Abstract

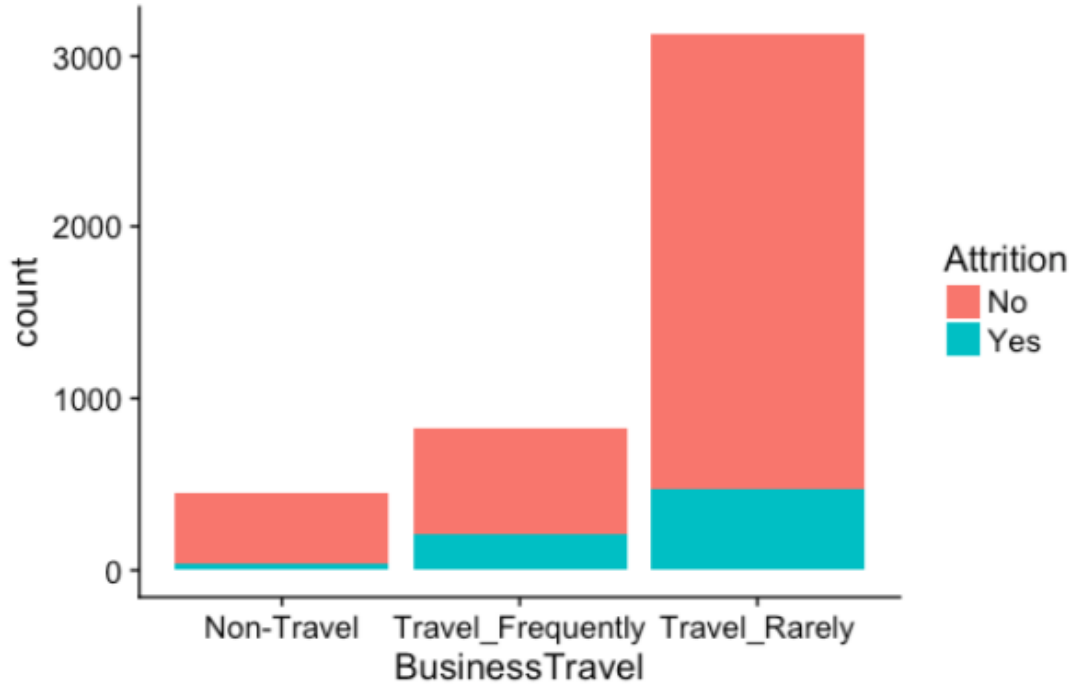
- The objective of this case study is to identify the important factors causing the attrition of employees
- Based on the identified factors, the management will take steps to control the attrition

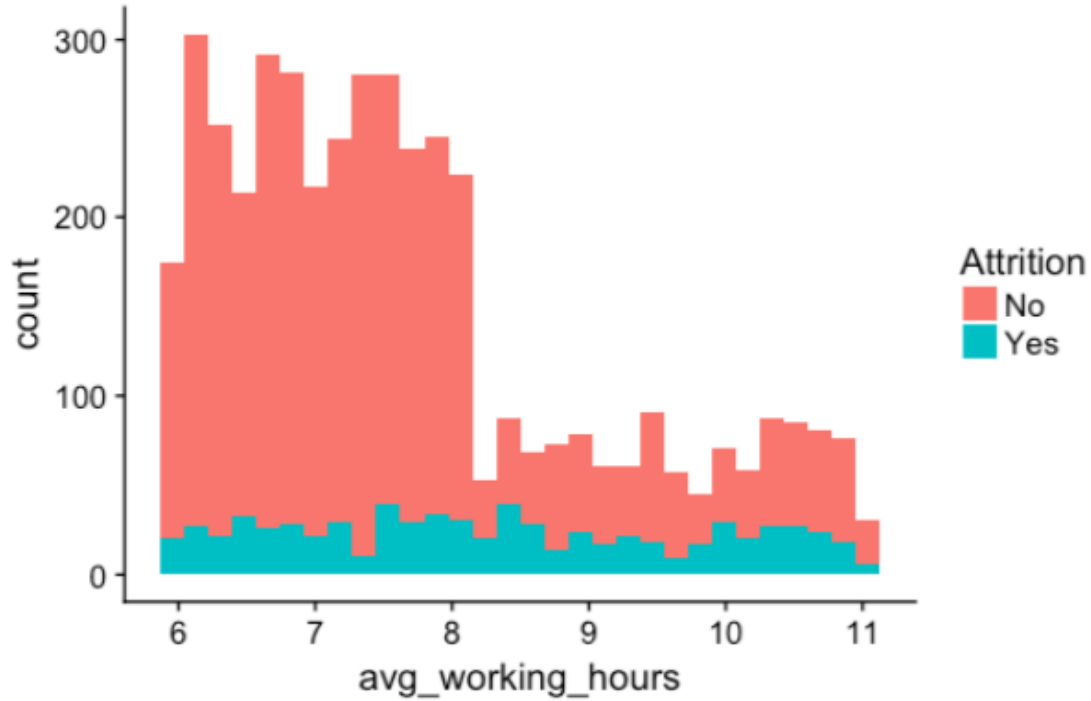
Problem Solving Methodology

- Understand the business requirement – Identify factors causing attrition
- Understand the data – Files contain employee in & out time, manager feedback, employee feedback and general data
- Clean the data
- Check for duplicates
- Impute the missing values – EnvironmentSatisfaction, JobSatisfaction, WorkLifeBalance, TotalWorkingYears, NumCompaniesWorked
- Derive new metrics - Average working hours, Total leaves.
- Perform EDA to understand the data better
- Convert categorical variables into numeric variables
- Discard insignificant variables
- Build the logistic regression model
- Train the model
- Evaluate the accuracy of the model

