TeamX Project Changing test

## Introduction

*TeamX* This project was completed by:

* Amy McVicar
* Angela Garcia
* Jennifer Mead
* Jonathan Lee

Businesses lose money when they lose employees. Employee attrition impacts businesses due to the costs of hiring and training new employees. Because of this, data-driven HR departments use data to identify who is likely to quit and to find trends in what factors influence quitting decisions, such as particular departments or locations. [1]

Companies have always been concerned about attrition, but “in many industries the cost of losing good workers is rising” [2]. Exact numbers vary by industry. For example, “[estimates] of annual turnover among U.S. salespeople run as high as 27%—twice the rate in the overall labor force.” [3]

A high attrition rate adds up: “U.S. firms spend $15 billion a year training salespeople and another $800 billion on incentives, and attrition reduces the return on those investments.” [4] In some cases, the cost of losing an employee can be as much as twice their yearly salary. [5]

When employees see other employees leave, attrition can increase. “In settings with high voluntary turnover, employees often lose faith in the company’s strategic direction (because they see others jumping ship), and they tend to be more aware of outside job opportunities, partly because their networks include former colleagues who recently defected. And when there’s lots of involuntary turnover, employees may lack trust in managers, feel little job security, and move on.” [6]

Those costs add up. “It takes an average of 24 days to fill a job, costing employers up to $4,000 per hire– maybe more, depending on your industry.”[7]

Another study “estimates that 42 million, or one in four, employees will leave their jobs in 2018, and that nearly 77 percent, or three-fourths, of that turnover could be prevented by employers.”[8]

Indicators to look for Researchers have found many factors that can be used to identify an increased likelihood of quitting. One study found that these “… include leaving work early, showing less focus or effort, and being reluctant to commit to long-term assignments.” [9]

Another study found that among people who left within the first six months, common issues were: not having clear priorities, a lack of effective training, and not feeling recognized for their contributions. [10]

Some research has been done on specific groups. Executives may have different motivators than sales people. One study identified key factors for executives leaving jobs in less than a year, including pay, a work culture that doesn’t recognize performance, and a lack of synergy among bosses, peers, and direct reports. [11]

Because there are many potential factors that influence voluntary attrition and because there is known variation between industries, roles, and companies, it is useful for companies to analyze their own data to determine patterns in their attrition.

## Analysis and Models

This analysis looks at data from IBM that shows common attrition factors for a fictional company.

Analysis will include using a variety of visualization and machine learning methods and then comparing the results. Combining methods helps to reduce bias [12] and gives a more comprehensive view of the data.

### About the data

Download the data from <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

Before running models on the data, the following steps were performed:

* identify variables to remove because the data is bad or not useful
* set appropriate types for each column (e.g. factor, numeric)
* visualize the variables to give a sense of where to focus analysis
* look for associations/correlations between variables
* perform data transformations for each method, such as creating transactions before using ARM and converting values to numbers before using k-means

#### Load the data

# load in the data  
HR\_original <- read.csv("http://www.creativecubecompany.com/syracuse/ist707/Attrition\_ORIGINAL.csv", fileEncoding ="UTF-8-BOM")

#### Clean the data

Look at the range and typical values for all variables to identify if any should be eliminated due to not being useful.

HR\_clean <- HR\_original  
summary(HR\_clean)

## Age Attrition BusinessTravel DailyRate   
## Min. :18.00 No :1233 Non-Travel : 150 Min. : 102.0   
## 1st Qu.:30.00 Yes: 237 Travel\_Frequently: 277 1st Qu.: 465.0   
## Median :36.00 Travel\_Rarely :1043 Median : 802.0   
## Mean :36.92 Mean : 802.5   
## 3rd Qu.:43.00 3rd Qu.:1157.0   
## Max. :60.00 Max. :1499.0   
##   
## Department DistanceFromHome Education   
## Human Resources : 63 Min. : 1.000 Min. :1.000   
## Research & Development:961 1st Qu.: 2.000 1st Qu.:2.000   
## Sales :446 Median : 7.000 Median :3.000   
## Mean : 9.193 Mean :2.913   
## 3rd Qu.:14.000 3rd Qu.:4.000   
## Max. :29.000 Max. :5.000   
##   
## EducationField EmployeeCount EmployeeNumber EnvironmentSatisfaction  
## Human Resources : 27 Min. :1 Min. : 1.0 Min. :1.000   
## Life Sciences :606 1st Qu.:1 1st Qu.: 491.2 1st Qu.:2.000   
## Marketing :159 Median :1 Median :1020.5 Median :3.000   
## Medical :464 Mean :1 Mean :1024.9 Mean :2.722   
## Other : 82 3rd Qu.:1 3rd Qu.:1555.8 3rd Qu.:4.000   
## Technical Degree:132 Max. :1 Max. :2068.0 Max. :4.000   
##   
## Gender HourlyRate JobInvolvement JobLevel   
## Female:588 Min. : 30.00 Min. :1.00 Min. :1.000   
## Male :882 1st Qu.: 48.00 1st Qu.:2.00 1st Qu.:1.000   
## Median : 66.00 Median :3.00 Median :2.000   
## Mean : 65.89 Mean :2.73 Mean :2.064   
## 3rd Qu.: 83.75 3rd Qu.:3.00 3rd Qu.:3.000   
## Max. :100.00 Max. :4.00 Max. :5.000   
##   
## JobRole JobSatisfaction MaritalStatus MonthlyIncome   
## Sales Executive :326 Min. :1.000 Divorced:327 Min. : 1009   
## Research Scientist :292 1st Qu.:2.000 Married :673 1st Qu.: 2911   
## Laboratory Technician :259 Median :3.000 Single :470 Median : 4919   
## Manufacturing Director :145 Mean :2.729 Mean : 6503   
## Healthcare Representative:131 3rd Qu.:4.000 3rd Qu.: 8379   
## Manager :102 Max. :4.000 Max. :19999   
## (Other) :215   
## MonthlyRate NumCompaniesWorked Over18 OverTime PercentSalaryHike  
## Min. : 2094 Min. :0.000 Y:1470 No :1054 Min. :11.00   
## 1st Qu.: 8047 1st Qu.:1.000 Yes: 416 1st Qu.:12.00   
## Median :14236 Median :2.000 Median :14.00   
## Mean :14313 Mean :2.693 Mean :15.21   
## 3rd Qu.:20462 3rd Qu.:4.000 3rd Qu.:18.00   
## Max. :26999 Max. :9.000 Max. :25.00   
##   
## PerformanceRating RelationshipSatisfaction StandardHours StockOptionLevel  
## Min. :3.000 Min. :1.000 Min. :80 Min. :0.0000   
## 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:80 1st Qu.:0.0000   
## Median :3.000 Median :3.000 Median :80 Median :1.0000   
## Mean :3.154 Mean :2.712 Mean :80 Mean :0.7939   
## 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:80 3rd Qu.:1.0000   
## Max. :4.000 Max. :4.000 Max. :80 Max. :3.0000   
##   
## TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany   
## Min. : 0.00 Min. :0.000 Min. :1.000 Min. : 0.000   
## 1st Qu.: 6.00 1st Qu.:2.000 1st Qu.:2.000 1st Qu.: 3.000   
## Median :10.00 Median :3.000 Median :3.000 Median : 5.000   
## Mean :11.28 Mean :2.799 Mean :2.761 Mean : 7.008   
## 3rd Qu.:15.00 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.: 9.000   
## Max. :40.00 Max. :6.000 Max. :4.000 Max. :40.000   
##   
## YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager  
## Min. : 0.000 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 2.000 1st Qu.: 0.000 1st Qu.: 2.000   
## Median : 3.000 Median : 1.000 Median : 3.000   
## Mean : 4.229 Mean : 2.188 Mean : 4.123   
## 3rd Qu.: 7.000 3rd Qu.: 3.000 3rd Qu.: 7.000   
## Max. :18.000 Max. :15.000 Max. :17.000   
##

str(HR\_clean)

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

Actions:

* remove EmployeeCount because it is always 1
* remove Over18 because it is always Y
* remove StandardHours because it is always 80
* change the name of i..Age to fix a typo

# reference-- drop columns by name: https://stackoverflow.com/questions/5234117/how-to-drop-columns-by-name-in-a-data-frame  
# reference -- move column to the first column: https://stackoverflow.com/questions/22286419/move-a-column-to-first-position-in-a-data-frame  
  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

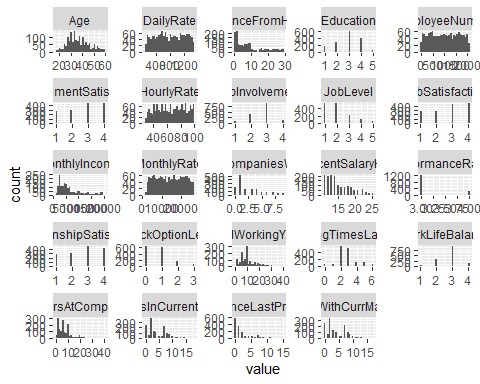
HR\_clean <- subset(HR\_clean, select=-c(EmployeeCount, StandardHours, Over18))  
  
HR\_clean <- HR\_clean %>%  
 select(EmployeeNumber, everything())  
head(HR\_clean, 10)

## EmployeeNumber Age Attrition BusinessTravel DailyRate  
## 1 1 41 Yes Travel\_Rarely 1102  
## 2 2 49 No Travel\_Frequently 279  
## 3 4 37 Yes Travel\_Rarely 1373  
## 4 5 33 No Travel\_Frequently 1392  
## 5 7 27 No Travel\_Rarely 591  
## 6 8 32 No Travel\_Frequently 1005  
## 7 10 59 No Travel\_Rarely 1324  
## 8 11 30 No Travel\_Rarely 1358  
## 9 12 38 No Travel\_Frequently 216  
## 10 13 36 No Travel\_Rarely 1299  
## Department DistanceFromHome Education EducationField  
## 1 Sales 1 2 Life Sciences  
## 2 Research & Development 8 1 Life Sciences  
## 3 Research & Development 2 2 Other  
## 4 Research & Development 3 4 Life Sciences  
## 5 Research & Development 2 1 Medical  
## 6 Research & Development 2 2 Life Sciences  
## 7 Research & Development 3 3 Medical  
## 8 Research & Development 24 1 Life Sciences  
## 9 Research & Development 23 3 Life Sciences  
## 10 Research & Development 27 3 Medical  
## EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel  
## 1 2 Female 94 3 2  
## 2 3 Male 61 2 2  
## 3 4 Male 92 2 1  
## 4 4 Female 56 3 1  
## 5 1 Male 40 3 1  
## 6 4 Male 79 3 1  
## 7 3 Female 81 4 1  
## 8 4 Male 67 3 1  
## 9 4 Male 44 2 3  
## 10 3 Male 94 3 2  
## JobRole JobSatisfaction MaritalStatus MonthlyIncome  
## 1 Sales Executive 4 Single 5993  
## 2 Research Scientist 2 Married 5130  
## 3 Laboratory Technician 3 Single 2090  
## 4 Research Scientist 3 Married 2909  
## 5 Laboratory Technician 2 Married 3468  
## 6 Laboratory Technician 4 Single 3068  
## 7 Laboratory Technician 1 Married 2670  
## 8 Laboratory Technician 3 Divorced 2693  
## 9 Manufacturing Director 3 Single 9526  
## 10 Healthcare Representative 3 Married 5237  
## MonthlyRate NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating  
## 1 19479 8 Yes 11 3  
## 2 24907 1 No 23 4  
## 3 2396 6 Yes 15 3  
## 4 23159 1 Yes 11 3  
## 5 16632 9 No 12 3  
## 6 11864 0 No 13 3  
## 7 9964 4 Yes 20 4  
## 8 13335 1 No 22 4  
## 9 8787 0 No 21 4  
## 10 16577 6 No 13 3  
## RelationshipSatisfaction StockOptionLevel TotalWorkingYears  
## 1 1 0 8  
## 2 4 1 10  
## 3 2 0 7  
## 4 3 0 8  
## 5 4 1 6  
## 6 3 0 8  
## 7 1 3 12  
## 8 2 1 1  
## 9 2 0 10  
## 10 2 2 17  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0 1 6 4  
## 2 3 3 10 7  
## 3 3 3 0 0  
## 4 3 3 8 7  
## 5 3 3 2 2  
## 6 2 2 7 7  
## 7 3 2 1 0  
## 8 2 3 1 0  
## 9 2 3 9 7  
## 10 3 2 7 7  
## YearsSinceLastPromotion YearsWithCurrManager  
## 1 0 5  
## 2 1 7  
## 3 0 0  
## 4 3 0  
## 5 2 2  
## 6 3 6  
## 7 0 0  
## 8 0 0  
## 9 1 8  
## 10 7 7

Look at histograms of all numeric variables to identify which should be categorical instead

# reference-- histogram of all variables: https://drsimonj.svbtle.com/quick-plot-of-all-variables  
library(purrr)  
library(tidyr)  
library(ggplot2)  
  
HR\_clean %>%  
 keep(is.numeric) %>%   
 gather() %>%   
 ggplot(aes(value)) +  
 facet\_wrap(~ key, scales = "free") +  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 Histograms with parallel lines instead of of distributions with close bins are typically factors. It looks like the following columns are actually factors instead of integers. Some models need numerical inputs and some need factors, but for the moment they should be converted.

Numerical columns that should be factors:

* Education
* EnvironmentSatisfaction
* JobInvolvement
* JobLevel
* JobSatisfaction
* PerformanceRating
* RelationshipSatisfaction
* StockOptionLevel
* WorkLifeBalance

HR\_clean$Education <- as.factor(HR\_clean$Education)  
HR\_clean$EnvironmentSatisfaction <- as.factor(HR\_clean$EnvironmentSatisfaction)  
HR\_clean$JobInvolvement <- as.factor(HR\_clean$JobInvolvement)  
HR\_clean$JobLevel <- as.factor(HR\_clean$JobLevel)  
HR\_clean$JobSatisfaction <- as.factor(HR\_clean$JobSatisfaction)  
HR\_clean$PerformanceRating <- as.factor(HR\_clean$PerformanceRating)  
HR\_clean$RelationshipSatisfaction <- as.factor(HR\_clean$RelationshipSatisfaction)  
HR\_clean$StockOptionLevel <- as.factor(HR\_clean$StockOptionLevel)  
HR\_clean$WorkLifeBalance <- as.factor(HR\_clean$WorkLifeBalance)  
  
head(HR\_clean)

## EmployeeNumber Age Attrition BusinessTravel DailyRate  
## 1 1 41 Yes Travel\_Rarely 1102  
## 2 2 49 No Travel\_Frequently 279  
## 3 4 37 Yes Travel\_Rarely 1373  
## 4 5 33 No Travel\_Frequently 1392  
## 5 7 27 No Travel\_Rarely 591  
## 6 8 32 No Travel\_Frequently 1005  
## Department DistanceFromHome Education EducationField  
## 1 Sales 1 2 Life Sciences  
## 2 Research & Development 8 1 Life Sciences  
## 3 Research & Development 2 2 Other  
## 4 Research & Development 3 4 Life Sciences  
## 5 Research & Development 2 1 Medical  
## 6 Research & Development 2 2 Life Sciences  
## EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel  
## 1 2 Female 94 3 2  
## 2 3 Male 61 2 2  
## 3 4 Male 92 2 1  
## 4 4 Female 56 3 1  
## 5 1 Male 40 3 1  
## 6 4 Male 79 3 1  
## JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate  
## 1 Sales Executive 4 Single 5993 19479  
## 2 Research Scientist 2 Married 5130 24907  
## 3 Laboratory Technician 3 Single 2090 2396  
## 4 Research Scientist 3 Married 2909 23159  
## 5 Laboratory Technician 2 Married 3468 16632  
## 6 Laboratory Technician 4 Single 3068 11864  
## NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating  
## 1 8 Yes 11 3  
## 2 1 No 23 4  
## 3 6 Yes 15 3  
## 4 1 Yes 11 3  
## 5 9 No 12 3  
## 6 0 No 13 3  
## RelationshipSatisfaction StockOptionLevel TotalWorkingYears  
## 1 1 0 8  
## 2 4 1 10  
## 3 2 0 7  
## 4 3 0 8  
## 5 4 1 6  
## 6 3 0 8  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0 1 6 4  
## 2 3 3 10 7  
## 3 3 3 0 0  
## 4 3 3 8 7  
## 5 3 3 2 2  
## 6 2 2 7 7  
## YearsSinceLastPromotion YearsWithCurrManager  
## 1 0 5  
## 2 1 7  
## 3 0 0  
## 4 3 0  
## 5 2 2  
## 6 3 6

Check for blanks

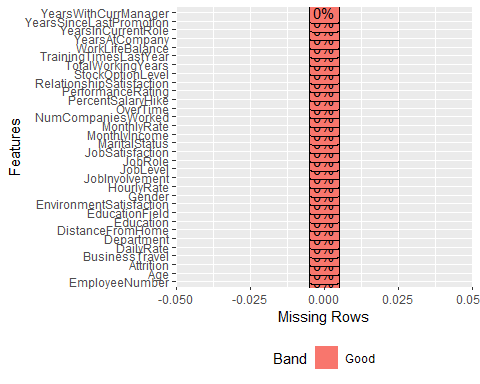
#reference -- checking for blanks: https://stackoverflow.com/questions/40715508/r-count-cells-with-missing-values-across-each-row  
  
colSums(is.na(HR\_clean) | HR\_clean == "" | HR\_clean == " ")

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

#### Visualize the variables

There are 32 variables in total. We can check again for any missing variables, and there are none.

if("DataExplorer" %in% rownames(installed.packages()) == FALSE) {install.packages('DataExplorer') }  
library(DataExplorer)  
HR\_eda <- HR\_clean  
plot\_str(HR\_eda)  
plot\_missing(HR\_eda)



From correlating the attributes we can see pockets of correlation.

Most notably are:

Years with Current Manager

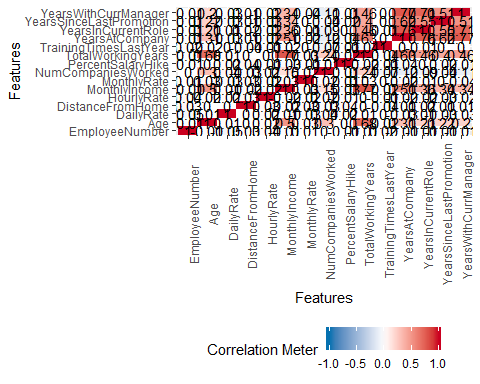
Years Since Last Promotion

Years in Current Role

Years at Company

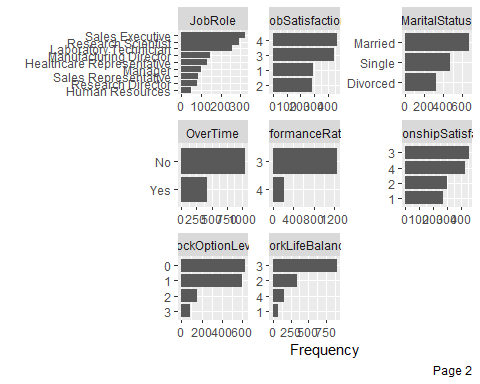
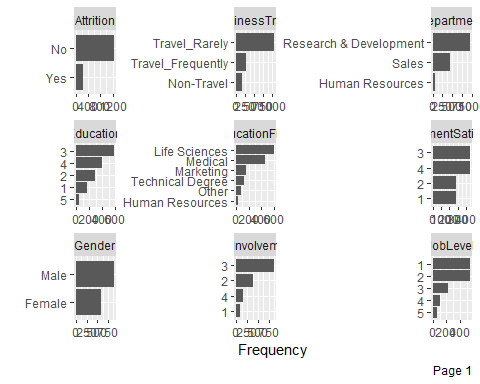
And no surprise, these correlate with Age, Income, and Total Working Years.

plot\_correlation(HR\_eda, type = 'continuous')



Simple barcharts of the attributes show us some interesting facts that we can use for deeper analysis. For example, most of the universe is ‘no’ to attrition. The Education Field and Department are limited in the selections available. This might help us understand the context of the findings of models. For example, there are only three department types (R&D, Sales, & HR). We might find that the weight of this attribute in models may only be relevant to this limited dataset and not as applicable to datasets that are more representative of real organizations. This is something we might not notice without this simple exploratory examination of the data first.

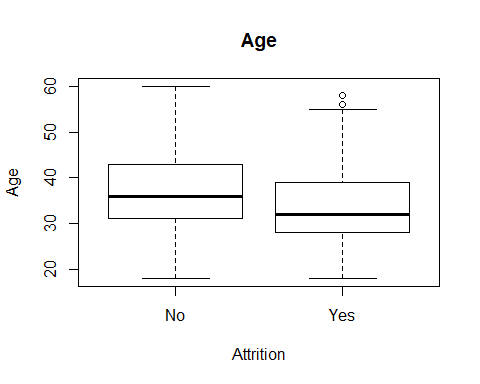
plot\_bar(HR\_eda)



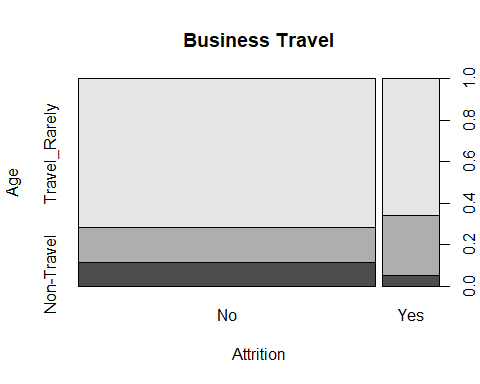
#create\_report(HR\_eda)

Each variable, except EmployeeNumber, in the data set is examined for significant variance in the attrition yes versus no segments using simple analysis and plotting.

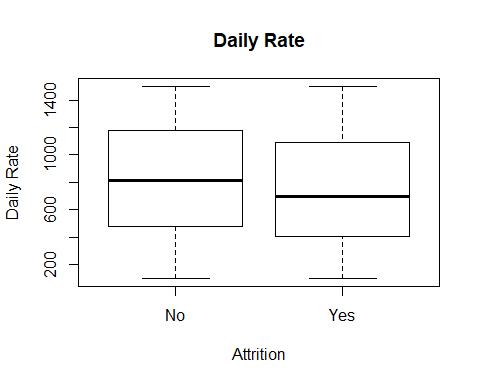
plot(HR\_eda$Attrition, HR\_eda$Age, main = "Age", ylab = "Age", xlab = "Attrition")



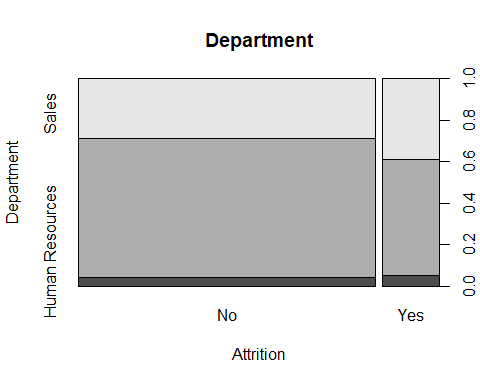
plot(HR\_eda$Attrition, HR\_eda$BusinessTravel, main = "Business Travel", ylab = "Age", xlab = "Attrition")



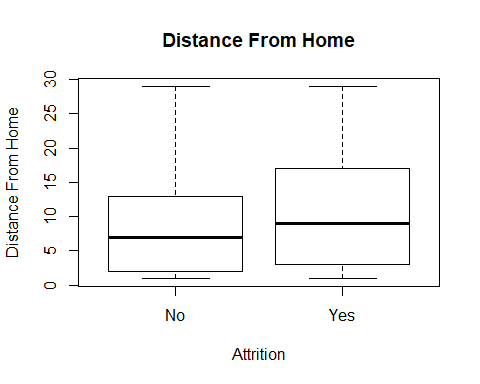
plot(HR\_eda$Attrition, HR\_eda$DailyRate, main = "Daily Rate", ylab = "Daily Rate", xlab = "Attrition")



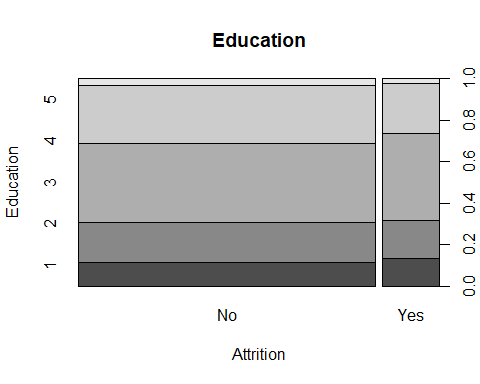
plot(HR\_eda$Attrition, HR\_eda$Department, main = "Department", ylab = "Department", xlab = "Attrition")



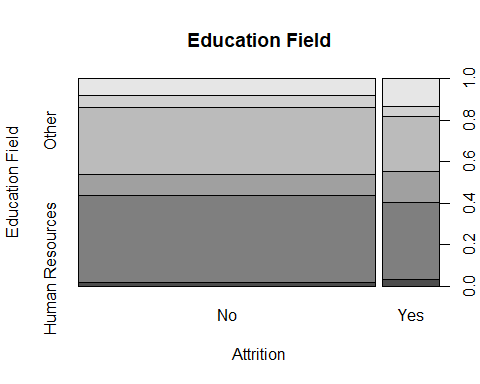
plot(HR\_eda$Attrition, HR\_eda$DistanceFromHome, main = "Distance From Home", ylab = "Distance From Home", xlab = "Attrition")



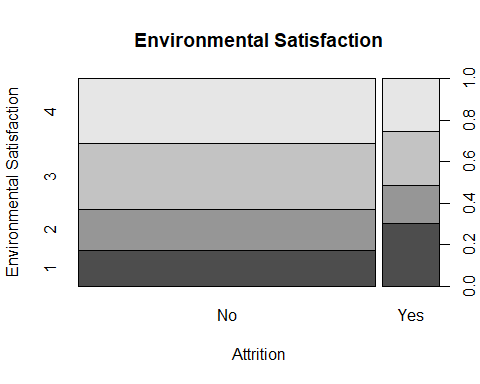
plot(HR\_eda$Attrition, HR\_eda$Education, main = "Education", ylab = "Education", xlab = "Attrition")



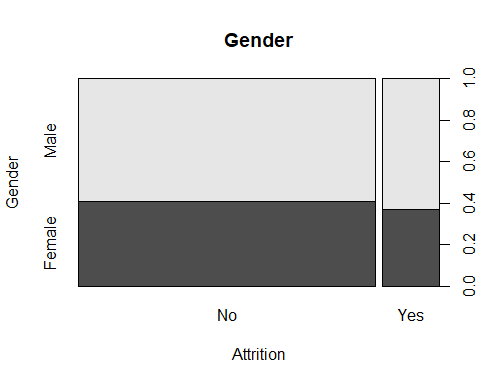
plot(HR\_eda$Attrition, HR\_eda$EducationField, main = "Education Field", ylab = "Education Field", xlab = "Attrition")



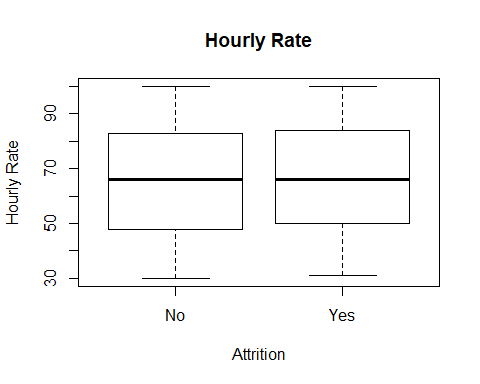
plot(HR\_eda$Attrition, HR\_eda$EnvironmentSatisfaction, main = "Environmental Satisfaction", ylab = "Environmental Satisfaction", xlab = "Attrition")



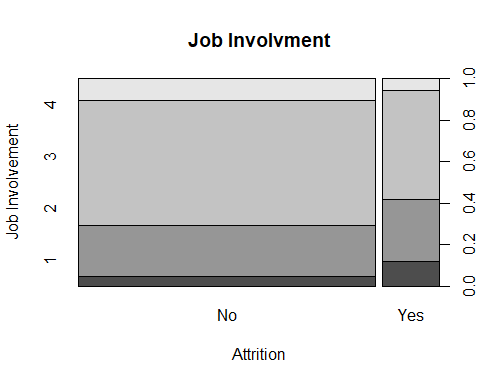
plot(HR\_eda$Attrition, HR\_eda$Gender, main = "Gender", ylab = "Gender", xlab = "Attrition")



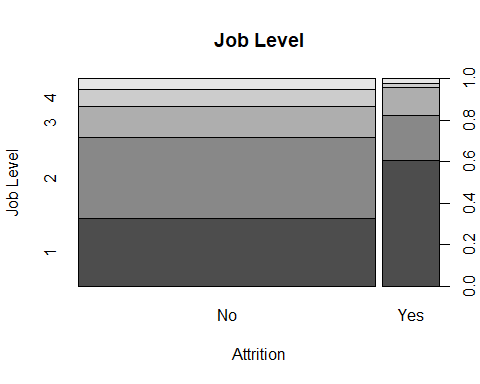
plot(HR\_eda$Attrition, HR\_eda$HourlyRate, main = "Hourly Rate", ylab = "Hourly Rate", xlab = "Attrition")



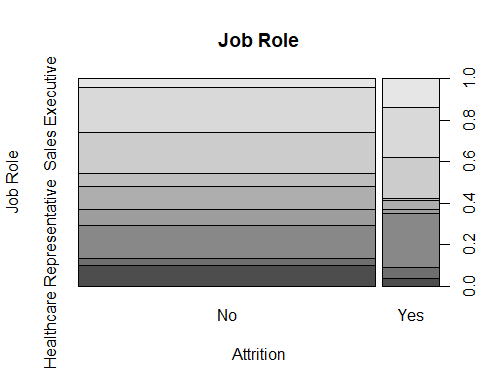
plot(HR\_eda$Attrition, HR\_eda$JobInvolvement, main = "Job Involvment", ylab = "Job Involvement", xlab = "Attrition")



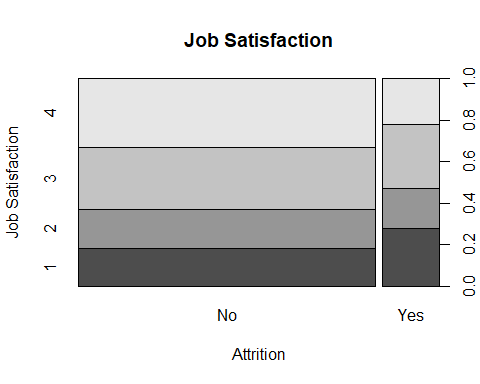
plot(HR\_eda$Attrition, HR\_eda$JobLevel, main = "Job Level", ylab = "Job Level", xlab = "Attrition")



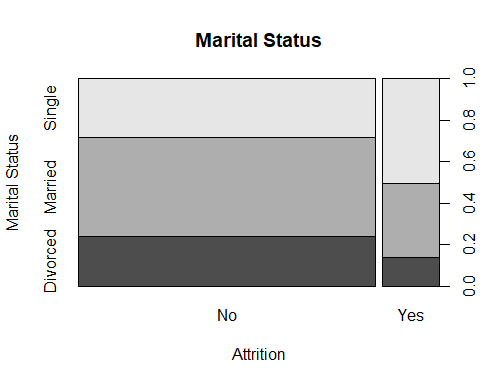
plot(HR\_eda$Attrition, HR\_eda$JobRole, main = "Job Role", ylab = "Job Role", xlab = "Attrition")



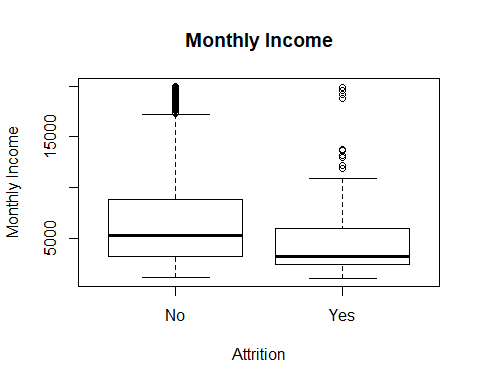
plot(HR\_eda$Attrition, HR\_eda$JobSatisfaction, main = "Job Satisfaction", ylab = "Job Satisfaction", xlab = "Attrition")



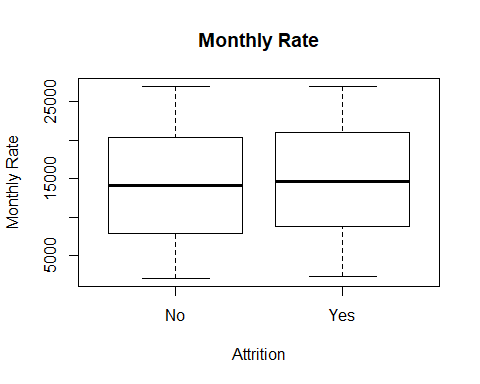
plot(HR\_eda$Attrition, HR\_eda$MaritalStatus, main = "Marital Status", ylab = "Marital Status", xlab = "Attrition")



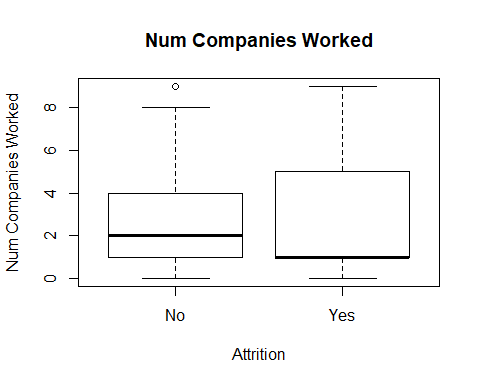
plot(HR\_eda$Attrition, HR\_eda$MonthlyIncome, main = "Monthly Income", ylab = "Monthly Income", xlab = "Attrition")



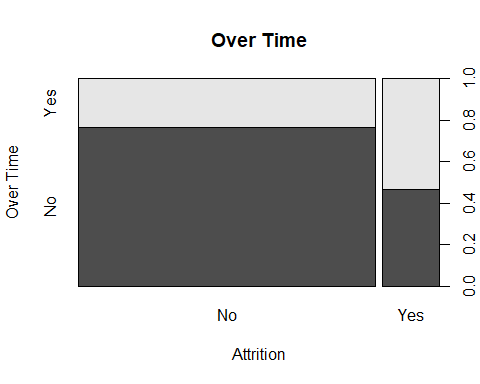
plot(HR\_eda$Attrition, HR\_eda$MonthlyRate, main = "Monthly Rate", ylab = "Monthly Rate", xlab = "Attrition")



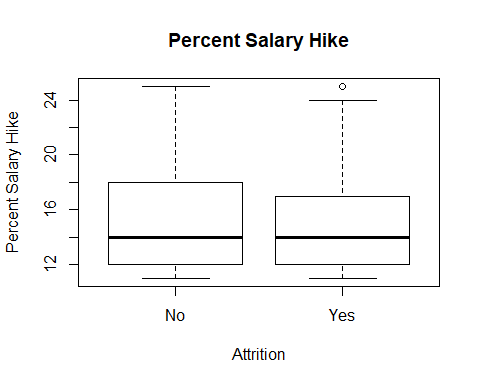
plot(HR\_eda$Attrition, HR\_eda$NumCompaniesWorked, main = "Num Companies Worked", ylab = "Num Companies Worked", xlab = "Attrition")



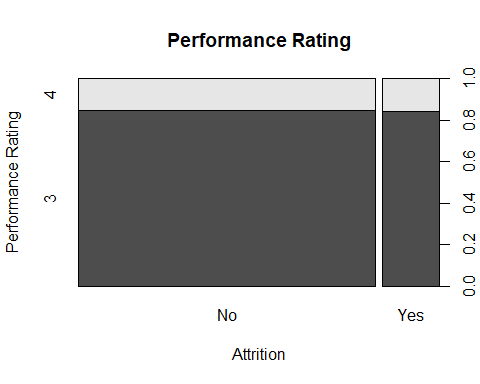
plot(HR\_eda$Attrition, HR\_eda$OverTime, main = "Over Time", ylab = "Over Time", xlab = "Attrition")



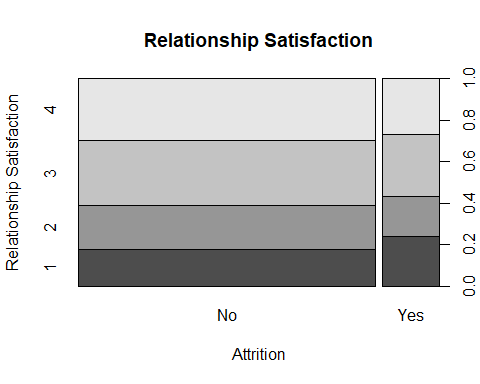
plot(HR\_eda$Attrition, HR\_eda$PercentSalaryHike, main = "Percent Salary Hike", ylab = "Percent Salary Hike", xlab = "Attrition")



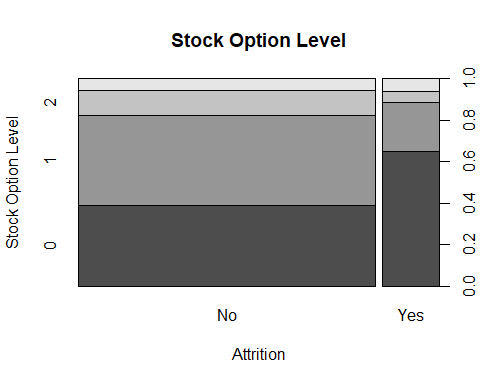
plot(HR\_eda$Attrition, HR\_eda$PerformanceRating, main = "Performance Rating", ylab = "Performance Rating", xlab = "Attrition")



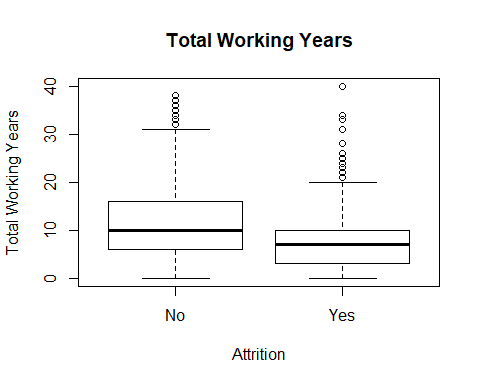
plot(HR\_eda$Attrition, HR\_eda$RelationshipSatisfaction, main = "Relationship Satisfaction", ylab = "Relationship Satisfaction", xlab = "Attrition")



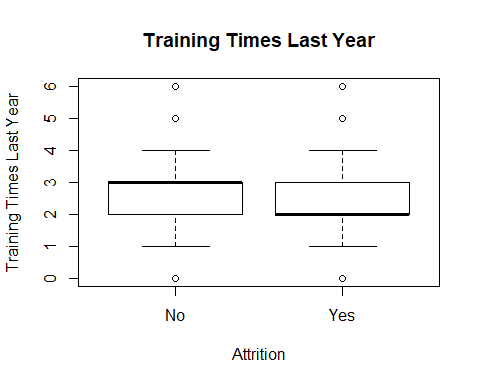
plot(HR\_eda$Attrition, HR\_eda$StockOptionLevel, main = "Stock Option Level", ylab = "Stock Option Level", xlab = "Attrition")



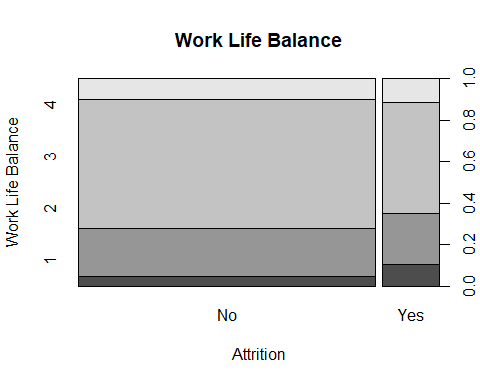
plot(HR\_eda$Attrition, HR\_eda$TotalWorkingYears, main = "Total Working Years", ylab = "Total Working Years", xlab = "Attrition")



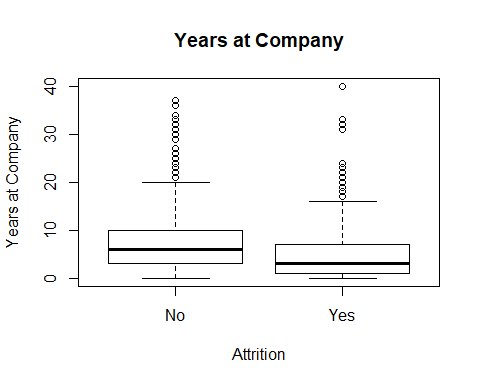
plot(HR\_eda$Attrition, HR\_eda$TrainingTimesLastYear, main = "Training Times Last Year", ylab = "Training Times Last Year", xlab = "Attrition")



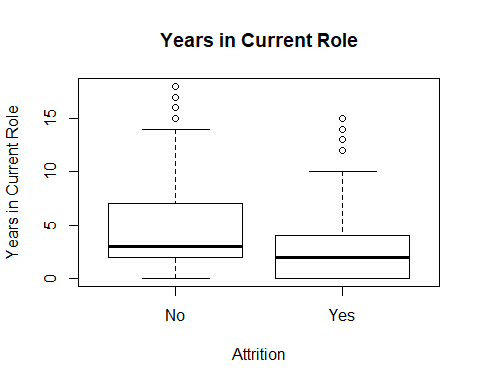
plot(HR\_eda$Attrition, HR\_eda$WorkLifeBalance, main = "Work Life Balance", ylab = "Work Life Balance", xlab = "Attrition")



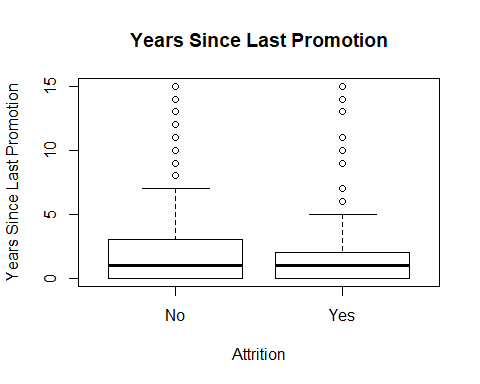
plot(HR\_eda$Attrition, HR\_eda$YearsAtCompany, main = "Years at Company", ylab = "Years at Company", xlab = "Attrition")



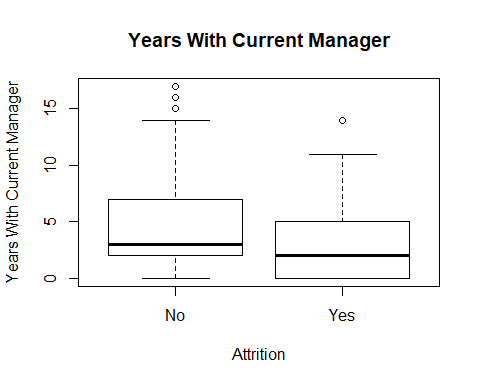
plot(HR\_eda$Attrition, HR\_eda$YearsInCurrentRole, main = "Years in Current Role", ylab = "Years in Current Role", xlab = "Attrition")



plot(HR\_eda$Attrition, HR\_eda$YearsSinceLastPromotion, main = "Years Since Last Promotion", ylab = "Years Since Last Promotion", xlab = "Attrition")



plot(HR\_eda$Attrition, HR\_eda$YearsWithCurrManager, main = "Years With Current Manager", ylab = "Years With Current Manager", xlab = "Attrition")

 On visual inspection the following variables appear to have a significant difference in the attrition yes and no segments:

EnvironmentalSatisfaction JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome NumCompaniesWorked OverTime RelationshipSatisfaction

On initial visual analysis and inspection, the following attributes may have significance:

StopOptionLevel TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentCompany YearsInCurrentRole YearsWithCurrentManager

Additionally, initial inspection shows that more than a few attributes appear to be highly correlated with each other. This information may be used for further analysis and refining the attributes used in models for simplification.

#Install packages if they dont exist  
  
if("formattable" %in% rownames(installed.packages()) == FALSE) {install.packages("formattable")}  
library(formattable)  
  
if("gridExtra" %in% rownames(installed.packages()) == FALSE) {install.packages("gridExtra")}  
library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

if("grid" %in% rownames(installed.packages()) == FALSE) {install.packages("grid")}  
library(grid)  
  
if("corrplot" %in% rownames(installed.packages()) == FALSE) {install.packages("corrplot")}  
library(corrplot)

## corrplot 0.84 loaded

if("rquery" %in% rownames(installed.packages()) == FALSE) {install.packages("rquery")}  
library(rquery)

## Loading required package: wrapr

##   
## Attaching package: 'wrapr'

## The following object is masked from 'package:tidyr':  
##   
## unpack

## The following object is masked from 'package:dplyr':  
##   
## coalesce

##   
## Attaching package: 'rquery'

## The following object is masked from 'package:grid':  
##   
## arrow

## The following object is masked from 'package:DataExplorer':  
##   
## drop\_columns

## The following object is masked from 'package:ggplot2':  
##   
## arrow

## The following object is masked from 'package:tidyr':  
##   
## expand\_grid

if("GoodmanKruskal" %in% rownames(installed.packages()) == FALSE) {install.packages("GoodmanKruskal")}  
library(GoodmanKruskal)

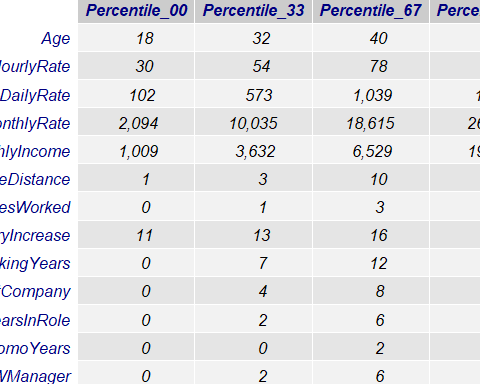
# Data Transformation  
  
# Data Assessment  
HR\_linear<-HR\_clean  
  
#Create Categories for numeric values with high number of records (Based on Percentiles)  
  
## Categoric Age  
  
# Age Percentiles  
Percentile\_00 = min(HR\_linear$Age)  
Percentile\_33 = quantile(HR\_linear$Age, 0.33333)  
Percentile\_67 = quantile(HR\_linear$Age, 0.66667)  
Percentile\_100 = max(HR\_linear$Age)  
  
# Values  
HR.BindA = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindA)[[2]] = "Value"  
#HR.BindA  
  
#Age:   
HR\_linear$AgeRange[HR\_linear$Age >= Percentile\_00 & HR\_linear$Age < Percentile\_33] = "Lower\_Range"  
HR\_linear$AgeRange[HR\_linear$Age >= Percentile\_33 & HR\_linear$Age < Percentile\_67] = "Mid\_Range"  
HR\_linear$AgeRange[HR\_linear$Age >= Percentile\_67 & HR\_linear$Age <= Percentile\_100] = "Higher\_Range"  
  
## Categoric Hourly Rate  
  
# Hourly Rate Percentiles  
Percentile\_00 = min(HR\_linear$HourlyRate)  
Percentile\_33 = quantile(HR\_linear$HourlyRate, 0.33333)  
Percentile\_67 = quantile(HR\_linear$HourlyRate, 0.66667)  
Percentile\_100 = max(HR\_linear$HourlyRate)  
  
# Values  
HR.BindH = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindH)[[2]] = "Value"  
#HR.BindH  
  
#Hourly Rate Ranges:   
HR\_linear$HourlyRateRange[HR\_linear$HourlyRate >= Percentile\_00 & HR\_linear$HourlyRate < Percentile\_33] = "Low\_Range"  
HR\_linear$HourlyRateRange[HR\_linear$HourlyRate >= Percentile\_33 & HR\_linear$HourlyRate < Percentile\_67] = "Mid\_Range"  
HR\_linear$HourlyRateRange[HR\_linear$HourlyRate >= Percentile\_67 & HR\_linear$HourlyRate <= Percentile\_100] = "High\_Range"  
  
## Categoric Daily Rate  
  
# Daily Rate Percentiles  
Percentile\_00 = min(HR\_linear$DailyRate)  
Percentile\_33 = quantile(HR\_linear$DailyRate, 0.33333)  
Percentile\_67 = quantile(HR\_linear$DailyRate, 0.66667)  
Percentile\_100 = max(HR\_linear$DailyRate)  
  
# Values  
HR.BindDR = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindDR)[[2]] = "Value"  
#HR.BindDR  
  
# Daily Rate Ranges:   
HR\_linear$DailyRateRange[HR\_linear$DailyRate >= Percentile\_00 & HR\_linear$DailyRate < Percentile\_33] = "Low\_Range"  
HR\_linear$DailyRateRange[HR\_linear$DailyRate >= Percentile\_33 & HR\_linear$DailyRate < Percentile\_67] = "Mid\_Range"  
HR\_linear$DailyRateRange[HR\_linear$DailyRate >= Percentile\_67 & HR\_linear$DailyRate <= Percentile\_100] = "High\_Range"  
  
  
## Categoric Monthly Rate  
  
# Monthly Rate Percentiles  
Percentile\_00 = min(HR\_linear$MonthlyRate)  
Percentile\_33 = quantile(HR\_linear$MonthlyRate, 0.33333)  
Percentile\_67 = quantile(HR\_linear$MonthlyRate, 0.66667)  
Percentile\_100 = max(HR\_linear$MonthlyRate)  
  
# Values  
HR.BindMR = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindMR)[[2]] = "Value"  
#HR.BindMR  
  
# Monthly Rate Level  
HR\_linear$MonthRateLevel[HR\_linear$MonthlyRate >= Percentile\_00 & HR\_linear$MonthlyRate < Percentile\_33] = "Low\_Income"  
HR\_linear$MonthRateLevel[HR\_linear$MonthlyRate >= Percentile\_33 & HR\_linear$MonthlyRate < Percentile\_67] = "Mid\_Income"  
HR\_linear$MonthRateLevel[HR\_linear$MonthlyRate >= Percentile\_67 & HR\_linear$MonthlyRate <= Percentile\_100] = "High\_Income"  
  
  
# Categoric Monthly Income  
  
# Monthly Income Percentiles  
Percentile\_00 = min(HR\_linear$MonthlyIncome)  
Percentile\_33 = quantile(HR\_linear$MonthlyIncome, 0.33333)  
Percentile\_67 = quantile(HR\_linear$MonthlyIncome, 0.66667)  
Percentile\_100 = max(HR\_linear$MonthlyIncome)  
  
# Values  
HR.BindI = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindI)[[2]] = "Value"  
#HR.BindI  
  
# Monthly Income Level  
HR\_linear$MonthIncomeLevel[HR\_linear$MonthlyIncome >= Percentile\_00 & HR\_linear$MonthlyIncome < Percentile\_33] = "Low\_Income"  
HR\_linear$MonthIncomeLevel[HR\_linear$MonthlyIncome >= Percentile\_33 & HR\_linear$MonthlyIncome < Percentile\_67] = "Mid\_Income"  
HR\_linear$MonthIncomeLevel[HR\_linear$MonthlyIncome >= Percentile\_67 & HR\_linear$MonthlyIncome <= Percentile\_100] = "High\_Income"  
  
# Categoric Distance From Home  
  
# Distance From Home Percentiles  
Percentile\_00 = min(HR\_linear$DistanceFromHome)  
Percentile\_33 = quantile(HR\_linear$DistanceFromHome, 0.33333)  
Percentile\_67 = quantile(HR\_linear$DistanceFromHome, 0.66667)  
Percentile\_100 = max(HR\_linear$DistanceFromHome)  
  
# Values  
HR.BindD = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindD)[[2]] = "Value"  
#HR.BindD  
  
# Distance From Home Ranges:   
HR\_linear$DistHomeRange[HR\_linear$DistanceFromHome >= Percentile\_00 & HR\_linear$DistanceFromHome < Percentile\_33] = "Low\_Distance"  
HR\_linear$DistHomeRange[HR\_linear$DistanceFromHome >= Percentile\_33 & HR\_linear$DistanceFromHome < Percentile\_67] = "Mid\_Distance"  
HR\_linear$DistHomeRange[HR\_linear$DistanceFromHome >= Percentile\_67 & HR\_linear$DistanceFromHome <= Percentile\_100] = "High\_Distance"  
  
  
# Categoric Number of Companies Worked  
  
# Number of Companies worked Percentiles  
Percentile\_00 = min(HR\_linear$NumCompaniesWorked)  
Percentile\_33 = quantile(HR\_linear$NumCompaniesWorked, 0.33333)  
Percentile\_67 = quantile(HR\_linear$NumCompaniesWorked, 0.66667)  
Percentile\_100 = max(HR\_linear$NumCompaniesWorked)  
  
# Values  
HR.BindC = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindC)[[2]] = "Value"  
#HR.BindC  
  
# Number of Companies worked Ranges:   
HR\_linear$NumCompWorked[HR\_linear$NumCompaniesWorked >= Percentile\_00 & HR\_linear$NumCompaniesWorked < Percentile\_33] = "Low\_Number"  
HR\_linear$NumCompWorked[HR\_linear$NumCompaniesWorked >= Percentile\_33 & HR\_linear$NumCompaniesWorked < Percentile\_67] = "Mid\_Number"  
HR\_linear$NumCompWorked[HR\_linear$NumCompaniesWorked >= Percentile\_67 & HR\_linear$NumCompaniesWorked <= Percentile\_100] = "High\_Number"  
  
# Categoric Salary Increase  
  
# Salary Increase Percentiles  
Percentile\_00 = min(HR\_linear$PercentSalaryHike)  
Percentile\_33 = quantile(HR\_linear$PercentSalaryHike, 0.33333)  
Percentile\_67 = quantile(HR\_linear$PercentSalaryHike, 0.66667)  
Percentile\_100 = max(HR\_linear$PercentSalaryHike)  
  
# Values  
HR.BindS = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindS)[[2]] = "Value"  
#HR.BindS  
  
# Salary Increase worked Ranges:   
HR\_linear$SalaryIncreaseLevel[HR\_linear$PercentSalaryHike >= Percentile\_00 & HR\_linear$PercentSalaryHike < Percentile\_33] = "Low\_Increase"  
HR\_linear$SalaryIncreaseLevel[HR\_linear$PercentSalaryHike >= Percentile\_33 & HR\_linear$PercentSalaryHike < Percentile\_67] = "Avg\_Increase"  
HR\_linear$SalaryIncreaseLevel[HR\_linear$PercentSalaryHike >= Percentile\_67 & HR\_linear$PercentSalaryHike <= Percentile\_100] = "High\_Increase"  
  
# Categoric Working Years  
  
# Working Years Percentiles  
Percentile\_00 = min(HR\_linear$TotalWorkingYears)  
Percentile\_33 = quantile(HR\_linear$TotalWorkingYears, 0.33333)  
Percentile\_67 = quantile(HR\_linear$TotalWorkingYears, 0.66667)  
Percentile\_100 = max(HR\_linear$TotalWorkingYears)  
  
# Values  
HR.BindW = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindW)[[2]] = "Value"  
#HR.BindW  
  
# Working Years Ranges:   
HR\_linear$WorkingYears[HR\_linear$TotalWorkingYears >= Percentile\_00 & HR\_linear$TotalWorkingYears < Percentile\_33] = "Lower\_Range"  
HR\_linear$WorkingYears[HR\_linear$TotalWorkingYears >= Percentile\_33 & HR\_linear$TotalWorkingYears < Percentile\_67] = "Mid\_Range"  
HR\_linear$WorkingYears[HR\_linear$TotalWorkingYears >= Percentile\_67 & HR\_linear$TotalWorkingYears <= Percentile\_100] = "Higher\_Range"  
  
