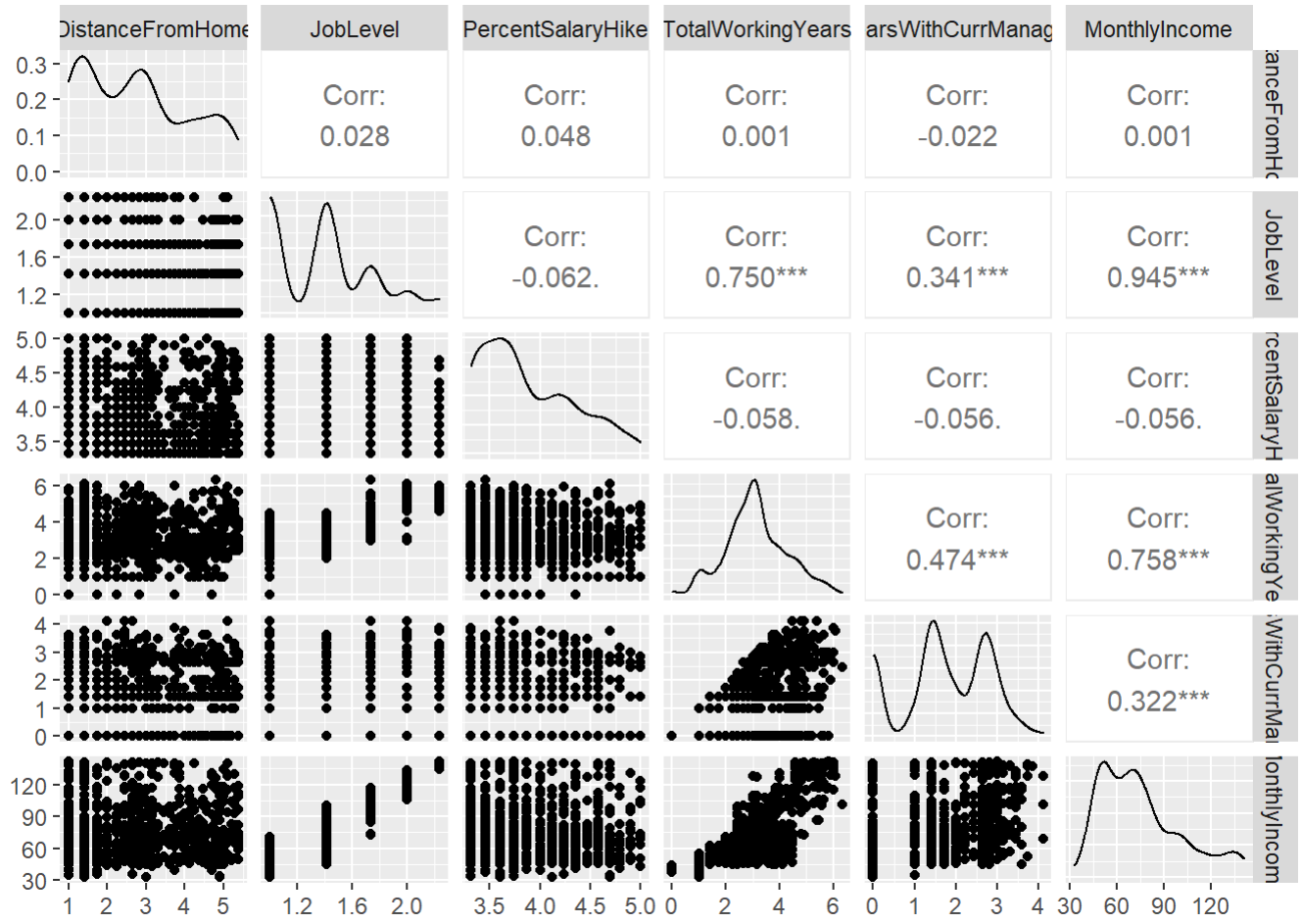
```
#dev.new()
ggpairs(employee.dB.MI)
```



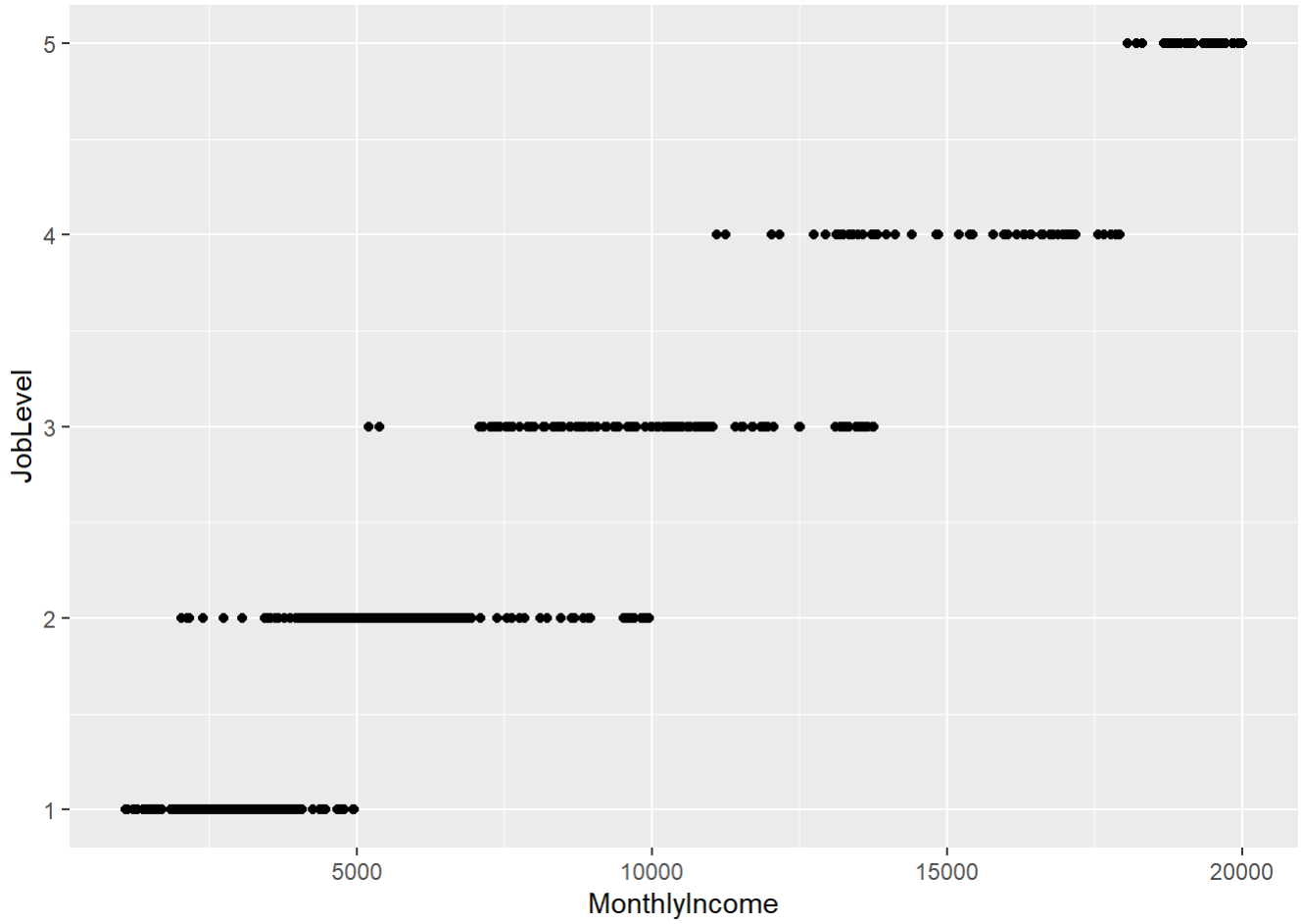

```
#dev.new()
ggpairs(employee.dB.MI.log)
```



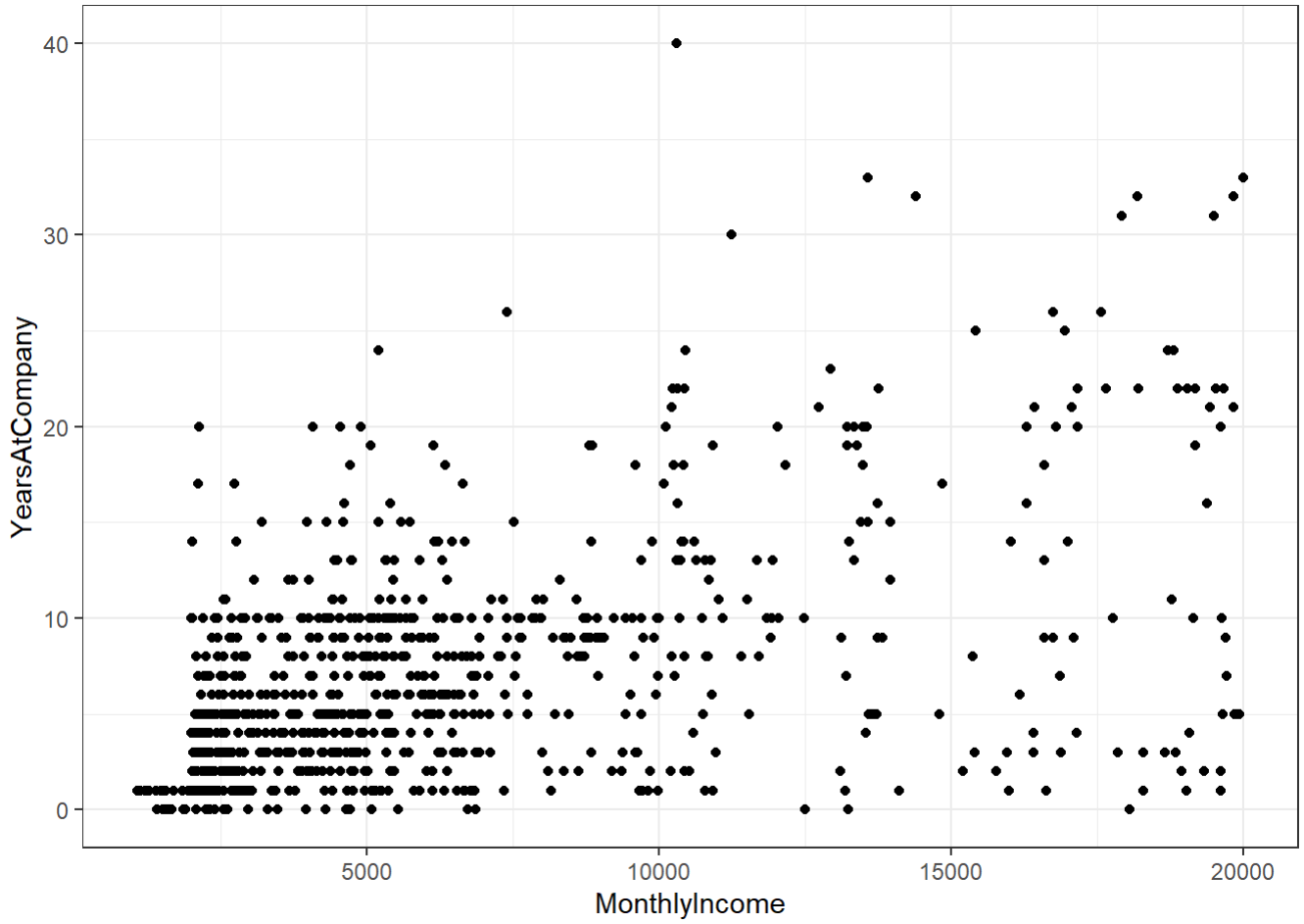

```
#dev.new()
ggpairs(employee.dB.MI.sqrt)
```



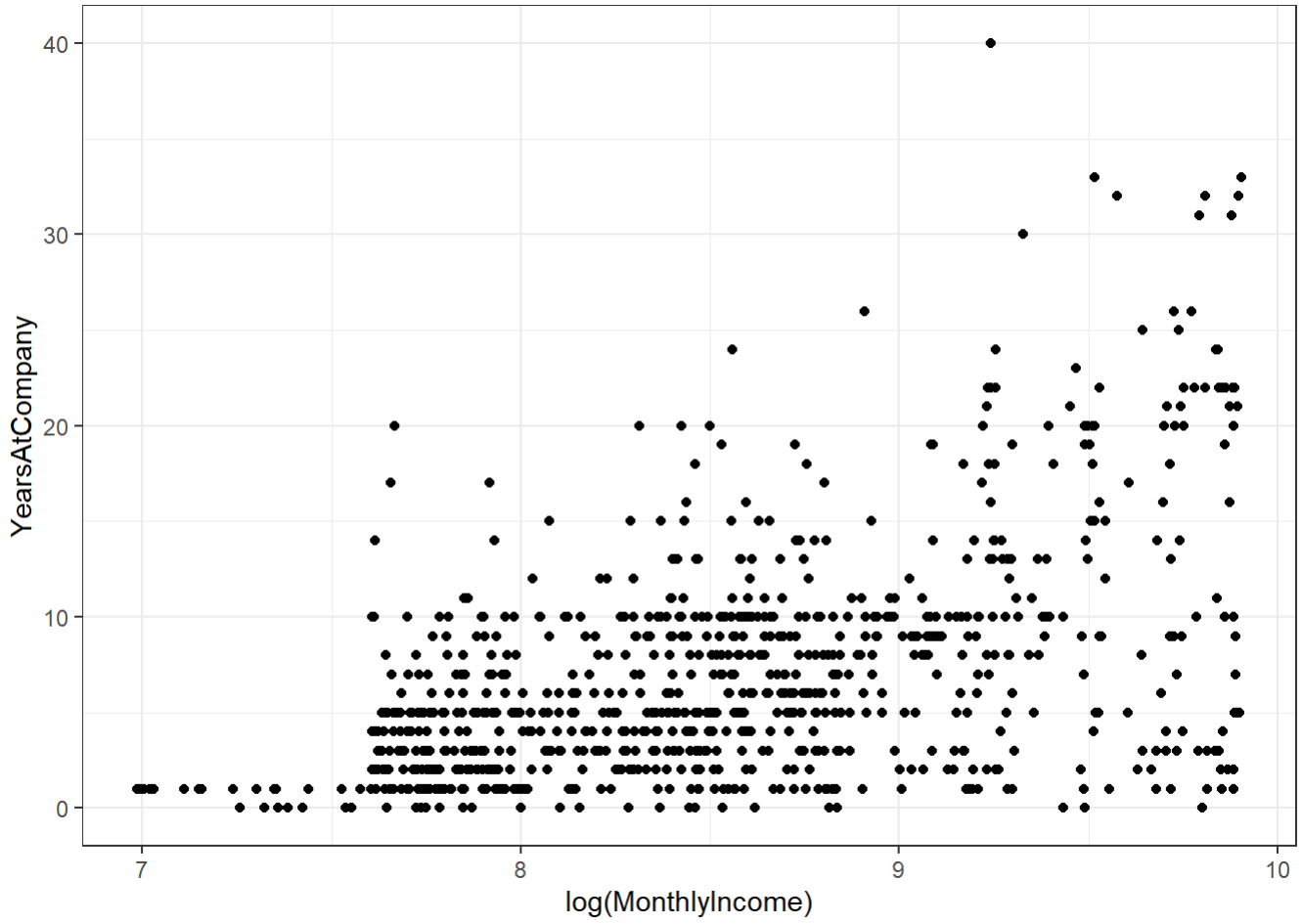

```
#dev.new()
employee.dB.MI %>% ggplot(aes(x = MonthlyIncome, y = JobLevel))+
  geom_point()
```