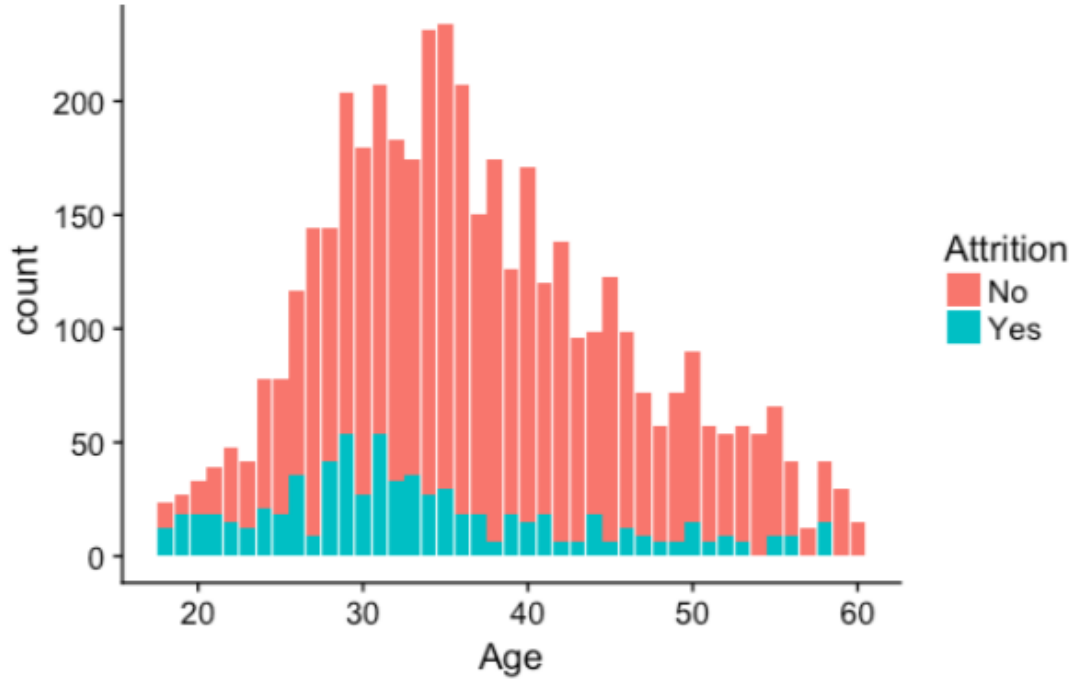
EDA Observations

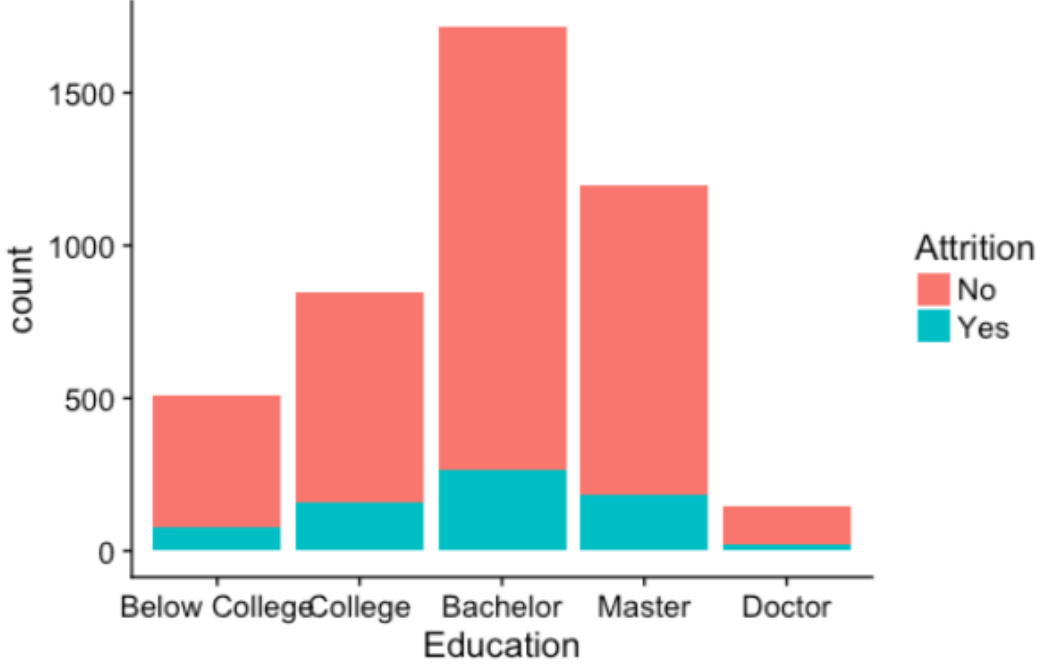
- Segmented Univariate analysis performed on Attrition by all the categorical variables
- There is a strong correlation between TotalWorkingYears and Age
- YearsSinceLastPromotion, Age, Average working hours could be significant factors

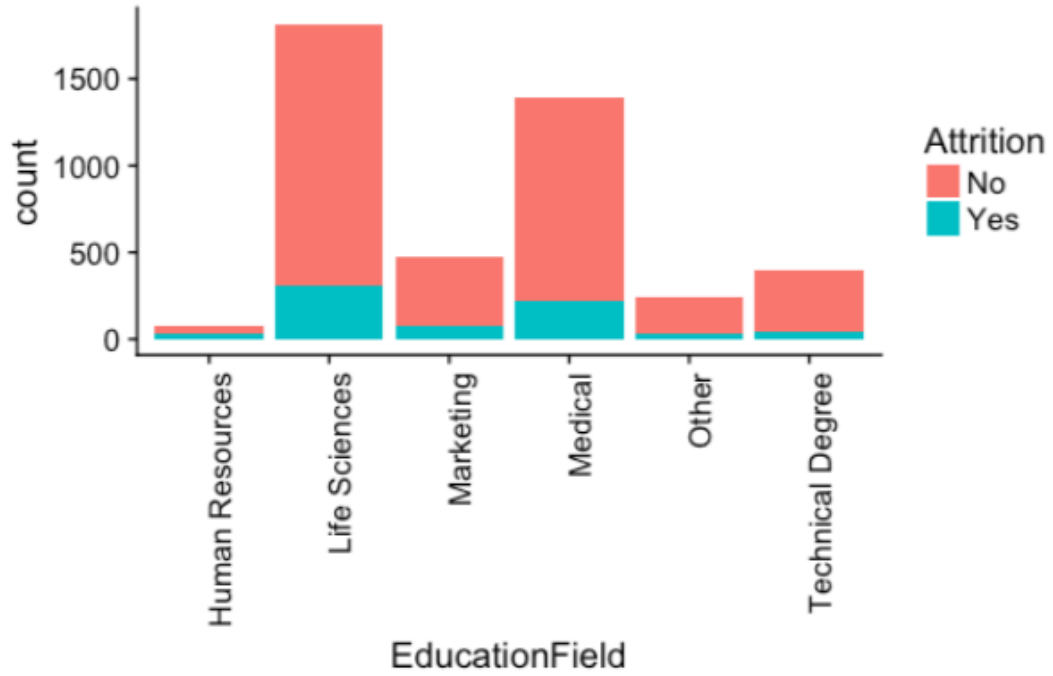

3000
2000
1000
0
count
Non-Travel
Travel_Frequently
Travel_Rarely
BusinessTravel
Attrition
No
Yes