# Categoric Years At Company  
  
# Years At Company Percentiles  
Percentile\_00 = min(HR\_linear$YearsAtCompany)  
Percentile\_33 = quantile(HR\_linear$YearsAtCompany, 0.33333)  
Percentile\_67 = quantile(HR\_linear$YearsAtCompany, 0.66667)  
Percentile\_100 = max(HR\_linear$YearsAtCompany)  
  
# Values  
HR.BindY = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindY)[[2]] = "Value"  
#HR.BindY  
  
# Years At Company Ranges:   
HR\_linear$CompanyYears[HR\_linear$YearsAtCompany >= Percentile\_00 & HR\_linear$YearsAtCompany < Percentile\_33] = "Lower\_Range"  
HR\_linear$CompanyYears[HR\_linear$YearsAtCompany >= Percentile\_33 & HR\_linear$YearsAtCompany < Percentile\_67] = "Mid\_Range"  
HR\_linear$CompanyYears[HR\_linear$YearsAtCompany >= Percentile\_67 & HR\_linear$YearsAtCompany <= Percentile\_100] = "Higher\_Range"  
  
# Categoric Years in Current Role  
  
# Years in Current Role Percentiles  
Percentile\_00 = min(HR\_linear$YearsInCurrentRole)  
Percentile\_33 = quantile(HR\_linear$YearsInCurrentRole, 0.33333)  
Percentile\_67 = quantile(HR\_linear$YearsInCurrentRole, 0.66667)  
Percentile\_100 = max(HR\_linear$YearsInCurrentRole)  
  
# Values  
HR.BindR = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindR)[[2]] = "Value"  
#HR.BindR  
  
# Years in Current Role Ranges:   
HR\_linear$RoleYear[HR\_linear$YearsInCurrentRole >= Percentile\_00 & HR\_linear$YearsInCurrentRole < Percentile\_33] = "Lower\_Range"  
HR\_linear$RoleYear[HR\_linear$YearsInCurrentRole >= Percentile\_33 & HR\_linear$YearsInCurrentRole < Percentile\_67] = "Mid\_Range"  
HR\_linear$RoleYear[HR\_linear$YearsInCurrentRole >= Percentile\_67 & HR\_linear$YearsInCurrentRole <= Percentile\_100] = "Higher\_Range"  
  
# Categoric Years No Promotion  
  
# Years No Promotion Percentiles  
Percentile\_00 = min(HR\_linear$YearsSinceLastPromotion)  
Percentile\_33 = quantile(HR\_linear$YearsSinceLastPromotion, 0.33333)  
Percentile\_67 = quantile(HR\_linear$YearsSinceLastPromotion, 0.66667)  
Percentile\_100 = max(HR\_linear$YearsSinceLastPromotion)  
  
# Values  
HR.BindP = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindP)[[2]] = "Value"  
#HR.BindP  
  
# Years No Promotion Ranges:   
HR\_linear$NoPromoYears[HR\_linear$YearsSinceLastPromotion >= Percentile\_00 & HR\_linear$YearsSinceLastPromotion < Percentile\_33] = "Lower\_Range"  
HR\_linear$NoPromoYears[HR\_linear$YearsSinceLastPromotion >= Percentile\_33 & HR\_linear$YearsSinceLastPromotion < Percentile\_67] = "Mid\_Range"  
HR\_linear$NoPromoYears[HR\_linear$YearsSinceLastPromotion >= Percentile\_67 & HR\_linear$YearsSinceLastPromotion <= Percentile\_100] = "Higher\_Range"  
  
  
# Categoric Years Current Manager  
  
# Years Current Manager Percentiles  
Percentile\_00 = min(HR\_linear$YearsWithCurrManager)  
Percentile\_33 = quantile(HR\_linear$YearsWithCurrManager, 0.33333)  
Percentile\_67 = quantile(HR\_linear$YearsWithCurrManager, 0.66667)  
Percentile\_100 = max(HR\_linear$YearsWithCurrManager)  
  
# Values  
HR.BindM = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.BindM)[[2]] = "Value"  
#HR.BindM  
  
# Years Current Manager Ranges:   
HR\_linear$ManagerYears[HR\_linear$YearsWithCurrManager >= Percentile\_00 & HR\_linear$YearsWithCurrManager < Percentile\_33] = "Lower\_Range"  
HR\_linear$ManagerYears[HR\_linear$YearsWithCurrManager >= Percentile\_33 & HR\_linear$YearsWithCurrManager < Percentile\_67] = "Mid\_Range"  
HR\_linear$ManagerYears[HR\_linear$YearsWithCurrManager >= Percentile\_67 & HR\_linear$YearsWithCurrManager <= Percentile\_100] = "Higher\_Range"  
  
# Remove Numerical values categorized  
HR\_linear<-HR\_linear[c(-1,-2,-5,-7,-12,-18,-19,-20,-22,-26,-29,-30,-31,-32)]  
  
# Convert all other Numerical values to factors  
HR\_linear<-lapply(HR\_linear, function(x){as.factor(x)})  
HR\_linear = as.data.frame(HR\_linear)  
str(HR\_linear)

## 'data.frame': 1470 obs. of 31 variables:  
## $ Attrition : Factor w/ 2 levels "No","Yes": 2 1 2 1 1 1 1 1 1 1 ...  
## $ BusinessTravel : Factor w/ 3 levels "Non-Travel","Travel\_Frequently",..: 3 2 3 2 3 2 3 3 2 3 ...  
## $ Department : Factor w/ 3 levels "Human Resources",..: 3 2 2 2 2 2 2 2 2 2 ...  
## $ Education : Factor w/ 5 levels "1","2","3","4",..: 2 1 2 4 1 2 3 1 3 3 ...  
## $ EducationField : Factor w/ 6 levels "Human Resources",..: 2 2 5 2 4 2 4 2 2 4 ...  
## $ EnvironmentSatisfaction : Factor w/ 4 levels "1","2","3","4": 2 3 4 4 1 4 3 4 4 3 ...  
## $ Gender : Factor w/ 2 levels "Female","Male": 1 2 2 1 2 2 1 2 2 2 ...  
## $ JobInvolvement : Factor w/ 4 levels "1","2","3","4": 3 2 2 3 3 3 4 3 2 3 ...  
## $ JobLevel : Factor w/ 5 levels "1","2","3","4",..: 2 2 1 1 1 1 1 1 3 2 ...  
## $ JobRole : Factor w/ 9 levels "Healthcare Representative",..: 8 7 3 7 3 3 3 3 5 1 ...  
## $ JobSatisfaction : Factor w/ 4 levels "1","2","3","4": 4 2 3 3 2 4 1 3 3 3 ...  
## $ MaritalStatus : Factor w/ 3 levels "Divorced","Married",..: 3 2 3 2 2 3 2 1 3 2 ...  
## $ OverTime : Factor w/ 2 levels "No","Yes": 2 1 2 2 1 1 2 1 1 1 ...  
## $ PerformanceRating : Factor w/ 2 levels "3","4": 1 2 1 1 1 1 2 2 2 1 ...  
## $ RelationshipSatisfaction: Factor w/ 4 levels "1","2","3","4": 1 4 2 3 4 3 1 2 2 2 ...  
## $ StockOptionLevel : Factor w/ 4 levels "0","1","2","3": 1 2 1 1 2 1 4 2 1 3 ...  
## $ TrainingTimesLastYear : Factor w/ 7 levels "0","1","2","3",..: 1 4 4 4 4 3 4 3 3 4 ...  
## $ WorkLifeBalance : Factor w/ 4 levels "1","2","3","4": 1 3 3 3 3 2 2 3 3 2 ...  
## $ AgeRange : Factor w/ 3 levels "Higher\_Range",..: 1 1 3 3 2 3 1 2 3 3 ...  
## $ HourlyRateRange : Factor w/ 3 levels "High\_Range","Low\_Range",..: 1 3 1 3 2 1 1 3 2 1 ...  
## $ DailyRateRange : Factor w/ 3 levels "High\_Range","Low\_Range",..: 1 2 1 1 3 3 1 1 2 1 ...  
## $ MonthRateLevel : Factor w/ 3 levels "High\_Income",..: 1 1 2 1 3 3 2 3 2 3 ...  
## $ MonthIncomeLevel : Factor w/ 3 levels "High\_Income",..: 3 3 2 2 2 2 2 2 1 3 ...  
## $ DistHomeRange : Factor w/ 3 levels "High\_Distance",..: 2 3 2 3 2 2 3 1 1 1 ...  
## $ NumCompWorked : Factor w/ 3 levels "High\_Number",..: 1 3 1 3 1 2 1 3 2 1 ...  
## $ SalaryIncreaseLevel : Factor w/ 3 levels "Avg\_Increase",..: 3 2 1 3 3 1 2 2 2 1 ...  
## $ WorkingYears : Factor w/ 3 levels "Higher\_Range",..: 3 3 3 3 2 3 1 2 3 1 ...  
## $ CompanyYears : Factor w/ 3 levels "Higher\_Range",..: 3 1 2 1 2 3 2 2 1 3 ...  
## $ RoleYear : Factor w/ 3 levels "Higher\_Range",..: 3 1 2 1 3 1 2 2 1 1 ...  
## $ NoPromoYears : Factor w/ 2 levels "Higher\_Range",..: 2 2 2 1 1 1 2 2 2 1 ...  
## $ ManagerYears : Factor w/ 3 levels "Higher\_Range",..: 3 1 2 2 3 1 2 2 1 1 ...

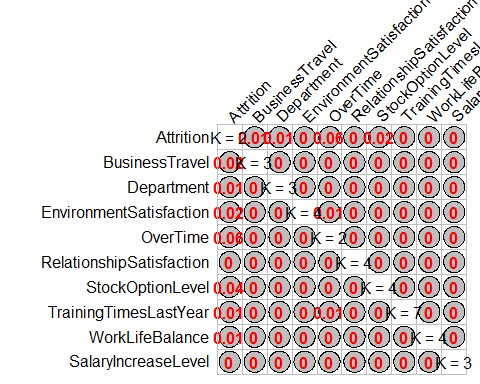
#summary(HR\_linear)  
  
Percentiles.HR<-cbind(HR.BindA,HR.BindH,HR.BindDR,HR.BindMR,HR.BindI,HR.BindD,HR.BindC,HR.BindS,HR.BindW,HR.BindY,HR.BindR,HR.BindP,HR.BindM)  
colnames(Percentiles.HR)<-c("Age","HourlyRate","DailyRate","MonthlyRate","MonthlyIncome","HomeDistance","CompaniesWorked","SalaryIncrease","WorkingYears","YearsAtCompany","YearsInRole","NoPromoYears","YearsWManager")  
if("knitr" %in% rownames(installed.packages()) == FALSE) {install.packages('knitr') }  
library(knitr)  
kable(t(Percentiles.HR),digits=0, format="markdown", padding =2, format.args = list(big.mark = ","))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Percentile\_00 | Percentile\_33 | Percentile\_67 | Percentile\_100 |
| Age | 18 | 32 | 40 | 60 |
| HourlyRate | 30 | 54 | 78 | 100 |
| DailyRate | 102 | 573 | 1,039 | 1,499 |
| MonthlyRate | 2,094 | 10,035 | 18,615 | 26,999 |
| MonthlyIncome | 1,009 | 3,632 | 6,529 | 19,999 |
| HomeDistance | 1 | 3 | 10 | 29 |
| CompaniesWorked | 0 | 1 | 3 | 9 |
| SalaryIncrease | 11 | 13 | 16 | 25 |
| WorkingYears | 0 | 7 | 12 | 40 |
| YearsAtCompany | 0 | 4 | 8 | 40 |
| YearsInRole | 0 | 2 | 6 | 18 |
| NoPromoYears | 0 | 0 | 2 | 15 |
| YearsWManager | 0 | 2 | 6 | 17 |

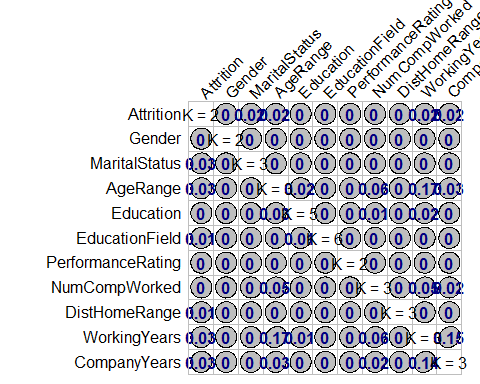
grid.arrange(tableGrob(t(format(Percentiles.HR,digits=0,big.mark=",")),   
 theme=ttheme\_default(core=list(fg\_params=list(fontface=3),big.mark = ","),  
 colhead=list(fg\_params=list(col="navyblue", fontface=4L)), rowhead=list(fg\_params=list(col="navyblue", fontface=3L)))))



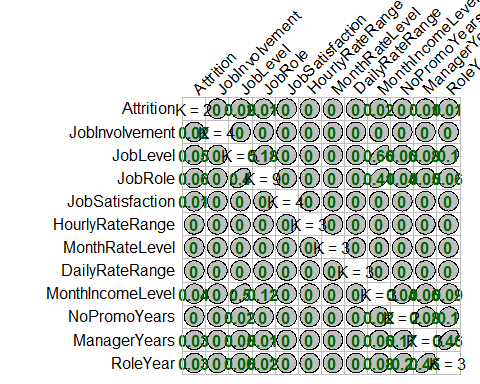
varCompany.set<- c("Attrition","BusinessTravel","Department","EnvironmentSatisfaction","OverTime","RelationshipSatisfaction","StockOptionLevel","TrainingTimesLastYear", "WorkLifeBalance", "SalaryIncreaseLevel")  
varPerson.set<- c("Attrition","Gender","MaritalStatus","AgeRange","Education","EducationField","PerformanceRating", "NumCompWorked","DistHomeRange","WorkingYears","CompanyYears")  
varJob.set<- c("Attrition","JobInvolvement","JobLevel","JobRole","JobSatisfaction","HourlyRateRange","MonthRateLevel","DailyRateRange", "MonthIncomeLevel", "NoPromoYears","ManagerYears","RoleYear")  
  
Frame1<- subset(HR\_linear, select = varCompany.set)  
Frame2<- subset(HR\_linear, select = varPerson.set)  
Frame3<- subset(HR\_linear, select = varJob.set)  
  
GKmatrix1<- GKtauDataframe(Frame1)  
plot(GKmatrix1, corrColors = "red")



GKmatrix1<- GKtauDataframe(Frame2)  
plot(GKmatrix1, corrColors = "navyblue")



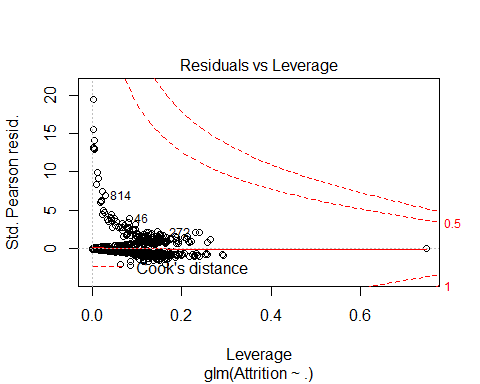
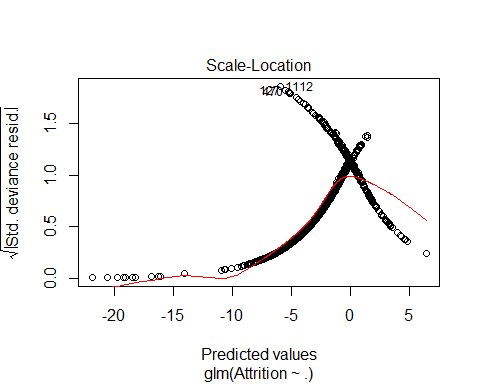
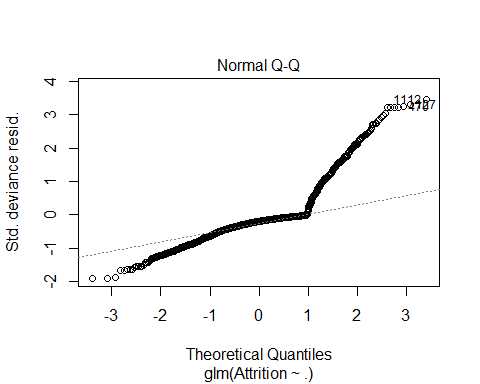
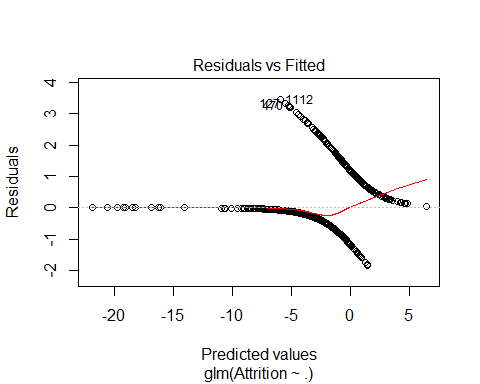
GKmatrix1<- GKtauDataframe(Frame3)  
plot(GKmatrix1, corrColors = "darkgreen")



#Logistic Regression Model  
Attrition.Model<-glm(Attrition~.,data=HR\_linear, family = binomial())  
summary(Attrition.Model)

##   
## Call:  
## glm(formula = Attrition ~ ., family = binomial(), data = HR\_linear)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8221 -0.4239 -0.1935 -0.0608 3.4447   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -10.75546 591.97139 -0.018 0.985504   
## BusinessTravelTravel\_Frequently 1.97433 0.44861 4.401 1.08e-05 \*\*\*  
## BusinessTravelTravel\_Rarely 0.91970 0.41169 2.234 0.025485 \*   
## DepartmentResearch & Development 14.56879 591.97061 0.025 0.980366   
## DepartmentSales 13.64583 591.97079 0.023 0.981609   
## Education2 0.13897 0.35400 0.393 0.694637   
## Education3 0.15916 0.31322 0.508 0.611349   
## Education4 0.21933 0.34823 0.630 0.528791   
## Education5 0.13534 0.66229 0.204 0.838085   
## EducationFieldLife Sciences -1.21045 0.90971 -1.331 0.183325   
## EducationFieldMarketing -0.54666 0.96623 -0.566 0.571551   
## EducationFieldMedical -1.15912 0.90834 -1.276 0.201927   
## EducationFieldOther -1.05483 0.97815 -1.078 0.280856   
## EducationFieldTechnical Degree -0.07002 0.92176 -0.076 0.939451   
## EnvironmentSatisfaction2 -1.11827 0.29978 -3.730 0.000191 \*\*\*  
## EnvironmentSatisfaction3 -1.20927 0.27365 -4.419 9.91e-06 \*\*\*  
## EnvironmentSatisfaction4 -1.53958 0.28182 -5.463 4.68e-08 \*\*\*  
## GenderMale 0.47825 0.20025 2.388 0.016925 \*   
## JobInvolvement2 -1.48650 0.39191 -3.793 0.000149 \*\*\*  
## JobInvolvement3 -1.74282 0.36722 -4.746 2.08e-06 \*\*\*  
## JobInvolvement4 -2.52586 0.51498 -4.905 9.35e-07 \*\*\*  
## JobLevel2 -1.54845 0.52632 -2.942 0.003261 \*\*   
## JobLevel3 -0.63142 0.71967 -0.877 0.380283   
## JobLevel4 -1.63014 0.99342 -1.641 0.100810   
## JobLevel5 0.72121 1.25275 0.576 0.564817   
## JobRoleHuman Resources 15.00953 591.97064 0.025 0.979772   
## JobRoleLaboratory Technician 0.76607 0.63039 1.215 0.224276   
## JobRoleManager -0.52825 1.05965 -0.499 0.618120   
## JobRoleManufacturing Director 0.36403 0.57706 0.631 0.528151   
## JobRoleResearch Director -2.14713 1.10191 -1.949 0.051349 .   
## JobRoleResearch Scientist -0.52874 0.65182 -0.811 0.417269   
## JobRoleSales Executive 2.26362 1.23758 1.829 0.067388 .   
## JobRoleSales Representative 2.03437 1.33341 1.526 0.127086   
## JobSatisfaction2 -0.63516 0.29573 -2.148 0.031733 \*   
## JobSatisfaction3 -0.67078 0.26230 -2.557 0.010549 \*   
## JobSatisfaction4 -1.32957 0.27730 -4.795 1.63e-06 \*\*\*  
## MaritalStatusMarried 0.37415 0.29782 1.256 0.209006   
## MaritalStatusSingle 0.86860 0.43003 2.020 0.043397 \*   
## OverTimeYes 2.18386 0.21598 10.111 < 2e-16 \*\*\*  
## PerformanceRating4 -0.14814 0.33200 -0.446 0.655450   
## RelationshipSatisfaction2 -0.77408 0.30556 -2.533 0.011300 \*   
## RelationshipSatisfaction3 -0.95383 0.27403 -3.481 0.000500 \*\*\*  
## RelationshipSatisfaction4 -0.90046 0.27236 -3.306 0.000946 \*\*\*  
## StockOptionLevel1 -1.02983 0.33514 -3.073 0.002121 \*\*   
## StockOptionLevel2 -0.89991 0.47163 -1.908 0.056380 .   
## StockOptionLevel3 -0.09687 0.49895 -0.194 0.846057   
## TrainingTimesLastYear1 -1.21408 0.61198 -1.984 0.047271 \*   
## TrainingTimesLastYear2 -1.29649 0.45547 -2.846 0.004420 \*\*   
## TrainingTimesLastYear3 -1.43809 0.46161 -3.115 0.001837 \*\*   
## TrainingTimesLastYear4 -1.18525 0.52739 -2.247 0.024614 \*   
## TrainingTimesLastYear5 -1.76607 0.57347 -3.080 0.002073 \*\*   
## TrainingTimesLastYear6 -2.13607 0.68857 -3.102 0.001921 \*\*   
## WorkLifeBalance2 -1.23098 0.40292 -3.055 0.002249 \*\*   
## WorkLifeBalance3 -1.73684 0.37493 -4.632 3.61e-06 \*\*\*  
## WorkLifeBalance4 -1.14898 0.44940 -2.557 0.010568 \*   
## AgeRangeLower\_Range 0.73932 0.30033 2.462 0.013829 \*   
## AgeRangeMid\_Range 0.04254 0.27367 0.155 0.876474   
## HourlyRateRangeLow\_Range -0.17638 0.24021 -0.734 0.462767   
## HourlyRateRangeMid\_Range -0.20467 0.23583 -0.868 0.385476   
## DailyRateRangeLow\_Range 0.55288 0.24346 2.271 0.023152 \*   
## DailyRateRangeMid\_Range 0.47637 0.24559 1.940 0.052416 .   
## MonthRateLevelLow\_Income -0.21245 0.24084 -0.882 0.377727   
## MonthRateLevelMid\_Income 0.07881 0.23373 0.337 0.735978   
## MonthIncomeLevelLow\_Income 0.29736 0.58747 0.506 0.612736   
## MonthIncomeLevelMid\_Income -0.18229 0.46001 -0.396 0.691911   
## DistHomeRangeLow\_Distance -1.07635 0.26194 -4.109 3.97e-05 \*\*\*  
## DistHomeRangeMid\_Distance -0.66039 0.22723 -2.906 0.003658 \*\*   
## NumCompWorkedLow\_Number -1.19190 0.35831 -3.326 0.000879 \*\*\*  
## NumCompWorkedMid\_Number -0.64634 0.25362 -2.548 0.010819 \*   
## SalaryIncreaseLevelHigh\_Increase 0.33967 0.26592 1.277 0.201484   
## SalaryIncreaseLevelLow\_Increase 0.49711 0.24994 1.989 0.046707 \*   
## WorkingYearsLower\_Range 0.65099 0.42321 1.538 0.123995   
## WorkingYearsMid\_Range 0.46081 0.30686 1.502 0.133173   
## CompanyYearsLower\_Range -0.17100 0.46833 -0.365 0.715017   
## CompanyYearsMid\_Range -0.08981 0.41168 -0.218 0.827306   
## RoleYearLower\_Range 0.83311 0.43739 1.905 0.056817 .   
## RoleYearMid\_Range 0.27031 0.37770 0.716 0.474182   
## NoPromoYearsMid\_Range -0.71065 0.24977 -2.845 0.004439 \*\*   
## ManagerYearsLower\_Range 0.37542 0.41858 0.897 0.369776   
## ManagerYearsMid\_Range -0.51274 0.40720 -1.259 0.207968   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1298.58 on 1469 degrees of freedom  
## Residual deviance: 764.83 on 1390 degrees of freedom  
## AIC: 924.83  
##   
## Number of Fisher Scoring iterations: 15

plot(Attrition.Model)



coef(Attrition.Model)

## (Intercept) BusinessTravelTravel\_Frequently   
## -10.75545923 1.97433246   
## BusinessTravelTravel\_Rarely DepartmentResearch & Development   
## 0.91969692 14.56878649   
## DepartmentSales Education2   
## 13.64583260 0.13896917   
## Education3 Education4   
## 0.15916403 0.21933477   
## Education5 EducationFieldLife Sciences   
## 0.13533513 -1.21044721   
## EducationFieldMarketing EducationFieldMedical   
## -0.54666188 -1.15912067   
## EducationFieldOther EducationFieldTechnical Degree   
## -1.05483288 -0.07001695   
## EnvironmentSatisfaction2 EnvironmentSatisfaction3   
## -1.11826902 -1.20927040   
## EnvironmentSatisfaction4 GenderMale   
## -1.53957646 0.47825472   
## JobInvolvement2 JobInvolvement3   
## -1.48649767 -1.74281547   
## JobInvolvement4 JobLevel2   
## -2.52586061 -1.54844825   
## JobLevel3 JobLevel4   
## -0.63142346 -1.63013741   
## JobLevel5 JobRoleHuman Resources   
## 0.72121257 15.00952722   
## JobRoleLaboratory Technician JobRoleManager   
## 0.76606788 -0.52825057   
## JobRoleManufacturing Director JobRoleResearch Director   
## 0.36402611 -2.14713495   
## JobRoleResearch Scientist JobRoleSales Executive   
## -0.52873627 2.26361927   
## JobRoleSales Representative JobSatisfaction2   
## 2.03437428 -0.63515621   
## JobSatisfaction3 JobSatisfaction4   
## -0.67078062 -1.32956972   
## MaritalStatusMarried MaritalStatusSingle   
## 0.37415193 0.86860374   
## OverTimeYes PerformanceRating4   
## 2.18385879 -0.14814193   
## RelationshipSatisfaction2 RelationshipSatisfaction3   
## -0.77408249 -0.95383487   
## RelationshipSatisfaction4 StockOptionLevel1   
## -0.90046361 -1.02982608   
## StockOptionLevel2 StockOptionLevel3   
## -0.89990610 -0.09687238   
## TrainingTimesLastYear1 TrainingTimesLastYear2   
## -1.21407782 -1.29649205   
## TrainingTimesLastYear3 TrainingTimesLastYear4   
## -1.43809486 -1.18525270   
## TrainingTimesLastYear5 TrainingTimesLastYear6   
## -1.76606510 -2.13607320   
## WorkLifeBalance2 WorkLifeBalance3   
## -1.23097788 -1.73683700   
## WorkLifeBalance4 AgeRangeLower\_Range   
## -1.14897878 0.73932292   
## AgeRangeMid\_Range HourlyRateRangeLow\_Range   
## 0.04253911 -0.17638421   
## HourlyRateRangeMid\_Range DailyRateRangeLow\_Range   
## -0.20466616 0.55288194   
## DailyRateRangeMid\_Range MonthRateLevelLow\_Income   
## 0.47637205 -0.21244541   
## MonthRateLevelMid\_Income MonthIncomeLevelLow\_Income   
## 0.07881118 0.29736241   
## MonthIncomeLevelMid\_Income DistHomeRangeLow\_Distance   
## -0.18228606 -1.07634754   
## DistHomeRangeMid\_Distance NumCompWorkedLow\_Number   
## -0.66038728 -1.19189862   
## NumCompWorkedMid\_Number SalaryIncreaseLevelHigh\_Increase   
## -0.64633805 0.33967312   
## SalaryIncreaseLevelLow\_Increase WorkingYearsLower\_Range   
## 0.49711458 0.65098501   
## WorkingYearsMid\_Range CompanyYearsLower\_Range   
## 0.46081012 -0.17100049   
## CompanyYearsMid\_Range RoleYearLower\_Range   
## -0.08981189 0.83310711   
## RoleYearMid\_Range NoPromoYearsMid\_Range   
## 0.27031415 -0.71064885   
## ManagerYearsLower\_Range ManagerYearsMid\_Range   
## 0.37542318 -0.51273654

Linear regression of categorical data doesn’t show high associations to attrition, with the hisghest one being Attrition and Overtime at 0.06.

Considering associations of other variables the highest association were:

Age Range to Working Years Working Years to Years in Company Job level and job role to Monthly Income level and, Time in a role to time with a manager

### Models

#### Association Rule Mining

# References:

* <https://rcompanion.org/handbook/E_05.html>
* <https://towardsdatascience.com/association-rule-mining-in-r-ddf2d044ae50>
* <https://www.datacamp.com/community/tutorials/market-basket-analysis-r>

#Install packages if they dont exist  
  
 # Package arules  
if("arules" %in% rownames(installed.packages()) == FALSE) {install.packages("arules")}  
library(arules)

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following object is masked from 'package:wrapr':  
##   
## unpack

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

##   
## Attaching package: 'arules'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following objects are masked from 'package:base':  
##   
## abbreviate, write

# Package arulesViz  
if("arulesViz" %in% rownames(installed.packages()) == FALSE) {install.packages("arulesViz")}  
library(arulesViz)

## Registered S3 method overwritten by 'seriation':  
## method from   
## reorder.hclust gclus

# RColorBrewer  
if("RColorBrewer" %in% rownames(installed.packages()) == FALSE) {install.packages("RColorBrewer")}  
library(RColorBrewer)  
  
if("gridExtra" %in% rownames(installed.packages()) == FALSE) {install.packages("gridExtra")}  
library(gridExtra)  
  
if("grid" %in% rownames(installed.packages()) == FALSE) {install.packages("grid")}  
library(grid)  
  
if("ggplot2" %in% rownames(installed.packages()) == FALSE) {install.packages("ggplot2")}  
library(ggplot2)  
  
if("lattice" %in% rownames(installed.packages()) == FALSE) {install.packages("lattice")}  
library(lattice)  
  
#Data Assessment  
HR\_arm <- HR\_clean  
str(HR\_arm)

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

Actions required: 1. Eliminate Employee ID 2. Redundant Attributes removed Daily Rate Hourly Rate Monthly Rate 3. Other Numerical Values need to be converted to Factors

# Data Transformation  
  
# Remove Redundant, None added value attributes  
HR\_arm<-HR\_arm[c(-1,-5,-12,-19)]  
  
#Create a Categoric Income Label based on Percentiles  
  
 # Determining percentiles  
Percentile\_00 = min(HR\_arm$MonthlyIncome)  
Percentile\_33 = quantile(HR\_arm$MonthlyIncome, 0.33333)  
Percentile\_67 = quantile(HR\_arm$MonthlyIncome, 0.66667)  
Percentile\_100 = max(HR\_arm$MonthlyIncome)  
  
 # Values  
HR.Bind = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.Bind)[[2]] = "Value"  
HR.Bind

## Value  
## Percentile\_00 1009.000  
## Percentile\_33 3631.647  
## Percentile\_67 6528.735  
## Percentile\_100 19999.000

# Grouping  
HR\_arm$Group[HR\_arm$MonthlyIncome >= Percentile\_00 & HR\_arm$MonthlyIncome < Percentile\_33] = "Low\_Income"  
HR\_arm$Group[HR\_arm$MonthlyIncome >= Percentile\_33 & HR\_arm$MonthlyIncome < Percentile\_67] = "Mid\_Income"  
HR\_arm$Group[HR\_arm$MonthlyIncome >= Percentile\_67 & HR\_arm$MonthlyIncome <= Percentile\_100] = "High\_Income"  
  
 # Remove Numerical "values"Monthly Income"  
HR\_arm<-HR\_arm[-15]  
  
 # Convert all other Numerical values to factors  
HR\_arm<-lapply(HR\_arm, function(x){as.factor(x)})  
HR\_arm = as.data.frame(HR\_arm)  
str(HR\_arm)

## 'data.frame': 1470 obs. of 28 variables:  
## $ Age : Factor w/ 43 levels "18","19","20",..: 24 32 20 16 10 15 42 13 21 19 ...  
## $ Attrition : Factor w/ 2 levels "No","Yes": 2 1 2 1 1 1 1 1 1 1 ...  
## $ BusinessTravel : Factor w/ 3 levels "Non-Travel","Travel\_Frequently",..: 3 2 3 2 3 2 3 3 2 3 ...  
## $ Department : Factor w/ 3 levels "Human Resources",..: 3 2 2 2 2 2 2 2 2 2 ...  
## $ DistanceFromHome : Factor w/ 29 levels "1","2","3","4",..: 1 8 2 3 2 2 3 24 23 27 ...  
## $ Education : Factor w/ 5 levels "1","2","3","4",..: 2 1 2 4 1 2 3 1 3 3 ...  
## $ EducationField : Factor w/ 6 levels "Human Resources",..: 2 2 5 2 4 2 4 2 2 4 ...  
## $ EnvironmentSatisfaction : Factor w/ 4 levels "1","2","3","4": 2 3 4 4 1 4 3 4 4 3 ...  
## $ Gender : Factor w/ 2 levels "Female","Male": 1 2 2 1 2 2 1 2 2 2 ...  
## $ JobInvolvement : Factor w/ 4 levels "1","2","3","4": 3 2 2 3 3 3 4 3 2 3 ...  
## $ JobLevel : Factor w/ 5 levels "1","2","3","4",..: 2 2 1 1 1 1 1 1 3 2 ...  
## $ JobRole : Factor w/ 9 levels "Healthcare Representative",..: 8 7 3 7 3 3 3 3 5 1 ...  
## $ JobSatisfaction : Factor w/ 4 levels "1","2","3","4": 4 2 3 3 2 4 1 3 3 3 ...  
## $ MaritalStatus : Factor w/ 3 levels "Divorced","Married",..: 3 2 3 2 2 3 2 1 3 2 ...  
## $ NumCompaniesWorked : Factor w/ 10 levels "0","1","2","3",..: 9 2 7 2 10 1 5 2 1 7 ...  
## $ OverTime : Factor w/ 2 levels "No","Yes": 2 1 2 2 1 1 2 1 1 1 ...  
## $ PercentSalaryHike : Factor w/ 15 levels "11","12","13",..: 1 13 5 1 2 3 10 12 11 3 ...  
## $ PerformanceRating : Factor w/ 2 levels "3","4": 1 2 1 1 1 1 2 2 2 1 ...  
## $ RelationshipSatisfaction: Factor w/ 4 levels "1","2","3","4": 1 4 2 3 4 3 1 2 2 2 ...  
## $ StockOptionLevel : Factor w/ 4 levels "0","1","2","3": 1 2 1 1 2 1 4 2 1 3 ...  
## $ TotalWorkingYears : Factor w/ 40 levels "0","1","2","3",..: 9 11 8 9 7 9 13 2 11 18 ...  
## $ TrainingTimesLastYear : Factor w/ 7 levels "0","1","2","3",..: 1 4 4 4 4 3 4 3 3 4 ...  
## $ WorkLifeBalance : Factor w/ 4 levels "1","2","3","4": 1 3 3 3 3 2 2 3 3 2 ...  
## $ YearsAtCompany : Factor w/ 37 levels "0","1","2","3",..: 7 11 1 9 3 8 2 2 10 8 ...  
## $ YearsInCurrentRole : Factor w/ 19 levels "0","1","2","3",..: 5 8 1 8 3 8 1 1 8 8 ...  
## $ YearsSinceLastPromotion : Factor w/ 16 levels "0","1","2","3",..: 1 2 1 4 3 4 1 1 2 8 ...  
## $ YearsWithCurrManager : Factor w/ 18 levels "0","1","2","3",..: 6 8 1 1 3 7 1 1 9 8 ...  
## $ Group : Factor w/ 3 levels "High\_Income",..: 3 3 2 2 2 2 2 2 1 3 ...

# Convert to Transactional Data  
HR\_Trans = as(HR\_arm, "transactions")  
HR\_Trans

## transactions in sparse format with  
## 1470 transactions (rows) and  
## 303 items (columns)

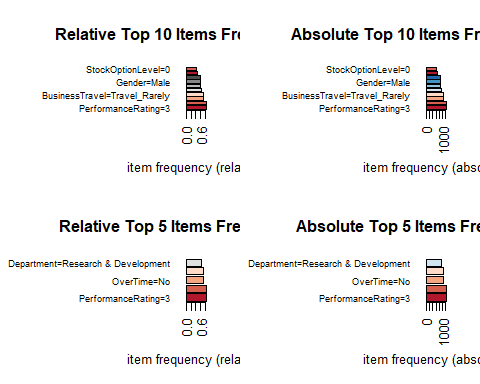
Data set as transactions! Lets take a look

# Information about the transactions data  
  
summary(HR\_Trans)

## transactions as itemMatrix in sparse format with  
## 1470 rows (elements/itemsets/transactions) and  
## 303 columns (items) and a density of 0.09240924   
##   
## most frequent items:  
## PerformanceRating=3 Attrition=No   
## 1244 1233   
## OverTime=No BusinessTravel=Travel\_Rarely   
## 1054 1043   
## Department=Research & Development (Other)   
## 961 35625   
##   
## element (itemset/transaction) length distribution:  
## sizes  
## 28   
## 1470   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 28 28 28 28 28 28   
##   
## includes extended item information - examples:  
## labels variables levels  
## 1 Age=18 Age 18  
## 2 Age=19 Age 19  
## 3 Age=20 Age 20  
##   
## includes extended transaction information - examples:  
## transactionID  
## 1 1  
## 2 2  
## 3 3

* 1470 rows (elements/itemsets/transactions) and 303 columns (items)
* Dense Matrix based on Percent of Non-Zero cells in the matrix (Density) of 0.09240924.
* Most Frequency Items:
  + PerformanceRating=3: 1244
  + Attrition=No: 1233
  + OverTime=No: 1054
  + BusinessTravel=Travel\_Rarely: 1043
  + Department=Research & Development: 961

par(mfrow=c(2,2))  
  
# Item Frequency Plot Top 10 Relative  
arules::itemFrequencyPlot(HR\_Trans,support = 0.2, cex.names=0.7, topN=10, col=brewer.pal(8,'RdGy'), type="relative",main="Relative Top 10 Items Frequency Plot", horiz=TRUE)  
  
# Item Frequency Plot Top 10 Absolute  
itemFrequencyPlot(HR\_Trans,support = 0.2, cex.names=0.7, topN=10, col=brewer.pal(8,'RdBu'), type="absolute", main="Absolute Top 10 Items Frequency Plot",horiz=TRUE)  
  
# Item Frequency Plot for top 5 Relative  
itemFrequencyPlot(HR\_Trans,support = 0.2, cex.names=0.7, topN=5, col=brewer.pal(8,'RdGy'),type="relative", main="Relative Top 5 Items Frequency Plot", horiz=TRUE)  
  
# Item Frequency Plot for top 5 most frequent items  
itemFrequencyPlot(HR\_Trans,support = 0.2, cex.names=0.7, topN= 5,col=brewer.pal(8,'RdBu'), type="absolute", main="Absolute Top 5 Items Frequency Plot",horiz=TRUE)

 “Attrition= No” is in the top of the list along with No Overtime, Travel Rarely and Performance Rating =3

# Apriori Rules with Support = 0.1 and Confidence 0.5  
HR\_Rules1<-apriori(HR\_Trans,parameter = list(support=0.1, confidence =0.5, maxlen = 305))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.5 0.1 1 none FALSE TRUE 5 0.1 1  
## maxlen target ext  
## 305 rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 147   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[303 item(s), 1470 transaction(s)] done [0.00s].  
## sorting and recoding items ... [75 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 6 7 done [0.02s].  
## writing ... [10478 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

HR\_Rules1

## set of 10478 rules

## Changing some parameters  
 ### For stronger rules: Increased confidence.  
 ### For lenghtier rules increase the maxlen parameter.  
 ### To eliminate shorter rules decrease the minlen parameter.  
  
# Apriori Rules with Support = 0.1 and Confidence 0.9 max items 30 min items 3  
HR\_Rules2<-apriori(HR\_Trans,parameter = list(support=0.1, confidence =0.9, maxlen = 30, minlen = 3))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.9 0.1 1 none FALSE TRUE 5 0.1 3  
## maxlen target ext  
## 30 rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 147   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[303 item(s), 1470 transaction(s)] done [0.00s].  
## sorting and recoding items ... [75 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 6 7 done [0.02s].  
## writing ... [921 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

HR\_Rules2

## set of 921 rules

# Apriori Rules with Support = 0.01 and Confidence 0.8 and RHS fixed to Attrition =Yes  
HR\_Rules3<-apriori(HR\_Trans,parameter = list(support=0.01, confidence =0.8, maxlen = 30), appearance = list(rhs="Attrition=Yes"))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.8 0.1 1 none FALSE TRUE 5 0.01 1  
## maxlen target ext  
## 30 rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 14   
##   
## set item appearances ...[1 item(s)] done [0.00s].  
## set transactions ...[303 item(s), 1470 transaction(s)] done [0.00s].  
## sorting and recoding items ... [239 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 6 7 8 9 10 11 12 done [3.47s].  
## writing ... [243 rule(s)] done [0.12s].  
## creating S4 object ... done [0.06s].

HR\_Rules3

## set of 243 rules

# Apriori Rules with Support = 0.1 and Confidence 0.9 and RHS fixed to Attrition =No  
HR\_Rules4<-apriori(HR\_Trans,parameter = list(support=0.1, confidence =0.8, maxlen = 30), appearance = list(rhs="Attrition=No"))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.8 0.1 1 none FALSE TRUE 5 0.1 1  
## maxlen target ext  
## 30 rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 147   
##   
## set item appearances ...[1 item(s)] done [0.00s].  
## set transactions ...[303 item(s), 1470 transaction(s)] done [0.00s].  
## sorting and recoding items ... [75 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 6 7 done [0.02s].  
## writing ... [1557 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

HR\_Rules4

## set of 1557 rules

Based on 303 items and 1,470 transactions and changing parameters created rules: \* First set of rules, created 10,478 rules \* Second set of rules created 921 rules \* Third set of rules (fixing the RHS to Attrition=No) 243 rules \* Fourth set of rules (fixing the RHS to Attrition=Yes) 1557 rules

# Support is an indication of how frequently the itemset appears in the dataset. For Attrition = Yes support was reduced to 0.01 as opposed to the other models that considered support = 0.1

# Confidence is an indication of how often the rule has been found to be true. All rules generated with confidence >= 80%

#Rules Summaries (just for Rules with Attrition Fixed)  
  
# Attrition = Yes  
  
summary(HR\_Rules3)

## set of 243 rules  
##   
## rule length distribution (lhs + rhs):sizes  
## 4 5 6 7 8 9 10   
## 3 40 100 70 23 6 1   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.000 6.000 6.000 6.379 7.000 10.000   
##   
## summary of quality measures:  
## support confidence lift count   
## Min. :0.01020 Min. :0.8000 Min. :4.962 Min. :15.00   
## 1st Qu.:0.01020 1st Qu.:0.8333 1st Qu.:5.169 1st Qu.:15.00   
## Median :0.01088 Median :0.8421 Median :5.223 Median :16.00   
## Mean :0.01157 Mean :0.8602 Mean :5.335 Mean :17.01   
## 3rd Qu.:0.01224 3rd Qu.:0.8824 3rd Qu.:5.473 3rd Qu.:18.00   
## Max. :0.01973 Max. :1.0000 Max. :6.203 Max. :29.00   
##   
## mining info:  
## data ntransactions support confidence  
## HR\_Trans 1470 0.01 0.8

# Attrition = No  
  
summary(HR\_Rules4)

## set of 1557 rules  
##   
## rule length distribution (lhs + rhs):sizes  
## 1 2 3 4 5 6   
## 1 52 387 688 384 45   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 3.000 4.000 3.987 5.000 6.000   
##   
## summary of quality measures:  
## support confidence lift count   
## Min. :0.1000 Min. :0.8000 Min. :0.9538 Min. : 147.0   
## 1st Qu.:0.1102 1st Qu.:0.8547 1st Qu.:1.0190 1st Qu.: 162.0   
## Median :0.1272 Median :0.8851 Median :1.0552 Median : 187.0   
## Mean :0.1474 Mean :0.8824 Mean :1.0520 Mean : 216.7   
## 3rd Qu.:0.1599 3rd Qu.:0.9118 3rd Qu.:1.0870 3rd Qu.: 235.0   
## Max. :0.8388 Max. :0.9755 Max. :1.1630 Max. :1233.0   
##   
## mining info:  
## data ntransactions support confidence  
## HR\_Trans 1470 0.1 0.8

For Attrition = Yes \* Parameter Specification: Support= 0.01 and Confidence = 0.8 \* A length of 6 items has the most rules (100) while a length of 10 items has only one \* Summary of Quality Measures: Min and Max Values for Support, Confidence and Lift shown For Attrition = No \* Parameter Specification: Support= 0.1 and Confidence = 0.8 (Same confidence much lower Support than the prior one) \* A length of 4 items has the most rules (688) while a length of 1 item has only one \* Summary of Quality Measures: Min and Max Values for Support, Confidence and Lift shown

Next: Looking at the top 20 rules considering 1 set of rules created without a fix RHS and the 2 RHS fixed rules

# Top 100 Rules for second set of Rules (Not Fixed)  
  
inspect(head(sort(HR\_Rules2, by = "confidence"), 100))

## lhs rhs support confidence lift count  
## [1] {Attrition=No,   
## PercentSalaryHike=12} => {PerformanceRating=3} 0.1122449 1 1.181672 165  
## [2] {Attrition=No,   
## PercentSalaryHike=14} => {PerformanceRating=3} 0.1204082 1 1.181672 177  
## [3] {BusinessTravel=Travel\_Rarely,   
## PercentSalaryHike=13} => {PerformanceRating=3} 0.1054422 1 1.181672 155  
## [4] {Attrition=No,   
## PercentSalaryHike=13} => {PerformanceRating=3} 0.1190476 1 1.181672 175  
## [5] {BusinessTravel=Travel\_Rarely,   
## PercentSalaryHike=11} => {PerformanceRating=3} 0.1013605 1 1.181672 149  
## [6] {OverTime=No,   
## PercentSalaryHike=11} => {PerformanceRating=3} 0.1013605 1 1.181672 149  
## [7] {Attrition=No,   
## PercentSalaryHike=11} => {PerformanceRating=3} 0.1149660 1 1.181672 169  
## [8] {JobRole=Laboratory Technician,   
## Group=Low\_Income} => {Department=Research & Development} 0.1176871 1 1.529657 173  
## [9] {JobLevel=1,   
## JobRole=Laboratory Technician} => {Department=Research & Development} 0.1360544 1 1.529657 200  
## [10] {JobInvolvement=3,   
## JobRole=Laboratory Technician} => {Department=Research & Development} 0.1000000 1 1.529657 147  
## [11] {Gender=Male,   
## JobRole=Laboratory Technician} => {Department=Research & Development} 0.1183673 1 1.529657 174  
## [12] {JobRole=Laboratory Technician,   
## WorkLifeBalance=3} => {Department=Research & Development} 0.1061224 1 1.529657 156  
## [13] {BusinessTravel=Travel\_Rarely,   
## JobRole=Laboratory Technician} => {Department=Research & Development} 0.1224490 1 1.529657 180  
## [14] {JobRole=Laboratory Technician,   
## OverTime=No} => {Department=Research & Development} 0.1340136 1 1.529657 197  
## [15] {Attrition=No,   
## JobRole=Laboratory Technician} => {Department=Research & Development} 0.1340136 1 1.529657 197  
## [16] {JobRole=Laboratory Technician,   
## PerformanceRating=3} => {Department=Research & Development} 0.1476190 1 1.529657 217  
## [17] {JobRole=Research Scientist,   
## Group=Low\_Income} => {Department=Research & Development} 0.1428571 1 1.529657 210  
## [18] {JobLevel=1,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1591837 1 1.529657 234  
## [19] {JobInvolvement=3,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1176871 1 1.529657 173  
## [20] {Gender=Male,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1210884 1 1.529657 178  
## [21] {JobRole=Research Scientist,   
## WorkLifeBalance=3} => {Department=Research & Development} 0.1129252 1 1.529657 166  
## [22] {BusinessTravel=Travel\_Rarely,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1428571 1 1.529657 210  
## [23] {JobRole=Research Scientist,   
## OverTime=No} => {Department=Research & Development} 0.1326531 1 1.529657 195  
## [24] {Attrition=No,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1666667 1 1.529657 245  
## [25] {JobRole=Research Scientist,   
## PerformanceRating=3} => {Department=Research & Development} 0.1653061 1 1.529657 243  
## [26] {JobRole=Sales Executive,   
## Group=Mid\_Income} => {Department=Sales} 0.1210884 1 3.295964 178  
## [27] {JobRole=Sales Executive,   
## Group=High\_Income} => {Department=Sales} 0.1006803 1 3.295964 148  
## [28] {JobLevel=2,   
## JobRole=Sales Executive} => {Department=Sales} 0.1585034 1 3.295964 233  
## [29] {JobRole=Sales Executive,   
## MaritalStatus=Married} => {Department=Sales} 0.1027211 1 3.295964 151  
## [30] {JobInvolvement=3,   
## JobRole=Sales Executive} => {Department=Sales} 0.1333333 1 3.295964 196  
## [31] {Gender=Male,   
## JobRole=Sales Executive} => {Department=Sales} 0.1319728 1 3.295964 194  
## [32] {JobRole=Sales Executive,   
## WorkLifeBalance=3} => {Department=Sales} 0.1374150 1 3.295964 202  
## [33] {BusinessTravel=Travel\_Rarely,   
## JobRole=Sales Executive} => {Department=Sales} 0.1551020 1 3.295964 228  
## [34] {JobRole=Sales Executive,   
## OverTime=No} => {Department=Sales} 0.1578231 1 3.295964 232  
## [35] {Attrition=No,   
## JobRole=Sales Executive} => {Department=Sales} 0.1829932 1 3.295964 269  
## [36] {JobRole=Sales Executive,   
## PerformanceRating=3} => {Department=Sales} 0.1938776 1 3.295964 285  
## [37] {JobRole=Sales Executive,   
## Group=Mid\_Income} => {JobLevel=2} 0.1210884 1 2.752809 178  
## [38] {MaritalStatus=Single,   
## RelationshipSatisfaction=4} => {StockOptionLevel=0} 0.1034014 1 2.329635 152  
## [39] {EnvironmentSatisfaction=4,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1047619 1 2.329635 154  
## [40] {Department=Sales,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1040816 1 2.329635 153  
## [41] {JobSatisfaction=4,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1068027 1 2.329635 157  
## [42] {EducationField=Medical,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1000000 1 2.329635 147  
## [43] {MaritalStatus=Single,   
## Group=Low\_Income} => {StockOptionLevel=0} 0.1217687 1 2.329635 179  
## [44] {MaritalStatus=Single,   
## Group=Mid\_Income} => {StockOptionLevel=0} 0.1020408 1 2.329635 150  
## [45] {MaritalStatus=Single,   
## TrainingTimesLastYear=3} => {StockOptionLevel=0} 0.1040816 1 2.329635 153  
## [46] {MaritalStatus=Single,   
## NumCompaniesWorked=1} => {StockOptionLevel=0} 0.1244898 1 2.329635 183  
## [47] {JobLevel=2,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1095238 1 2.329635 161  
## [48] {JobLevel=1,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1374150 1 2.329635 202  
## [49] {MaritalStatus=Single,   
## TrainingTimesLastYear=2} => {StockOptionLevel=0} 0.1170068 1 2.329635 172  
## [50] {Education=3,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1319728 1 2.329635 194  
## [51] {MaritalStatus=Single,   
## YearsSinceLastPromotion=0} => {StockOptionLevel=0} 0.1326531 1 2.329635 195  
## [52] {Gender=Female,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1353741 1 2.329635 199  
## [53] {EducationField=Life Sciences,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1367347 1 2.329635 201  
## [54] {JobInvolvement=3,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1884354 1 2.329635 277  
## [55] {Gender=Male,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.1843537 1 2.329635 271  
## [56] {MaritalStatus=Single,   
## WorkLifeBalance=3} => {StockOptionLevel=0} 0.2000000 1 2.329635 294  
## [57] {Department=Research & Development,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.2068027 1 2.329635 304  
## [58] {BusinessTravel=Travel\_Rarely,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.2224490 1 2.329635 327  
## [59] {MaritalStatus=Single,   
## OverTime=No} => {StockOptionLevel=0} 0.2306122 1 2.329635 339  
## [60] {Attrition=No,   
## MaritalStatus=Single} => {StockOptionLevel=0} 0.2380952 1 2.329635 350  
## [61] {MaritalStatus=Single,   
## PerformanceRating=3} => {StockOptionLevel=0} 0.2707483 1 2.329635 398  
## [62] {JobLevel=1,   
## JobRole=Laboratory Technician,   
## Group=Low\_Income} => {Department=Research & Development} 0.1081633 1 1.529657 159  
## [63] {JobLevel=1,   
## JobRole=Laboratory Technician,   
## OverTime=No} => {Department=Research & Development} 0.1034014 1 1.529657 152  
## [64] {JobLevel=1,   
## JobRole=Laboratory Technician,   
## PerformanceRating=3} => {Department=Research & Development} 0.1122449 1 1.529657 165  
## [65] {BusinessTravel=Travel\_Rarely,   
## JobRole=Laboratory Technician,   
## PerformanceRating=3} => {Department=Research & Development} 0.1034014 1 1.529657 152  
## [66] {Attrition=No,   
## JobRole=Laboratory Technician,   
## OverTime=No} => {Department=Research & Development} 0.1129252 1 1.529657 166  
## [67] {JobRole=Laboratory Technician,   
## OverTime=No,   
## PerformanceRating=3} => {Department=Research & Development} 0.1122449 1 1.529657 165  
## [68] {Attrition=No,   
## JobRole=Laboratory Technician,   
## PerformanceRating=3} => {Department=Research & Development} 0.1142857 1 1.529657 168  
## [69] {JobLevel=1,   
## JobRole=Research Scientist,   
## Group=Low\_Income} => {Department=Research & Development} 0.1387755 1 1.529657 204  
## [70] {BusinessTravel=Travel\_Rarely,   
## JobRole=Research Scientist,   
## Group=Low\_Income} => {Department=Research & Development} 0.1061224 1 1.529657 156  
## [71] {Attrition=No,   
## JobRole=Research Scientist,   
## Group=Low\_Income} => {Department=Research & Development} 0.1163265 1 1.529657 171  
## [72] {JobRole=Research Scientist,   
## PerformanceRating=3,   
## Group=Low\_Income} => {Department=Research & Development} 0.1170068 1 1.529657 172  
## [73] {Gender=Male,   
## JobLevel=1,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1000000 1 1.529657 147  
## [74] {BusinessTravel=Travel\_Rarely,   
## JobLevel=1,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1163265 1 1.529657 171  
## [75] {JobLevel=1,   
## JobRole=Research Scientist,   
## OverTime=No} => {Department=Research & Development} 0.1061224 1 1.529657 156  
## [76] {Attrition=No,   
## JobLevel=1,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1285714 1 1.529657 189  
## [77] {JobLevel=1,   
## JobRole=Research Scientist,   
## PerformanceRating=3} => {Department=Research & Development} 0.1319728 1 1.529657 194  
## [78] {Attrition=No,   
## JobInvolvement=3,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1013605 1 1.529657 149  
## [79] {Attrition=No,   
## Gender=Male,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1006803 1 1.529657 148  
## [80] {Gender=Male,   
## JobRole=Research Scientist,   
## PerformanceRating=3} => {Department=Research & Development} 0.1006803 1 1.529657 148  
## [81] {BusinessTravel=Travel\_Rarely,   
## JobRole=Research Scientist,   
## OverTime=No} => {Department=Research & Development} 0.1000000 1 1.529657 147  
## [82] {Attrition=No,   
## BusinessTravel=Travel\_Rarely,   
## JobRole=Research Scientist} => {Department=Research & Development} 0.1238095 1 1.529657 182  
## [83] {BusinessTravel=Travel\_Rarely,   
## JobRole=Research Scientist,   
## PerformanceRating=3} => {Department=Research & Development} 0.1190476 1 1.529657 175  
## [84] {Attrition=No,   
## JobRole=Research Scientist,   
## OverTime=No} => {Department=Research & Development} 0.1231293 1 1.529657 181  
## [85] {JobRole=Research Scientist,   
## OverTime=No,   
## PerformanceRating=3} => {Department=Research & Development} 0.1115646 1 1.529657 164  
## [86] {Attrition=No,   
## JobRole=Research Scientist,   
## PerformanceRating=3} => {Department=Research & Development} 0.1414966 1 1.529657 208  
## [87] {Department=Sales,   
## JobRole=Sales Executive,   
## Group=Mid\_Income} => {JobLevel=2} 0.1210884 1 2.752809 178  
## [88] {JobLevel=2,   
## JobRole=Sales Executive,   
## Group=Mid\_Income} => {Department=Sales} 0.1210884 1 3.295964 178  
## [89] {Attrition=No,   
## JobRole=Sales Executive,   
## Group=Mid\_Income} => {Department=Sales} 0.1027211 1 3.295964 151  
## [90] {JobRole=Sales Executive,   
## PerformanceRating=3,   
## Group=Mid\_Income} => {Department=Sales} 0.1020408 1 3.295964 150  
## [91] {BusinessTravel=Travel\_Rarely,   
## JobLevel=2,   
## JobRole=Sales Executive} => {Department=Sales} 0.1108844 1 3.295964 163  
## [92] {JobLevel=2,   
## JobRole=Sales Executive,   
## OverTime=No} => {Department=Sales} 0.1129252 1 3.295964 166  
## [93] {Attrition=No,   
## JobLevel=2,   
## JobRole=Sales Executive} => {Department=Sales} 0.1340136 1 3.295964 197  
## [94] {JobLevel=2,   
## JobRole=Sales Executive,   
## PerformanceRating=3} => {Department=Sales} 0.1360544 1 3.295964 200  
## [95] {Attrition=No,   
## JobInvolvement=3,   
## JobRole=Sales Executive} => {Department=Sales} 0.1142857 1 3.295964 168  
## [96] {JobInvolvement=3,   
## JobRole=Sales Executive,   
## PerformanceRating=3} => {Department=Sales} 0.1149660 1 3.295964 169  
## [97] {Attrition=No,   
## Gender=Male,   
## JobRole=Sales Executive} => {Department=Sales} 0.1068027 1 3.295964 157  
## [98] {Gender=Male,   
## JobRole=Sales Executive,   
## PerformanceRating=3} => {Department=Sales} 0.1183673 1 3.295964 174  
## [99] {Attrition=No,   
## JobRole=Sales Executive,   
## WorkLifeBalance=3} => {Department=Sales} 0.1176871 1 3.295964 173  
## [100] {JobRole=Sales Executive,   
## PerformanceRating=3,   
## WorkLifeBalance=3} => {Department=Sales} 0.1231293 1 3.295964 181