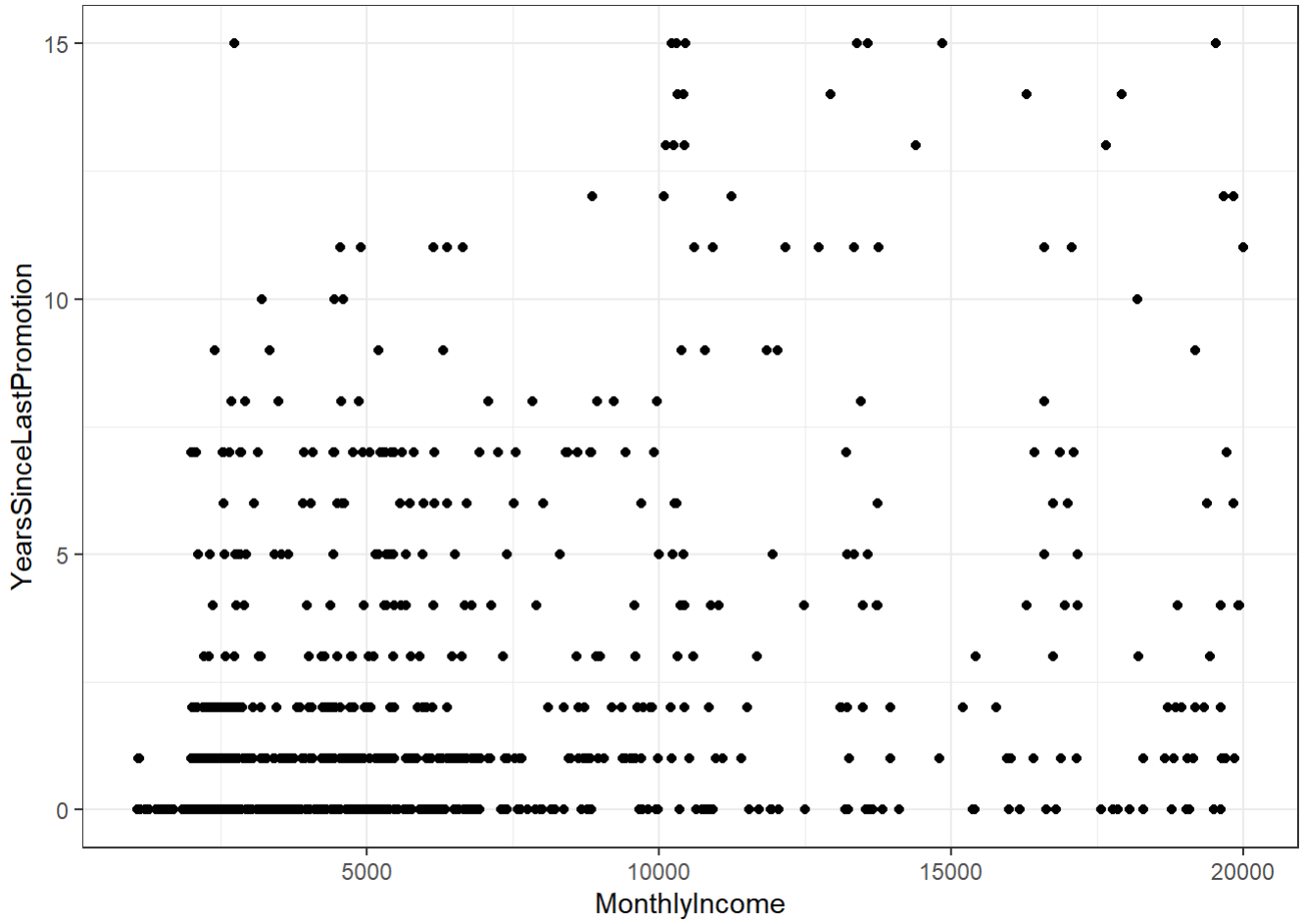

```
#dev.new()
employee.dB.MI %>% ggplot(aes(x = MonthlyIncome, y = YearsAtCompany))+
  theme_bw()+
  geom_point()
```




```
#dev.new()
employee.dB.MI %>% ggplot(aes(x = log(MonthlyIncome), y = YearsAtCompany))+
  theme_bw()+
  geom_point()
```




```
#dev.new()
employee.dB.MI %>% ggplot(aes(x = MonthlyIncome, y = YearsSinceLastPromotion))+
  theme_bw()+
  geom_point()
```




```
################ MLR MODEL BUILDING (USING ALL GIVEN DATA) ################

# All integer column model
fit1 <- lm(MonthlyIncome ~ Age + DailyRate + DistanceFromHome + Education + EmployeeNumber + Env
ironmentSatisfaction + HourlyRate +
             JobInvolvement + JobLevel + JobSatisfaction + MonthlyRate + NumCompaniesWorked + Pe
rcentSalaryHike + PerformanceRating +
             RelationshipSatisfaction + StockOptionLevel + TotalWorkingYears + TrainingTimesLast
Year + WorkLifeBalance + YearsAtCompany + YearsInCurrentRole +
             YearsSinceLastPromotion + YearsWithCurrManager, data = employee.dB)
summary(fit1)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ Age + DailyRate + DistanceFromHome +
##     Education + EmployeeNumber + EnvironmentSatisfaction + HourlyRate +
##     JobInvolvement + JobLevel + JobSatisfaction + MonthlyRate +
##     NumCompaniesWorked + PercentSalaryHike + PerformanceRating +
##     RelationshipSatisfaction + StockOptionLevel + TotalWorkingYears +
##     TrainingTimesLastYear + WorkLifeBalance + YearsAtCompany +
##     YearsInCurrentRole + YearsSinceLastPromotion + YearsWithCurrManager,
##     data = employee.dB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5781.6  -841.4    36.2   692.0  4026.8
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.077e+03  6.697e+02  -1.608   0.1081
## Age                      -3.440e+00  7.237e+00  -0.475   0.6347
## DailyRate                -4.737e-02  1.180e-01  -0.402   0.6882
## DistanceFromHome         -1.665e+01  5.830e+00  -2.856   0.0044 **
## Education                 2.772e+00  4.766e+01   0.058   0.9536
## EmployeeNumber            6.912e-02  7.819e-02   0.884   0.3769
## EnvironmentSatisfaction  -6.656e+01  4.298e+01  -1.549   0.1218
## HourlyRate                1.566e+00  2.359e+00   0.664   0.5071
## JobInvolvement            1.021e+02  6.737e+01   1.516   0.1300
## JobLevel                  3.732e+03  6.975e+01  53.501  < 2e-16 ***
## JobSatisfaction          -6.436e+00  4.272e+01  -0.151   0.8803
## MonthlyRate              -5.185e-03  6.676e-03  -0.777   0.4376
## NumCompaniesWorked       -1.228e+01  2.159e+01  -0.568   0.5698
## PercentSalaryHike         3.542e+01  2.035e+01   1.740   0.0821 .
## PerformanceRating        -3.277e+02  2.077e+02  -1.578   0.1150
## RelationshipSatisfaction  1.674e+01  4.289e+01   0.390   0.6964
## StockOptionLevel          2.174e+00  5.566e+01   0.039   0.9688
## TotalWorkingYears         7.539e+01  1.386e+01   5.439 7.03e-08 ***
## TrainingTimesLastYear     1.795e+01  3.717e+01   0.483   0.6292
## WorkLifeBalance          -3.941e+01  6.653e+01  -0.592   0.5538
## YearsAtCompany           -1.048e+01  1.747e+01  -0.600   0.5487
## YearsInCurrentRole       -5.681e+00  2.177e+01  -0.261   0.7941
## YearsSinceLastPromotion   3.451e+00  1.954e+01   0.177   0.8599
## YearsWithCurrManager     -5.153e+01  2.138e+01  -2.410   0.0162 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1378 on 846 degrees of freedom
## Multiple R-squared:  0.9125, Adjusted R-squared:  0.9101
## F-statistic: 383.6 on 23 and 846 DF,  p-value: < 2.2e-16
```

```
vif(fit1)
```

```
##                     Age               DailyRate        DistanceFromHome
##                  1.9085                  1.0242                  1.0293
##               Education          EmployeeNumber  EnvironmentSatisfaction
##                  1.0872                  1.0227                  1.0213
##              HourlyRate          JobInvolvement                JobLevel
##                  1.0314                  1.0278                  2.6440
##         JobSatisfaction             MonthlyRate       NumCompaniesWorked
##                  1.0367                  1.0300                  1.3550
##       PercentSalaryHike       PerformanceRating RelationshipSatisfaction
##                  2.5595                  2.5423                  1.0228
##        StockOptionLevel       TotalWorkingYears    TrainingTimesLastYear
##                  1.0425                  4.9611                  1.0235
##         WorkLifeBalance          YearsAtCompany       YearsInCurrentRole
##                  1.0262                  5.0611                  2.8697
##  YearsSinceLastPromotion     YearsWithCurrManager
##                  1.7731                  2.6720
```

```
AIC(fit1)
```

```
## [1] 15072.51
```

```
BIC(fit1)
```

```
## [1] 15191.72
```

```
press(fit1)
```

```
## [1] 1706123885
```

```
# Only statistically significant variables
fit2 <- lm(MonthlyIncome ~ DistanceFromHome + JobLevel + PercentSalaryHike + TotalWorkingYears +
YearsWithCurrManager, data = employee.dB)
summary(fit2)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ DistanceFromHome + JobLevel + PercentSalaryHike +
##     TotalWorkingYears + YearsWithCurrManager, data = employee.dB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5758.7  -871.9    16.4   739.8  4035.5
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1707.304    227.298  -7.511 1.46e-13 ***
## DistanceFromHome       -15.572      5.737  -2.714  0.00678 **
## JobLevel              3723.772     68.435  54.413  < 2e-16 ***
## PercentSalaryHike        9.575     12.723   0.753  0.45194
## TotalWorkingYears       68.123     10.408   6.545 1.02e-10 ***
## YearsWithCurrManager   -60.036     14.696  -4.085 4.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1373 on 864 degrees of freedom
## Multiple R-squared:  0.9113, Adjusted R-squared:  0.9108
## F-statistic:  1776 on 5 and 864 DF,  p-value: < 2.2e-16
```

```
vif(fit2)
```

```
##     DistanceFromHome             JobLevel    PercentSalaryHike
##               1.0044               2.5650               1.0079
##    TotalWorkingYears YearsWithCurrManager
##               2.8188               1.2717
```

```
AIC(fit2)
```

```
## [1] 15048.17
```

```
BIC(fit2)
```

```
## [1] 15081.55
```

```
press(fit2)
```

```
## [1] 1654520853
```

```
fit2b <- lm(MonthlyIncome ~ DistanceFromHome + JobLevel + TotalWorkingYears + YearsWithCurrManag
er, data = employee.dB)
summary(fit2b)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ DistanceFromHome + JobLevel + TotalWorkingYears +
##     YearsWithCurrManager, data = employee.dB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5775.8  -859.8    21.0   727.6  4022.3
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1559.495    114.362 -13.637  < 2e-16 ***
## DistanceFromHome       -15.335      5.727  -2.677  0.00756 **
## JobLevel              3722.700     68.403  54.423  < 2e-16 ***
## TotalWorkingYears       67.981     10.404   6.534 1.09e-10 ***
## YearsWithCurrManager   -60.215     14.690  -4.099 4.54e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1373 on 865 degrees of freedom
## Multiple R-squared:  0.9113, Adjusted R-squared:  0.9108
## F-statistic:  2221 on 4 and 865 DF,  p-value: < 2.2e-16
```

```
vif(fit2b)
```

```
##     DistanceFromHome              JobLevel    TotalWorkingYears
##               1.0014               2.5639               2.8179
## YearsWithCurrManager
##               1.2714
```

```
AIC(fit2b)
```

```
## [1] 15046.74
```

```
BIC(fit2b)
```

```
## [1] 15075.35
```

```
press(fit2b)
```

```
## [1] 1651820568
```

```
# Log
#fit2c <- lm(MonthlyIncome ~ DistanceFromHome + JobLevel + TotalWorkingYears + YearsWithCurrMana
ger, data = employee.dB.MI.log)
#summary(fit2c)
#vif(fit2c)
#AIC(fit2c)
#BIC(fit2c)
#press(fit2c)

# Sqrt
fit2d <- lm(MonthlyIncome ~ DistanceFromHome + JobLevel + TotalWorkingYears + YearsWithCurrManag
er, data = employee.dB.MI.sqrt)
summary(fit2d)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ DistanceFromHome + JobLevel + TotalWorkingYears +
##     YearsWithCurrManager, data = employee.dB.MI.sqrt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.628  -5.031  -0.026   4.938  24.171
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -16.6268     1.2462 -13.342  < 2e-16 ***
## DistanceFromHome      -0.4810     0.2122  -2.266  0.02368 *
## JobLevel              62.0347     1.1755  52.772  < 2e-16 ***
## TotalWorkingYears      2.9501     0.4005   7.366  4.1e-13 ***
## YearsWithCurrManager  -0.8087     0.2999  -2.696  0.00715 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.3 on 865 degrees of freedom
## Multiple R-squared:  0.8998, Adjusted R-squared:  0.8993
## F-statistic:  1941 on 4 and 865 DF,  p-value: < 2.2e-16
```

```
vif(fit2d)
```

```
##     DistanceFromHome             JobLevel    TotalWorkingYears
##               1.0024               2.2950               2.6120
## YearsWithCurrManager
##               1.2912
```

```
AIC(fit2d)
```

```
## [1] 6158.224
```

```
BIC(fit2d)
```

```
## [1] 6186.835
```

```
press(fit2d)
```

```
## [1] 60275.97
```

```
# Chi-Squared Impactful Variable Model
fit3 <- lm(MonthlyIncome ~ JobLevel + YearsAtCompany + YearsSinceLastPromotion + YearsWithCurrMa
nager + Age, data = employee.dB)
summary(fit3)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ JobLevel + YearsAtCompany + YearsSinceLastPromotion +
##     YearsWithCurrManager + Age, data = employee.dB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5112.3  -960.7    53.5   745.0  3769.4
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -2251.583    205.015 -10.983  < 2e-16 ***
## JobLevel                  3951.274     55.928  70.650  < 2e-16 ***
## YearsAtCompany              15.347     14.859   1.033  0.30197
## YearsSinceLastPromotion      8.716     19.570   0.445  0.65617
## YearsWithCurrManager       -51.385     20.753  -2.476  0.01347 *
## Age                         18.243      6.099   2.991  0.00286 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1403 on 864 degrees of freedom
## Multiple R-squared:  0.9074, Adjusted R-squared:  0.9068
## F-statistic:  1692 on 5 and 864 DF,  p-value: < 2.2e-16
```

```
vif(fit3)
```

```
##                JobLevel          YearsAtCompany YearsSinceLastPromotion
##                  1.6398                  3.5312                  1.7150
##    YearsWithCurrManager                     Age
##                  2.4276                  1.3074
```

```
AIC(fit3)
```

```
## [1] 15086.22
```

```
BIC(fit3)
```

```
## [1] 15119.6
```

```
press(fit3)
```

```
## [1] 1729161252
```

```
#
Q2_fit_forward <- ols_step_forward_p(fit1, penter = 0.05, details = TRUE)
```

```
## Forward Selection Method
## --------------------------
##
## Candidate Terms:
##
## 1. Age
## 2. DailyRate
## 3. DistanceFromHome
## 4. Education
## 5. EmployeeNumber
## 6. EnvironmentSatisfaction
## 7. HourlyRate
## 8. JobInvolvement
## 9. JobLevel
## 10. JobSatisfaction
## 11. MonthlyRate
## 12. NumCompaniesWorked
## 13. PercentSalaryHike
## 14. PerformanceRating
## 15. RelationshipSatisfaction
## 16. StockOptionLevel
## 17. TotalWorkingYears
## 18. TrainingTimesLastYear
## 19. WorkLifeBalance
## 20. YearsAtCompany
## 21. YearsInCurrentRole
## 22. YearsSinceLastPromotion
## 23. YearsWithCurrManager
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - JobLevel
##
##                            Model Summary
## ----------------------------------------------------------------------
## R                       0.952       RMSE                   1413.296
## R-Squared               0.906       Coef. Var                22.116
## Adj. R-Squared          0.906       MSE                 1997404.971
## Pred R-Squared          0.905       MAE                    1073.683
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF        Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16635876872.790          1    16635876872.790    8328.745    0.0000
```

```
## Residual       1733747514.405        868        1997404.971
## Total         18369624387.195        869
## ----------------------------------------------------------------------------
##
##                                     Parameter Estimates
## ------------------------------------------------------------------------------------------
-----
##       model         Beta    Std. Error    Std. Beta       t        Sig         lower        u
pper
## ------------------------------------------------------------------------------------------
-----
## (Intercept)    -1793.934       101.676                -17.644    0.000    -1993.494    -159
4.375
##    JobLevel     4013.671        43.980        0.952     91.262    0.000     3927.352     409
9.990
## ------------------------------------------------------------------------------------------
-----
##
##
##
## Forward Selection: Step 2
##
## - TotalWorkingYears
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.953       RMSE                 1389.696
## R-Squared               0.909       Coef. Var              21.747
## Adj. R-Squared          0.909       MSE                 1931256.053
## Pred R-Squared          0.908       MAE                    1054.184
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## ---------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## ---------------------------------------------------------------------------
## Regression    16695225388.963          2    8347612694.481    4322.375    0.0000
## Residual       1674398998.232        867       1931256.053
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                         Parameter Estimates
## ------------------------------------------------------------------------------------------
-----------
##             model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ------------------------------------------------------------------------------------------
-----------
##       (Intercept)    -1798.376        99.982                -17.987    0.000    -1994.610
-1602.142
```

```
##          JobLevel     3714.122        69.210         0.881     53.664    0.000     3578.283
3849.961
## TotalWorkingYears       55.664        10.041         0.091      5.544    0.000       35.956
75.372
## ----------------------------------------------------------------------------------------
-----------
##
##
##
## Forward Selection: Step 3
##
## - YearsWithCurrManager
##
##                            Model Summary
## ---------------------------------------------------------------------
## R                       0.954       RMSE                   1377.669
## R-Squared               0.911       Coef. Var                21.559
## Adj. R-Squared          0.910       MSE                 1897970.749
## Pred R-Squared          0.910       MAE                    1035.798
## ---------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ---------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## ---------------------------------------------------------------------------------
## Regression    16725981718.483          3    5575327239.494     2937.52    0.0000
## Residual       1643642668.712        866       1897970.749
## Total         18369624387.195        869
## ---------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------
--------------
##                model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ----------------------------------------------------------------------------------------
--------------
##          (Intercept)    -1699.158       102.135                -16.636    0.000    -1899.618
-1498.697
##             JobLevel     3717.242        68.615         0.881     54.175    0.000     3582.570
3851.914
##    TotalWorkingYears       68.336        10.440         0.112      6.545    0.000       47.845
88.827
## YearsWithCurrManager      -59.331        14.739        -0.046     -4.026    0.000      -88.259
-30.403
## ----------------------------------------------------------------------------------------
--------------
##
##
##
```

```
## Forward Selection: Step 4
##
## - DistanceFromHome
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1372.788
## R-Squared               0.911       Coef. Var              21.482
## Adj. R-Squared          0.911       MSE                 1884546.340
## Pred R-Squared          0.910       MAE                    1032.363
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16739491803.372          4    4184872950.843    2220.626    0.0000
## Residual       1630132583.824        865       1884546.340
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                        Parameter Estimates
## ----------------------------------------------------------------------------------------------
--------------
##                model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ----------------------------------------------------------------------------------------------
--------------
##          (Intercept)    -1559.495       114.362                -13.637    0.000    -1783.954
-1335.037
##             JobLevel     3722.700        68.403        0.883     54.423    0.000     3588.445
3856.955
##    TotalWorkingYears       67.981        10.404        0.111      6.534    0.000       47.561
88.402
## YearsWithCurrManager      -60.215        14.690       -0.047     -4.099    0.000      -89.047
-31.382
##     DistanceFromHome      -15.335         5.727       -0.027     -2.677    0.008      -26.576
-4.094
## ----------------------------------------------------------------------------------------------
--------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + JobLevel
## + TotalWorkingYears
## + YearsWithCurrManager
```

```
## + DistanceFromHome
##
##
## Final Model Output
## ------------------
##
##                             Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1372.788
## R-Squared               0.911       Coef. Var              21.482
## Adj. R-Squared          0.911       MSE                1884546.340
## Pred R-Squared          0.910       MAE                   1032.363
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16739491803.372          4    4184872950.843    2220.626    0.0000
## Residual       1630132583.824        865       1884546.340
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------------
--------------
##                model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ----------------------------------------------------------------------------------------------------
--------------
##          (Intercept)    -1559.495       114.362                -13.637    0.000    -1783.954
-1335.037
##              JobLevel     3722.700        68.403        0.883     54.423    0.000     3588.445
3856.955
##    TotalWorkingYears       67.981        10.404        0.111      6.534    0.000       47.561
88.402
## YearsWithCurrManager      -60.215        14.690       -0.047     -4.099    0.000      -89.047
-31.382
##     DistanceFromHome      -15.335         5.727       -0.027     -2.677    0.008      -26.576
-4.094
## ----------------------------------------------------------------------------------------------------
--------------
```

```
Q2_fit_backward <- ols_step_backward_p(fit1, penter = 0.05, details = TRUE)
```

```
## Backward Elimination Method
## --------------------------
##
## Candidate Terms:
##
## 1 . Age
## 2 . DailyRate
## 3 . DistanceFromHome
## 4 . Education
## 5 . EmployeeNumber
## 6 . EnvironmentSatisfaction
## 7 . HourlyRate
## 8 . JobInvolvement
## 9 . JobLevel
## 10 . JobSatisfaction
## 11 . MonthlyRate
## 12 . NumCompaniesWorked
## 13 . PercentSalaryHike
## 14 . PerformanceRating
## 15 . RelationshipSatisfaction
## 16 . StockOptionLevel
## 17 . TotalWorkingYears
## 18 . TrainingTimesLastYear
## 19 . WorkLifeBalance
## 20 . YearsAtCompany
## 21 . YearsInCurrentRole
## 22 . YearsSinceLastPromotion
## 23 . YearsWithCurrManager
##
## We are eliminating variables based on p value...
##
## - StockOptionLevel
##
## Backward Elimination: Step 1
##
##  Variable StockOptionLevel Removed
##
##                            Model Summary
## ----------------------------------------------------------------------
## R                        0.955       RMSE                  1377.580
## R-Squared                0.912       Coef. Var               21.557
## Adj. R-Squared           0.910       MSE                1897725.643
## Pred R-Squared           0.907       MAE                   1029.927
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF      Mean Square       F         Sig.
## --------------------------------------------------------------------------------
```

```
## Regression    16762250767.827         22    761920489.447    401.491    0.0000
## Residual       1607373619.368        847      1897725.643
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ------------------------------------------------------------------------------------------
----------------
##                     model         Beta    Std. Error    Std. Beta      t        Sig         lo
wer       upper
## ------------------------------------------------------------------------------------------
----------------
##               (Intercept)    -1077.144       669.346                -1.609    0.108    -2390.
915     236.628
##                       Age       -3.433         7.230       -0.007    -0.475    0.635      -17.
624      10.759
##                 DailyRate       -0.047         0.118       -0.004    -0.401    0.688       -0.
279       0.184
##           DistanceFromHome      -16.635         5.811       -0.029    -2.863    0.004      -28.
040      -5.229
##                 Education        2.789        47.626        0.001     0.059    0.953      -90.
690      96.268
##             EmployeeNumber        0.069         0.078        0.009     0.892    0.373       -0.
083       0.222
##  EnvironmentSatisfaction      -66.516        42.939       -0.016    -1.549    0.122     -150.
796      17.764
##                 HourlyRate        1.571         2.354        0.007     0.667    0.505       -3.
050       6.192
##             JobInvolvement      102.278        67.196        0.016     1.522    0.128      -29.
611     234.168
##                  JobLevel     3731.491        69.702        0.885    53.535    0.000     3594.
682    3868.299
##             JobSatisfaction       -6.413        42.688       -0.002    -0.150    0.881      -90.
199      77.373
##                MonthlyRate       -0.005         0.007       -0.008    -0.780    0.436       -0.
018       0.008
##       NumCompaniesWorked      -12.252        21.573       -0.007    -0.568    0.570      -54.
594      30.091
##        PercentSalaryHike       35.433        20.340        0.028     1.742    0.082       -4.
489      75.355
##        PerformanceRating     -327.851       207.525       -0.026    -1.580    0.115     -735.
174      79.472
## RelationshipSatisfaction       16.681        42.841        0.004     0.389    0.697      -67.
406     100.768
##        TotalWorkingYears       75.386        13.853        0.123     5.442    0.000       48.
195     102.576
##    TrainingTimesLastYear       17.992        37.134        0.005     0.485    0.628      -54.
894      90.878
##           WorkLifeBalance      -39.306        66.439       -0.006    -0.592    0.554     -169.
712      91.099
##            YearsAtCompany      -10.500        17.454       -0.014    -0.602    0.548      -44.
759      23.759
##       YearsInCurrentRole       -5.607        21.668       -0.004    -0.259    0.796      -48.
137      36.924
```

```
##  YearsSinceLastPromotion        3.438        19.529        0.002     0.176    0.860      -34.
893      41.769
##     YearsWithCurrManager      -51.536        21.370       -0.040    -2.412    0.016      -93.
481      -9.591
## ----------------------------------------------------------------------------------------
----------------
##
##
## - Education
##
## Backward Elimination: Step 2
##
##  Variable Education Removed
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1376.770
## R-Squared               0.912       Coef. Var              21.545
## Adj. R-Squared          0.910       MSE                 1895495.433
## Pred R-Squared          0.908       MAE                    1029.843
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## ---------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF      Mean Square       F         Sig.
## ---------------------------------------------------------------------------
## Regression    16762244260.177         21    798202107.627    421.105    0.0000
## Residual       1607380127.019        848      1895495.433
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------
----------------
##                    model         Beta    Std. Error    Std. Beta      t        Sig         lo
wer       upper
## ----------------------------------------------------------------------------------------
----------------
##              (Intercept)    -1070.892       660.388                -1.622    0.105    -2367.
079     225.295
##                      Age       -3.368         7.142       -0.007    -0.472    0.637      -17.
386      10.649
##                DailyRate       -0.048         0.118       -0.004    -0.403    0.687       -0.
279       0.184
##         DistanceFromHome      -16.616         5.799       -0.029    -2.865    0.004      -27.
998      -5.234
##            EmployeeNumber        0.069         0.078        0.009     0.894    0.372       -0.
083       0.222
##  EnvironmentSatisfaction      -66.609        42.885       -0.016    -1.553    0.121     -150.
781      17.564
```

```
##               HourlyRate        1.570         2.353        0.007     0.667    0.505       -3.
048       6.189
##           JobInvolvement      102.397        67.126        0.016     1.525    0.128      -29.
355     234.149
##                JobLevel     3731.711        69.559        0.885    53.648    0.000     3595.
182    3868.240
##          JobSatisfaction       -6.340        42.644       -0.002    -0.149    0.882      -90.
041      77.361
##             MonthlyRate       -0.005         0.007       -0.008    -0.783    0.434       -0.
018       0.008
##       NumCompaniesWorked      -12.100        21.404       -0.007    -0.565    0.572      -54.
111      29.911
##        PercentSalaryHike       35.475        20.315        0.028     1.746    0.081       -4.
398      75.349
##        PerformanceRating     -328.355       207.225       -0.026    -1.585    0.113     -735.
088      78.379
## RelationshipSatisfaction       16.601        42.794        0.004     0.388    0.698      -67.
393     100.595
##        TotalWorkingYears       75.340        13.823        0.123     5.450    0.000       48.
208     102.473
##    TrainingTimesLastYear       17.899        37.079        0.005     0.483    0.629      -54.
877      90.676
##          WorkLifeBalance      -39.254        66.394       -0.006    -0.591    0.555     -169.
571      91.062
##           YearsAtCompany      -10.529        17.437       -0.014    -0.604    0.546      -44.
754      23.697
##       YearsInCurrentRole       -5.604        21.656       -0.004    -0.259    0.796      -48.
109      36.901
##  YearsSinceLastPromotion        3.473        19.508        0.002     0.178    0.859      -34.
817      41.763
##     YearsWithCurrManager      -51.445        21.301       -0.040    -2.415    0.016      -93.
253      -9.636
## ----------------------------------------------------------------------------------------
----------------
##
##
## - JobSatisfaction
##
## Backward Elimination: Step 3
##
##  Variable JobSatisfaction Removed
##
##                            Model Summary
## ------------------------------------------------------------------
## R                       0.955       RMSE                 1375.977
## R-Squared               0.912       Coef. Var              21.532
## Adj. R-Squared          0.910       MSE                 1893312.160
## Pred R-Squared          0.908       MAE                    1029.735
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
```

```
## ---------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF      Mean Square       F         Sig.
## ---------------------------------------------------------------------------
## Regression    16762202363.048         20    838110118.152    442.669    0.0000
## Residual       1607422024.147        849      1893312.160
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ------------------------------------------------------------------------------------------------
----------------
##                     model         Beta    Std. Error    Std. Beta      t        Sig         lo
wer       upper
## ------------------------------------------------------------------------------------------------
----------------
##               (Intercept)    -1094.432       640.756                -1.708    0.088    -2352.
084     163.220
##                       Age       -3.403         7.134       -0.007    -0.477    0.633      -17.
405      10.598
##                DailyRate       -0.048         0.118       -0.004    -0.404    0.686       -0.
279       0.183
##          DistanceFromHome      -16.601         5.795       -0.029    -2.865    0.004      -27.
975      -5.228
##             EmployeeNumber        0.070         0.078        0.009     0.906    0.365       -0.
082       0.222
##  EnvironmentSatisfaction      -66.492        42.853       -0.016    -1.552    0.121     -150.
602      17.618
##                HourlyRate        1.600         2.343        0.007     0.683    0.495       -3.
000       6.199
##             JobInvolvement      102.833        67.023        0.016     1.534    0.125      -28.
717     234.383
##                  JobLevel     3732.055        69.481        0.885    53.713    0.000     3595.
681    3868.429
##               MonthlyRate       -0.005         0.007       -0.008    -0.789    0.430       -0.
018       0.008
##       NumCompaniesWorked      -11.966        21.373       -0.007    -0.560    0.576      -53.
915      29.983
##        PercentSalaryHike       35.428        20.301        0.028     1.745    0.081       -4.
417      75.274
##        PerformanceRating     -327.835       207.076       -0.026    -1.583    0.114     -734.
276      78.605
## RelationshipSatisfaction       16.806        42.747        0.004     0.393    0.694      -67.
096     100.708
##        TotalWorkingYears       75.417        13.806        0.123     5.463    0.000       48.
320     102.515
##    TrainingTimesLastYear       18.141        37.022        0.005     0.490    0.624      -54.
523      90.806
##           WorkLifeBalance      -39.078        66.346       -0.006    -0.589    0.556     -169.
298      91.142
##            YearsAtCompany      -10.745        17.367       -0.014    -0.619    0.536      -44.
831      23.342
##       YearsInCurrentRole       -5.509        21.634       -0.004    -0.255    0.799      -47.
971      36.953
```

```
##  YearsSinceLastPromotion        3.604        19.477        0.002     0.185    0.853      -34.
625      41.833
##     YearsWithCurrManager      -51.399        21.286       -0.040    -2.415    0.016      -93.
179      -9.619
## ----------------------------------------------------------------------------------------
----------------
##
##
## - YearsSinceLastPromotion
##
## Backward Elimination: Step 4
##
##  Variable YearsSinceLastPromotion Removed
##
##                            Model Summary
## ---------------------------------------------------------------------
## R                       0.955       RMSE                   1375.195
## R-Squared               0.912       Coef. Var                21.520
## Adj. R-Squared          0.911       MSE                 1891161.000
## Pred R-Squared          0.908       MAE                    1029.487
## ---------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## ---------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF      Mean Square       F         Sig.
## ---------------------------------------------------------------------------
## Regression    16762137537.015         19    882217765.106    466.495    0.0000
## Residual       1607486850.181        850      1891161.000
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------
----------------
##                    model         Beta    Std. Error    Std. Beta      t        Sig         lo
wer       upper
## ----------------------------------------------------------------------------------------
----------------
##              (Intercept)    -1094.408       640.392                -1.709    0.088    -2351.
343     162.527
##                      Age       -3.397         7.129       -0.007    -0.476    0.634      -17.
391      10.596
##                DailyRate       -0.049         0.118       -0.004    -0.414    0.679       -0.
279       0.182
##          DistanceFromHome      -16.607         5.791       -0.029    -2.868    0.004      -27.
974      -5.241
##             EmployeeNumber        0.070         0.077        0.009     0.909    0.364       -0.
082       0.223
##  EnvironmentSatisfaction      -66.354        42.822       -0.016    -1.550    0.122     -150.
403      17.695
```

```
##               HourlyRate        1.603         2.342        0.007     0.684    0.494       -2.
994       6.200
##           JobInvolvement      102.658        66.978        0.016     1.533    0.126      -28.
804     234.120
##                JobLevel     3731.152        69.270        0.885    53.864    0.000     3595.
192    3867.112
##              MonthlyRate       -0.005         0.007       -0.008    -0.787    0.431       -0.
018       0.008
##       NumCompaniesWorked      -12.026        21.358       -0.007    -0.563    0.574      -53.
947      29.894
##        PercentSalaryHike       35.281        20.274        0.028     1.740    0.082       -4.
511      75.073
##        PerformanceRating     -327.392       206.944       -0.026    -1.582    0.114     -733.
574      78.790
## RelationshipSatisfaction       17.126        42.688        0.004     0.401    0.688      -66.
660     100.911
##        TotalWorkingYears       75.610        13.759        0.124     5.495    0.000       48.
605     102.615
##    TrainingTimesLastYear       17.683        36.918        0.005     0.479    0.632      -54.
777      90.143
##          WorkLifeBalance      -38.842        66.296       -0.006    -0.586    0.558     -168.
965      91.280
##           YearsAtCompany       -9.800        16.590       -0.013    -0.591    0.555      -42.
361      22.762
##       YearsInCurrentRole       -5.093        21.505       -0.004    -0.237    0.813      -47.
302      37.115
##     YearsWithCurrManager      -51.383        21.274       -0.040    -2.415    0.016      -93.
139      -9.628
## ----------------------------------------------------------------------------------------------
----------------
##
##
## - YearsInCurrentRole
##
## Backward Elimination: Step 5
##
##  Variable YearsInCurrentRole Removed
##
##                            Model Summary
## ------------------------------------------------------------------
## R                       0.955       RMSE                 1374.432
## R-Squared               0.912       Coef. Var              21.508
## Adj. R-Squared          0.911       MSE                 1889063.387
## Pred R-Squared          0.908       MAE                    1029.712
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF      Mean Square       F         Sig.
## --------------------------------------------------------------------------------
```

```
## Regression    16762031444.853         18    931223969.159    492.955    0.0000
## Residual       1607592942.342        851      1889063.387
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ---------------------------------------------------------------------------------------------
----------------
##                     model         Beta    Std. Error    Std. Beta      t        Sig         lo
wer       upper
## ---------------------------------------------------------------------------------------------
----------------
##               (Intercept)    -1085.823       639.011                -1.699    0.090    -2340.
045     168.399
##                       Age       -3.366         7.124       -0.007    -0.472    0.637      -17.
349      10.617
##                 DailyRate       -0.050         0.117       -0.004    -0.426    0.670       -0.
280       0.180
##           DistanceFromHome      -16.630         5.787       -0.029    -2.873    0.004      -27.
989      -5.271
##              EmployeeNumber        0.071         0.077        0.009     0.913    0.362       -0.
081       0.223
##  EnvironmentSatisfaction      -66.875        42.742       -0.016    -1.565    0.118     -150.
766      17.017
##                HourlyRate        1.603         2.341        0.007     0.685    0.494       -2.
992       6.197
##             JobInvolvement      101.809        66.845        0.016     1.523    0.128      -29.
392     233.009
##                  JobLevel     3731.450        69.220        0.885    53.907    0.000     3595.
588    3867.312
##                MonthlyRate       -0.005         0.007       -0.008    -0.806    0.420       -0.
018       0.008
##        NumCompaniesWorked      -12.116        21.343       -0.007    -0.568    0.570      -54.
007      29.774
##         PercentSalaryHike       35.269        20.262        0.028     1.741    0.082       -4.
501      75.039
##         PerformanceRating     -328.526       206.774       -0.026    -1.589    0.112     -734.
373      77.321
## RelationshipSatisfaction       17.192        42.663        0.004     0.403    0.687      -66.
545     100.929
##         TotalWorkingYears       75.585        13.751        0.124     5.497    0.000       48.
596     102.575
##    TrainingTimesLastYear       18.091        36.857        0.005     0.491    0.624      -54.
251      90.432
##           WorkLifeBalance      -40.358        65.949       -0.006    -0.612    0.541     -169.
801      89.084
##            YearsAtCompany      -11.582        14.777       -0.015    -0.784    0.433      -40.
585      17.421
##     YearsWithCurrManager      -52.805        20.398       -0.041    -2.589    0.010      -92.
842     -12.768
## ---------------------------------------------------------------------------------------------
----------------
##
##
```

```
## - RelationshipSatisfaction
##
## Backward Elimination: Step 6
##
##  Variable RelationshipSatisfaction Removed
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1373.756
## R-Squared               0.912       Coef. Var              21.498
## Adj. R-Squared          0.911       MSE                1887206.217
## Pred R-Squared          0.909       MAE                   1029.237
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF      Mean Square       F         Sig.
## --------------------------------------------------------------------------------
## Regression    16761724690.675         17    985983805.334    522.457    0.0000
## Residual       1607899696.520        852      1887206.217
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ----------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1043.405       629.971                -1.656    0.098    -2279.8
83     193.073
##                     Age       -3.341         7.121       -0.006    -0.469    0.639      -17.3
17      10.635
##               DailyRate       -0.049         0.117       -0.004    -0.422    0.673       -0.2
80       0.181
##        DistanceFromHome      -16.527         5.779       -0.029    -2.860    0.004      -27.8
69      -5.184
##           EmployeeNumber        0.069         0.077        0.009     0.890    0.374       -0.0
83       0.220
## EnvironmentSatisfaction      -66.803        42.720       -0.016    -1.564    0.118     -150.6
52      17.047
##              HourlyRate        1.614         2.339        0.007     0.690    0.490       -2.9
78       6.206
##          JobInvolvement      102.366        66.798        0.016     1.532    0.126      -28.7
42     233.474
##                JobLevel     3731.941        69.175        0.885    53.949    0.000     3596.1
67    3867.714
##             MonthlyRate       -0.005         0.007       -0.008    -0.814    0.416       -0.0
18       0.008
```

```
##      NumCompaniesWorked      -11.543        21.285       -0.006    -0.542    0.588      -53.3
19      30.234
##       PercentSalaryHike       34.937        20.236        0.028     1.726    0.085       -4.7
81      74.654
##       PerformanceRating     -327.558       206.658       -0.026    -1.585    0.113     -733.1
77      78.062
##       TotalWorkingYears       75.292        13.725        0.123     5.486    0.000       48.3
54     102.230
##   TrainingTimesLastYear       18.355        36.833        0.005     0.498    0.618      -53.9
39      90.650
##          WorkLifeBalance      -39.405        65.874       -0.006    -0.598    0.550     -168.7
00      89.890
##           YearsAtCompany      -11.211        14.741       -0.015    -0.761    0.447      -40.1
43      17.721
##    YearsWithCurrManager      -53.109        20.374       -0.041    -2.607    0.009      -93.0
99     -13.119
## ----------------------------------------------------------------------------------------------
---------------
##
##
## - DailyRate
##
## Backward Elimination: Step 7
##
##  Variable DailyRate Removed
##
##                            Model Summary
## ---------------------------------------------------------------------
## R                       0.955       RMSE                  1373.094
## R-Squared               0.912       Coef. Var               21.487
## Adj. R-Squared          0.911       MSE                1885387.313
## Pred R-Squared          0.909       MAE                   1030.016
## ---------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ---------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F         Sig.
## ---------------------------------------------------------------------------------
## Regression    16761389008.995         16    1047586813.062    555.635    0.0000
## Residual       1608235378.200        853       1885387.313
## Total         18369624387.195        869
## ---------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ----------------------------------------------------------------------------------------------
---------------
```

```
##             (Intercept)    -1089.871       619.964                -1.758    0.079    -2306.7
04     126.962
##                     Age       -3.357         7.117        -0.007    -0.472    0.637      -17.3
26      10.612
##        DistanceFromHome      -16.550         5.776        -0.029    -2.865    0.004      -27.8
87      -5.214
##          EmployeeNumber        0.070         0.077         0.009     0.903    0.367       -0.0
82       0.221
## EnvironmentSatisfaction      -66.640        42.698        -0.016    -1.561    0.119     -150.4
45      17.166
##              HourlyRate        1.569         2.336         0.007     0.672    0.502       -3.0
16       6.153
##          JobInvolvement      100.813        66.664         0.015     1.512    0.131      -30.0
32     231.658
##                JobLevel     3731.226        69.121         0.885    53.981    0.000     3595.5
59    3866.894
##             MonthlyRate       -0.005         0.007        -0.008    -0.803    0.422       -0.0
18       0.008
##       NumCompaniesWorked      -11.945        21.253        -0.007    -0.562    0.574      -53.6
60      29.769
##        PercentSalaryHike       34.313        20.172         0.027     1.701    0.089       -5.2
79      73.905
##        PerformanceRating     -321.254       206.018        -0.025    -1.559    0.119     -725.6
16      83.107
##        TotalWorkingYears       75.391        13.716         0.123     5.497    0.000       48.4
70     102.312
##   TrainingTimesLastYear       18.438        36.815         0.005     0.501    0.617      -53.8
21      90.696
##          WorkLifeBalance      -38.643        65.818        -0.006    -0.587    0.557     -167.8
27      90.541
##           YearsAtCompany      -11.112        14.732        -0.015    -0.754    0.451      -40.0
27      17.802
##    YearsWithCurrManager      -53.171        20.364        -0.041    -2.611    0.009      -93.1
41     -13.202
## ----------------------------------------------------------------------------------------------
---------------
##
##
## - Age
##
## Backward Elimination: Step 8
##
##  Variable Age Removed
##
##                            Model Summary
## ------------------------------------------------------------------
## R                       0.955       RMSE                  1372.469
## R-Squared               0.912       Coef. Var               21.477
## Adj. R-Squared          0.911       MSE                1883670.769
## Pred R-Squared          0.909       MAE                   1030.615
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
```

```
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF       Mean Square       F         Sig.
## --------------------------------------------------------------------------------
## Regression    16760969550.680         15    1117397970.045    593.202    0.0000
## Residual       1608654836.515        854       1883670.769
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1188.753       583.178                -2.038    0.042    -2333.3
83     -44.123
##        DistanceFromHome      -16.558         5.773       -0.029    -2.868    0.004      -27.8
89      -5.227
##           EmployeeNumber        0.070         0.077        0.009     0.905    0.366       -0.0
82       0.221
## EnvironmentSatisfaction      -66.780        42.678       -0.016    -1.565    0.118     -150.5
45      16.986
##              HourlyRate        1.527         2.333        0.007     0.655    0.513       -3.0
52       6.107
##           JobInvolvement       99.960        66.609        0.015     1.501    0.134      -30.7
77     230.697
##                JobLevel     3732.742        69.015        0.885    54.086    0.000     3597.2
83    3868.201
##             MonthlyRate       -0.005         0.007       -0.008    -0.821    0.412       -0.0
18       0.008
##       NumCompaniesWorked      -12.852        21.156       -0.007    -0.607    0.544      -54.3
76      28.672
##        PercentSalaryHike       33.945        20.147        0.027     1.685    0.092       -5.6
00      73.489
##        PerformanceRating     -317.285       205.752       -0.025    -1.542    0.123     -721.1
24      86.555
##        TotalWorkingYears       72.293        12.036        0.118     6.006    0.000       48.6
70      95.917
##   TrainingTimesLastYear       18.584        36.797        0.005     0.505    0.614      -53.6
39      90.806
##          WorkLifeBalance      -37.738        65.760       -0.006    -0.574    0.566     -166.8
08      91.332
##           YearsAtCompany      -10.466        14.661       -0.014    -0.714    0.476      -39.2
42      18.310
##    YearsWithCurrManager      -52.826        20.341       -0.041    -2.597    0.010      -92.7
51     -12.900
## ------------------------------------------------------------------------------------------
---------------
##
##
```

```
## - TrainingTimesLastYear
##
## Backward Elimination: Step 9
##
##  Variable TrainingTimesLastYear Removed
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1371.871
## R-Squared               0.912       Coef. Var              21.468
## Adj. R-Squared          0.911       MSE                 1882029.576
## Pred R-Squared          0.909       MAE                    1031.783
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F         Sig.
## --------------------------------------------------------------------------------
## Regression    16760489099.782         14    1197177792.842     636.11    0.0000
## Residual       1609135287.414        855       1882029.576
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ----------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1129.014       570.807                -1.978    0.048    -2249.3
62      -8.667
##        DistanceFromHome      -16.677         5.766       -0.030    -2.892    0.004      -27.9
94      -5.361
##           EmployeeNumber        0.070         0.077        0.009     0.909    0.363       -0.0
81       0.221
## EnvironmentSatisfaction      -67.149        42.653       -0.016    -1.574    0.116     -150.8
65      16.568
##              HourlyRate        1.541         2.332        0.007     0.661    0.509       -3.0
36       6.118
##          JobInvolvement       99.232        66.565        0.015     1.491    0.136      -31.4
17     229.882
##               JobLevel     3731.806        68.960        0.885    54.116    0.000     3596.4
55    3867.156
##            MonthlyRate       -0.005         0.007       -0.008    -0.825    0.410       -0.0
18       0.008
##      NumCompaniesWorked      -13.353        21.124       -0.007    -0.632    0.527      -54.8
13      28.108
##       PercentSalaryHike       34.077        20.137        0.027     1.692    0.091       -5.4
47      73.600
```

```
##       PerformanceRating     -318.892       205.638       -0.025    -1.551    0.121     -722.5
06      84.723
##       TotalWorkingYears       72.205        12.030        0.118     6.002    0.000       48.5
94      95.816
##          WorkLifeBalance      -36.918        65.711       -0.006    -0.562    0.574     -165.8
93      92.056
##          YearsAtCompany      -10.172        14.643       -0.013    -0.695    0.487      -38.9
13      18.569
##    YearsWithCurrManager      -53.037        20.328       -0.041    -2.609    0.009      -92.9
37     -13.138
## ----------------------------------------------------------------------------------------------
---------------
##
##
## - WorkLifeBalance
##
## Backward Elimination: Step 10
##
##  Variable WorkLifeBalance Removed
##
##                             Model Summary
## ------------------------------------------------------------------
## R                       0.955       RMSE                  1371.322
## R-Squared               0.912       Coef. Var               21.460
## Adj. R-Squared          0.911       MSE                1880524.941
## Pred R-Squared          0.909       MAE                   1031.586
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ---------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F         Sig.
## ---------------------------------------------------------------------------------
## Regression    16759895037.548         13    1289222695.196    685.565    0.0000
## Residual       1609729349.647        856       1880524.941
## Total         18369624387.195        869
## ---------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ----------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1219.090       547.610                -2.226    0.026    -2293.9
05    -144.274
##        DistanceFromHome      -16.653         5.763       -0.029    -2.890    0.004      -27.9
65      -5.342
##           EmployeeNumber        0.070         0.077        0.009     0.905    0.366       -0.0
82       0.221
```

```
## EnvironmentSatisfaction      -69.113        42.492       -0.017    -1.626    0.104     -152.5
14      14.288
##              HourlyRate        1.570         2.330        0.007     0.674    0.501       -3.0
04       6.144
##          JobInvolvement       98.801        66.534        0.015     1.485    0.138      -31.7
87     229.389
##               JobLevel     3730.902        68.914        0.885    54.139    0.000     3595.6
42    3866.161
##             MonthlyRate       -0.005         0.007       -0.008    -0.826    0.409       -0.0
18       0.008
##       NumCompaniesWorked      -13.702        21.106       -0.008    -0.649    0.516      -55.1
28      27.724
##        PercentSalaryHike       34.171        20.128        0.027     1.698    0.090       -5.3
35      73.678
##        PerformanceRating     -321.098       205.518       -0.025    -1.562    0.119     -724.4
76      82.281
##        TotalWorkingYears       72.364        12.021        0.118     6.020    0.000       48.7
69      95.959
##           YearsAtCompany      -10.399        14.632       -0.014    -0.711    0.477      -39.1
17      18.320
##     YearsWithCurrManager      -52.971        20.320       -0.041    -2.607    0.009      -92.8
53     -13.088
## ----------------------------------------------------------------------------------------
---------------
##
##
## - NumCompaniesWorked
##
## Backward Elimination: Step 11
##
##  Variable NumCompaniesWorked Removed
##
##                            Model Summary
## ------------------------------------------------------------------
## R                       0.955       RMSE                  1370.859
## R-Squared               0.912       Coef. Var               21.452
## Adj. R-Squared          0.911       MSE                1879255.452
## Pred R-Squared          0.909       MAE                   1032.368
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ---------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF       Mean Square       F         Sig.
## ---------------------------------------------------------------------------
## Regression    16759102465.163         12    1396591872.097    743.162    0.0000
## Residual       1610521922.032        857       1879255.452
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                       Parameter Estimates
```

```
## ----------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ----------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1260.589       543.683                -2.319    0.021    -2327.6
94    -193.483
##        DistanceFromHome      -16.404         5.748       -0.029    -2.854    0.004      -27.6
86      -5.121
##           EmployeeNumber        0.069         0.077        0.009     0.899    0.369       -0.0
82       0.220
## EnvironmentSatisfaction      -69.666        42.469       -0.017    -1.640    0.101     -153.0
23      13.690
##              HourlyRate        1.573         2.330        0.007     0.675    0.500       -3.0
00       6.145
##          JobInvolvement       99.777        66.494        0.015     1.501    0.134      -30.7
34     230.287
##               JobLevel     3734.723        68.639        0.886    54.411    0.000     3600.0
03    3869.442
##             MonthlyRate       -0.005         0.007       -0.008    -0.808    0.419       -0.0
18       0.008
##       PercentSalaryHike       34.141        20.121        0.027     1.697    0.090       -5.3
52      73.634
##       PerformanceRating     -319.214       205.428       -0.025    -1.554    0.121     -722.4
16      83.988
##       TotalWorkingYears       69.234        11.008        0.113     6.289    0.000       47.6
27      90.841
##          YearsAtCompany       -7.442        13.900       -0.010    -0.535    0.593      -34.7
25      19.840
##    YearsWithCurrManager      -52.974        20.313       -0.041    -2.608    0.009      -92.8
43     -13.105
## ----------------------------------------------------------------------------------------------
---------------
##
##
## - YearsAtCompany
##
## Backward Elimination: Step 12
##
##  Variable YearsAtCompany Removed
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1370.289
## R-Squared               0.912       Coef. Var              21.443
## Adj. R-Squared          0.911       MSE                 1877693.056
## Pred R-Squared          0.910       MAE                    1032.516
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
```

```
## ----------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F         Sig.
## ----------------------------------------------------------------------------
## Regression    16758563744.741         11    1523505794.976    811.371    0.0000
## Residual       1611060642.454        858       1877693.056
## Total         18369624387.195        869
## ----------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ------------------------------------------------------------------------------------------
---------------
##                    model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ------------------------------------------------------------------------------------------
---------------
##              (Intercept)    -1269.286       543.214                -2.337    0.020    -2335.4
70    -203.102
##         DistanceFromHome      -16.335         5.745       -0.029    -2.844    0.005      -27.6
10      -5.060
##          EmployeeNumber        0.069         0.077        0.009     0.896    0.370       -0.0
82       0.220
## EnvironmentSatisfaction      -69.918        42.449       -0.017    -1.647    0.100     -153.2
35      13.398
##               HourlyRate        1.581         2.329        0.007     0.679    0.497       -2.9
90       6.152
##          JobInvolvement      102.517        66.269        0.016     1.547    0.122      -27.5
52     232.586
##                JobLevel     3732.532        68.488        0.885    54.499    0.000     3598.1
08    3866.956
##             MonthlyRate       -0.005         0.007       -0.008    -0.781    0.435       -0.0
18       0.008
##       PercentSalaryHike       34.124        20.113        0.027     1.697    0.090       -5.3
52      73.601
##       PerformanceRating     -318.150       205.333       -0.025    -1.549    0.122     -721.1
65      84.864
##       TotalWorkingYears       67.328        10.413        0.110     6.466    0.000       46.8
91      87.765
##    YearsWithCurrManager      -60.469        14.712       -0.047    -4.110    0.000      -89.3
45     -31.594
## ------------------------------------------------------------------------------------------
---------------
##
##
## - HourlyRate
##
## Backward Elimination: Step 13
##
##  Variable HourlyRate Removed
##
##                             Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                  1369.859
## R-Squared               0.912       Coef. Var               21.437
```

```
## Adj. R-Squared          0.911       MSE                1876514.800
## Pred R-Squared          0.910       MAE                   1032.003
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                       Sum of
##                      Squares          DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16757698174.311         10    1675769817.431    893.022    0.0000
## Residual       1611926212.885        859       1876514.800
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1179.517       526.712                -2.239    0.025    -2213.3
11    -145.723
##        DistanceFromHome      -16.065         5.729       -0.028    -2.804    0.005      -27.3
10      -4.821
##           EmployeeNumber        0.070         0.077        0.009     0.903    0.367       -0.0
82       0.221
## EnvironmentSatisfaction      -70.764        42.418       -0.017    -1.668    0.096     -154.0
18      12.490
##           JobInvolvement      105.736        66.079        0.016     1.600    0.110      -23.9
59     235.431
##                JobLevel     3730.170        68.378        0.884    54.552    0.000     3595.9
62    3864.378
##             MonthlyRate       -0.005         0.007       -0.008    -0.791    0.429       -0.0
18       0.008
##       PercentSalaryHike       33.529        20.088        0.027     1.669    0.095       -5.8
98      72.955
##       PerformanceRating     -313.325       205.146       -0.024    -1.527    0.127     -715.9
71      89.321
##       TotalWorkingYears       67.760        10.390        0.111     6.522    0.000       47.3
67      88.152
##    YearsWithCurrManager      -60.678        14.704       -0.047    -4.127    0.000      -89.5
38     -31.818
## ------------------------------------------------------------------------------------------
---------------
##
##
## - MonthlyRate
##
## Backward Elimination: Step 14
##
```

```
##  Variable MonthlyRate Removed
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1369.561
## R-Squared               0.912       Coef. Var              21.432
## Adj. R-Squared          0.911       MSE                1875696.593
## Pred R-Squared          0.910       MAE                   1031.568
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F         Sig.
## --------------------------------------------------------------------------------
## Regression    16756525316.870          9    1861836146.319    992.611    0.0000
## Residual       1613099070.326        860       1875696.593
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ----------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1246.142       519.813                -2.397    0.017    -2266.3
93    -225.890
##        DistanceFromHome      -16.025         5.728       -0.028    -2.798    0.005      -27.2
67      -4.784
##          EmployeeNumber        0.068         0.077        0.009     0.878    0.380       -0.0
83       0.219
## EnvironmentSatisfaction      -72.686        42.339       -0.017    -1.717    0.086     -155.7
85      10.413
##          JobInvolvement      106.586        66.056        0.016     1.614    0.107      -23.0
63     236.235
##               JobLevel     3727.682        68.291        0.884    54.585    0.000     3593.6
46    3861.718
##       PercentSalaryHike       33.388        20.082        0.027     1.663    0.097       -6.0
28      72.805
##       PerformanceRating     -312.416       205.098       -0.024    -1.523    0.128     -714.9
68      90.135
##       TotalWorkingYears       67.616        10.386        0.110     6.510    0.000       47.2
31      88.001
##    YearsWithCurrManager      -60.052        14.679       -0.047    -4.091    0.000      -88.8
64     -31.241
## ----------------------------------------------------------------------------------------------
---------------
##
##
```

```
## - EmployeeNumber
##
## Backward Elimination: Step 15
##
##  Variable EmployeeNumber Removed
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                1369.379
## R-Squared               0.912       Coef. Var             21.429
## Adj. R-Squared          0.911       MSE                1875197.506
## Pred R-Squared          0.910       MAE                   1030.211
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16755079334.790          8    2094384916.849    1116.888    0.0000
## Residual       1614545052.405        861       1875197.506
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
## ----------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1177.303       513.798                -2.291    0.022    -2185.7
46    -168.859
##        DistanceFromHome      -16.035         5.727       -0.028    -2.800    0.005      -27.2
75      -4.795
## EnvironmentSatisfaction      -71.645        42.316       -0.017    -1.693    0.091     -154.7
01      11.410
##          JobInvolvement      106.627        66.047        0.016     1.614    0.107      -23.0
04     236.259
##               JobLevel     3728.472        68.276        0.884    54.609    0.000     3594.4
65    3862.479
##       PercentSalaryHike       32.999        20.075        0.026     1.644    0.101       -6.4
03      72.400
##       PerformanceRating     -311.914       205.070       -0.024    -1.521    0.129     -714.4
10      90.581
##       TotalWorkingYears       67.612        10.385        0.110     6.511    0.000       47.2
29      87.994
##    YearsWithCurrManager      -59.891        14.676       -0.047    -4.081    0.000      -88.6
97     -31.086
## ----------------------------------------------------------------------------------------------
---------------
```

```
##
##
##
## No more variables satisfy the condition of p value = 0.3
##
##
## Variables Removed:
##
## - StockOptionLevel
## - Education
## - JobSatisfaction
## - YearsSinceLastPromotion
## - YearsInCurrentRole
## - RelationshipSatisfaction
## - DailyRate
## - Age
## - TrainingTimesLastYear
## - WorkLifeBalance
## - NumCompaniesWorked
## - YearsAtCompany
## - HourlyRate
## - MonthlyRate
## - EmployeeNumber
##
##
## Final Model Output
## ------------------
##
##                             Model Summary
## ----------------------------------------------------------------------
## R                       0.955       RMSE                   1369.379
## R-Squared               0.912       Coef. Var                21.429
## Adj. R-Squared          0.911       MSE                 1875197.506
## Pred R-Squared          0.910       MAE                    1030.211
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16755079334.790          8    2094384916.849    1116.888    0.0000
## Residual       1614545052.405        861       1875197.506
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------
---------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er       upper
```

```
## ------------------------------------------------------------------------------------------
---------------
##             (Intercept)    -1177.303       513.798               -2.291    0.022    -2185.7
46    -168.859
##        DistanceFromHome      -16.035         5.727       -0.028    -2.800    0.005      -27.2
75      -4.795
## EnvironmentSatisfaction      -71.645        42.316       -0.017    -1.693    0.091     -154.7
01      11.410
##          JobInvolvement      106.627        66.047        0.016     1.614    0.107      -23.0
04     236.259
##               JobLevel     3728.472        68.276        0.884    54.609    0.000     3594.4
65    3862.479
##       PercentSalaryHike       32.999        20.075        0.026     1.644    0.101       -6.4
03      72.400
##       PerformanceRating     -311.914       205.070       -0.024    -1.521    0.129     -714.4
10      90.581
##       TotalWorkingYears       67.612        10.385        0.110     6.511    0.000       47.2
29      87.994
##    YearsWithCurrManager      -59.891        14.676       -0.047    -4.081    0.000      -88.6
97     -31.086
## ------------------------------------------------------------------------------------------
---------------
```

```
Q2_fit_stepwise <- ols_step_both_p(fit1, penter = 0.05, details = TRUE)
```

```
## Stepwise Selection Method
## --------------------------
##
## Candidate Terms:
##
## 1. Age
## 2. DailyRate
## 3. DistanceFromHome
## 4. Education
## 5. EmployeeNumber
## 6. EnvironmentSatisfaction
## 7. HourlyRate
## 8. JobInvolvement
## 9. JobLevel
## 10. JobSatisfaction
## 11. MonthlyRate
## 12. NumCompaniesWorked
## 13. PercentSalaryHike
## 14. PerformanceRating
## 15. RelationshipSatisfaction
## 16. StockOptionLevel
## 17. TotalWorkingYears
## 18. TrainingTimesLastYear
## 19. WorkLifeBalance
## 20. YearsAtCompany
## 21. YearsInCurrentRole
## 22. YearsSinceLastPromotion
## 23. YearsWithCurrManager
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - JobLevel added
##
##                            Model Summary
## ----------------------------------------------------------------------
## R                       0.952       RMSE                   1413.296
## R-Squared               0.906       Coef. Var                22.116
## Adj. R-Squared          0.906       MSE                 1997404.971
## Pred R-Squared          0.905       MAE                    1073.683
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF        Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16635876872.790          1    16635876872.790    8328.745    0.0000
```

```
## Residual       1733747514.405        868        1997404.971
## Total         18369624387.195        869
## ----------------------------------------------------------------------------
##
##                                     Parameter Estimates
## ------------------------------------------------------------------------------------------
-----
##       model         Beta    Std. Error    Std. Beta       t        Sig         lower        u
pper
## ------------------------------------------------------------------------------------------
-----
## (Intercept)    -1793.934       101.676                -17.644    0.000    -1993.494    -159
4.375
##    JobLevel     4013.671        43.980        0.952     91.262    0.000     3927.352     409
9.990
## ------------------------------------------------------------------------------------------
-----
##
##
##
## Stepwise Selection: Step 2
##
## - TotalWorkingYears added
##
##                            Model Summary
## ---------------------------------------------------------------------
## R                       0.953       RMSE                  1389.696
## R-Squared               0.909       Coef. Var               21.747
## Adj. R-Squared          0.909       MSE                  1931256.053
## Pred R-Squared          0.908       MAE                     1054.184
## ---------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## ---------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## ---------------------------------------------------------------------------
## Regression    16695225388.963          2    8347612694.481    4322.375    0.0000
## Residual       1674398998.232        867       1931256.053
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                         Parameter Estimates
## ------------------------------------------------------------------------------------------
-----------
##             model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ------------------------------------------------------------------------------------------
-----------
##       (Intercept)    -1798.376        99.982                -17.987    0.000    -1994.610
-1602.142
```

```
##          JobLevel     3714.122        69.210         0.881     53.664    0.000     3578.283
3849.961
## TotalWorkingYears       55.664        10.041         0.091      5.544    0.000       35.956
75.372
## ----------------------------------------------------------------------------------------
-----------
##
##
##
##                             Model Summary
## ------------------------------------------------------------------
## R                       0.953       RMSE                  1389.696
## R-Squared               0.909       Coef. Var               21.747
## Adj. R-Squared          0.909       MSE                1931256.053
## Pred R-Squared          0.908       MAE                   1054.184
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                       ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16695225388.963          2    8347612694.481    4322.375    0.0000
## Residual       1674398998.232        867       1931256.053
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------
-----------
##             model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ----------------------------------------------------------------------------------------
-----------
##       (Intercept)    -1798.376        99.982                -17.987    0.000    -1994.610
-1602.142
##          JobLevel     3714.122        69.210         0.881     53.664    0.000     3578.283
3849.961
## TotalWorkingYears       55.664        10.041         0.091      5.544    0.000       35.956
75.372
## ----------------------------------------------------------------------------------------
-----------
##
##
##
## Stepwise Selection: Step 3
##
## - YearsWithCurrManager added
##
##                             Model Summary
## ------------------------------------------------------------------
```

```
## R                       0.954       RMSE                   1377.669
## R-Squared               0.911       Coef. Var                21.559
## Adj. R-Squared          0.910       MSE                 1897970.749
## Pred R-Squared          0.910       MAE                    1035.798
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16725981718.483          3    5575327239.494     2937.52    0.0000
## Residual       1643642668.712        866       1897970.749
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------------
--------------
##                model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ----------------------------------------------------------------------------------------------------
--------------
##          (Intercept)    -1699.158       102.135                -16.636    0.000    -1899.618
-1498.697
##             JobLevel     3717.242        68.615        0.881     54.175    0.000     3582.570
3851.914
##    TotalWorkingYears       68.336        10.440        0.112      6.545    0.000       47.845
88.827
## YearsWithCurrManager      -59.331        14.739       -0.046     -4.026    0.000      -88.259
-30.403
## ----------------------------------------------------------------------------------------------------
--------------
##
##
##
##                            Model Summary
## ----------------------------------------------------------------------
## R                       0.954       RMSE                   1377.669
## R-Squared               0.911       Coef. Var                21.559
## Adj. R-Squared          0.910       MSE                 1897970.749
## Pred R-Squared          0.910       MAE                    1035.798
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
```

```
## ----------------------------------------------------------------------------
## Regression    16725981718.483          3    5575327239.494     2937.52    0.0000
## Residual       1643642668.712        866       1897970.749
## Total         18369624387.195        869
## ----------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
--------------
##                model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ----------------------------------------------------------------------------------------------
--------------
##          (Intercept)    -1699.158       102.135                -16.636    0.000    -1899.618
-1498.697
##             JobLevel     3717.242        68.615        0.881     54.175    0.000     3582.570
3851.914
##    TotalWorkingYears       68.336        10.440        0.112      6.545    0.000       47.845
88.827
## YearsWithCurrManager      -59.331        14.739       -0.046     -4.026    0.000      -88.259
-30.403
## ----------------------------------------------------------------------------------------------
--------------
##
##
##
## Stepwise Selection: Step 4
##
## - DistanceFromHome added
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1372.788
## R-Squared               0.911       Coef. Var              21.482
## Adj. R-Squared          0.911       MSE                 1884546.340
## Pred R-Squared          0.910       MAE                    1032.363
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ----------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## ----------------------------------------------------------------------------
## Regression    16739491803.372          4    4184872950.843    2220.626    0.0000
## Residual       1630132583.824        865       1884546.340
## Total         18369624387.195        869
## ----------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
--------------
```

```
##                model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ------------------------------------------------------------------------------------------
--------------
##          (Intercept)    -1559.495       114.362                -13.637    0.000    -1783.954
-1335.037
##              JobLevel     3722.700        68.403        0.883     54.423    0.000     3588.445
3856.955
##    TotalWorkingYears       67.981        10.404        0.111      6.534    0.000       47.561
88.402
## YearsWithCurrManager      -60.215        14.690       -0.047     -4.099    0.000      -89.047
-31.382
##     DistanceFromHome      -15.335         5.727       -0.027     -2.677    0.008      -26.576
-4.094
## ------------------------------------------------------------------------------------------
--------------
##
##
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1372.788
## R-Squared               0.911       Coef. Var              21.482
## Adj. R-Squared          0.911       MSE                 1884546.340
## Pred R-Squared          0.910       MAE                    1032.363
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ---------------------------------------------------------------------------
##                       Sum of
##                      Squares         DF       Mean Square       F          Sig.
## ---------------------------------------------------------------------------
## Regression    16739491803.372          4    4184872950.843    2220.626    0.0000
## Residual       1630132583.824        865       1884546.340
## Total         18369624387.195        869
## ---------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ------------------------------------------------------------------------------------------
--------------
##                model         Beta    Std. Error    Std. Beta       t        Sig         lower
upper
## ------------------------------------------------------------------------------------------
--------------
##          (Intercept)    -1559.495       114.362                -13.637    0.000    -1783.954
-1335.037
##              JobLevel     3722.700        68.403        0.883     54.423    0.000     3588.445
3856.955
##    TotalWorkingYears       67.981        10.404        0.111      6.534    0.000       47.561
88.402
## YearsWithCurrManager      -60.215        14.690       -0.047     -4.099    0.000      -89.047
```

```
-31.382
##     DistanceFromHome      -15.335         5.727       -0.027     -2.677    0.008      -26.576
-4.094
## ----------------------------------------------------------------------------------------------
--------------
##
##
##
## Stepwise Selection: Step 5
##
## - EnvironmentSatisfaction added
##
##                            Model Summary
## ---------------------------------------------------------------------
## R                       0.955       RMSE                  1371.354
## R-Squared               0.912       Coef. Var               21.460
## Adj. R-Squared          0.911       MSE                1880610.661
## Pred R-Squared          0.910       MAE                   1030.966
## ---------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16744776775.699          5    3348955355.140    1780.781    0.0000
## Residual       1624847611.496        864       1880610.661
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                            Parameter Estimates
## ----------------------------------------------------------------------------------------------
----------------
##                    model         Beta    Std. Error    Std. Beta      t        Sig         low
er        upper
## ----------------------------------------------------------------------------------------------
----------------
##             (Intercept)    -1363.449       163.486                -8.340    0.000    -1684.3
26    -1042.571
##                JobLevel     3725.745        68.355        0.883    54.505    0.000     3591.5
82     3859.907
##       TotalWorkingYears       67.523        10.397        0.110     6.495    0.000       47.1
17       87.929
##    YearsWithCurrManager      -60.635        14.677       -0.047    -4.131    0.000      -89.4
42      -31.828
##        DistanceFromHome      -15.716         5.726       -0.028    -2.745    0.006      -26.9
55       -4.478
## EnvironmentSatisfaction      -71.038        42.376       -0.017    -1.676    0.094     -154.2
09       12.134
## ----------------------------------------------------------------------------------------------
----------------
```

```
##
##
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1371.354
## R-Squared               0.912       Coef. Var              21.460
## Adj. R-Squared          0.911       MSE                 1880610.661
## Pred R-Squared          0.910       MAE                    1030.966
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares          DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16744776775.699           5    3348955355.140    1780.781    0.0000
## Residual       1624847611.496         864       1880610.661
## Total         18369624387.195         869
## --------------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------
----------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er        upper
## ----------------------------------------------------------------------------------------------
----------------
##             (Intercept)    -1363.449       163.486                -8.340    0.000    -1684.3
26    -1042.571
##                JobLevel     3725.745        68.355        0.883    54.505    0.000     3591.5
82     3859.907
##       TotalWorkingYears       67.523        10.397        0.110     6.495    0.000       47.1
17       87.929
##    YearsWithCurrManager      -60.635        14.677       -0.047    -4.131    0.000      -89.4
42      -31.828
##        DistanceFromHome      -15.716         5.726       -0.028    -2.745    0.006      -26.9
55       -4.478
## EnvironmentSatisfaction      -71.038        42.376       -0.017    -1.676    0.094     -154.2
09       12.134
## ----------------------------------------------------------------------------------------------
----------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## ------------------
##
```

```
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.955       RMSE                 1371.354
## R-Squared               0.912       Coef. Var              21.460
## Adj. R-Squared          0.911       MSE                 1880610.661
## Pred R-Squared          0.910       MAE                    1030.966
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## --------------------------------------------------------------------------------
##                        Sum of
##                       Squares         DF       Mean Square       F          Sig.
## --------------------------------------------------------------------------------
## Regression    16744776775.699          5    3348955355.140    1780.781    0.0000
## Residual       1624847611.496        864       1880610.661
## Total         18369624387.195        869
## --------------------------------------------------------------------------------
##
##                                            Parameter Estimates
## ----------------------------------------------------------------------------------------------
----------------
##                   model         Beta    Std. Error    Std. Beta      t        Sig         low
er        upper
## ----------------------------------------------------------------------------------------------
----------------
##             (Intercept)    -1363.449       163.486                -8.340    0.000    -1684.3
26    -1042.571
##                JobLevel     3725.745        68.355        0.883    54.505    0.000     3591.5
82     3859.907
##       TotalWorkingYears       67.523        10.397        0.110     6.495    0.000       47.1
17       87.929
##    YearsWithCurrManager      -60.635        14.677       -0.047    -4.131    0.000      -89.4
42      -31.828
##        DistanceFromHome      -15.716         5.726       -0.028    -2.745    0.006      -26.9
55       -4.478
## EnvironmentSatisfaction      -71.038        42.376       -0.017    -1.676    0.094     -154.2
09       12.134
## ----------------------------------------------------------------------------------------------
----------------
```

```
# Stepwise Variable Selection Model
fit4 <- lm(MonthlyIncome ~ JobLevel + TotalWorkingYears + YearsWithCurrManager + DistanceFromHom
e + EnvironmentSatisfaction, data = employee.dB)
summary(fit4)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ JobLevel + TotalWorkingYears + YearsWithCurrManager +
##     DistanceFromHome + EnvironmentSatisfaction, data = employee.dB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5899.1  -877.3    22.3   762.0  4045.0
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1363.449    163.486  -8.340 2.92e-16 ***
## JobLevel                 3725.745     68.355  54.505  < 2e-16 ***
## TotalWorkingYears          67.523     10.397   6.495 1.40e-10 ***
## YearsWithCurrManager      -60.635     14.677  -4.131 3.96e-05 ***
## DistanceFromHome          -15.716      5.726  -2.745  0.00618 **
## EnvironmentSatisfaction   -71.038     42.376  -1.676  0.09403 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1371 on 864 degrees of freedom
## Multiple R-squared:  0.9115, Adjusted R-squared:  0.911
## F-statistic:  1781 on 5 and 864 DF,  p-value: < 2.2e-16
```

```
vif(fit4)
```

```
##               JobLevel       TotalWorkingYears    YearsWithCurrManager
##                 2.5657                  2.8199                  1.2718
##        DistanceFromHome EnvironmentSatisfaction
##                 1.0030                  1.0029
```

```
AIC(fit4)
```

```
## [1] 15045.92
```

```
BIC(fit4)
```

```
## [1] 15079.29
```

```
press(fit4)
```

```
## [1] 1650612731
```

```
################ PREFERRED MLR MODEL TEST & TRAIN ################

# Build test & train
set.seed(1236)
splitPerc <- 0.70 #Use a 70/30 train to test ratio
trainIndicies <- sample( 1:dim(employee.dB)[1], round(splitPerc * dim(employee.dB)[1]))
train.mlr <- employee.dB[trainIndicies, ]
test.mlr  <- employee.dB[-trainIndicies, ]

# compare the dimention of splitted dataset
dim(employee.dB)
```

```
## [1] 870  83
```

```
dim(train.mlr)
```

```
## [1] 609  83
```

