HR ANALYTICS CASE STUDY

SUBMISSION

Group Name: EXEMPLARY ANALYSTS

1. SACHIN SHINGARI
2. SHAILENDRA KUMAR
3. UJWALA ISHWARAPPAGOL
4. USHA HARIRAM

Basic Understanding: XYZ Company

- 4410 employee data for a span of 1 year
- Annual attrition rate close to 15%
- Attrition is high and management is concerned

Our Approach:

- Problem Definition – “Higher Attrition Rate”
- Understanding Objectives and expectations
- Data collection, importing and Data Understanding
- Assumptions
- Data Cleaning and creation of derived metrics, if any
- Conducting EDA along with relevant Data preparation
- Collating Data in a Master File for Model Development
- Identifying Predictor and Response Variables
- Splitting Data in to Training and Test Data set randomly
- Developing the model based on training set
- Fine Tuning the model
- Selection of most appropriate model based on algorithm and Statistical criterion
- Prediction from test set
- Assessment of Quality of Prediction of Response Variable
- Final Recommendations

Why do we need to study the causes?

15% of attrition per year leads to:

- Project delays
- Missing Deadlines and loss of business
- Loss of reputation among accounts
- Increased new recruitment cost and high HR department overheads
- Further additional Training and Development expenses
- Negative motivation levels among current employees

Why HR Analytics intervention?

To understand :

- Factors for curbing attrition rate from growing
- What changes are to be made in workplace environment to minimise attrition
- Priorities among factors attributing high attrition to attend to them immediately.

Objective of case study:

- Develop a model to find the probability of attrition
- Method for designing model - Logistic Regression

The logistic regression modelling method helps us:

- Understand major drivers for high attrition rate
- What fine tuning is required in workplace settings, to keep the employees from quitting.

About the dataset:

- The Manager Survey Data – EmployeeID is the primary key
- The Employee Survey Data – EmployeeID is the primary key
- In-Time Data – Collected from company's attendance Log sheet, EmployeeID is the primary key
- Out-Time Data – Collected from company's attendance Log sheet, EmployeeID is the primary key
- General Data – General data includes employee 's personal data along with education and their satisfaction level for association with XYZ org, etc.

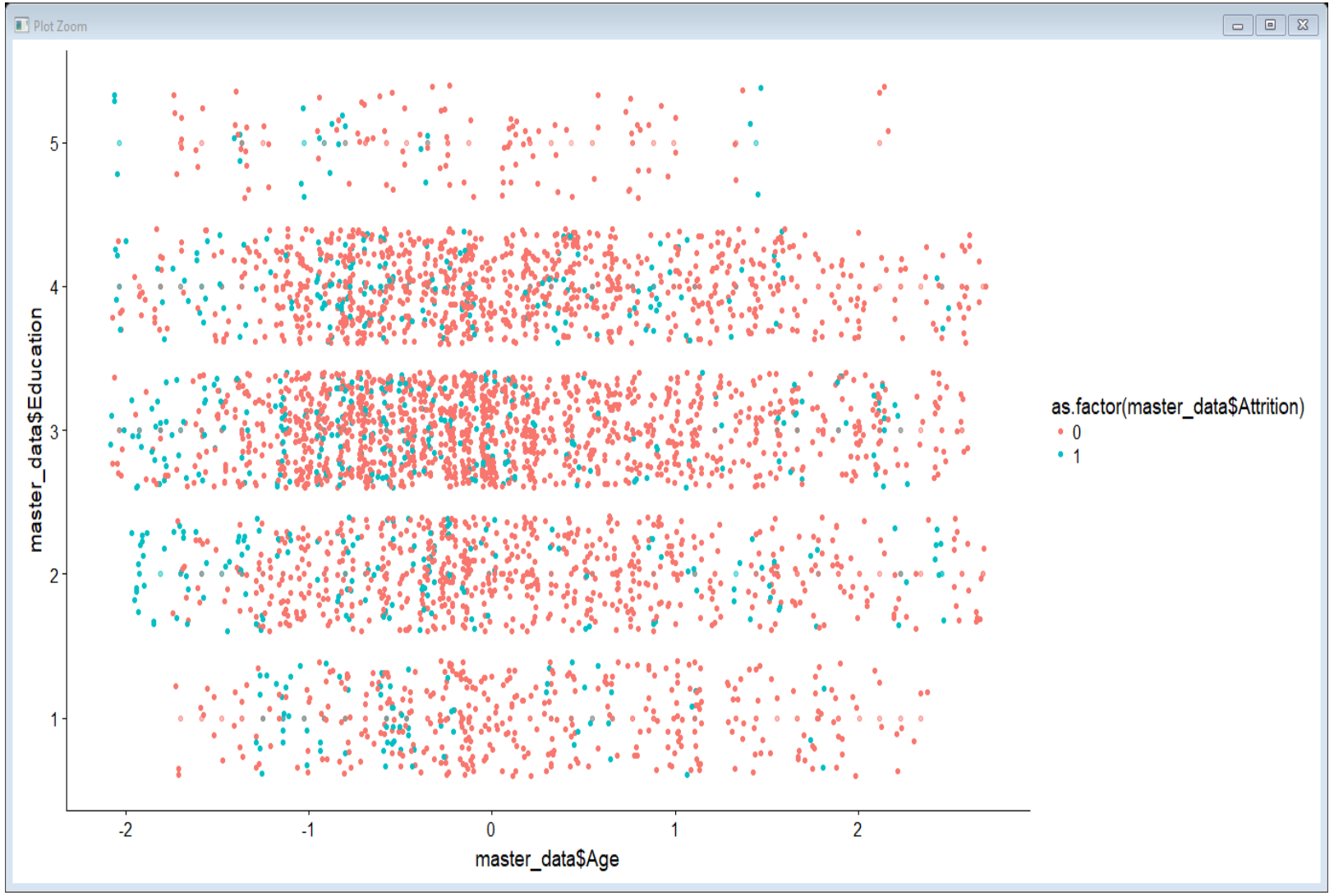
Assumptions:

In the in_time and out_time files, the columns with all missing values or NA are considered as holiday.