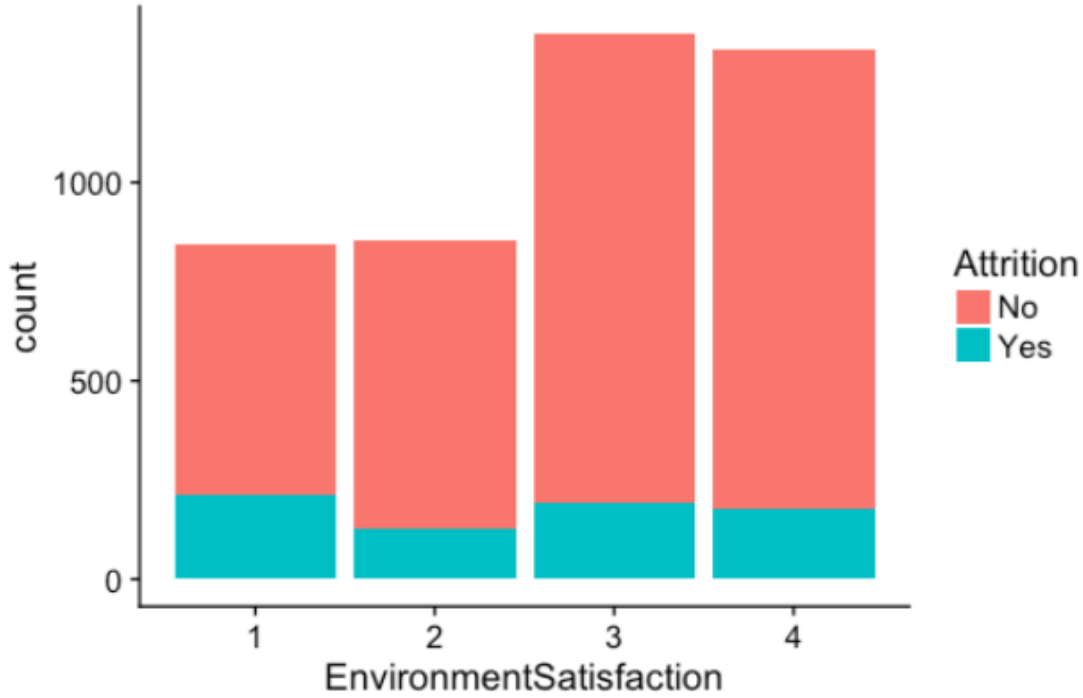

300
200
100
0
count
6
7
8
9
10
11
avg_working_hours
Attrition
No
Yes

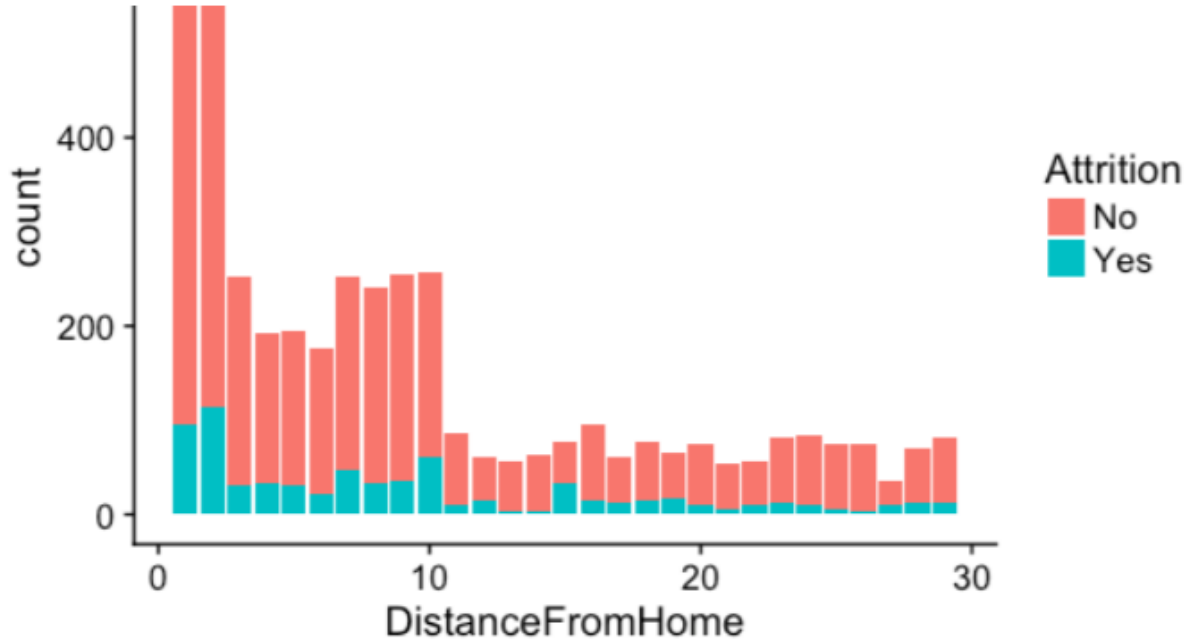

3000
2000
1000
0
count
Human Resources
Research & Development
Sales
Department
Attrition
No
Yes

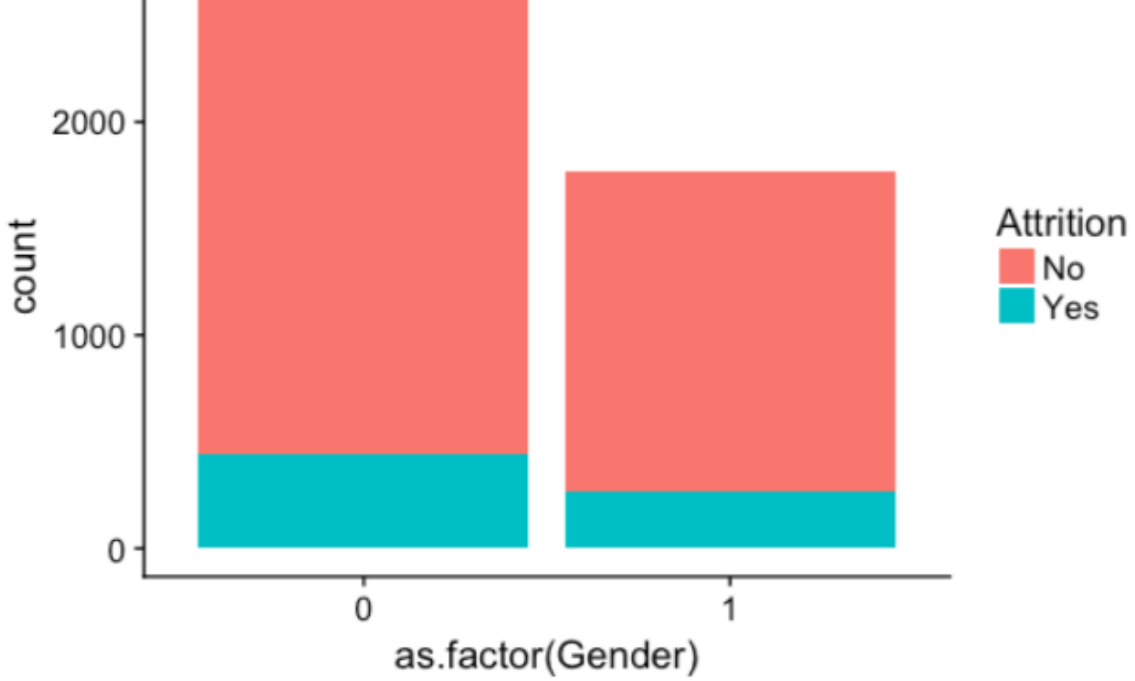

200
150
100
50
0
count
20
30
40
50
60
Age
Attrition
No
Yes

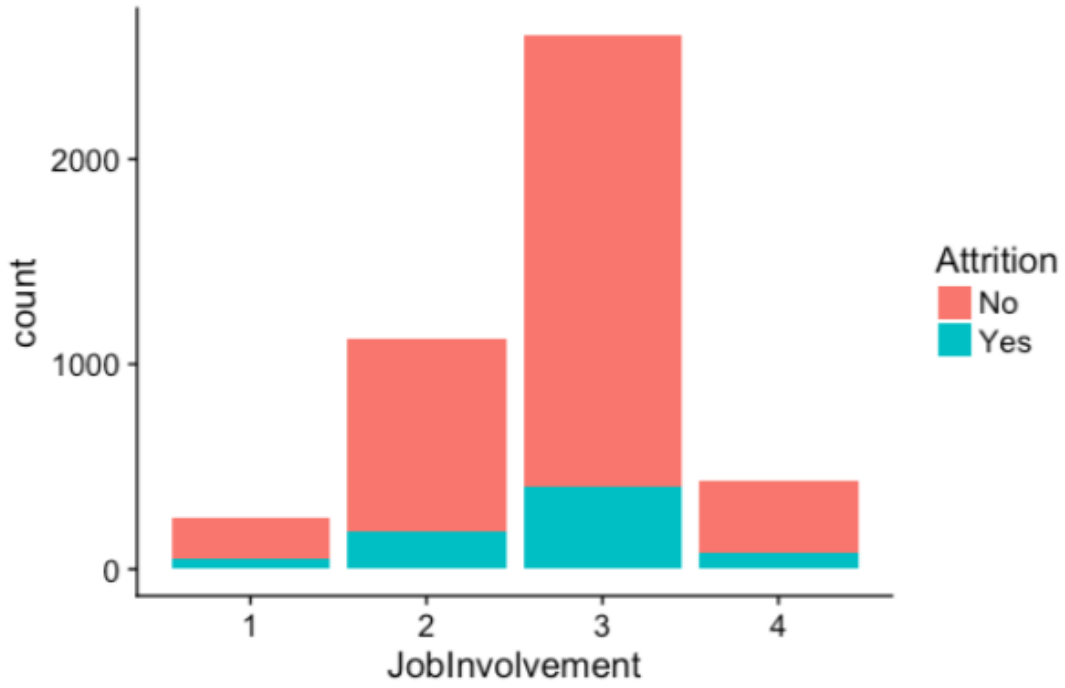

count
1500
1000
500
0
Below College
College
Bachelor
Master
Doctor
Education
Attrition
No
Yes