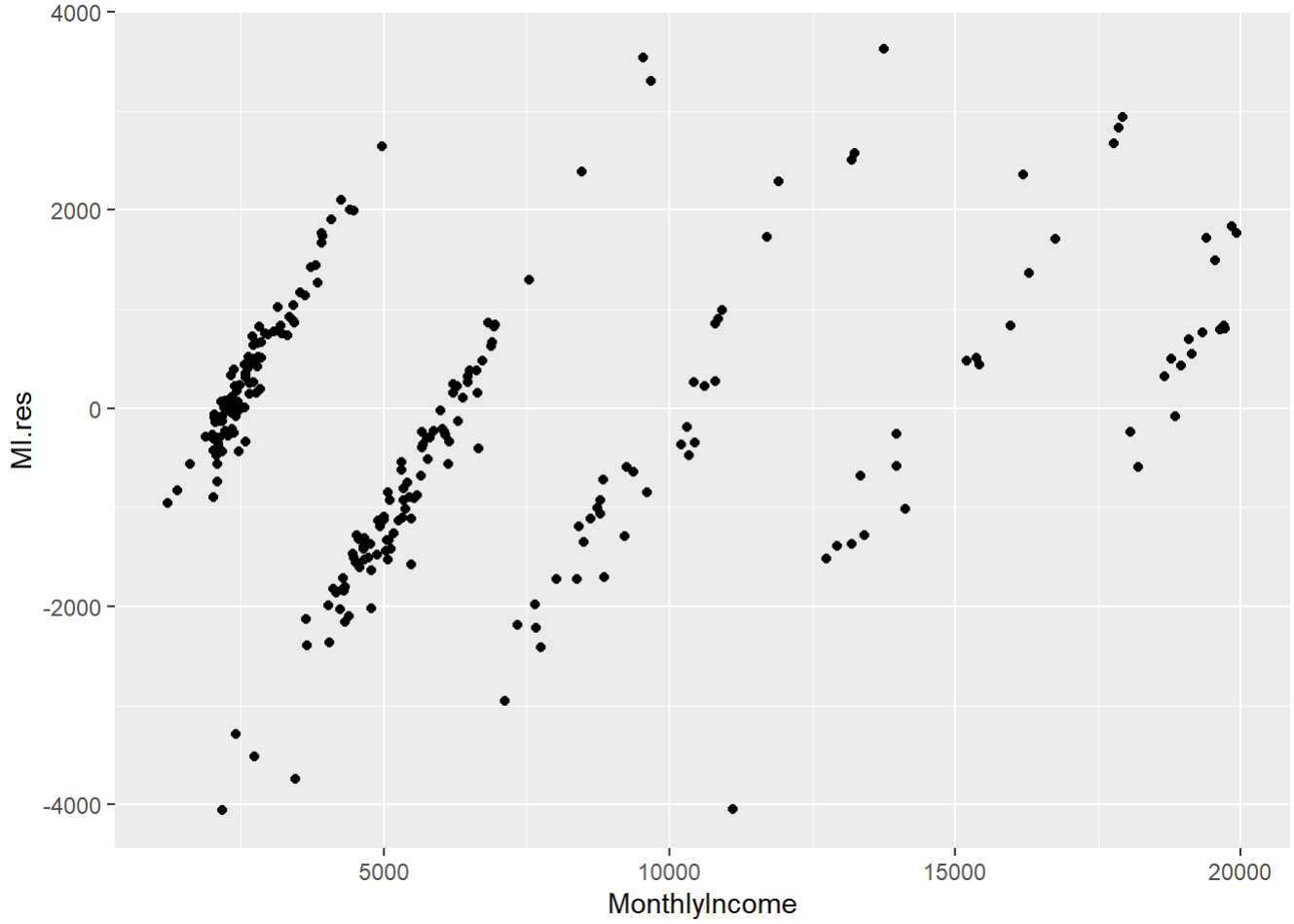
```
dim(test.mlr)
```

```
## [1] 261  83
```