# Top 20 Rules for rules with RHS at Attrition = Yes  
  
inspect(head(sort(HR\_Rules3, by = "confidence"), 20))

## lhs rhs support confidence lift count  
## [1] {MaritalStatus=Single,   
## OverTime=Yes,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [2] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## YearsWithCurrManager=0} => {Attrition=Yes} 0.01156463 1.0000000 6.202532 17  
## [3] {MaritalStatus=Single,   
## OverTime=Yes,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01020408 1.0000000 6.202532 15  
## [4] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [5] {JobLevel=1,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0} => {Attrition=Yes} 0.01156463 1.0000000 6.202532 17  
## [6] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [7] {MaritalStatus=Single,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [8] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsWithCurrManager=0} => {Attrition=Yes} 0.01156463 1.0000000 6.202532 17  
## [9] {JobLevel=1,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsSinceLastPromotion=0,   
## YearsWithCurrManager=0} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [10] {BusinessTravel=Travel\_Frequently,   
## JobLevel=1,   
## PerformanceRating=3,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [11] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01020408 1.0000000 6.202532 15  
## [12] {MaritalStatus=Single,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01020408 1.0000000 6.202532 15  
## [13] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [14] {JobLevel=1,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [15] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01088435 1.0000000 6.202532 16  
## [16] {JobLevel=1,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsSinceLastPromotion=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01020408 1.0000000 6.202532 15  
## [17] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsInCurrentRole=0,   
## YearsWithCurrManager=0,   
## Group=Low\_Income} => {Attrition=Yes} 0.01020408 1.0000000 6.202532 15  
## [18] {JobLevel=1,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsInCurrentRole=0} => {Attrition=Yes} 0.01292517 0.9500000 5.892405 19  
## [19] {JobLevel=1,   
## OverTime=Yes,   
## StockOptionLevel=0,   
## YearsWithCurrManager=0} => {Attrition=Yes} 0.01292517 0.9500000 5.892405 19  
## [20] {JobLevel=1,   
## MaritalStatus=Single,   
## OverTime=Yes,   
## YearsInCurrentRole=0} => {Attrition=Yes} 0.01224490 0.9473684 5.876083 18

# Top 20 Rules for rules with RHS at Attrition = No  
  
inspect(head(sort(HR\_Rules4, by = "confidence"), 20))

## lhs rhs support confidence lift count  
## [1] {Department=Research & Development,   
## OverTime=No,   
## StockOptionLevel=1,   
## WorkLifeBalance=3} => {Attrition=No} 0.1081633 0.9754601 1.162957 159  
## [2] {BusinessTravel=Travel\_Rarely,   
## Department=Research & Development,   
## OverTime=No,   
## Group=High\_Income} => {Attrition=No} 0.1006803 0.9673203 1.153253 148  
## [3] {OverTime=No,   
## StockOptionLevel=1,   
## WorkLifeBalance=3} => {Attrition=No} 0.1680272 0.9648438 1.150300 247  
## [4] {EnvironmentSatisfaction=4,   
## OverTime=No,   
## WorkLifeBalance=3} => {Attrition=No} 0.1224490 0.9625668 1.147586 180  
## [5] {Department=Research & Development,   
## OverTime=No,   
## YearsWithCurrManager=2} => {Attrition=No} 0.1034014 0.9620253 1.146940 152  
## [6] {JobLevel=2,   
## StockOptionLevel=1,   
## Group=Mid\_Income} => {Attrition=No} 0.1034014 0.9620253 1.146940 152  
## [7] {BusinessTravel=Travel\_Rarely,   
## JobLevel=2,   
## OverTime=No,   
## WorkLifeBalance=3} => {Attrition=No} 0.1034014 0.9620253 1.146940 152  
## [8] {EnvironmentSatisfaction=4,   
## OverTime=No,   
## PerformanceRating=3,   
## WorkLifeBalance=3} => {Attrition=No} 0.1027211 0.9617834 1.146652 151  
## [9] {EducationField=Life Sciences,   
## OverTime=No,   
## StockOptionLevel=1} => {Attrition=No} 0.1190476 0.9615385 1.146360 175  
## [10] {JobLevel=2,   
## OverTime=No,   
## StockOptionLevel=1} => {Attrition=No} 0.1000000 0.9607843 1.145461 147  
## [11] {Department=Research & Development,   
## MaritalStatus=Married,   
## OverTime=No,   
## WorkLifeBalance=3} => {Attrition=No} 0.1156463 0.9604520 1.145064 170  
## [12] {Department=Research & Development,   
## JobLevel=2,   
## WorkLifeBalance=3} => {Attrition=No} 0.1142857 0.9600000 1.144526 168  
## [13] {Department=Research & Development,   
## JobLevel=2,   
## OverTime=No,   
## PerformanceRating=3} => {Attrition=No} 0.1136054 0.9597701 1.144251 167  
## [14] {Department=Research & Development,   
## Gender=Female,   
## OverTime=No,   
## WorkLifeBalance=3} => {Attrition=No} 0.1074830 0.9575758 1.141635 158  
## [15] {OverTime=No,   
## PerformanceRating=3,   
## StockOptionLevel=1,   
## WorkLifeBalance=3} => {Attrition=No} 0.1380952 0.9575472 1.141601 203  
## [16] {JobSatisfaction=4,   
## OverTime=No,   
## WorkLifeBalance=3} => {Attrition=No} 0.1210884 0.9569892 1.140936 178  
## [17] {Department=Research & Development,   
## TrainingTimesLastYear=3,   
## WorkLifeBalance=3} => {Attrition=No} 0.1176871 0.9558011 1.139520 173  
## [18] {MaritalStatus=Married,   
## OverTime=No,   
## StockOptionLevel=1,   
## WorkLifeBalance=3} => {Attrition=No} 0.1176871 0.9558011 1.139520 173  
## [19] {Department=Research & Development,   
## JobLevel=2,   
## PerformanceRating=3} => {Attrition=No} 0.1551020 0.9539749 1.137342 228  
## [20] {BusinessTravel=Travel\_Rarely,   
## Department=Research & Development,   
## Group=High\_Income} => {Attrition=No} 0.1360544 0.9523810 1.135442 200

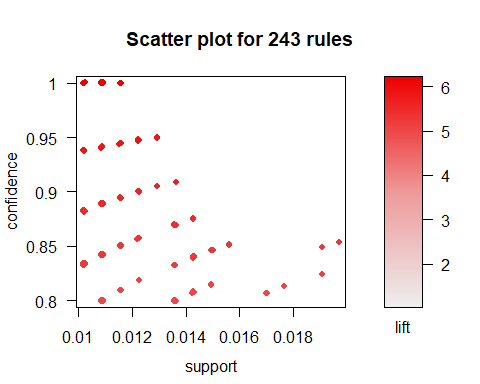
The first set of rules provides insight in regards to performance rating,department information,stock option level and job level but no information about attrition. By fixing the RHS to Attrition = Yes and Attrition = No rules provide more insight.

With Attrition = Yes, the most frequent factors in the top 20 rules are: \* Marital Status = Single. In 13 out of the 20 rules \* Overtime = Yes. In 18 out of the 20 rules \* Years with current Manager = 0. In 16 out of the 20 rules \* Years in current Role = 0. In 12 out of the 20 rules \* Low Income. In 10 out of the 20 rules

With Attrition = No, the most frequent factors in the top 20 rules are: \* Department=Research & Development. In 10 out of the 20 rules  
\* OverTime=No. In 15 out of the 20 rules  
\* StockOptionLevel=1. In 6 out of the 20 rules  
\* WorkLifeBalance=3. In 11 out of the 20 rules

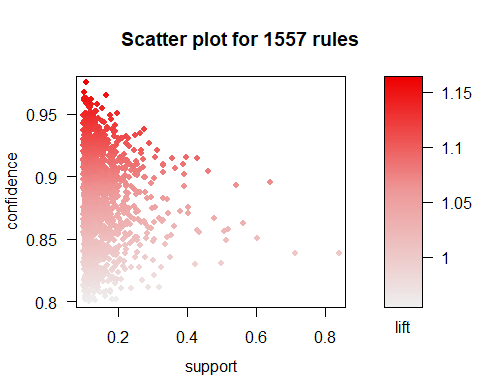
### Rules with Confidence > 40 and 50%  
   
## Attrition = Yes  
subsetRulesYes<-HR\_Rules3[quality(HR\_Rules3)$confidence>0.4]  
   
## Attrition = No  
subsetRulesNo<-HR\_Rules4[quality(HR\_Rules4)$confidence>0.5]  
  
### Plots  
  
## Scatter  
  
# Attrition = Yes  
plot(subsetRulesYes)

## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



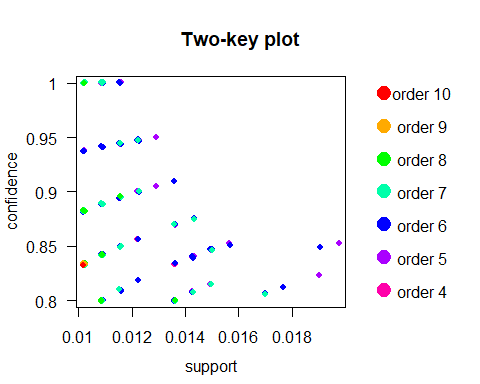
# Attrition = No  
plot(subsetRulesNo)

## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



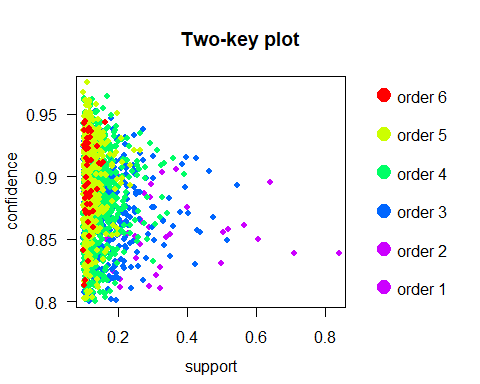
## Two-Key   
  
# Attrition = Yes  
plot(subsetRulesYes, method = "two-key plot")

## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



# Attrition = No  
plot(subsetRulesNo, method = "two-key plot")

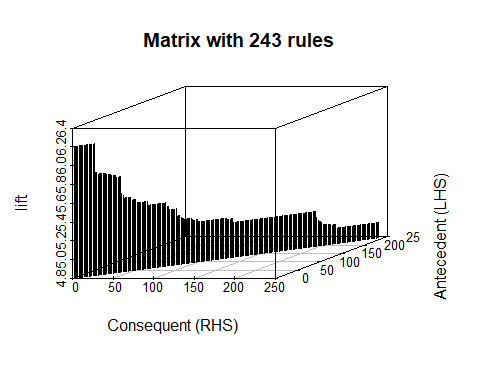
## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



## Matrix 3D  
  
# Attrition = Yes  
plot(subsetRulesYes, method = "matrix3d")

## Warning in plot.rules(subsetRulesYes, method = "matrix3d"): method 'matrix3D' is  
## deprecated use method 'matrix' with engine '3d'

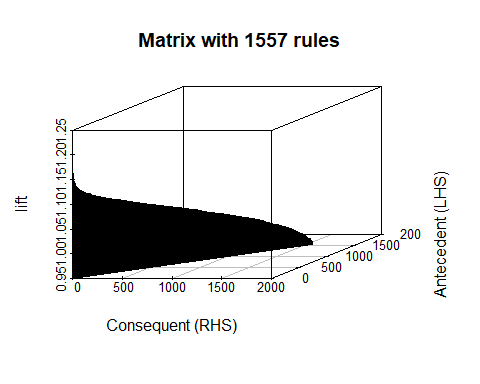
## Itemsets in Antecedent (LHS)  
## NULL  
## Itemsets in Consequent (RHS)  
## NULL



# Attrition = No  
plot(subsetRulesNo, method = "matrix3d")

## Warning in plot.rules(subsetRulesNo, method = "matrix3d"): method 'matrix3D' is  
## deprecated use method 'matrix' with engine '3d'

## Itemsets in Antecedent (LHS)  
## NULL  
## Itemsets in Consequent (RHS)  
## NULL



### Interactive Scatter-Plot  
  
# Attrition = Yes  
plotly\_arules(subsetRulesYes)

## Warning: 'plotly\_arules' is deprecated.  
## Use 'plot' instead.  
## See help("Deprecated")

## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.

## PhantomJS not found. You can install it with webshot::install\_phantomjs(). If it is installed, please make sure the phantomjs executable can be found via the PATH variable.

# Attrition = No  
plotly\_arules(subsetRulesNo)

## Warning: 'plotly\_arules' is deprecated.  
## Use 'plot' instead.  
## See help("Deprecated")

## Warning: plot: Too many rules supplied. Only plotting the best 1000 rules using  
## measure lift (change parameter max if needed)

## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.

#### Graph Based Visualizations  
  
### Subrules  
### Selecting 20 Rules with the Highest Confidence for each set  
  
 ## Attrition = Yes  
top20.subRulesYes<-head(subsetRulesYes, n = 20, by ="confidence")  
  
 ## Attrition = No  
top20.subRulesNo<-head(subsetRulesNo, n = 20, by ="confidence")  
  
### 20 Rules Plots  
  
 ## Attrition = Yes  
plot(top20.subRulesYes, method = "graph", engine = "htmlwidget")

## Attrition = No  
plot(top20.subRulesNo, method = "graph", engine = "htmlwidget")

### Selecting 10 Rules with the Highest Confidence for each set  
  
 ## Attrition = Yes  
top10.subRulesYes<-head(subsetRulesYes, n = 10, by ="confidence")  
  
 ## Attrition = No  
top10.subRulesNo<-head(subsetRulesNo, n = 10, by ="confidence")  
  
### 10 Rules Plots  
  
 ## Attrition = Yes  
plot(top10.subRulesYes, method = "graph", engine = "htmlwidget")

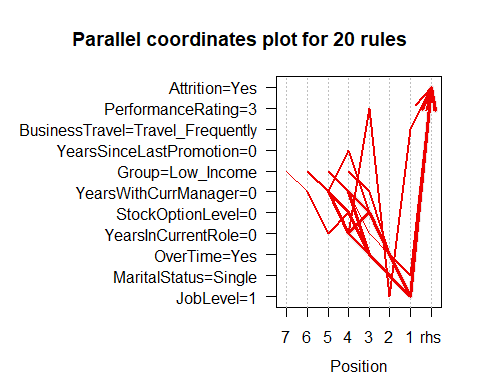
## Attrition = No  
plot(top10.subRulesNo, method = "graph", engine = "htmlwidget")

### Selecting 5 Rules with the Highest Confidence for each set  
  
 ## Attrition = Yes  
top5.subRulesYes<-head(subsetRulesYes, n = 5, by ="confidence")  
  
 ## Attrition = No  
top5.subRulesNo<-head(subsetRulesNo, n = 5, by ="confidence")  
  
## 5 Rules Plots  
  
 ## Attrition = Yes  
plot(top5.subRulesYes, method = "graph", engine = "htmlwidget")

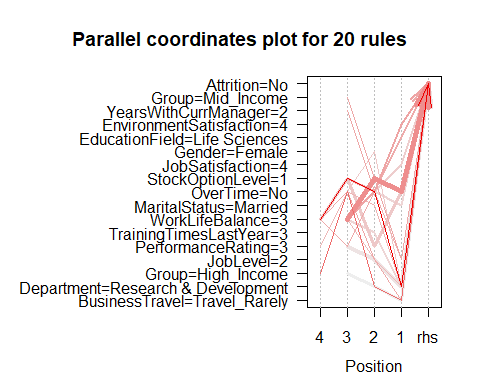
## Attrition = No  
plot(top5.subRulesNo, method = "graph", engine = "htmlwidget")

Graphs 1 and 2: Rules with high lift have low support Graphs 3 and 4: Rules with High confidence and low support have around 7 or 8 items. High support 5 or 6 items

### Selecting 20 Rules with the Highest Lift  
  
 ## Attrition = Yes  
top20.subRulesYesL<-head(subsetRulesYes, n = 20, by ="lift")  
  
 ## Attrition = No  
top20.subRulesNoL<-head(subsetRulesNo, n = 20, by ="lift")  
  
### 20 Rules Plots  
  
 ## Attrition = Yes  
plot(top20.subRulesYesL, method = "paracoord")



## Attrition = No  
plot(top20.subRulesNoL, method = "paracoord")



#### K-means Clustering

K-means clustering is used to visualize patterns in how the attributes contribute to the creation of groups of employees.

xc <- HR\_clean  
x\_factors <- Filter(is.factor, xc)  
head(x\_factors)

## Attrition BusinessTravel Department Education EducationField  
## 1 Yes Travel\_Rarely Sales 2 Life Sciences  
## 2 No Travel\_Frequently Research & Development 1 Life Sciences  
## 3 Yes Travel\_Rarely Research & Development 2 Other  
## 4 No Travel\_Frequently Research & Development 4 Life Sciences  
## 5 No Travel\_Rarely Research & Development 1 Medical  
## 6 No Travel\_Frequently Research & Development 2 Life Sciences  
## EnvironmentSatisfaction Gender JobInvolvement JobLevel JobRole  
## 1 2 Female 3 2 Sales Executive  
## 2 3 Male 2 2 Research Scientist  
## 3 4 Male 2 1 Laboratory Technician  
## 4 4 Female 3 1 Research Scientist  
## 5 1 Male 3 1 Laboratory Technician  
## 6 4 Male 3 1 Laboratory Technician  
## JobSatisfaction MaritalStatus OverTime PerformanceRating  
## 1 4 Single Yes 3  
## 2 2 Married No 4  
## 3 3 Single Yes 3  
## 4 3 Married Yes 3  
## 5 2 Married No 3  
## 6 4 Single No 3  
## RelationshipSatisfaction StockOptionLevel WorkLifeBalance  
## 1 1 0 1  
## 2 4 1 3  
## 3 2 0 3  
## 4 3 0 3  
## 5 4 1 3  
## 6 3 0 2

## Kmeans needs a matrix/dataframe of all numbers  
# remove employee number and attrition yes/no to start with  
xc <-HR\_clean  
xc\_att <-HR\_clean  
xc\_att <- xc[,c(2:32)] # keep a version of the data with attrition so we can compare the impact of attrition on groups  
xc <- xc[,c(2,4:32)]  
xc[] <- lapply(xc, function(x) as.numeric(x))  
head(xc)

## Age BusinessTravel DailyRate Department DistanceFromHome Education  
## 1 41 3 1102 3 1 2  
## 2 49 2 279 2 8 1  
## 3 37 3 1373 2 2 2  
## 4 33 2 1392 2 3 4  
## 5 27 3 591 2 2 1  
## 6 32 2 1005 2 2 2  
## EducationField EnvironmentSatisfaction Gender HourlyRate JobInvolvement  
## 1 2 2 1 94 3  
## 2 2 3 2 61 2  
## 3 5 4 2 92 2  
## 4 2 4 1 56 3  
## 5 4 1 2 40 3  
## 6 2 4 2 79 3  
## JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate  
## 1 2 8 4 3 5993 19479  
## 2 2 7 2 2 5130 24907  
## 3 1 3 3 3 2090 2396  
## 4 1 7 3 2 2909 23159  
## 5 1 3 2 2 3468 16632  
## 6 1 3 4 3 3068 11864  
## NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating  
## 1 8 2 11 1  
## 2 1 1 23 2  
## 3 6 2 15 1  
## 4 1 2 11 1  
## 5 9 1 12 1  
## 6 0 1 13 1  
## RelationshipSatisfaction StockOptionLevel TotalWorkingYears  
## 1 1 1 8  
## 2 4 2 10  
## 3 2 1 7  
## 4 3 1 8  
## 5 4 2 6  
## 6 3 1 8  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0 1 6 4  
## 2 3 3 10 7  
## 3 3 3 0 0  
## 4 3 3 8 7  
## 5 3 3 2 2  
## 6 2 2 7 7  
## YearsSinceLastPromotion YearsWithCurrManager  
## 1 0 5  
## 2 1 7  
## 3 0 0  
## 4 3 0  
## 5 2 2  
## 6 3 6

# make all numeric  
xc\_att[] <- lapply(xc\_att, function(x) as.numeric(x))  
# reorder columns so attrition is last  
xc\_att <- xc\_att[,c(1, 3:31, 2)]  
head(xc\_att)

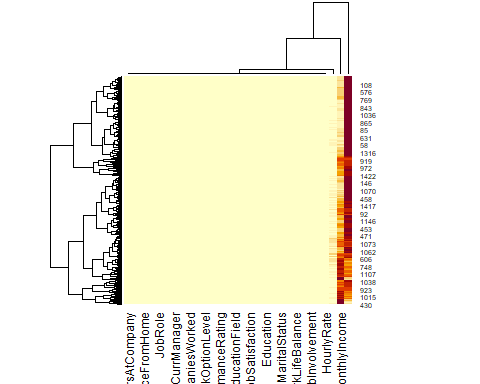
## Age BusinessTravel DailyRate Department DistanceFromHome Education  
## 1 41 3 1102 3 1 2  
## 2 49 2 279 2 8 1  
## 3 37 3 1373 2 2 2  
## 4 33 2 1392 2 3 4  
## 5 27 3 591 2 2 1  
## 6 32 2 1005 2 2 2  
## EducationField EnvironmentSatisfaction Gender HourlyRate JobInvolvement  
## 1 2 2 1 94 3  
## 2 2 3 2 61 2  
## 3 5 4 2 92 2  
## 4 2 4 1 56 3  
## 5 4 1 2 40 3  
## 6 2 4 2 79 3  
## JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate  
## 1 2 8 4 3 5993 19479  
## 2 2 7 2 2 5130 24907  
## 3 1 3 3 3 2090 2396  
## 4 1 7 3 2 2909 23159  
## 5 1 3 2 2 3468 16632  
## 6 1 3 4 3 3068 11864  
## NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating  
## 1 8 2 11 1  
## 2 1 1 23 2  
## 3 6 2 15 1  
## 4 1 2 11 1  
## 5 9 1 12 1  
## 6 0 1 13 1  
## RelationshipSatisfaction StockOptionLevel TotalWorkingYears  
## 1 1 1 8  
## 2 4 2 10  
## 3 2 1 7  
## 4 3 1 8  
## 5 4 2 6  
## 6 3 1 8  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0 1 6 4  
## 2 3 3 10 7  
## 3 3 3 0 0  
## 4 3 3 8 7  
## 5 3 3 2 2  
## 6 2 2 7 7  
## YearsSinceLastPromotion YearsWithCurrManager Attrition  
## 1 0 5 2  
## 2 1 7 1  
## 3 0 0 2  
## 4 3 0 1  
## 5 2 2 1  
## 6 3 6 1

# Some parts of kmeans don't work well with NAs, so make sure those are gone  
colSums(is.na(xc))

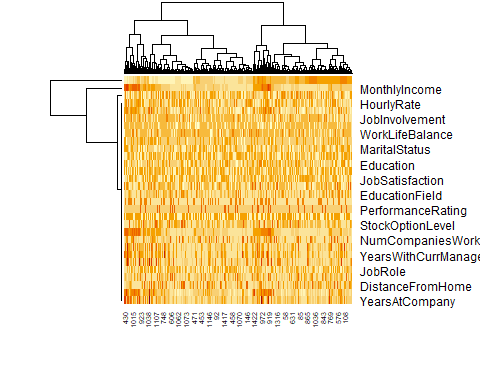
## Age BusinessTravel DailyRate   
## 0 0 0   
## Department DistanceFromHome Education   
## 0 0 0   
## EducationField EnvironmentSatisfaction Gender   
## 0 0 0   
## HourlyRate JobInvolvement JobLevel   
## 0 0 0   
## JobRole JobSatisfaction MaritalStatus   
## 0 0 0   
## MonthlyIncome MonthlyRate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

## Depending on the data, we may need a scaled or transformed matrix. Make all three so we can visualize them.   
xc.m <- as.matrix(xc) # m stands for matrix  
xc.sm <-scale(xc.m) # sm for scaled matrix  
xc.tm <-t(xc.m) # tm for transformed matrix

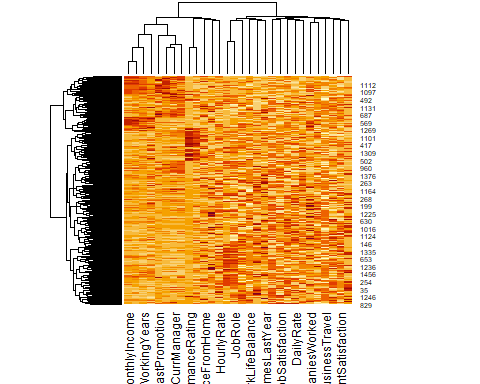
## visualize matrix  
### result: this matrix isn't useful. It needs to be scaled so that income isn't much higher.   
heatmap(xc.m)



## Visualize transformed matrix  
### result: there is a lot of variety in the data, but too many groups to be useful  
heatmap(xc.tm)



#colSums(is.na(xc.sm))  
heatmap(xc.sm)

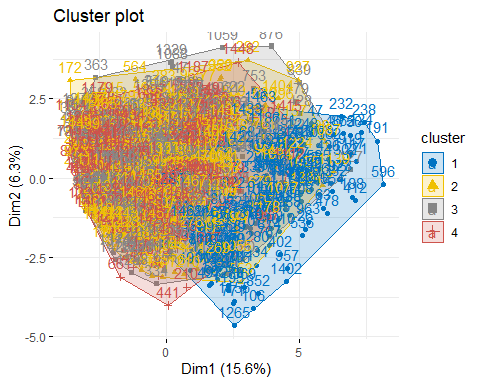


model\_xc4m <- kmeans(xc.m, 4)  
model\_xc4sm <- kmeans(xc.sm, 4)  
model\_xc4tm <- kmeans(xc.tm, 4)

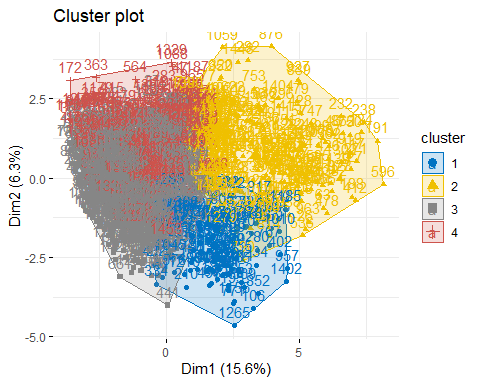
if("factoextra" %in% rownames(installed.packages()) == FALSE) {install.packages('factoextra') }  
library(factoextra)

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

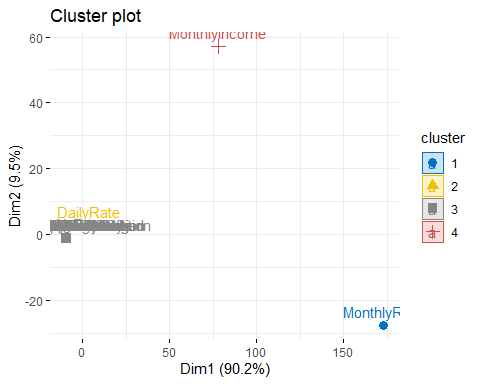
## visualizing kmeans 4 groups with a cluster plot  
### because the numbers aren't scaled, the groups overlap  
fviz\_cluster(model\_xc4m, data = xc.m,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



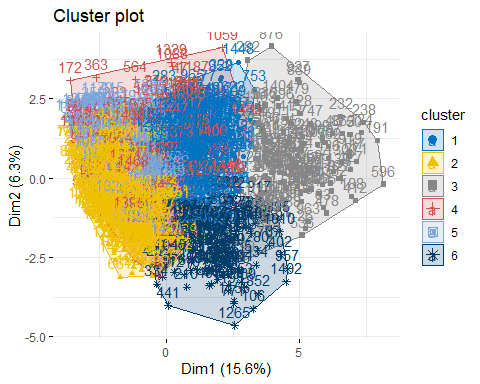
### scaled matrix has three groups, but two overlap a lot  
fviz\_cluster(model\_xc4sm, data = xc.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



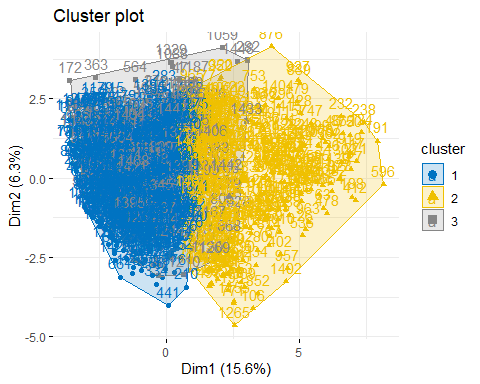
### this isn't useful. Not using transformed matrix going forward.   
fviz\_cluster(model\_xc4tm, data = xc.tm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



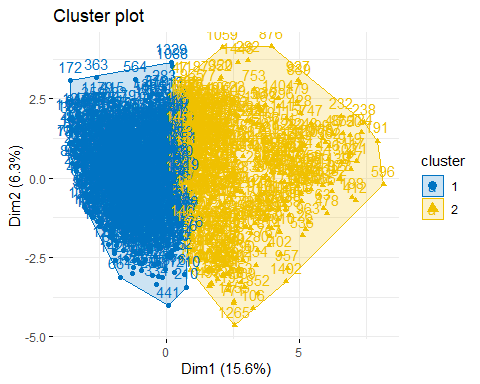
model\_xc6sm <- kmeans(xc.sm, 6)  
fviz\_cluster(model\_xc6sm, data = xc.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



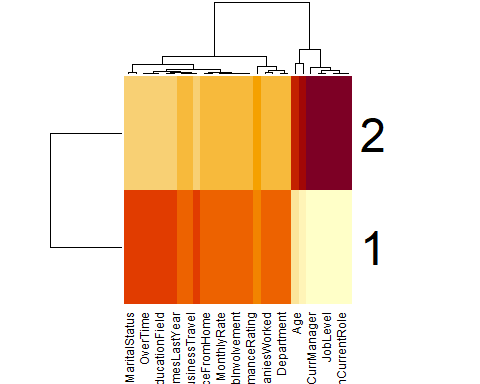
model\_xc3sm <- kmeans(xc.sm, 3)  
fviz\_cluster(model\_xc3sm, data = xc.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



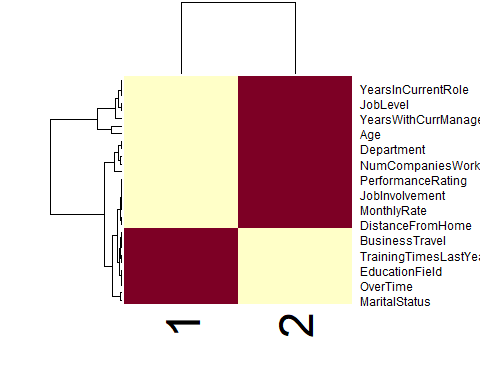
model\_xc2sm <- kmeans(xc.sm, 2)  
fviz\_cluster(model\_xc2sm, data = xc.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



heatmap(model\_xc2sm$centers)

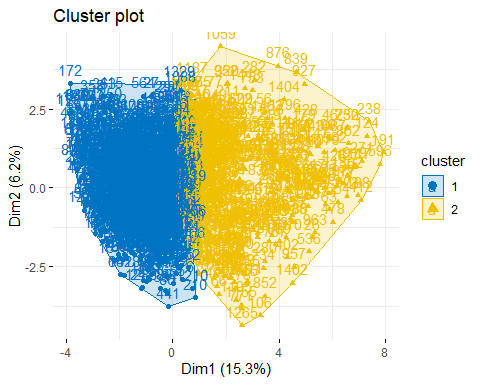


### this isn't useful because it is too coarse  
centers2 <- t(model\_xc2sm$centers)  
heatmap(centers2)

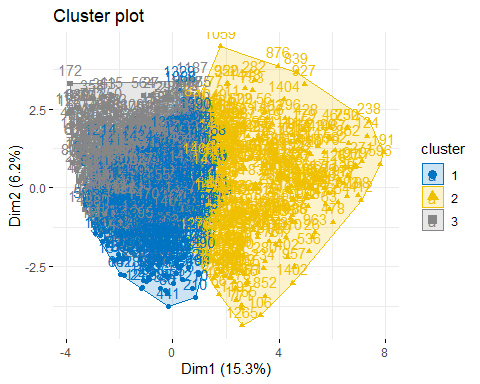


## does including attrition change the clusters?  
  
xc\_att.sm <-scale(as.matrix(xc\_att))  
model\_attsm2 <- kmeans(xc\_att.sm, 2)  
model\_attsm3 <- kmeans(xc\_att.sm, 3)  
model\_attsm4 <- kmeans(xc\_att.sm, 4)  
model\_attsm6 <- kmeans(xc\_att.sm, 6)

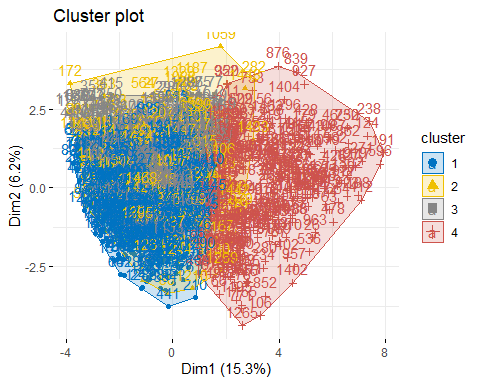
# no change at 2 clusters  
fviz\_cluster(model\_attsm2, data = xc\_att.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



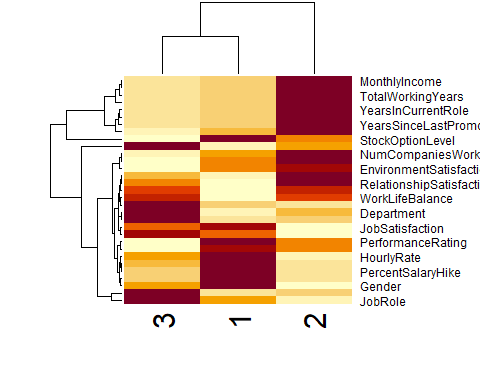
# with 3 clusters, there is some separation  
fviz\_cluster(model\_attsm3, data = xc\_att.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



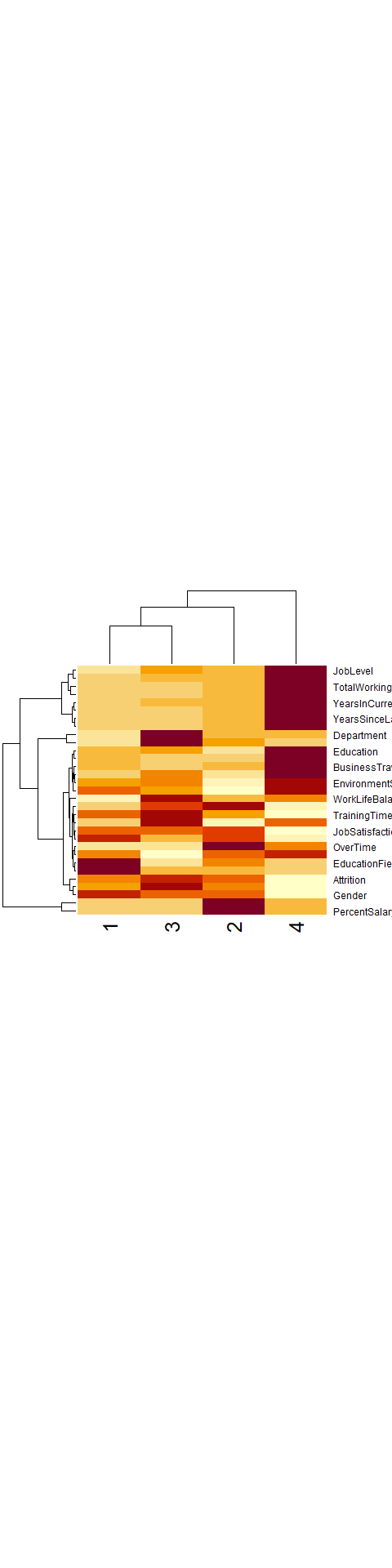
### with 4 clusters, there is too much overlap with three clusters  
### but one cluster is still separate  
fviz\_cluster(model\_attsm4, data = xc\_att.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



centers\_att3 <- t(model\_attsm3$centers)  
heatmap(centers\_att3)

 The centers for the 4-cluster model gives a clear heatmap.

### it looks like the cluster that was separate is for people with high job level, education, and business travel.  
### the three overlapping clusters differ in department, worklife balance, and overtime, among others  
### attrition seems high in one department and with worklife balance and training last year  
centers\_att4 <- t(model\_attsm4$centers)  
heatmap(centers\_att4)



head(xc\_att)

## Age BusinessTravel DailyRate Department DistanceFromHome Education  
## 1 41 3 1102 3 1 2  
## 2 49 2 279 2 8 1  
## 3 37 3 1373 2 2 2  
## 4 33 2 1392 2 3 4  
## 5 27 3 591 2 2 1  
## 6 32 2 1005 2 2 2  
## EducationField EnvironmentSatisfaction Gender HourlyRate JobInvolvement  
## 1 2 2 1 94 3  
## 2 2 3 2 61 2  
## 3 5 4 2 92 2  
## 4 2 4 1 56 3  
## 5 4 1 2 40 3  
## 6 2 4 2 79 3  
## JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate  
## 1 2 8 4 3 5993 19479  
## 2 2 7 2 2 5130 24907  
## 3 1 3 3 3 2090 2396  
## 4 1 7 3 2 2909 23159  
## 5 1 3 2 2 3468 16632  
## 6 1 3 4 3 3068 11864  
## NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating  
## 1 8 2 11 1  
## 2 1 1 23 2  
## 3 6 2 15 1  
## 4 1 2 11 1  
## 5 9 1 12 1  
## 6 0 1 13 1  
## RelationshipSatisfaction StockOptionLevel TotalWorkingYears  
## 1 1 1 8  
## 2 4 2 10  
## 3 2 1 7  
## 4 3 1 8  
## 5 4 2 6  
## 6 3 1 8  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0 1 6 4  
## 2 3 3 10 7  
## 3 3 3 0 0  
## 4 3 3 8 7  
## 5 3 3 2 2  
## 6 2 2 7 7  
## YearsSinceLastPromotion YearsWithCurrManager Attrition  
## 1 0 5 2  
## 2 1 7 1  
## 3 0 0 2  
## 4 3 0 1  
## 5 2 2 1  
## 6 3 6 1

att\_YES <- xc\_att[which(xc\_att$Attrition == 2) , ]  
head(att\_YES)

## Age BusinessTravel DailyRate Department DistanceFromHome Education  
## 1 41 3 1102 3 1 2  
## 3 37 3 1373 2 2 2  
## 15 28 3 103 2 24 3  
## 22 36 3 1218 3 9 4  
## 25 34 3 699 2 6 1  
## 27 32 2 1125 2 16 1  
## EducationField EnvironmentSatisfaction Gender HourlyRate JobInvolvement  
## 1 2 2 1 94 3  
## 3 5 4 2 92 2  
## 15 2 3 2 50 2  
## 22 2 3 2 82 2  
## 25 4 2 2 83 3  
## 27 2 2 1 72 1  
## JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate  
## 1 2 8 4 3 5993 19479  
## 3 1 3 3 3 2090 2396  
## 15 1 3 3 3 2028 12947  
## 22 1 9 1 3 3407 6986  
## 25 1 7 1 3 2960 17102  
## 27 1 7 1 3 3919 4681  
## NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating  
## 1 8 2 11 1  
## 3 6 2 15 1  
## 15 5 2 14 1  
## 22 7 1 23 2  
## 25 2 1 11 1  
## 27 1 2 22 2  
## RelationshipSatisfaction StockOptionLevel TotalWorkingYears  
## 1 1 1 8  
## 3 2 1 7  
## 15 2 1 6  
## 22 2 1 10  
## 25 3 1 8  
## 27 2 1 10  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0 1 6 4  
## 3 3 3 0 0  
## 15 4 3 4 2  
## 22 4 3 5 3  
## 25 2 3 4 2  
## 27 5 3 10 2  
## YearsSinceLastPromotion YearsWithCurrManager Attrition  
## 1 0 5 2  
## 3 0 0 2  
## 15 0 3 2  
## 22 0 3 2  
## 25 1 3 2  
## 27 6 7 2

str(att\_YES)

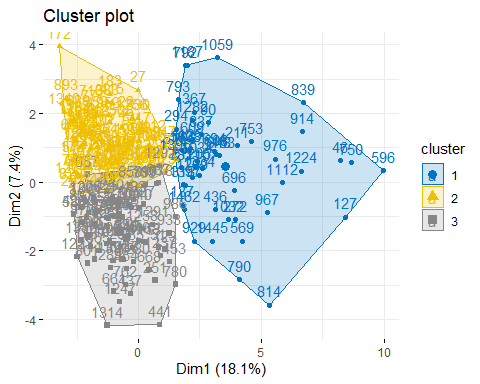
## 'data.frame': 237 obs. of 31 variables:  
## $ Age : num 41 37 28 36 34 32 39 24 50 26 ...  
## $ BusinessTravel : num 3 3 3 3 3 2 3 3 3 3 ...  
## $ DailyRate : num 1102 1373 103 1218 699 ...  
## $ Department : num 3 2 2 3 2 2 3 2 3 2 ...  
## $ DistanceFromHome : num 1 2 24 9 6 16 5 1 3 25 ...  
## $ Education : num 2 2 3 4 1 1 3 3 2 3 ...  
## $ EducationField : num 2 5 2 2 4 2 6 4 3 2 ...  
## $ EnvironmentSatisfaction : num 2 4 3 3 2 2 4 2 1 1 ...  
## $ Gender : num 1 2 2 2 2 1 2 2 2 2 ...  
## $ HourlyRate : num 94 92 50 82 83 72 56 61 86 48 ...  
## $ JobInvolvement : num 3 2 2 2 3 1 3 3 2 1 ...  
## $ JobLevel : num 2 1 1 1 1 1 2 1 1 1 ...  
## $ JobRole : num 8 3 3 9 7 7 9 7 9 3 ...  
## $ JobSatisfaction : num 4 3 3 1 1 1 4 4 3 3 ...  
## $ MaritalStatus : num 3 3 3 3 3 3 2 2 2 3 ...  
## $ MonthlyIncome : num 5993 2090 2028 3407 2960 ...  
## $ MonthlyRate : num 19479 2396 12947 6986 17102 ...  
## $ NumCompaniesWorked : num 8 6 5 7 2 1 3 2 1 1 ...  
## $ OverTime : num 2 2 2 1 1 2 1 2 2 1 ...  
## $ PercentSalaryHike : num 11 15 14 23 11 22 14 16 14 12 ...  
## $ PerformanceRating : num 1 1 1 2 1 2 1 1 1 1 ...  
## $ RelationshipSatisfaction: num 1 2 2 2 3 2 3 1 3 3 ...  
## $ StockOptionLevel : num 1 1 1 1 1 1 2 2 1 1 ...  
## $ TotalWorkingYears : num 8 7 6 10 8 10 19 6 3 1 ...  
## $ TrainingTimesLastYear : num 0 3 4 4 2 5 6 2 2 2 ...  
## $ WorkLifeBalance : num 1 3 3 3 3 3 4 2 3 2 ...  
## $ YearsAtCompany : num 6 0 4 5 4 10 1 2 3 1 ...  
## $ YearsInCurrentRole : num 4 0 2 3 2 2 0 0 2 0 ...  
## $ YearsSinceLastPromotion : num 0 0 0 0 1 6 0 2 0 0 ...  
## $ YearsWithCurrManager : num 5 0 3 3 3 7 0 0 2 1 ...  
## $ Attrition : num 2 2 2 2 2 2 2 2 2 2 ...

att\_YES.sm <-scale(as.matrix(att\_YES[,1:30]))  
head(att\_YES.sm)

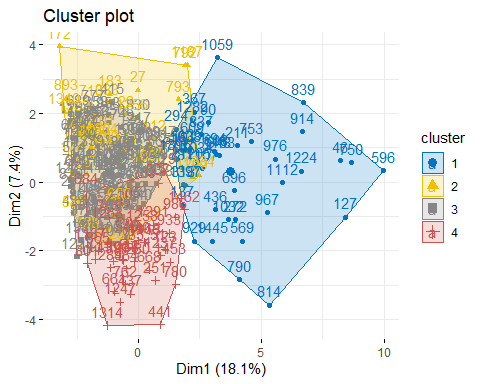
## Age BusinessTravel DailyRate Department DistanceFromHome Education  
## 1 0.7629413 0.6718551 0.8749379 1.1597753 -1.1396489 -0.8327969  
## 3 0.3501169 0.6718551 1.5492358 -0.5909683 -1.0213411 -0.8327969  
## 15 -0.5787380 0.6718551 -1.6107580 -0.5909683 1.5814314 0.1590265  
## 22 0.2469108 0.6718551 1.1635673 1.1597753 -0.1931862 1.1508500  
## 25 0.0404986 0.6718551 -0.1278003 -0.5909683 -0.5481097 -1.8246204  
## 27 -0.1659136 -1.0402917 0.9321662 -0.5909683 0.6349687 -1.8246204  
## EducationField EnvironmentSatisfaction Gender HourlyRate JobInvolvement  
## 1 -0.9258765 -0.3967674 -1.3102912 1.4142398 0.6219417  
## 3 1.1639591 1.3129393 0.7599689 1.3147371 -0.6710424  
## 15 -0.9258765 0.4580860 0.7599689 -0.7748195 -0.6710424  
## 22 -0.9258765 0.4580860 0.7599689 0.8172236 -0.6710424  
## 25 0.4673472 -0.3967674 0.7599689 0.8669750 0.6219417  
## 27 -0.9258765 -0.3967674 -1.3102912 0.3197101 -1.9640264  
## JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome  
## 1 0.3857873 0.8390632 1.3699161 0.8836752 0.3312740  
## 3 -0.6773707 -1.0991237 0.4755081 0.8836752 -0.7409167  
## 15 -0.6773707 -1.0991237 0.4755081 0.8836752 -0.7579487  
## 22 -0.6773707 1.2267006 -1.3133080 0.8836752 -0.3791245  
## 25 -0.6773707 0.4514258 -1.3133080 0.8836752 -0.5019196  
## 27 -0.6773707 0.4514258 -1.3133080 0.8836752 -0.2384733  
## MonthlyRate NumCompaniesWorked OverTime PercentSalaryHike  
## 1 0.6825177 1.8887574 0.9287018 -1.08666491  
## 3 -1.6874375 1.1420760 0.9287018 -0.02573975  
## 15 -0.2236784 0.7687353 0.9287018 -0.29097104  
## 22 -1.0506586 1.5154167 -1.0722285 2.09611059  
## 25 0.3527522 -0.3512868 -1.0722285 -1.08666491  
## 27 -1.3704353 -0.7246275 0.9287018 1.83087929  
## PerformanceRating RelationshipSatisfaction StockOptionLevel  
## 1 -0.4292079 -1.4209196 -0.6158921  
## 3 -0.4292079 -0.5323762 -0.6158921  
## 15 -0.4292079 -0.5323762 -0.6158921  
## 22 2.3200426 -0.5323762 -0.6158921  
## 25 -0.4292079 0.3561672 -0.6158921  
## 27 2.3200426 -0.5323762 -0.6158921  
## TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany  
## 1 -0.03413569 -2.0915729 -2.0310150 0.14608414  
## 3 -0.17362120 0.2992765 0.4186061 -0.86232193  
## 15 -0.31310670 1.0962263 0.4186061 -0.19005121  
## 22 0.24483531 1.0962263 0.4186061 -0.02198354  
## 25 -0.03413569 -0.4976733 0.4186061 -0.19005121  
## 27 0.24483531 1.8931761 0.4186061 0.81835485  
## YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager  
## 1 0.34554528 -0.6169046 0.68324566  
## 3 -0.91436597 -0.6169046 -0.90741467  
## 15 -0.28441035 -0.6169046 0.04698153  
## 22 0.03056747 -0.6169046 0.04698153  
## 25 -0.28441035 -0.2997541 0.04698153  
## 27 -0.28441035 1.2859986 1.31950979

model\_YES3 <- kmeans(att\_YES.sm, 3)  
model\_YES4 <- kmeans(att\_YES.sm, 4)  
model\_YES5 <- kmeans(att\_YES.sm, 5)  
model\_YES6 <- kmeans(att\_YES.sm, 6)

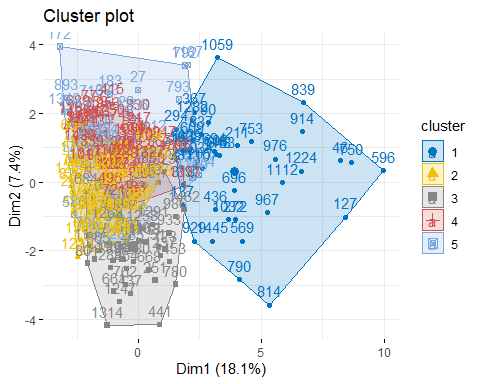
fviz\_cluster(model\_YES3, data = att\_YES.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())

 4 clusters appears to be the most useful Notice how few people left in the right group (cluster 1)

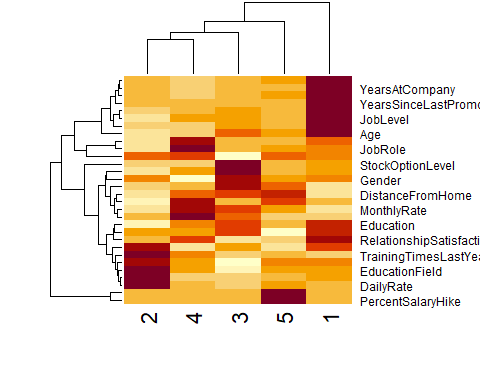
fviz\_cluster(model\_YES4, data = att\_YES.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



fviz\_cluster(model\_YES5, data = att\_YES.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



centers\_yes5 <- t(model\_YES5$centers)  
heatmap(centers\_yes5)

 Looking at which attributes most distinguish between attrition = YES and attrition = NO.

model\_attsm4$centers

## Age BusinessTravel DailyRate Department DistanceFromHome  
## 1 -0.2515959 -0.01729307 -0.008251035 -0.61976593 -0.01649139  
## 2 -0.1613853 -0.01731657 -0.013432306 -0.06303898 0.06471734  
## 3 -0.2390566 -0.03822896 0.018062243 1.38930376 0.03697398  
## 4 0.8967014 0.08893716 0.007114391 -0.11846405 -0.04654518  
## Education EducationField EnvironmentSatisfaction Gender HourlyRate  
## 1 -0.0510113686 0.06050488 -0.014533540 0.05180336 0.02065917  
## 2 -0.0974392430 -0.01148879 -0.087373191 0.01236695 0.01156482  
## 3 0.0003783799 -0.06041629 0.005292806 0.01389846 -0.01642353  
## 4 0.1720267246 -0.05708551 0.082438887 -0.13334268 -0.03409524  
## JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus  
## 1 0.05104536 -0.52624486 -0.429590323 0.010306653 0.0003346135  
## 2 -0.03922610 -0.27677187 -0.003659263 0.017140486 0.0258624404  
## 3 -0.03814935 -0.06336164 1.111592882 0.007499393 0.1127041671  
## 4 -0.04252903 1.37036573 -0.266681908 -0.041169036 -0.1377569950  
## MonthlyIncome MonthlyRate NumCompaniesWorked OverTime PercentSalaryHike  
## 1 -0.4968266 -0.02440074 -0.02776542 -0.009665699 -0.3118666  
## 2 -0.2870826 -0.04267619 -0.07329688 0.033258735 1.7960154  
## 3 -0.1497667 0.01274773 -0.06680580 -0.009665699 -0.2991780  
## 4 1.4064917 0.06632095 0.17830864 0.009177666 -0.1898044  
## PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears  
## 1 -0.4260850 0.01559113 0.005731948 -0.4270696  
## 2 2.3453529 -0.10755119 0.010741700 -0.2462420  
## 3 -0.4260850 -0.02020185 -0.045136307 -0.3538560  
## 4 -0.1699785 0.05857579 0.028875685 1.4486415  
## TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole  
## 1 0.011573798 -0.076144359 -0.3750310 -0.3371420  
## 2 -0.008873963 -0.005164695 -0.2156679 -0.1023052  
## 3 0.045192538 0.118867122 -0.2883653 -0.1854680  
## 4 -0.067052207 0.039002591 1.2479019 0.9833543  
## YearsSinceLastPromotion YearsWithCurrManager Attrition  
## 1 -0.2768375 -0.2978467 0.01198866  
## 2 -0.1570048 -0.1180193 0.06971126  
## 3 -0.2100648 -0.2323374 0.18451889  
## 4 0.9167488 0.9598080 -0.26781223

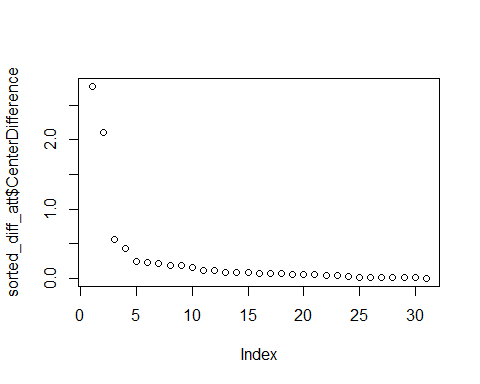
diff\_att <- data.frame(t(model\_attsm4$centers[1:2, ]))  
#rownames(diff\_att) <- c("Group1", "Group2")  
#difference.list <- abs(diff(diff\_att))  
diff\_att$CenterDifference <- round(abs(diff\_att$X1 - diff\_att$X2),2)  
diff\_att

## X1 X2 CenterDifference  
## Age -0.2515959013 -0.161385254 0.09  
## BusinessTravel -0.0172930731 -0.017316570 0.00  
## DailyRate -0.0082510346 -0.013432306 0.01  
## Department -0.6197659270 -0.063038984 0.56  
## DistanceFromHome -0.0164913918 0.064717341 0.08  
## Education -0.0510113686 -0.097439243 0.05  
## EducationField 0.0605048795 -0.011488786 0.07  
## EnvironmentSatisfaction -0.0145335404 -0.087373191 0.07  
## Gender 0.0518033612 0.012366952 0.04  
## HourlyRate 0.0206591674 0.011564820 0.01  
## JobInvolvement 0.0510453635 -0.039226103 0.09  
## JobLevel -0.5262448577 -0.276771865 0.25  
## JobRole -0.4295903229 -0.003659263 0.43  
## JobSatisfaction 0.0103066527 0.017140486 0.01  
## MaritalStatus 0.0003346135 0.025862440 0.03  
## MonthlyIncome -0.4968265877 -0.287082649 0.21  
## MonthlyRate -0.0244007409 -0.042676187 0.02  
## NumCompaniesWorked -0.0277654202 -0.073296881 0.05  
## OverTime -0.0096656991 0.033258735 0.04  
## PercentSalaryHike -0.3118666062 1.796015392 2.11  
## PerformanceRating -0.4260850142 2.345352910 2.77  
## RelationshipSatisfaction 0.0155911308 -0.107551190 0.12  
## StockOptionLevel 0.0057319477 0.010741700 0.01  
## TotalWorkingYears -0.4270696221 -0.246242030 0.18  
## TrainingTimesLastYear 0.0115737983 -0.008873963 0.02  
## WorkLifeBalance -0.0761443591 -0.005164695 0.07  
## YearsAtCompany -0.3750310155 -0.215667862 0.16  
## YearsInCurrentRole -0.3371420266 -0.102305223 0.23  
## YearsSinceLastPromotion -0.2768374767 -0.157004769 0.12  
## YearsWithCurrManager -0.2978467465 -0.118019334 0.18  
## Attrition 0.0119886608 0.069711264 0.06

sorted\_diff\_att <- diff\_att[order(-diff\_att$CenterDifference),]  
sorted\_diff\_att[1:11, ]

## X1 X2 CenterDifference  
## PerformanceRating -0.42608501 2.345352910 2.77  
## PercentSalaryHike -0.31186661 1.796015392 2.11  
## Department -0.61976593 -0.063038984 0.56  
## JobRole -0.42959032 -0.003659263 0.43  
## JobLevel -0.52624486 -0.276771865 0.25  
## YearsInCurrentRole -0.33714203 -0.102305223 0.23  
## MonthlyIncome -0.49682659 -0.287082649 0.21  
## TotalWorkingYears -0.42706962 -0.246242030 0.18  
## YearsWithCurrManager -0.29784675 -0.118019334 0.18  
## YearsAtCompany -0.37503102 -0.215667862 0.16  
## RelationshipSatisfaction 0.01559113 -0.107551190 0.12

plot(sorted\_diff\_att$CenterDifference)



Looking at the group with the higest attrition and the group with the lowest attrition, the attributes with the biggest difference between those groups are:

* TotalWorkingYears
* NumCompaniesWorked
* JobLevel
* MonthlyIncome
* Education
* JobRole
* MaritalStatus
* StockOptionLevel
* JobInvolvement

#### Decision Trees

# Getting Set Up  
HR\_tree <- HR\_clean  
HR\_tree <- HR\_tree[,2:length(HR\_tree)]  
  
# Dataset 1/3  
# set Seed for randomizer to always pick the same  
seedNum1 <- 23  
seedNum2 <- 465  
seedNum3 <- 1  
seedNum4 <- 987  
seedNum5 <- 307  
  
set.seed(23)  
# Generate random sample of rows  
randIndex1 <- sample(1:nrow(HR\_clean))  
# Set 2/3 Cutpoint of total rows  
cutPoint <- floor(nrow(HR\_clean)\*2/3)  
# Create train data based on the 2/3 value  
trainData1 <- HR\_tree[randIndex1[1:cutPoint],]  
# Create test data based on the remaining 1/3  
testData1 <- HR\_tree[randIndex1[(cutPoint+1):length(randIndex1)],]  
  
# Dataset 2/3  
set.seed(465)  
# Generate random sample of rows  
randIndex2 <- sample(1:nrow(HR\_clean))  
  
# Dataset 2/3  
set.seed(1)  
# Generate random sample of rows  
randIndex3 <- sample(1:nrow(HR\_clean))

To start, we’re running a decision tree with cp=0 on all the data to see how it plays out.