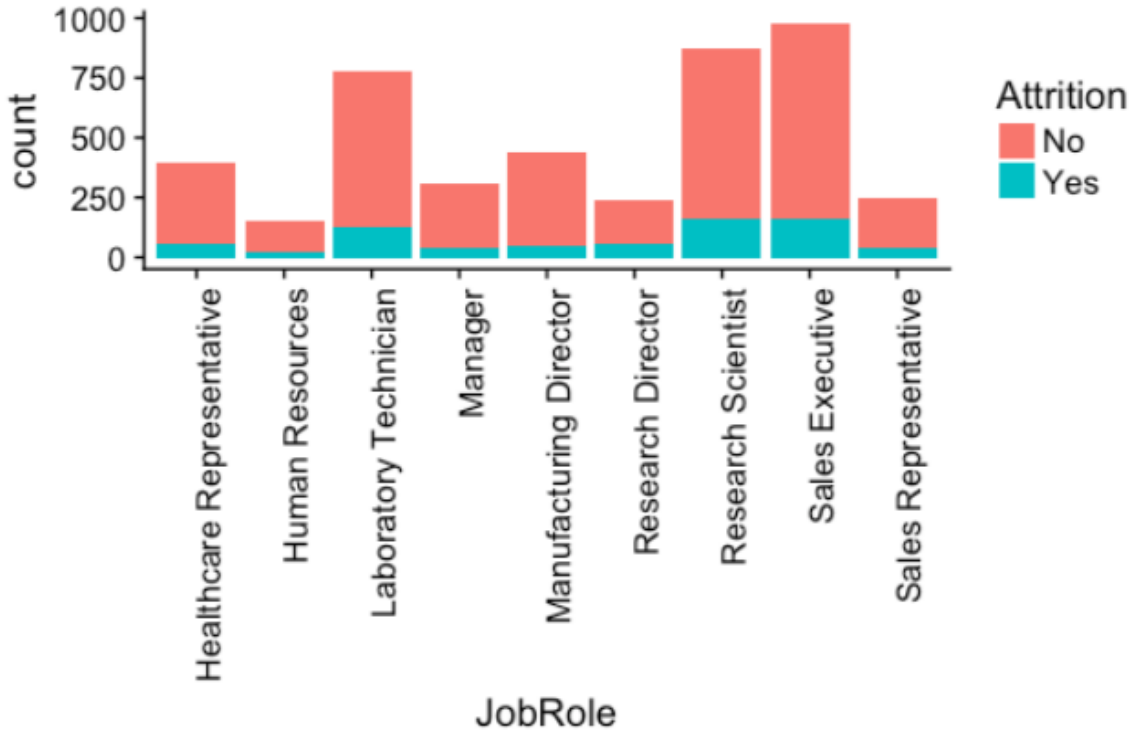

count
1500
1000
500
0
Human Resources
Life Sciences
Marketing
Medical
Other
Technical Degree
EducationField
Attrition
No
Yes