Building the model in RStudio:

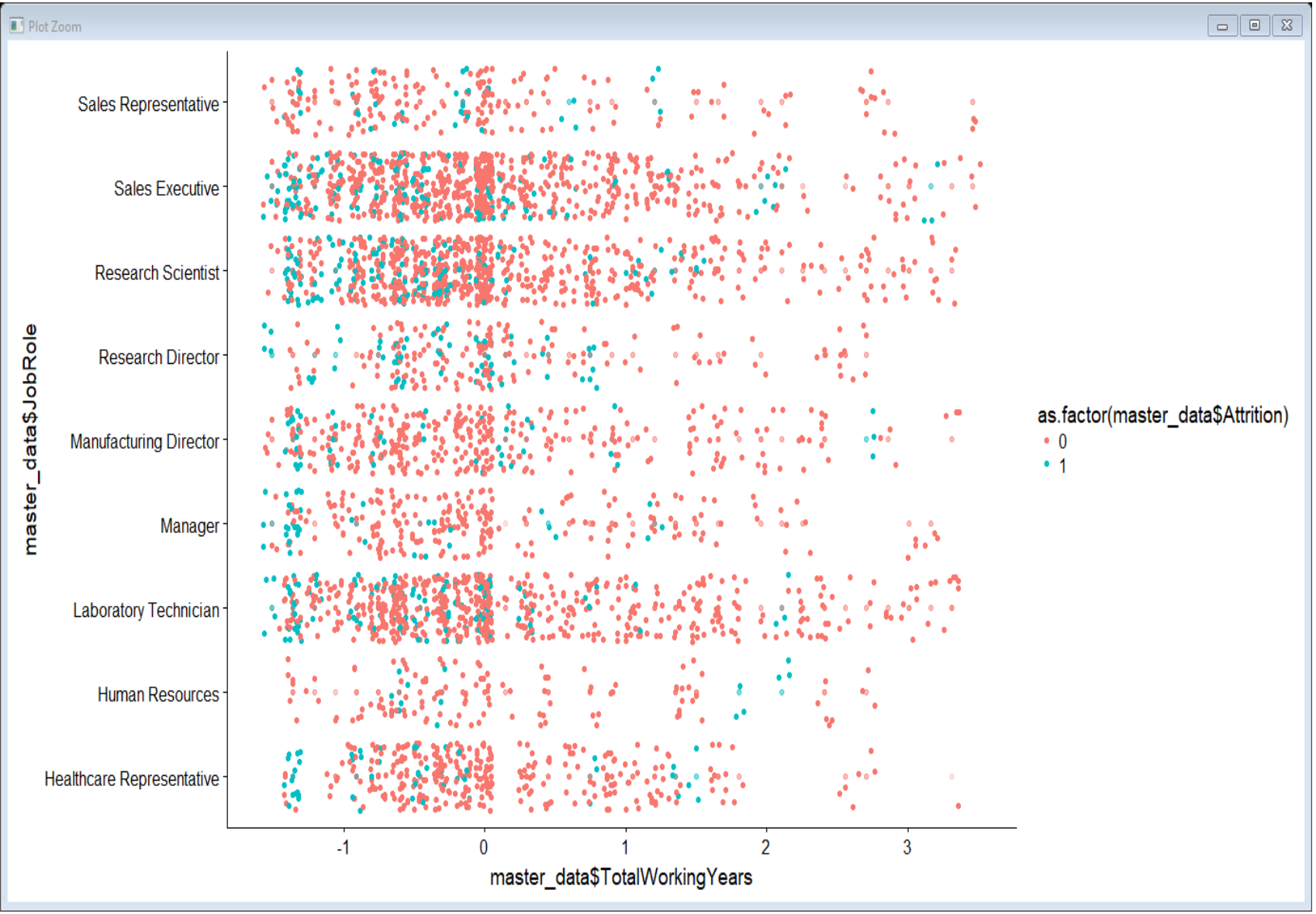
- Data Import – Importing data from files onto R.
- Data Cleansing –Handling mostly NA values and other inapplicable data.
- Data Merging – Merge all 5 files to get amalgamated dataset based on the unique key.
- Creation of dummy variables for categorical data
- Splitting the data into training and test data in the ratio 70:30
- Create model using ‘glm’ function
- Use stepAIC method to find the most preferable variables
- Reconstruct model using ‘glm’ until the p-value and vif for all variables are significantly low
- Use the final model on the test data set to find the predicted value
- Calculate the accuracy, sensitivity and specificity

While Analysing the data through EDA, these were the observations:



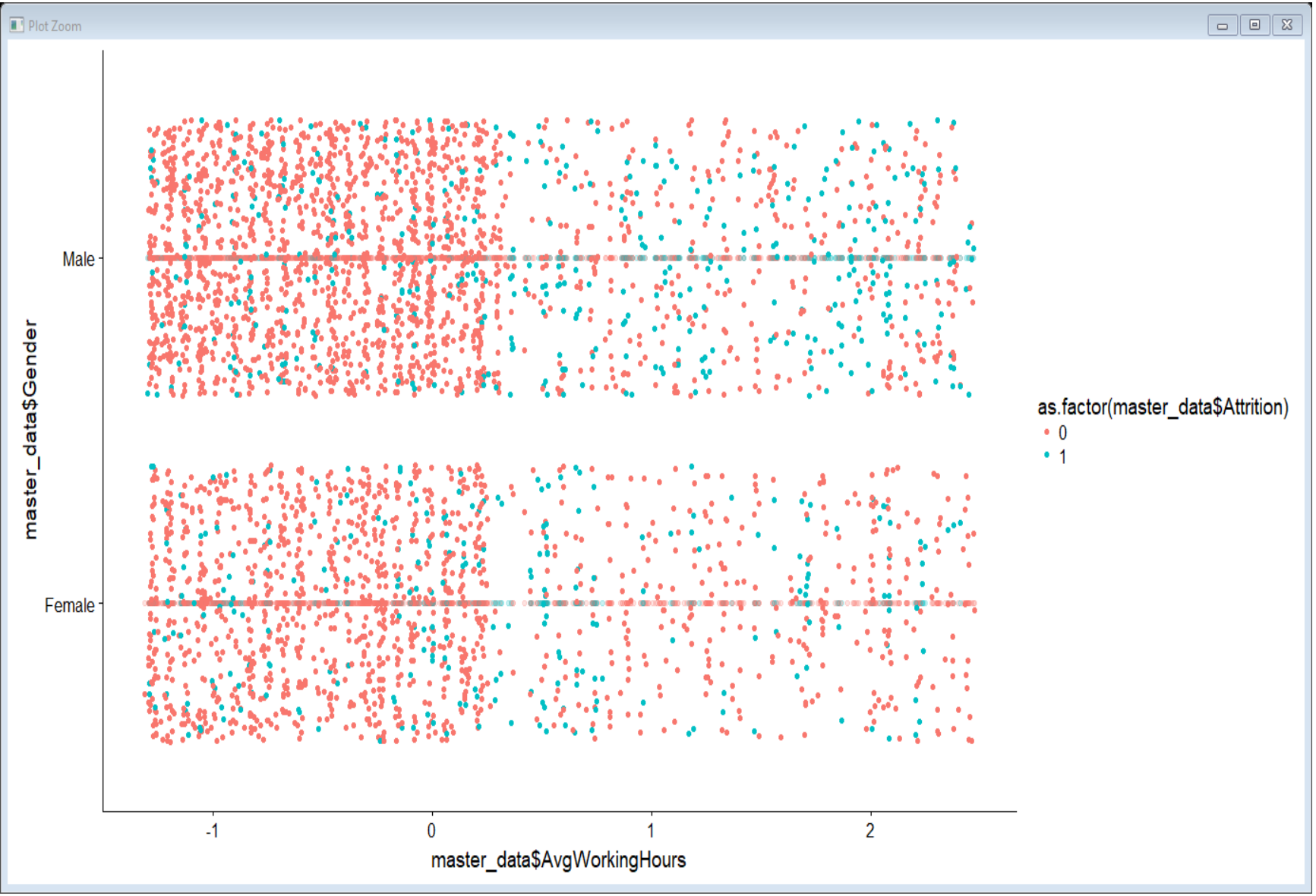
Inference: It appears many employees who quit are having education below Master's.

While Analysing the data through EDA, these were the observations:



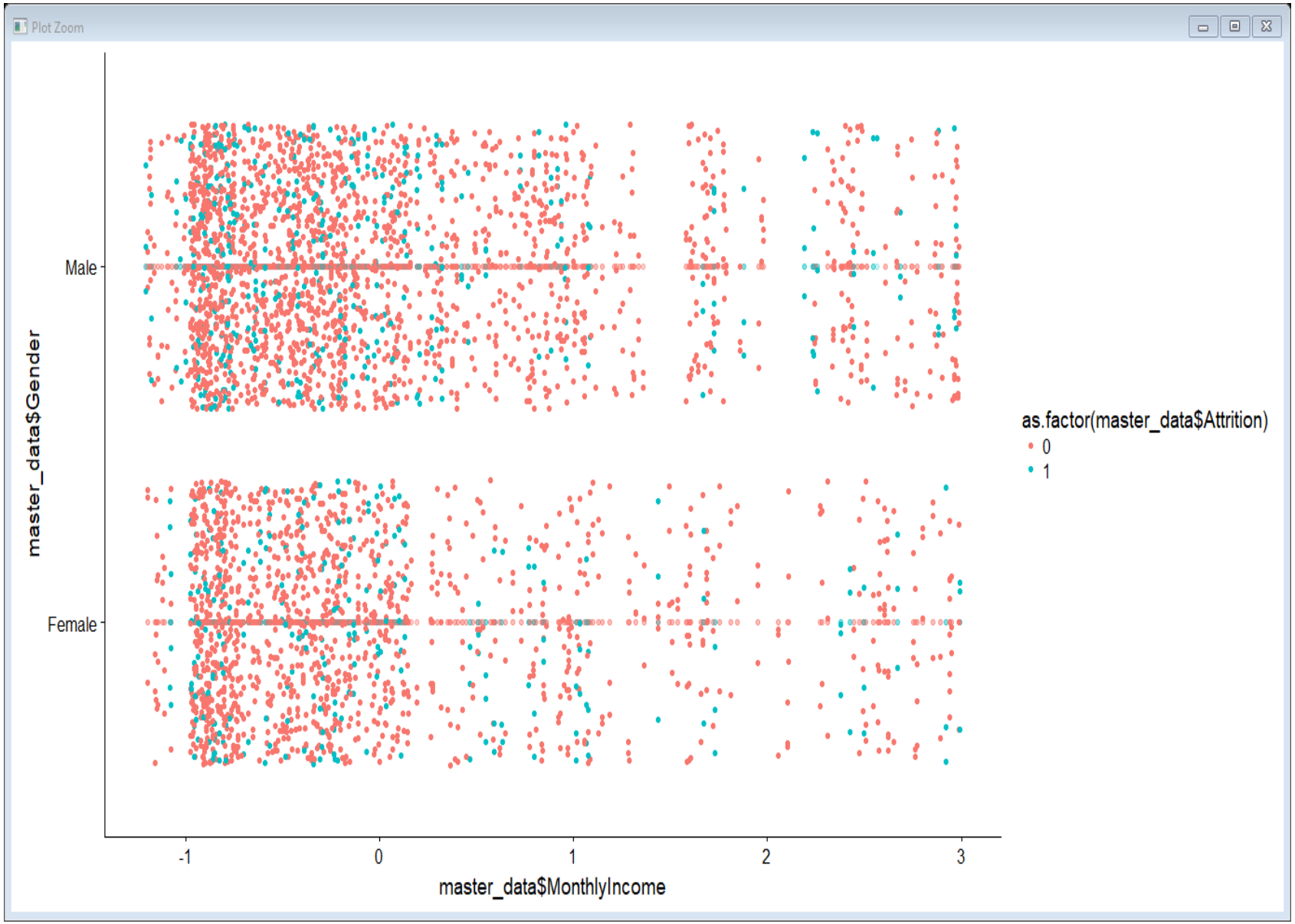
Inference: it appears many employees with less than 10 years experience are likely to quit