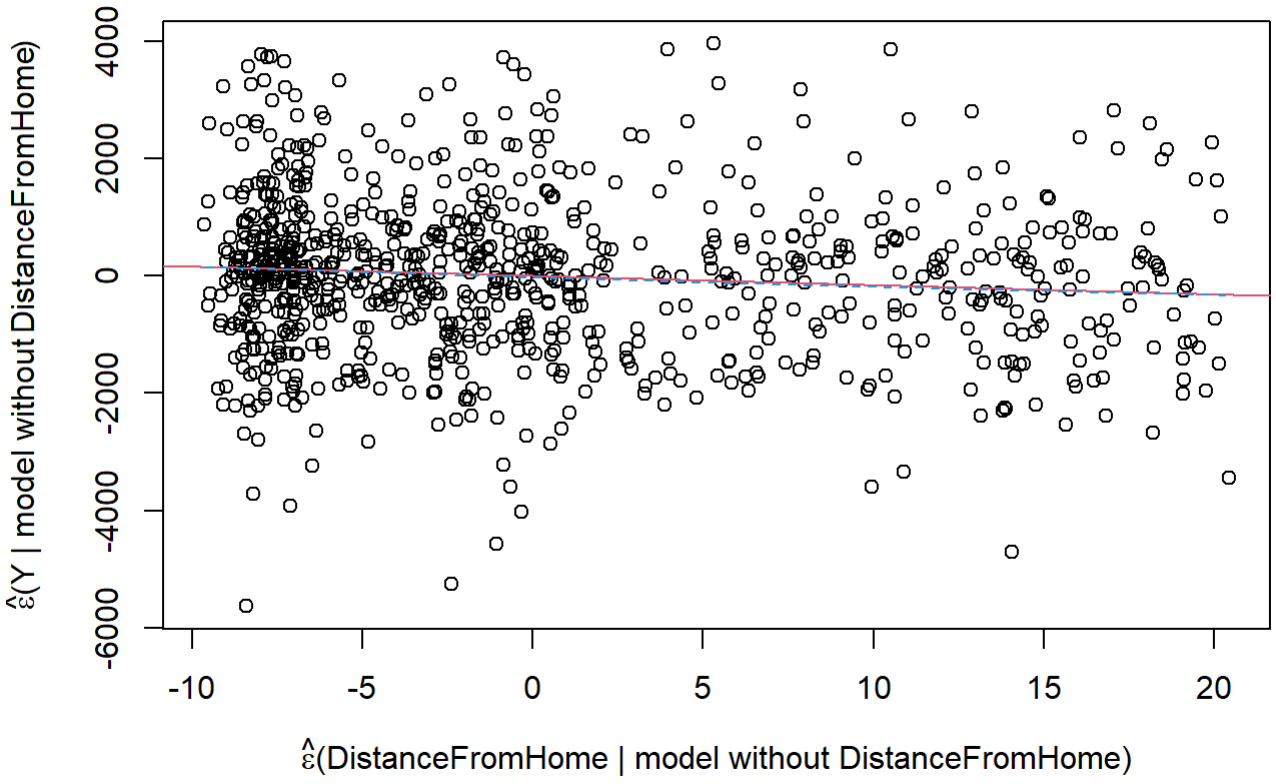
```
#fit.train <- lm(MonthlyIncome ~ DistanceFromHome + JobLevel + PercentSalaryHike + TotalWorkingY
ears + YearsWithCurrManager, data = train.mlr) # Original Preferred Model before trying Stepwise
fit.train <- lm(MonthlyIncome ~ JobLevel + TotalWorkingYears + YearsWithCurrManager + DistanceFr
omHome + EnvironmentSatisfaction, data = train.mlr)
summary(fit.train)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ JobLevel + TotalWorkingYears + YearsWithCurrManager +
##     DistanceFromHome + EnvironmentSatisfaction, data = train.mlr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5778.5  -893.4     7.9   728.5  3919.2
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1485.954    207.524  -7.160 2.35e-12 ***
## JobLevel                  3755.053     83.995  44.706  < 2e-16 ***
## TotalWorkingYears           61.447     12.793   4.803 1.97e-06 ***
## YearsWithCurrManager       -57.901     17.719  -3.268  0.00115 **
## DistanceFromHome           -11.763      6.955  -1.691  0.09128 .
## EnvironmentSatisfaction    -21.276     53.172  -0.400  0.68920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1409 on 603 degrees of freedom
## Multiple R-squared:  0.9017, Adjusted R-squared:  0.9009
## F-statistic:  1107 on 5 and 603 DF,  p-value: < 2.2e-16
```