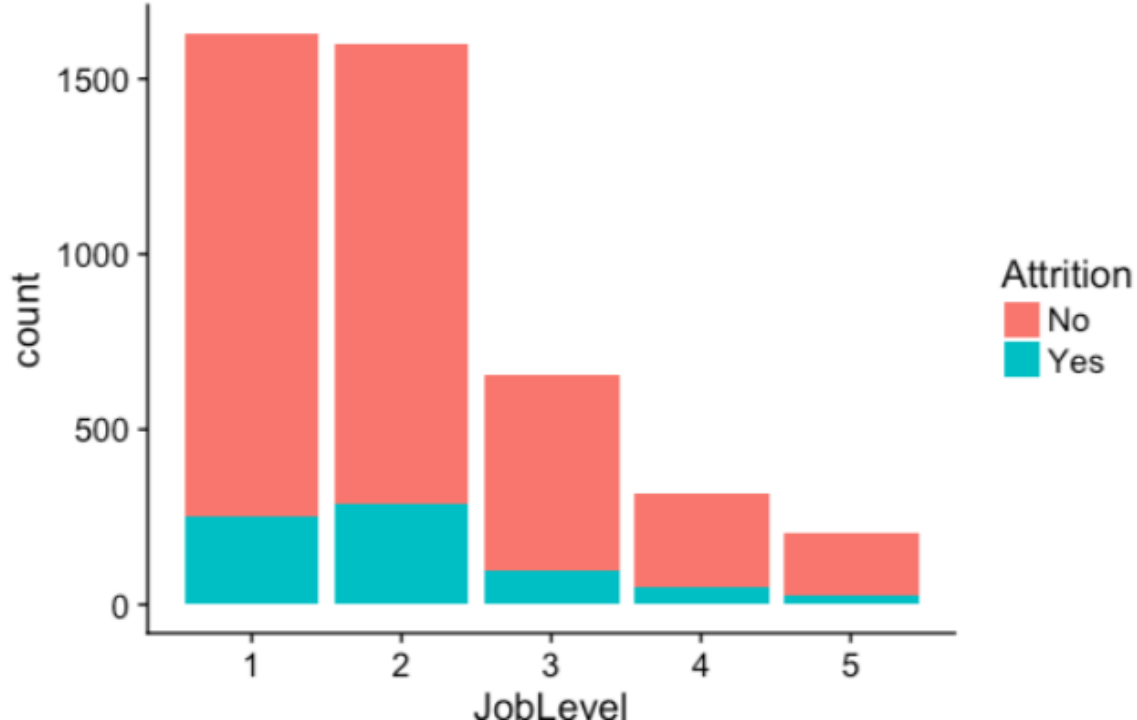

count
1000
500
0
1
2
3
4
EnvironmentSatisfaction
Attrition
No
Yes

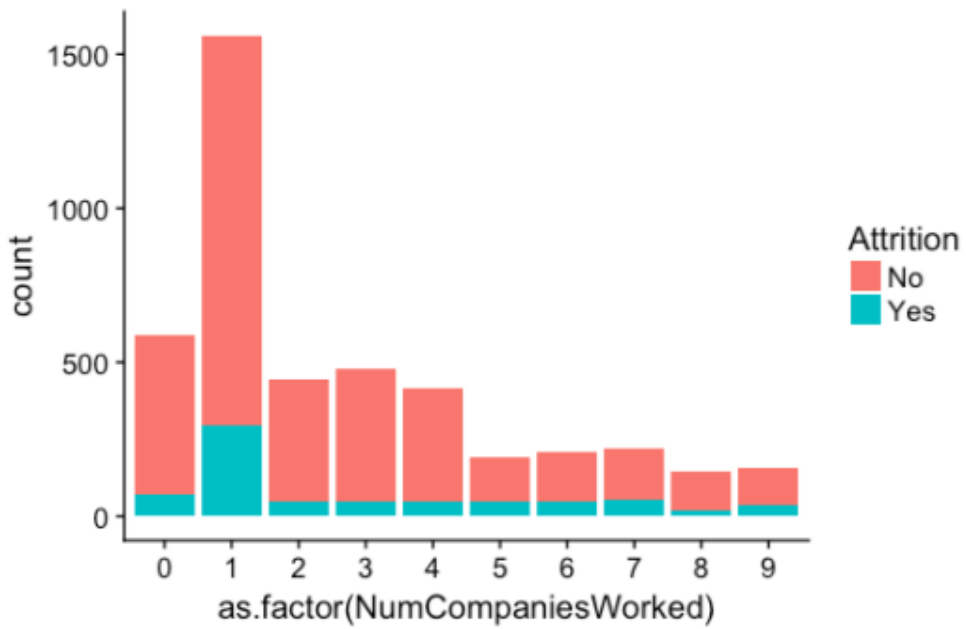

count
400
200
0
0
10
20
30
DistanceFromHome
Attrition
No
Yes

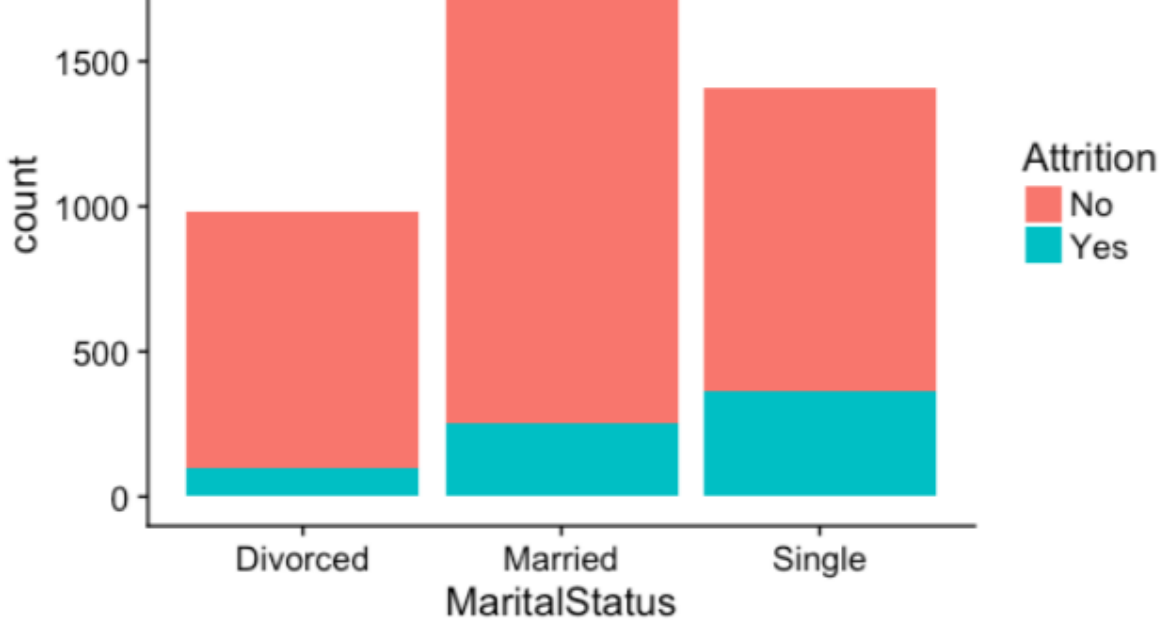

count
2000
1000
0
0
1
as.factor(Gender)
Attrition
No
Yes

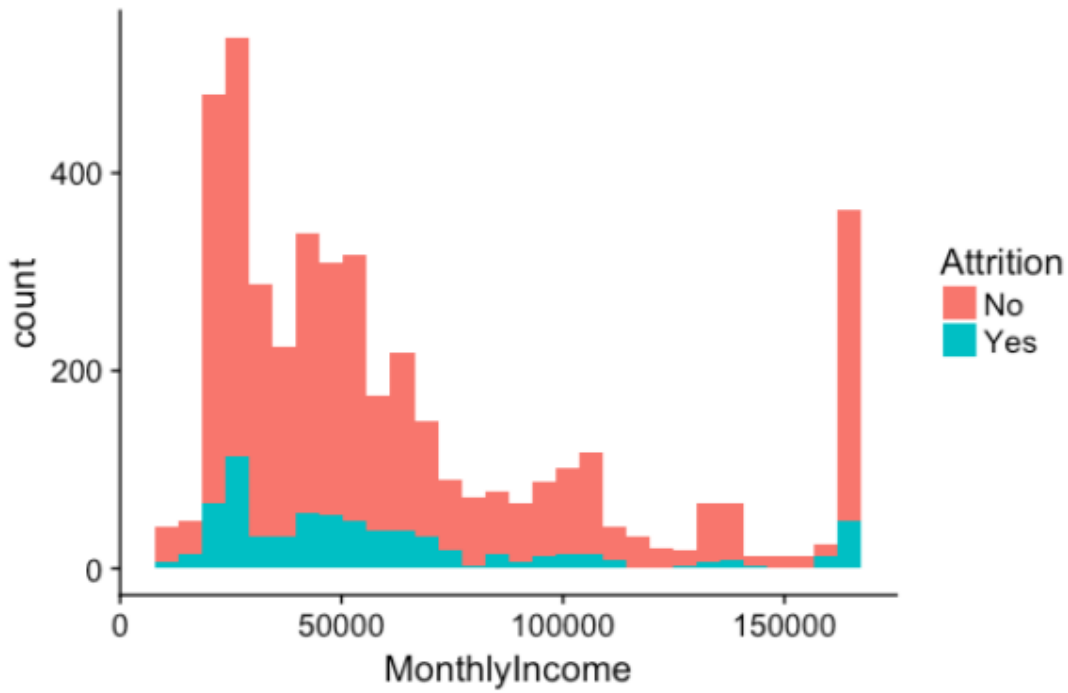

count
2000
1000
0
1
2
3
4
JobInvolvement
Attrition
No
Yes

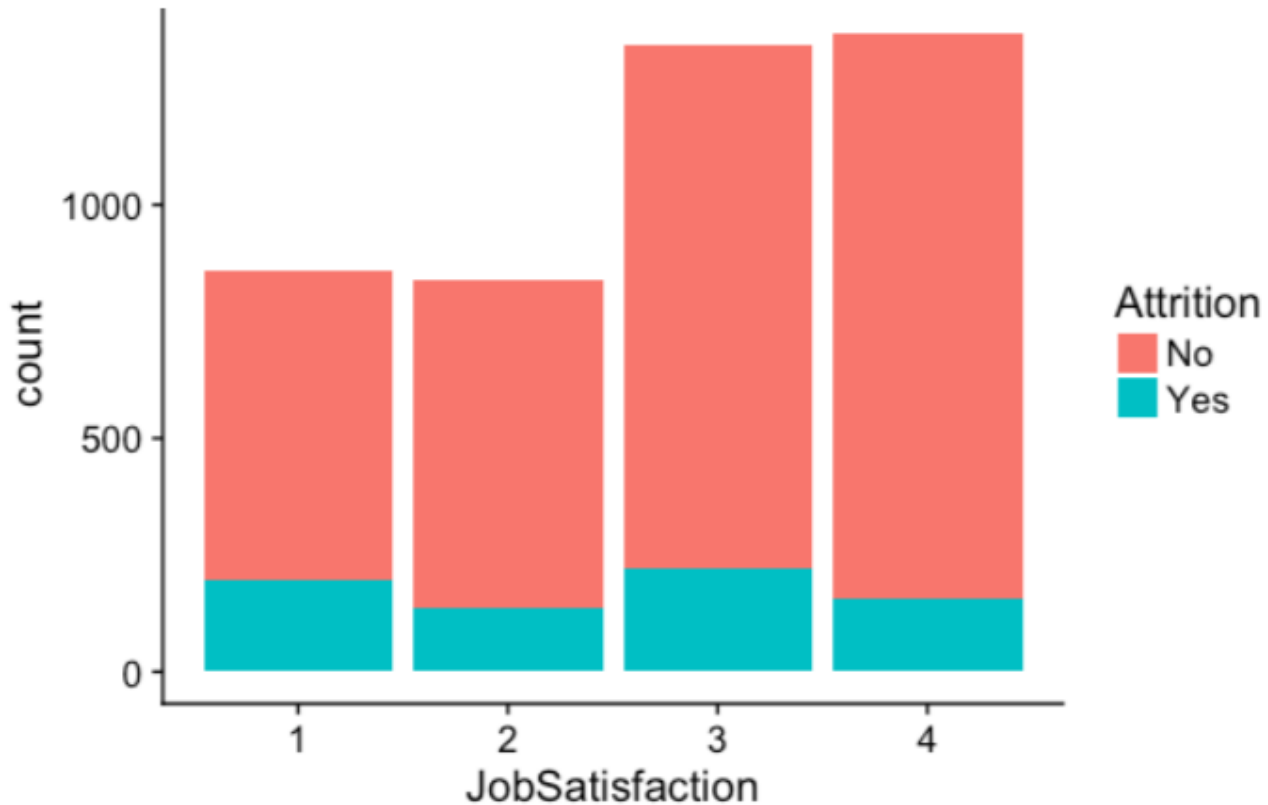

count
1000
750
500
250
0
Healthcare Representative
Human Resources
Laboratory Technician
Manager
Manufacturing Director
Research Director
Research Scientist
Sales Executive
Sales Representative
JobRole
Attrition
No
Yes