# Function  
# Decision Tree Function:   
# First variable is putting in the seedNumber. I've set 5 variables labeled as seedNum1 - seedNum5  
# Second variable is whichever dataset that is being generated  
printDecision <- function(seedNum, dataSet, depth=5){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 decisionTree <- rpart(Attrition ~ ., data = train, method="class", control=rpart.control(cp=0, minsplit = 5, maxdepth = depth))  
 summary(decisionTree)  
 # plot number of splits  
 rpart.plot(decisionTree, tweak=1.6)  
 # Predictions  
 predicted <- predict(decisionTree, test, type="class")  
 print(summary(predicted))  
 print(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
 set.seed(NULL)  
}

if("rpart" %in% rownames(installed.packages()) == FALSE) {install.packages('rpart') }  
if("rattle" %in% rownames(installed.packages()) == FALSE) {install.packages('rattle') }  
if("rpart.plot" %in% rownames(installed.packages()) == FALSE) {install.packages('rpart.plot') }  
library(rpart)  
library(rattle)

## Rattle: A free graphical interface for data science with R.  
## Version 5.2.0 Copyright (c) 2006-2018 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

library(rpart.plot)  
basicTree <- rpart(Attrition ~ ., data = trainData1, method="class", control=rpart.control(cp=0))  
summary(basicTree)

## Call:  
## rpart(formula = Attrition ~ ., data = trainData1, method = "class",   
## control = rpart.control(cp = 0))  
## n= 980   
##   
## CP nsplit rel error xerror xstd  
## 1 0.059375000 0 1.00000 1.00000 0.07231592  
## 2 0.031250000 2 0.88125 0.87500 0.06846532  
## 3 0.025000000 4 0.81875 0.90625 0.06946951  
## 4 0.020833333 5 0.79375 0.93750 0.07044524  
## 5 0.018750000 8 0.73125 0.94375 0.07063708  
## 6 0.012500000 9 0.71250 0.93750 0.07044524  
## 7 0.008333333 14 0.65000 0.99375 0.07213352  
## 8 0.006250000 17 0.62500 1.01250 0.07267769  
## 9 0.002083333 20 0.60625 1.08750 0.07476686  
## 10 0.000000000 23 0.60000 1.13750 0.07608589  
##   
## Variable importance  
## MonthlyIncome OverTime TotalWorkingYears   
## 13 10 6   
## JobRole DailyRate DistanceFromHome   
## 6 6 5   
## YearsAtCompany MonthlyRate MaritalStatus   
## 4 4 4   
## YearsWithCurrManager Department EducationField   
## 4 4 4   
## YearsInCurrentRole EnvironmentSatisfaction HourlyRate   
## 4 4 4   
## Age StockOptionLevel JobLevel   
## 3 3 3   
## BusinessTravel JobInvolvement YearsSinceLastPromotion   
## 2 2 1   
## JobSatisfaction Gender NumCompaniesWorked   
## 1 1 1   
## PercentSalaryHike   
## 1   
##   
## Node number 1: 980 observations, complexity param=0.059375  
## predicted class=No expected loss=0.1632653 P(node) =1  
## class counts: 820 160  
## probabilities: 0.837 0.163   
## left son=2 (767 obs) right son=3 (213 obs)  
## Primary splits:  
## MonthlyIncome < 2780 to the right, improve=19.41164, (0 missing)  
## OverTime splits as LR, improve=19.34035, (0 missing)  
## TotalWorkingYears < 1.5 to the right, improve=14.55748, (0 missing)  
## JobLevel splits as RLLLL, improve=14.47392, (0 missing)  
## JobRole splits as LRRLLLRRR, improve=12.10966, (0 missing)  
## Surrogate splits:  
## TotalWorkingYears < 3.5 to the right, agree=0.841, adj=0.268, (0 split)  
## JobLevel splits as RLLLL, agree=0.834, adj=0.235, (0 split)  
## Age < 23.5 to the right, agree=0.809, adj=0.122, (0 split)  
## JobRole splits as LLLLLLLLR, agree=0.801, adj=0.085, (0 split)  
## YearsAtCompany < 0.5 to the right, agree=0.785, adj=0.009, (0 split)  
##   
## Node number 2: 767 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.1108214 P(node) =0.7826531  
## class counts: 682 85  
## probabilities: 0.889 0.111   
## left son=4 (558 obs) right son=5 (209 obs)  
## Primary splits:  
## OverTime splits as LR, improve=7.474748, (0 missing)  
## StockOptionLevel splits as RLLL, improve=6.348036, (0 missing)  
## MaritalStatus splits as LLR, improve=4.600851, (0 missing)  
## JobRole splits as LRLLLLLRR, improve=4.578610, (0 missing)  
## Department splits as LLR, improve=3.972311, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 26.5 to the left, agree=0.729, adj=0.005, (0 split)  
##   
## Node number 3: 213 observations, complexity param=0.059375  
## predicted class=No expected loss=0.3521127 P(node) =0.2173469  
## class counts: 138 75  
## probabilities: 0.648 0.352   
## left son=6 (150 obs) right son=7 (63 obs)  
## Primary splits:  
## OverTime splits as LR, improve=15.961510, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve= 8.052241, (0 missing)  
## MonthlyRate < 25073 to the left, improve= 4.817714, (0 missing)  
## Age < 21.5 to the right, improve= 4.695013, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve= 4.511393, (0 missing)  
## Surrogate splits:  
## PercentSalaryHike < 11.5 to the right, agree=0.718, adj=0.048, (0 split)  
## DailyRate < 107.5 to the right, agree=0.714, adj=0.032, (0 split)  
## YearsSinceLastPromotion < 6.5 to the left, agree=0.714, adj=0.032, (0 split)  
## Education splits as LLLLR, agree=0.709, adj=0.016, (0 split)  
## MonthlyRate < 3046 to the right, agree=0.709, adj=0.016, (0 split)  
##   
## Node number 4: 558 observations, complexity param=0.008333333  
## predicted class=No expected loss=0.06810036 P(node) =0.5693878  
## class counts: 520 38  
## probabilities: 0.932 0.068   
## left son=8 (447 obs) right son=9 (111 obs)  
## Primary splits:  
## JobSatisfaction splits as RLLL, improve=2.004734, (0 missing)  
## StockOptionLevel splits as RLLR, improve=1.702476, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=1.301085, (0 missing)  
## Age < 33.5 to the right, improve=1.242657, (0 missing)  
## JobRole splits as LRRLLLLRR, improve=1.112509, (0 missing)  
## Surrogate splits:  
## Age < 59.5 to the left, agree=0.805, adj=0.018, (0 split)  
## PercentSalaryHike < 24.5 to the left, agree=0.803, adj=0.009, (0 split)  
## YearsWithCurrManager < 15.5 to the left, agree=0.803, adj=0.009, (0 split)  
##   
## Node number 5: 209 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.2248804 P(node) =0.2132653  
## class counts: 162 47  
## probabilities: 0.775 0.225   
## left son=10 (146 obs) right son=11 (63 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=8.695338, (0 missing)  
## StockOptionLevel splits as RLLL, improve=7.655439, (0 missing)  
## JobRole splits as LLRLLLLRR, improve=5.659909, (0 missing)  
## Department splits as LLR, improve=4.921394, (0 missing)  
## DistanceFromHome < 11.5 to the left, improve=3.682416, (0 missing)  
## Surrogate splits:  
## StockOptionLevel splits as RLLL, agree=0.876, adj=0.587, (0 split)  
## HourlyRate < 98.5 to the left, agree=0.713, adj=0.048, (0 split)  
## MonthlyRate < 2582 to the right, agree=0.713, adj=0.048, (0 split)  
## Age < 24.5 to the right, agree=0.708, adj=0.032, (0 split)  
## JobRole splits as LLLLLLLLR, agree=0.708, adj=0.032, (0 split)  
##   
## Node number 6: 150 observations, complexity param=0.03125  
## predicted class=No expected loss=0.2266667 P(node) =0.1530612  
## class counts: 116 34  
## probabilities: 0.773 0.227   
## left son=12 (96 obs) right son=13 (54 obs)  
## Primary splits:  
## YearsWithCurrManager < 0.5 to the right, improve=9.422315, (0 missing)  
## YearsAtCompany < 1.5 to the right, improve=6.140827, (0 missing)  
## TotalWorkingYears < 2.5 to the right, improve=5.819890, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=4.997185, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=4.650030, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 1.5 to the right, agree=0.947, adj=0.852, (0 split)  
## YearsInCurrentRole < 0.5 to the right, agree=0.893, adj=0.704, (0 split)  
## TotalWorkingYears < 1.5 to the right, agree=0.867, adj=0.630, (0 split)  
## MonthlyIncome < 1976 to the right, agree=0.760, adj=0.333, (0 split)  
## YearsSinceLastPromotion < 0.5 to the right, agree=0.720, adj=0.222, (0 split)  
##   
## Node number 7: 63 observations, complexity param=0.025  
## predicted class=Yes expected loss=0.3492063 P(node) =0.06428571  
## class counts: 22 41  
## probabilities: 0.349 0.651   
## left son=14 (18 obs) right son=15 (45 obs)  
## Primary splits:  
## MonthlyIncome < 2469.5 to the right, improve=3.457143, (0 missing)  
## DailyRate < 1129 to the right, improve=3.262580, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=3.250305, (0 missing)  
## DistanceFromHome < 16.5 to the left, improve=2.777778, (0 missing)  
## EducationField splits as LLRLRL, improve=1.920635, (0 missing)  
## Surrogate splits:  
## Age < 39.5 to the right, agree=0.778, adj=0.222, (0 split)  
## StockOptionLevel splits as RRLR, agree=0.746, adj=0.111, (0 split)  
## YearsInCurrentRole < 5 to the right, agree=0.746, adj=0.111, (0 split)  
## YearsSinceLastPromotion < 6 to the right, agree=0.746, adj=0.111, (0 split)  
## TotalWorkingYears < 13.5 to the right, agree=0.730, adj=0.056, (0 split)  
##   
## Node number 8: 447 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.04697987 P(node) =0.4561224  
## class counts: 426 21  
## probabilities: 0.953 0.047   
## left son=16 (226 obs) right son=17 (221 obs)  
## Primary splits:  
## StockOptionLevel splits as RLLR, improve=1.3292120, (0 missing)  
## BusinessTravel splits as LRL, improve=1.0482950, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=0.7610717, (0 missing)  
## YearsSinceLastPromotion < 5.5 to the left, improve=0.6421145, (0 missing)  
## JobInvolvement splits as RLLL, improve=0.6305365, (0 missing)  
## Surrogate splits:  
## MaritalStatus splits as LLR, agree=0.857, adj=0.710, (0 split)  
## HourlyRate < 53.5 to the right, agree=0.582, adj=0.154, (0 split)  
## YearsAtCompany < 6.5 to the right, agree=0.566, adj=0.122, (0 split)  
## JobRole splits as RLRLRRLLR, agree=0.555, adj=0.100, (0 split)  
## TotalWorkingYears < 7.5 to the right, agree=0.553, adj=0.095, (0 split)  
##   
## Node number 9: 111 observations, complexity param=0.008333333  
## predicted class=No expected loss=0.1531532 P(node) =0.1132653  
## class counts: 94 17  
## probabilities: 0.847 0.153   
## left son=18 (89 obs) right son=19 (22 obs)  
## Primary splits:  
## DailyRate < 417.5 to the right, improve=3.594631, (0 missing)  
## DistanceFromHome < 21.5 to the left, improve=3.409459, (0 missing)  
## JobRole splits as LRRLLLLRR, improve=3.117468, (0 missing)  
## Department splits as RLR, improve=1.723803, (0 missing)  
## TotalWorkingYears < 7.5 to the right, improve=1.621752, (0 missing)  
## Surrogate splits:  
## Department splits as RLL, agree=0.820, adj=0.091, (0 split)  
## JobRole splits as LRLLLLLLL, agree=0.820, adj=0.091, (0 split)  
## EducationField splits as RLLLLL, agree=0.811, adj=0.045, (0 split)  
##   
## Node number 10: 146 observations, complexity param=0.0125  
## predicted class=No expected loss=0.130137 P(node) =0.1489796  
## class counts: 127 19  
## probabilities: 0.870 0.130   
## left son=20 (124 obs) right son=21 (22 obs)  
## Primary splits:  
## DistanceFromHome < 21.5 to the left, improve=2.824589, (0 missing)  
## NumCompaniesWorked < 5.5 to the left, improve=2.154795, (0 missing)  
## YearsAtCompany < 3.5 to the right, improve=1.817952, (0 missing)  
## MonthlyRate < 21041.5 to the left, improve=1.733797, (0 missing)  
## HourlyRate < 71.5 to the right, improve=1.228609, (0 missing)  
##   
## Node number 11: 63 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.4444444 P(node) =0.06428571  
## class counts: 35 28  
## probabilities: 0.556 0.444   
## left son=22 (27 obs) right son=23 (36 obs)  
## Primary splits:  
## JobRole splits as LLRLLLLRR, improve=6.351852, (0 missing)  
## Department splits as LLR, improve=5.656566, (0 missing)  
## EducationField splits as -RRLRR, improve=4.424957, (0 missing)  
## TotalWorkingYears < 9.5 to the right, improve=3.968254, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=2.533983, (0 missing)  
## Surrogate splits:  
## Department splits as LLR, agree=0.873, adj=0.704, (0 split)  
## EducationField splits as -RRLLR, agree=0.683, adj=0.259, (0 split)  
## EnvironmentSatisfaction splits as RRLR, agree=0.683, adj=0.259, (0 split)  
## Gender splits as LR, agree=0.683, adj=0.259, (0 split)  
## MonthlyRate < 4437.5 to the left, agree=0.651, adj=0.185, (0 split)  
##   
## Node number 12: 96 observations  
## predicted class=No expected loss=0.09375 P(node) =0.09795918  
## class counts: 87 9  
## probabilities: 0.906 0.094   
##   
## Node number 13: 54 observations, complexity param=0.03125  
## predicted class=No expected loss=0.462963 P(node) =0.05510204  
## class counts: 29 25  
## probabilities: 0.537 0.463   
## left son=26 (36 obs) right son=27 (18 obs)  
## Primary splits:  
## HourlyRate < 56.5 to the right, improve=5.351852, (0 missing)  
## BusinessTravel splits as LRL, improve=3.188808, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.918059, (0 missing)  
## RelationshipSatisfaction splits as RLRL, improve=2.687079, (0 missing)  
## JobRole splits as -RR---L-R, improve=2.572043, (0 missing)  
## Surrogate splits:  
## EducationField splits as LLRLLL, agree=0.722, adj=0.167, (0 split)  
## WorkLifeBalance splits as LRLL, agree=0.722, adj=0.167, (0 split)  
## BusinessTravel splits as LRL, agree=0.704, adj=0.111, (0 split)  
## DailyRate < 1429 to the left, agree=0.704, adj=0.111, (0 split)  
## MonthlyRate < 25042.5 to the left, agree=0.704, adj=0.111, (0 split)  
##   
## Node number 14: 18 observations  
## predicted class=No expected loss=0.3888889 P(node) =0.01836735  
## class counts: 11 7  
## probabilities: 0.611 0.389   
##   
## Node number 15: 45 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.2444444 P(node) =0.04591837  
## class counts: 11 34  
## probabilities: 0.244 0.756   
## left son=30 (15 obs) right son=31 (30 obs)  
## Primary splits:  
## DailyRate < 1067.5 to the right, improve=3.755556, (0 missing)  
## DistanceFromHome < 12 to the left, improve=2.428674, (0 missing)  
## Education splits as LLRLL, improve=2.140741, (0 missing)  
## EnvironmentSatisfaction splits as RLRL, improve=2.029337, (0 missing)  
## TrainingTimesLastYear < 3.5 to the right, improve=1.679365, (0 missing)  
## Surrogate splits:  
## Age < 36 to the right, agree=0.711, adj=0.133, (0 split)  
## HourlyRate < 35 to the left, agree=0.711, adj=0.133, (0 split)  
## MonthlyIncome < 1349 to the left, agree=0.711, adj=0.133, (0 split)  
## Education splits as RRRRL, agree=0.689, adj=0.067, (0 split)  
## EnvironmentSatisfaction splits as RRRL, agree=0.689, adj=0.067, (0 split)  
##   
## Node number 16: 226 observations  
## predicted class=No expected loss=0.008849558 P(node) =0.2306122  
## class counts: 224 2  
## probabilities: 0.991 0.009   
##   
## Node number 17: 221 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.08597285 P(node) =0.2255102  
## class counts: 202 19  
## probabilities: 0.914 0.086   
## left son=34 (174 obs) right son=35 (47 obs)  
## Primary splits:  
## EnvironmentSatisfaction splits as RLLL, improve=1.329265, (0 missing)  
## YearsSinceLastPromotion < 6.5 to the left, improve=1.319341, (0 missing)  
## BusinessTravel splits as LRL, improve=1.201945, (0 missing)  
## DailyRate < 1334.5 to the left, improve=1.183280, (0 missing)  
## Age < 31.5 to the right, improve=1.142622, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 2506.5 to the right, agree=0.796, adj=0.043, (0 split)  
## TotalWorkingYears < 1.5 to the right, agree=0.792, adj=0.021, (0 split)  
## YearsInCurrentRole < 11.5 to the left, agree=0.792, adj=0.021, (0 split)  
##   
## Node number 18: 89 observations  
## predicted class=No expected loss=0.08988764 P(node) =0.09081633  
## class counts: 81 8  
## probabilities: 0.910 0.090   
##   
## Node number 19: 22 observations, complexity param=0.008333333  
## predicted class=No expected loss=0.4090909 P(node) =0.02244898  
## class counts: 13 9  
## probabilities: 0.591 0.409   
## left son=38 (8 obs) right son=39 (14 obs)  
## Primary splits:  
## DistanceFromHome < 8.5 to the left, improve=4.207792, (0 missing)  
## DailyRate < 300 to the left, improve=4.122078, (0 missing)  
## Department splits as LLR, improve=4.122078, (0 missing)  
## JobRole splits as LLL-LLLRR, improve=4.122078, (0 missing)  
## YearsInCurrentRole < 2.5 to the right, improve=3.103030, (0 missing)  
## Surrogate splits:  
## JobRole splits as LRL-RRLRR, agree=0.818, adj=0.500, (0 split)  
## EducationField splits as RRRLRL, agree=0.773, adj=0.375, (0 split)  
## Age < 41 to the right, agree=0.727, adj=0.250, (0 split)  
## DailyRate < 217.5 to the left, agree=0.727, adj=0.250, (0 split)  
## Department splits as RLR, agree=0.727, adj=0.250, (0 split)  
##   
## Node number 20: 124 observations, complexity param=0.0125  
## predicted class=No expected loss=0.08870968 P(node) =0.1265306  
## class counts: 113 11  
## probabilities: 0.911 0.089   
## left son=40 (99 obs) right son=41 (25 obs)  
## Primary splits:  
## MonthlyRate < 21715 to the left, improve=2.291619, (0 missing)  
## YearsAtCompany < 2.5 to the right, improve=1.505410, (0 missing)  
## JobInvolvement splits as RLLL, improve=1.401835, (0 missing)  
## NumCompaniesWorked < 2.5 to the left, improve=1.396495, (0 missing)  
## TotalWorkingYears < 5.5 to the right, improve=1.225010, (0 missing)  
## Surrogate splits:  
## NumCompaniesWorked < 8.5 to the left, agree=0.815, adj=0.08, (0 split)  
## YearsInCurrentRole < 11.5 to the left, agree=0.815, adj=0.08, (0 split)  
##   
## Node number 21: 22 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3636364 P(node) =0.02244898  
## class counts: 14 8  
## probabilities: 0.636 0.364   
## left son=42 (14 obs) right son=43 (8 obs)  
## Primary splits:  
## JobRole splits as RRLRLLLR-, improve=3.753247, (0 missing)  
## YearsInCurrentRole < 7.5 to the right, improve=2.715152, (0 missing)  
## EducationField splits as RRRLLL, improve=2.048485, (0 missing)  
## Gender splits as LR, improve=1.431818, (0 missing)  
## MonthlyIncome < 5542 to the left, improve=1.431818, (0 missing)  
## Surrogate splits:  
## Department splits as RLR, agree=0.909, adj=0.750, (0 split)  
## EducationField splits as RLRLLL, agree=0.818, adj=0.500, (0 split)  
## NumCompaniesWorked < 3.5 to the left, agree=0.773, adj=0.375, (0 split)  
## MonthlyRate < 12845 to the right, agree=0.727, adj=0.250, (0 split)  
## TrainingTimesLastYear < 2.5 to the right, agree=0.727, adj=0.250, (0 split)  
##   
## Node number 22: 27 observations, complexity param=0.00625  
## predicted class=No expected loss=0.1851852 P(node) =0.02755102  
## class counts: 22 5  
## probabilities: 0.815 0.185   
## left son=44 (20 obs) right son=45 (7 obs)  
## Primary splits:  
## DailyRate < 1011 to the left, improve=2.819577, (0 missing)  
## HourlyRate < 69.5 to the left, improve=1.481481, (0 missing)  
## JobLevel splits as RLRLR, improve=1.481481, (0 missing)  
## EducationField splits as -RRLRL, improve=1.273148, (0 missing)  
## MonthlyIncome < 4000 to the right, improve=1.119577, (0 missing)  
## Surrogate splits:  
## JobInvolvement splits as RLLL, agree=0.815, adj=0.286, (0 split)  
## Department splits as LLR, agree=0.778, adj=0.143, (0 split)  
## EducationField splits as -LRLLL, agree=0.778, adj=0.143, (0 split)  
## HourlyRate < 90.5 to the left, agree=0.778, adj=0.143, (0 split)  
## JobLevel splits as RLLLL, agree=0.778, adj=0.143, (0 split)  
##   
## Node number 23: 36 observations, complexity param=0.01875  
## predicted class=Yes expected loss=0.3611111 P(node) =0.03673469  
## class counts: 13 23  
## probabilities: 0.361 0.639   
## left son=46 (17 obs) right son=47 (19 obs)  
## Primary splits:  
## TotalWorkingYears < 9.5 to the right, improve=3.323185, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=2.777778, (0 missing)  
## MonthlyRate < 8860.5 to the left, improve=2.400202, (0 missing)  
## YearsAtCompany < 8.5 to the right, improve=2.400202, (0 missing)  
## JobInvolvement splits as RRLR, improve=2.312929, (0 missing)  
## Surrogate splits:  
## MonthlyIncome < 6489.5 to the right, agree=0.750, adj=0.471, (0 split)  
## YearsAtCompany < 8.5 to the right, agree=0.722, adj=0.412, (0 split)  
## YearsInCurrentRole < 4.5 to the right, agree=0.722, adj=0.412, (0 split)  
## JobLevel splits as RRLL-, agree=0.694, adj=0.353, (0 split)  
## MonthlyRate < 17153 to the left, agree=0.694, adj=0.353, (0 split)  
##   
## Node number 26: 36 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3055556 P(node) =0.03673469  
## class counts: 25 11  
## probabilities: 0.694 0.306   
## left son=52 (26 obs) right son=53 (10 obs)  
## Primary splits:  
## DistanceFromHome < 11 to the left, improve=2.400855, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.207544, (0 missing)  
## HourlyRate < 84.5 to the left, improve=2.177778, (0 missing)  
## RelationshipSatisfaction splits as RLLL, improve=2.099206, (0 missing)  
## Education splits as RRLR-, improve=1.525397, (0 missing)  
## Surrogate splits:  
## JobInvolvement splits as RLLR, agree=0.806, adj=0.3, (0 split)  
## DailyRate < 158 to the right, agree=0.778, adj=0.2, (0 split)  
## EducationField splits as RL-LRL, agree=0.778, adj=0.2, (0 split)  
## HourlyRate < 60 to the right, agree=0.778, adj=0.2, (0 split)  
## MonthlyIncome < 2543 to the left, agree=0.778, adj=0.2, (0 split)  
##   
## Node number 27: 18 observations  
## predicted class=Yes expected loss=0.2222222 P(node) =0.01836735  
## class counts: 4 14  
## probabilities: 0.222 0.778   
##   
## Node number 30: 15 observations  
## predicted class=No expected loss=0.4666667 P(node) =0.01530612  
## class counts: 8 7  
## probabilities: 0.533 0.467   
##   
## Node number 31: 30 observations  
## predicted class=Yes expected loss=0.1 P(node) =0.03061224  
## class counts: 3 27  
## probabilities: 0.100 0.900   
##   
## Node number 34: 174 observations  
## predicted class=No expected loss=0.05747126 P(node) =0.177551  
## class counts: 164 10  
## probabilities: 0.943 0.057   
##   
## Node number 35: 47 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.1914894 P(node) =0.04795918  
## class counts: 38 9  
## probabilities: 0.809 0.191   
## left son=70 (36 obs) right son=71 (11 obs)  
## Primary splits:  
## BusinessTravel splits as LRL, improve=3.598646, (0 missing)  
## HourlyRate < 52.5 to the left, improve=1.953191, (0 missing)  
## EducationField splits as RRRLLL, improve=1.764101, (0 missing)  
## JobRole splits as RRLLLLLR-, improve=1.633837, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=1.424537, (0 missing)  
## Surrogate splits:  
## EducationField splits as RLLLLL, agree=0.787, adj=0.091, (0 split)  
##   
## Node number 38: 8 observations  
## predicted class=No expected loss=0 P(node) =0.008163265  
## class counts: 8 0  
## probabilities: 1.000 0.000   
##   
## Node number 39: 14 observations  
## predicted class=Yes expected loss=0.3571429 P(node) =0.01428571  
## class counts: 5 9  
## probabilities: 0.357 0.643   
##   
## Node number 40: 99 observations  
## predicted class=No expected loss=0.04040404 P(node) =0.1010204  
## class counts: 95 4  
## probabilities: 0.960 0.040   
##   
## Node number 41: 25 observations, complexity param=0.0125  
## predicted class=No expected loss=0.28 P(node) =0.0255102  
## class counts: 18 7  
## probabilities: 0.720 0.280   
## left son=82 (17 obs) right son=83 (8 obs)  
## Primary splits:  
## EnvironmentSatisfaction splits as RRLL, improve=5.197647, (0 missing)  
## YearsAtCompany < 4.5 to the right, improve=2.768312, (0 missing)  
## JobRole splits as L-LRL-RR-, improve=2.613333, (0 missing)  
## EducationField splits as -LLRLR, improve=2.233846, (0 missing)  
## TotalWorkingYears < 7.5 to the right, improve=2.135556, (0 missing)  
## Surrogate splits:  
## Age < 31.5 to the right, agree=0.80, adj=0.375, (0 split)  
## JobInvolvement splits as RLLL, agree=0.80, adj=0.375, (0 split)  
## DistanceFromHome < 1.5 to the right, agree=0.76, adj=0.250, (0 split)  
## EducationField splits as -LLRLL, agree=0.76, adj=0.250, (0 split)  
## MonthlyIncome < 11825 to the left, agree=0.72, adj=0.125, (0 split)  
##   
## Node number 42: 14 observations  
## predicted class=No expected loss=0.1428571 P(node) =0.01428571  
## class counts: 12 2  
## probabilities: 0.857 0.143   
##   
## Node number 43: 8 observations  
## predicted class=Yes expected loss=0.25 P(node) =0.008163265  
## class counts: 2 6  
## probabilities: 0.250 0.750   
##   
## Node number 44: 20 observations  
## predicted class=No expected loss=0.05 P(node) =0.02040816  
## class counts: 19 1  
## probabilities: 0.950 0.050   
##   
## Node number 45: 7 observations  
## predicted class=Yes expected loss=0.4285714 P(node) =0.007142857  
## class counts: 3 4  
## probabilities: 0.429 0.571   
##   
## Node number 46: 17 observations  
## predicted class=No expected loss=0.4117647 P(node) =0.01734694  
## class counts: 10 7  
## probabilities: 0.588 0.412   
##   
## Node number 47: 19 observations  
## predicted class=Yes expected loss=0.1578947 P(node) =0.01938776  
## class counts: 3 16  
## probabilities: 0.158 0.842   
##   
## Node number 52: 26 observations, complexity param=0.00625  
## predicted class=No expected loss=0.1923077 P(node) =0.02653061  
## class counts: 21 5  
## probabilities: 0.808 0.192   
## left son=104 (19 obs) right son=105 (7 obs)  
## Primary splits:  
## MonthlyRate < 20229 to the left, improve=2.7536150, (0 missing)  
## HourlyRate < 84.5 to the left, improve=2.6223780, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=1.7501260, (0 missing)  
## JobSatisfaction splits as RRLL, improve=0.8864469, (0 missing)  
## MaritalStatus splits as LLR, improve=0.8864469, (0 missing)  
## Surrogate splits:  
## BusinessTravel splits as LRL, agree=0.808, adj=0.286, (0 split)  
## HourlyRate < 93 to the left, agree=0.808, adj=0.286, (0 split)  
## PercentSalaryHike < 19 to the left, agree=0.808, adj=0.286, (0 split)  
## PerformanceRating splits as LR, agree=0.808, adj=0.286, (0 split)  
## JobInvolvement splits as RLL-, agree=0.769, adj=0.143, (0 split)  
##   
## Node number 53: 10 observations  
## predicted class=Yes expected loss=0.4 P(node) =0.01020408  
## class counts: 4 6  
## probabilities: 0.400 0.600   
##   
## Node number 70: 36 observations  
## predicted class=No expected loss=0.08333333 P(node) =0.03673469  
## class counts: 33 3  
## probabilities: 0.917 0.083   
##   
## Node number 71: 11 observations  
## predicted class=Yes expected loss=0.4545455 P(node) =0.01122449  
## class counts: 5 6  
## probabilities: 0.455 0.545   
##   
## Node number 82: 17 observations  
## predicted class=No expected loss=0.05882353 P(node) =0.01734694  
## class counts: 16 1  
## probabilities: 0.941 0.059   
##   
## Node number 83: 8 observations  
## predicted class=Yes expected loss=0.25 P(node) =0.008163265  
## class counts: 2 6  
## probabilities: 0.250 0.750   
##   
## Node number 104: 19 observations  
## predicted class=No expected loss=0.05263158 P(node) =0.01938776  
## class counts: 18 1  
## probabilities: 0.947 0.053   
##   
## Node number 105: 7 observations  
## predicted class=Yes expected loss=0.4285714 P(node) =0.007142857  
## class counts: 3 4  
## probabilities: 0.429 0.571

#predict the test dataset using the model for train tree No. 1  
basicPredict <- predict(basicTree, testData1, type="class")  
#plot number of splits  
summary(basicPredict)

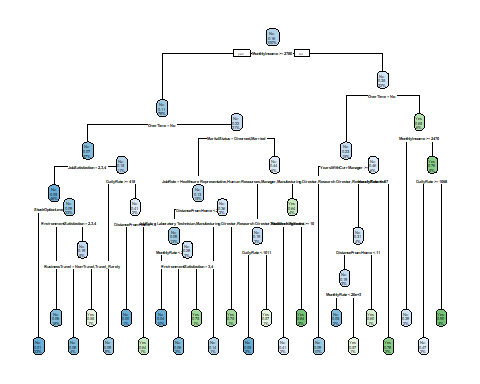
## No Yes   
## 432 58

table(predictedAttrition=basicPredict, actualAttrition=testData1$Attrition)

## actualAttrition  
## predictedAttrition No Yes  
## No 375 57  
## Yes 38 20

#Disputed Prediction  
rpart.plot(basicTree, tweak=1.6)