While Analysing the data through EDA, these were the observations:



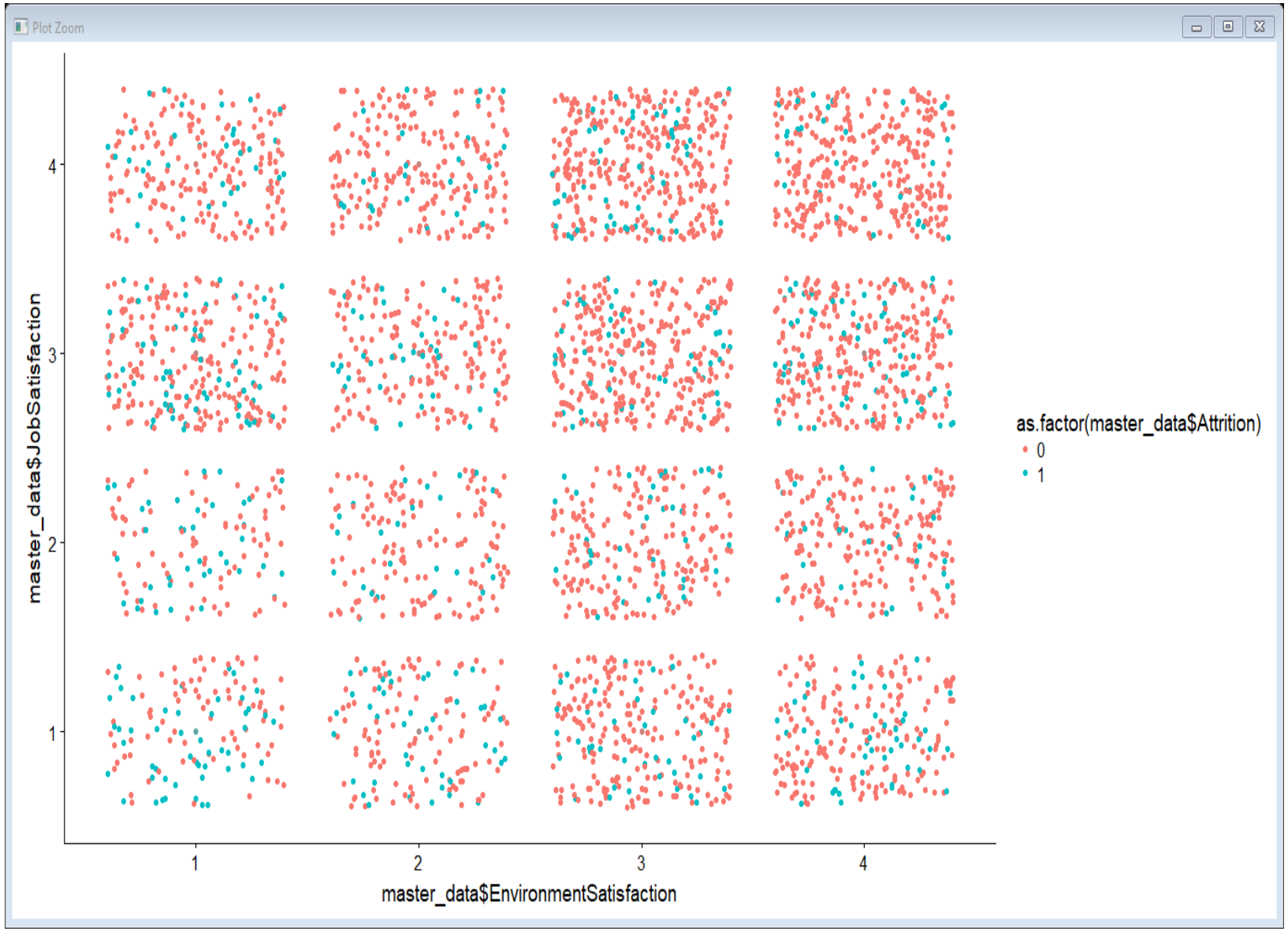
Inference: higher the working hours(average), higher is the attrition