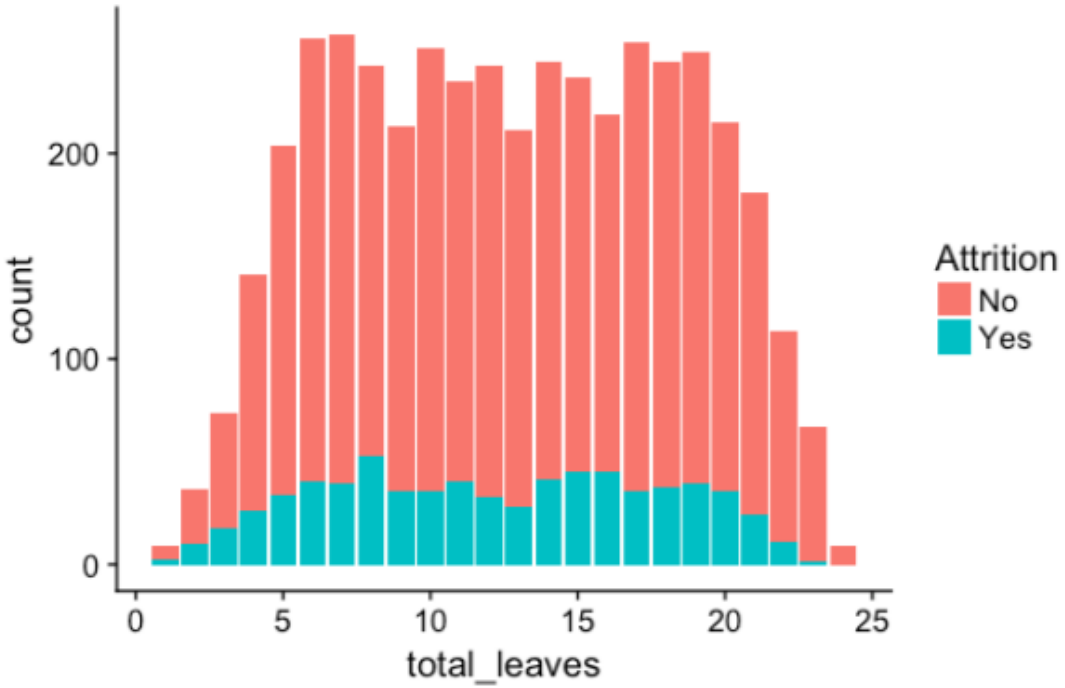

count
1500
1000
500
0
1
2
3
4
5
JobLevel
Attrition
No
Yes

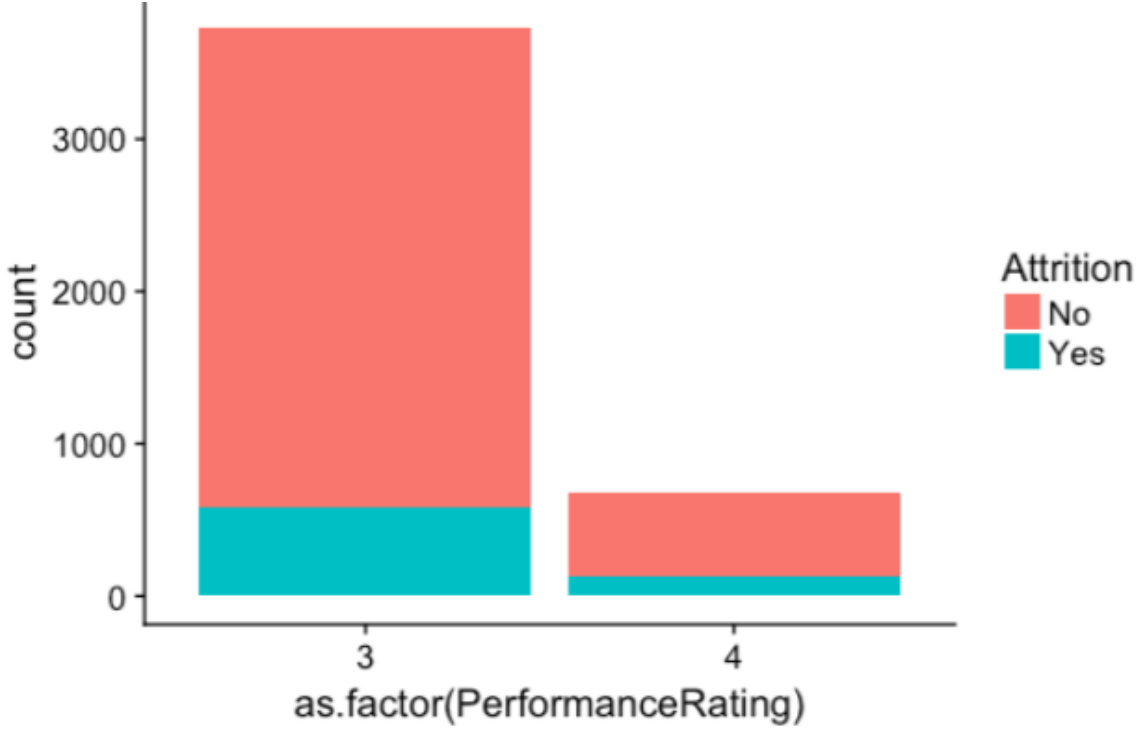

count
1500
1000
500
0
0 1 2 3 4 5 6 7 8 9
as.factor(NumCompaniesWorked)
Attrition
No
Yes