## Warning: labs do not fit even at cex 0.15, there may be some overplotting

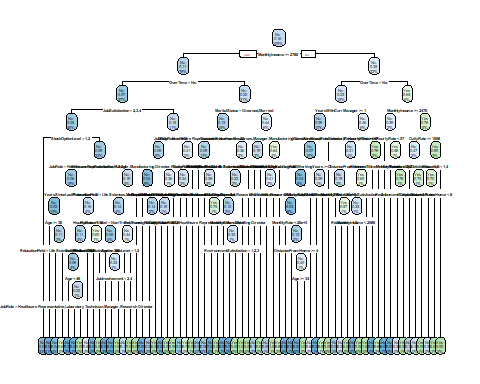


Prediction Accuracy: 395/490 = ~.806%

# Increase minSplit and maxDepth  
advancedTree <- printDecision(seedNum1, HR\_tree, 10)

## Call:  
## rpart(formula = Attrition ~ ., data = train, method = "class",   
## control = rpart.control(cp = 0, minsplit = 5, maxdepth = depth))  
## n= 980   
##   
## CP nsplit rel error xerror xstd  
## 1 0.059375000 0 1.00000 1.00000 0.07231592  
## 2 0.031250000 2 0.88125 0.90000 0.06927099  
## 3 0.025000000 4 0.81875 0.93750 0.07044524  
## 4 0.020833333 5 0.79375 0.91875 0.06986315  
## 5 0.015625000 10 0.67500 0.94375 0.07063708  
## 6 0.012500000 14 0.61250 0.93750 0.07044524  
## 7 0.010416667 26 0.46250 0.99375 0.07213352  
## 8 0.009375000 30 0.41875 1.03125 0.07321292  
## 9 0.006250000 37 0.35000 1.07500 0.07442809  
## 10 0.004166667 55 0.23125 1.17500 0.07703862  
## 11 0.003125000 58 0.21875 1.18125 0.07719446  
## 12 0.000000000 64 0.20000 1.20625 0.07780957  
##   
## Variable importance  
## MonthlyIncome TotalWorkingYears DailyRate   
## 11 7 7   
## MonthlyRate JobRole OverTime   
## 6 5 5   
## EducationField EnvironmentSatisfaction DistanceFromHome   
## 4 4 4   
## Age HourlyRate YearsAtCompany   
## 4 4 4   
## YearsWithCurrManager Department JobLevel   
## 3 3 3   
## NumCompaniesWorked WorkLifeBalance YearsInCurrentRole   
## 3 3 2   
## MaritalStatus JobInvolvement Education   
## 2 2 2   
## BusinessTravel RelationshipSatisfaction StockOptionLevel   
## 2 2 2   
## YearsSinceLastPromotion JobSatisfaction Gender   
## 2 1 1   
## PercentSalaryHike   
## 1   
##   
## Node number 1: 980 observations, complexity param=0.059375  
## predicted class=No expected loss=0.1632653 P(node) =1  
## class counts: 820 160  
## probabilities: 0.837 0.163   
## left son=2 (767 obs) right son=3 (213 obs)  
## Primary splits:  
## MonthlyIncome < 2780 to the right, improve=19.41164, (0 missing)  
## OverTime splits as LR, improve=19.34035, (0 missing)  
## TotalWorkingYears < 1.5 to the right, improve=14.55748, (0 missing)  
## JobLevel splits as RLLLL, improve=14.47392, (0 missing)  
## JobRole splits as LRRLLLRRR, improve=12.10966, (0 missing)  
## Surrogate splits:  
## TotalWorkingYears < 3.5 to the right, agree=0.841, adj=0.268, (0 split)  
## JobLevel splits as RLLLL, agree=0.834, adj=0.235, (0 split)  
## Age < 23.5 to the right, agree=0.809, adj=0.122, (0 split)  
## JobRole splits as LLLLLLLLR, agree=0.801, adj=0.085, (0 split)  
## YearsAtCompany < 0.5 to the right, agree=0.785, adj=0.009, (0 split)  
##   
## Node number 2: 767 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.1108214 P(node) =0.7826531  
## class counts: 682 85  
## probabilities: 0.889 0.111   
## left son=4 (558 obs) right son=5 (209 obs)  
## Primary splits:  
## OverTime splits as LR, improve=7.474748, (0 missing)  
## StockOptionLevel splits as RLLL, improve=6.348036, (0 missing)  
## MaritalStatus splits as LLR, improve=4.600851, (0 missing)  
## JobRole splits as LRLLLLLRR, improve=4.578610, (0 missing)  
## Department splits as LLR, improve=3.972311, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 26.5 to the left, agree=0.729, adj=0.005, (0 split)  
##   
## Node number 3: 213 observations, complexity param=0.059375  
## predicted class=No expected loss=0.3521127 P(node) =0.2173469  
## class counts: 138 75  
## probabilities: 0.648 0.352   
## left son=6 (150 obs) right son=7 (63 obs)  
## Primary splits:  
## OverTime splits as LR, improve=15.961510, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve= 8.052241, (0 missing)  
## MonthlyRate < 25073 to the left, improve= 4.817714, (0 missing)  
## Age < 21.5 to the right, improve= 4.695013, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve= 4.511393, (0 missing)  
## Surrogate splits:  
## PercentSalaryHike < 11.5 to the right, agree=0.718, adj=0.048, (0 split)  
## DailyRate < 107.5 to the right, agree=0.714, adj=0.032, (0 split)  
## YearsSinceLastPromotion < 6.5 to the left, agree=0.714, adj=0.032, (0 split)  
## Education splits as LLLLR, agree=0.709, adj=0.016, (0 split)  
## MonthlyRate < 3046 to the right, agree=0.709, adj=0.016, (0 split)  
##   
## Node number 4: 558 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.06810036 P(node) =0.5693878  
## class counts: 520 38  
## probabilities: 0.932 0.068   
## left son=8 (447 obs) right son=9 (111 obs)  
## Primary splits:  
## JobSatisfaction splits as RLLL, improve=2.004734, (0 missing)  
## StockOptionLevel splits as RLLR, improve=1.702476, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=1.301085, (0 missing)  
## Age < 33.5 to the right, improve=1.242657, (0 missing)  
## JobRole splits as LRRLLLLRR, improve=1.112509, (0 missing)  
## Surrogate splits:  
## Age < 59.5 to the left, agree=0.805, adj=0.018, (0 split)  
## PercentSalaryHike < 24.5 to the left, agree=0.803, adj=0.009, (0 split)  
## YearsWithCurrManager < 15.5 to the left, agree=0.803, adj=0.009, (0 split)  
##   
## Node number 5: 209 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.2248804 P(node) =0.2132653  
## class counts: 162 47  
## probabilities: 0.775 0.225   
## left son=10 (146 obs) right son=11 (63 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=8.695338, (0 missing)  
## StockOptionLevel splits as RLLL, improve=7.655439, (0 missing)  
## JobRole splits as LLRLLLLRR, improve=5.659909, (0 missing)  
## Department splits as LLR, improve=4.921394, (0 missing)  
## DistanceFromHome < 11.5 to the left, improve=3.682416, (0 missing)  
## Surrogate splits:  
## StockOptionLevel splits as RLLL, agree=0.876, adj=0.587, (0 split)  
## HourlyRate < 98.5 to the left, agree=0.713, adj=0.048, (0 split)  
## MonthlyRate < 2582 to the right, agree=0.713, adj=0.048, (0 split)  
## Age < 24.5 to the right, agree=0.708, adj=0.032, (0 split)  
## JobRole splits as LLLLLLLLR, agree=0.708, adj=0.032, (0 split)  
##   
## Node number 6: 150 observations, complexity param=0.03125  
## predicted class=No expected loss=0.2266667 P(node) =0.1530612  
## class counts: 116 34  
## probabilities: 0.773 0.227   
## left son=12 (96 obs) right son=13 (54 obs)  
## Primary splits:  
## YearsWithCurrManager < 0.5 to the right, improve=9.422315, (0 missing)  
## YearsAtCompany < 1.5 to the right, improve=6.140827, (0 missing)  
## TotalWorkingYears < 2.5 to the right, improve=5.819890, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=4.997185, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=4.650030, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 1.5 to the right, agree=0.947, adj=0.852, (0 split)  
## YearsInCurrentRole < 0.5 to the right, agree=0.893, adj=0.704, (0 split)  
## TotalWorkingYears < 1.5 to the right, agree=0.867, adj=0.630, (0 split)  
## MonthlyIncome < 1976 to the right, agree=0.760, adj=0.333, (0 split)  
## YearsSinceLastPromotion < 0.5 to the right, agree=0.720, adj=0.222, (0 split)  
##   
## Node number 7: 63 observations, complexity param=0.025  
## predicted class=Yes expected loss=0.3492063 P(node) =0.06428571  
## class counts: 22 41  
## probabilities: 0.349 0.651   
## left son=14 (18 obs) right son=15 (45 obs)  
## Primary splits:  
## MonthlyIncome < 2469.5 to the right, improve=3.457143, (0 missing)  
## DailyRate < 1129 to the right, improve=3.262580, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=3.250305, (0 missing)  
## NumCompaniesWorked < 0.5 to the left, improve=3.108605, (0 missing)  
## DistanceFromHome < 16.5 to the left, improve=2.777778, (0 missing)  
## Surrogate splits:  
## Age < 39.5 to the right, agree=0.778, adj=0.222, (0 split)  
## StockOptionLevel splits as RRLR, agree=0.746, adj=0.111, (0 split)  
## YearsInCurrentRole < 5 to the right, agree=0.746, adj=0.111, (0 split)  
## YearsSinceLastPromotion < 6 to the right, agree=0.746, adj=0.111, (0 split)  
## TotalWorkingYears < 13.5 to the right, agree=0.730, adj=0.056, (0 split)  
##   
## Node number 8: 447 observations, complexity param=0.00625  
## predicted class=No expected loss=0.04697987 P(node) =0.4561224  
## class counts: 426 21  
## probabilities: 0.953 0.047   
## left son=16 (226 obs) right son=17 (221 obs)  
## Primary splits:  
## StockOptionLevel splits as RLLR, improve=1.3292120, (0 missing)  
## BusinessTravel splits as LRL, improve=1.0482950, (0 missing)  
## YearsAtCompany < 29.5 to the left, improve=0.8245984, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=0.7610717, (0 missing)  
## YearsSinceLastPromotion < 5.5 to the left, improve=0.6421145, (0 missing)  
## Surrogate splits:  
## MaritalStatus splits as LLR, agree=0.857, adj=0.710, (0 split)  
## HourlyRate < 53.5 to the right, agree=0.582, adj=0.154, (0 split)  
## YearsAtCompany < 6.5 to the right, agree=0.566, adj=0.122, (0 split)  
## JobRole splits as RLRLRRLLR, agree=0.555, adj=0.100, (0 split)  
## TotalWorkingYears < 7.5 to the right, agree=0.553, adj=0.095, (0 split)  
##   
## Node number 9: 111 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.1531532 P(node) =0.1132653  
## class counts: 94 17  
## probabilities: 0.847 0.153   
## left son=18 (89 obs) right son=19 (22 obs)  
## Primary splits:  
## DailyRate < 417.5 to the right, improve=3.594631, (0 missing)  
## DistanceFromHome < 21.5 to the left, improve=3.409459, (0 missing)  
## JobRole splits as LRRLLLLRR, improve=3.117468, (0 missing)  
## Department splits as RLR, improve=1.723803, (0 missing)  
## TotalWorkingYears < 7.5 to the right, improve=1.621752, (0 missing)  
## Surrogate splits:  
## Department splits as RLL, agree=0.820, adj=0.091, (0 split)  
## JobRole splits as LRLLLLLLL, agree=0.820, adj=0.091, (0 split)  
## EducationField splits as RLLLLL, agree=0.811, adj=0.045, (0 split)  
##   
## Node number 10: 146 observations, complexity param=0.0125  
## predicted class=No expected loss=0.130137 P(node) =0.1489796  
## class counts: 127 19  
## probabilities: 0.870 0.130   
## left son=20 (124 obs) right son=21 (22 obs)  
## Primary splits:  
## DistanceFromHome < 21.5 to the left, improve=2.824589, (0 missing)  
## NumCompaniesWorked < 5.5 to the left, improve=2.154795, (0 missing)  
## YearsAtCompany < 3.5 to the right, improve=1.817952, (0 missing)  
## MonthlyRate < 21041.5 to the left, improve=1.733797, (0 missing)  
## TrainingTimesLastYear < 0.5 to the right, improve=1.711937, (0 missing)  
##   
## Node number 11: 63 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.4444444 P(node) =0.06428571  
## class counts: 35 28  
## probabilities: 0.556 0.444   
## left son=22 (27 obs) right son=23 (36 obs)  
## Primary splits:  
## JobRole splits as LLRLLLLRR, improve=6.351852, (0 missing)  
## Department splits as LLR, improve=5.656566, (0 missing)  
## EducationField splits as -RRLRR, improve=4.424957, (0 missing)  
## TotalWorkingYears < 9.5 to the right, improve=3.968254, (0 missing)  
## DailyRate < 1412.5 to the left, improve=2.636535, (0 missing)  
## Surrogate splits:  
## Department splits as LLR, agree=0.873, adj=0.704, (0 split)  
## EducationField splits as -RRLLR, agree=0.683, adj=0.259, (0 split)  
## EnvironmentSatisfaction splits as RRLR, agree=0.683, adj=0.259, (0 split)  
## Gender splits as LR, agree=0.683, adj=0.259, (0 split)  
## MonthlyRate < 4437.5 to the left, agree=0.651, adj=0.185, (0 split)  
##   
## Node number 12: 96 observations, complexity param=0.0125  
## predicted class=No expected loss=0.09375 P(node) =0.09795918  
## class counts: 87 9  
## probabilities: 0.906 0.094   
## left son=24 (94 obs) right son=25 (2 obs)  
## Primary splits:  
## YearsSinceLastPromotion < 8 to the left, improve=3.355053, (0 missing)  
## EducationField splits as LLLLLR, improve=1.809826, (0 missing)  
## JobSatisfaction splits as RLRL, improve=1.397156, (0 missing)  
## MonthlyRate < 4005 to the right, improve=1.377717, (0 missing)  
## YearsInCurrentRole < 8 to the left, improve=1.377717, (0 missing)  
##   
## Node number 13: 54 observations, complexity param=0.03125  
## predicted class=No expected loss=0.462963 P(node) =0.05510204  
## class counts: 29 25  
## probabilities: 0.537 0.463   
## left son=26 (36 obs) right son=27 (18 obs)  
## Primary splits:  
## HourlyRate < 56.5 to the right, improve=5.351852, (0 missing)  
## BusinessTravel splits as LRL, improve=3.188808, (0 missing)  
## MonthlyRate < 24118 to the left, improve=3.178382, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.918059, (0 missing)  
## RelationshipSatisfaction splits as RLRL, improve=2.687079, (0 missing)  
## Surrogate splits:  
## EducationField splits as LLRLLL, agree=0.722, adj=0.167, (0 split)  
## WorkLifeBalance splits as LRLL, agree=0.722, adj=0.167, (0 split)  
## BusinessTravel splits as LRL, agree=0.704, adj=0.111, (0 split)  
## DailyRate < 1429 to the left, agree=0.704, adj=0.111, (0 split)  
## MonthlyRate < 25042.5 to the left, agree=0.704, adj=0.111, (0 split)  
##   
## Node number 14: 18 observations, complexity param=0.015625  
## predicted class=No expected loss=0.3888889 P(node) =0.01836735  
## class counts: 11 7  
## probabilities: 0.611 0.389   
## left son=28 (6 obs) right son=29 (12 obs)  
## Primary splits:  
## HourlyRate < 56.5 to the left, improve=2.722222, (0 missing)  
## MonthlyIncome < 2624 to the left, improve=2.722222, (0 missing)  
## YearsInCurrentRole < 6.5 to the left, improve=2.340171, (0 missing)  
## EducationField splits as -LRLRL, improve=1.680556, (0 missing)  
## JobInvolvement splits as RLLR, improve=1.680556, (0 missing)  
## Surrogate splits:  
## DailyRate < 347.5 to the left, agree=0.778, adj=0.333, (0 split)  
## Education splits as LRRR-, agree=0.778, adj=0.333, (0 split)  
## TrainingTimesLastYear < 2.5 to the right, agree=0.778, adj=0.333, (0 split)  
## YearsInCurrentRole < 1.5 to the left, agree=0.778, adj=0.333, (0 split)  
## DistanceFromHome < 2.5 to the left, agree=0.722, adj=0.167, (0 split)  
##   
## Node number 15: 45 observations, complexity param=0.015625  
## predicted class=Yes expected loss=0.2444444 P(node) =0.04591837  
## class counts: 11 34  
## probabilities: 0.244 0.756   
## left son=30 (15 obs) right son=31 (30 obs)  
## Primary splits:  
## DailyRate < 1067.5 to the right, improve=3.755556, (0 missing)  
## NumCompaniesWorked < 0.5 to the left, improve=3.669841, (0 missing)  
## DistanceFromHome < 12 to the left, improve=2.428674, (0 missing)  
## JobInvolvement splits as RRRL, improve=2.244173, (0 missing)  
## Education splits as LLRLL, improve=2.140741, (0 missing)  
## Surrogate splits:  
## Age < 36 to the right, agree=0.711, adj=0.133, (0 split)  
## HourlyRate < 35 to the left, agree=0.711, adj=0.133, (0 split)  
## MonthlyIncome < 1349 to the left, agree=0.711, adj=0.133, (0 split)  
## Education splits as RRRRL, agree=0.689, adj=0.067, (0 split)  
## EnvironmentSatisfaction splits as RRRL, agree=0.689, adj=0.067, (0 split)  
##   
## Node number 16: 226 observations  
## predicted class=No expected loss=0.008849558 P(node) =0.2306122  
## class counts: 224 2  
## probabilities: 0.991 0.009   
##   
## Node number 17: 221 observations, complexity param=0.00625  
## predicted class=No expected loss=0.08597285 P(node) =0.2255102  
## class counts: 202 19  
## probabilities: 0.914 0.086   
## left son=34 (174 obs) right son=35 (47 obs)  
## Primary splits:  
## EnvironmentSatisfaction splits as RLLL, improve=1.329265, (0 missing)  
## YearsSinceLastPromotion < 6.5 to the left, improve=1.319341, (0 missing)  
## BusinessTravel splits as LRL, improve=1.201945, (0 missing)  
## DailyRate < 1334.5 to the left, improve=1.183280, (0 missing)  
## Age < 31.5 to the right, improve=1.142622, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 2506.5 to the right, agree=0.796, adj=0.043, (0 split)  
## TotalWorkingYears < 1.5 to the right, agree=0.792, adj=0.021, (0 split)  
## YearsInCurrentRole < 11.5 to the left, agree=0.792, adj=0.021, (0 split)  
##   
## Node number 18: 89 observations, complexity param=0.009375  
## predicted class=No expected loss=0.08988764 P(node) =0.09081633  
## class counts: 81 8  
## probabilities: 0.910 0.090   
## left son=36 (75 obs) right son=37 (14 obs)  
## Primary splits:  
## JobRole splits as LRRLLLLLR, improve=2.3732260, (0 missing)  
## DailyRate < 1360 to the left, improve=1.8811450, (0 missing)  
## NumCompaniesWorked < 8.5 to the left, improve=1.4088570, (0 missing)  
## JobInvolvement splits as LRLL, improve=0.8430478, (0 missing)  
## EnvironmentSatisfaction splits as RRLL, improve=0.7670888, (0 missing)  
## Surrogate splits:  
## MonthlyIncome < 3579 to the right, agree=0.888, adj=0.286, (0 split)  
## JobLevel splits as RLLLL, agree=0.876, adj=0.214, (0 split)  
## HourlyRate < 96.5 to the left, agree=0.865, adj=0.143, (0 split)  
## Department splits as RLL, agree=0.854, adj=0.071, (0 split)  
##   
## Node number 19: 22 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.4090909 P(node) =0.02244898  
## class counts: 13 9  
## probabilities: 0.591 0.409   
## left son=38 (17 obs) right son=39 (5 obs)  
## Primary splits:  
## DailyRate < 333 to the left, improve=4.518717, (0 missing)  
## DistanceFromHome < 8.5 to the left, improve=4.207792, (0 missing)  
## Department splits as LLR, improve=4.122078, (0 missing)  
## JobRole splits as LLL-LLLRR, improve=4.122078, (0 missing)  
## YearsInCurrentRole < 2.5 to the right, improve=3.103030, (0 missing)  
## Surrogate splits:  
## DistanceFromHome < 17.5 to the left, agree=0.818, adj=0.2, (0 split)  
## EducationField splits as RLLLLL, agree=0.818, adj=0.2, (0 split)  
## JobRole splits as LLL-LLLLR, agree=0.818, adj=0.2, (0 split)  
## NumCompaniesWorked < 0.5 to the right, agree=0.818, adj=0.2, (0 split)  
##   
## Node number 20: 124 observations, complexity param=0.0125  
## predicted class=No expected loss=0.08870968 P(node) =0.1265306  
## class counts: 113 11  
## probabilities: 0.911 0.089   
## left son=40 (99 obs) right son=41 (25 obs)  
## Primary splits:  
## MonthlyRate < 21715 to the left, improve=2.291619, (0 missing)  
## YearsAtCompany < 2.5 to the right, improve=1.505410, (0 missing)  
## JobInvolvement splits as RLLL, improve=1.401835, (0 missing)  
## NumCompaniesWorked < 2.5 to the left, improve=1.396495, (0 missing)  
## TotalWorkingYears < 5.5 to the right, improve=1.225010, (0 missing)  
## Surrogate splits:  
## NumCompaniesWorked < 8.5 to the left, agree=0.815, adj=0.08, (0 split)  
## YearsInCurrentRole < 11.5 to the left, agree=0.815, adj=0.08, (0 split)  
##   
## Node number 21: 22 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3636364 P(node) =0.02244898  
## class counts: 14 8  
## probabilities: 0.636 0.364   
## left son=42 (18 obs) right son=43 (4 obs)  
## Primary splits:  
## EducationField splits as RLRLLL, improve=3.959596, (0 missing)  
## JobRole splits as RRLRLLLR-, improve=3.753247, (0 missing)  
## YearsInCurrentRole < 7.5 to the right, improve=2.715152, (0 missing)  
## YearsAtCompany < 11 to the right, improve=1.711230, (0 missing)  
## Department splits as RLR, improve=1.515152, (0 missing)  
## Surrogate splits:  
## Department splits as RLR, agree=0.909, adj=0.5, (0 split)  
## JobRole splits as LRLLLLLR-, agree=0.909, adj=0.5, (0 split)  
##   
## Node number 22: 27 observations, complexity param=0.0125  
## predicted class=No expected loss=0.1851852 P(node) =0.02755102  
## class counts: 22 5  
## probabilities: 0.815 0.185   
## left son=44 (25 obs) right son=45 (2 obs)  
## Primary splits:  
## JobInvolvement splits as RLLL, improve=2.868148, (0 missing)  
## DailyRate < 1011 to the left, improve=2.819577, (0 missing)  
## YearsSinceLastPromotion < 5 to the left, improve=2.111785, (0 missing)  
## Education splits as RLLLL, improve=1.564815, (0 missing)  
## RelationshipSatisfaction splits as RLLL, improve=1.529101, (0 missing)  
##   
## Node number 23: 36 observations, complexity param=0.02083333  
## predicted class=Yes expected loss=0.3611111 P(node) =0.03673469  
## class counts: 13 23  
## probabilities: 0.361 0.639   
## left son=46 (17 obs) right son=47 (19 obs)  
## Primary splits:  
## TotalWorkingYears < 9.5 to the right, improve=3.323185, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=2.777778, (0 missing)  
## MonthlyRate < 8860.5 to the left, improve=2.400202, (0 missing)  
## YearsAtCompany < 8.5 to the right, improve=2.400202, (0 missing)  
## JobInvolvement splits as RRLR, improve=2.312929, (0 missing)  
## Surrogate splits:  
## MonthlyIncome < 6489.5 to the right, agree=0.750, adj=0.471, (0 split)  
## YearsAtCompany < 8.5 to the right, agree=0.722, adj=0.412, (0 split)  
## YearsInCurrentRole < 4.5 to the right, agree=0.722, adj=0.412, (0 split)  
## JobLevel splits as RRLL-, agree=0.694, adj=0.353, (0 split)  
## MonthlyRate < 17153 to the left, agree=0.694, adj=0.353, (0 split)  
##   
## Node number 24: 94 observations, complexity param=0.009375  
## predicted class=No expected loss=0.07446809 P(node) =0.09591837  
## class counts: 87 7  
## probabilities: 0.926 0.074   
## left son=48 (80 obs) right son=49 (14 obs)  
## Primary splits:  
## TotalWorkingYears < 2.5 to the right, improve=1.468161, (0 missing)  
## EducationField splits as LLLLLR, improve=1.138399, (0 missing)  
## Age < 21.5 to the right, improve=1.119245, (0 missing)  
## MonthlyRate < 18752 to the left, improve=1.073389, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=1.048488, (0 missing)  
## Surrogate splits:  
## Age < 20.5 to the right, agree=0.883, adj=0.214, (0 split)  
## YearsAtCompany < 1.5 to the right, agree=0.883, adj=0.214, (0 split)  
## EducationField splits as LLRLLL, agree=0.862, adj=0.071, (0 split)  
##   
## Node number 25: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 26: 36 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3055556 P(node) =0.03673469  
## class counts: 25 11  
## probabilities: 0.694 0.306   
## left son=52 (26 obs) right son=53 (10 obs)  
## Primary splits:  
## DistanceFromHome < 11 to the left, improve=2.400855, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.207544, (0 missing)  
## HourlyRate < 84.5 to the left, improve=2.177778, (0 missing)  
## RelationshipSatisfaction splits as RLLL, improve=2.099206, (0 missing)  
## MonthlyRate < 24118 to the left, improve=2.042484, (0 missing)  
## Surrogate splits:  
## JobInvolvement splits as RLLR, agree=0.806, adj=0.3, (0 split)  
## DailyRate < 158 to the right, agree=0.778, adj=0.2, (0 split)  
## EducationField splits as RL-LRL, agree=0.778, adj=0.2, (0 split)  
## HourlyRate < 60 to the right, agree=0.778, adj=0.2, (0 split)  
## MonthlyIncome < 2543 to the left, agree=0.778, adj=0.2, (0 split)  
##   
## Node number 27: 18 observations, complexity param=0.0125  
## predicted class=Yes expected loss=0.2222222 P(node) =0.01836735  
## class counts: 4 14  
## probabilities: 0.222 0.778   
## left son=54 (2 obs) right son=55 (16 obs)  
## Primary splits:  
## BusinessTravel splits as LRR, improve=2.722222, (0 missing)  
## DailyRate < 1382.5 to the right, improve=2.722222, (0 missing)  
## Age < 34.5 to the right, improve=1.976068, (0 missing)  
## JobRole splits as -RR---L-R, improve=1.976068, (0 missing)  
## YearsAtCompany < 0.5 to the left, improve=1.976068, (0 missing)  
##   
## Node number 28: 6 observations  
## predicted class=No expected loss=0 P(node) =0.006122449  
## class counts: 6 0  
## probabilities: 1.000 0.000   
##   
## Node number 29: 12 observations, complexity param=0.015625  
## predicted class=Yes expected loss=0.4166667 P(node) =0.0122449  
## class counts: 5 7  
## probabilities: 0.417 0.583   
## left son=58 (3 obs) right son=59 (9 obs)  
## Primary splits:  
## MonthlyIncome < 2621 to the left, improve=2.722222, (0 missing)  
## DistanceFromHome < 4 to the right, improve=2.083333, (0 missing)  
## JobSatisfaction splits as LLRL, improve=2.083333, (0 missing)  
## PercentSalaryHike < 14.5 to the right, improve=1.633333, (0 missing)  
## BusinessTravel splits as -RL, improve=1.388889, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 20652 to the right, agree=0.833, adj=0.333, (0 split)  
## RelationshipSatisfaction splits as -LRR, agree=0.833, adj=0.333, (0 split)  
##   
## Node number 30: 15 observations, complexity param=0.015625  
## predicted class=No expected loss=0.4666667 P(node) =0.01530612  
## class counts: 8 7  
## probabilities: 0.533 0.467   
## left son=60 (7 obs) right son=61 (8 obs)  
## Primary splits:  
## RelationshipSatisfaction splits as LRLR, improve=2.752381, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=2.133333, (0 missing)  
## MonthlyRate < 4623.5 to the right, improve=2.133333, (0 missing)  
## NumCompaniesWorked < 4.5 to the left, improve=2.133333, (0 missing)  
## DailyRate < 1301.5 to the left, improve=1.800000, (0 missing)  
## Surrogate splits:  
## DistanceFromHome < 5.5 to the left, agree=0.867, adj=0.714, (0 split)  
## WorkLifeBalance splits as RLRL, agree=0.867, adj=0.714, (0 split)  
## DailyRate < 1301.5 to the left, agree=0.800, adj=0.571, (0 split)  
## EducationField splits as LLRRRR, agree=0.733, adj=0.429, (0 split)  
## HourlyRate < 64.5 to the left, agree=0.733, adj=0.429, (0 split)  
##   
## Node number 31: 30 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.1 P(node) =0.03061224  
## class counts: 3 27  
## probabilities: 0.100 0.900   
## left son=62 (10 obs) right son=63 (20 obs)  
## Primary splits:  
## Education splits as LRRL-, improve=1.2000000, (0 missing)  
## EducationField splits as RRRLRR, improve=0.9000000, (0 missing)  
## PercentSalaryHike < 11.5 to the left, improve=0.8166667, (0 missing)  
## JobInvolvement splits as RRRL, improve=0.6857143, (0 missing)  
## DistanceFromHome < 7.5 to the left, improve=0.6000000, (0 missing)  
## Surrogate splits:  
## EnvironmentSatisfaction splits as RRRL, agree=0.767, adj=0.3, (0 split)  
## WorkLifeBalance splits as RLRR, agree=0.767, adj=0.3, (0 split)  
## EducationField splits as LRLRRR, agree=0.733, adj=0.2, (0 split)  
## MonthlyRate < 23430.5 to the right, agree=0.733, adj=0.2, (0 split)  
## RelationshipSatisfaction splits as RRRL, agree=0.733, adj=0.2, (0 split)  
##   
## Node number 34: 174 observations, complexity param=0.00625  
## predicted class=No expected loss=0.05747126 P(node) =0.177551  
## class counts: 164 10  
## probabilities: 0.943 0.057   
## left son=68 (131 obs) right son=69 (43 obs)  
## Primary splits:  
## YearsSinceLastPromotion < 3.5 to the left, improve=1.2670490, (0 missing)  
## DailyRate < 1358 to the left, improve=0.8438857, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=0.6986367, (0 missing)  
## YearsAtCompany < 6.5 to the left, improve=0.5749337, (0 missing)  
## YearsWithCurrManager < 6.5 to the left, improve=0.5454611, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 12.5 to the left, agree=0.805, adj=0.209, (0 split)  
## MonthlyIncome < 18941.5 to the left, agree=0.787, adj=0.140, (0 split)  
## JobLevel splits as LLLLR, agree=0.776, adj=0.093, (0 split)  
## YearsInCurrentRole < 9.5 to the left, agree=0.776, adj=0.093, (0 split)  
## YearsWithCurrManager < 5.5 to the left, agree=0.776, adj=0.093, (0 split)  
##   
## Node number 35: 47 observations, complexity param=0.00625  
## predicted class=No expected loss=0.1914894 P(node) =0.04795918  
## class counts: 38 9  
## probabilities: 0.809 0.191   
## left son=70 (44 obs) right son=71 (3 obs)  
## Primary splits:  
## EducationField splits as RLRLLL, improve=4.189555, (0 missing)  
## BusinessTravel splits as LRL, improve=3.598646, (0 missing)  
## HourlyRate < 52.5 to the left, improve=1.953191, (0 missing)  
## JobRole splits as RRLLLLLR-, improve=1.633837, (0 missing)  
## JobInvolvement splits as RLLL, improve=1.447131, (0 missing)  
##   
## Node number 36: 75 observations, complexity param=0.00625  
## predicted class=No expected loss=0.04 P(node) =0.07653061  
## class counts: 72 3  
## probabilities: 0.960 0.040   
## left son=72 (54 obs) right son=73 (21 obs)  
## Primary splits:  
## JobInvolvement splits as LRLL, improve=0.6171429, (0 missing)  
## NumCompaniesWorked < 8.5 to the left, improve=0.5377778, (0 missing)  
## YearsAtCompany < 9.5 to the left, improve=0.4028571, (0 missing)  
## MonthlyIncome < 8557 to the left, improve=0.3806897, (0 missing)  
## Age < 30.5 to the right, improve=0.3642155, (0 missing)  
## Surrogate splits:  
## Age < 54.5 to the left, agree=0.747, adj=0.095, (0 split)  
## DailyRate < 1346.5 to the left, agree=0.747, adj=0.095, (0 split)  
## MaritalStatus splits as LLR, agree=0.747, adj=0.095, (0 split)  
## MonthlyRate < 3122 to the right, agree=0.733, adj=0.048, (0 split)  
## StockOptionLevel splits as RLLL, agree=0.733, adj=0.048, (0 split)  
##   
## Node number 37: 14 observations, complexity param=0.009375  
## predicted class=No expected loss=0.3571429 P(node) =0.01428571  
## class counts: 9 5  
## probabilities: 0.643 0.357   
## left son=74 (11 obs) right son=75 (3 obs)  
## Primary splits:  
## NumCompaniesWorked < 4.5 to the left, improve=3.155844, (0 missing)  
## MonthlyIncome < 3969 to the left, improve=2.011905, (0 missing)  
## DailyRate < 1412 to the left, improve=1.928571, (0 missing)  
## PercentSalaryHike < 19 to the left, improve=1.928571, (0 missing)  
## PerformanceRating splits as LR, improve=1.928571, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 10848.5 to the right, agree=0.929, adj=0.667, (0 split)  
## Department splits as RLL, agree=0.857, adj=0.333, (0 split)  
## JobRole splits as -RL-----L, agree=0.857, adj=0.333, (0 split)  
##   
## Node number 38: 17 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.2352941 P(node) =0.01734694  
## class counts: 13 4  
## probabilities: 0.765 0.235   
## left son=76 (13 obs) right son=77 (4 obs)  
## Primary splits:  
## Department splits as LLR, improve=2.771493, (0 missing)  
## JobRole splits as LLL-LLLR-, improve=2.771493, (0 missing)  
## YearsInCurrentRole < 2.5 to the right, improve=2.689076, (0 missing)  
## DistanceFromHome < 15 to the left, improve=1.884314, (0 missing)  
## HourlyRate < 67.5 to the left, improve=1.673203, (0 missing)  
## Surrogate splits:  
## DistanceFromHome < 12 to the left, agree=0.882, adj=0.50, (0 split)  
## TotalWorkingYears < 5.5 to the right, agree=0.882, adj=0.50, (0 split)  
## DailyRate < 127 to the right, agree=0.824, adj=0.25, (0 split)  
## EducationField splits as -LRLLL, agree=0.824, adj=0.25, (0 split)  
## JobInvolvement splits as RLLL, agree=0.824, adj=0.25, (0 split)  
##   
## Node number 39: 5 observations  
## predicted class=Yes expected loss=0 P(node) =0.005102041  
## class counts: 0 5  
## probabilities: 0.000 1.000   
##   
## Node number 40: 99 observations  
## predicted class=No expected loss=0.04040404 P(node) =0.1010204  
## class counts: 95 4  
## probabilities: 0.960 0.040   
##   
## Node number 41: 25 observations, complexity param=0.0125  
## predicted class=No expected loss=0.28 P(node) =0.0255102  
## class counts: 18 7  
## probabilities: 0.720 0.280   
## left son=82 (17 obs) right son=83 (8 obs)  
## Primary splits:  
## EnvironmentSatisfaction splits as RRLL, improve=5.197647, (0 missing)  
## JobInvolvement splits as RLLL, improve=3.534545, (0 missing)  
## YearsAtCompany < 4.5 to the right, improve=2.768312, (0 missing)  
## JobRole splits as L-LRL-RR-, improve=2.613333, (0 missing)  
## MonthlyRate < 22203 to the right, improve=2.253913, (0 missing)  
## Surrogate splits:  
## Age < 31.5 to the right, agree=0.80, adj=0.375, (0 split)  
## JobInvolvement splits as RLLL, agree=0.80, adj=0.375, (0 split)  
## DistanceFromHome < 1.5 to the right, agree=0.76, adj=0.250, (0 split)  
## EducationField splits as -LLRLL, agree=0.76, adj=0.250, (0 split)  
## MonthlyIncome < 11825 to the left, agree=0.72, adj=0.125, (0 split)  
##   
## Node number 42: 18 observations, complexity param=0.0125  
## predicted class=No expected loss=0.2222222 P(node) =0.01836735  
## class counts: 14 4  
## probabilities: 0.778 0.222   
## left son=84 (16 obs) right son=85 (2 obs)  
## Primary splits:  
## JobRole splits as R-LRLLLL-, improve=2.722222, (0 missing)  
## DistanceFromHome < 28.5 to the left, improve=1.422222, (0 missing)  
## PercentSalaryHike < 18.5 to the left, improve=1.422222, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=1.422222, (0 missing)  
## Gender splits as LR, improve=1.131313, (0 missing)  
## Surrogate splits:  
## DistanceFromHome < 28.5 to the left, agree=0.944, adj=0.5, (0 split)  
##   
## Node number 43: 4 observations  
## predicted class=Yes expected loss=0 P(node) =0.004081633  
## class counts: 0 4  
## probabilities: 0.000 1.000   
##   
## Node number 44: 25 observations  
## predicted class=No expected loss=0.12 P(node) =0.0255102  
## class counts: 22 3  
## probabilities: 0.880 0.120   
##   
## Node number 45: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 46: 17 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.4117647 P(node) =0.01734694  
## class counts: 10 7  
## probabilities: 0.588 0.412   
## left son=92 (11 obs) right son=93 (6 obs)  
## Primary splits:  
## WorkLifeBalance splits as LRLL, improve=6.417112, (0 missing)  
## MonthlyIncome < 8044 to the left, improve=4.721008, (0 missing)  
## YearsSinceLastPromotion < 2.5 to the left, improve=3.619910, (0 missing)  
## JobLevel splits as LLRR-, improve=3.457516, (0 missing)  
## Education splits as LLRL-, improve=3.295900, (0 missing)  
## Surrogate splits:  
## MonthlyIncome < 8044 to the left, agree=0.941, adj=0.833, (0 split)  
## JobLevel splits as LLRR-, agree=0.882, adj=0.667, (0 split)  
## Age < 51.5 to the left, agree=0.824, adj=0.500, (0 split)  
## TotalWorkingYears < 21.5 to the left, agree=0.824, adj=0.500, (0 split)  
## Education splits as LLRL-, agree=0.765, adj=0.333, (0 split)  
##   
## Node number 47: 19 observations  
## predicted class=Yes expected loss=0.1578947 P(node) =0.01938776  
## class counts: 3 16  
## probabilities: 0.158 0.842   
##   
## Node number 48: 80 observations, complexity param=0.003125  
## predicted class=No expected loss=0.0375 P(node) =0.08163265  
## class counts: 77 3  
## probabilities: 0.963 0.037   
## left son=96 (78 obs) right son=97 (2 obs)  
## Primary splits:  
## HourlyRate < 32.5 to the right, improve=0.8775641, (0 missing)  
## EducationField splits as LLLLLR, improve=0.6920579, (0 missing)  
## RelationshipSatisfaction splits as RLLL, improve=0.6920579, (0 missing)  
## MonthlyRate < 18752 to the left, improve=0.6321429, (0 missing)  
## WorkLifeBalance splits as LRLR, improve=0.4416667, (0 missing)  
##   
## Node number 49: 14 observations, complexity param=0.009375  
## predicted class=No expected loss=0.2857143 P(node) =0.01428571  
## class counts: 10 4  
## probabilities: 0.714 0.286   
## left son=98 (9 obs) right son=99 (5 obs)  
## Primary splits:  
## JobSatisfaction splits as LLRL, improve=4.114286, (0 missing)  
## EnvironmentSatisfaction splits as R-LL, improve=2.380952, (0 missing)  
## MonthlyRate < 21567 to the left, improve=2.380952, (0 missing)  
## EducationField splits as -LLRLR, improve=1.536508, (0 missing)  
## WorkLifeBalance splits as RRL-, improve=1.536508, (0 missing)  
## Surrogate splits:  
## EnvironmentSatisfaction splits as R-LL, agree=0.786, adj=0.4, (0 split)  
## MonthlyRate < 21567 to the left, agree=0.786, adj=0.4, (0 split)  
## BusinessTravel splits as -RL, agree=0.714, adj=0.2, (0 split)  
## DailyRate < 577 to the right, agree=0.714, adj=0.2, (0 split)  
## Department splits as RLL, agree=0.714, adj=0.2, (0 split)  
##   
## Node number 52: 26 observations, complexity param=0.009375  
## predicted class=No expected loss=0.1923077 P(node) =0.02653061  
## class counts: 21 5  
## probabilities: 0.808 0.192   
## left son=104 (19 obs) right son=105 (7 obs)  
## Primary splits:  
## MonthlyRate < 20229 to the left, improve=2.753615, (0 missing)  
## HourlyRate < 84.5 to the left, improve=2.622378, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=1.750126, (0 missing)  
## JobSatisfaction splits as RLLL, improve=1.476923, (0 missing)  
## RelationshipSatisfaction splits as RLLL, improve=1.476923, (0 missing)  
## Surrogate splits:  
## BusinessTravel splits as LRL, agree=0.808, adj=0.286, (0 split)  
## HourlyRate < 93 to the left, agree=0.808, adj=0.286, (0 split)  
## PercentSalaryHike < 19 to the left, agree=0.808, adj=0.286, (0 split)  
## PerformanceRating splits as LR, agree=0.808, adj=0.286, (0 split)  
## JobInvolvement splits as RLL-, agree=0.769, adj=0.143, (0 split)  
##   
## Node number 53: 10 observations, complexity param=0.0125  
## predicted class=Yes expected loss=0.4 P(node) =0.01020408  
## class counts: 4 6  
## probabilities: 0.400 0.600   
## left son=106 (6 obs) right son=107 (4 obs)  
## Primary splits:  
## RelationshipSatisfaction splits as RLRL, improve=2.133333, (0 missing)  
## EducationField splits as RR--RL, improve=1.800000, (0 missing)  
## NumCompaniesWorked < 0.5 to the left, improve=1.800000, (0 missing)  
## TotalWorkingYears < 1.5 to the right, improve=1.633333, (0 missing)  
## Department splits as RLR, improve=1.371429, (0 missing)  
## Surrogate splits:  
## EnvironmentSatisfaction splits as LRLR, agree=0.8, adj=0.5, (0 split)  
## MonthlyIncome < 2080 to the right, agree=0.8, adj=0.5, (0 split)  
## StockOptionLevel splits as LR-L, agree=0.8, adj=0.5, (0 split)  
## TotalWorkingYears < 1.5 to the right, agree=0.8, adj=0.5, (0 split)  
## WorkLifeBalance splits as RLLR, agree=0.8, adj=0.5, (0 split)  
##   
## Node number 54: 2 observations  
## predicted class=No expected loss=0 P(node) =0.002040816  
## class counts: 2 0  
## probabilities: 1.000 0.000   
##   
## Node number 55: 16 observations  
## predicted class=Yes expected loss=0.125 P(node) =0.01632653  
## class counts: 2 14  
## probabilities: 0.125 0.875   
##   
## Node number 58: 3 observations  
## predicted class=No expected loss=0 P(node) =0.003061224  
## class counts: 3 0  
## probabilities: 1.000 0.000   
##   
## Node number 59: 9 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.2222222 P(node) =0.009183673  
## class counts: 2 7  
## probabilities: 0.222 0.778   
## left son=118 (3 obs) right son=119 (6 obs)  
## Primary splits:  
## EnvironmentSatisfaction splits as RRLR, improve=1.777778, (0 missing)  
## JobRole splits as -RL---R-R, improve=1.777778, (0 missing)  
## JobSatisfaction splits as R-RL, improve=1.111111, (0 missing)  
## MaritalStatus splits as RLR, improve=1.111111, (0 missing)  
## StockOptionLevel splits as RLR-, improve=1.111111, (0 missing)  
## Surrogate splits:  
## DistanceFromHome < 18 to the right, agree=0.889, adj=0.667, (0 split)  
## NumCompaniesWorked < 5.5 to the right, agree=0.889, adj=0.667, (0 split)  
## DailyRate < 887 to the right, agree=0.778, adj=0.333, (0 split)  
## Education splits as -RLR-, agree=0.778, adj=0.333, (0 split)  
## EducationField splits as -RRLRR, agree=0.778, adj=0.333, (0 split)  
##   
## Node number 60: 7 observations  
## predicted class=No expected loss=0.1428571 P(node) =0.007142857  
## class counts: 6 1  
## probabilities: 0.857 0.143   
##   
## Node number 61: 8 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.25 P(node) =0.008163265  
## class counts: 2 6  
## probabilities: 0.250 0.750   
## left son=122 (3 obs) right son=123 (5 obs)  
## Primary splits:  
## Department splits as -RL, improve=1.666667, (0 missing)  
## EducationField splits as -LRRRL, improve=1.666667, (0 missing)  
## EnvironmentSatisfaction splits as RLLR, improve=1.666667, (0 missing)  
## JobRole splits as --R---R-L, improve=1.666667, (0 missing)  
## MaritalStatus splits as RLR, improve=1.666667, (0 missing)  
## Surrogate splits:  
## NumCompaniesWorked < 1.5 to the left, agree=0.875, adj=0.667, (0 split)  
## Age < 35.5 to the left, agree=0.750, adj=0.333, (0 split)  
## BusinessTravel splits as -LR, agree=0.750, adj=0.333, (0 split)  
## Gender splits as LR, agree=0.750, adj=0.333, (0 split)  
## MaritalStatus splits as RLR, agree=0.750, adj=0.333, (0 split)  
##   
## Node number 62: 10 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.3 P(node) =0.01020408  
## class counts: 3 7  
## probabilities: 0.300 0.700   
## left son=124 (4 obs) right son=125 (6 obs)  
## Primary splits:  
## DistanceFromHome < 7.5 to the left, improve=2.7, (0 missing)  
## EducationField splits as RRRLRR, improve=2.7, (0 missing)  
## MonthlyIncome < 2136.5 to the right, improve=1.2, (0 missing)  
## PercentSalaryHike < 17.5 to the left, improve=1.2, (0 missing)  
## YearsAtCompany < 4.5 to the left, improve=1.2, (0 missing)  
## Surrogate splits:  
## Age < 36 to the right, agree=0.8, adj=0.50, (0 split)  
## NumCompaniesWorked < 2.5 to the right, agree=0.8, adj=0.50, (0 split)  
## YearsAtCompany < 3.5 to the left, agree=0.8, adj=0.50, (0 split)  
## DailyRate < 409.5 to the right, agree=0.7, adj=0.25, (0 split)  
## Department splits as LRR, agree=0.7, adj=0.25, (0 split)  
##   
## Node number 63: 20 observations  
## predicted class=Yes expected loss=0 P(node) =0.02040816  
## class counts: 0 20  
## probabilities: 0.000 1.000   
##   
## Node number 68: 131 observations, complexity param=0.003125  
## predicted class=No expected loss=0.02290076 P(node) =0.1336735  
## class counts: 128 3  
## probabilities: 0.977 0.023   
## left son=136 (103 obs) right son=137 (28 obs)  
## Primary splits:  
## Age < 29.5 to the right, improve=0.5054526, (0 missing)  
## DailyRate < 125.5 to the right, improve=0.4255875, (0 missing)  
## DistanceFromHome < 26.5 to the left, improve=0.4255875, (0 missing)  
## EducationField splits as LLRLLR, improve=0.2289400, (0 missing)  
## HourlyRate < 98.5 to the left, improve=0.2128258, (0 missing)  
## Surrogate splits:  
## TotalWorkingYears < 3.5 to the right, agree=0.824, adj=0.179, (0 split)  
## TrainingTimesLastYear < 0.5 to the right, agree=0.802, adj=0.071, (0 split)  
##   
## Node number 69: 43 observations, complexity param=0.00625  
## predicted class=No expected loss=0.1627907 P(node) =0.04387755  
## class counts: 36 7  
## probabilities: 0.837 0.163   
## left son=138 (38 obs) right son=139 (5 obs)  
## Primary splits:  
## HourlyRate < 37.5 to the right, improve=2.163035, (0 missing)  
## JobRole splits as LLLLRLRRL, improve=1.593737, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=1.466385, (0 missing)  
## TotalWorkingYears < 14.5 to the right, improve=1.240411, (0 missing)  
## DailyRate < 1357.5 to the left, improve=1.002982, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 25056.5 to the left, agree=0.930, adj=0.4, (0 split)  
## WorkLifeBalance splits as RLLL, agree=0.907, adj=0.2, (0 split)  
##   
## Node number 70: 44 observations, complexity param=0.00625  
## predicted class=No expected loss=0.1363636 P(node) =0.04489796  
## class counts: 38 6  
## probabilities: 0.864 0.136   
## left son=140 (35 obs) right son=141 (9 obs)  
## Primary splits:  
## BusinessTravel splits as LRL, improve=2.1477630, (0 missing)  
## Age < 33.5 to the right, improve=1.2803030, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=1.1136360, (0 missing)  
## HourlyRate < 52.5 to the left, improve=1.0303030, (0 missing)  
## MonthlyRate < 22756 to the left, improve=0.8779221, (0 missing)  
##   
## Node number 71: 3 observations  
## predicted class=Yes expected loss=0 P(node) =0.003061224  
## class counts: 0 3  
## probabilities: 0.000 1.000   
##   
## Node number 72: 54 observations  
## predicted class=No expected loss=0 P(node) =0.05510204  
## class counts: 54 0  
## probabilities: 1.000 0.000   
##   
## Node number 73: 21 observations, complexity param=0.00625  
## predicted class=No expected loss=0.1428571 P(node) =0.02142857  
## class counts: 18 3  
## probabilities: 0.857 0.143   
## left son=146 (19 obs) right son=147 (2 obs)  
## Primary splits:  
## EnvironmentSatisfaction splits as LRLL, improve=3.2481200, (0 missing)  
## YearsSinceLastPromotion < 5.5 to the left, improve=1.2605040, (0 missing)  
## MonthlyIncome < 8154.5 to the left, improve=1.1428570, (0 missing)  
## YearsAtCompany < 9.5 to the left, improve=0.9428571, (0 missing)  
## Age < 35.5 to the right, improve=0.6428571, (0 missing)  
##   
## Node number 74: 11 observations, complexity param=0.009375  
## predicted class=No expected loss=0.1818182 P(node) =0.01122449  
## class counts: 9 2  
## probabilities: 0.818 0.182   
## left son=148 (9 obs) right son=149 (2 obs)  
## Primary splits:  
## DailyRate < 1412 to the left, improve=3.2727270, (0 missing)  
## RelationshipSatisfaction splits as RRLL, improve=0.8727273, (0 missing)  
## YearsAtCompany < 5 to the right, improve=0.8727273, (0 missing)  
## Education splits as RRLL-, improve=0.6060606, (0 missing)  
## MonthlyRate < 15646.5 to the left, improve=0.6060606, (0 missing)  
##   
## Node number 75: 3 observations  
## predicted class=Yes expected loss=0 P(node) =0.003061224  
## class counts: 0 3  
## probabilities: 0.000 1.000   
##   
## Node number 76: 13 observations  
## predicted class=No expected loss=0.07692308 P(node) =0.01326531  
## class counts: 12 1  
## probabilities: 0.923 0.077   
##   
## Node number 77: 4 observations  
## predicted class=Yes expected loss=0.25 P(node) =0.004081633  
## class counts: 1 3  
## probabilities: 0.250 0.750   
##   
## Node number 82: 17 observations  
## predicted class=No expected loss=0.05882353 P(node) =0.01734694  
## class counts: 16 1  
## probabilities: 0.941 0.059   
##   
## Node number 83: 8 observations, complexity param=0.0125  
## predicted class=Yes expected loss=0.25 P(node) =0.008163265  
## class counts: 2 6  
## probabilities: 0.250 0.750   
## left son=166 (2 obs) right son=167 (6 obs)  
## Primary splits:  
## JobRole splits as L--RL-RR-, improve=3.000000, (0 missing)  
## BusinessTravel splits as LLR, improve=1.666667, (0 missing)  
## MonthlyRate < 23087 to the left, improve=1.666667, (0 missing)  
## DailyRate < 712 to the left, improve=1.000000, (0 missing)  
## HourlyRate < 59.5 to the right, improve=1.000000, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 23087 to the left, agree=0.875, adj=0.5, (0 split)  
##   
## Node number 84: 16 observations, complexity param=0.00625  
## predicted class=No expected loss=0.125 P(node) =0.01632653  
## class counts: 14 2  
## probabilities: 0.875 0.125   
## left son=168 (10 obs) right son=169 (6 obs)  
## Primary splits:  
## MonthlyRate < 17961 to the left, improve=0.8333333, (0 missing)  
## RelationshipSatisfaction splits as LLLR, improve=0.8333333, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=0.8333333, (0 missing)  
## Age < 32 to the right, improve=0.6428571, (0 missing)  
## HourlyRate < 93.5 to the left, improve=0.6428571, (0 missing)  
## Surrogate splits:  
## Education splits as LLLRR, agree=0.812, adj=0.500, (0 split)  
## YearsAtCompany < 1.5 to the right, agree=0.812, adj=0.500, (0 split)  
## YearsWithCurrManager < 1 to the right, agree=0.812, adj=0.500, (0 split)  
## Age < 34.5 to the right, agree=0.750, adj=0.333, (0 split)  
## BusinessTravel splits as LRL, agree=0.750, adj=0.333, (0 split)  
##   
## Node number 85: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 92: 11 observations  
## predicted class=No expected loss=0.09090909 P(node) =0.01122449  
## class counts: 10 1  
## probabilities: 0.909 0.091   
##   
## Node number 93: 6 observations  
## predicted class=Yes expected loss=0 P(node) =0.006122449  
## class counts: 0 6  
## probabilities: 0.000 1.000   
##   
## Node number 96: 78 observations, complexity param=0.003125  
## predicted class=No expected loss=0.02564103 P(node) =0.07959184  
## class counts: 76 2  
## probabilities: 0.974 0.026   
## left son=192 (62 obs) right son=193 (16 obs)  
## Primary splits:  
## MonthlyRate < 19747 to the left, improve=0.3974359, (0 missing)  
## YearsWithCurrManager < 4.5 to the left, improve=0.3184885, (0 missing)  
## MonthlyIncome < 2060 to the right, improve=0.2585470, (0 missing)  
## WorkLifeBalance splits as LRLR, improve=0.2174359, (0 missing)  
## YearsAtCompany < 5.5 to the left, improve=0.2174359, (0 missing)  
## Surrogate splits:  
## PercentSalaryHike < 23.5 to the left, agree=0.808, adj=0.063, (0 split)  
##   
## Node number 97: 2 observations  
## predicted class=No expected loss=0.5 P(node) =0.002040816  
## class counts: 1 1  
## probabilities: 0.500 0.500   
##   
## Node number 98: 9 observations  
## predicted class=No expected loss=0 P(node) =0.009183673  
## class counts: 9 0  
## probabilities: 1.000 0.000   
##   
## Node number 99: 5 observations  
## predicted class=Yes expected loss=0.2 P(node) =0.005102041  
## class counts: 1 4  
## probabilities: 0.200 0.800   
##   
## Node number 104: 19 observations  
## predicted class=No expected loss=0.05263158 P(node) =0.01938776  
## class counts: 18 1  
## probabilities: 0.947 0.053   
##   
## Node number 105: 7 observations, complexity param=0.009375  
## predicted class=Yes expected loss=0.4285714 P(node) =0.007142857  
## class counts: 3 4  
## probabilities: 0.429 0.571   
## left son=210 (4 obs) right son=211 (3 obs)  
## Primary splits:  
## Education splits as LRLR-, improve=1.928571, (0 missing)  
## JobSatisfaction splits as RLLL, improve=1.928571, (0 missing)  
## MaritalStatus splits as LLR, improve=1.928571, (0 missing)  
## MonthlyRate < 22242 to the right, improve=1.928571, (0 missing)  
## RelationshipSatisfaction splits as RLL-, improve=1.928571, (0 missing)  
## Surrogate splits:  
## Age < 31 to the left, agree=0.857, adj=0.667, (0 split)  
## NumCompaniesWorked < 2 to the left, agree=0.857, adj=0.667, (0 split)  
## PercentSalaryHike < 18.5 to the right, agree=0.857, adj=0.667, (0 split)  
## TotalWorkingYears < 3 to the left, agree=0.857, adj=0.667, (0 split)  
## BusinessTravel splits as RLL, agree=0.714, adj=0.333, (0 split)  
##   
## Node number 106: 6 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3333333 P(node) =0.006122449  
## class counts: 4 2  
## probabilities: 0.667 0.333   
## left son=212 (4 obs) right son=213 (2 obs)  
## Primary splits:  
## MonthlyIncome < 2586 to the left, improve=2.666667, (0 missing)  
## EducationField splits as -R--LL, improve=1.333333, (0 missing)  
## HourlyRate < 89.5 to the left, improve=1.333333, (0 missing)  
## JobInvolvement splits as LLRR, improve=1.333333, (0 missing)  
## JobSatisfaction splits as LRLR, improve=1.333333, (0 missing)  
## Surrogate splits:  
## HourlyRate < 89.5 to the left, agree=0.833, adj=0.5, (0 split)  
## MonthlyRate < 10148.5 to the left, agree=0.833, adj=0.5, (0 split)  
##   
## Node number 107: 4 observations  
## predicted class=Yes expected loss=0 P(node) =0.004081633  
## class counts: 0 4  
## probabilities: 0.000 1.000   
##   
## Node number 118: 3 observations  
## predicted class=No expected loss=0.3333333 P(node) =0.003061224  
## class counts: 2 1  
## probabilities: 0.667 0.333   
##   
## Node number 119: 6 observations  
## predicted class=Yes expected loss=0 P(node) =0.006122449  
## class counts: 0 6  
## probabilities: 0.000 1.000   
##   
## Node number 122: 3 observations  
## predicted class=No expected loss=0.3333333 P(node) =0.003061224  
## class counts: 2 1  
## probabilities: 0.667 0.333   
##   
## Node number 123: 5 observations  
## predicted class=Yes expected loss=0 P(node) =0.005102041  
## class counts: 0 5  
## probabilities: 0.000 1.000   
##   
## Node number 124: 4 observations  
## predicted class=No expected loss=0.25 P(node) =0.004081633  
## class counts: 3 1  
## probabilities: 0.750 0.250   
##   
## Node number 125: 6 observations  
## predicted class=Yes expected loss=0 P(node) =0.006122449  
## class counts: 0 6  
## probabilities: 0.000 1.000   
##   
## Node number 136: 103 observations  
## predicted class=No expected loss=0 P(node) =0.105102  
## class counts: 103 0  
## probabilities: 1.000 0.000   
##   
## Node number 137: 28 observations, complexity param=0.003125  
## predicted class=No expected loss=0.1071429 P(node) =0.02857143  
## class counts: 25 3  
## probabilities: 0.893 0.107   
## left son=274 (25 obs) right son=275 (3 obs)  
## Primary splits:  
## EducationField splits as -LRLLR, improve=2.1038100, (0 missing)  
## YearsAtCompany < 2.5 to the right, improve=1.4404760, (0 missing)  
## NumCompaniesWorked < 4 to the left, improve=1.0440990, (0 missing)  
## PercentSalaryHike < 15.5 to the left, improve=0.7417582, (0 missing)  
## DailyRate < 140.5 to the right, improve=0.6648352, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 23928.5 to the left, agree=0.964, adj=0.667, (0 split)  
## DailyRate < 1292.5 to the left, agree=0.929, adj=0.333, (0 split)  
##   
## Node number 138: 38 observations, complexity param=0.004166667  
## predicted class=No expected loss=0.1052632 P(node) =0.03877551  
## class counts: 34 4  
## probabilities: 0.895 0.105   
## left son=276 (34 obs) right son=277 (4 obs)  
## Primary splits:  
## DailyRate < 1357.5 to the left, improve=1.3931890, (0 missing)  
## JobRole splits as LLLLRRRRL, improve=0.6817043, (0 missing)  
## Education splits as LLLLR, improve=0.6578947, (0 missing)  
## YearsInCurrentRole < 13 to the left, improve=0.6578947, (0 missing)  
## YearsSinceLastPromotion < 13.5 to the left, improve=0.6578947, (0 missing)  
##   
## Node number 139: 5 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.4 P(node) =0.005102041  
## class counts: 2 3  
## probabilities: 0.400 0.600   
## left son=278 (2 obs) right son=279 (3 obs)  
## Primary splits:  
## RelationshipSatisfaction splits as R-RL, improve=2.400000, (0 missing)  
## Age < 34 to the right, improve=1.066667, (0 missing)  
## EnvironmentSatisfaction splits as -LLR, improve=1.066667, (0 missing)  
## HourlyRate < 32.5 to the left, improve=1.066667, (0 missing)  
## TotalWorkingYears < 11 to the left, improve=1.066667, (0 missing)  
## Surrogate splits:  
## Age < 34 to the right, agree=0.8, adj=0.5, (0 split)  
## HourlyRate < 32.5 to the left, agree=0.8, adj=0.5, (0 split)  
## TotalWorkingYears < 11 to the left, agree=0.8, adj=0.5, (0 split)  
## YearsInCurrentRole < 5 to the left, agree=0.8, adj=0.5, (0 split)  
##   
## Node number 140: 35 observations, complexity param=0.00625  
## predicted class=No expected loss=0.05714286 P(node) =0.03571429  
## class counts: 33 2  
## probabilities: 0.943 0.057   
## left son=280 (29 obs) right son=281 (6 obs)  
## Primary splits:  
## Age < 28.5 to the right, improve=1.1047620, (0 missing)  
## TotalWorkingYears < 5.5 to the right, improve=1.1047620, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=1.1047620, (0 missing)  
## YearsAtCompany < 2.5 to the right, improve=0.9142857, (0 missing)  
## NumCompaniesWorked < 7.5 to the left, improve=0.8320346, (0 missing)  
## Surrogate splits:  
## Education splits as RLLLL, agree=0.886, adj=0.333, (0 split)  
## TotalWorkingYears < 7 to the right, agree=0.886, adj=0.333, (0 split)  
## JobRole splits as LLLLLLLR-, agree=0.857, adj=0.167, (0 split)  
## MonthlyIncome < 3557 to the right, agree=0.857, adj=0.167, (0 split)  
## NumCompaniesWorked < 0.5 to the right, agree=0.857, adj=0.167, (0 split)  
##   
## Node number 141: 9 observations, complexity param=0.00625  
## predicted class=No expected loss=0.4444444 P(node) =0.009183673  
## class counts: 5 4  
## probabilities: 0.556 0.444   
## left son=282 (6 obs) right son=283 (3 obs)  
## Primary splits:  
## JobLevel splits as LLRR-, improve=2.777778, (0 missing)  
## MonthlyIncome < 8790 to the left, improve=2.777778, (0 missing)  
## Education splits as LLRR-, improve=1.777778, (0 missing)  
## JobRole splits as R-L-R-RL-, improve=1.777778, (0 missing)  
## TotalWorkingYears < 11 to the left, improve=1.777778, (0 missing)  
## Surrogate splits:  
## MonthlyIncome < 8790 to the left, agree=1.000, adj=1.000, (0 split)  
## DailyRate < 576 to the right, agree=0.889, adj=0.667, (0 split)  
## Education splits as LLRL-, agree=0.778, adj=0.333, (0 split)  
## EducationField splits as -R-L--, agree=0.778, adj=0.333, (0 split)  
## Gender splits as LR, agree=0.778, adj=0.333, (0 split)  
##   
## Node number 146: 19 observations  
## predicted class=No expected loss=0.05263158 P(node) =0.01938776  
## class counts: 18 1  
## probabilities: 0.947 0.053   
##   
## Node number 147: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 148: 9 observations  
## predicted class=No expected loss=0 P(node) =0.009183673  
## class counts: 9 0  
## probabilities: 1.000 0.000   
##   
## Node number 149: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 166: 2 observations  
## predicted class=No expected loss=0 P(node) =0.002040816  
## class counts: 2 0  
## probabilities: 1.000 0.000   
##   
## Node number 167: 6 observations  
## predicted class=Yes expected loss=0 P(node) =0.006122449  
## class counts: 0 6  
## probabilities: 0.000 1.000   
##   
## Node number 168: 10 observations  
## predicted class=No expected loss=0 P(node) =0.01020408  
## class counts: 10 0  
## probabilities: 1.000 0.000   
##   
## Node number 169: 6 observations, complexity param=0.00625  
## predicted class=No expected loss=0.3333333 P(node) =0.006122449  
## class counts: 4 2  
## probabilities: 0.667 0.333   
## left son=338 (4 obs) right son=339 (2 obs)  
## Primary splits:  
## EnvironmentSatisfaction splits as LLLR, improve=2.666667, (0 missing)  
## MonthlyRate < 21130 to the right, improve=2.666667, (0 missing)  
## DistanceFromHome < 25 to the right, improve=1.333333, (0 missing)  
## JobRole splits as --R-RLL--, improve=1.333333, (0 missing)  
## JobSatisfaction splits as L-RL, improve=1.333333, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 21130 to the right, agree=1.000, adj=1.0, (0 split)  
## DistanceFromHome < 25 to the right, agree=0.833, adj=0.5, (0 split)  
##   
## Node number 192: 62 observations  
## predicted class=No expected loss=0 P(node) =0.06326531  
## class counts: 62 0  
## probabilities: 1.000 0.000   
##   
## Node number 193: 16 observations, complexity param=0.003125  
## predicted class=No expected loss=0.125 P(node) =0.01632653  
## class counts: 14 2  
## probabilities: 0.875 0.125   
## left son=386 (11 obs) right son=387 (5 obs)  
## Primary splits:  
## DistanceFromHome < 4 to the right, improve=1.1000000, (0 missing)  
## MonthlyRate < 21620 to the right, improve=1.1000000, (0 missing)  
## RelationshipSatisfaction splits as RLLR, improve=1.1000000, (0 missing)  
## YearsAtCompany < 5.5 to the left, improve=0.8333333, (0 missing)  
## YearsWithCurrManager < 4.5 to the left, improve=0.8333333, (0 missing)  
## Surrogate splits:  
## DailyRate < 272 to the right, agree=0.812, adj=0.4, (0 split)  
## EducationField splits as LL-RLR, agree=0.812, adj=0.4, (0 split)  
## EnvironmentSatisfaction splits as LLLR, agree=0.812, adj=0.4, (0 split)  
## JobSatisfaction splits as RLLL, agree=0.812, adj=0.4, (0 split)  
## MonthlyRate < 20505 to the right, agree=0.812, adj=0.4, (0 split)  
##   
## Node number 210: 4 observations  
## predicted class=No expected loss=0.25 P(node) =0.004081633  
## class counts: 3 1  
## probabilities: 0.750 0.250   
##   
## Node number 211: 3 observations  
## predicted class=Yes expected loss=0 P(node) =0.003061224  
## class counts: 0 3  
## probabilities: 0.000 1.000   
##   
## Node number 212: 4 observations  
## predicted class=No expected loss=0 P(node) =0.004081633  
## class counts: 4 0  
## probabilities: 1.000 0.000   
##   
## Node number 213: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 274: 25 observations  
## predicted class=No expected loss=0.04 P(node) =0.0255102  
## class counts: 24 1  
## probabilities: 0.960 0.040   
##   
## Node number 275: 3 observations  
## predicted class=Yes expected loss=0.3333333 P(node) =0.003061224  
## class counts: 1 2  
## probabilities: 0.333 0.667   
##   
## Node number 276: 34 observations, complexity param=0.004166667  
## predicted class=No expected loss=0.05882353 P(node) =0.03469388  
## class counts: 32 2  
## probabilities: 0.941 0.059   
## left son=552 (25 obs) right son=553 (9 obs)  
## Primary splits:  
## Age < 45.5 to the left, improve=0.6535948, (0 missing)  
## JobRole splits as LLLLLLRRL, improve=0.5647059, (0 missing)  
## EducationField splits as -LLRLL, improve=0.4313725, (0 missing)  
## TotalWorkingYears < 11.5 to the right, improve=0.4313725, (0 missing)  
## YearsAtCompany < 9.5 to the right, improve=0.4313725, (0 missing)  
## Surrogate splits:  
## TotalWorkingYears < 25 to the left, agree=0.882, adj=0.556, (0 split)  
## MonthlyIncome < 19645.5 to the left, agree=0.794, adj=0.222, (0 split)  
## MonthlyRate < 4595 to the right, agree=0.794, adj=0.222, (0 split)  
## NumCompaniesWorked < 3.5 to the left, agree=0.794, adj=0.222, (0 split)  
## BusinessTravel splits as LRL, agree=0.765, adj=0.111, (0 split)  
##   
## Node number 277: 4 observations  
## predicted class=No expected loss=0.5 P(node) =0.004081633  
## class counts: 2 2  
## probabilities: 0.500 0.500   
##   
## Node number 278: 2 observations  
## predicted class=No expected loss=0 P(node) =0.002040816  
## class counts: 2 0  
## probabilities: 1.000 0.000   
##   
## Node number 279: 3 observations  
## predicted class=Yes expected loss=0 P(node) =0.003061224  
## class counts: 0 3  
## probabilities: 0.000 1.000   
##   
## Node number 280: 29 observations  
## predicted class=No expected loss=0 P(node) =0.02959184  
## class counts: 29 0  
## probabilities: 1.000 0.000   
##   
## Node number 281: 6 observations, complexity param=0.00625  
## predicted class=No expected loss=0.3333333 P(node) =0.006122449  
## class counts: 4 2  
## probabilities: 0.667 0.333   
## left son=562 (4 obs) right son=563 (2 obs)  
## Primary splits:  
## JobInvolvement splits as -LRL, improve=2.666667, (0 missing)  
## RelationshipSatisfaction splits as RLRL, improve=2.666667, (0 missing)  
## YearsAtCompany < 2.5 to the right, improve=2.666667, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=2.666667, (0 missing)  
## Education splits as LR-R-, improve=1.333333, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 2.5 to the right, agree=1.000, adj=1.0, (0 split)  
## YearsWithCurrManager < 0.5 to the right, agree=1.000, adj=1.0, (0 split)  
## MonthlyRate < 13706 to the left, agree=0.833, adj=0.5, (0 split)  
## NumCompaniesWorked < 1 to the left, agree=0.833, adj=0.5, (0 split)  
## TotalWorkingYears < 5.5 to the right, agree=0.833, adj=0.5, (0 split)  
##   
## Node number 282: 6 observations  
## predicted class=No expected loss=0.1666667 P(node) =0.006122449  
## class counts: 5 1  
## probabilities: 0.833 0.167   
##   
## Node number 283: 3 observations  
## predicted class=Yes expected loss=0 P(node) =0.003061224  
## class counts: 0 3  
## probabilities: 0.000 1.000   
##   
## Node number 338: 4 observations  
## predicted class=No expected loss=0 P(node) =0.004081633  
## class counts: 4 0  
## probabilities: 1.000 0.000   
##   
## Node number 339: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 386: 11 observations  
## predicted class=No expected loss=0 P(node) =0.01122449  
## class counts: 11 0  
## probabilities: 1.000 0.000   
##   
## Node number 387: 5 observations, complexity param=0.003125  
## predicted class=No expected loss=0.4 P(node) =0.005102041  
## class counts: 3 2  
## probabilities: 0.600 0.400   
## left son=774 (3 obs) right son=775 (2 obs)  
## Primary splits:  
## Age < 34 to the right, improve=2.4, (0 missing)  
## EnvironmentSatisfaction splits as RLRL, improve=2.4, (0 missing)  
## RelationshipSatisfaction splits as RL-R, improve=2.4, (0 missing)  
## WorkLifeBalance splits as -RLR, improve=2.4, (0 missing)  
## YearsAtCompany < 5.5 to the left, improve=2.4, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 5.5 to the left, agree=1.0, adj=1.0, (0 split)  
## YearsWithCurrManager < 4 to the left, agree=1.0, adj=1.0, (0 split)  
## DailyRate < 333 to the left, agree=0.8, adj=0.5, (0 split)  
## MonthlyRate < 21620 to the right, agree=0.8, adj=0.5, (0 split)  
## TotalWorkingYears < 5.5 to the left, agree=0.8, adj=0.5, (0 split)  
##   
## Node number 552: 25 observations  
## predicted class=No expected loss=0 P(node) =0.0255102  
## class counts: 25 0  
## probabilities: 1.000 0.000   
##   
## Node number 553: 9 observations, complexity param=0.004166667  
## predicted class=No expected loss=0.2222222 P(node) =0.009183673  
## class counts: 7 2  
## probabilities: 0.778 0.222   
## left son=1106 (7 obs) right son=1107 (2 obs)  
## Primary splits:  
## JobRole splits as L-LL-LRR-, improve=3.111111, (0 missing)  
## DailyRate < 845.5 to the right, improve=1.777778, (0 missing)  
## EnvironmentSatisfaction splits as -LRL, improve=1.777778, (0 missing)  
## MonthlyIncome < 10032 to the right, improve=1.777778, (0 missing)  
## NumCompaniesWorked < 2.5 to the right, improve=1.777778, (0 missing)  
## Surrogate splits:  
## DailyRate < 845.5 to the right, agree=0.889, adj=0.5, (0 split)  
## MonthlyIncome < 10032 to the right, agree=0.889, adj=0.5, (0 split)  
## NumCompaniesWorked < 2.5 to the right, agree=0.889, adj=0.5, (0 split)  
## TotalWorkingYears < 14.5 to the right, agree=0.889, adj=0.5, (0 split)  
##   
## Node number 562: 4 observations  
## predicted class=No expected loss=0 P(node) =0.004081633  
## class counts: 4 0  
## probabilities: 1.000 0.000   
##   
## Node number 563: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 774: 3 observations  
## predicted class=No expected loss=0 P(node) =0.003061224  
## class counts: 3 0  
## probabilities: 1.000 0.000   
##   
## Node number 775: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 1106: 7 observations  
## predicted class=No expected loss=0 P(node) =0.007142857  
## class counts: 7 0  
## probabilities: 1.000 0.000   
##   
## Node number 1107: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000