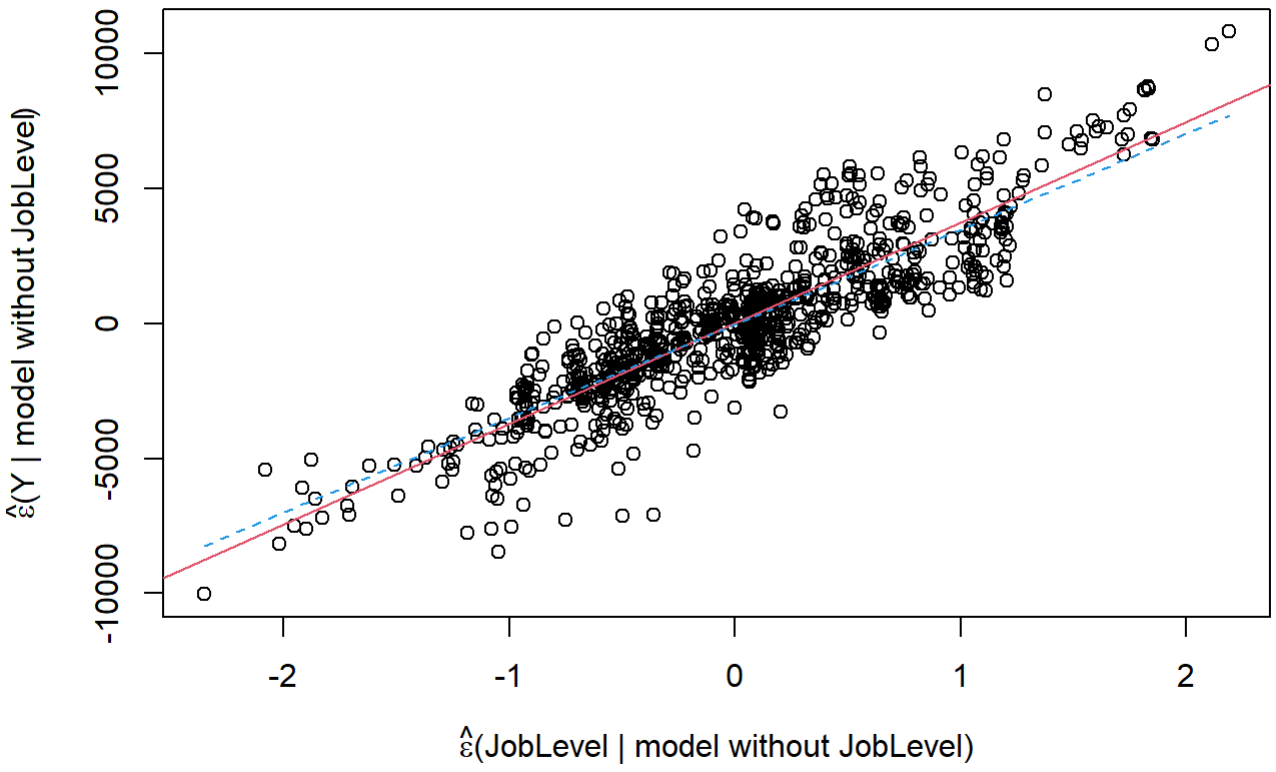
```
test.predict <- predict(fit.train, test.mlr, interval = "confidence")
test.mlr$fit <- test.predict[, "fit"]
test.mlr$MI.res <- test.mlr$MonthlyIncome - test.mlr$fit

#dev.new()
test.mlr %>% ggplot(aes(x = MonthlyIncome, y = MI.res))+
  geom_point()
```




```
partial.resid.plot(fit2, smooth.span = 2, lf.col = 2)
```

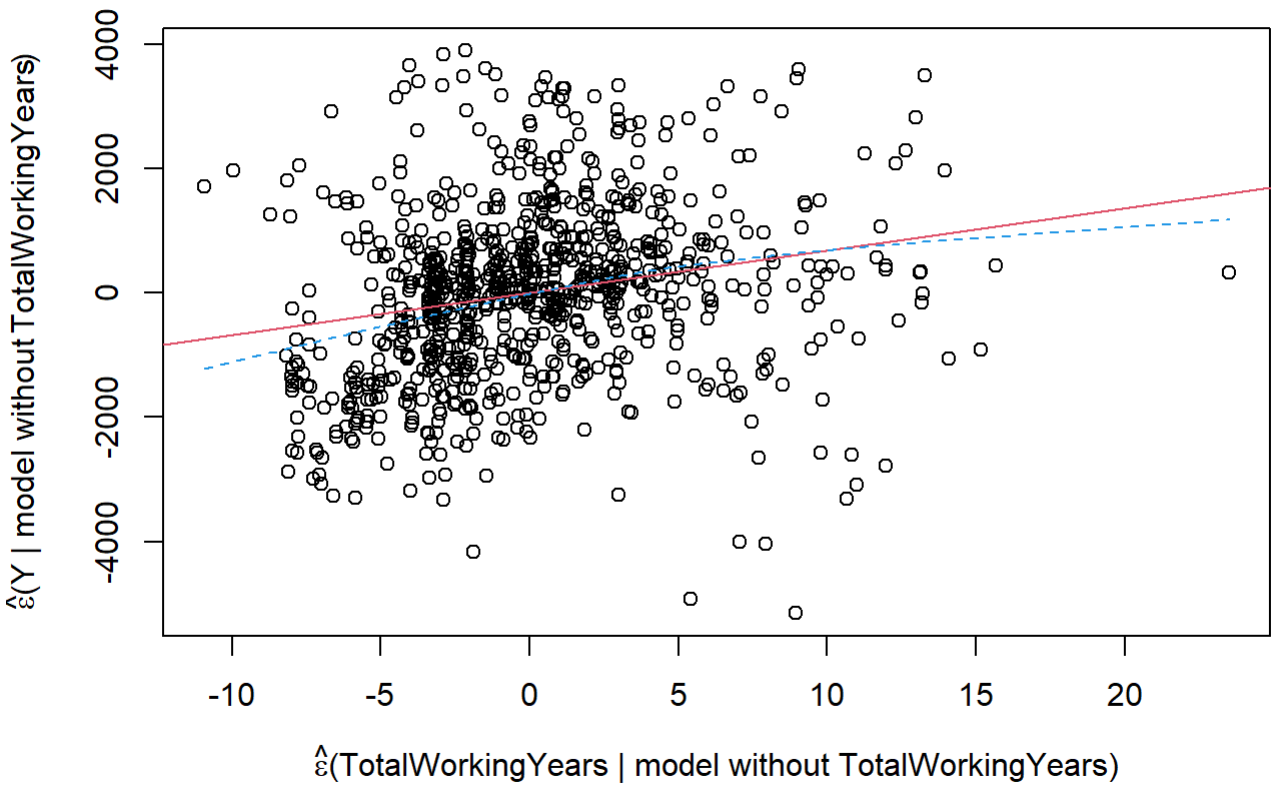



```
## Press return for next plot
```

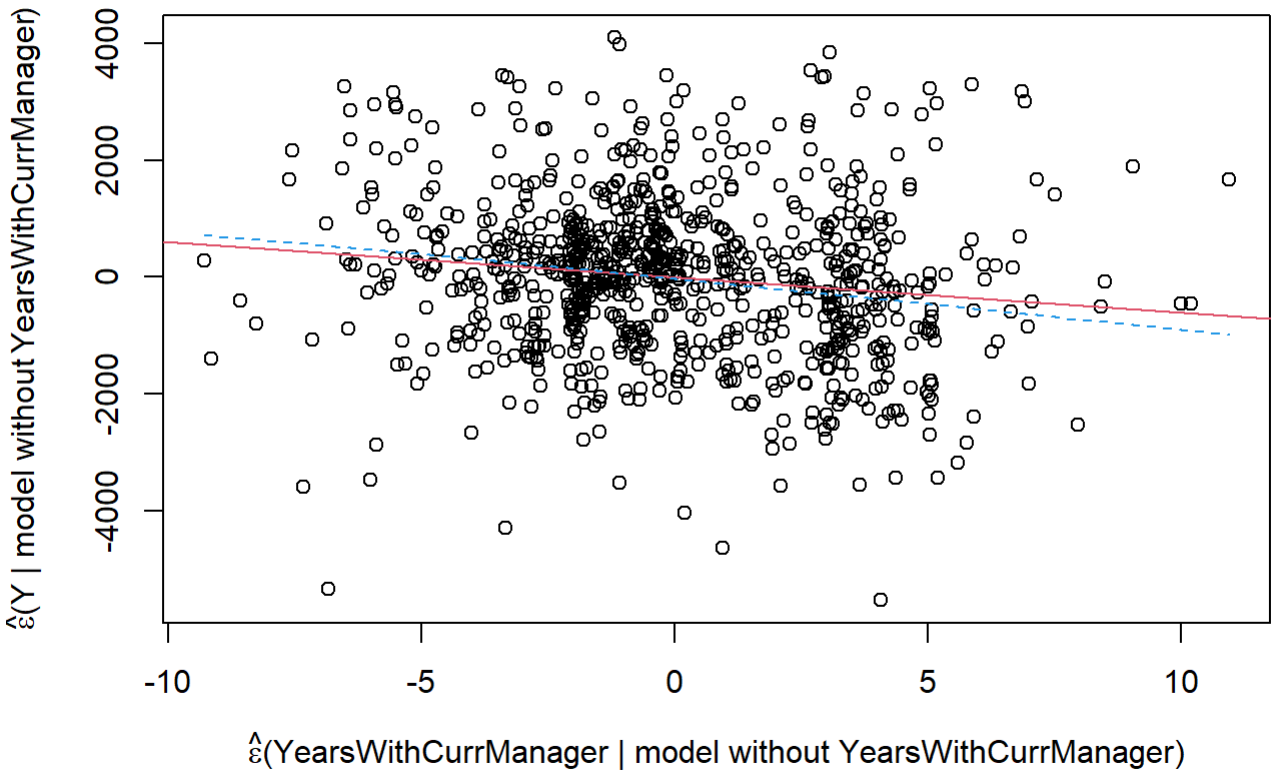



```
## Press return for next plot
```


```
## Press return for next plot
```



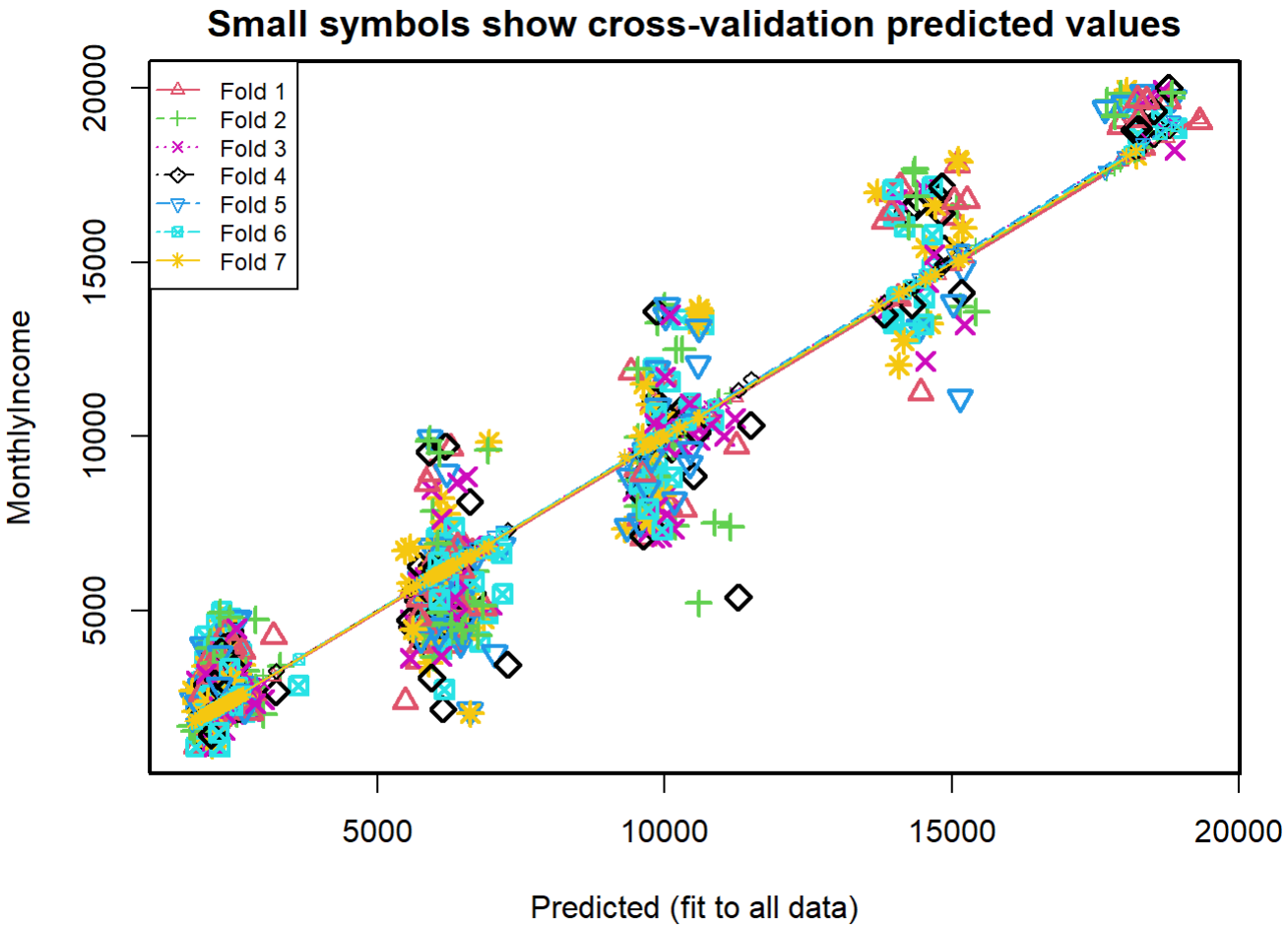

```
## Press return for next plot
```




```
## Press return for next plot
```

```
### INTERNAL CROSS-VALIDATION of GLM ###

fit2.glm <- glm(MonthlyIncome ~ JobLevel + TotalWorkingYears + YearsWithCurrManager + DistanceFr
omHome + EnvironmentSatisfaction, data = employee.dB)
cv.err <- cv.glm(employee.dB, fit2.glm)$delta
cv.err.2 <- cv.glm(employee.dB, fit2.glm, K = 2)$delta
cv.err.3 <- cv.glm(employee.dB, fit2.glm, K = 3)$delta
cv.err.6 <- cv.glm(employee.dB, fit2.glm, K = 6)$delta

fit2.cv.lm <- cv.lm(employee.dB, m = 7, plotit = "Observed", form.lm = formula(MonthlyIncome ~ J
obLevel + TotalWorkingYears + YearsWithCurrManager + DistanceFromHome + EnvironmentSatisfaction)
)
```

```
## Analysis of Variance Table
##
## Response: MonthlyIncome
##                           Df   Sum Sq  Mean Sq F value  Pr(>F)
## JobLevel                   1 1.66e+10 1.66e+10 8846.00 < 2e-16 ***
## TotalWorkingYears          1 5.93e+07 5.93e+07   31.56 2.6e-08 ***
## YearsWithCurrManager       1 3.08e+07 3.08e+07   16.35 5.7e-05 ***
## DistanceFromHome           1 1.35e+07 1.35e+07    7.18  0.0075 **
## EnvironmentSatisfaction    1 5.28e+06 5.28e+06    2.81  0.0940 .
## Residuals                864 1.62e+09 1.88e+06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
##
## fold 1
## Observations in test set: 124
##                  7   10     34   35   40    41    44   60    62    76   84   89
## Predicted     2432 5762 2313.2 6265 1916 15015 18400 1840  9623  6245 6234 9587
## cvpred        2431 5768 2329.8 6202 1912 14942 18222 1823  9603  6241 6224 9572
## MonthlyIncome 2127 5063 2258.0 9679 2791 16291 19613 1118  7094  4011 5974 8726
## CV residual   -304 -705  -71.8 3477  879  1349  1391 -705 -2509 -2230 -250 -846
##                  97  106   108  109   114   116   121   123  126  127   135
## Predicted     14772 2360 18764 5927  6853  9574  9989 10482 5894 2220  6325
## cvpred        14623 2401 18567 5906  6755  9586  9965 10376 5917 2241  6261
## MonthlyIncome 16606 4420 19636 8926  5093  7642 11713 10761 5204 3919  5257
## CV residual    1983 2019  1069 3020 -1662 -1944  1748   385 -713 1678 -1004
##                 138  147  154   168  182  208   212   218  232   235   237  243
## Predicted     10334 2034 6489 13825 6233 6638 18321 14106 2060 13878 17934 2280
## cvpred        10273 2045 6443 13787 6238 6536 18148 14001 2062 13824 17907 2290
## MonthlyIncome 10739 2756 6870 16184 6893 6673 18303 17174 2014 16959 18880 2703
## CV residual     466  711  427  2397  655  137   155  3173  -48  3135   973  413
##                 256  263  266   273  291    298  299  302  309  310  315   327
## Predicted     2501 2227 2263  5707 2481 6150.9 2157 9511 6315 5827 6319 17960
## cvpred        2494 2223 2268  5724 2456 6158.9 2152 9513 6283 5800 6279 17890
## MonthlyIncome 2438 2436 2426  3540 2270 6074.0 2321 9069 6385 4851 5405 19049
## CV residual    -56  213  158 -2184 -186  -84.9  169 -444  102 -949 -874  1159
##                 355  382  403   404  428   431  438  448   449  458  459  474
## Predicted     10788 1872 2551 10080 2234  5868 2310 6417  5478 6288 2613 2119
## cvpred        10696 1894 2519 10026 2210  5859 2331 6352  5487 6210 2598 2095
## MonthlyIncome 10447 2942 2105  8008 3230  4478 2867 5538  2406 5441 2066 3452
## CV residual    -249 1048 -414 -2018 1020 -1381  536 -814 -3081 -769 -532 1357
##                   475  477  481  484  493  503  507   512  517   523    535
## Predicted     2204.33 6160 6285 3185 6015 2827 2067  9782 1918 11250 2142.7
## cvpred        2208.33 6088 6268 3114 5996 2801 2052  9728 1903 11103 2180.9
## MonthlyIncome 2207.00 6334 5373 4257 5869 2655 2851 10920 2413  9724 2119.0
## CV residual     -1.33  246 -895 1143 -127 -146  799  1192  510 -1379  -61.9
##                 538  539  545  546  550  561    566     570  578  589   592  600
## Predicted     2388 2438 2341 2172 2006 2788 2488.4 14060.5 2361 2537  6114 2610
## cvpred        2366 2458 2340 2193 2012 2739 2483.1 13944.1 2355 2525  6125 2584
## MonthlyIncome 2696 3202 4381 2929 2177 2088 2436.0 13973.0 3162 3629  4599 2293
## CV residual    330  744 2041  736  165 -651  -47.1    28.9  807 1104 -1526 -291
##                 601  608  617   622   642   643  645  647  652  660   667   670
## Predicted     1941 6023 2647  6549  6346 15267 2550 2023 6101 5996 10338 15093
## cvpred        1923 6045 2637  6504  6330 15093 2533 2017 6120 5981 10267 14962
## MonthlyIncome 2033 5605 3838  5067  4448 16756 2107 1904 5220 6499  7918 17779
## CV residual    110 -440 1201 -1437 -1882  1663 -426 -113 -900  518 -2349  2817
##                 691   692  696   700  709  722  725   738  741   753   757   762
## Predicted     6272  9416 2128 18252 2079 6063 2493  6402 5861 14456  5725 15051
## cvpred        6223  9425 2125 18088 2086 6015 2445  6373 5834 14387  5717 14879
## MonthlyIncome 6077 11849 3294 19081 3702 5768 3968  5171 8639 11245  4014 16752
## CV residual   -146  2424 1169   993 1616 -247 1523 -1202 2805 -3142 -1703  1873
##                 763  774  781   786     789  799  800  802  809  810  824   826
## Predicted     6455 2072 2055 13950 19316.5 2202 6392 6287 2159 2227 1927 18397
## cvpred        6437 2051 2052 13864 19092.5 2201 6347 6276 2191 2200 1929 18237
## MonthlyIncome 6151 3388 3140 16422 19038.0 1859 6842 6582 2862 3058 2610 19627
## CV residual   -286 1337 1088  2558   -54.5 -342  495  306  671  858  681  1390
```

```
##                838   840   846  854  855  858
## Predicted     5725 18236  5798 5878 5701 9620
## cvpred        5717 18117  5807 5887 5672 9626
## MonthlyIncome 5337 19665  4558 5661 4788 8943
## CV residual   -380  1548 -1249 -226 -884 -683
##
## Sum of squares = 2.19e+08    Mean square = 1766391    n = 124
##
## fold 2
## Observations in test set: 125
##                    2    9   15    26   27    31      32   38   46    52    59
## Predicted     17704 2060 2313 18194 5906  5846 2047.24 6701 2475 18823 10619
## cvpred        17543 2051 2320 18092 5849  5775 2034.71 6738 2520 18775 10629
## MonthlyIncome 19626 2220 3423 18213 9924  4668 2028.00 6134 3673 19701 13603
## CV residual    2083  169 1103   121 4075 -1107   -6.71 -604 1153   926  2974
##                  69   71   73    78    86   93   99  100   105   107   125
## Predicted     2421 2873 6064 14335 15139 2519 2196 2414 15102  5918  6044
## cvpred        2457 2942 6034 14258 15160 2552 2193 2434 15108  5854  6009
## MonthlyIncome 2194 2782 5916 17567 13726 2090 3180 3294 16598  4319  4554
## CV residual   -263 -160 -118  3309 -1434 -462  987  860  1490 -1535 -1455
##                  139   140   141   146   151  152  153  155  162  164  178  183
## Predicted     2402.8 10182 14576  5921  6228 5837 6087 2131 9424 6091 2544 2874
## cvpred        2421.5 10130 14524  5863  6216 5765 6041 2137 9295 6048 2601 2955
## MonthlyIncome 2404.0  7428 13402  4649  4233 5396 5484 2008 8722 6545 2042 4739
## CV residual    -17.5 -2702 -1122 -1214 -1983 -369 -557 -129 -573  497 -559 1784
##                 205   210  214  216   229   240   252   255   258   259  264
## Predicted     2370  9845 6207 2537  6119  9878  6277  9999  6113 10146 1916
## cvpred        2393  9755 6179 2577  6104  9780  6275  9917  6074 10092 1890
## MonthlyIncome 3931  8189 6062 2345  4999 13269  5079 13757  4779  9713 2157
## CV residual   1538 -1566 -117 -232 -1105  3489 -1196  3840 -1295  -379  267
##                 275   280  281    283  286  313  328    333  340   347   363
## Predicted     18353  5890 5963 6487.6 6094 2079 6007 2356.7 2114  5989 18845
## cvpred        18255  5828 5904 6495.4 6063 2061 5956 2378.2 2105  5938 18820
## MonthlyIncome 18722  4035 7847 6513.0 5207 4084 5410 2342.0 2341  4033 19845
## CV residual     467 -1793 1943   17.6 -856 2023 -546  -36.2  236 -1905  1025
##                 365   373   381  396  415  419   433    446   456   462  465
## Predicted     17836 10586 18260 2199 2186 6067  6745 3302.2 17946  3012 2287
## cvpred        17680 10586 18155 2205 2184 6017  6792 3438.1 17819  3117 2302
## MonthlyIncome 19187  5210 18300 2695 2351 6474  4284 3420.0 19833  2013 2437
## CV residual    1507 -5376   145  490  167  457 -2508  -18.1  2014 -1104  135
##                469   492  495    498  504  509   524   526  528   547   551
## Predicted     5842 14382 6040 2227.6 9532 2621 18246 10867 6047 11151 17845
## cvpred        5787 14314 6005 2243.2 9415 2682 18133 10881 6009 11212 17691
## MonthlyIncome 4969 16880 5673 2326.0 9980 2793 19502  7525 6833  7403 19189
## CV residual   -818  2566 -332   82.8  565  111  1369 -3356  824 -3809  1498
##                552  556  588   602   603   611  612    613  623     629   635
## Predicted     2257 6920 2501  6275  9729  6463 6236 2713.2 2698 10930.0 10166
## cvpred        2274 6967 2537  6253  9635  6463 6231 2784.2 2763 10949.4 10114
## MonthlyIncome 4936 9602 3407  4087 11416  5363 6811 2707.0 3280 10976.0  9985
## CV residual   2662 2635  870 -2166  1781 -1100  580  -77.2  517    26.6  -129
##                 637  644   654   657   658  662  668   679  684    688  695
## Predicted     14353 2164 10300 10202 14253 2568 2492  6835 9791 2498.3 6082
## cvpred        14281 2158 10275 10157 14181 2600 2542  6872 9696 2535.7 6048
## MonthlyIncome 17665 2307 12504 12490 16032 2768 3408  5131 9439 2622.0 9525
```

```
## CV residual    3384  149  2229  2333  1851  168  866 -1741 -257   86.3 3477
##                 702   711  716   718  724   732  735  749  756  776  780   782
## Predicted      6028  9527 1800  5969 9869  6567 2249 2048 2279 2148 2424  6517
## cvpred         5995  9410 1755  5918 9772  6577 2276 2043 2305 2163 2469  6515
## MonthlyIncome 5714 11916 1555  3660 8858  5238 2972 3907 3708 3505 4787  4553
## CV residual    -281  2506 -200 -2258 -914 -1339  696 1864 1403 1342 2318 -1962
##                 785   788     794   813  820  822   825     832   835   842
## Predicted      6114  6068 10137.8  5802 5883 6044  6098 1732.38 15425  9534
## cvpred         6085  6028 10077.5  5730 5820 5996  6066 1680.92 15470  9417
## MonthlyIncome  4978  4941 10096.0  4260 9854 6931  4663 1675.00 13577  7988
## CV residual   -1107 -1087    18.5 -1470 4034  935 -1403   -5.92 -1893 -1429
##                 851   868   870
## Predicted      5780  6276  6468
## cvpred         5721  6261  6483
## MonthlyIncome  4559  4591  4425
## CV residual   -1162 -1670 -2058
##
## Sum of squares = 3.45e+08    Mean square = 2756228    n = 125
##
## fold 3
## Observations in test set: 125
##                   5   12   17   21   43   53    55   70   72    77   82   94
## Predicted      2483 2215 6186 6561 2853 2511  6007 2786 5949  9825 6103 6293
## cvpred         2501 2253 6250 6597 2770 2466  6001 2678 6019  9919 6035 6336
## MonthlyIncome 3760 2706 6725 8847 2587 2011  4424 2115 8463  7083 6883 6142
## CV residual    1259  453  475 2250 -183 -455 -1577 -563 2444 -2836  848 -194
##                  98  111   112   113   117   118  120   124   131  142  156
## Predicted      6991 2136 14510 10777  5995  6093 6122  5931  6068 1944 5560
## cvpred         7020 2168 14548 10871  6088  6125 6077  5967  6075 1972 5628
## MonthlyIncome  5163 2703 17099 10445  4950  4025 6687  4028  4187 2372 4736
## CV residual   -1857  535  2551  -426 -1138 -2100  610 -1939 -1888  400 -892
##                157   158   173  180   181   184   199  209   211  215  219  223
## Predicted      6321 14552 11229 6387 10340 10064 10222 2418 13840 2521 2103 5975
## cvpred         6283 14665 11228 6403 10420 10204 10133 2411 13878 2562 2100 6038
## MonthlyIncome 5677 12169 10527 8686 10274  8376 10248 3204 13570 2045 2632 5368
## CV residual    -606 -2496  -701 2283  -146 -1828   115  793  -308 -517  532 -670
##                224  226  228  257  268  269   276  307  314   318    319  320
## Predicted      2343 2448 2148 2013 2114 5817 18688 2690 6511  6079 5984.6 1834
## cvpred         2423 2375 2169 2048 2177 5820 18832 2711 6588  6188 6028.7 1870
## MonthlyIncome 1563 3748 2328 2789 1393 5968 19943 2380 6132  4617 6120.0 2994
## CV residual    -860 1373  159  741 -784  148  1111 -331 -456 -1571   91.3 1124
##                 330   343   354  360   366   372   376  380  385   388   400
## Predicted      14042  9937  6097 2397 15239  9445 14596 2358 3019  5924 10599
## cvpred         14099 10068  6158 2255 15214  9501 14562 2361 2917  5972 10696
## MonthlyIncome 16799  7119  4069 3210 13212  8412 14411 2132 2461  4262 10209
## CV residual    2700 -2949 -2089  955 -2002 -1089  -151 -229 -456 -1710  -487
##                  411  416  427  430  432   435  440  453  454   457  467  468
## Predicted      2388.8 6503 2107 6105 2303 10328 6120 5701 5918  6492 2469 2464
## cvpred         2399.5 6491 2168 6078 2319 10472 6136 5687 5914  6560 2478 2504
## MonthlyIncome 2342.0 5762 1274 5957 3544  9705 7625 5813 5487  5476 2644 3038
## CV residual    -57.5 -729 -894 -121 1225  -767 1489  126 -427 -1084  166  534
##                 470   487  490  497  499   501  502  518  521  533   537   584
## Predicted      5755  2146 6314 2553 6174 11038 1995 2262 2121 2038 10173  6152
## cvpred         5822  2211 6363 2556 6111 10958 1975 2246 2111 1957 10318  6205
```

```
## MonthlyIncome  4450  1081 6142 2654 5237 10008 2028 2766 2553 2700  7351  4779
## CV residual   -1372 -1130 -221   98 -874  -950   53  520  442  743 -2967 -1426
##                  599  604  610   615   632  636  648   653  655   677  678  680
## Predicted     6106.3 2316 2176  9726 10612 2029 1892 10092 2235  6231 6064 2867
## cvpred        6188.8 2253 2210  9808 10737 2065 1905 10057 2262  6267 6090 2833
## MonthlyIncome 6142.0 3131 2342  7264  9888 2376 2305 13499 3894  5228 6323 2329
## CV residual    -46.8  878  132 -2544  -849  311  400  3442 1632 -1039  233 -504
##                 683  685    690   697  698   703   714  723   726   740  748
## Predicted     6162 2035 2652.5 14695 6381  9828 10048 2556  6218  6111 6188
## cvpred        6204 2035 2578.9 14807 6347  9825 10062 2531  6325  6160 6156
## MonthlyIncome 5347 2713 2592.0 15202 5914 10377  9699 2996  4898  3681 5343
## CV residual   -857  678   13.1   395 -433   552  -363  465 -1427 -2479 -813
##                 767   773  777   791  792   793   801  812   814  817   818
## Predicted     2300 10435 2357  6382 2140 10052  6353 2627 18396 6654 10012
## cvpred        2236 10446 2365  6441 2099 10179  6405 2648 18536 6717 10067
## MonthlyIncome 3376 10938 2064  4789 2311  7756  5376 3229 19658 6877 11691
## CV residual   1140   492 -301 -1652  212 -2423 -1029  581  1122  160  1624
##                 830   831  833   839  852    859   860   862  869
## Predicted     18883  5565 2003 10822 2127 2237.9  9879  9989 2533
## cvpred        18974  5549 1983 10758 2049 2220.8  9980  9894 2546
## MonthlyIncome 18200  3633 3196 10333 2206 2238.0  7978 10231 4477
## CV residual    -774 -1916 1213  -425  157   17.2 -2002   337 1931
##
## Sum of squares = 1.95e+08    Mean square = 1558045    n = 125
##
## fold 4
## Observations in test set: 124
##                   6    8    20   22   23   25    33   42   45     51   56   57
## Predicted     9556 6187  6484 6604 5913 6093 14761 6462 2069 2673.7 2103 2202
## cvpred        9532 6224  6479 6609 5903 6060 14758 6464 2059 2691.3 2109 2263
## MonthlyIncome 8793 6694  5033 8120 5679 6949 16872 6632 1223 2619.0 2289 2759
## CV residual   -739  470 -1446 1511 -224  889  2114  168 -836  -72.3  180  496
##                 65    79    87    90   133  134   136   143    145   159   161
## Predicted     1878 10503  6510 18517  5594 6088  5959  9873 6442.3 10331  6463
## cvpred        1929 10505  6549 18523  5660 6128  5917  9853 6413.8 10437  6476
## MonthlyIncome 2210  8865  4960 18947  4523 5295  4244 10845 6377.0 10820  5126
## CV residual    281 -1640 -1589   424 -1137 -833 -1673   992  -36.8   383 -1350
##                  166   167  171   174   192  196  202  221   236  238  239  244
## Predicted     2131.5 14825 5701  9563 10119 2201 6060 5935 10450 5978 2402 2434
## cvpred        2118.5 14919 5750  9535 10117 2196 6004 5909 10418 5963 2402 2454
## MonthlyIncome 2044.0 16413 5304  8380  9637 2723 6220 6274 10609 6322 2187 4400
## CV residual    -74.5  1494 -446 -1155  -480  527  216  365   191  359 -215 1946
##                 245    246   247     248   250  260  271  285  287   295  296
## Predicted      6374 2539.5  9627 2285.80 14366 2398 6094 2606 2036 14226 2122
## cvpred         6393 2552.6  9577 2271.03 14335 2403 6058 2620 2052 14213 2151
## MonthlyIncome  5056 2532.0  7143 2269.00 16595 3977 5154 2844 2472 16437 2660
## CV residual   -1337  -20.6 -2434   -2.03  2260 1574 -904  224  420  2224  509
##                 297  317    321   331   332   334  341    344   345   361  379
## Predicted     5902 2402 2349.8 14864  9927  9768 2547 2169.6  5777 18787 2208
## cvpred        5853 2402 2343.2 14891  9892  9713 2570 2171.1  5856 18772 2243
## MonthlyIncome 9547 2231 2332.0 15379 10400  7412 3419 2083.0  4115 19999 2168
## CV residual   3694 -171  -11.2   488   508 -2301  849  -88.1 -1741  1227  -75
##                 383   386  391   392  401   414  418   422  424  425  442   445
## Predicted      6656  9860 6033  5851 6465  5917 2136  6141 2062 5561 2423 18520
```