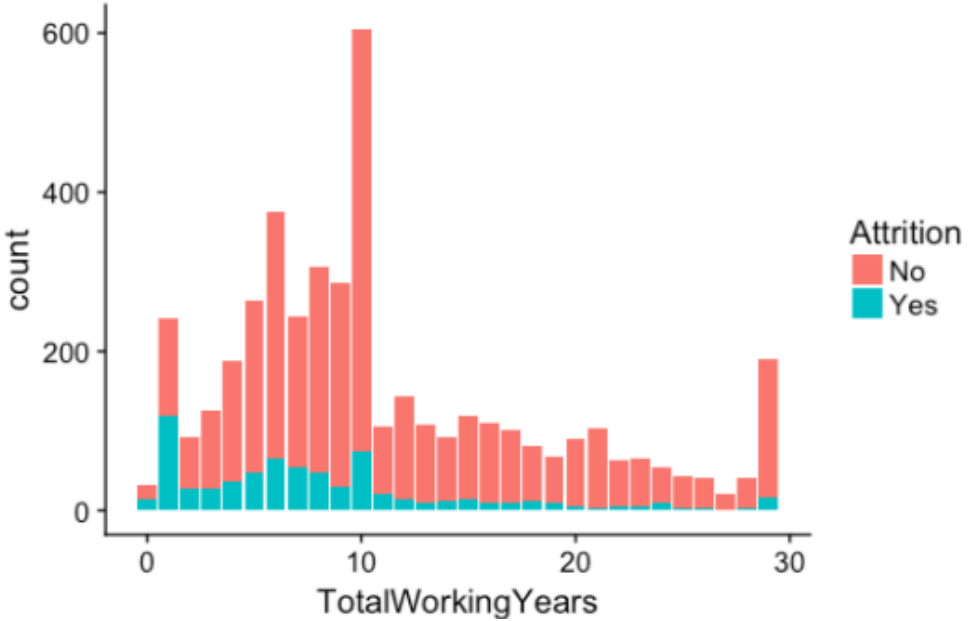

count
1500
1000
500
0
Divorced
Married
Single
MaritalStatus
Attrition
No
Yes

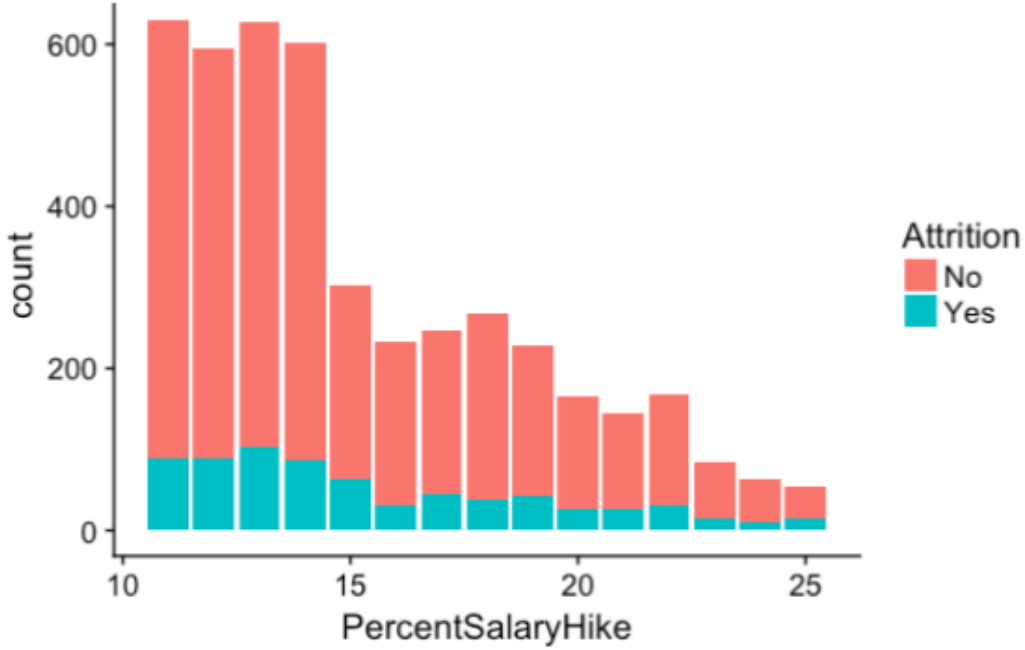

count
400
200
0
0
50000
100000
150000
MonthlyIncome
Attrition
No
Yes

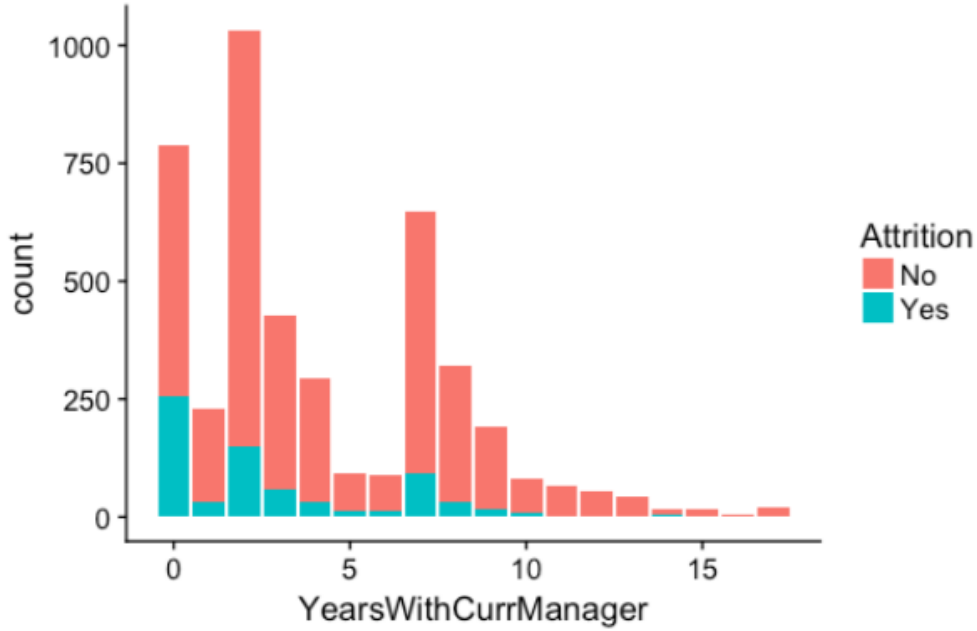

count
1000
500
0
1
2
3
4
JobSatisfaction
Attrition
No
Yes

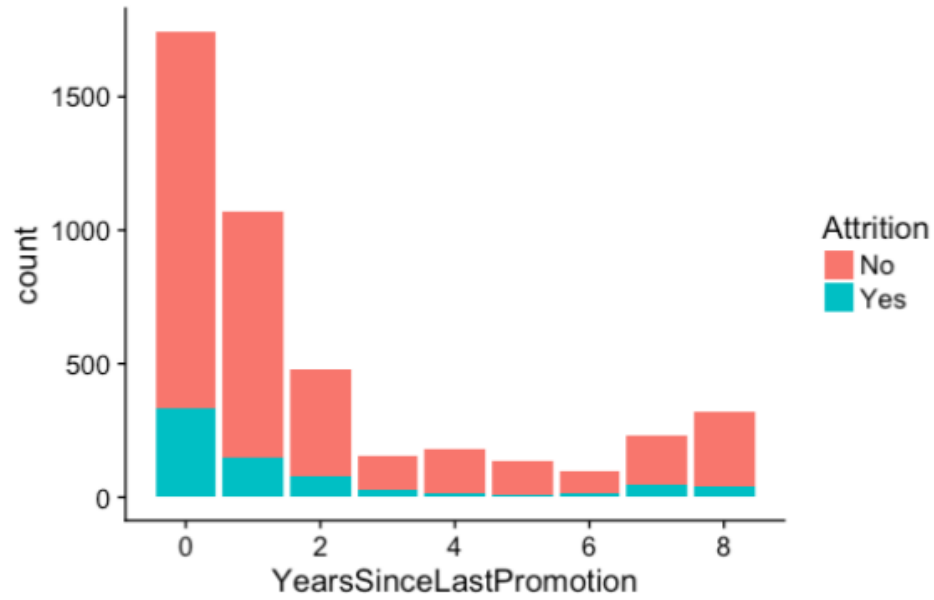

count
200
100
0
0
5
10
15
20
25
total_leaves
Attrition
No
Yes

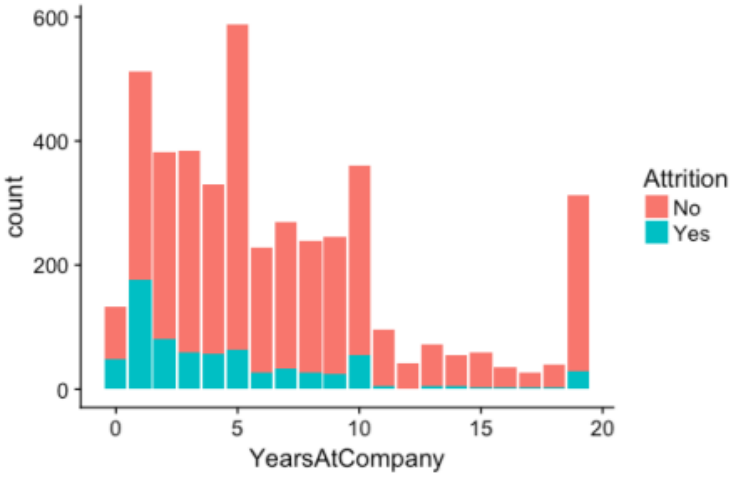

count
3000
2000
1000
0
3
4
as.factor(PerformanceRating)
Attrition
No
Yes

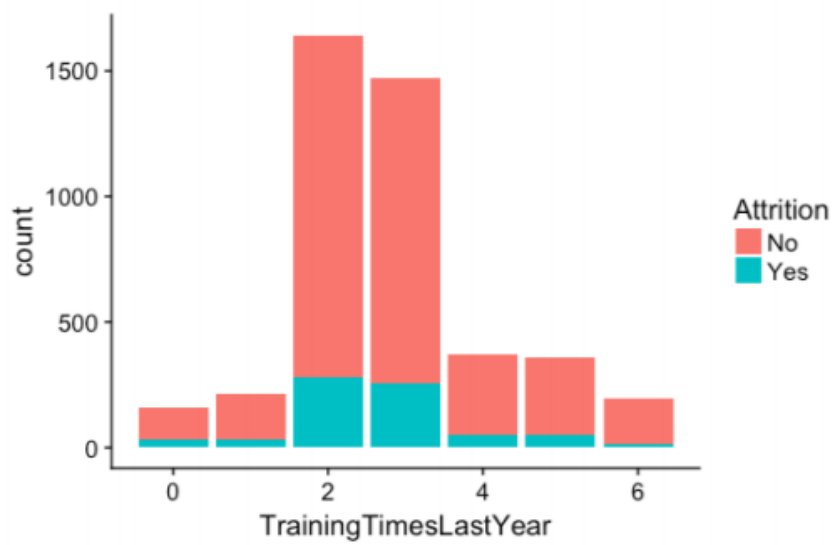

count
600
400
200
0
0
10
20
30
TotalWorkingYears
Attrition
No
Yes

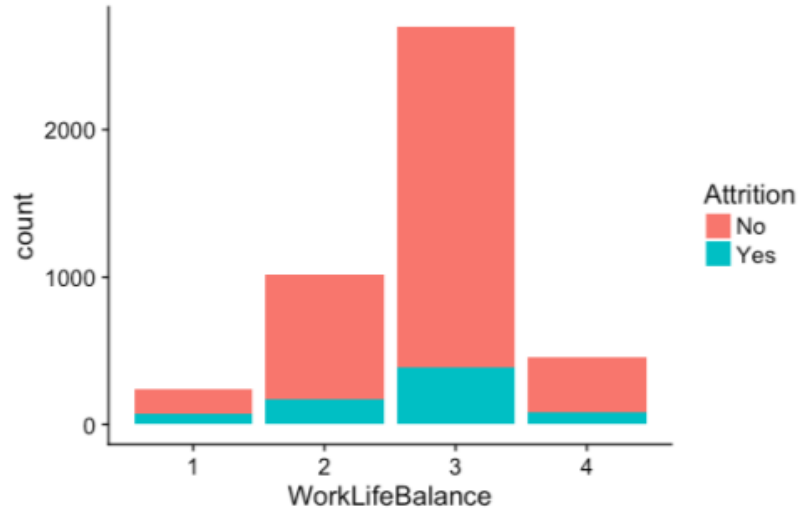

count
600
400
200
0
10
15
20
25
PercentSalaryHike
Attrition
No
Yes

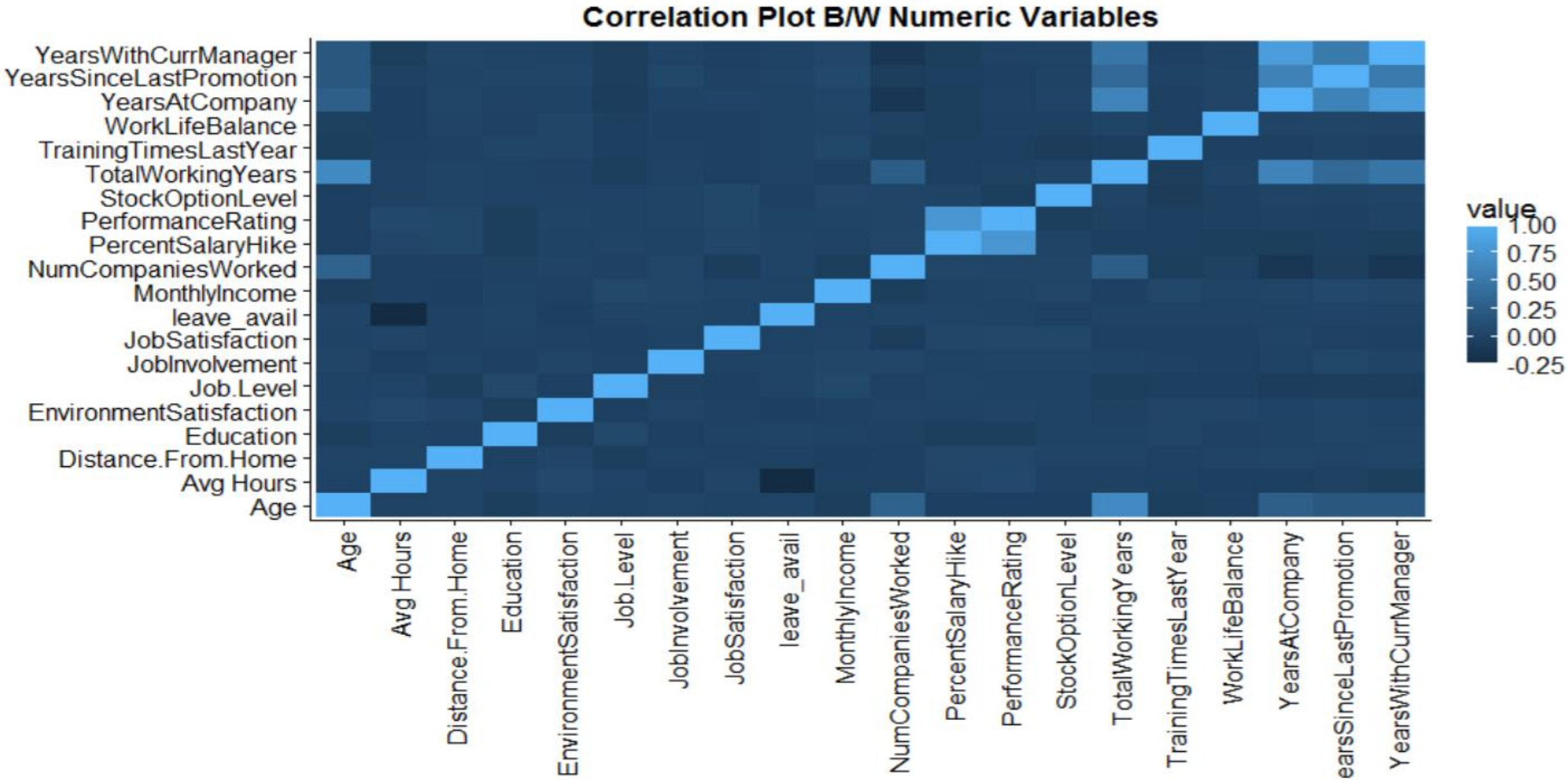

count
1000
750
500
250
0
0
5
10
15
YearsWithCurrManager
Attrition
No
Yes


count
1500
1000
500
0
0
2
4
6
8
YearsSinceLastPromotion