## Warning: labs do not fit even at cex 0.15, there may be some overplotting

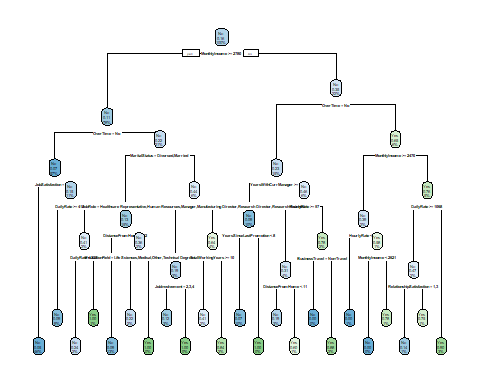


## No Yes   
## 428 62   
## actualAttrition  
## predictedAttrition No Yes  
## No 372 56  
## Yes 41 21

Accuracy 393/490

# Increase minSplit and maxDepth  
advancedTree <- printDecision(seedNum1, HR\_tree)

## Call:  
## rpart(formula = Attrition ~ ., data = train, method = "class",   
## control = rpart.control(cp = 0, minsplit = 5, maxdepth = depth))  
## n= 980   
##   
## CP nsplit rel error xerror xstd  
## 1 0.05937500 0 1.00000 1.00000 0.07231592  
## 2 0.03125000 2 0.88125 0.90000 0.06927099  
## 3 0.02500000 4 0.81875 0.93750 0.07044524  
## 4 0.02083333 5 0.79375 0.93125 0.07025231  
## 5 0.01875000 8 0.73125 0.95000 0.07082783  
## 6 0.01562500 9 0.71250 0.95000 0.07082783  
## 7 0.01250000 13 0.65000 0.91875 0.06986315  
## 8 0.01041667 19 0.57500 0.95625 0.07101751  
## 9 0.00000000 22 0.54375 0.99375 0.07213352  
##   
## Variable importance  
## MonthlyIncome OverTime DailyRate   
## 14 11 7   
## TotalWorkingYears JobRole HourlyRate   
## 7 5 5   
## YearsAtCompany YearsWithCurrManager YearsInCurrentRole   
## 4 4 4   
## EducationField MaritalStatus DistanceFromHome   
## 4 4 4   
## Department YearsSinceLastPromotion JobLevel   
## 3 3 3   
## StockOptionLevel MonthlyRate Age   
## 3 2 2   
## RelationshipSatisfaction JobInvolvement BusinessTravel   
## 2 2 2   
## WorkLifeBalance JobSatisfaction EnvironmentSatisfaction   
## 1 1 1   
## Gender Education   
## 1 1   
##   
## Node number 1: 980 observations, complexity param=0.059375  
## predicted class=No expected loss=0.1632653 P(node) =1  
## class counts: 820 160  
## probabilities: 0.837 0.163   
## left son=2 (767 obs) right son=3 (213 obs)  
## Primary splits:  
## MonthlyIncome < 2780 to the right, improve=19.41164, (0 missing)  
## OverTime splits as LR, improve=19.34035, (0 missing)  
## TotalWorkingYears < 1.5 to the right, improve=14.55748, (0 missing)  
## JobLevel splits as RLLLL, improve=14.47392, (0 missing)  
## JobRole splits as LRRLLLRRR, improve=12.10966, (0 missing)  
## Surrogate splits:  
## TotalWorkingYears < 3.5 to the right, agree=0.841, adj=0.268, (0 split)  
## JobLevel splits as RLLLL, agree=0.834, adj=0.235, (0 split)  
## Age < 23.5 to the right, agree=0.809, adj=0.122, (0 split)  
## JobRole splits as LLLLLLLLR, agree=0.801, adj=0.085, (0 split)  
## YearsAtCompany < 0.5 to the right, agree=0.785, adj=0.009, (0 split)  
##   
## Node number 2: 767 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.1108214 P(node) =0.7826531  
## class counts: 682 85  
## probabilities: 0.889 0.111   
## left son=4 (558 obs) right son=5 (209 obs)  
## Primary splits:  
## OverTime splits as LR, improve=7.474748, (0 missing)  
## StockOptionLevel splits as RLLL, improve=6.348036, (0 missing)  
## MaritalStatus splits as LLR, improve=4.600851, (0 missing)  
## JobRole splits as LRLLLLLRR, improve=4.578610, (0 missing)  
## Department splits as LLR, improve=3.972311, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 26.5 to the left, agree=0.729, adj=0.005, (0 split)  
##   
## Node number 3: 213 observations, complexity param=0.059375  
## predicted class=No expected loss=0.3521127 P(node) =0.2173469  
## class counts: 138 75  
## probabilities: 0.648 0.352   
## left son=6 (150 obs) right son=7 (63 obs)  
## Primary splits:  
## OverTime splits as LR, improve=15.961510, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve= 8.052241, (0 missing)  
## MonthlyRate < 25073 to the left, improve= 4.817714, (0 missing)  
## Age < 21.5 to the right, improve= 4.695013, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve= 4.511393, (0 missing)  
## Surrogate splits:  
## PercentSalaryHike < 11.5 to the right, agree=0.718, adj=0.048, (0 split)  
## DailyRate < 107.5 to the right, agree=0.714, adj=0.032, (0 split)  
## YearsSinceLastPromotion < 6.5 to the left, agree=0.714, adj=0.032, (0 split)  
## Education splits as LLLLR, agree=0.709, adj=0.016, (0 split)  
## MonthlyRate < 3046 to the right, agree=0.709, adj=0.016, (0 split)  
##   
## Node number 4: 558 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.06810036 P(node) =0.5693878  
## class counts: 520 38  
## probabilities: 0.932 0.068   
## left son=8 (447 obs) right son=9 (111 obs)  
## Primary splits:  
## JobSatisfaction splits as RLLL, improve=2.004734, (0 missing)  
## StockOptionLevel splits as RLLR, improve=1.702476, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=1.301085, (0 missing)  
## Age < 33.5 to the right, improve=1.242657, (0 missing)  
## JobRole splits as LRRLLLLRR, improve=1.112509, (0 missing)  
## Surrogate splits:  
## Age < 59.5 to the left, agree=0.805, adj=0.018, (0 split)  
## PercentSalaryHike < 24.5 to the left, agree=0.803, adj=0.009, (0 split)  
## YearsWithCurrManager < 15.5 to the left, agree=0.803, adj=0.009, (0 split)  
##   
## Node number 5: 209 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.2248804 P(node) =0.2132653  
## class counts: 162 47  
## probabilities: 0.775 0.225   
## left son=10 (146 obs) right son=11 (63 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=8.695338, (0 missing)  
## StockOptionLevel splits as RLLL, improve=7.655439, (0 missing)  
## JobRole splits as LLRLLLLRR, improve=5.659909, (0 missing)  
## Department splits as LLR, improve=4.921394, (0 missing)  
## DistanceFromHome < 11.5 to the left, improve=3.682416, (0 missing)  
## Surrogate splits:  
## StockOptionLevel splits as RLLL, agree=0.876, adj=0.587, (0 split)  
## HourlyRate < 98.5 to the left, agree=0.713, adj=0.048, (0 split)  
## MonthlyRate < 2582 to the right, agree=0.713, adj=0.048, (0 split)  
## Age < 24.5 to the right, agree=0.708, adj=0.032, (0 split)  
## JobRole splits as LLLLLLLLR, agree=0.708, adj=0.032, (0 split)  
##   
## Node number 6: 150 observations, complexity param=0.03125  
## predicted class=No expected loss=0.2266667 P(node) =0.1530612  
## class counts: 116 34  
## probabilities: 0.773 0.227   
## left son=12 (96 obs) right son=13 (54 obs)  
## Primary splits:  
## YearsWithCurrManager < 0.5 to the right, improve=9.422315, (0 missing)  
## YearsAtCompany < 1.5 to the right, improve=6.140827, (0 missing)  
## TotalWorkingYears < 2.5 to the right, improve=5.819890, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=4.997185, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=4.650030, (0 missing)  
## Surrogate splits:  
## YearsAtCompany < 1.5 to the right, agree=0.947, adj=0.852, (0 split)  
## YearsInCurrentRole < 0.5 to the right, agree=0.893, adj=0.704, (0 split)  
## TotalWorkingYears < 1.5 to the right, agree=0.867, adj=0.630, (0 split)  
## MonthlyIncome < 1976 to the right, agree=0.760, adj=0.333, (0 split)  
## YearsSinceLastPromotion < 0.5 to the right, agree=0.720, adj=0.222, (0 split)  
##   
## Node number 7: 63 observations, complexity param=0.025  
## predicted class=Yes expected loss=0.3492063 P(node) =0.06428571  
## class counts: 22 41  
## probabilities: 0.349 0.651   
## left son=14 (18 obs) right son=15 (45 obs)  
## Primary splits:  
## MonthlyIncome < 2469.5 to the right, improve=3.457143, (0 missing)  
## DailyRate < 1129 to the right, improve=3.262580, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=3.250305, (0 missing)  
## NumCompaniesWorked < 0.5 to the left, improve=3.108605, (0 missing)  
## DistanceFromHome < 16.5 to the left, improve=2.777778, (0 missing)  
## Surrogate splits:  
## Age < 39.5 to the right, agree=0.778, adj=0.222, (0 split)  
## StockOptionLevel splits as RRLR, agree=0.746, adj=0.111, (0 split)  
## YearsInCurrentRole < 5 to the right, agree=0.746, adj=0.111, (0 split)  
## YearsSinceLastPromotion < 6 to the right, agree=0.746, adj=0.111, (0 split)  
## TotalWorkingYears < 13.5 to the right, agree=0.730, adj=0.056, (0 split)  
##   
## Node number 8: 447 observations  
## predicted class=No expected loss=0.04697987 P(node) =0.4561224  
## class counts: 426 21  
## probabilities: 0.953 0.047   
##   
## Node number 9: 111 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.1531532 P(node) =0.1132653  
## class counts: 94 17  
## probabilities: 0.847 0.153   
## left son=18 (89 obs) right son=19 (22 obs)  
## Primary splits:  
## DailyRate < 417.5 to the right, improve=3.594631, (0 missing)  
## DistanceFromHome < 21.5 to the left, improve=3.409459, (0 missing)  
## JobRole splits as LRRLLLLRR, improve=3.117468, (0 missing)  
## Department splits as RLR, improve=1.723803, (0 missing)  
## TotalWorkingYears < 7.5 to the right, improve=1.621752, (0 missing)  
## Surrogate splits:  
## Department splits as RLL, agree=0.820, adj=0.091, (0 split)  
## JobRole splits as LRLLLLLLL, agree=0.820, adj=0.091, (0 split)  
## EducationField splits as RLLLLL, agree=0.811, adj=0.045, (0 split)  
##   
## Node number 10: 146 observations, complexity param=0.0125  
## predicted class=No expected loss=0.130137 P(node) =0.1489796  
## class counts: 127 19  
## probabilities: 0.870 0.130   
## left son=20 (124 obs) right son=21 (22 obs)  
## Primary splits:  
## DistanceFromHome < 21.5 to the left, improve=2.824589, (0 missing)  
## NumCompaniesWorked < 5.5 to the left, improve=2.154795, (0 missing)  
## YearsAtCompany < 3.5 to the right, improve=1.817952, (0 missing)  
## MonthlyRate < 21041.5 to the left, improve=1.733797, (0 missing)  
## TrainingTimesLastYear < 0.5 to the right, improve=1.711937, (0 missing)  
##   
## Node number 11: 63 observations, complexity param=0.02083333  
## predicted class=No expected loss=0.4444444 P(node) =0.06428571  
## class counts: 35 28  
## probabilities: 0.556 0.444   
## left son=22 (27 obs) right son=23 (36 obs)  
## Primary splits:  
## JobRole splits as LLRLLLLRR, improve=6.351852, (0 missing)  
## Department splits as LLR, improve=5.656566, (0 missing)  
## EducationField splits as -RRLRR, improve=4.424957, (0 missing)  
## TotalWorkingYears < 9.5 to the right, improve=3.968254, (0 missing)  
## DailyRate < 1412.5 to the left, improve=2.636535, (0 missing)  
## Surrogate splits:  
## Department splits as LLR, agree=0.873, adj=0.704, (0 split)  
## EducationField splits as -RRLLR, agree=0.683, adj=0.259, (0 split)  
## EnvironmentSatisfaction splits as RRLR, agree=0.683, adj=0.259, (0 split)  
## Gender splits as LR, agree=0.683, adj=0.259, (0 split)  
## MonthlyRate < 4437.5 to the left, agree=0.651, adj=0.185, (0 split)  
##   
## Node number 12: 96 observations, complexity param=0.0125  
## predicted class=No expected loss=0.09375 P(node) =0.09795918  
## class counts: 87 9  
## probabilities: 0.906 0.094   
## left son=24 (94 obs) right son=25 (2 obs)  
## Primary splits:  
## YearsSinceLastPromotion < 8 to the left, improve=3.355053, (0 missing)  
## EducationField splits as LLLLLR, improve=1.809826, (0 missing)  
## JobSatisfaction splits as RLRL, improve=1.397156, (0 missing)  
## MonthlyRate < 4005 to the right, improve=1.377717, (0 missing)  
## YearsInCurrentRole < 8 to the left, improve=1.377717, (0 missing)  
##   
## Node number 13: 54 observations, complexity param=0.03125  
## predicted class=No expected loss=0.462963 P(node) =0.05510204  
## class counts: 29 25  
## probabilities: 0.537 0.463   
## left son=26 (36 obs) right son=27 (18 obs)  
## Primary splits:  
## HourlyRate < 56.5 to the right, improve=5.351852, (0 missing)  
## BusinessTravel splits as LRL, improve=3.188808, (0 missing)  
## MonthlyRate < 24118 to the left, improve=3.178382, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.918059, (0 missing)  
## RelationshipSatisfaction splits as RLRL, improve=2.687079, (0 missing)  
## Surrogate splits:  
## EducationField splits as LLRLLL, agree=0.722, adj=0.167, (0 split)  
## WorkLifeBalance splits as LRLL, agree=0.722, adj=0.167, (0 split)  
## BusinessTravel splits as LRL, agree=0.704, adj=0.111, (0 split)  
## DailyRate < 1429 to the left, agree=0.704, adj=0.111, (0 split)  
## MonthlyRate < 25042.5 to the left, agree=0.704, adj=0.111, (0 split)  
##   
## Node number 14: 18 observations, complexity param=0.015625  
## predicted class=No expected loss=0.3888889 P(node) =0.01836735  
## class counts: 11 7  
## probabilities: 0.611 0.389   
## left son=28 (6 obs) right son=29 (12 obs)  
## Primary splits:  
## HourlyRate < 56.5 to the left, improve=2.722222, (0 missing)  
## MonthlyIncome < 2624 to the left, improve=2.722222, (0 missing)  
## YearsInCurrentRole < 6.5 to the left, improve=2.340171, (0 missing)  
## EducationField splits as -LRLRL, improve=1.680556, (0 missing)  
## JobInvolvement splits as RLLR, improve=1.680556, (0 missing)  
## Surrogate splits:  
## DailyRate < 347.5 to the left, agree=0.778, adj=0.333, (0 split)  
## Education splits as LRRR-, agree=0.778, adj=0.333, (0 split)  
## TrainingTimesLastYear < 2.5 to the right, agree=0.778, adj=0.333, (0 split)  
## YearsInCurrentRole < 1.5 to the left, agree=0.778, adj=0.333, (0 split)  
## DistanceFromHome < 2.5 to the left, agree=0.722, adj=0.167, (0 split)  
##   
## Node number 15: 45 observations, complexity param=0.015625  
## predicted class=Yes expected loss=0.2444444 P(node) =0.04591837  
## class counts: 11 34  
## probabilities: 0.244 0.756   
## left son=30 (15 obs) right son=31 (30 obs)  
## Primary splits:  
## DailyRate < 1067.5 to the right, improve=3.755556, (0 missing)  
## NumCompaniesWorked < 0.5 to the left, improve=3.669841, (0 missing)  
## DistanceFromHome < 12 to the left, improve=2.428674, (0 missing)  
## JobInvolvement splits as RRRL, improve=2.244173, (0 missing)  
## Education splits as LLRLL, improve=2.140741, (0 missing)  
## Surrogate splits:  
## Age < 36 to the right, agree=0.711, adj=0.133, (0 split)  
## HourlyRate < 35 to the left, agree=0.711, adj=0.133, (0 split)  
## MonthlyIncome < 1349 to the left, agree=0.711, adj=0.133, (0 split)  
## Education splits as RRRRL, agree=0.689, adj=0.067, (0 split)  
## EnvironmentSatisfaction splits as RRRL, agree=0.689, adj=0.067, (0 split)  
##   
## Node number 18: 89 observations  
## predicted class=No expected loss=0.08988764 P(node) =0.09081633  
## class counts: 81 8  
## probabilities: 0.910 0.090   
##   
## Node number 19: 22 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.4090909 P(node) =0.02244898  
## class counts: 13 9  
## probabilities: 0.591 0.409   
## left son=38 (17 obs) right son=39 (5 obs)  
## Primary splits:  
## DailyRate < 333 to the left, improve=4.518717, (0 missing)  
## DistanceFromHome < 8.5 to the left, improve=4.207792, (0 missing)  
## Department splits as LLR, improve=4.122078, (0 missing)  
## JobRole splits as LLL-LLLRR, improve=4.122078, (0 missing)  
## YearsInCurrentRole < 2.5 to the right, improve=3.103030, (0 missing)  
## Surrogate splits:  
## DistanceFromHome < 17.5 to the left, agree=0.818, adj=0.2, (0 split)  
## EducationField splits as RLLLLL, agree=0.818, adj=0.2, (0 split)  
## JobRole splits as LLL-LLLLR, agree=0.818, adj=0.2, (0 split)  
## NumCompaniesWorked < 0.5 to the right, agree=0.818, adj=0.2, (0 split)  
##   
## Node number 20: 124 observations  
## predicted class=No expected loss=0.08870968 P(node) =0.1265306  
## class counts: 113 11  
## probabilities: 0.911 0.089   
##   
## Node number 21: 22 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3636364 P(node) =0.02244898  
## class counts: 14 8  
## probabilities: 0.636 0.364   
## left son=42 (18 obs) right son=43 (4 obs)  
## Primary splits:  
## EducationField splits as RLRLLL, improve=3.959596, (0 missing)  
## JobRole splits as RRLRLLLR-, improve=3.753247, (0 missing)  
## YearsInCurrentRole < 7.5 to the right, improve=2.715152, (0 missing)  
## YearsAtCompany < 11 to the right, improve=1.711230, (0 missing)  
## Department splits as RLR, improve=1.515152, (0 missing)  
## Surrogate splits:  
## Department splits as RLR, agree=0.909, adj=0.5, (0 split)  
## JobRole splits as LRLLLLLR-, agree=0.909, adj=0.5, (0 split)  
##   
## Node number 22: 27 observations, complexity param=0.0125  
## predicted class=No expected loss=0.1851852 P(node) =0.02755102  
## class counts: 22 5  
## probabilities: 0.815 0.185   
## left son=44 (25 obs) right son=45 (2 obs)  
## Primary splits:  
## JobInvolvement splits as RLLL, improve=2.868148, (0 missing)  
## DailyRate < 1011 to the left, improve=2.819577, (0 missing)  
## YearsSinceLastPromotion < 5 to the left, improve=2.111785, (0 missing)  
## Education splits as RLLLL, improve=1.564815, (0 missing)  
## RelationshipSatisfaction splits as RLLL, improve=1.529101, (0 missing)  
##   
## Node number 23: 36 observations, complexity param=0.01875  
## predicted class=Yes expected loss=0.3611111 P(node) =0.03673469  
## class counts: 13 23  
## probabilities: 0.361 0.639   
## left son=46 (17 obs) right son=47 (19 obs)  
## Primary splits:  
## TotalWorkingYears < 9.5 to the right, improve=3.323185, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=2.777778, (0 missing)  
## MonthlyRate < 8860.5 to the left, improve=2.400202, (0 missing)  
## YearsAtCompany < 8.5 to the right, improve=2.400202, (0 missing)  
## JobInvolvement splits as RRLR, improve=2.312929, (0 missing)  
## Surrogate splits:  
## MonthlyIncome < 6489.5 to the right, agree=0.750, adj=0.471, (0 split)  
## YearsAtCompany < 8.5 to the right, agree=0.722, adj=0.412, (0 split)  
## YearsInCurrentRole < 4.5 to the right, agree=0.722, adj=0.412, (0 split)  
## JobLevel splits as RRLL-, agree=0.694, adj=0.353, (0 split)  
## MonthlyRate < 17153 to the left, agree=0.694, adj=0.353, (0 split)  
##   
## Node number 24: 94 observations  
## predicted class=No expected loss=0.07446809 P(node) =0.09591837  
## class counts: 87 7  
## probabilities: 0.926 0.074   
##   
## Node number 25: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 26: 36 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3055556 P(node) =0.03673469  
## class counts: 25 11  
## probabilities: 0.694 0.306   
## left son=52 (26 obs) right son=53 (10 obs)  
## Primary splits:  
## DistanceFromHome < 11 to the left, improve=2.400855, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.207544, (0 missing)  
## HourlyRate < 84.5 to the left, improve=2.177778, (0 missing)  
## RelationshipSatisfaction splits as RLLL, improve=2.099206, (0 missing)  
## MonthlyRate < 24118 to the left, improve=2.042484, (0 missing)  
## Surrogate splits:  
## JobInvolvement splits as RLLR, agree=0.806, adj=0.3, (0 split)  
## DailyRate < 158 to the right, agree=0.778, adj=0.2, (0 split)  
## EducationField splits as RL-LRL, agree=0.778, adj=0.2, (0 split)  
## HourlyRate < 60 to the right, agree=0.778, adj=0.2, (0 split)  
## MonthlyIncome < 2543 to the left, agree=0.778, adj=0.2, (0 split)  
##   
## Node number 27: 18 observations, complexity param=0.0125  
## predicted class=Yes expected loss=0.2222222 P(node) =0.01836735  
## class counts: 4 14  
## probabilities: 0.222 0.778   
## left son=54 (2 obs) right son=55 (16 obs)  
## Primary splits:  
## BusinessTravel splits as LRR, improve=2.722222, (0 missing)  
## DailyRate < 1382.5 to the right, improve=2.722222, (0 missing)  
## Age < 34.5 to the right, improve=1.976068, (0 missing)  
## JobRole splits as -RR---L-R, improve=1.976068, (0 missing)  
## YearsAtCompany < 0.5 to the left, improve=1.976068, (0 missing)  
##   
## Node number 28: 6 observations  
## predicted class=No expected loss=0 P(node) =0.006122449  
## class counts: 6 0  
## probabilities: 1.000 0.000   
##   
## Node number 29: 12 observations, complexity param=0.015625  
## predicted class=Yes expected loss=0.4166667 P(node) =0.0122449  
## class counts: 5 7  
## probabilities: 0.417 0.583   
## left son=58 (3 obs) right son=59 (9 obs)  
## Primary splits:  
## MonthlyIncome < 2621 to the left, improve=2.722222, (0 missing)  
## DistanceFromHome < 4 to the right, improve=2.083333, (0 missing)  
## JobSatisfaction splits as LLRL, improve=2.083333, (0 missing)  
## PercentSalaryHike < 14.5 to the right, improve=1.633333, (0 missing)  
## BusinessTravel splits as -RL, improve=1.388889, (0 missing)  
## Surrogate splits:  
## MonthlyRate < 20652 to the right, agree=0.833, adj=0.333, (0 split)  
## RelationshipSatisfaction splits as -LRR, agree=0.833, adj=0.333, (0 split)  
##   
## Node number 30: 15 observations, complexity param=0.015625  
## predicted class=No expected loss=0.4666667 P(node) =0.01530612  
## class counts: 8 7  
## probabilities: 0.533 0.467   
## left son=60 (7 obs) right son=61 (8 obs)  
## Primary splits:  
## RelationshipSatisfaction splits as LRLR, improve=2.752381, (0 missing)  
## EnvironmentSatisfaction splits as RLLL, improve=2.133333, (0 missing)  
## MonthlyRate < 4623.5 to the right, improve=2.133333, (0 missing)  
## NumCompaniesWorked < 4.5 to the left, improve=2.133333, (0 missing)  
## DailyRate < 1301.5 to the left, improve=1.800000, (0 missing)  
## Surrogate splits:  
## DistanceFromHome < 5.5 to the left, agree=0.867, adj=0.714, (0 split)  
## WorkLifeBalance splits as RLRL, agree=0.867, adj=0.714, (0 split)  
## DailyRate < 1301.5 to the left, agree=0.800, adj=0.571, (0 split)  
## EducationField splits as LLRRRR, agree=0.733, adj=0.429, (0 split)  
## HourlyRate < 64.5 to the left, agree=0.733, adj=0.429, (0 split)  
##   
## Node number 31: 30 observations  
## predicted class=Yes expected loss=0.1 P(node) =0.03061224  
## class counts: 3 27  
## probabilities: 0.100 0.900   
##   
## Node number 38: 17 observations  
## predicted class=No expected loss=0.2352941 P(node) =0.01734694  
## class counts: 13 4  
## probabilities: 0.765 0.235   
##   
## Node number 39: 5 observations  
## predicted class=Yes expected loss=0 P(node) =0.005102041  
## class counts: 0 5  
## probabilities: 0.000 1.000   
##   
## Node number 42: 18 observations  
## predicted class=No expected loss=0.2222222 P(node) =0.01836735  
## class counts: 14 4  
## probabilities: 0.778 0.222   
##   
## Node number 43: 4 observations  
## predicted class=Yes expected loss=0 P(node) =0.004081633  
## class counts: 0 4  
## probabilities: 0.000 1.000   
##   
## Node number 44: 25 observations  
## predicted class=No expected loss=0.12 P(node) =0.0255102  
## class counts: 22 3  
## probabilities: 0.880 0.120   
##   
## Node number 45: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000   
##   
## Node number 46: 17 observations  
## predicted class=No expected loss=0.4117647 P(node) =0.01734694  
## class counts: 10 7  
## probabilities: 0.588 0.412   
##   
## Node number 47: 19 observations  
## predicted class=Yes expected loss=0.1578947 P(node) =0.01938776  
## class counts: 3 16  
## probabilities: 0.158 0.842   
##   
## Node number 52: 26 observations  
## predicted class=No expected loss=0.1923077 P(node) =0.02653061  
## class counts: 21 5  
## probabilities: 0.808 0.192   
##   
## Node number 53: 10 observations  
## predicted class=Yes expected loss=0.4 P(node) =0.01020408  
## class counts: 4 6  
## probabilities: 0.400 0.600   
##   
## Node number 54: 2 observations  
## predicted class=No expected loss=0 P(node) =0.002040816  
## class counts: 2 0  
## probabilities: 1.000 0.000   
##   
## Node number 55: 16 observations  
## predicted class=Yes expected loss=0.125 P(node) =0.01632653  
## class counts: 2 14  
## probabilities: 0.125 0.875   
##   
## Node number 58: 3 observations  
## predicted class=No expected loss=0 P(node) =0.003061224  
## class counts: 3 0  
## probabilities: 1.000 0.000   
##   
## Node number 59: 9 observations  
## predicted class=Yes expected loss=0.2222222 P(node) =0.009183673  
## class counts: 2 7  
## probabilities: 0.222 0.778   
##   
## Node number 60: 7 observations  
## predicted class=No expected loss=0.1428571 P(node) =0.007142857  
## class counts: 6 1  
## probabilities: 0.857 0.143   
##   
## Node number 61: 8 observations  
## predicted class=Yes expected loss=0.25 P(node) =0.008163265  
## class counts: 2 6  
## probabilities: 0.250 0.750



## No Yes   
## 445 45   
## actualAttrition  
## predictedAttrition No Yes  
## No 386 59  
## Yes 27 18

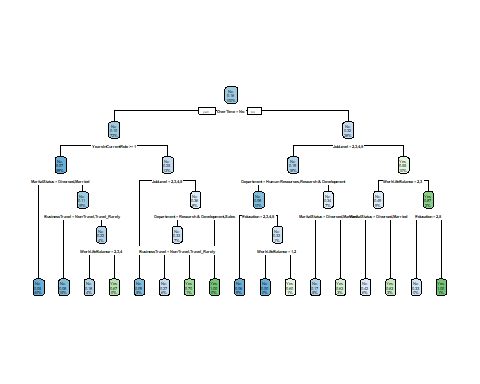
Accuracy: 406/490 = ~.829

An increase Based on previous information from KMeans and from Apriori, let’s select/remove fields. Selecting:

* Attrition
* BusinessTravel
* Department
* Education
* JobLevel
* MaritalStatus
* MonthlyIncome
* OverTime
* WorkLifeBalance
* YearsWithCurrManager
* YearsInCurrentRole

treeSpecific <- data.frame(HR\_tree$Attrition, HR\_tree$BusinessTravel, HR\_tree$Department, HR\_tree$Education, HR\_tree$JobLevel, HR\_tree$MaritalStatus, HR\_tree$OverTime, HR\_tree$WorkLifeBalance, HR\_tree$YearsInCurrentRole, HR\_tree$YearsWithCurrManager )  
  
# Picking specific attributes based on what the previous analysis  
colnames(treeSpecific) <- c("Attrition","BusinessTravel","Department","Education","JobLevel","MaritalStatus","OverTime","WorkLifeBalance","YearsWithCurrManager","YearsInCurrentRole")  
  
specificTree <- printDecision(seedNum1, treeSpecific)

## Call:  
## rpart(formula = Attrition ~ ., data = train, method = "class",   
## control = rpart.control(cp = 0, minsplit = 5, maxdepth = depth))  
## n= 980   
##   
## CP nsplit rel error xerror xstd  
## 1 0.031250000 0 1.00000 1.00000 0.07231592  
## 2 0.028125000 2 0.93750 1.01875 0.07285709  
## 3 0.021875000 4 0.88125 1.04375 0.07356487  
## 4 0.012500000 6 0.83750 1.02500 0.07303550  
## 5 0.006250000 10 0.78750 1.03750 0.07338938  
## 6 0.003125000 11 0.78125 1.05000 0.07373941  
## 7 0.002083333 13 0.77500 1.10000 0.07510197  
## 8 0.000000000 16 0.76875 1.10625 0.07526817  
##   
## Variable importance  
## JobLevel OverTime MaritalStatus   
## 22 20 12   
## Department YearsInCurrentRole WorkLifeBalance   
## 11 10 7   
## YearsWithCurrManager Education BusinessTravel   
## 7 6 5   
##   
## Node number 1: 980 observations, complexity param=0.03125  
## predicted class=No expected loss=0.1632653 P(node) =1  
## class counts: 820 160  
## probabilities: 0.837 0.163   
## left son=2 (708 obs) right son=3 (272 obs)  
## Primary splits:  
## OverTime splits as LR, improve=19.340350, (0 missing)  
## JobLevel splits as RLLLL, improve=14.473920, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=11.862970, (0 missing)  
## MaritalStatus splits as LLR, improve= 8.850673, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve= 5.865829, (0 missing)  
##   
## Node number 2: 708 observations, complexity param=0.0125  
## predicted class=No expected loss=0.1016949 P(node) =0.722449  
## class counts: 636 72  
## probabilities: 0.898 0.102   
## left son=4 (581 obs) right son=5 (127 obs)  
## Primary splits:  
## YearsInCurrentRole < 0.5 to the right, improve=6.989662, (0 missing)  
## JobLevel splits as RLLLL, improve=3.848129, (0 missing)  
## YearsWithCurrManager < 1.5 to the right, improve=3.663303, (0 missing)  
## MaritalStatus splits as LLR, improve=2.517111, (0 missing)  
## BusinessTravel splits as LRL, improve=1.754625, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 0.5 to the right, agree=0.908, adj=0.488, (0 split)  
##   
## Node number 3: 272 observations, complexity param=0.03125  
## predicted class=No expected loss=0.3235294 P(node) =0.277551  
## class counts: 184 88  
## probabilities: 0.676 0.324   
## left son=6 (172 obs) right son=7 (100 obs)  
## Primary splits:  
## JobLevel splits as RLLLL, improve=16.221610, (0 missing)  
## MaritalStatus splits as LLR, improve=11.235870, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve= 5.125997, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve= 5.110382, (0 missing)  
## Department splits as LLR, improve= 1.904673, (0 missing)  
## Surrogate splits:  
## YearsInCurrentRole < 2.5 to the right, agree=0.684, adj=0.14, (0 split)  
## YearsWithCurrManager < 2.5 to the right, agree=0.665, adj=0.09, (0 split)  
##   
## Node number 4: 581 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.06884682 P(node) =0.5928571  
## class counts: 541 40  
## probabilities: 0.931 0.069   
## left son=8 (390 obs) right son=9 (191 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=0.9613378, (0 missing)  
## BusinessTravel splits as LRL, improve=0.6394274, (0 missing)  
## JobLevel splits as RLRLL, improve=0.5082975, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=0.4528595, (0 missing)  
## Department splits as RLR, improve=0.4522810, (0 missing)  
##   
## Node number 5: 127 observations, complexity param=0.0125  
## predicted class=No expected loss=0.2519685 P(node) =0.1295918  
## class counts: 95 32  
## probabilities: 0.748 0.252   
## left son=10 (53 obs) right son=11 (74 obs)  
## Primary splits:  
## JobLevel splits as RLLLL, improve=4.520115, (0 missing)  
## Department splits as RLL, improve=3.126475, (0 missing)  
## BusinessTravel splits as LRL, improve=2.920745, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=1.968316, (0 missing)  
## MaritalStatus splits as LLR, improve=1.397728, (0 missing)  
## Surrogate splits:  
## Education splits as RRRLL, agree=0.685, adj=0.245, (0 split)  
## YearsWithCurrManager < 2.5 to the right, agree=0.677, adj=0.226, (0 split)  
## MaritalStatus splits as RLR, agree=0.614, adj=0.075, (0 split)  
## Department splits as RRL, agree=0.606, adj=0.057, (0 split)  
## WorkLifeBalance splits as RRRL, agree=0.606, adj=0.057, (0 split)  
##   
## Node number 6: 172 observations, complexity param=0.021875  
## predicted class=No expected loss=0.1918605 P(node) =0.1755102  
## class counts: 139 33  
## probabilities: 0.808 0.192   
## left son=12 (99 obs) right son=13 (73 obs)  
## Primary splits:  
## Department splits as LLR, improve=5.7534260, (0 missing)  
## MaritalStatus splits as LLR, improve=5.5401660, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=1.4887240, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve=1.3325820, (0 missing)  
## JobLevel splits as -LRLL, improve=0.6820912, (0 missing)  
## Surrogate splits:  
## MaritalStatus splits as LLR, agree=0.610, adj=0.082, (0 split)  
## YearsWithCurrManager < 0.5 to the right, agree=0.605, adj=0.068, (0 split)  
## YearsInCurrentRole < 0.5 to the right, agree=0.599, adj=0.055, (0 split)  
## Education splits as RLLLL, agree=0.581, adj=0.014, (0 split)  
##   
## Node number 7: 100 observations, complexity param=0.028125  
## predicted class=Yes expected loss=0.45 P(node) =0.1020408  
## class counts: 45 55  
## probabilities: 0.450 0.550   
## left son=14 (85 obs) right son=15 (15 obs)  
## Primary splits:  
## WorkLifeBalance splits as RLLR, improve=3.539216, (0 missing)  
## MaritalStatus splits as LLR, improve=3.158744, (0 missing)  
## Education splits as LRRRL, improve=1.590716, (0 missing)  
## YearsWithCurrManager < 8.5 to the right, improve=1.289474, (0 missing)  
## Department splits as LLR, improve=1.186275, (0 missing)  
##   
## Node number 8: 390 observations  
## predicted class=No expected loss=0.04871795 P(node) =0.3979592  
## class counts: 371 19  
## probabilities: 0.951 0.049   
##   
## Node number 9: 191 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.1099476 P(node) =0.194898  
## class counts: 170 21  
## probabilities: 0.890 0.110   
## left son=18 (150 obs) right son=19 (41 obs)  
## Primary splits:  
## BusinessTravel splits as LRL, improve=1.2534180, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=0.3735579, (0 missing)  
## JobLevel splits as RRRLL, improve=0.3652498, (0 missing)  
## YearsWithCurrManager < 3.5 to the right, improve=0.3490413, (0 missing)  
## YearsInCurrentRole < 3.5 to the right, improve=0.2502361, (0 missing)  
##   
## Node number 10: 53 observations  
## predicted class=No expected loss=0.09433962 P(node) =0.05408163  
## class counts: 48 5  
## probabilities: 0.906 0.094   
##   
## Node number 11: 74 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3648649 P(node) =0.0755102  
## class counts: 47 27  
## probabilities: 0.635 0.365   
## left son=22 (70 obs) right son=23 (4 obs)  
## Primary splits:  
## Department splits as RLL, improve=3.4115830, (0 missing)  
## BusinessTravel splits as LRL, improve=3.3939780, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.4413330, (0 missing)  
## MaritalStatus splits as LLR, improve=1.6160110, (0 missing)  
## Education splits as LLLRR, improve=0.8427518, (0 missing)  
## Surrogate splits:  
## Education splits as LLLLR, agree=0.959, adj=0.25, (0 split)  
##   
## Node number 12: 99 observations, complexity param=0.003125  
## predicted class=No expected loss=0.08080808 P(node) =0.1010204  
## class counts: 91 8  
## probabilities: 0.919 0.081   
## left son=24 (90 obs) right son=25 (9 obs)  
## Primary splits:  
## Education splits as RLLLL, improve=1.2626260, (0 missing)  
## JobLevel splits as -LRLL, improve=0.7463061, (0 missing)  
## YearsInCurrentRole < 5.5 to the left, improve=0.3103360, (0 missing)  
## WorkLifeBalance splits as LRLR, improve=0.2375055, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve=0.1760362, (0 missing)  
##   
## Node number 13: 73 observations, complexity param=0.021875  
## predicted class=No expected loss=0.3424658 P(node) =0.0744898  
## class counts: 48 25  
## probabilities: 0.658 0.342   
## left son=26 (46 obs) right son=27 (27 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=7.066728, (0 missing)  
## Education splits as LRRRL, improve=1.639176, (0 missing)  
## WorkLifeBalance splits as RRRL, improve=1.259065, (0 missing)  
## YearsWithCurrManager < 10.5 to the right, improve=1.259065, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=1.215174, (0 missing)  
## Surrogate splits:  
## BusinessTravel splits as RLL, agree=0.644, adj=0.037, (0 split)  
## WorkLifeBalance splits as RLLL, agree=0.644, adj=0.037, (0 split)  
## YearsWithCurrManager < 1 to the right, agree=0.644, adj=0.037, (0 split)  
##   
## Node number 14: 85 observations, complexity param=0.028125  
## predicted class=No expected loss=0.4941176 P(node) =0.08673469  
## class counts: 43 42  
## probabilities: 0.506 0.494   
## left son=28 (55 obs) right son=29 (30 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=1.7971480, (0 missing)  
## YearsInCurrentRole < 6.5 to the left, improve=1.0845940, (0 missing)  
## YearsWithCurrManager < 9.5 to the right, improve=1.0001420, (0 missing)  
## Department splits as LLR, improve=0.8320172, (0 missing)  
## Education splits as LRRRL, improve=0.5593350, (0 missing)  
## Surrogate splits:  
## Education splits as RLLLL, agree=0.706, adj=0.167, (0 split)  
## YearsWithCurrManager < 7.5 to the left, agree=0.659, adj=0.033, (0 split)  
##   
## Node number 15: 15 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.1333333 P(node) =0.01530612  
## class counts: 2 13  
## probabilities: 0.133 0.867   
## left son=30 (3 obs) right son=31 (12 obs)  
## Primary splits:  
## Education splits as -LRRL, improve=2.1333330, (0 missing)  
## MaritalStatus splits as LRR, improve=0.6205128, (0 missing)  
## YearsWithCurrManager < 2.5 to the right, improve=0.6205128, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=0.2666667, (0 missing)  
## BusinessTravel splits as RRL, improve=0.1939394, (0 missing)  
## Surrogate splits:  
## Department splits as LRR, agree=0.867, adj=0.333, (0 split)  
##   
## Node number 18: 150 observations  
## predicted class=No expected loss=0.08 P(node) =0.1530612  
## class counts: 138 12  
## probabilities: 0.920 0.080   
##   
## Node number 19: 41 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.2195122 P(node) =0.04183673  
## class counts: 32 9  
## probabilities: 0.780 0.220   
## left son=38 (38 obs) right son=39 (3 obs)  
## Primary splits:  
## WorkLifeBalance splits as RLLL, improve=1.29439500, (0 missing)  
## Education splits as LLLRR, improve=0.43958510, (0 missing)  
## YearsWithCurrManager < 3.5 to the right, improve=0.37735190, (0 missing)  
## YearsInCurrentRole < 1.5 to the right, improve=0.33083180, (0 missing)  
## Department splits as LRL, improve=0.04271988, (0 missing)  
##   
## Node number 22: 70 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3285714 P(node) =0.07142857  
## class counts: 47 23  
## probabilities: 0.671 0.329   
## left son=44 (60 obs) right son=45 (10 obs)  
## Primary splits:  
## BusinessTravel splits as LRL, improve=3.2190480, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=3.0857140, (0 missing)  
## MaritalStatus splits as LLR, improve=2.3142860, (0 missing)  
## Education splits as LLLR-, improve=0.6857143, (0 missing)  
## Department splits as -LR, improve=0.4460858, (0 missing)  
##   
## Node number 23: 4 observations  
## predicted class=Yes expected loss=0 P(node) =0.004081633  
## class counts: 0 4  
## probabilities: 0.000 1.000   
##   
## Node number 24: 90 observations  
## predicted class=No expected loss=0.05555556 P(node) =0.09183673  
## class counts: 85 5  
## probabilities: 0.944 0.056   
##   
## Node number 25: 9 observations, complexity param=0.003125  
## predicted class=No expected loss=0.3333333 P(node) =0.009183673  
## class counts: 6 3  
## probabilities: 0.667 0.333   
## left son=50 (4 obs) right son=51 (5 obs)  
## Primary splits:  
## WorkLifeBalance splits as LLRR, improve=1.6000000, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve=1.6000000, (0 missing)  
## JobLevel splits as -LRLL, improve=1.0000000, (0 missing)  
## MaritalStatus splits as LRR, improve=0.5714286, (0 missing)  
## YearsInCurrentRole < 7.5 to the right, improve=0.5714286, (0 missing)  
## Surrogate splits:  
## MaritalStatus splits as RLR, agree=0.778, adj=0.50, (0 split)  
## BusinessTravel splits as LRR, agree=0.667, adj=0.25, (0 split)  
## YearsInCurrentRole < 4 to the left, agree=0.667, adj=0.25, (0 split)  
##   
## Node number 26: 46 observations  
## predicted class=No expected loss=0.173913 P(node) =0.04693878  
## class counts: 38 8  
## probabilities: 0.826 0.174   
##   
## Node number 27: 27 observations  
## predicted class=Yes expected loss=0.3703704 P(node) =0.02755102  
## class counts: 10 17  
## probabilities: 0.370 0.630   
##   
## Node number 28: 55 observations  
## predicted class=No expected loss=0.4181818 P(node) =0.05612245  
## class counts: 32 23  
## probabilities: 0.582 0.418   
##   
## Node number 29: 30 observations  
## predicted class=Yes expected loss=0.3666667 P(node) =0.03061224  
## class counts: 11 19  
## probabilities: 0.367 0.633   
##   
## Node number 30: 3 observations  
## predicted class=No expected loss=0.3333333 P(node) =0.003061224  
## class counts: 2 1  
## probabilities: 0.667 0.333   
##   
## Node number 31: 12 observations  
## predicted class=Yes expected loss=0 P(node) =0.0122449  
## class counts: 0 12  
## probabilities: 0.000 1.000   
##   
## Node number 38: 38 observations  
## predicted class=No expected loss=0.1842105 P(node) =0.03877551  
## class counts: 31 7  
## probabilities: 0.816 0.184   
##   
## Node number 39: 3 observations  
## predicted class=Yes expected loss=0.3333333 P(node) =0.003061224  
## class counts: 1 2  
## probabilities: 0.333 0.667   
##   
## Node number 44: 60 observations  
## predicted class=No expected loss=0.2666667 P(node) =0.06122449  
## class counts: 44 16  
## probabilities: 0.733 0.267   
##   
## Node number 45: 10 observations  
## predicted class=Yes expected loss=0.3 P(node) =0.01020408  
## class counts: 3 7  
## probabilities: 0.300 0.700   
##   
## Node number 50: 4 observations  
## predicted class=No expected loss=0 P(node) =0.004081633  
## class counts: 4 0  
## probabilities: 1.000 0.000   
##   
## Node number 51: 5 observations  
## predicted class=Yes expected loss=0.4 P(node) =0.005102041  
## class counts: 2 3  
## probabilities: 0.400 0.600



## No Yes   
## 445 45   
## actualAttrition  
## predictedAttrition No Yes  
## No 391 54  
## Yes 22 23

Predicted Accuracy: 414/490 = ~ .845

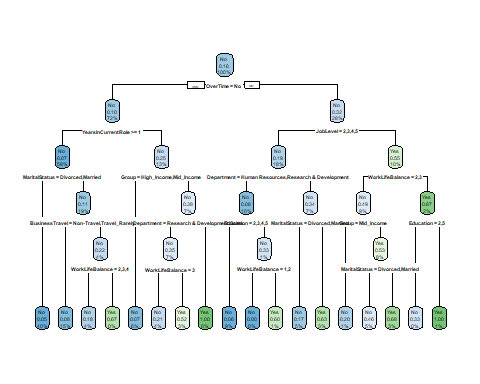
Adding in income to see if it changes.

# Determining percentiles  
Percentile\_00 = min(HR\_tree$MonthlyIncome)  
Percentile\_33 = quantile(HR\_tree$MonthlyIncome, 0.33333)  
Percentile\_67 = quantile(HR\_tree$MonthlyIncome, 0.66667)  
Percentile\_100 = max(HR\_tree$MonthlyIncome)  
  
# Values  
HR.Bind = rbind(Percentile\_00, Percentile\_33, Percentile\_67, Percentile\_100)  
dimnames(HR.Bind)[[2]] = "Value"  
HR.Bind

## Value  
## Percentile\_00 1009.000  
## Percentile\_33 3631.647  
## Percentile\_67 6528.735  
## Percentile\_100 19999.000

# Grouping  
treeIncome <- treeSpecific  
treeIncome$income <- HR\_tree$MonthlyIncome  
treeIncome$Group[treeIncome$income >= Percentile\_00 & treeIncome$income < Percentile\_33] = "Low\_Income"  
treeIncome$Group[treeIncome$income >= Percentile\_33 & treeIncome$income < Percentile\_67] = "Mid\_Income"  
treeIncome$Group[treeIncome$income >= Percentile\_67 & treeIncome$income <= Percentile\_100] = "High\_Income"  
treeIncome$income <- NULL  
  
incomeTree <- printDecision(seedNum1, treeIncome)