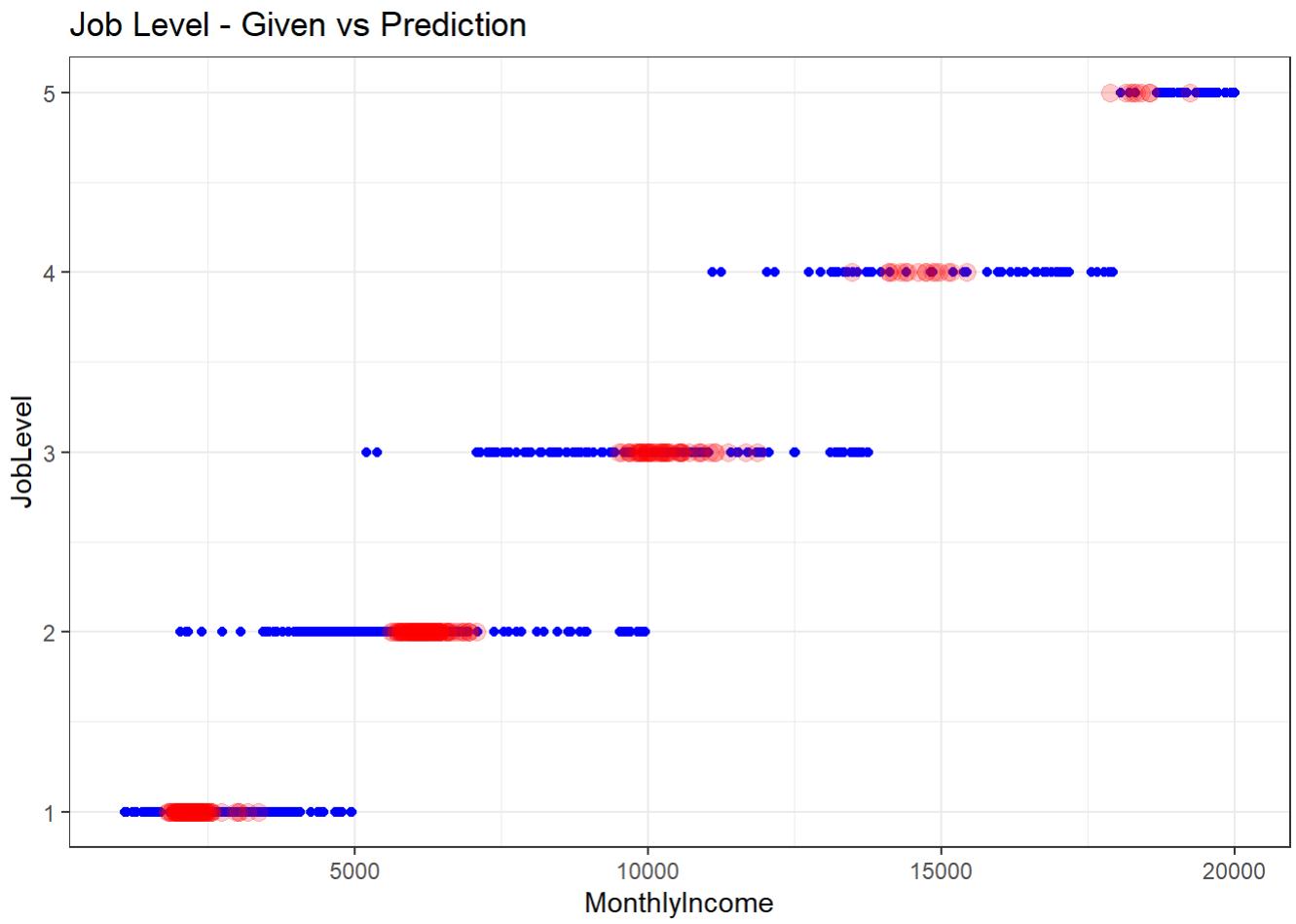
```
## cvpred         6690  9797 6041  5921 6468  5941 2199  6198 2056 5612 2458 18527
## MonthlyIncome  5486 13582 6500  4157 6091  4648 3688  4449 1611 4724 2974 19328
## CV residual   -1204  3785  459 -1764 -377 -1293 1489 -1749 -445 -888  516   801
##                450  466  471   478   485   489   505  506  520  525  534   540
## Predicted      5980 2331 2473  6101 11505 11280 14829 6232 2252 2575 6126  6188
## cvpred         6037 2332 2471  6198 11634 11317 14823 6229 2334 2596 6167  6168
## MonthlyIncome  4385 3944 2226  4306 10312  5381 17169 5878 2587 2177 6667  5155
## CV residual   -1652 1612 -245 -1892 -1322 -5936  2346 -351  253 -419  500 -1013
##                543   555  558  563   564   567  568  575  581  585  593   598
## Predicted     2109  5879 2428 2047  9836  6138 6181 5740 2131 6073 2492  7265
## cvpred        2103  5888 2416 2088  9817  6122 6140 5797 2194 6091 2482  7311
## MonthlyIncome 2566  4876 3622 2811 11031  2176 9714 6272 3669 6465 2909  3448
## CV residual    463 -1012 1206  723  1214 -3946 3574  475 1475  374  427 -3863
##                631  640   656   663    666   669  672   676    682    694  699
## Predicted     3229 2334  6000 10265 6342.1  6308 5935 18189 2167.0 2288.9 2173
## cvpred        3256 2331  6004 10255 6340.6  6335 5909 18167 2239.8 2300.4 2170
## MonthlyIncome 2662 3041  4741 10368 6294.0  4883 5505 18789 2258.0 2285.0 1569
## CV residual   -594  710 -1263   113  -46.6 -1452 -404   622   18.2  -15.4 -601
##                704  706   712   719  728  730  733   734   745   760   775  796
## Predicted     1926 6121 15173  6069 2259 2260 2013  5915  6054 18251 10573 2495
## cvpred        1975 6112 15322  6078 2299 2257 2076  5934  6044 18197 10557 2472
## MonthlyIncome 2148 6230 14118  4581 3815 2523 2863  4707  4907 18824 10124 3348
## CV residual    173  118 -1204 -1497 1516  266  787 -1227 -1137   627  -433  876
##                797  807   811  815  841  850  857   864
## Predicted     13824 6018  5937 2160 2102 2171 2329 14311
## cvpred        13837 6056  5910 2172 2100 2155 2326 14269
## MonthlyIncome 13496 6232  3072 2899 1420 2544 2657 13770
## CV residual    -341  176 -2838  727 -680  389  331  -499
##
## Sum of squares = 2.37e+08    Mean square = 1913945    n = 124
##
## fold 5
## Observations in test set: 124
##                   1     4    11   16    29   36    37    39     47   49   50
## Predicted      6100 10106 17669 6097 10257 2266 10385 18862 6676.5 2142 2221
## cvpred         6079 10095 17607 6145 10287 2231 10462 19042 6672.5 2081 2234
## MonthlyIncome  4403 10422 19392 6932 10448 2794 10306 19717 6735.0 1878 2024
## CV residual   -1676   327  1785  787   161  563  -156   675   62.5 -203 -210
##                  54   58   68   83   91  103  115  122  128  148  170    176
## Predicted      5935 2545 1985 6173 2114 6096 6196 2084 5743 6310 2027 2076.2
## cvpred         5891 2538 1976 6120 2089 6071 6229 2042 5714 6278 2052 2044.4
## MonthlyIncome  4447 4723 2647 5321 1702 6538 8966 2380 5906 6392 2514 2070.0
## CV residual   -1444 2185  671 -799 -387  467 2737  338  192  114  462   25.6
##                177  185   194  195   198  200  225  231   241  262  270  278
## Predicted     18328 2139  6435 2537 14433 1878 2572 5902  9343 2594 5912 2175
## cvpred        18412 2117  6430 2498 14531 1868 2586 5873  9322 2564 5911 2164
## MonthlyIncome 19859 3755  4051 3904 13120 2127 4771 5770  7406 2362 9957 2455
## CV residual    1447 1638 -2379 1406 -1411  259 2185 -103 -1916 -202 4046  291
##                289  292  300  311   316  322   324   329   336  342    349
## Predicted     9689 5826 2476 6477  5924 2395  6537  6604  6021 2060 2216.1
## cvpred        9627 5806 2430 6471  5903 2363  6614  6691  6019 2030 2252.4
## MonthlyIncome 9738 5674 3464 5484  4163 2691  4615  2133  4508 3867 2290.0
## CV residual    111 -132 1034 -987 -1740  328 -1999 -4558 -1511 1837   37.6
##                350  368     369   371  374  390  393  399  405  407    417
```

```
## Predicted     15150 6423 2069.80  5947 2179 2357 2190 6191 2069 6015 2503.6
## cvpred        15326 6409 2065.21  5893 2144 2290 2157 6175 2092 6010 2513.8
## MonthlyIncome 11103 5473 2073.00  4148 3936 2174 3065 6306 2559 5468 2564.0
## CV residual   -4223 -936    7.79 -1745 1792 -116  908  131  467 -542   50.2
##                 420  429  437  439   441   447   463  464  472  479   480
## Predicted      9875 2211 2395 2325  6023  6525 10432 5792 1768 2591 10039
## cvpred         9814 2198 2385 2261  6010  6581 10497 5813 1764 2630 10048
## MonthlyIncome 11935 3730 4382 2683  4998  4735  9208 5238 2439 2950 13744
## CV residual    2121 1532 1997  422 -1012 -1846 -1289 -575  675  320  3696
##                   482  486   508   510   514  515  544   548   549  554   557
## Predicted     2324.18 6167 10464 18003 10012 2231 6718  6441 10572 2642  7032
## cvpred        2358.36 6120 10479 18026  9943 2172 6797  6451 10648 2747  7149
## MonthlyIncome 2356.00 5993 10596 19545 13458 2610 6553  5094 12061 2259  3780
## CV residual     -2.36 -127   117  1519  3515  438 -244 -1357  1413 -488 -3369
##                 571  572  573    576  580   587  590  591   596  605   606  618
## Predicted      6452 1863 5907 6049.5 6170  6454 1937 1900  6124 2093 15195 6356
## cvpred         6491 1853 5850 5962.7 6163  6465 1929 1873  6048 2039 15372 6334
## MonthlyIncome  4393 2693 6834 5993.0 6623  5329 4014 2814  4037 2760 14814 5454
## CV residual   -2098  840  984   30.3  460 -1136 2085  941 -2011  721  -558 -880
##                 621  624  633  634   646   649  650  674   687  707  708     713
## Predicted     2051 9375 7182 2717  9884  6454 2675 1875  9564 6215 2371 2519.37
## cvpred        2030 9341 7279 2706  9854  6474 2672 1895  9518 6244 2403 2551.33
## MonthlyIncome 2926 8837 6861 2835 10855  4312 2166 2022  7553 6854 2109 2559.00
## CV residual    896 -504 -418  129  1001 -2162 -506  127 -1965  610 -294    7.67
##                 717   739   743  750  761  765  770   771  778   784   787
## Predicted     6150  5930  6063 2267 2124 2251 6090  6336 2185  6168 10587
## cvpred        6153  5923  6016 2240 2125 2225 6062  6338 2146  6193 10597
## MonthlyIncome 6578  4601  4285 2500 2404 3737 5593  4682 2827  4539 13116
## CV residual    425 -1322 -1731  260  279 1512 -469 -1656  681 -1654  2519
##                  804  808   827   828   843   844   848   853  861  865
## Predicted     9650.0 1864 10167  5992 15027  9700 10443  5758 2242 9770
## cvpred        9609.6 1834 10157  5946 15234  9681 10480  5749 2203 9771
## MonthlyIncome 9582.0 2323  8161  4538 13826  8500  9613  4193 2838 9380
## CV residual    -27.6  489 -1996 -1408 -1408 -1181  -867 -1556  635 -391
##
## Sum of squares = 2.5e+08    Mean square = 2019304    n = 124
##
## fold 6
## Observations in test set: 124
##                   3   14   18   19   24   30   48   63     67   74    75    96
## Predicted     9872 2185 6238 2331 6266 1996 9627 2341 6514.5 2437  9844  6917
## cvpred        9826 2171 6200 2330 6272 1951 9670 2364 6579.8 2395  9906  6965
## MonthlyIncome 9362 2476 5258 2932 5332 4258 8740 3067 6646.0 4766  8446  4907
## CV residual   -464  305 -942  602 -940 2307 -930  703   66.2 2371 -1460 -2058
##                 104  110   129   132  149  150   160  163   165   175    186
## Predicted     2445 2196  6788  5929 1985 5927 10685 6033  9614 14659 2031.8
## cvpred        2459 2198  6856  5926 1925 5843 10590 6021  9664 14639 2040.1
## MonthlyIncome 2819 2439  4081  4507 3597 4877 13191 7104  8020 14852 2070.0
## CV residual    360  241 -2775 -1419 1672 -966  2601 1083 -1644   213   29.9
##                 189   193   201  203  204    206  207  213  227   230   233
## Predicted     18658  6390 14506 6272 2526 2558.3 6166 2191 2154 14530 14267
## cvpred        18689  6414 14406 6341 2609 2535.7 6176 2145 2134 14622 14330
## MonthlyIncome 19144  4810 13194 6162 2216 2559.0 5346 4680 1951 13966 12936
## CV residual     455 -1604 -1212 -179 -393   23.3 -830 2535 -183  -656 -1394
```