While Analysing the data through EDA, these were the observations:



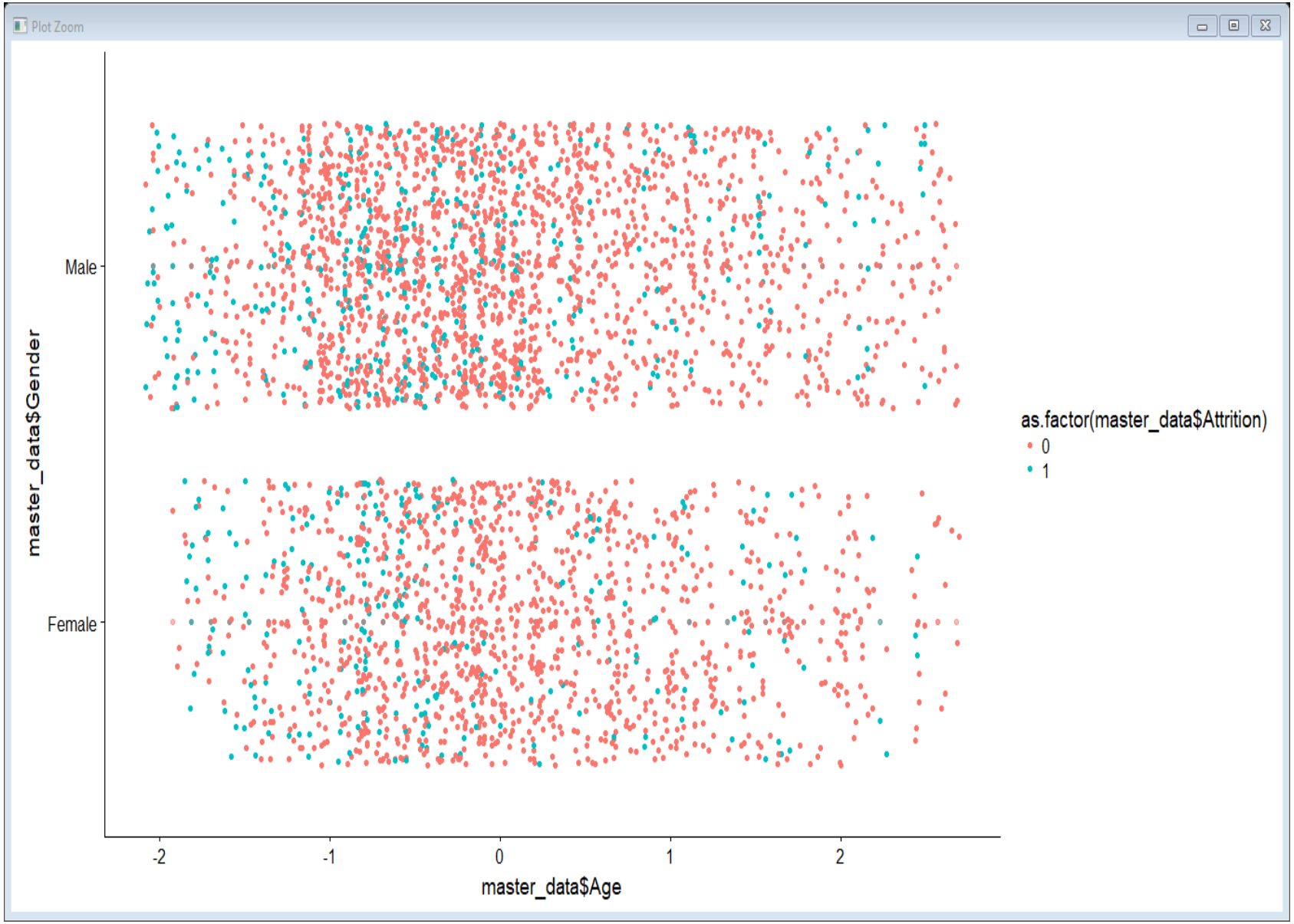
Inference: employees with Lower monthly income quit, may be to find a better job opportunity

While Analysing the data through EDA, these were the observations:



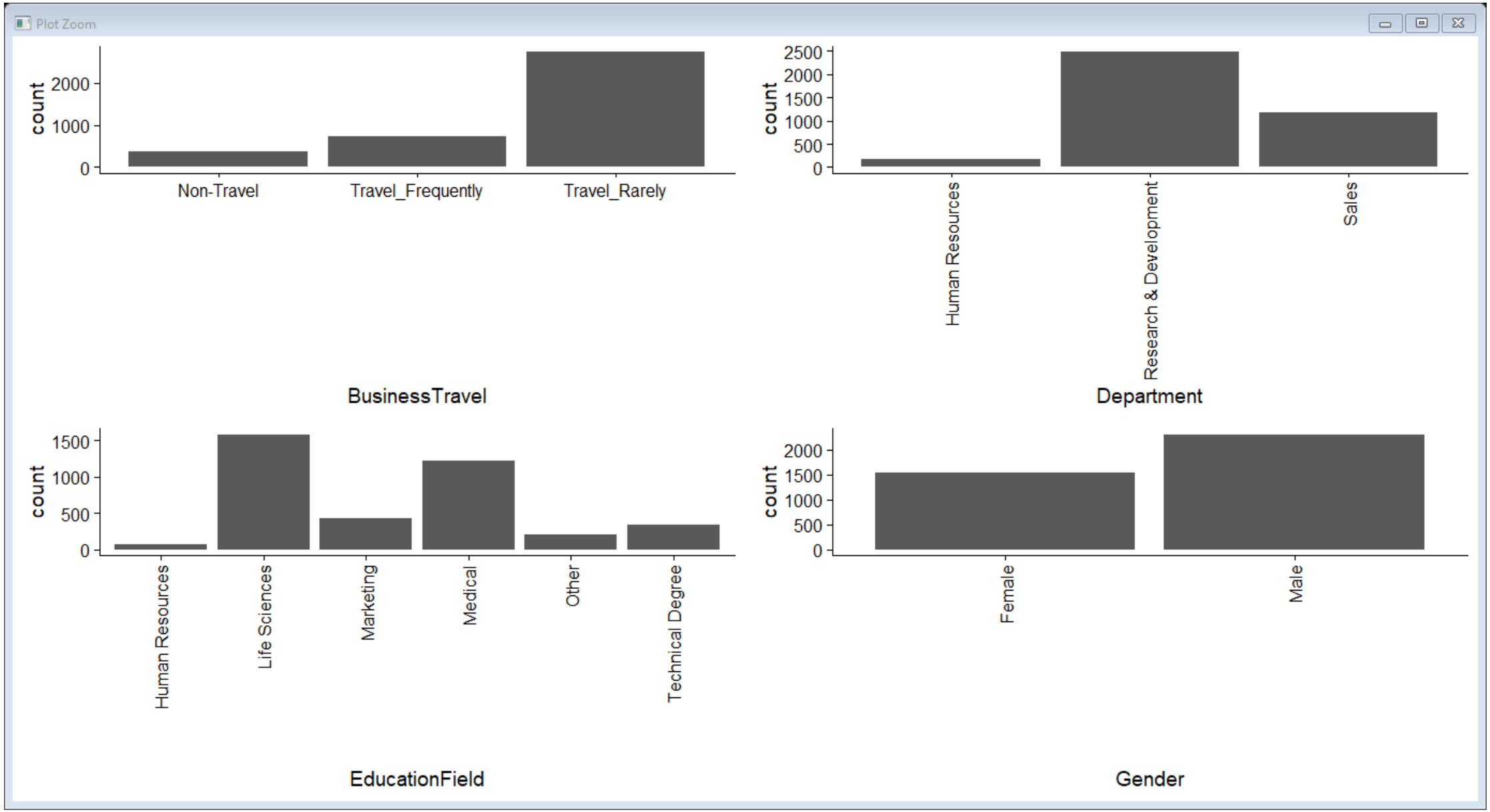
Inference: low job satisfaction and environment satisfaction is leading to high attrition

While Analysing the data through EDA, these were the observations:

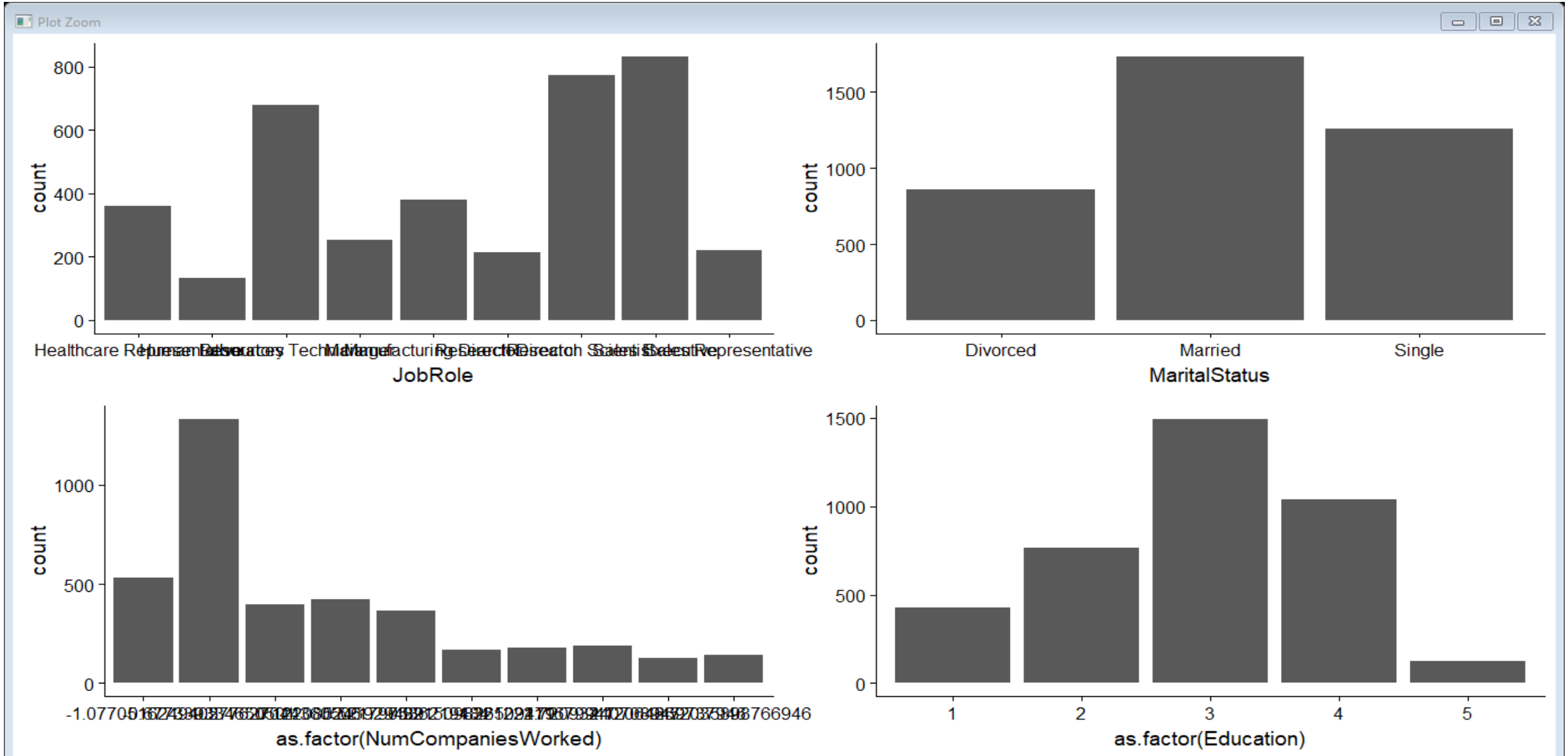


Inference: Attrition seems to subside over time. More Aged an employee, lesser is the Attrition

While Analysing the data through EDA, these were the observations:



While Analysing the data through EDA, these were the observations:



Final Logistic Regression Model

We are left with 14 predictor variables now.

1. JobRole.xManufacturing.Director
2. Age
3. TrainingTimesLastYear
4. EnvironmentSatisfaction.x3
5. EnvironmentSatisfaction.x2
6. NumCompaniesWorked
7. YearsWithCurrManager
8. TotalWorkingYears
9. YearsSinceLastPromotion
10. MaritalStatus.xSingle
11. BusinessTravel.xTravel_Frequently
12. JobSatisfaction.x4
13. EnvironmentSatisfaction.x4
14. AvgWorkingHours

Prediction based on final model:

	Reference	
Prediction	No	Yes
No	717	48
Yes	243	146

→ Confusion Matrix

Accuracy : 0.7478

95% CI : (0.7217, 0.7727)