## Call:  
## rpart(formula = Attrition ~ ., data = train, method = "class",   
## control = rpart.control(cp = 0, minsplit = 5, maxdepth = depth))  
## n= 980   
##   
## CP nsplit rel error xerror xstd  
## 1 0.031250000 0 1.00000 1.00000 0.07231592  
## 2 0.021875000 2 0.93750 1.01875 0.07285709  
## 3 0.018750000 4 0.89375 1.05625 0.07391299  
## 4 0.008333333 7 0.83125 1.03750 0.07338938  
## 5 0.006250000 10 0.80625 1.06875 0.07425732  
## 6 0.003125000 12 0.79375 1.07500 0.07442809  
## 7 0.002083333 14 0.78750 1.11250 0.07543348  
## 8 0.000000000 17 0.78125 1.11875 0.07559790  
##   
## Variable importance  
## Group JobLevel OverTime   
## 18 18 16   
## MaritalStatus WorkLifeBalance YearsInCurrentRole   
## 9 8 8   
## Department YearsWithCurrManager Education   
## 8 6 5   
## BusinessTravel   
## 2   
##   
## Node number 1: 980 observations, complexity param=0.03125  
## predicted class=No expected loss=0.1632653 P(node) =1  
## class counts: 820 160  
## probabilities: 0.837 0.163   
## left son=2 (708 obs) right son=3 (272 obs)  
## Primary splits:  
## OverTime splits as LR, improve=19.340350, (0 missing)  
## JobLevel splits as RLLLL, improve=14.473920, (0 missing)  
## Group splits as LRL, improve=12.147050, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=11.862970, (0 missing)  
## MaritalStatus splits as LLR, improve= 8.850673, (0 missing)  
##   
## Node number 2: 708 observations, complexity param=0.008333333  
## predicted class=No expected loss=0.1016949 P(node) =0.722449  
## class counts: 636 72  
## probabilities: 0.898 0.102   
## left son=4 (581 obs) right son=5 (127 obs)  
## Primary splits:  
## YearsInCurrentRole < 0.5 to the right, improve=6.989662, (0 missing)  
## JobLevel splits as RLLLL, improve=3.848129, (0 missing)  
## YearsWithCurrManager < 1.5 to the right, improve=3.663303, (0 missing)  
## Group splits as LRL, improve=2.618150, (0 missing)  
## MaritalStatus splits as LLR, improve=2.517111, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 0.5 to the right, agree=0.908, adj=0.488, (0 split)  
##   
## Node number 3: 272 observations, complexity param=0.03125  
## predicted class=No expected loss=0.3235294 P(node) =0.277551  
## class counts: 184 88  
## probabilities: 0.676 0.324   
## left son=6 (172 obs) right son=7 (100 obs)  
## Primary splits:  
## JobLevel splits as RLLLL, improve=16.221610, (0 missing)  
## Group splits as LRL, improve=15.902780, (0 missing)  
## MaritalStatus splits as LLR, improve=11.235870, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve= 5.125997, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve= 5.110382, (0 missing)  
## Surrogate splits:  
## Group splits as LRL, agree=0.949, adj=0.86, (0 split)  
## YearsInCurrentRole < 2.5 to the right, agree=0.684, adj=0.14, (0 split)  
## YearsWithCurrManager < 2.5 to the right, agree=0.665, adj=0.09, (0 split)  
##   
## Node number 4: 581 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.06884682 P(node) =0.5928571  
## class counts: 541 40  
## probabilities: 0.931 0.069   
## left son=8 (390 obs) right son=9 (191 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=0.9613378, (0 missing)  
## BusinessTravel splits as LRL, improve=0.6394274, (0 missing)  
## JobLevel splits as RLRLL, improve=0.5082975, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=0.4528595, (0 missing)  
## Department splits as RLR, improve=0.4522810, (0 missing)  
##   
## Node number 5: 127 observations, complexity param=0.008333333  
## predicted class=No expected loss=0.2519685 P(node) =0.1295918  
## class counts: 95 32  
## probabilities: 0.748 0.252   
## left son=10 (54 obs) right son=11 (73 obs)  
## Primary splits:  
## Group splits as LRL, improve=5.946060, (0 missing)  
## JobLevel splits as RLLLL, improve=4.520115, (0 missing)  
## Department splits as RLL, improve=3.126475, (0 missing)  
## BusinessTravel splits as LRL, improve=2.920745, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=1.968316, (0 missing)  
## Surrogate splits:  
## JobLevel splits as RLLLL, agree=0.945, adj=0.870, (0 split)  
## Education splits as RRRLL, agree=0.677, adj=0.241, (0 split)  
## YearsWithCurrManager < 2.5 to the right, agree=0.669, adj=0.222, (0 split)  
## MaritalStatus splits as RLR, agree=0.591, adj=0.037, (0 split)  
## WorkLifeBalance splits as RRRL, agree=0.583, adj=0.019, (0 split)  
##   
## Node number 6: 172 observations, complexity param=0.021875  
## predicted class=No expected loss=0.1918605 P(node) =0.1755102  
## class counts: 139 33  
## probabilities: 0.808 0.192   
## left son=12 (99 obs) right son=13 (73 obs)  
## Primary splits:  
## Department splits as LLR, improve=5.7534260, (0 missing)  
## MaritalStatus splits as LLR, improve=5.5401660, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=1.4887240, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve=1.3325820, (0 missing)  
## JobLevel splits as -LRLL, improve=0.6820912, (0 missing)  
## Surrogate splits:  
## MaritalStatus splits as LLR, agree=0.610, adj=0.082, (0 split)  
## YearsWithCurrManager < 0.5 to the right, agree=0.605, adj=0.068, (0 split)  
## YearsInCurrentRole < 0.5 to the right, agree=0.599, adj=0.055, (0 split)  
## Education splits as RLLLL, agree=0.581, adj=0.014, (0 split)  
##   
## Node number 7: 100 observations, complexity param=0.01875  
## predicted class=Yes expected loss=0.45 P(node) =0.1020408  
## class counts: 45 55  
## probabilities: 0.450 0.550   
## left son=14 (85 obs) right son=15 (15 obs)  
## Primary splits:  
## WorkLifeBalance splits as RLLR, improve=3.539216, (0 missing)  
## MaritalStatus splits as LLR, improve=3.158744, (0 missing)  
## Education splits as LRRRL, improve=1.590716, (0 missing)  
## YearsWithCurrManager < 8.5 to the right, improve=1.289474, (0 missing)  
## Group splits as -RL, improve=1.280303, (0 missing)  
##   
## Node number 8: 390 observations  
## predicted class=No expected loss=0.04871795 P(node) =0.3979592  
## class counts: 371 19  
## probabilities: 0.951 0.049   
##   
## Node number 9: 191 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.1099476 P(node) =0.194898  
## class counts: 170 21  
## probabilities: 0.890 0.110   
## left son=18 (150 obs) right son=19 (41 obs)  
## Primary splits:  
## BusinessTravel splits as LRL, improve=1.2534180, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=0.3735579, (0 missing)  
## JobLevel splits as RRRLL, improve=0.3652498, (0 missing)  
## YearsWithCurrManager < 3.5 to the right, improve=0.3490413, (0 missing)  
## Group splits as LLR, improve=0.3330036, (0 missing)  
##   
## Node number 10: 54 observations  
## predicted class=No expected loss=0.07407407 P(node) =0.05510204  
## class counts: 50 4  
## probabilities: 0.926 0.074   
##   
## Node number 11: 73 observations, complexity param=0.008333333  
## predicted class=No expected loss=0.3835616 P(node) =0.0744898  
## class counts: 45 28  
## probabilities: 0.616 0.384   
## left son=22 (69 obs) right son=23 (4 obs)  
## Primary splits:  
## Department splits as RLL, improve=3.2162000, (0 missing)  
## BusinessTravel splits as LRL, improve=3.0601370, (0 missing)  
## WorkLifeBalance splits as RRLR, improve=2.4854870, (0 missing)  
## MaritalStatus splits as LLR, improve=1.4032550, (0 missing)  
## Education splits as LLLRR, improve=0.6789057, (0 missing)  
## Surrogate splits:  
## Education splits as LLLLR, agree=0.959, adj=0.25, (0 split)  
##   
## Node number 12: 99 observations, complexity param=0.003125  
## predicted class=No expected loss=0.08080808 P(node) =0.1010204  
## class counts: 91 8  
## probabilities: 0.919 0.081   
## left son=24 (90 obs) right son=25 (9 obs)  
## Primary splits:  
## Education splits as RLLLL, improve=1.2626260, (0 missing)  
## JobLevel splits as -LRLL, improve=0.7463061, (0 missing)  
## YearsInCurrentRole < 5.5 to the left, improve=0.3103360, (0 missing)  
## WorkLifeBalance splits as LRLR, improve=0.2375055, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve=0.1760362, (0 missing)  
##   
## Node number 13: 73 observations, complexity param=0.021875  
## predicted class=No expected loss=0.3424658 P(node) =0.0744898  
## class counts: 48 25  
## probabilities: 0.658 0.342   
## left son=26 (46 obs) right son=27 (27 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=7.066728, (0 missing)  
## Education splits as LRRRL, improve=1.639176, (0 missing)  
## WorkLifeBalance splits as RRRL, improve=1.259065, (0 missing)  
## YearsWithCurrManager < 10.5 to the right, improve=1.259065, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=1.215174, (0 missing)  
## Surrogate splits:  
## BusinessTravel splits as RLL, agree=0.644, adj=0.037, (0 split)  
## WorkLifeBalance splits as RLLL, agree=0.644, adj=0.037, (0 split)  
## YearsWithCurrManager < 1 to the right, agree=0.644, adj=0.037, (0 split)  
##   
## Node number 14: 85 observations, complexity param=0.01875  
## predicted class=No expected loss=0.4941176 P(node) =0.08673469  
## class counts: 43 42  
## probabilities: 0.506 0.494   
## left son=28 (10 obs) right son=29 (75 obs)  
## Primary splits:  
## Group splits as -RL, improve=1.9607840, (0 missing)  
## MaritalStatus splits as LLR, improve=1.7971480, (0 missing)  
## YearsInCurrentRole < 6.5 to the left, improve=1.0845940, (0 missing)  
## YearsWithCurrManager < 9.5 to the right, improve=1.0001420, (0 missing)  
## Department splits as LLR, improve=0.8320172, (0 missing)  
##   
## Node number 15: 15 observations, complexity param=0.00625  
## predicted class=Yes expected loss=0.1333333 P(node) =0.01530612  
## class counts: 2 13  
## probabilities: 0.133 0.867   
## left son=30 (3 obs) right son=31 (12 obs)  
## Primary splits:  
## Education splits as -LRRL, improve=2.1333330, (0 missing)  
## MaritalStatus splits as LRR, improve=0.6205128, (0 missing)  
## YearsWithCurrManager < 2.5 to the right, improve=0.6205128, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=0.2666667, (0 missing)  
## BusinessTravel splits as RRL, improve=0.1939394, (0 missing)  
## Surrogate splits:  
## Department splits as LRR, agree=0.867, adj=0.333, (0 split)  
##   
## Node number 18: 150 observations  
## predicted class=No expected loss=0.08 P(node) =0.1530612  
## class counts: 138 12  
## probabilities: 0.920 0.080   
##   
## Node number 19: 41 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.2195122 P(node) =0.04183673  
## class counts: 32 9  
## probabilities: 0.780 0.220   
## left son=38 (38 obs) right son=39 (3 obs)  
## Primary splits:  
## WorkLifeBalance splits as RLLL, improve=1.2943950, (0 missing)  
## Education splits as LLLRR, improve=0.4395851, (0 missing)  
## YearsWithCurrManager < 3.5 to the right, improve=0.3773519, (0 missing)  
## YearsInCurrentRole < 1.5 to the right, improve=0.3308318, (0 missing)  
## Group splits as RLL, improve=0.2960332, (0 missing)  
##   
## Node number 22: 69 observations, complexity param=0.00625  
## predicted class=No expected loss=0.3478261 P(node) =0.07040816  
## class counts: 45 24  
## probabilities: 0.652 0.348   
## left son=44 (38 obs) right son=45 (31 obs)  
## Primary splits:  
## WorkLifeBalance splits as RRLR, improve=3.1888980, (0 missing)  
## BusinessTravel splits as LRL, improve=2.9009580, (0 missing)  
## MaritalStatus splits as LLR, improve=2.0203140, (0 missing)  
## Education splits as LLLR-, improve=0.5416360, (0 missing)  
## Department splits as -LR, improve=0.3936335, (0 missing)  
## Surrogate splits:  
## Education splits as LLLR-, agree=0.580, adj=0.065, (0 split)  
## YearsWithCurrManager < 3 to the left, agree=0.565, adj=0.032, (0 split)  
##   
## Node number 23: 4 observations  
## predicted class=Yes expected loss=0 P(node) =0.004081633  
## class counts: 0 4  
## probabilities: 0.000 1.000   
##   
## Node number 24: 90 observations  
## predicted class=No expected loss=0.05555556 P(node) =0.09183673  
## class counts: 85 5  
## probabilities: 0.944 0.056   
##   
## Node number 25: 9 observations, complexity param=0.003125  
## predicted class=No expected loss=0.3333333 P(node) =0.009183673  
## class counts: 6 3  
## probabilities: 0.667 0.333   
## left son=50 (4 obs) right son=51 (5 obs)  
## Primary splits:  
## WorkLifeBalance splits as LLRR, improve=1.6000000, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve=1.6000000, (0 missing)  
## Group splits as R-L, improve=1.0000000, (0 missing)  
## JobLevel splits as -LRLL, improve=1.0000000, (0 missing)  
## MaritalStatus splits as LRR, improve=0.5714286, (0 missing)  
## Surrogate splits:  
## MaritalStatus splits as RLR, agree=0.778, adj=0.50, (0 split)  
## BusinessTravel splits as LRR, agree=0.667, adj=0.25, (0 split)  
## YearsInCurrentRole < 4 to the left, agree=0.667, adj=0.25, (0 split)  
##   
## Node number 26: 46 observations  
## predicted class=No expected loss=0.173913 P(node) =0.04693878  
## class counts: 38 8  
## probabilities: 0.826 0.174   
##   
## Node number 27: 27 observations  
## predicted class=Yes expected loss=0.3703704 P(node) =0.02755102  
## class counts: 10 17  
## probabilities: 0.370 0.630   
##   
## Node number 28: 10 observations  
## predicted class=No expected loss=0.2 P(node) =0.01020408  
## class counts: 8 2  
## probabilities: 0.800 0.200   
##   
## Node number 29: 75 observations, complexity param=0.01875  
## predicted class=Yes expected loss=0.4666667 P(node) =0.07653061  
## class counts: 35 40  
## probabilities: 0.467 0.533   
## left son=58 (50 obs) right son=59 (25 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=1.6133330, (0 missing)  
## Education splits as LRRRL, improve=1.1428570, (0 missing)  
## YearsInCurrentRole < 2.5 to the left, improve=1.0370370, (0 missing)  
## Department splits as LLR, improve=0.5079365, (0 missing)  
## YearsWithCurrManager < 6.5 to the left, improve=0.4129586, (0 missing)  
## Surrogate splits:  
## Education splits as RLLLL, agree=0.68, adj=0.04, (0 split)  
##   
## Node number 30: 3 observations  
## predicted class=No expected loss=0.3333333 P(node) =0.003061224  
## class counts: 2 1  
## probabilities: 0.667 0.333   
##   
## Node number 31: 12 observations  
## predicted class=Yes expected loss=0 P(node) =0.0122449  
## class counts: 0 12  
## probabilities: 0.000 1.000   
##   
## Node number 38: 38 observations  
## predicted class=No expected loss=0.1842105 P(node) =0.03877551  
## class counts: 31 7  
## probabilities: 0.816 0.184   
##   
## Node number 39: 3 observations  
## predicted class=Yes expected loss=0.3333333 P(node) =0.003061224  
## class counts: 1 2  
## probabilities: 0.333 0.667   
##   
## Node number 44: 38 observations  
## predicted class=No expected loss=0.2105263 P(node) =0.03877551  
## class counts: 30 8  
## probabilities: 0.789 0.211   
##   
## Node number 45: 31 observations  
## predicted class=Yes expected loss=0.483871 P(node) =0.03163265  
## class counts: 15 16  
## probabilities: 0.484 0.516   
##   
## Node number 50: 4 observations  
## predicted class=No expected loss=0 P(node) =0.004081633  
## class counts: 4 0  
## probabilities: 1.000 0.000   
##   
## Node number 51: 5 observations  
## predicted class=Yes expected loss=0.4 P(node) =0.005102041  
## class counts: 2 3  
## probabilities: 0.400 0.600   
##   
## Node number 58: 50 observations  
## predicted class=No expected loss=0.46 P(node) =0.05102041  
## class counts: 27 23  
## probabilities: 0.540 0.460   
##   
## Node number 59: 25 observations  
## predicted class=Yes expected loss=0.32 P(node) =0.0255102  
## class counts: 8 17  
## probabilities: 0.320 0.680



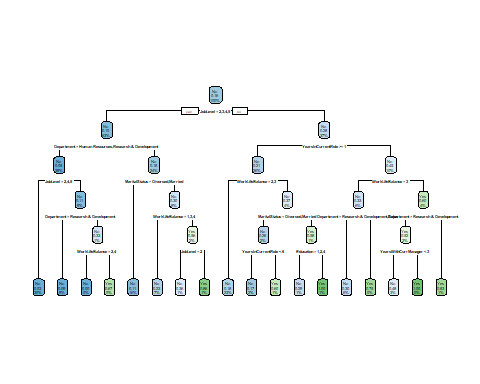
## No Yes   
## 443 47   
## actualAttrition  
## predictedAttrition No Yes  
## No 388 55  
## Yes 25 22

410/490 - worse off.

Removing income. Removing overtime because worklifebalance and travel and it are pretty intertwined according to previous analysis.

# Grouping  
reducedFields <- treeSpecific  
reducedFields$Group <- NULL  
reducedFields$OverTime <- NULL  
reducedFields$BusinessTravel<- NULL  
  
printDecision(seedNum1, reducedFields)

## Call:  
## rpart(formula = Attrition ~ ., data = train, method = "class",   
## control = rpart.control(cp = 0, minsplit = 5, maxdepth = depth))  
## n= 980   
##   
## CP nsplit rel error xerror xstd  
## 1 0.018750000 0 1.00000 1.00000 0.07231592  
## 2 0.012500000 3 0.94375 0.99375 0.07213352  
## 3 0.010416667 4 0.93125 1.03750 0.07338938  
## 4 0.007812500 7 0.90000 1.07500 0.07442809  
## 5 0.006250000 11 0.86875 1.07500 0.07442809  
## 6 0.003125000 12 0.86250 1.08750 0.07476686  
## 7 0.002083333 14 0.85625 1.08750 0.07476686  
## 8 0.000000000 17 0.85000 1.11250 0.07543348  
##   
## Variable importance  
## JobLevel YearsInCurrentRole WorkLifeBalance   
## 25 17 15   
## YearsWithCurrManager Department MaritalStatus   
## 15 14 9   
## Education   
## 6   
##   
## Node number 1: 980 observations, complexity param=0.01875  
## predicted class=No expected loss=0.1632653 P(node) =1  
## class counts: 820 160  
## probabilities: 0.837 0.163   
## left son=2 (622 obs) right son=3 (358 obs)  
## Primary splits:  
## JobLevel splits as RLLLL, improve=14.473920, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=11.862970, (0 missing)  
## MaritalStatus splits as LLR, improve= 8.850673, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve= 5.865829, (0 missing)  
## Department splits as RLR, improve= 3.279613, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 2.5 to the right, agree=0.680, adj=0.123, (0 split)  
## YearsInCurrentRole < 2.5 to the right, agree=0.663, adj=0.078, (0 split)  
## WorkLifeBalance splits as RLLL, agree=0.637, adj=0.006, (0 split)  
##   
## Node number 2: 622 observations, complexity param=0.0078125  
## predicted class=No expected loss=0.09807074 P(node) =0.6346939  
## class counts: 561 61  
## probabilities: 0.902 0.098   
## left son=4 (386 obs) right son=5 (236 obs)  
## Primary splits:  
## Department splits as LLR, improve=4.8549890, (0 missing)  
## MaritalStatus splits as LLR, improve=2.7558750, (0 missing)  
## JobLevel splits as -LRLL, improve=1.0246140, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=0.9661681, (0 missing)  
## WorkLifeBalance splits as LRLL, improve=0.7247592, (0 missing)  
## Surrogate splits:  
## YearsInCurrentRole < 16.5 to the left, agree=0.622, adj=0.004, (0 split)  
##   
## Node number 3: 358 observations, complexity param=0.01875  
## predicted class=No expected loss=0.2765363 P(node) =0.3653061  
## class counts: 259 99  
## probabilities: 0.723 0.277   
## left son=6 (257 obs) right son=7 (101 obs)  
## Primary splits:  
## YearsInCurrentRole < 0.5 to the right, improve=8.037427, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=5.378697, (0 missing)  
## WorkLifeBalance splits as RLLR, improve=5.002546, (0 missing)  
## MaritalStatus splits as LLR, improve=4.592086, (0 missing)  
## Department splits as RLR, improve=3.082292, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 0.5 to the right, agree=0.891, adj=0.614, (0 split)  
##   
## Node number 4: 386 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.0492228 P(node) =0.3938776  
## class counts: 367 19  
## probabilities: 0.951 0.049   
## left son=8 (294 obs) right son=9 (92 obs)  
## Primary splits:  
## JobLevel splits as -LRLL, improve=0.8544671, (0 missing)  
## YearsInCurrentRole < 4.5 to the left, improve=0.3025831, (0 missing)  
## Education splits as RLLRR, improve=0.2084054, (0 missing)  
## MaritalStatus splits as LRR, improve=0.1940921, (0 missing)  
## YearsWithCurrManager < 7.5 to the right, improve=0.1535140, (0 missing)  
##   
## Node number 5: 236 observations, complexity param=0.0078125  
## predicted class=No expected loss=0.1779661 P(node) =0.2408163  
## class counts: 194 42  
## probabilities: 0.822 0.178   
## left son=10 (153 obs) right son=11 (83 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=3.8888660, (0 missing)  
## YearsInCurrentRole < 0.5 to the right, improve=1.5273220, (0 missing)  
## WorkLifeBalance splits as RRLL, improve=1.2659990, (0 missing)  
## Education splits as RRRRL, improve=0.5245317, (0 missing)  
## JobLevel splits as -LRLR, improve=0.4370770, (0 missing)  
## Surrogate splits:  
## YearsInCurrentRole < 13.5 to the left, agree=0.653, adj=0.012, (0 split)  
##   
## Node number 6: 257 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.2101167 P(node) =0.2622449  
## class counts: 203 54  
## probabilities: 0.790 0.210   
## left son=12 (222 obs) right son=13 (35 obs)  
## Primary splits:  
## WorkLifeBalance splits as RLLR, improve=2.1086800, (0 missing)  
## MaritalStatus splits as LLR, improve=1.9223410, (0 missing)  
## YearsInCurrentRole < 6.5 to the left, improve=1.5070630, (0 missing)  
## Department splits as LLR, improve=0.9160886, (0 missing)  
## Education splits as LRLLL, improve=0.5616588, (0 missing)  
##   
## Node number 7: 101 observations, complexity param=0.01875  
## predicted class=No expected loss=0.4455446 P(node) =0.1030612  
## class counts: 56 45  
## probabilities: 0.554 0.446   
## left son=14 (58 obs) right son=15 (43 obs)  
## Primary splits:  
## WorkLifeBalance splits as RRLR, improve=3.7911260, (0 missing)  
## MaritalStatus splits as LLR, improve=2.3451690, (0 missing)  
## Department splits as RLL, improve=1.9185340, (0 missing)  
## Education splits as LLRRR, improve=1.2091870, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=0.3133278, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 3 to the left, agree=0.594, adj=0.047, (0 split)  
##   
## Node number 8: 294 observations  
## predicted class=No expected loss=0.03061224 P(node) =0.3  
## class counts: 285 9  
## probabilities: 0.969 0.031   
##   
## Node number 9: 92 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.1086957 P(node) =0.09387755  
## class counts: 82 10  
## probabilities: 0.891 0.109   
## left son=18 (86 obs) right son=19 (6 obs)  
## Primary splits:  
## Department splits as RL-, improve=0.6477924, (0 missing)  
## Education splits as RRLRR, improve=0.4876254, (0 missing)  
## YearsWithCurrManager < 9.5 to the right, improve=0.3901895, (0 missing)  
## YearsInCurrentRole < 9.5 to the right, improve=0.3577325, (0 missing)  
## MaritalStatus splits as LRR, improve=0.2917732, (0 missing)  
##   
## Node number 10: 153 observations  
## predicted class=No expected loss=0.1111111 P(node) =0.1561224  
## class counts: 136 17  
## probabilities: 0.889 0.111   
##   
## Node number 11: 83 observations, complexity param=0.0078125  
## predicted class=No expected loss=0.3012048 P(node) =0.08469388  
## class counts: 58 25  
## probabilities: 0.699 0.301   
## left son=22 (65 obs) right son=23 (18 obs)  
## Primary splits:  
## WorkLifeBalance splits as LRLL, improve=2.9739470, (0 missing)  
## Education splits as LRRLL, improve=1.2397590, (0 missing)  
## JobLevel splits as -LLRR, improve=0.6997590, (0 missing)  
## YearsInCurrentRole < 3.5 to the right, improve=0.6333263, (0 missing)  
## YearsWithCurrManager < 2.5 to the right, improve=0.1529844, (0 missing)  
##   
## Node number 12: 222 observations  
## predicted class=No expected loss=0.1846847 P(node) =0.2265306  
## class counts: 181 41  
## probabilities: 0.815 0.185   
##   
## Node number 13: 35 observations, complexity param=0.01041667  
## predicted class=No expected loss=0.3714286 P(node) =0.03571429  
## class counts: 22 13  
## probabilities: 0.629 0.371   
## left son=26 (23 obs) right son=27 (12 obs)  
## Primary splits:  
## MaritalStatus splits as LLR, improve=1.6399590, (0 missing)  
## Education splits as LRRRL, improve=1.6095240, (0 missing)  
## YearsInCurrentRole < 5.5 to the left, improve=1.2623970, (0 missing)  
## Department splits as LRR, improve=0.9053571, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=0.5720238, (0 missing)  
##   
## Node number 14: 58 observations, complexity param=0.0125  
## predicted class=No expected loss=0.3275862 P(node) =0.05918367  
## class counts: 39 19  
## probabilities: 0.672 0.328   
## left son=28 (54 obs) right son=29 (4 obs)  
## Primary splits:  
## Department splits as RLL, improve=1.5332060, (0 missing)  
## MaritalStatus splits as LRR, improve=1.5269180, (0 missing)  
## Education splits as LLRRR, improve=1.4305120, (0 missing)  
## YearsWithCurrManager < 0.5 to the right, improve=0.6587009, (0 missing)  
## Surrogate splits:  
## Education splits as LLLLR, agree=0.948, adj=0.25, (0 split)  
##   
## Node number 15: 43 observations, complexity param=0.003125  
## predicted class=Yes expected loss=0.3953488 P(node) =0.04387755  
## class counts: 17 26  
## probabilities: 0.395 0.605   
## left son=30 (31 obs) right son=31 (12 obs)  
## Primary splits:  
## Department splits as RLR, improve=1.7409350, (0 missing)  
## MaritalStatus splits as RLR, improve=0.6660941, (0 missing)  
## Education splits as RLLL-, improve=0.1863447, (0 missing)  
## YearsWithCurrManager < 3 to the left, improve=0.1863447, (0 missing)  
## WorkLifeBalance splits as RL-R, improve=0.1537917, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 5.5 to the left, agree=0.744, adj=0.083, (0 split)  
##   
## Node number 18: 86 observations  
## predicted class=No expected loss=0.09302326 P(node) =0.0877551  
## class counts: 78 8  
## probabilities: 0.907 0.093   
##   
## Node number 19: 6 observations, complexity param=0.002083333  
## predicted class=No expected loss=0.3333333 P(node) =0.006122449  
## class counts: 4 2  
## probabilities: 0.667 0.333   
## left son=38 (3 obs) right son=39 (3 obs)  
## Primary splits:  
## WorkLifeBalance splits as -LRL, improve=1.3333330, (0 missing)  
## Education splits as -RRL-, improve=0.6666667, (0 missing)  
## MaritalStatus splits as RRL, improve=0.6666667, (0 missing)  
## YearsWithCurrManager < 1 to the left, improve=0.6666667, (0 missing)  
## YearsInCurrentRole < 6.5 to the right, improve=0.6666667, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 6.5 to the right, agree=0.833, adj=0.667, (0 split)  
## Education splits as -RLR-, agree=0.667, adj=0.333, (0 split)  
## MaritalStatus splits as RLR, agree=0.667, adj=0.333, (0 split)  
## YearsInCurrentRole < 4.5 to the left, agree=0.667, adj=0.333, (0 split)  
##   
## Node number 22: 65 observations  
## predicted class=No expected loss=0.2307692 P(node) =0.06632653  
## class counts: 50 15  
## probabilities: 0.769 0.231   
##   
## Node number 23: 18 observations, complexity param=0.0078125  
## predicted class=Yes expected loss=0.4444444 P(node) =0.01836735  
## class counts: 8 10  
## probabilities: 0.444 0.556   
## left son=46 (11 obs) right son=47 (7 obs)  
## Primary splits:  
## JobLevel splits as -LRR-, improve=2.083694, (0 missing)  
## YearsWithCurrManager < 2.5 to the right, improve=2.031746, (0 missing)  
## Education splits as LLRL-, improve=1.088889, (0 missing)  
## YearsInCurrentRole < 3.5 to the right, improve=1.088889, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 1 to the right, agree=0.722, adj=0.286, (0 split)  
## YearsInCurrentRole < 0.5 to the right, agree=0.722, adj=0.286, (0 split)  
##   
## Node number 26: 23 observations, complexity param=0.00625  
## predicted class=No expected loss=0.2608696 P(node) =0.02346939  
## class counts: 17 6  
## probabilities: 0.739 0.261   
## left son=52 (18 obs) right son=53 (5 obs)  
## Primary splits:  
## YearsInCurrentRole < 5.5 to the left, improve=1.4695650, (0 missing)  
## MaritalStatus splits as LR-, improve=0.8695652, (0 missing)  
## Education splits as LRRRL, improve=0.6590389, (0 missing)  
## YearsWithCurrManager < 3.5 to the left, improve=0.5659938, (0 missing)  
## Department splits as LRL, improve=0.4695652, (0 missing)  
## Surrogate splits:  
## YearsWithCurrManager < 5.5 to the left, agree=0.87, adj=0.4, (0 split)  
##   
## Node number 27: 12 observations, complexity param=0.01041667  
## predicted class=Yes expected loss=0.4166667 P(node) =0.0122449  
## class counts: 5 7  
## probabilities: 0.417 0.583   
## left son=54 (7 obs) right son=55 (5 obs)  
## Primary splits:  
## Education splits as LLRL-, improve=2.97619000, (0 missing)  
## YearsWithCurrManager < 3 to the right, improve=0.50000000, (0 missing)  
## YearsInCurrentRole < 2.5 to the right, improve=0.16666670, (0 missing)  
## WorkLifeBalance splits as R--L, improve=0.08333333, (0 missing)  
## Surrogate splits:  
## Department splits as LLR, agree=0.667, adj=0.2, (0 split)  
##   
## Node number 28: 54 observations  
## predicted class=No expected loss=0.2962963 P(node) =0.05510204  
## class counts: 38 16  
## probabilities: 0.704 0.296   
##   
## Node number 29: 4 observations  
## predicted class=Yes expected loss=0.25 P(node) =0.004081633  
## class counts: 1 3  
## probabilities: 0.250 0.750   
##   
## Node number 30: 31 observations, complexity param=0.003125  
## predicted class=Yes expected loss=0.483871 P(node) =0.03163265  
## class counts: 15 16  
## probabilities: 0.484 0.516   
## left son=60 (29 obs) right son=61 (2 obs)  
## Primary splits:  
## YearsWithCurrManager < 3 to the left, improve=1.0011120, (0 missing)  
## MaritalStatus splits as RLR, improve=0.8475073, (0 missing)  
## WorkLifeBalance splits as RL-L, improve=0.5747801, (0 missing)  
## Education splits as RLRR-, improve=0.4295231, (0 missing)  
##   
## Node number 31: 12 observations  
## predicted class=Yes expected loss=0.1666667 P(node) =0.0122449  
## class counts: 2 10  
## probabilities: 0.167 0.833   
##   
## Node number 38: 3 observations  
## predicted class=No expected loss=0 P(node) =0.003061224  
## class counts: 3 0  
## probabilities: 1.000 0.000   
##   
## Node number 39: 3 observations  
## predicted class=Yes expected loss=0.3333333 P(node) =0.003061224  
## class counts: 1 2  
## probabilities: 0.333 0.667   
##   
## Node number 46: 11 observations  
## predicted class=No expected loss=0.3636364 P(node) =0.01122449  
## class counts: 7 4  
## probabilities: 0.636 0.364   
##   
## Node number 47: 7 observations  
## predicted class=Yes expected loss=0.1428571 P(node) =0.007142857  
## class counts: 1 6  
## probabilities: 0.143 0.857   
##   
## Node number 52: 18 observations  
## predicted class=No expected loss=0.1666667 P(node) =0.01836735  
## class counts: 15 3  
## probabilities: 0.833 0.167   
##   
## Node number 53: 5 observations  
## predicted class=Yes expected loss=0.4 P(node) =0.005102041  
## class counts: 2 3  
## probabilities: 0.400 0.600   
##   
## Node number 54: 7 observations  
## predicted class=No expected loss=0.2857143 P(node) =0.007142857  
## class counts: 5 2  
## probabilities: 0.714 0.286   
##   
## Node number 55: 5 observations  
## predicted class=Yes expected loss=0 P(node) =0.005102041  
## class counts: 0 5  
## probabilities: 0.000 1.000   
##   
## Node number 60: 29 observations  
## predicted class=No expected loss=0.4827586 P(node) =0.02959184  
## class counts: 15 14  
## probabilities: 0.517 0.483   
##   
## Node number 61: 2 observations  
## predicted class=Yes expected loss=0 P(node) =0.002040816  
## class counts: 0 2  
## probabilities: 0.000 1.000



## No Yes   
## 480 10   
## actualAttrition  
## predictedAttrition No Yes  
## No 406 74  
## Yes 7 3

409/490 Worse overall

Time for averaging across them

confusionTable <- function(seedNum, dataSet){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 decisionTree <- rpart(Attrition ~ ., data = train, method="class", control=rpart.control(cp=0, minsplit = 5, maxdepth = 5))  
 predicted <- predict(decisionTree, test, type="class")  
 set.seed(NULL)  
 return(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
}  
  
tableCalc <- function(table){  
 calcTable <- as.data.frame(table)  
 accuracy <- (calcTable[which(calcTable$predictedAttrition=="Yes" & calcTable$actualAttrition=="Yes"), 3] + calcTable[which(calcTable$predictedAttrition=="No" && calcTable$actualAttrition=="No"), 3])/sum(calcTable$Freq)  
 precisionYes <- (calcTable[which(calcTable$predictedAttrition=="Yes" & calcTable$actualAttrition=="Yes"), 3])/(calcTable[which(calcTable$predictedAttrition=="Yes" & calcTable$actualAttrition=="Yes"),3] + calcTable[which(calcTable$predictedAttrition=="Yes" & calcTable$actualAttrition=="No"),3])  
 precisionNo <- (calcTable[which(calcTable$predictedAttrition=="No" & calcTable$actualAttrition=="No"), 3])/(calcTable[which(calcTable$predictedAttrition=="No" & calcTable$actualAttrition=="Yes"),3] + calcTable[which(calcTable$predictedAttrition=="No" & calcTable$actualAttrition=="No"),3])  
 recallYes <- (calcTable[which(calcTable$predictedAttrition=="Yes" & calcTable$actualAttrition=="Yes"), 3])/(calcTable[which(calcTable$predictedAttrition=="Yes" & calcTable$actualAttrition=="Yes"),3] + calcTable[which(calcTable$predictedAttrition=="No" & calcTable$actualAttrition=="Yes"),3])  
 recallNo <- (calcTable[which(calcTable$predictedAttrition=="No" & calcTable$actualAttrition=="No"), 3])/(calcTable[which(calcTable$predictedAttrition=="Yes" & calcTable$actualAttrition=="No"),3] + calcTable[which(calcTable$predictedAttrition=="No" & calcTable$actualAttrition=="No"),3])  
 dataFrame <- data.frame(accuracy,precisionYes,precisionNo,recallYes,recallNo)  
 return(dataFrame)  
}  
  
averageTableCalc <- function(dataFrame){  
 avgAccuracy <- mean(dataFrame$accuracy)  
 avgPrecisionYes <- mean(dataFrame$precisionYes)  
 avgPrecisionNo <- mean(dataFrame$precisionNo)  
 avgRecallYes <- mean(dataFrame$recallYes)  
 avgRecallNo <- mean(dataFrame$recallNo)  
 newDF <- data.frame(avgAccuracy, avgPrecisionYes, avgPrecisionNo, avgRecallYes, avgRecallNo)  
 return(newDF)  
}  
  
completeTreeFunc <- function(dataSet){  
 treeTable1 <- confusionTable(seedNum1, dataSet)  
 treeTable2 <- confusionTable(seedNum2, dataSet)  
 treeTable3 <- confusionTable(seedNum3, dataSet)  
 treeTable4 <- confusionTable(seedNum4, dataSet)  
 treeTable5 <- confusionTable(seedNum5, dataSet)  
   
 treeTableCalc1 <- tableCalc(treeTable1)  
 treeTableCalc2 <- tableCalc(treeTable2)  
 treeTableCalc3 <- tableCalc(treeTable3)  
 treeTableCalc4 <- tableCalc(treeTable4)  
 treeTableCalc5 <- tableCalc(treeTable5)  
   
 treeTableCalc <- data.frame(rbind(as.matrix(treeTableCalc1),as.matrix(treeTableCalc2),as.matrix(treeTableCalc3),as.matrix(treeTableCalc4),as.matrix(treeTableCalc5)))  
   
 avgTreeCalc <- averageTableCalc(treeTableCalc)  
 print(avgTreeCalc)  
}

hrTree <- completeTreeFunc(HR\_tree)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.824898 0.4087265 0.8681532 0.241628 0.934175

hrTree$type <- "decisionTrees\_hrTree"  
hrTreeSpecific <- completeTreeFunc(treeSpecific)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8416327 0.4991805 0.8788818 0.3053566 0.9419136

hrTreeSpecific$type <- "decisionTrees\_treeSpecific"  
hrTreeIncome <- completeTreeFunc(treeIncome)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.837551 0.4811203 0.8776329 0.3002276 0.9380277

hrTreeIncome$type <- "decisionTrees\_treeIncome"  
hrTreeReduced <- completeTreeFunc(reducedFields)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8261224 0.3762607 0.854591 0.1302459 0.9570162

hrTreeReduced$type <- "decisionTrees\_treeReduced"  
completeModels <- rbind(hrTree, hrTreeSpecific, hrTreeIncome, hrTreeReduced)  
completeModels

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8248980 0.4087265 0.8681532 0.2416280 0.9341750  
## 2 0.8416327 0.4991805 0.8788818 0.3053566 0.9419136  
## 3 0.8375510 0.4811203 0.8776329 0.3002276 0.9380277  
## 4 0.8261224 0.3762607 0.8545910 0.1302459 0.9570162  
## type  
## 1 decisionTrees\_hrTree  
## 2 decisionTrees\_treeSpecific  
## 3 decisionTrees\_treeIncome  
## 4 decisionTrees\_treeReduced

#### Support Vector Machines

if("kernlab" %in% rownames(installed.packages()) == FALSE) {install.packages('kernlab') }  
if("e1071" %in% rownames(installed.packages()) == FALSE) {install.packages('e1071') }  
library(kernlab)

##   
## Attaching package: 'kernlab'

## The following object is masked from 'package:arules':  
##   
## size

## The following object is masked from 'package:ggplot2':  
##   
## alpha

## The following object is masked from 'package:purrr':  
##   
## cross

library(e1071)  
  
printSVM <- function(seedNum, dataSet, kernelType="radial", cost=1){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 svmModel <- svm(Attrition ~ ., data = train, kernel=kernelType, cost=cost)  
 # Predictions  
 predicted <- predict(svmModel, test, type="votes")  
 print(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
 set.seed(NULL)  
}

kernelName <- "radial"  
print(kernelName)

## [1] "radial"

dataFrame <- HR\_tree  
printSVM(seedNum1, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 76  
## Yes 0 1

printSVM(seedNum1, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 66  
## Yes 0 3

printSVM(seedNum2, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 78  
## Yes 0 3

printSVM(seedNum3, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

kernelName <- "sigmoid"  
print(kernelName)

## [1] "sigmoid"

dataFrame <- HR\_tree  
printSVM(seedNum1, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

kernelName <- "polynomial"  
print(kernelName)

## [1] "polynomial"

dataFrame <- HR\_tree  
printSVM(seedNum1, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

kernelName <- "linear"  
print(kernelName)

## [1] "linear"

dataFrame <- HR\_tree  
printSVM(seedNum1, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 390 37  
## Yes 23 40

printSVM(seedNum1, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 391 37  
## Yes 22 40

printSVM(seedNum1, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 393 39  
## Yes 20 38

printSVM(seedNum1, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 397 42  
## Yes 16 35

printSVM(seedNum1, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 405 45  
## Yes 8 32

printSVM(seedNum2, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 401 29  
## Yes 20 40

printSVM(seedNum2, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 403 30  
## Yes 18 39

printSVM(seedNum2, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 32  
## Yes 12 37

printSVM(seedNum2, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 33  
## Yes 12 36

printSVM(seedNum2, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 415 44  
## Yes 6 25

printSVM(seedNum3, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 386 44  
## Yes 23 37

printSVM(seedNum3, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 389 43  
## Yes 20 38

printSVM(seedNum3, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 387 45  
## Yes 22 36

printSVM(seedNum3, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 394 46  
## Yes 15 35

printSVM(seedNum3, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 402 58  
## Yes 7 23

kernelName <- "sigmoid"  
print(kernelName)

## [1] "sigmoid"

print("treeSpecific")

## [1] "treeSpecific"

dataFrame <- treeSpecific  
printSVM(seedNum1, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum1, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 413 77  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum2, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 421 69  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.7)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.5)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.3)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

printSVM(seedNum3, dataFrame, kernelName, cost=.1)

## actualAttrition  
## predictedAttrition No Yes  
## No 409 81  
## Yes 0 0

Based on our tests, SVM does not seem like a potential as it will always guess no unless it’s linear and has all the parameters in place.

confusionTableSVM <- function(seedNum, dataSet){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 algorithm <- svm(Attrition ~ ., data = train, kernel="linear", cost=.5)  
 predicted <- predict(algorithm, test, type="class")  
 set.seed(NULL)  
 return(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
}  
  
  
completeSVMFunc <- function(dataSet){  
 table1 <- confusionTableSVM(seedNum1, dataSet)  
 table2 <- confusionTableSVM(seedNum2, dataSet)  
 table3 <- confusionTableSVM(seedNum3, dataSet)  
 table4 <- confusionTableSVM(seedNum4, dataSet)  
 table5 <- confusionTableSVM(seedNum5, dataSet)  
   
 tableCalc1 <- tableCalc(table1)  
 tableCalc2 <- tableCalc(table2)  
 tableCalc3 <- tableCalc(table3)  
 tableCalc4 <- tableCalc(table4)  
 tableCalc5 <- tableCalc(table5)  
   
 tableCalc <- data.frame(rbind(as.matrix(tableCalc1),as.matrix(tableCalc2),as.matrix(tableCalc3),as.matrix(tableCalc4),as.matrix(tableCalc5)))  
   
 avgTableCalc <- averageTableCalc(tableCalc)  
 print(avgTableCalc)  
}

hrSVM <- completeSVMFunc(HR\_tree)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8861224 0.7035583 0.909154 0.4885391 0.9607056

hrSVM$type <- "svm\_hrTree"  
completeModels <- rbind(completeModels, hrSVM)  
completeModels

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8248980 0.4087265 0.8681532 0.2416280 0.9341750  
## 2 0.8416327 0.4991805 0.8788818 0.3053566 0.9419136  
## 3 0.8375510 0.4811203 0.8776329 0.3002276 0.9380277  
## 4 0.8261224 0.3762607 0.8545910 0.1302459 0.9570162  
## 5 0.8861224 0.7035583 0.9091540 0.4885391 0.9607056  
## type  
## 1 decisionTrees\_hrTree  
## 2 decisionTrees\_treeSpecific  
## 3 decisionTrees\_treeIncome  
## 4 decisionTrees\_treeReduced  
## 5 svm\_hrTree

#### Naive Bayes

#  
printNB <- function(seedNum, dataSet, laplaceNum=1){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 model=naiveBayes(Attrition~., data = train, laplace = laplaceNum, na.action = na.pass)  
 # Predictions  
 predicted <- predict(model, test)  
 print(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
 set.seed(NULL)  
}

printNB(seedNum1, HR\_tree)

## actualAttrition  
## predictedAttrition No Yes  
## No 342 26  
## Yes 71 51

printNB(seedNum1, HR\_tree, 2)

## actualAttrition  
## predictedAttrition No Yes  
## No 343 25  
## Yes 70 52

printNB(seedNum1, HR\_tree, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 347 26  
## Yes 66 51

printNB(seedNum1, HR\_tree, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 352 27  
## Yes 61 50

printNB(seedNum1, HR\_tree, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 358 29  
## Yes 55 48

printNB(seedNum2, HR\_tree)

## actualAttrition  
## predictedAttrition No Yes  
## No 335 20  
## Yes 86 49

printNB(seedNum2, HR\_tree, 2)

## actualAttrition  
## predictedAttrition No Yes  
## No 334 20  
## Yes 87 49

printNB(seedNum2, HR\_tree, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 338 22  
## Yes 83 47

printNB(seedNum2, HR\_tree, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 343 24  
## Yes 78 45

printNB(seedNum2, HR\_tree, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 353 27  
## Yes 68 42

printNB(seedNum3, HR\_tree)

## actualAttrition  
## predictedAttrition No Yes  
## No 345 38  
## Yes 64 43

printNB(seedNum3, HR\_tree, 2)

## actualAttrition  
## predictedAttrition No Yes  
## No 346 37  
## Yes 63 44

printNB(seedNum3, HR\_tree, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 350 37  
## Yes 59 44

printNB(seedNum3, HR\_tree, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 357 43  
## Yes 52 38

printNB(seedNum3, HR\_tree, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 362 47  
## Yes 47 34

printNB(seedNum1, treeSpecific)

## actualAttrition  
## predictedAttrition No Yes  
## No 396 50  
## Yes 17 27

printNB(seedNum1, treeSpecific, 2)

## actualAttrition  
## predictedAttrition No Yes  
## No 396 51  
## Yes 17 26

printNB(seedNum1, treeSpecific, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 396 52  
## Yes 17 25

printNB(seedNum1, treeSpecific, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 399 54  
## Yes 14 23

printNB(seedNum1, treeSpecific, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 401 58  
## Yes 12 19

printNB(seedNum2, treeSpecific)

## actualAttrition  
## predictedAttrition No Yes  
## No 399 43  
## Yes 22 26

printNB(seedNum2, treeSpecific, 2)

## actualAttrition  
## predictedAttrition No Yes  
## No 400 45  
## Yes 21 24

printNB(seedNum2, treeSpecific, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 399 44  
## Yes 22 25

printNB(seedNum2, treeSpecific, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 402 46  
## Yes 19 23

printNB(seedNum2, treeSpecific, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 405 49  
## Yes 16 20

printNB(seedNum3, treeSpecific)

## actualAttrition  
## predictedAttrition No Yes  
## No 389 57  
## Yes 20 24

printNB(seedNum3, treeSpecific, 2)

## actualAttrition  
## predictedAttrition No Yes  
## No 392 59  
## Yes 17 22

printNB(seedNum3, treeSpecific, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 394 60  
## Yes 15 21

printNB(seedNum3, treeSpecific, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 399 60  
## Yes 10 21

printNB(seedNum3, treeSpecific, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 400 62  
## Yes 9 19

confusionTableNB <- function(seedNum, dataSet, laplaceNum=1){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 algorithm <- naiveBayes(Attrition~., data = train, laplace = laplaceNum, na.action = na.pass)  
 predicted <- predict(algorithm, test, type="class")  
 set.seed(NULL)  
 return(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
}  
  
  
completeNBFunc <- function(dataSet, laplaceNum=1){  
 table1 <- confusionTableNB(seedNum1, dataSet, laplaceNum)  
 table2 <- confusionTableNB(seedNum2, dataSet, laplaceNum)  
 table3 <- confusionTableNB(seedNum3, dataSet, laplaceNum)  
 table4 <- confusionTableNB(seedNum4, dataSet, laplaceNum)  
 table5 <- confusionTableNB(seedNum5, dataSet, laplaceNum)  
   
 tableCalc1 <- tableCalc(table1)  
 tableCalc2 <- tableCalc(table2)  
 tableCalc3 <- tableCalc(table3)  
 tableCalc4 <- tableCalc(table4)  
 tableCalc5 <- tableCalc(table5)  
   
 tableCalc <- data.frame(rbind(as.matrix(tableCalc1),as.matrix(tableCalc2),as.matrix(tableCalc3),as.matrix(tableCalc4),as.matrix(tableCalc5)))  
   
 avgTableCalc <- averageTableCalc(tableCalc)  
 print(avgTableCalc)  
}

nbModel <- completeNBFunc(HR\_tree, 1)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8028571 0.4203825 0.9265368 0.64722 0.8325573

nbModel$type <- "nb\_hrTree"  
completeModels <- rbind(completeModels, nbModel)  
completeModels

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8248980 0.4087265 0.8681532 0.2416280 0.9341750  
## 2 0.8416327 0.4991805 0.8788818 0.3053566 0.9419136  
## 3 0.8375510 0.4811203 0.8776329 0.3002276 0.9380277  
## 4 0.8261224 0.3762607 0.8545910 0.1302459 0.9570162  
## 5 0.8861224 0.7035583 0.9091540 0.4885391 0.9607056  
## 6 0.8028571 0.4203825 0.9265368 0.6472200 0.8325573  
## type  
## 1 decisionTrees\_hrTree  
## 2 decisionTrees\_treeSpecific  
## 3 decisionTrees\_treeIncome  
## 4 decisionTrees\_treeReduced  
## 5 svm\_hrTree  
## 6 nb\_hrTree

#### Random Forest

if("randomForest" %in% rownames(installed.packages()) == FALSE) {install.packages('randomForest') }  
library(randomForest)

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:rattle':  
##   
## importance

## The following object is masked from 'package:gridExtra':  
##   
## combine

## The following object is masked from 'package:ggplot2':  
##   
## margin

## The following object is masked from 'package:dplyr':  
##   
## combine

printRF <- function(seedNum, dataSet, trees=3){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 model=randomForest(Attrition~., data = train, ntree=trees)  
 # Predictions  
 predicted <- predict(model, test, type=c("class"))  
 print(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
 set.seed(NULL)  
}

printRF(seedNum1, HR\_tree)

## actualAttrition  
## predictedAttrition No Yes  
## No 364 54  
## Yes 49 23

printRF(seedNum1, HR\_tree, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 389 60  
## Yes 24 17

printRF(seedNum1, HR\_tree, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 399 60  
## Yes 14 17

printRF(seedNum1, HR\_tree, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 408 62  
## Yes 5 15

printRF(seedNum1, HR\_tree, 25)

## actualAttrition  
## predictedAttrition No Yes  
## No 404 64  
## Yes 9 13

printRF(seedNum2, HR\_tree)

## actualAttrition  
## predictedAttrition No Yes  
## No 394 52  
## Yes 27 17

printRF(seedNum2, HR\_tree, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 402 54  
## Yes 19 15

printRF(seedNum2, HR\_tree, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 410 55  
## Yes 11 14

printRF(seedNum2, HR\_tree, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 411 55  
## Yes 10 14

printRF(seedNum2, HR\_tree, 25)

## actualAttrition  
## predictedAttrition No Yes  
## No 415 56  
## Yes 6 13

printRF(seedNum3, HR\_tree)

## actualAttrition  
## predictedAttrition No Yes  
## No 375 64  
## Yes 34 17

printRF(seedNum3, HR\_tree, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 390 67  
## Yes 19 14

printRF(seedNum3, HR\_tree, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 401 70  
## Yes 8 11

printRF(seedNum3, HR\_tree, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 401 72  
## Yes 8 9

printRF(seedNum3, HR\_tree, 25)

## actualAttrition  
## predictedAttrition No Yes  
## No 403 68  
## Yes 6 13

printRF(seedNum1, treeSpecific)

## actualAttrition  
## predictedAttrition No Yes  
## No 382 56  
## Yes 31 21

printRF(seedNum1, treeSpecific, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 381 54  
## Yes 32 23

printRF(seedNum1, treeSpecific, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 388 58  
## Yes 25 19

printRF(seedNum1, treeSpecific, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 388 57  
## Yes 25 20

printRF(seedNum1, treeSpecific, 25)

## actualAttrition  
## predictedAttrition No Yes  
## No 385 58  
## Yes 28 19

printRF(seedNum2, treeSpecific)

## actualAttrition  
## predictedAttrition No Yes  
## No 377 46  
## Yes 44 23

printRF(seedNum2, treeSpecific, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 386 43  
## Yes 35 26

printRF(seedNum2, treeSpecific, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 395 49  
## Yes 26 20

printRF(seedNum2, treeSpecific, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 403 51  
## Yes 18 18

printRF(seedNum2, treeSpecific, 25)

## actualAttrition  
## predictedAttrition No Yes  
## No 403 51  
## Yes 18 18

printRF(seedNum3, treeSpecific)

## actualAttrition  
## predictedAttrition No Yes  
## No 385 61  
## Yes 24 20

printRF(seedNum3, treeSpecific, 5)

## actualAttrition  
## predictedAttrition No Yes  
## No 391 63  
## Yes 18 18

printRF(seedNum3, treeSpecific, 10)

## actualAttrition  
## predictedAttrition No Yes  
## No 397 66  
## Yes 12 15

printRF(seedNum3, treeSpecific, 15)

## actualAttrition  
## predictedAttrition No Yes  
## No 399 66  
## Yes 10 15

printRF(seedNum3, treeSpecific, 25)

## actualAttrition  
## predictedAttrition No Yes  
## No 399 63  
## Yes 10 18

confusionTableRF <- function(seedNum, dataSet, ntrees=3){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 train <- dataSet[randIndex[1:cutPoint],]  
 test <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 algorithm <- randomForest(Attrition~., data = train, ntree=ntrees)  
 print("importance")  
 print(importance(algorithm))  
 predicted <- predict(algorithm, test, type="class")  
 set.seed(NULL)  
 return(table(predictedAttrition=predicted, actualAttrition=test$Attrition))  
}  
  
  
completeRFFunc <- function(dataSet, ntrees=3){  
 table1 <- confusionTableRF(seedNum1, dataSet, ntrees)  
 table2 <- confusionTableRF(seedNum2, dataSet, ntrees)  
 table3 <- confusionTableRF(seedNum3, dataSet, ntrees)  
 table4 <- confusionTableRF(seedNum4, dataSet, ntrees)  
 table5 <- confusionTableRF(seedNum5, dataSet, ntrees)  
   
 tableCalc1 <- tableCalc(table1)  
 tableCalc2 <- tableCalc(table2)  
 tableCalc3 <- tableCalc(table3)  
 tableCalc4 <- tableCalc(table4)  
 tableCalc5 <- tableCalc(table5)  
   
 tableCalc <- data.frame(rbind(as.matrix(tableCalc1),as.matrix(tableCalc2),as.matrix(tableCalc3),as.matrix(tableCalc4),as.matrix(tableCalc5)))  
   
 avgTableCalc <- averageTableCalc(tableCalc)  
 print(avgTableCalc)  
}

rfHRTree <- completeRFFunc(HR\_tree, 3)

## [1] "importance"  
## MeanDecreaseGini  
## Age 16.568799  
## BusinessTravel 2.696075  
## DailyRate 12.403092  
## Department 7.567464  
## DistanceFromHome 13.754774  
## Education 4.579637  
## EducationField 10.214950  
## EnvironmentSatisfaction 5.912594  
## Gender 4.442336  
## HourlyRate 9.581976  
## JobInvolvement 6.841059  
## JobLevel 8.990705  
## JobRole 20.140894  
## JobSatisfaction 10.050587  
## MaritalStatus 1.235543  
## MonthlyIncome 17.683327  
## MonthlyRate 16.555428  
## NumCompaniesWorked 10.595557  
## OverTime 7.404556  
## PercentSalaryHike 6.850607  
## PerformanceRating 0.000000  
## RelationshipSatisfaction 5.843753  
## StockOptionLevel 16.441079  
## TotalWorkingYears 14.307302  
## TrainingTimesLastYear 9.069804  
## WorkLifeBalance 6.533326  
## YearsAtCompany 6.351438  
## YearsInCurrentRole 4.529337  
## YearsSinceLastPromotion 11.645180  
## YearsWithCurrManager 10.293177  
## [1] "importance"  
## MeanDecreaseGini  
## Age 11.7263153  
## BusinessTravel 4.1106387  
## DailyRate 14.3078637  
## Department 1.2158730  
## DistanceFromHome 9.3321644  
## Education 9.2597307  
## EducationField 14.3100424  
## EnvironmentSatisfaction 12.4578582  
## Gender 3.5752070  
## HourlyRate 14.6616177  
## JobInvolvement 5.3768469  
## JobLevel 5.9047992  
## JobRole 20.7384974  
## JobSatisfaction 4.6674494  
## MaritalStatus 4.4290591  
## MonthlyIncome 24.3333470  
## MonthlyRate 17.1750674  
## NumCompaniesWorked 6.8788692  
## OverTime 10.5522002  
## PercentSalaryHike 8.8600769  
## PerformanceRating 0.7700747  
## RelationshipSatisfaction 3.9585138  
## StockOptionLevel 7.4752438  
## TotalWorkingYears 12.8475909  
## TrainingTimesLastYear 6.3165043  
## WorkLifeBalance 6.4502592  
## YearsAtCompany 7.5261923  
## YearsInCurrentRole 5.9446659  
## YearsSinceLastPromotion 7.4908263  
## YearsWithCurrManager 12.5949043  
## [1] "importance"  
## MeanDecreaseGini  
## Age 16.04689192  
## BusinessTravel 7.01848517  
## DailyRate 10.36272199  
## Department 1.98789119  
## DistanceFromHome 18.97155705  
## Education 4.31341122  
## EducationField 6.42116867  
## EnvironmentSatisfaction 6.15496031  
## Gender 1.56620745  
## HourlyRate 13.99333342  
## JobInvolvement 6.01012883  
## JobLevel 5.61049475  
## JobRole 14.35717305  
## JobSatisfaction 8.27183405  
## MaritalStatus 1.49415458  
## MonthlyIncome 21.92780395  
## MonthlyRate 12.02456539  
## NumCompaniesWorked 11.45452529  
## OverTime 3.28869814  
## PercentSalaryHike 7.75223970  
## PerformanceRating 0.07936508  
## RelationshipSatisfaction 3.54316482  
## StockOptionLevel 6.26159150  
## TotalWorkingYears 11.12378284  
## TrainingTimesLastYear 7.12836021  
## WorkLifeBalance 6.64046461  
## YearsAtCompany 13.32241659  
## YearsInCurrentRole 7.56954604  
## YearsSinceLastPromotion 12.54499513  
## YearsWithCurrManager 2.87371332  
## [1] "importance"  
## MeanDecreaseGini  
## Age 12.34253928  
## BusinessTravel 3.30264180  
## DailyRate 16.34509940  
## Department 3.11950211  
## DistanceFromHome 13.90507421  
## Education 3.62987640  
## EducationField 4.41968075  
## EnvironmentSatisfaction 14.91680303  
## Gender 0.03809524  
## HourlyRate 13.92881372  
## JobInvolvement 7.93384085  
## JobLevel 4.81594825  
## JobRole 16.11695088  
## JobSatisfaction 6.73711258  
## MaritalStatus 7.75076680  
## MonthlyIncome 20.92410287  
## MonthlyRate 9.26220354  
## NumCompaniesWorked 4.04451438  
## OverTime 9.05690825  
## PercentSalaryHike 8.17191233  
## PerformanceRating 0.12698413  
## RelationshipSatisfaction 11.11352710  
## StockOptionLevel 6.70129475  
## TotalWorkingYears 10.84291147  
## TrainingTimesLastYear 4.42474142  
## WorkLifeBalance 8.56512834  
## YearsAtCompany 11.55165349  
## YearsInCurrentRole 10.43712365  
## YearsSinceLastPromotion 4.24018317  
## YearsWithCurrManager 6.68349893  
## [1] "importance"  
## MeanDecreaseGini  
## Age 14.5643840  
## BusinessTravel 1.5461245  
## DailyRate 12.2219322  
## Department 0.0000000  
## DistanceFromHome 7.2396125  
## Education 9.8421356  
## EducationField 8.2186435  
## EnvironmentSatisfaction 15.1978127  
## Gender 1.5000000  
## HourlyRate 12.8973586  
## JobInvolvement 5.7103782  
## JobLevel 10.7207703  
## JobRole 11.7944146  
## JobSatisfaction 7.4854208  
## MaritalStatus 2.7460696  
## MonthlyIncome 24.8099663  
## MonthlyRate 20.6101852  
## NumCompaniesWorked 5.1629693  
## OverTime 18.7415274  
## PercentSalaryHike 11.8632290  
## PerformanceRating 0.1978579  
## RelationshipSatisfaction 3.1258850  
## StockOptionLevel 5.7941290  
## TotalWorkingYears 10.9165377  
## TrainingTimesLastYear 8.2498290  
## WorkLifeBalance 4.8099305  
## YearsAtCompany 12.1553719  
## YearsInCurrentRole 1.8983177  
## YearsSinceLastPromotion 7.1892690  
## YearsWithCurrManager 12.4137475  
## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8183673 0.3966877 0.8724989 0.2817504 0.9185592

rfHRTree$type <- "rf\_hrTree\_3trees"  
rfTreeSpecific <- completeRFFunc(treeSpecific, 3)

## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 17.76600  
## Department 14.87304  
## Education 32.23117  
## JobLevel 24.64975  
## MaritalStatus 19.18004  
## OverTime 20.76941  
## WorkLifeBalance 21.08525  
## YearsWithCurrManager 31.64634  
## YearsInCurrentRole 44.82100  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 12.86621  
## Department 22.72904  
## Education 28.25008  
## JobLevel 30.47064  
## MaritalStatus 23.73631  
## OverTime 19.75361  
## WorkLifeBalance 28.37204  
## YearsWithCurrManager 30.73328  
## YearsInCurrentRole 38.86750  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 15.19219  
## Department 12.50108  
## Education 27.74661  
## JobLevel 26.34851  
## MaritalStatus 23.18696  
## OverTime 13.26965  
## WorkLifeBalance 25.99548  
## YearsWithCurrManager 31.29155  
## YearsInCurrentRole 34.34350  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 18.55807  
## Department 16.71112  
## Education 32.68188  
## JobLevel 19.98063  
## MaritalStatus 20.82644  
## OverTime 26.97002  
## WorkLifeBalance 30.70121  
## YearsWithCurrManager 29.95286  
## YearsInCurrentRole 35.16876  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 17.33322  
## Department 15.83450  
## Education 27.72673  
## JobLevel 26.33782  
## MaritalStatus 20.63323  
## OverTime 26.07227  
## WorkLifeBalance 21.36755  
## YearsWithCurrManager 34.73185  
## YearsInCurrentRole 40.48304  
## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8228571 0.4147345 0.8744216 0.2915126 0.9226324

rfTreeSpecific$type <- "rf\_treeSpecific\_3trees"  
rfHRTree10 <- completeRFFunc(HR\_tree, 10)

## [1] "importance"  
## MeanDecreaseGini  
## Age 16.5403933  
## BusinessTravel 4.7102330  
## DailyRate 15.5138540  
## Department 4.7718993  
## DistanceFromHome 14.5779545  
## Education 6.3243040  
## EducationField 8.8794899  
## EnvironmentSatisfaction 6.7814061  
## Gender 2.2519351  
## HourlyRate 10.7240732  
## JobInvolvement 6.7658349  
## JobLevel 9.9785587  
## JobRole 13.3710957  
## JobSatisfaction 7.4372504  
## MaritalStatus 4.5220750  
## MonthlyIncome 19.2312246  
## MonthlyRate 15.2204949  
## NumCompaniesWorked 8.8961307  
## OverTime 13.9745914  
## PercentSalaryHike 7.7093761  
## PerformanceRating 0.6097619  
## RelationshipSatisfaction 8.5816831  
## StockOptionLevel 11.0919229  
## TotalWorkingYears 14.2730348  
## TrainingTimesLastYear 5.6961960  
## WorkLifeBalance 6.4578316  
## YearsAtCompany 9.1778205  
## YearsInCurrentRole 5.6890665  
## YearsSinceLastPromotion 7.8188743  
## YearsWithCurrManager 9.7781643  
## [1] "importance"  
## MeanDecreaseGini  
## Age 17.1772612  
## BusinessTravel 4.5315064  
## DailyRate 16.2595866  
## Department 2.2119461  
## DistanceFromHome 13.1720059  
## Education 6.5925728  
## EducationField 11.3064355  
## EnvironmentSatisfaction 10.6046331  
## Gender 1.6892865  
## HourlyRate 13.1867917  
## JobInvolvement 4.7762763  
## JobLevel 6.6344146  
## JobRole 17.3231178  
## JobSatisfaction 6.4728153  
## MaritalStatus 3.9168425  
## MonthlyIncome 24.7567405  
## MonthlyRate 11.9693344  
## NumCompaniesWorked 8.2283481  
## OverTime 12.0722438  
## PercentSalaryHike 8.4406718  
## PerformanceRating 0.5243557  
## RelationshipSatisfaction 7.0722604  
## StockOptionLevel 7.1232444  
## TotalWorkingYears 15.8798443  
## TrainingTimesLastYear 4.9140072  
## WorkLifeBalance 6.5656841  
## YearsAtCompany 9.4622514  
## YearsInCurrentRole 6.1425143  
## YearsSinceLastPromotion 7.3090949  
## YearsWithCurrManager 10.1132998  
## [1] "importance"  
## MeanDecreaseGini  
## Age 17.2579988  
## BusinessTravel 6.3288357  
## DailyRate 10.4848487  
## Department 2.1203995  
## DistanceFromHome 13.4980425  
## Education 5.0222917  
## EducationField 7.2939161  
## EnvironmentSatisfaction 8.0386141  
## Gender 1.9560830  
## HourlyRate 14.7360384  
## JobInvolvement 7.0157361  
## JobLevel 8.5609877  
## JobRole 14.7462543  
## JobSatisfaction 7.7943626  
## MaritalStatus 3.9749808  
## MonthlyIncome 16.7518281  
## MonthlyRate 9.2107456  
## NumCompaniesWorked 9.3696712  
## OverTime 7.4131218  
## PercentSalaryHike 10.4155521  
## PerformanceRating 0.5770219  
## RelationshipSatisfaction 7.5873091  
## StockOptionLevel 8.9605591  
## TotalWorkingYears 15.4392486  
## TrainingTimesLastYear 6.9614919  
## WorkLifeBalance 8.1130977  
## YearsAtCompany 11.2374127  
## YearsInCurrentRole 5.6150011  
## YearsSinceLastPromotion 9.5105083  
## YearsWithCurrManager 4.5204897  
## [1] "importance"  
## MeanDecreaseGini  
## Age 13.8325641  
## BusinessTravel 3.7741476  
## DailyRate 13.1973467  
## Department 4.0886183  
## DistanceFromHome 14.4819046  
## Education 5.7283710  
## EducationField 7.7516129  
## EnvironmentSatisfaction 10.1679544  
## Gender 0.9197382  
## HourlyRate 12.1775511  
## JobInvolvement 7.8514736  
## JobLevel 5.9978745  
## JobRole 11.3680906  
## JobSatisfaction 10.3760315  
## MaritalStatus 7.5522538  
## MonthlyIncome 19.7842237  
## MonthlyRate 11.5097174  
## NumCompaniesWorked 4.5493866  
## OverTime 10.1605613  
## PercentSalaryHike 8.0901171  
## PerformanceRating 0.5716191  
## RelationshipSatisfaction 8.2556670  
## StockOptionLevel 5.2030379  
## TotalWorkingYears 13.4063756  
## TrainingTimesLastYear 6.5276390  
## WorkLifeBalance 7.5618281  
## YearsAtCompany 16.2058505  
## YearsInCurrentRole 8.9985792  
## YearsSinceLastPromotion 6.7628264  
## YearsWithCurrManager 6.0208475  
## [1] "importance"  
## MeanDecreaseGini  
## Age 17.3005806  
## BusinessTravel 4.5944768  
## DailyRate 14.1671338  
## Department 0.5360548  
## DistanceFromHome 10.7226878  
## Education 7.9319259  
## EducationField 8.3590199  
## EnvironmentSatisfaction 9.9297045  
## Gender 1.8477557  
## HourlyRate 10.7212129  
## JobInvolvement 6.0040747  
## JobLevel 6.3931463  
## JobRole 10.5437434  
## JobSatisfaction 9.3180904  
## MaritalStatus 5.0588837  
## MonthlyIncome 20.4950693  
## MonthlyRate 14.1362922  
## NumCompaniesWorked 7.6267218  
## OverTime 13.8482199  
## PercentSalaryHike 9.9759633  
## PerformanceRating 0.3175991  
## RelationshipSatisfaction 3.5898563  
## StockOptionLevel 6.3805949  
## TotalWorkingYears 15.1047317  
## TrainingTimesLastYear 6.6172487  
## WorkLifeBalance 7.1352384  
## YearsAtCompany 15.4026791  
## YearsInCurrentRole 4.7991175  
## YearsSinceLastPromotion 6.3036449  
## YearsWithCurrManager 6.9346540  
## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8485714 0.5564685 0.8641478 0.1821809 0.9733676

rfHRTree10$type <- "rf\_hrTree\_10trees"  
rfTreeSpecific10 <- completeRFFunc(treeSpecific, 10)

## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 15.84294  
## Department 15.93535  
## Education 27.78380  
## JobLevel 27.06234  
## MaritalStatus 16.39962  
## OverTime 22.60902  
## WorkLifeBalance 23.00375  
## YearsWithCurrManager 32.90919  
## YearsInCurrentRole 38.62155  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 14.53293  
## Department 18.14653  
## Education 32.85010  
## JobLevel 29.26883  
## MaritalStatus 26.90360  
## OverTime 21.97900  
## WorkLifeBalance 30.14144  
## YearsWithCurrManager 33.12022  
## YearsInCurrentRole 33.40180  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 17.92474  
## Department 14.64269  
## Education 28.75878  
## JobLevel 24.02155  
## MaritalStatus 22.46780  
## OverTime 19.32335  
## WorkLifeBalance 27.60865  
## YearsWithCurrManager 31.60862  
## YearsInCurrentRole 32.86493  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 16.69357  
## Department 15.37450  
## Education 30.45704  
## JobLevel 22.01949  
## MaritalStatus 22.98287  
## OverTime 22.49288  
## WorkLifeBalance 27.44667  
## YearsWithCurrManager 32.55641  
## YearsInCurrentRole 32.62379  
## [1] "importance"  
## MeanDecreaseGini  
## BusinessTravel 17.40195  
## Department 15.24161  
## Education 26.84357  
## JobLevel 23.29130  
## MaritalStatus 19.26776  
## OverTime 28.57994  
## WorkLifeBalance 23.28351  
## YearsWithCurrManager 35.17507  
## YearsInCurrentRole 35.07135  
## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8404082 0.4967064 0.8685259 0.22746 0.9555466

rfTreeSpecific10$type <- "rf\_treeSpecific\_10trees"  
completeModels <- rbind(completeModels, rfHRTree, rfHRTree10, rfTreeSpecific, rfTreeSpecific10)

### KNN

if("class" %in% rownames(installed.packages()) == FALSE) {install.packages('class') }  
library(class)  
#convert to numeric  
HR\_factor <- HR\_tree  
HR\_factor$Attrition <-as.numeric(HR\_factor$Attrition)  
HR\_factor$BusinessTravel <- as.numeric(HR\_factor$BusinessTravel)  
HR\_factor$Department <- as.numeric(HR\_factor$Department)  
HR\_factor$Education <- as.numeric(HR\_factor$Education)  
HR\_factor$EducationField <- as.numeric(HR\_factor$EducationField)  
HR\_factor$EnvironmentSatisfaction <- as.numeric(HR\_factor$EnvironmentSatisfaction)  
HR\_factor$Gender <- as.numeric(HR\_factor$Gender)  
HR\_factor$JobInvolvement <- as.numeric(HR\_factor$JobInvolvement)  
HR\_factor$JobLevel <- as.numeric(HR\_factor$JobLevel)  
HR\_factor$JobRole <- as.numeric(HR\_factor$JobRole)  
HR\_factor$JobSatisfaction <- as.numeric(HR\_factor$JobSatisfaction)  
HR\_factor$MaritalStatus <- as.numeric(HR\_factor$MaritalStatus)  
HR\_factor$OverTime <- as.numeric(HR\_factor$OverTime)  
HR\_factor$PerformanceRating <- as.numeric(HR\_factor$PerformanceRating)  
HR\_factor$RelationshipSatisfaction <- as.numeric(HR\_factor$RelationshipSatisfaction)  
HR\_factor$StockOptionLevel <- as.numeric(HR\_factor$StockOptionLevel)  
HR\_factor$WorkLifeBalance <- as.numeric(HR\_factor$WorkLifeBalance)  
  
printNN <- function(seedNum, dataSet, kGuess=3){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 cutPoint <- floor(nrow(dataSet)\*2/3)  
 newTrain <- dataSet[randIndex[1:cutPoint],]  
 newTest <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 testNoLabel <- newTest  
 testNoLabel$Attrion <- NULL  
   
 predicted <- knn(train=newTrain, test=testNoLabel, cl=newTrain$Attrition, k=kGuess, prob=FALSE)  
 print(table(predictedAttrition=predicted, actualAttrition=newTest$Attrition))  
 set.seed(NULL)  
}

printNN(seedNum1, HR\_factor, 3)

## actualAttrition  
## predictedAttrition 1 2  
## 1 373 65  
## 2 40 12

printNN(seedNum1, HR\_factor, 5)

## actualAttrition  
## predictedAttrition 1 2  
## 1 394 67  
## 2 19 10

printNN(seedNum2, HR\_factor, 3)

## actualAttrition  
## predictedAttrition 1 2  
## 1 389 57  
## 2 32 12

printNN(seedNum2, HR\_factor, 5)

## actualAttrition  
## predictedAttrition 1 2  
## 1 404 62  
## 2 17 7

printNN(seedNum3, HR\_factor, 3)

## actualAttrition  
## predictedAttrition 1 2  
## 1 391 67  
## 2 18 14

printNN(seedNum3, HR\_factor, 5)

## actualAttrition  
## predictedAttrition 1 2  
## 1 394 71  
## 2 15 10

factorSpecific <- data.frame("Attrition"=HR\_factor$Attrition, "BusinessTravel"=HR\_factor$BusinessTravel, "Department"=HR\_factor$Department, "Education"=HR\_factor$Education, "JobLevel"=HR\_factor$JobLevel, "MaritalStatus"=HR\_factor$MaritalStatus, "Overtime"=HR\_factor$OverTime, "WorkLifeBalance"=HR\_factor$WorkLifeBalance, "YearsInCurrentRole"=HR\_factor$YearsInCurrentRole, "YearsWithCurrManager"=HR\_factor$YearsWithCurrManager )  
printNN(seedNum1, factorSpecific, 3)

## actualAttrition  
## predictedAttrition 1 2  
## 1 412 37  
## 2 1 40

printNN(seedNum1, factorSpecific, 5)

## actualAttrition  
## predictedAttrition 1 2  
## 1 413 42  
## 2 0 35

printNN(seedNum2, factorSpecific, 3)

## actualAttrition  
## predictedAttrition 1 2  
## 1 420 34  
## 2 1 35

printNN(seedNum2, factorSpecific, 5)

## actualAttrition  
## predictedAttrition 1 2  
## 1 421 39  
## 2 0 30

printNN(seedNum3, factorSpecific, 3)

## actualAttrition  
## predictedAttrition 1 2  
## 1 408 44  
## 2 1 37

printNN(seedNum3, factorSpecific, 5)

## actualAttrition  
## predictedAttrition 1 2  
## 1 409 47  
## 2 0 34

confusionTableNN <- function(seedNum, dataSet, kGuess=3){  
 # set seed  
 set.seed(seedNum)  
 # Generate random sample of rows  
 randIndex <- sample(1:nrow(dataSet))  
 newTrain <- dataSet[randIndex[1:cutPoint],]  
 newTest <- dataSet[randIndex[(cutPoint+1):length(randIndex)],]  
 testNoLabel <- newTest  
 testNoLabel$Attrion <- NULL  
   
 predicted <- knn(train=newTrain, test=testNoLabel, cl=newTrain$Attrition, k=kGuess, prob=FALSE)  
 set.seed(NULL)  
 return(table(predictedAttrition=predicted, actualAttrition=newTest$Attrition))  
}  
  
tableCalc2 <- function(newTable){  
 calcTable <- as.data.frame(as.matrix.data.frame(newTable))  
 accurateNumbers <- 0  
 totalNumbers <- 0  
 precision <- data.frame()  
 recall <- data.frame()  
 for(i in 1:length(calcTable)){  
 columnSum <- sum(calcTable[,i])  
 rowSum <- sum(calcTable[i,])  
 cell <- calcTable[i,i]  
 accurateNumbers <- accurateNumbers + cell  
 totalNumbers <- totalNumbers + columnSum  
 precision[1,i] <- cell / columnSum  
 recall[1,i] <- cell / rowSum  
 }  
 dataFrame <- data.frame("precisionNo"=precision[1,1], "precisionYes"=precision[1,2], "recallNo"=recall[1,1],"recallYes"=recall[1,2], "accuracy"=accurateNumbers/totalNumbers)  
}  
  
averageTableCalc2 <- function(dataFrame){  
 avgAccuracy <- mean(dataFrame$accuracy)  
 avgPrecisionYes <- mean(dataFrame$precisionYes)  
 avgPrecisionNo <- mean(dataFrame$precisionNo)  
 avgRecallYes <- mean(dataFrame$recallYes)  
 avgRecallNo <- mean(dataFrame$recallNo)  
 newDF <- data.frame(avgAccuracy, avgPrecisionYes, avgPrecisionNo, avgRecallYes, avgRecallNo)  
 return(newDF)  
}  
  
  
completeNNFunc <- function(dataSet, kGuess=3){  
 table1 <- confusionTableNN(seedNum1, dataSet, kGuess)  
 table2 <- confusionTableNN(seedNum2, dataSet, kGuess)  
 table3 <- confusionTableNN(seedNum3, dataSet, kGuess)  
 table4 <- confusionTableNN(seedNum4, dataSet, kGuess)  
 table5 <- confusionTableNN(seedNum5, dataSet, kGuess)  
   
 tableCalc1 <- tableCalc2(table1)  
 tableCalc2 <- tableCalc2(table2)  
 tableCalc3 <- tableCalc2(table3)  
 tableCalc4 <- tableCalc2(table4)  
 tableCalc5 <- tableCalc2(table5)  
   
 tableCalc <- data.frame(rbind(as.matrix(tableCalc1),as.matrix(tableCalc2),as.matrix(tableCalc3),as.matrix(tableCalc4),as.matrix(tableCalc5)))  
   
 avgTableCalc <- averageTableCalc2(tableCalc)  
 print(avgTableCalc)  
}

nn3 <- completeNNFunc(factorSpecific, 3)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.9130612 0.4677794 0.9966005 0.9632506 0.9089811

nn3$type <- "nn\_treeSpecific\_3"  
nn10 <- completeNNFunc(factorSpecific, 10)

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8893878 0.2993576 1 1 0.8839559

nn10$type <- "nn\_treeSpecific\_10"  
completeModels <- rbind(completeModels, nn3, nn10)  
completeModels

## avgAccuracy avgPrecisionYes avgPrecisionNo avgRecallYes avgRecallNo  
## 1 0.8248980 0.4087265 0.8681532 0.2416280 0.9341750  
## 2 0.8416327 0.4991805 0.8788818 0.3053566 0.9419136  
## 3 0.8375510 0.4811203 0.8776329 0.3002276 0.9380277  
## 4 0.8261224 0.3762607 0.8545910 0.1302459 0.9570162  
## 5 0.8861224 0.7035583 0.9091540 0.4885391 0.9607056  
## 6 0.8028571 0.4203825 0.9265368 0.6472200 0.8325573  
## 7 0.8183673 0.3966877 0.8724989 0.2817504 0.9185592  
## 8 0.8485714 0.5564685 0.8641478 0.1821809 0.9733676  
## 9 0.8228571 0.4147345 0.8744216 0.2915126 0.9226324  
## 10 0.8404082 0.4967064 0.8685259 0.2274600 0.9555466  
## 11 0.9130612 0.4677794 0.9966005 0.9632506 0.9089811  
## 12 0.8893878 0.2993576 1.0000000 1.0000000 0.8839559  
## type  
## 1 decisionTrees\_hrTree  
## 2 decisionTrees\_treeSpecific  
## 3 decisionTrees\_treeIncome  
## 4 decisionTrees\_treeReduced  
## 5 svm\_hrTree  
## 6 nb\_hrTree  
## 7 rf\_hrTree\_3trees  
## 8 rf\_hrTree\_10trees  
## 9 rf\_treeSpecific\_3trees  
## 10 rf\_treeSpecific\_10trees  
## 11 nn\_treeSpecific\_3  
## 12 nn\_treeSpecific\_10

completeModels <- subset(completeModels, select=c(6,1:5))

formattable(completeModels, align = c("l",rep("r", NCOL("type") - 1)), list(  
 `type` = formatter("span", style = ~ style(color = "#000000",font.weight = "bold")),   
 area(col = 2:6) ~ color\_tile("#ff0000", "#71CA97")))

type

avgAccuracy

avgPrecisionYes

avgPrecisionNo

avgRecallYes

avgRecallNo

decisionTrees\_hrTree

0.8248980

0.4087265

0.8681532

0.2416280

0.9341750

decisionTrees\_treeSpecific

0.8416327

0.4991805

0.8788818

0.3053566

0.9419136

decisionTrees\_treeIncome

0.8375510

0.4811203

0.8776329

0.3002276

0.9380277

decisionTrees\_treeReduced

0.8261224

0.3762607

0.8545910

0.1302459

0.9570162

svm\_hrTree

0.8861224

0.7035583

0.9091540

0.4885391

0.9607056

nb\_hrTree

0.8028571

0.4203825

0.9265368

0.6472200

0.8325573

rf\_hrTree\_3trees

0.8183673

0.3966877

0.8724989

0.2817504

0.9185592

rf\_hrTree\_10trees

0.8485714

0.5564685

0.8641478

0.1821809

0.9733676

rf\_treeSpecific\_3trees

0.8228571

0.4147345

0.8744216

0.2915126

0.9226324

rf\_treeSpecific\_10trees

0.8404082

0.4967064

0.8685259

0.2274600

0.9555466

nn\_treeSpecific\_3

0.9130612

0.4677794

0.9966005

0.9632506

0.9089811

nn\_treeSpecific\_10

0.8893878

0.2993576

1.0000000

1.0000000

0.8839559

When looking at the complete table, we do need to define what is our success criteria for defining how well a model performs.

We can look at accuracy, precisionYes, precisionNo, recallYes, and recallNo and decide across a combination of metrics to best define what makes the most sense.

Because we ultimately want to maximize employees who are likely to leave, we should weight Yes.

When we look at recallYes, which provides us with insight on the percentage of correctly classified relevant results, K Nearest Neighbors immediately handles best followed by Naive Bayes.

The best Accuracy is KNN followed by SVM.

The best on precisionYes is SVM followed by RF with 10 trees on the complete dataset.

Each of these has its cons. KNN has fairly low precision on Yes meaning that only when it’s certain, will it make a move. And that certainty is between 30-47% of the of time. But when it does count, the data is indredibly accurate.

SVM was high on PrecisionYes but medium on Recall. That means that it modeled more employees as being likely to leave, but of those, only 49% were truly likely to leave.

Ultimately, this is a question of is it better to classify someone as leaving when they’re staying or to classify someone as staying when they’re leaving?

## Results

###Exploratory Data Analysis & Visualization

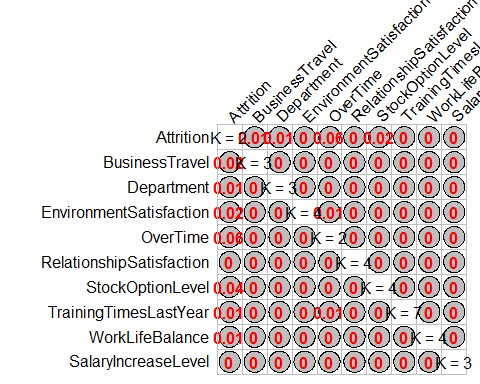
Exploratory Data Analysis and Visualization showed that there was not strong association between any one attribute and attrition. The Goodman and Kruskal Tau measure model was used to establish association of categorical values. To make these associations easier to visualize, we grouped the attributes in 3 groups:

Person/Profile Company Role/Job

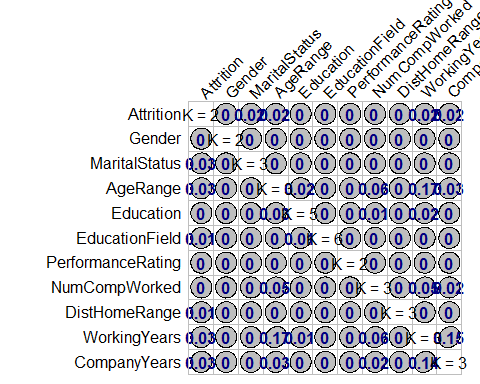
and compared them to Attrition.

Each group showed low association to attrition.

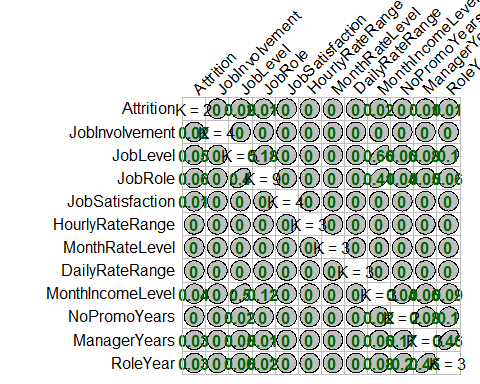
GKmatrix1<- GKtauDataframe(Frame1)  
plot(GKmatrix1, corrColors = "red")



GKmatrix1<- GKtauDataframe(Frame2)  
plot(GKmatrix1, corrColors = "navyblue")

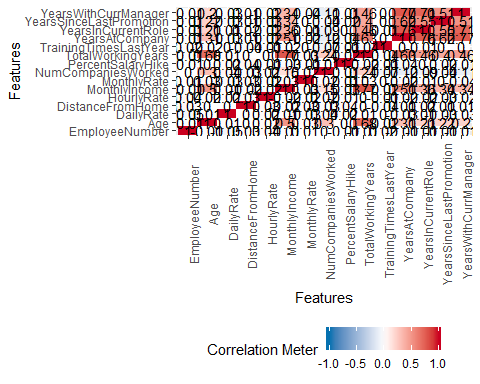


GKmatrix1<- GKtauDataframe(Frame3)  
plot(GKmatrix1, corrColors = "darkgreen")

 Attributes were correlated to each other, and we can see pockets of correlation between attributes. We can use this information to simplify models in later stages.

!Use graphic from slide 7!

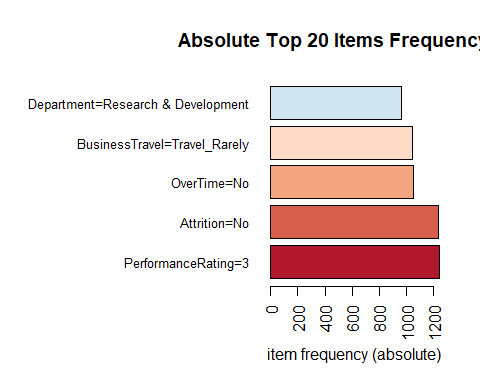
plot\_correlation(HR\_eda, type = 'continuous')

 What was interesting and unexpected was that the attribute to attribute correlation chart showed actionable information that could be used for simplifying models while the direct correlation chart was relatively inconclusive.

Additionally, each attribute was correlated to attrition individually. The findings confirmed that there was any singular or group of attributes that could be strongly correlated to attrition and further work with more advanced techniques should be used to identify important attributes.

### Association Rule Mining

# Item Frequency Plot Top 5 Absolute  
itemFrequencyPlot(HR\_Trans,support = 0.2, cex.names=0.8, topN=5, col=brewer.pal(8,'RdBu'), type="absolute", main="Absolute Top 20 Items Frequency Plot",horiz=TRUE)

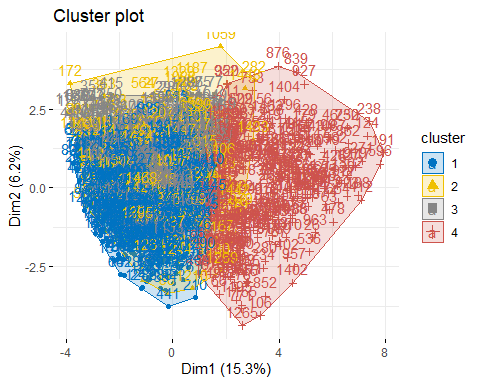
 Conversion of the data to transacions allowed for an initial assessment of the most frequent responses. Attrition, the attribute of interest in this study had 83.9% “No” responses (1233 out of the 1470 transactions), followed by Overtime with 71.7% “No” responses and Business Travel with 70.9% “Travel Rarely” responses. Considering other data analytics indicated Frequent business travel and overtime as a driver for attrition, knowing that only 30% or less of the respondants had those responses is key information when it comes to deciding what type of startegies to implement and who shoudld be the target audience.

By fixing the RHS to Attrition = Yes and Attrition = No rules provide more insight.

With Attrition = Yes, the most frequent factors in the top 20 rules are: \* Marital Status = Single. In 13 out of the 20 rules \* Overtime = Yes. In 18 out of the 20 rules \* Years with current Manager = 0. In 16 out of the 20 rules \* Years in current Role = 0. In 12 out of the 20 rules \* Low Income. In 10 out of the 20 rules

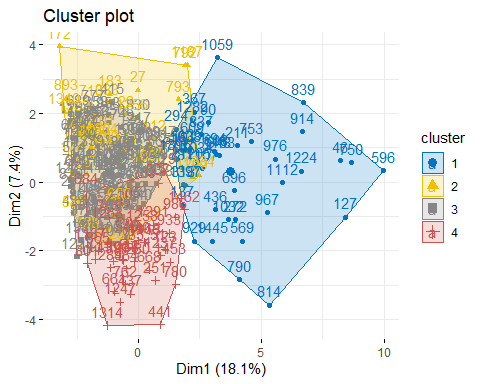
With Attrition = No, the most frequent factors in the top 20 rules are: \* Department=Research & Development. In 10 out of the 20 rules  
\* OverTime=No. In 15 out of the 20 rules  
\* StockOptionLevel=1. In 6 out of the 20 rules  
\* WorkLifeBalance=3. In 11 out of the 20 rules

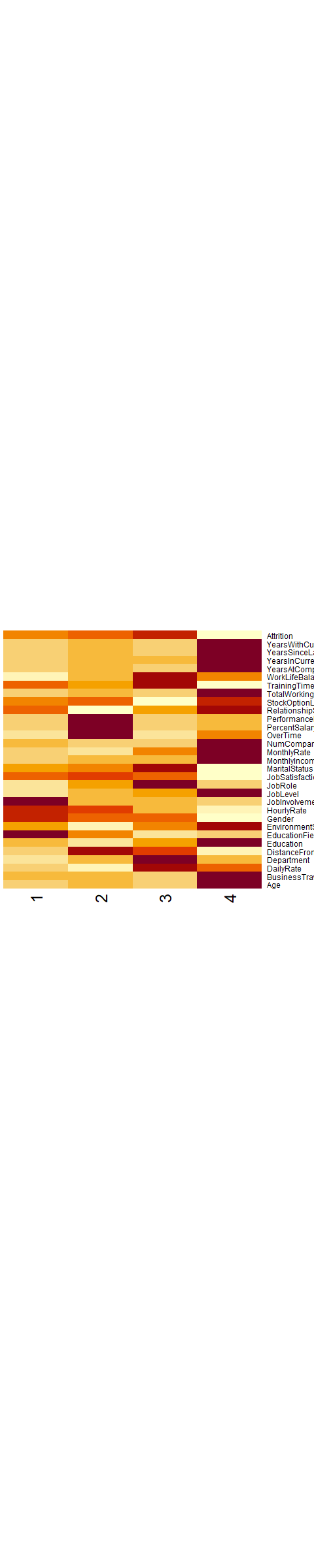
<!--html\_preserve--><div id="htmlwidget-5bc1878f89b504715ed5" style="width:480px;height:384px;" class="visNetwork html-widget"></div>  
<script type="application/json" data-for="htmlwidget-5bc1878f89b504715ed5">{"x":{"nodes":{"id":[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31],"label":["MaritalStatus=Single","OverTime=Yes","YearsWithCurrManager=0","Group=Low\_Income","JobLevel=1","YearsInCurrentRole=0","StockOptionLevel=0","YearsSinceLastPromotion=0","BusinessTravel=Travel\_Frequently","PerformanceRating=3","Attrition=Yes","rule 1","rule 2","rule 3","rule 4","rule 5","rule 6","rule 7","rule 8","rule 9","rule 10","rule 11","rule 12","rule 13","rule 14","rule 15","rule 16","rule 17","rule 18","rule 19","rule 20"],"group":["item","item","item","item","item","item","item","item","item","item","item","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule"],"value":[1,1,1,1,1,1,1,1,1,1,1,25.75,50.5,1,25.75,50.5,25.75,25.75,50.5,25.75,25.75,1,1,25.75,25.75,25.75,1,1,100,100,75.25],"color":["#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EE1B1B","#EED5D5","#EED5D5","#EEDCDC"],"title":["MaritalStatus=Single","OverTime=Yes","YearsWithCurrManager=0","Group=Low\_Income","JobLevel=1","YearsInCurrentRole=0","StockOptionLevel=0","YearsSinceLastPromotion=0","BusinessTravel=Travel\_Frequently","PerformanceRating=3","Attrition=Yes","<B>[1]<\/B><BR><B>{MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 5","<B>[2]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;YearsWithCurrManager=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0116<BR>confidence = 1<BR>lift = 6.2<BR>count = 17<BR>order = 5","<B>[3]<\/B><BR><B>{MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0102<BR>confidence = 1<BR>lift = 6.2<BR>count = 15<BR>order = 6","<B>[4]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 6","<B>[5]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0116<BR>confidence = 1<BR>lift = 6.2<BR>count = 17<BR>order = 6","<B>[6]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 6","<B>[7]<\/B><BR><B>{MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 6","<B>[8]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0116<BR>confidence = 1<BR>lift = 6.2<BR>count = 17<BR>order = 6","<B>[9]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsSinceLastPromotion=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 6","<B>[10]<\/B><BR><B>{BusinessTravel=Travel\_Frequently,<BR>&nbsp;&nbsp;JobLevel=1,<BR>&nbsp;&nbsp;PerformanceRating=3,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 7","<B>[11]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0102<BR>confidence = 1<BR>lift = 6.2<BR>count = 15<BR>order = 7","<B>[12]<\/B><BR><B>{MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0102<BR>confidence = 1<BR>lift = 6.2<BR>count = 15<BR>order = 7","<B>[13]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 7","<B>[14]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 7","<B>[15]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0109<BR>confidence = 1<BR>lift = 6.2<BR>count = 16<BR>order = 7","<B>[16]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsSinceLastPromotion=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0102<BR>confidence = 1<BR>lift = 6.2<BR>count = 15<BR>order = 7","<B>[17]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsInCurrentRole=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0,<BR>&nbsp;&nbsp;Group=Low\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0102<BR>confidence = 1<BR>lift = 6.2<BR>count = 15<BR>order = 8","<B>[18]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsInCurrentRole=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0129<BR>confidence = 0.95<BR>lift = 5.89<BR>count = 19<BR>order = 5","<B>[19]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;StockOptionLevel=0,<BR>&nbsp;&nbsp;YearsWithCurrManager=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0129<BR>confidence = 0.95<BR>lift = 5.89<BR>count = 19<BR>order = 5","<B>[20]<\/B><BR><B>{JobLevel=1,<BR>&nbsp;&nbsp;MaritalStatus=Single,<BR>&nbsp;&nbsp;OverTime=Yes,<BR>&nbsp;&nbsp;YearsInCurrentRole=0}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=Yes}<\/B><BR><BR>support = 0.0122<BR>confidence = 0.947<BR>lift = 5.88<BR>count = 18<BR>order = 5"],"shape":["box","box","box","box","box","box","box","box","box","box","box","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle"],"x":[-0.600957778561597,-0.49575269074701,-0.417442079132989,-0.422270568184076,-0.232024702669341,-0.338272526451312,-0.409005387234742,-0.221105367684205,0.728362895270825,1,-0.37417751231944,-1,-0.726456928830335,-0.653838865926093,-0.19401411218057,0.094008315248544,0.0211200148606956,-0.869783203412773,-0.108848055083313,-0.462001943462623,0.206364139430302,-0.680125653723969,-0.821069522816543,-0.816024480802355,-0.0907953456642309,-0.242583005393673,-0.176165506303663,-0.742849534538436,0.105072493656886,-0.375112148432987,-0.424754193008756],"y":[0.143867362694003,-0.00619062046507646,-0.0438188560546097,0.135821538727039,0.0160828747326762,0.320365593107318,-0.200564533539766,-1,1,0.723768262766066,0.0425599086600326,0.178034980250294,-0.301208283458874,0.452984170084479,0.398451131098795,0.0150428728168677,-0.0925774640729562,-0.192059257113307,0.18916757711451,-0.514588929653898,0.448137587664506,0.334129069676654,-0.0517088903100179,0.0593751231100097,-0.138150662930341,-0.247397475739124,-0.420526707696031,0.187638971264544,0.166464059462799,-0.417744219447846,0.511967718334289]},"edges":{"from":[1,2,3,4,5,1,2,3,1,2,6,3,4,5,1,2,6,3,5,2,7,6,3,5,1,2,3,4,1,2,7,3,4,5,1,2,7,3,5,2,7,8,3,9,5,10,6,3,4,5,1,2,6,3,4,1,2,7,6,3,4,5,1,2,7,6,3,5,2,7,6,3,4,5,1,2,7,3,4,5,2,7,8,3,4,5,1,2,7,6,3,4,5,2,7,6,5,2,7,3,5,1,2,6,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31],"to":[12,12,12,12,13,13,13,13,14,14,14,14,14,15,15,15,15,15,16,16,16,16,16,17,17,17,17,17,18,18,18,18,18,19,19,19,19,19,20,20,20,20,20,21,21,21,21,21,21,22,22,22,22,22,22,23,23,23,23,23,23,24,24,24,24,24,24,25,25,25,25,25,25,26,26,26,26,26,26,27,27,27,27,27,27,28,28,28,28,28,28,28,29,29,29,29,30,30,30,30,31,31,31,31,11,11,11,11,11,11,11,11,11,11,11,11,11,11,11,11,11,11,11,11],"arrows":["to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to"]},"nodesToDataframe":true,"edgesToDataframe":true,"options":{"width":"100%","height":"100%","nodes":{"shape":"dot","scaling":{"label":{"enabled":true}},"physics":false},"manipulation":{"enabled":false},"edges":{"smooth":false},"physics":{"stabilization":false},"interaction":{"hover":true}},"groups":["item","rule"],"width":null,"height":null,"idselection":{"enabled":true,"style":"width: 150px; height: 26px","useLabels":true,"main":"Select by id"},"byselection":{"enabled":false,"style":"width: 150px; height: 26px","multiple":false,"hideColor":"rgba(200,200,200,0.5)","highlight":false},"main":null,"submain":null,"footer":null,"background":"rgba(0, 0, 0, 0)","igraphlayout":{"type":"square"},"tooltipStay":300,"tooltipStyle":"position: fixed;visibility:hidden;padding: 5px;white-space: nowrap;font-family: verdana;font-size:14px;font-color:#000000;background-color: #f5f4ed;-moz-border-radius: 3px;-webkit-border-radius: 3px;border-radius: 3px;border: 1px solid #808074;box-shadow: 3px 3px 10px rgba(0, 0, 0, 0.2);","highlight":{"enabled":true,"hoverNearest":true,"degree":1,"algorithm":"all","hideColor":"rgba(200,200,200,0.5)","labelOnly":true},"collapse":{"enabled":false,"fit":false,"resetHighlight":true,"clusterOptions":null,"keepCoord":true,"labelSuffix":"(cluster)"}},"evals":[],"jsHooks":[]}</script><!--/html\_preserve--><!--html\_preserve--><div id="htmlwidget-e7533d2602bea58fac02" style="width:480px;height:384px;" class="visNetwork html-widget"></div>  
<script type="application/json" data-for="htmlwidget-e7533d2602bea58fac02">{"x":{"nodes":{"id":[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37],"label":["Department=Research & Development","OverTime=No","StockOptionLevel=1","WorkLifeBalance=3","BusinessTravel=Travel\_Rarely","Group=High\_Income","EnvironmentSatisfaction=4","YearsWithCurrManager=2","JobLevel=2","Group=Mid\_Income","PerformanceRating=3","EducationField=Life Sciences","MaritalStatus=Married","Gender=Female","JobSatisfaction=4","TrainingTimesLastYear=3","Attrition=No","rule 1","rule 2","rule 3","rule 4","rule 5","rule 6","rule 7","rule 8","rule 9","rule 10","rule 11","rule 12","rule 13","rule 14","rule 15","rule 16","rule 17","rule 18","rule 19","rule 20"],"group":["item","item","item","item","item","item","item","item","item","item","item","item","item","item","item","item","item","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule","rule"],"value":[1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,12.88,1.98999999999999,100,33.67,5.94999999999999,5.94999999999999,5.94999999999999,4.96,28.72,1,23.77,21.79,20.8,11.89,56.44,31.69,26.74,26.74,81.19,53.47],"color":["#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#CBD2FC","#EE1B1B","#EE7272","#EE8E8E","#EEA0A0","#EEA4A4","#EEA4A4","#EEA4A4","#EEA4A4","#EEA5A5","#EEABAB","#EEACAC","#EEAEAE","#EEB0B0","#EEBDBD","#EEBDBD","#EEC1C1","#EEC8C8","#EEC8C8","#EED2D2","#EEDCDC"],"title":["Department=Research & Development","OverTime=No","StockOptionLevel=1","WorkLifeBalance=3","BusinessTravel=Travel\_Rarely","Group=High\_Income","EnvironmentSatisfaction=4","YearsWithCurrManager=2","JobLevel=2","Group=Mid\_Income","PerformanceRating=3","EducationField=Life Sciences","MaritalStatus=Married","Gender=Female","JobSatisfaction=4","TrainingTimesLastYear=3","Attrition=No","<B>[1]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;StockOptionLevel=1,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.108<BR>confidence = 0.975<BR>lift = 1.16<BR>count = 159<BR>order = 5","<B>[2]<\/B><BR><B>{BusinessTravel=Travel\_Rarely,<BR>&nbsp;&nbsp;Department=Research & Development,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;Group=High\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.101<BR>confidence = 0.967<BR>lift = 1.15<BR>count = 148<BR>order = 5","<B>[3]<\/B><BR><B>{OverTime=No,<BR>&nbsp;&nbsp;StockOptionLevel=1,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.168<BR>confidence = 0.965<BR>lift = 1.15<BR>count = 247<BR>order = 4","<B>[4]<\/B><BR><B>{EnvironmentSatisfaction=4,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.122<BR>confidence = 0.963<BR>lift = 1.15<BR>count = 180<BR>order = 4","<B>[5]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;YearsWithCurrManager=2}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.103<BR>confidence = 0.962<BR>lift = 1.15<BR>count = 152<BR>order = 4","<B>[6]<\/B><BR><B>{JobLevel=2,<BR>&nbsp;&nbsp;StockOptionLevel=1,<BR>&nbsp;&nbsp;Group=Mid\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.103<BR>confidence = 0.962<BR>lift = 1.15<BR>count = 152<BR>order = 4","<B>[7]<\/B><BR><B>{BusinessTravel=Travel\_Rarely,<BR>&nbsp;&nbsp;JobLevel=2,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.103<BR>confidence = 0.962<BR>lift = 1.15<BR>count = 152<BR>order = 5","<B>[8]<\/B><BR><B>{EnvironmentSatisfaction=4,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;PerformanceRating=3,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.103<BR>confidence = 0.962<BR>lift = 1.15<BR>count = 151<BR>order = 5","<B>[9]<\/B><BR><B>{EducationField=Life Sciences,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;StockOptionLevel=1}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.119<BR>confidence = 0.962<BR>lift = 1.15<BR>count = 175<BR>order = 4","<B>[10]<\/B><BR><B>{JobLevel=2,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;StockOptionLevel=1}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.1<BR>confidence = 0.961<BR>lift = 1.15<BR>count = 147<BR>order = 4","<B>[11]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;MaritalStatus=Married,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.116<BR>confidence = 0.96<BR>lift = 1.15<BR>count = 170<BR>order = 5","<B>[12]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;JobLevel=2,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.114<BR>confidence = 0.96<BR>lift = 1.14<BR>count = 168<BR>order = 4","<B>[13]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;JobLevel=2,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;PerformanceRating=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.114<BR>confidence = 0.96<BR>lift = 1.14<BR>count = 167<BR>order = 5","<B>[14]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;Gender=Female,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.107<BR>confidence = 0.958<BR>lift = 1.14<BR>count = 158<BR>order = 5","<B>[15]<\/B><BR><B>{OverTime=No,<BR>&nbsp;&nbsp;PerformanceRating=3,<BR>&nbsp;&nbsp;StockOptionLevel=1,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.138<BR>confidence = 0.958<BR>lift = 1.14<BR>count = 203<BR>order = 5","<B>[16]<\/B><BR><B>{JobSatisfaction=4,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.121<BR>confidence = 0.957<BR>lift = 1.14<BR>count = 178<BR>order = 4","<B>[17]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;TrainingTimesLastYear=3,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.118<BR>confidence = 0.956<BR>lift = 1.14<BR>count = 173<BR>order = 4","<B>[18]<\/B><BR><B>{MaritalStatus=Married,<BR>&nbsp;&nbsp;OverTime=No,<BR>&nbsp;&nbsp;StockOptionLevel=1,<BR>&nbsp;&nbsp;WorkLifeBalance=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.118<BR>confidence = 0.956<BR>lift = 1.14<BR>count = 173<BR>order = 5","<B>[19]<\/B><BR><B>{Department=Research & Development,<BR>&nbsp;&nbsp;JobLevel=2,<BR>&nbsp;&nbsp;PerformanceRating=3}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.155<BR>confidence = 0.954<BR>lift = 1.14<BR>count = 228<BR>order = 4","<B>[20]<\/B><BR><B>{BusinessTravel=Travel\_Rarely,<BR>&nbsp;&nbsp;Department=Research & Development,<BR>&nbsp;&nbsp;Group=High\_Income}<\/B><BR>&nbsp;&nbsp; => <B>{Attrition=No}<\/B><BR><BR>support = 0.136<BR>confidence = 0.952<BR>lift = 1.14<BR>count = 200<BR>order = 4"],"shape":["box","box","box","box","box","box","box","box","box","box","box","box","box","box","box","box","box","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle","circle"],"x":[0.109831167432097,-0.12883941152168,-0.320881816726418,-0.203668973457435,0.627246369645532,0.542484869529657,-1,1,0.204995984580371,-0.147008279868371,-0.462553444244441,0.258246333422435,-0.723119523791593,0.771347979480098,-0.714310510139751,-0.047131144809814,-0.0595746358289112,-0.199046922214244,0.251781846087691,-0.403026758615111,-0.560873972030638,0.444645027099508,-0.0896354847795608,0.256430832245083,-0.573817451727819,0.0243507515906316,-0.109354884615286,-0.280031577109795,0.187776255488939,-0.0550777680217124,0.236027350871266,-0.383413805276888,-0.388140111229604,-0.0487902787470647,-0.452683941808862,0.0155787747197125,0.407988393854424],"y":[0.175855944049765,0.0816084567958508,-0.299490062369748,0.189186008420966,0.327446515804984,0.704349193589877,0.115524576775428,-0.129444718172417,-0.302652307301293,-1,-0.301912180831988,-0.837717196625235,0.433040461009671,0.561768488116354,0.849550169106171,1,0.0616148256644586,-0.0373163428414361,0.39994894952907,-0.0284308572736597,0.193856443656111,-0.00518494174183315,-0.539187427501746,0.0784645647503308,0.00982762654878799,-0.403745967477904,-0.285783326977332,0.335574073860386,-0.0441307322459368,-0.136275431864001,0.302131141190651,-0.129917581495368,0.457801504874154,0.531407296115816,0.118483409817094,-0.218686691570865,0.399792460532238]},"edges":{"from":[1,2,3,4,5,1,2,6,2,3,4,7,2,4,1,2,8,9,3,10,5,9,2,4,7,2,11,4,12,2,3,9,2,3,1,13,2,4,1,9,4,1,9,2,11,1,14,2,4,2,11,3,4,15,2,4,1,16,4,13,2,3,4,1,9,11,5,1,6,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37],"to":[18,18,18,18,19,19,19,19,20,20,20,21,21,21,22,22,22,23,23,23,24,24,24,24,25,25,25,25,26,26,26,27,27,27,28,28,28,28,29,29,29,30,30,30,30,31,31,31,31,32,32,32,32,33,33,33,34,34,34,35,35,35,35,36,36,36,37,37,37,17,17,17,17,17,17,17,17,17,17,17,17,17,17,17,17,17,17,17,17],"arrows":["to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to","to"]},"nodesToDataframe":true,"edgesToDataframe":true,"options":{"width":"100%","height":"100%","nodes":{"shape":"dot","scaling":{"label":{"enabled":true}},"physics":false},"manipulation":{"enabled":false},"edges":{"smooth":false},"physics":{"stabilization":false},"interaction":{"hover":true}},"groups":["item","rule"],"width":null,"height":null,"idselection":{"enabled":true,"style":"width: 150px; height: 26px","useLabels":true,"main":"Select by id"},"byselection":{"enabled":false,"style":"width: 150px; height: 26px","multiple":false,"hideColor":"rgba(200,200,200,0.5)","highlight":false},"main":null,"submain":null,"footer":null,"background":"rgba(0, 0, 0, 0)","igraphlayout":{"type":"square"},"tooltipStay":300,"tooltipStyle":"position: fixed;visibility:hidden;padding: 5px;white-space: nowrap;font-family: verdana;font-size:14px;font-color:#000000;background-color: #f5f4ed;-moz-border-radius: 3px;-webkit-border-radius: 3px;border-radius: 3px;border: 1px solid #808074;box-shadow: 3px 3px 10px rgba(0, 0, 0, 0.2);","highlight":{"enabled":true,"hoverNearest":true,"degree":1,"algorithm":"all","hideColor":"rgba(200,200,200,0.5)","labelOnly":true},"collapse":{"enabled":false,"fit":false,"resetHighlight":true,"clusterOptions":null,"keepCoord":true,"labelSuffix":"(cluster)"}},"evals":[],"jsHooks":[]}</script><!--/html\_preserve-->  
### K-means Clustering  
K-means was run with k=2, 3, 4, 5, and 6 clusters, both with and without the attribution attribute. The most interesting cluster was k=4, scaled data, with attribution included. This shows one group that separates clearly, and three that overlap.   
  
```r  
### with 4 clusters, there is too much overlap with three clusters  
### but one cluster is still separate  
fviz\_cluster(model\_attsm4, data = xc\_att.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



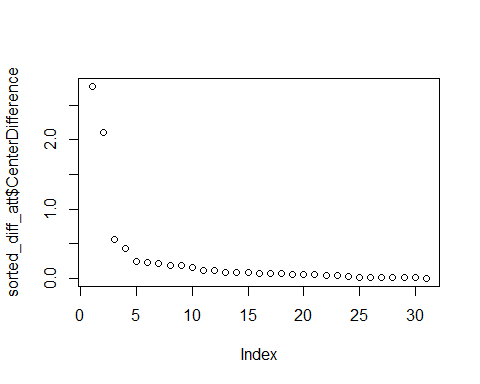
When looking at only the people who left (attrition = yes), notice how few people left in the right group (cluster 1).

fviz\_cluster(model\_YES4, data = att\_YES.sm,  
 ellipse.type = "convex",  
 palette = "jco",  
 ggtheme = theme\_minimal())



The 4-cluster version shows clear separation on several attributes between the group with the highest attrition and the one with the lowest attrition. 

plot(sorted\_diff\_att$CenterDifference)

 The ten most influential attributes were:

sorted\_diff\_att[1:11, ]

## X1 X2 CenterDifference  
## PerformanceRating -0.42608501 2.345352910 2.77  
## PercentSalaryHike -0.31186661 1.796015392 2.11  
## Department -0.61976593 -0.063038984 0.56  
## JobRole -0.42959032 -0.003659263 0.43  
## JobLevel -0.52624486 -0.276771865 0.25  
## YearsInCurrentRole -0.33714203 -0.102305223 0.23  
## MonthlyIncome -0.49682659 -0.287082649 0.21  
## TotalWorkingYears -0.42706962 -0.246242030 0.18  
## YearsWithCurrManager -0.29784675 -0.118019334 0.18  
## YearsAtCompany -0.37503102 -0.215667862 0.16  
## RelationshipSatisfaction 0.01559113 -0.107551190 0.12

### Decision Trees

### Support Vector Machines

### K nearest neighbors

### Naive Bayes

### Random Forest

## Conclusions

### Next steps in analysis

Analysis is best done iteratively. To further improve on the models, it is recommended that future analysis include these steps.

*Gather more data* \* More data leads to better results. It would be better to have several thousand observations. \* Collect more attributes. Research has shown that some other factors that weren’t studied here can impact attrition, including onboarding experience and the networking of employees. \* Get a balanced sample. Some models work better when the “yes” and “no” classes have similar numbers of observations. \* Compare the models’ predictions with actual attrition to see what parameters they may have chosen to be groups.

*Focus on the most successful models* All models had some good qualities. We recommend continuing with:

* KNN (very accurate at identifying who will quit but not as useful for large data, and may need to have cutoffs based on timeframes)
* SVM (very little processing work, scales well but doesn’t really provide insight into what business variables to improve)
* Random forest (good at handling a variety of attributes, provides attributes, but is less accurate)

*Run the models on the new data every quarter*

* This helps to identify if there are changes at the company, and whether new programs to influence attrition are working.

### Business decisions

In addition to using the models to predict if an individual employee is going to leave, the models also identified common contributions to attrition. The HR department can develop programs to target these factors.

Of the attributes that the models chose as influential, the most common were:

* Overtime (in 5 models)
* Environmental satisfaction (4)
* Job level (4)
* Marital status (4)
* Monthly income (4)
* Work-life balance (4)

The company cannot impact marital status, and it is illegal to hire based on that attribute. However, the company can influence the others. For example, it could reduce overtime…or pay people who work overtime more money. There are lots of possible ways of addressing these issues, and the HR department should look further at things like environmental satisfaction and work-life balance by interviewing at-risk employees.

## References

[1] “Why People Quit Their Jobs.” Harvard Business Review, Sept. 2016, hbr.org/2016/09/why-people-quit-their-jobs. Accessed 10 Mar. 2020.

[2] “Why People Quit Their Jobs.” Harvard Business Review, Sept. 2016, hbr.org/2016/09/why-people-quit-their-jobs. Accessed 10 Mar. 2020.

[3] “How to Predict Turnover on Your Sales Team.” Harvard Business Review, July 2017, hbr.org/2017/07/how-to-predict-turnover-on-your-sales-team. Accessed 10 Mar. 2020.

[4] “How to Predict Turnover on Your Sales Team.” Harvard Business Review, July 2017, hbr.org/2017/07/how-to-predict-turnover-on-your-sales-team. Accessed 10 Mar. 2020.

[5] “To Retain New Hires, Make Sure You Meet with Them in Their First Week.” Harvard Business Review, 14 June 2018, hbr.org/2018/06/to-retain-new-hires-make-sure-you-meet-with-them-in-their-first-week. Accessed 10 Mar. 2020.

[6] “How to Predict Turnover on Your Sales Team.” Harvard Business Review, July 2017, hbr.org/2017/07/how-to-predict-turnover-on-your-sales-team. Accessed 10 Mar. 2020.

[7] “8 Things Leaders Do That Make Employees Quit.” Harvard Business Review, 10 Sept. 2019, hbr.org/2019/09/8-things-leaders-do-that-make-employees-quit. Accessed 10 Mar. 2020.

[8] “Work Institute Releases National Employee Retention Report.” Businesswire.Com, May 2018, www.businesswire.com/news/home/20180501006594/en/Work-Institute-Releases-National-Employee-Retention-Report. Accessed 10 Mar. 2020.

[9] “How to Predict Turnover on Your Sales Team.” Harvard Business Review, July 2017, hbr.org/2017/07/how-to-predict-turnover-on-your-sales-team. Accessed 10 Mar. 2020.

[10] Maurer, Roy. “Onboarding Key to Retaining, Engaging Talent.” SHRM, SHRM, 16 Apr. 2015, www.shrm.org/ResourcesAndTools/hr-topics/talent-acquisition/Pages/Onboarding-Key-Retaining-Engaging-Talent.aspx. Accessed 10 Mar. 2020.

[11] “The Battle Against Executive Attrition.” Harvard Business Review, 17 July 2008, hbr.org/2008/07/the-battle-against-executive-a. Accessed 10 Mar. 2020.

[12] Dowsett, C. (2018, April). It’s Time to Talk About Organizational Bias in Data Use. Medium; Towards Data Science.