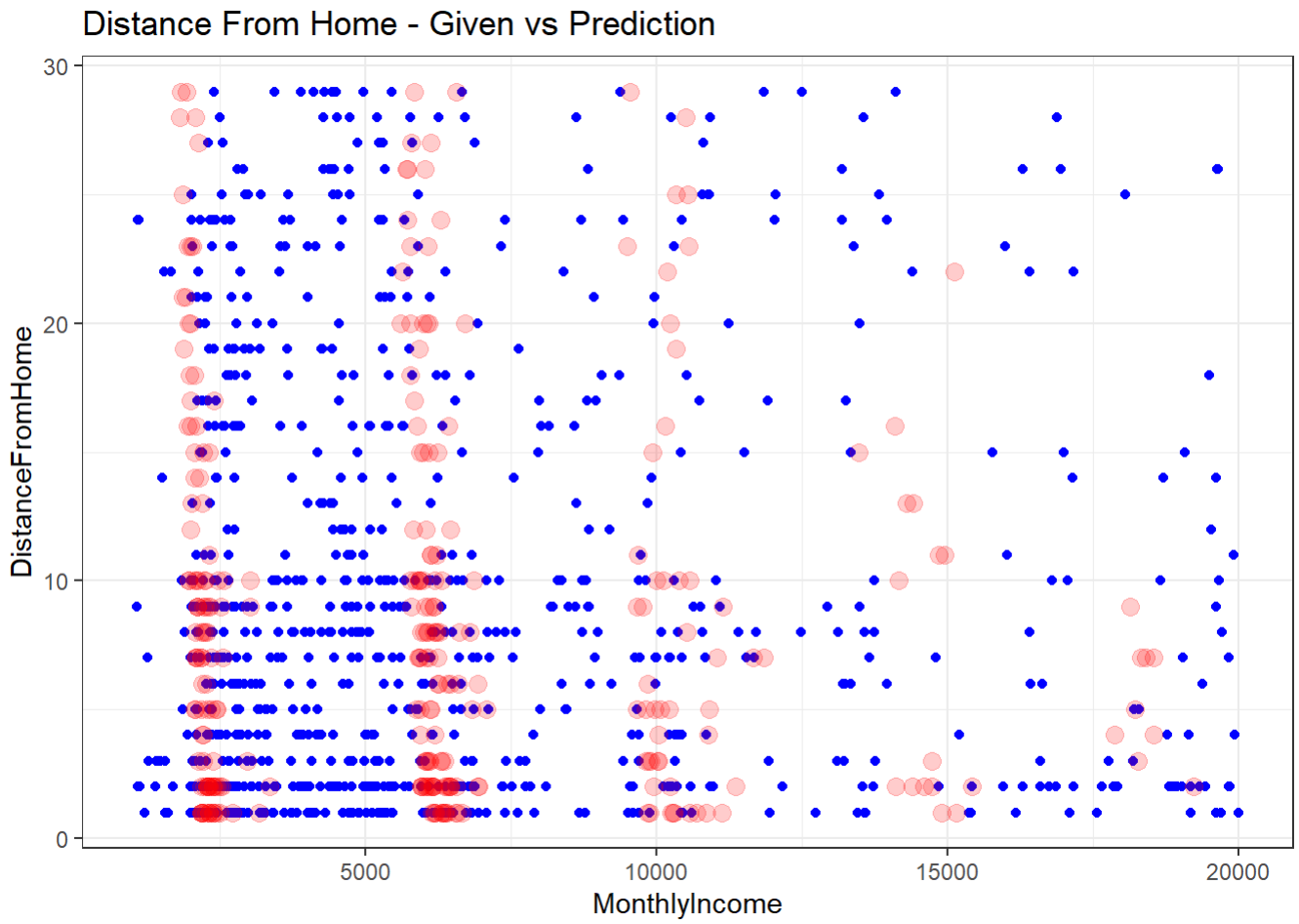
```
##                  253  254  265  272  288  290    293   294  301  304  308   312
## Predicted     2256.3 1930 2221 5954 2003 2247 2227.5  9803 6361 5995 2507 14000
## cvpred        2248.4 1865 2190 6035 1973 2253 2219.9  9931 6371 6061 2484 14122
## MonthlyIncome 2274.0 2956 2774 5433 2661 2572 2279.0 11957 6524 5098 2979 13237
## CV residual     25.6 1091  584 -602  688  319   59.1  2026  153 -963  495  -885
##                 323     335  337   346   351  352   356  358    359  362   364
## Predicted     2635 18925.1 5847 13973 10297 6087  6042 2101 6123.9 1840 18103
## cvpred        2647 18893.1 5822 14032 10340 6145  6158 2031 6219.6 1766 18187
## MonthlyIncome 3761 18844.0 4968 13341 13348 6929  4759 2546 6162.0 1091 19436
## CV residual   1114   -49.1 -854  -691  3008  784 -1399  515  -57.6 -675  1249
##                 375   394  402   410   421   426    434  443  444   451  460
## Predicted     1950  9979 6166 14663  6157 10423 2082.8 2597 2552  6325 6144
## cvpred        1936 10010 6113 14590  6191 10321 2056.7 2591 2558  6282 6231
## MonthlyIncome 2728 10266 5265 17159  4325 10798 2109.0 3485 3195  4878 5810
## CV residual    792   256 -848  2569 -1866   477   52.3  894  637 -1404 -421
##                 461  476  483  488   491   496  500   527  529  530   531  536
## Predicted     2682 6399 5943 2038 18259 13987 6241  6238 2367 2291 10138 2320
## cvpred        2657 6477 5999 2007 18216 14106 6201  6246 2409 2283 10175 2304
## MonthlyIncome 2093 5772 5747 2911 18665 16307 5736  5006 2859 4963  8823 2089
## CV residual   -564 -705 -252  904   449  2201 -465 -1240  450 2680 -1352 -215
##                 542   553  560   562  577   579   582  583   586   607   616
## Predicted     2534  6105 5975 14665 3626  6144 14187 6676  6165  7170  6092
## cvpred        2560  6190 6046 14621 3603  6231 14105 6605  6178  7185  6229
## MonthlyIncome 3143  3886 5296 15787 2858  5042 15992 5811  2741  5473  3986
## CV residual    583 -2304 -750  1166 -745 -1189  1887 -794 -3437 -1712 -2243
##                 630   638  651  664   665   673   686   689  693  701  705  736
## Predicted     9571  9872 2144 2144  5835  5921  9852  2256 6931 2051 2477 6169
## cvpred        9611  9821 2150 2139  5850  5926  9906  2248 6911 2020 2477 6248
## MonthlyIncome 8621  8628 2450 3917  4568  4898 10648  1102 6781 2804 2267 6500
## CV residual   -990 -1193  300 1778 -1282 -1028   742 -1146 -130  784 -210  252
##                 742    747   754  758  766   768  769   772  779    783   803
## Predicted     10105 2357.0  9964 7168 6229 10450 6340 10852 2016 2321.8 13981
## cvpred        10076 2321.9 10001 7158 6246 10351 6357 10907 2067 2330.5 14047
## MonthlyIncome 11557 2335.0  7314 6651 5769 10932 7379 10453 2570 2277.0 17068
## CV residual    1481   13.1 -2687 -507 -477   581 1022  -454  503  -53.5  3021
##                 816  819  821     829  834   836  847  849  863   866
## Predicted     2556 2473 2426 2422.28 2075  9785 2241 6069 6084  9724
## cvpred        2651 2485 2430 2464.15 2042  9844 2230 5990 6077  9788
## MonthlyIncome 2781 2720 3312 2468.00 2552  8834 1483 6380 5304  7898
## CV residual    130  235  882    3.85  510 -1010 -747  390 -773 -1890
##
## Sum of squares = 1.89e+08    Mean square = 1528155    n = 124
##
## fold 7
## Observations in test set: 124
##                    13   28    61   64   66   80    81   85   88    92    95
## Predicted     10274.7 6099 10209 2000 2382 2114  6857 6331 2145  9305 18055
## cvpred        10264.5 6093 10206 2009 2348 2119  6769 6321 2148  9411 18065
## MonthlyIncome 10221.0 8224 10325 1514 3298 1129  4777 5410 3743  7336 19926
## CV residual     -43.5 2131   119 -495  950 -990 -1992 -911 1595 -2075  1861
##                 101   102    119   130    137  144  169  172  179  187   188
## Predicted     2210  9949 2463.9 14624 2272.0 6193 6379 6311 9729 5821 15203
## cvpred        2201  9965 2427.9 14566 2248.5 6185 6368 6301 9771 5864 15083
## MonthlyIncome 2099 10793 2451.0 13225 2319.0 6502 5577 5467 8998 5206 15972
```

```
## CV residual   -102   828   23.1 -1341   70.5  317 -791 -834 -773 -658   889
##                190  191  197  217  220   222  234   242  249  251  261   267
## Predicted      5978 5561 2430 6933 2505  5946 6534 18216 6086 2272 2320 13696
## cvpred         5990 5642 2408 6849 2467  5958 6501 18203 6104 2248 2298 13746
## MonthlyIncome  4221 5775 2075 9824 2875  4286 5940 18061 5467 2008 2785 17007
## CV residual   -1769  133 -333 2975  408 -1672 -561  -142 -637 -240  487  3261
##                274  277  279  282  284  303     305  306   325  326  338   339
## Predicted     6245 2132 6020 2304 6147 2364 2519.37 2391 14173 2225 6200  6126
## cvpred        6237 2123 6032 2310 6158 2334 2482.46 2355 14179 2207 6183  6129
## MonthlyIncome 5647 2244 5744 4721 7547 3537 2479.00 3162 12742 2318 7756  4728
## CV residual   -590  121 -288 2411 1389 1203   -3.46  807 -1437  111 1573 -1401
##                348   353  357   367   370  377   378    384   387   389    395
## Predicted     6022  6547 2211  5983  6236 2264  6703 2320.3  9775  6567 2607.7
## cvpred        6034  6508 2205  6012  6218 2269  6637 2298.1  9817  6527 2551.9
## MonthlyIncome 6397  4422 2368  4444  5055 2743  5087 2322.0  8789  4502 2515.0
## CV residual    363 -2086  163 -1568 -1163  474 -1550   23.9 -1028 -2025  -36.9
##                397  398   406  408   409  412  413    423  436  452   455
## Predicted     15132 2164 14721 2103 15106 9686 2462 6237.5 5936 1987 14536
## cvpred        15012 2150 14650 2113 14996 9744 2424 6226.7 5957 2015 14498
## MonthlyIncome 17861 2742 16627 2741 15427 9241 3022 6172.0 6214 3420 15402
## CV residual    2849  592  1977  628   431 -503  598  -54.7  257 1405   904
##                473   494  511  513   516   519  522    532   541   559  565
## Predicted     10583  9839 5989 2627  6529  6173 1840 6943.1  5894  5680 2414
## cvpred        10540  9879 6002 2567  6493  6163 1874 6849.1  5928  5736 2390
## MonthlyIncome 13245  8606 5343 3433  5482  4627 2408 6815.0  3491  4434 3833
## CV residual    2705 -1273 -659  866 -1011 -1536  534  -34.1 -2437 -1302 1443
##                  569  574  594   595   597  609  614  619  620    625  626
## Predicted     6148.2 6010 2189  9995  9732 2155 1816 9619 2170 2657.9 2384
## cvpred        6150.5 6025 2192 10011  9792 2160 1858 9684 2163 2600.6 2347
## MonthlyIncome 6244.0 6799 3211 10435 10903 3578 2121 9991 1281 2561.0 2517
## CV residual     93.5  774 1019   424  1111 1418  263  307 -882  -39.6  170
##                 627   628  639  641  659  661   671  675   681   710   715
## Predicted      5897  9629 2021 1768 6129 5891 10592 2311  5968 15106  9747
## cvpred         5928  9692 2035 1817 6135 5915 10532 2288  5986 14996  9783
## MonthlyIncome  4162 11510 2725 2693 6032 5309 13664 3517  4256 17924  7587
## CV residual   -1766  1818  690  876 -103 -606  3132 1229 -1730  2928 -2196
##                 720  721  727  729   731   737  744  746  751  752   755  759
## Predicted      9733 9744 2494 1840  6161 14074 2168 1954 1950 2144 10610 2488
## cvpred         9785 9790 2460 1874  6145 14084 2164 1976 1977 2148 10548 2455
## MonthlyIncome  7655 9434 2973 2340  4765 12031 3295 2296 2323 2302 13549 3306
## CV residual   -2130 -356  513  466 -1380 -2053 1131  320  346  154  3001  851
##                  764    790  795  798   805   806  823  837    845  856   867
## Predicted     6282.9 2136.8 5569 2091  6611  9961 5480 2445 6040.6 2036  5615
## cvpred        6284.8 2140.6 5644 2094  6561  9991 5567 2414 6055.5 2051  5675
## MonthlyIncome 6347.0 2096.0 6804 2398  2042  8321 6712 2996 6029.0 2404  4465
## CV residual     62.2  -44.6 1160  304 -4519 -1670 1145  582  -26.5  353 -1210
##
## Sum of squares = 2.19e+08    Mean square = 1768845    n = 124
##
## Overall (Sum over all 124 folds)
##      ms
## 1902147
```

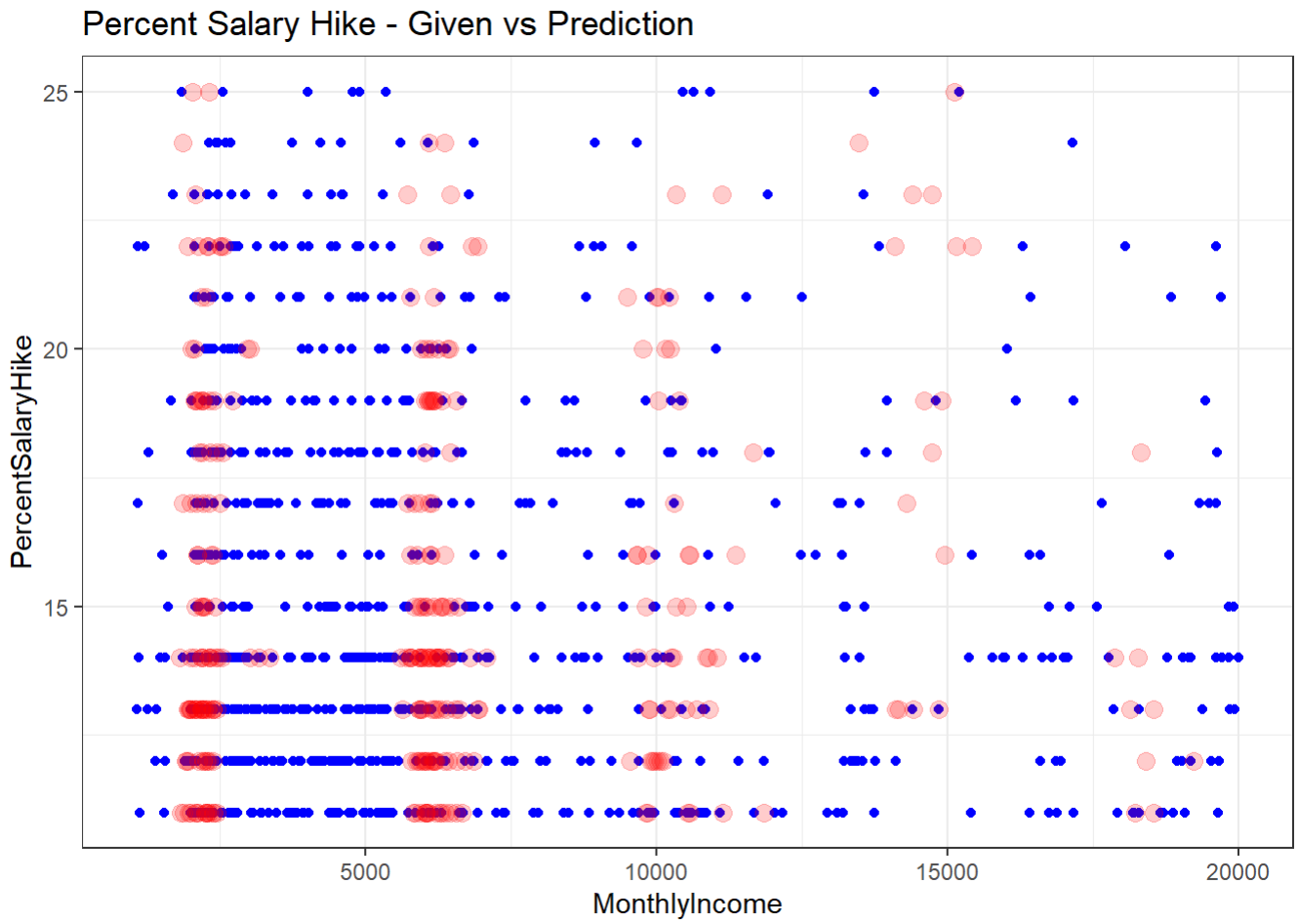
```
################ PREFERRED MLR MODEL RUN ON BLIND DATASET FOR GRADING SUBMISSION ###############

path.mlr.prediction.dB <- "C:\\Users\\dloveday\\Dropbox\\Family\\School\\SMU\\Courses\\Spring 20
21\\DS 6306 - Doing Data Science\\Lecture Notes\\Unit 14 and 15 Case Study 2\\CaseStudy2CompSet
 No Salary.csv"
mlr.prediction.input.dB <- read.csv(path.mlr.prediction.dB)

mlr.prediction <- predict(fit2, mlr.prediction.input.dB, interval = "confidence")
mlr.prediction.input.dB$MonthlyIncome <- mlr.prediction[, "fit"]

# Compare DistanceFromHome
#dev.new()
plot.JobLevel <-  ggplot()+
                    theme_bw()+
                    geom_point(data = employee.dB, aes(x = MonthlyIncome, y = JobLevel), color =
"blue")+
                    geom_point(data = mlr.prediction.input.dB, aes(x = MonthlyIncome, y = JobLev
el), color = "red", size = 3, alpha = 0.2)+
                    ggtitle("Job Level - Given vs Prediction")
multiplot(plot.JobLevel)
```



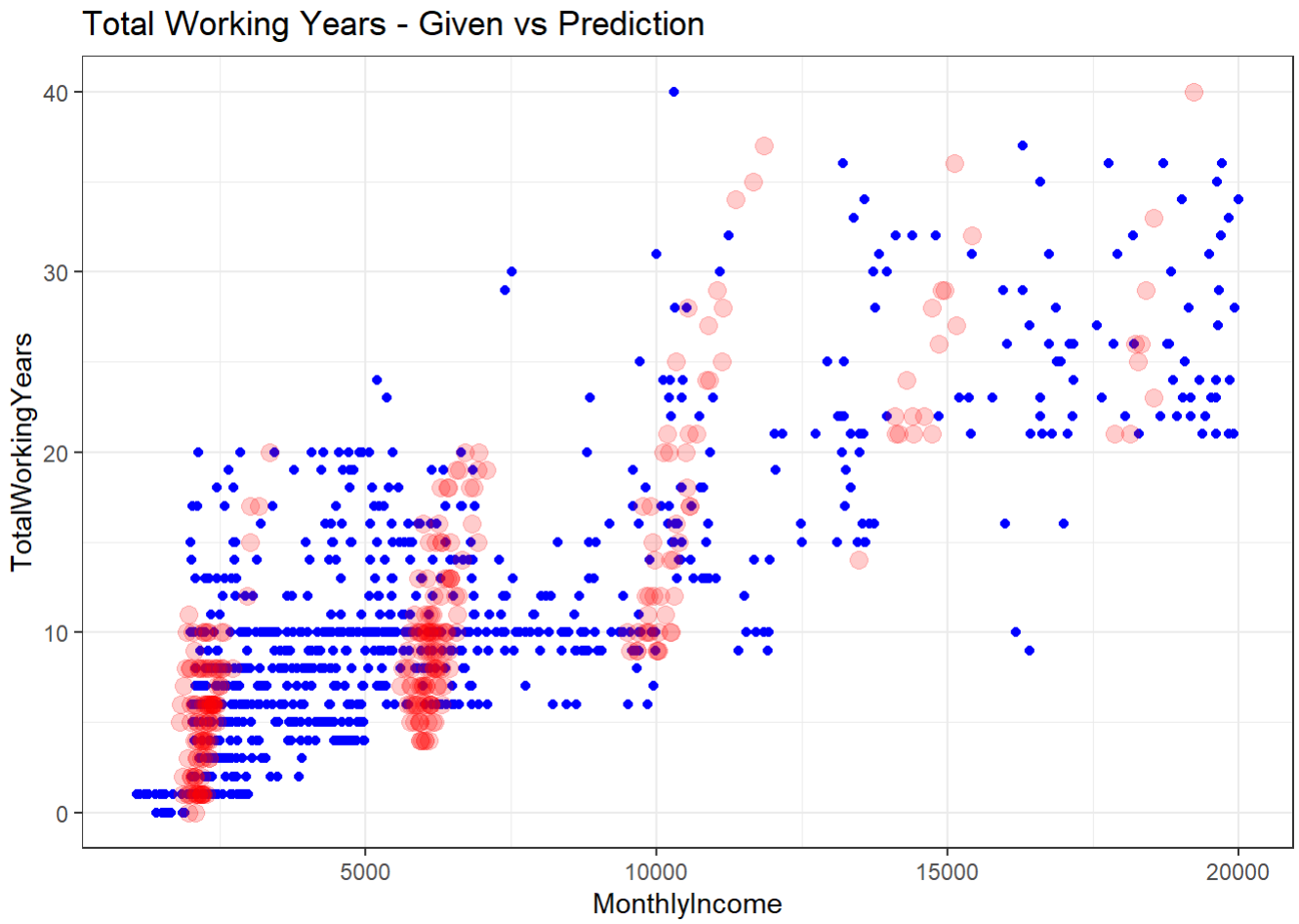
```
#dev.new()
plot.DistanceFromHome <-  ggplot()+
                            theme_bw()+
                            geom_point(data = employee.dB, aes(x = MonthlyIncome, y = DistanceFr
omHome), color = "blue")+
                            geom_point(data = mlr.prediction.input.dB, aes(x = MonthlyIncome, y
 = DistanceFromHome), color = "red", size = 3, alpha = 0.2)+
                            ggtitle("Distance From Home - Given vs Prediction")
multiplot(plot.DistanceFromHome)
```