No Information Rate : 0.8319

P-Value [Acc > NIR] : 1

Kappa : 0.3565

Mcnemar's Test P-Value : <2e-16

→ Prediction Statistics

Sensitivity : 0.7526

Specificity : 0.7469

Pos Pred Value : 0.3753

Neg Pred Value : 0.9373

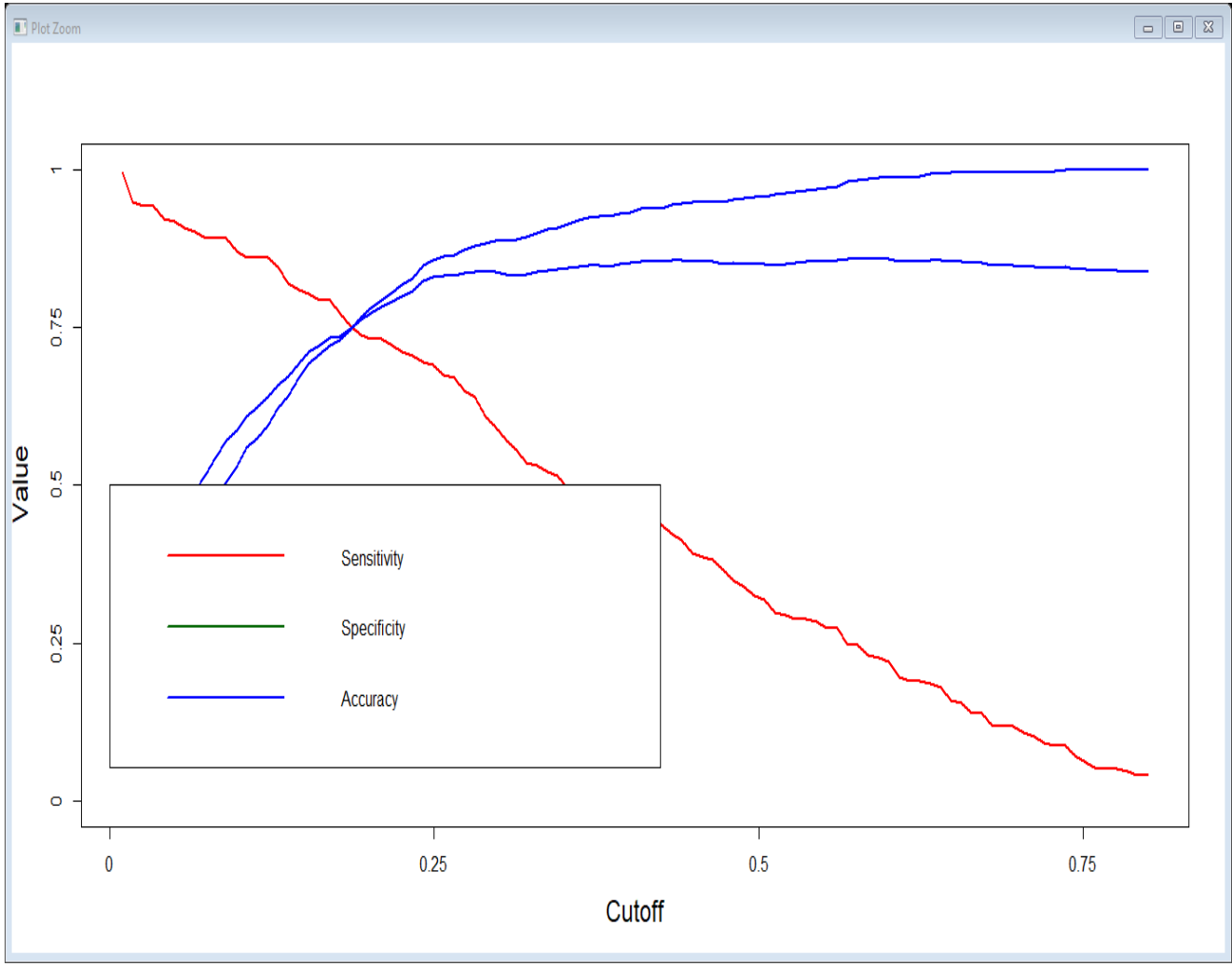
Prevalence : 0.1681

Detection Rate : 0.1265

Detection Prevalence : 0.3371

Balanced Accuracy : 0.7497

'Positive' Class : Yes



Model Assessment Summary:

Using the KS Statistic Method:

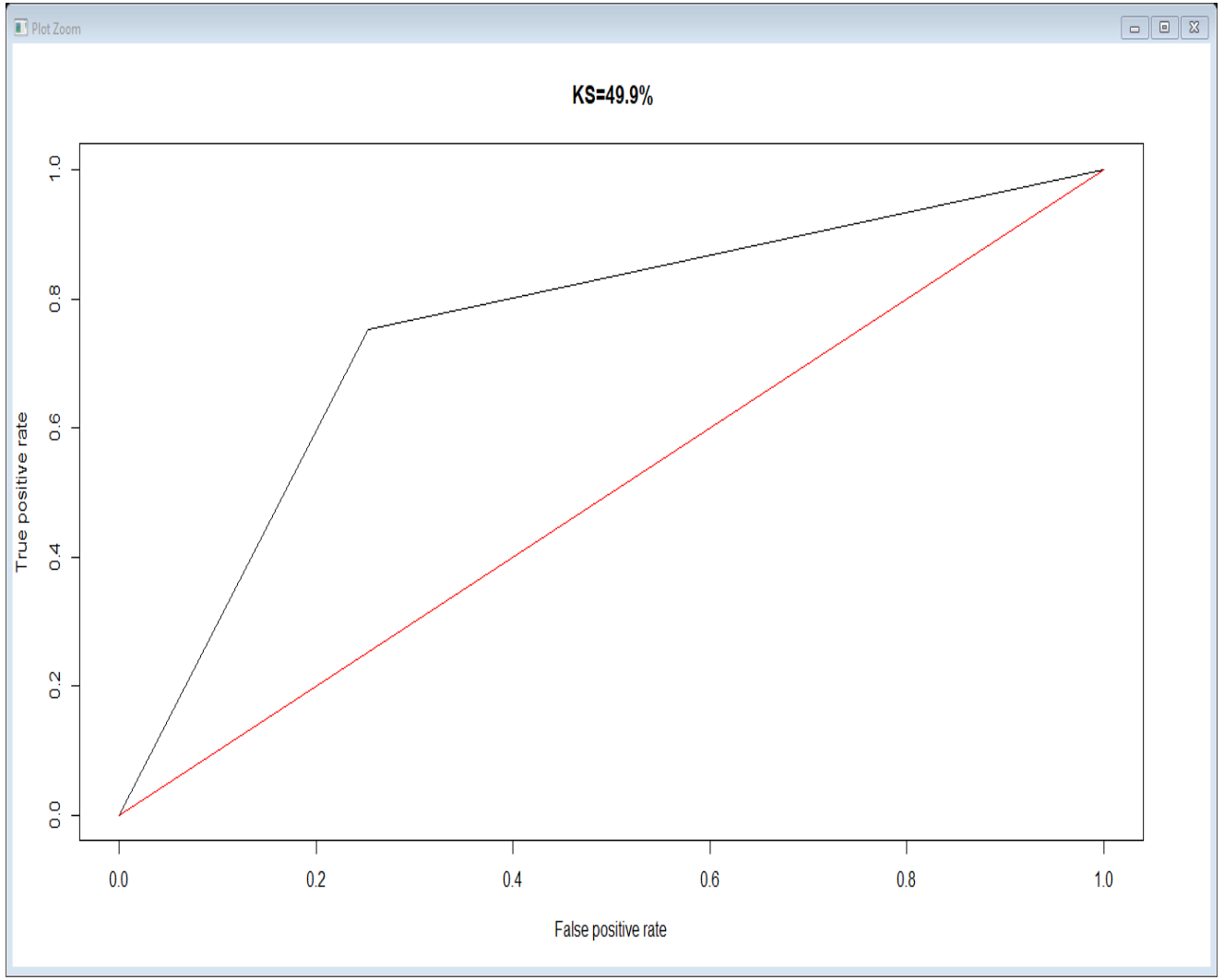
- KSS value = 0.4995 i.e. ~ 50% which is > 40%

Also,

- Area under the curve = 0.7497

These values indicate that the final model obtained is a fairly good model.

Our final model is ~75% close to ideal model.



Recommendations:

- Environment Satisfaction and Job Satisfaction, the better these are for employees the less are their chances of leaving the company.
- The more an employee works overtime on an average the more are the chances that he/she will leave the company.
- If an employee works with the same manager for a longer period of time the lesser are the chances that employee will leave the company.
- Hire people with more experience as they are less likely to leave the company. But if the person has worked in many companies then the chances that he/she will leave the company increases.
- Employees who are unmarried are prone to leaving the company.

Thank
you