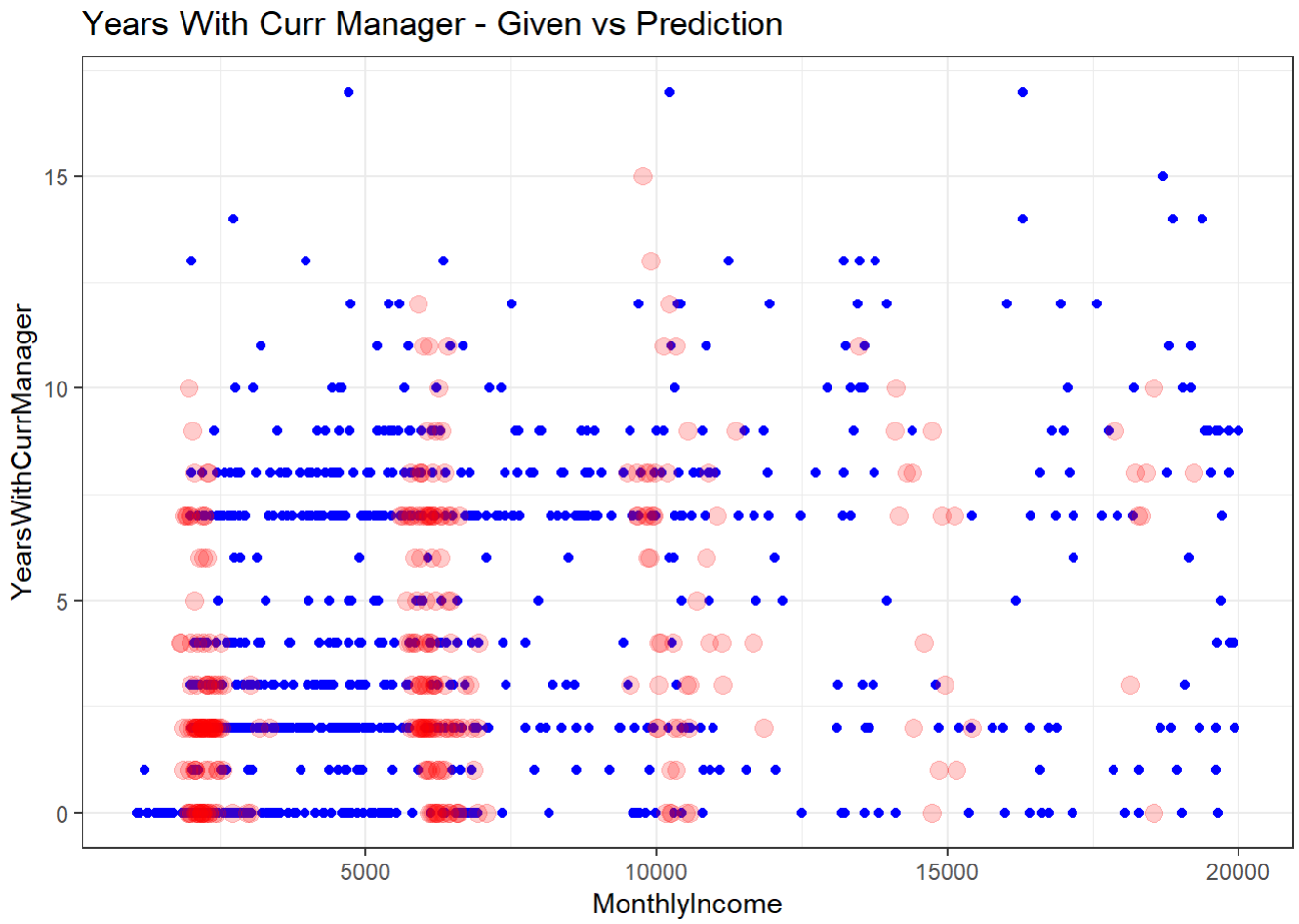

```
#dev.new()
plot.PercentSalaryHike <-  ggplot()+
                             theme_bw()+
                             geom_point(data = employee.dB, aes(x = MonthlyIncome, y = PercentS
alaryHike), color = "blue")+
                             geom_point(data = mlr.prediction.input.dB, aes(x = MonthlyIncome,
 y = PercentSalaryHike), color = "red", size = 3, alpha = 0.2)+
                             ggtitle("Percent Salary Hike - Given vs Prediction")
multiplot(plot.PercentSalaryHike)
```



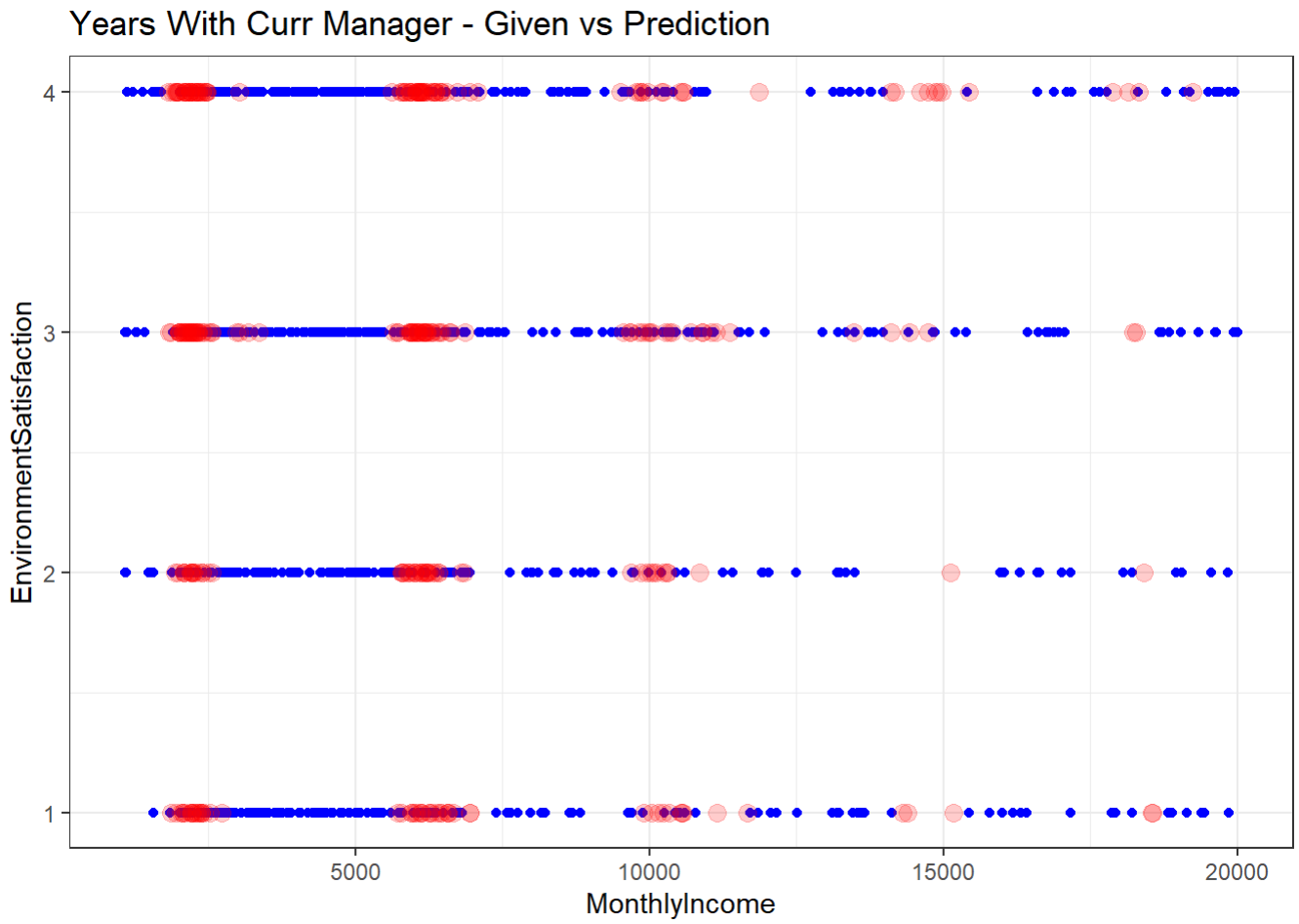

```
#dev.new()
plot.TotalWorkingYears <-  ggplot()+
                           theme_bw()+
                           geom_point(data = employee.dB, aes(x = MonthlyIncome, y = TotalWor
kingYears), color = "blue")+
                           geom_point(data = mlr.prediction.input.dB, aes(x = MonthlyIncome,
 y = TotalWorkingYears), color = "red", size = 3, alpha = 0.2)+
                           ggtitle("Total Working Years - Given vs Prediction")
multiplot(plot.TotalWorkingYears)
```




```
#dev.new()
plot.YearsWithCurrManager <-  ggplot()+
                                theme_bw()+
                                geom_point(data = employee.dB, aes(x = MonthlyIncome, y = YearsW
ithCurrManager), color = "blue")+
                                geom_point(data = mlr.prediction.input.dB, aes(x = MonthlyIncom
e, y = YearsWithCurrManager), color = "red", size = 3, alpha = 0.2)+
                                ggtitle("Years With Curr Manager - Given vs Prediction")
multiplot(plot.YearsWithCurrManager)
```




```
#dev.new()
plot.EnvironmentSatisfaction <-  ggplot()+
                                    theme_bw()+
                                    geom_point(data = employee.dB, aes(x = MonthlyIncome, y = En
vironmentSatisfaction), color = "blue")+
                                    geom_point(data = mlr.prediction.input.dB, aes(x = MonthlyIn
come, y = EnvironmentSatisfaction), color = "red", size = 3, alpha = 0.2)+
                                    ggtitle("Years With Curr Manager - Given vs Prediction")
multiplot(plot.EnvironmentSatisfaction)
```

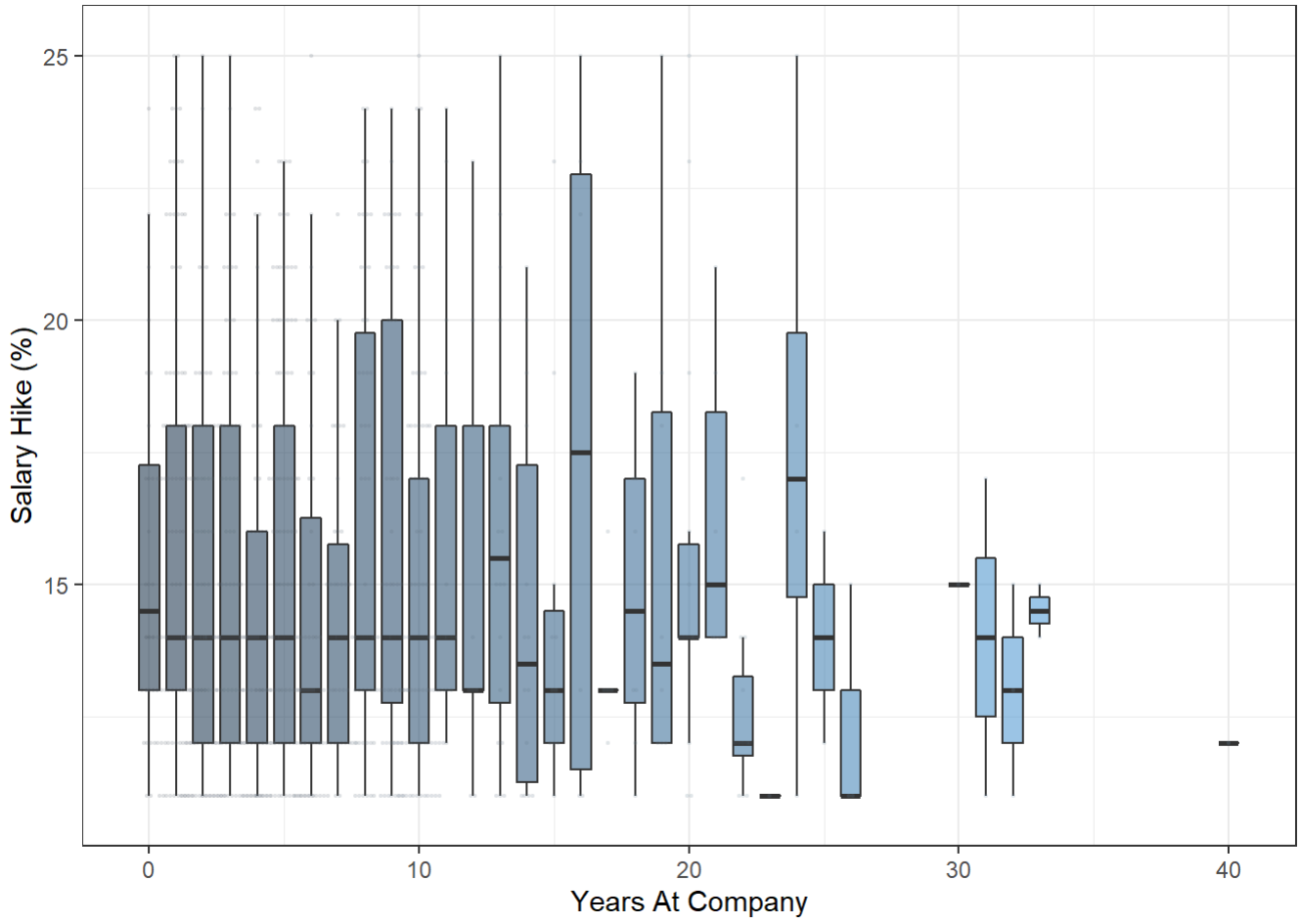



```
# Generate Outputs
mlr.prediction.output.dB <- mlr.prediction.input.dB[, c("ID", "MonthlyIncome")]

write.csv(mlr.prediction.input.dB[, c("ID", "MonthlyIncome")], file.path("C:\\Users\\dloveday\\D
ropbox\\Family\\School\\SMU\\Courses\\Spring 2021\\DS 6306 - Doing Data Science\\Lecture Notes
\\Unit 14 and 15 Case Study 2\\DL Work\\Outputs for Submission\\","Case2PredictionsLoveday Salar
y.csv"), row.names = FALSE)
```

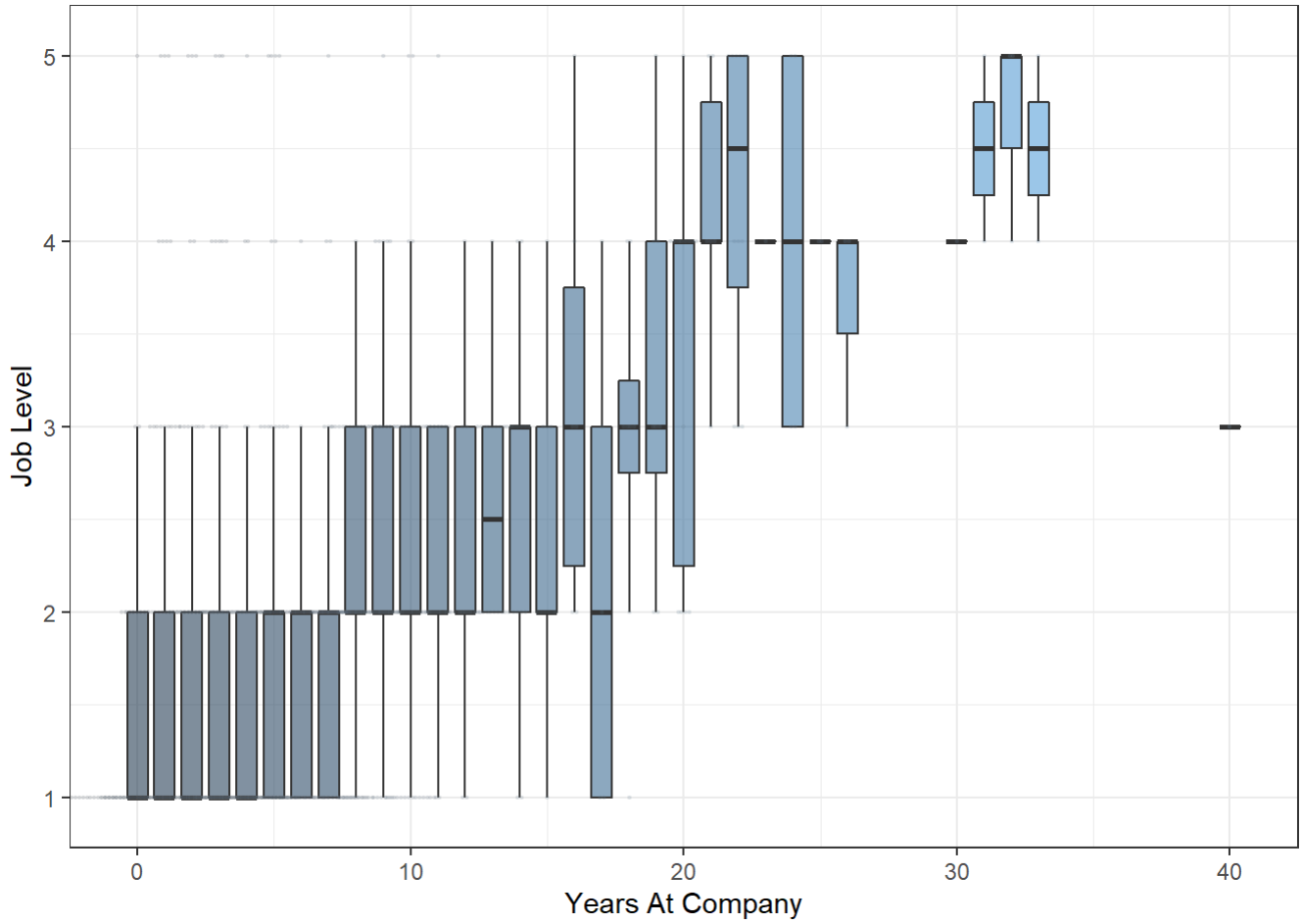
STEP 4 - PRESENTATION GRAPHICS

```
age.quartile.bins <- (max(employee.dB$Age) -  min(employee.dB$Age))/4
age.1st <- min(employee.dB$Age) + age.quartile.bins*1
age.2nd <- min(employee.dB$Age) + age.quartile.bins*2
age.3rd <- min(employee.dB$Age) + age.quartile.bins*3
age.4th <- min(employee.dB$Age) + age.quartile.bins*4


yearscurrentrole.quartile.bins <- (max(employee.dB$YearsInCurrentRole) -  min(employee.dB$YearsI
nCurrentRole))/4
yearscurrentrole.1st <- min(employee.dB$YearsInCurrentRole) + yearscurrentrole.quartile.bins*1
yearscurrentrole.2nd <- min(employee.dB$YearsInCurrentRole) + yearscurrentrole.quartile.bins*2
yearscurrentrole.3rd <- min(employee.dB$YearsInCurrentRole) + yearscurrentrole.quartile.bins*3
yearscurrentrole.4th <- min(employee.dB$YearsInCurrentRole) + yearscurrentrole.quartile.bins*4


monthlyincome.quartile.bins <- (max(employee.dB$MonthlyIncome) -  min(employee.dB$MonthlyIncom
e))/4
MonthlyIncome.1st <- min(employee.dB$MonthlyIncome) + monthlyincome.quartile.bins*1
MonthlyIncome.2nd <- min(employee.dB$MonthlyIncome) + monthlyincome.quartile.bins*2
MonthlyIncome.3rd <- min(employee.dB$MonthlyIncome) + monthlyincome.quartile.bins*3
MonthlyIncome.4th <- min(employee.dB$MonthlyIncome) + monthlyincome.quartile.bins*4


yearswithcurrentmanager.quartile.bins <- (max(employee.dB$YearsWithCurrManager) -  min(employee.
dB$YearsWithCurrManager))/4
yearswithcurrentmanager.1st <- min(employee.dB$YearsWithCurrManager) + yearswithcurrentmanager.q
uartile.bins*1
yearswithcurrentmanager.2nd <- min(employee.dB$YearsWithCurrManager) + yearswithcurrentmanager.q
uartile.bins*2
yearswithcurrentmanager.3rd <- min(employee.dB$YearsWithCurrManager) + yearswithcurrentmanager.q
uartile.bins*3
yearswithcurrentmanager.4th <- min(employee.dB$YearsWithCurrManager) + yearswithcurrentmanager.q
uartile.bins*4

environmental.quartile.bins <- (max(employee.dB$EnvironmentSatisfaction) -  min(employee.dB$Envi
ronmentSatisfaction))/4
environmental.1st <- min(employee.dB$EnvironmentSatisfaction) + environmental.quartile.bins*1
environmental.2nd <- min(employee.dB$EnvironmentSatisfaction) + environmental.quartile.bins*2
environmental.3rd <- min(employee.dB$EnvironmentSatisfaction) + environmental.quartile.bins*3
environmental.4th <- min(employee.dB$EnvironmentSatisfaction) + environmental.quartile.bins*4

############ Visuals ############

# Boxplots of Percentage Salary Hike by Years with Company


temp.plot.melt <- employee.dB[,c("PercentSalaryHike", "YearsAtCompany")]

#dev.new()
temp.plot.melt %>% ggplot(aes(x = YearsAtCompany, y = PercentSalaryHike, group = YearsAtCompany,
fill = YearsAtCompany))+
  theme_bw()+
  theme(legend.position="none")+
```

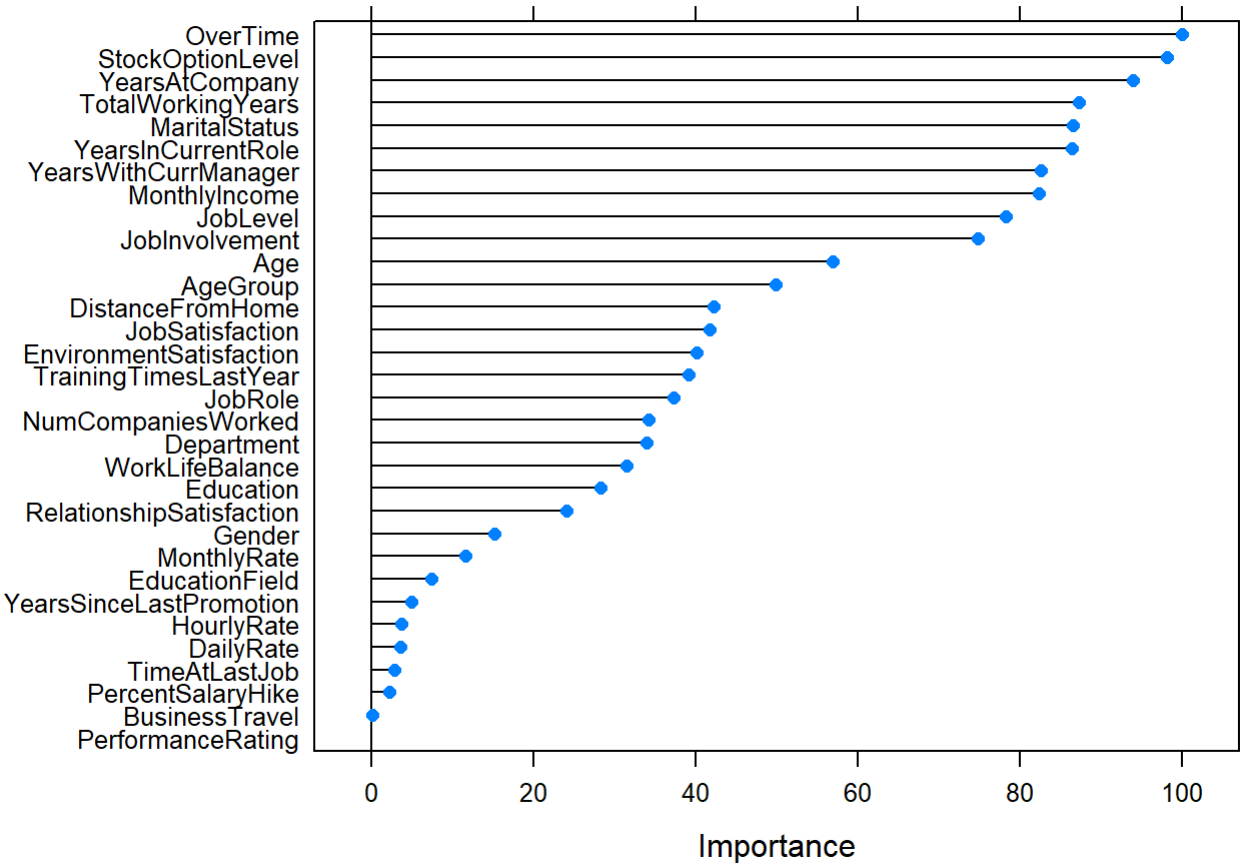
```
  geom_boxplot(outlier.shape = NA, alpha = 0.55)+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.15, color = "grey55", alpha = 0.1)+
  ylab("Salary Hike (%)")+
  xlab("Years At Company")
```




```
# Boxplots of Job Level by Years with Company

temp.plot.melt <- employee.dB[,c("JobLevel", "YearsAtCompany")]

#dev.new()
temp.plot.melt %>% ggplot(aes(x = YearsAtCompany, y = JobLevel, group = YearsAtCompany, fill = Y
earsAtCompany))+
  theme_bw()+
  theme(legend.position="none")+
  geom_boxplot(outlier.shape = NA, alpha = 0.55)+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.15, color = "grey55", alpha = 0.1)+
  ylab("Job Level")+
  xlab("Years At Company")
```




```
################### Naive Bayes variable importance

Grid <- data.frame(usekernel = TRUE, laplace = 0, adjust = 1)

mdl <- train(Attrition~., data=alt.employee.dB_cat_train, method = "naive_bayes", trControl=trai
nControl(method = "none"), tuneGrid=Grid)

mdl.variable.importance <- varImp(mdl)
plot(mdl.variable.importance)
```


OverTime
StockOptionLevel
YearsAtCompany
TotalWorkingYears
MaritalStatus
YearsInCurrentRole
YearsWithCurrManager
MonthlyIncome
JobLevel
JobInvolvement
Age
AgeGroup
DistanceFromHome
JobSatisfaction
EnvironmentSatisfaction
TrainingTimesLastYear
JobRole
NumCompaniesWorked
Department
WorkLifeBalance
Education
RelationshipSatisfaction
Gender
MonthlyRate
EducationField
YearsSinceLastPromotion
HourlyRate
DailyRate
TimeAtLastJob
PercentSalaryHike
BusinessTravel
PerformanceRating
0
20
40
60
80
100
Importance