Attrition
No
Yes


count
600
400
200
0
0
5
10
15
20
YearsAtCompany
Attrition
No
Yes


count
1500
1000
500
0
0
2
4
6
TrainingTimesLastYear
Attrition
No
Yes


count
2000
1000
0
1
2
3
4
WorkLifeBalance
Attrition
No
Yes


Correlation Plot B/W Numeric Variables
YearsWithCurrManager
YearsSinceLastPromotion
YearsAtCompany
WorkLifeBalance
TrainingTimesLastYear
TotalWorkingYears
StockOptionLevel
PerformanceRating
PercentSalaryHike
NumCompaniesWorked
MonthlyIncome
leave_avail
JobSatisfaction
JobInvolvement
Job.Level
EnvironmentSatisfaction
Education
Distance.From.Home
Avg Hours
Age
Age
Avg Hours
Distance.From.Home
Education
EnvironmentSatisfaction
Job.Level
JobInvolvement
JobSatisfaction
leave_avail
MonthlyIncome
NumCompaniesWorked
PercentSalaryHike
PerformanceRating
StockOptionLevel
TotalWorkingYears
TrainingTimesLastYear
WorkLifeBalance
YearsAtCompany
YearsSinceLastPromotion
YearsWithCurrManager
value
1.00
0.75
0.50
0.25
0.00
-0.25

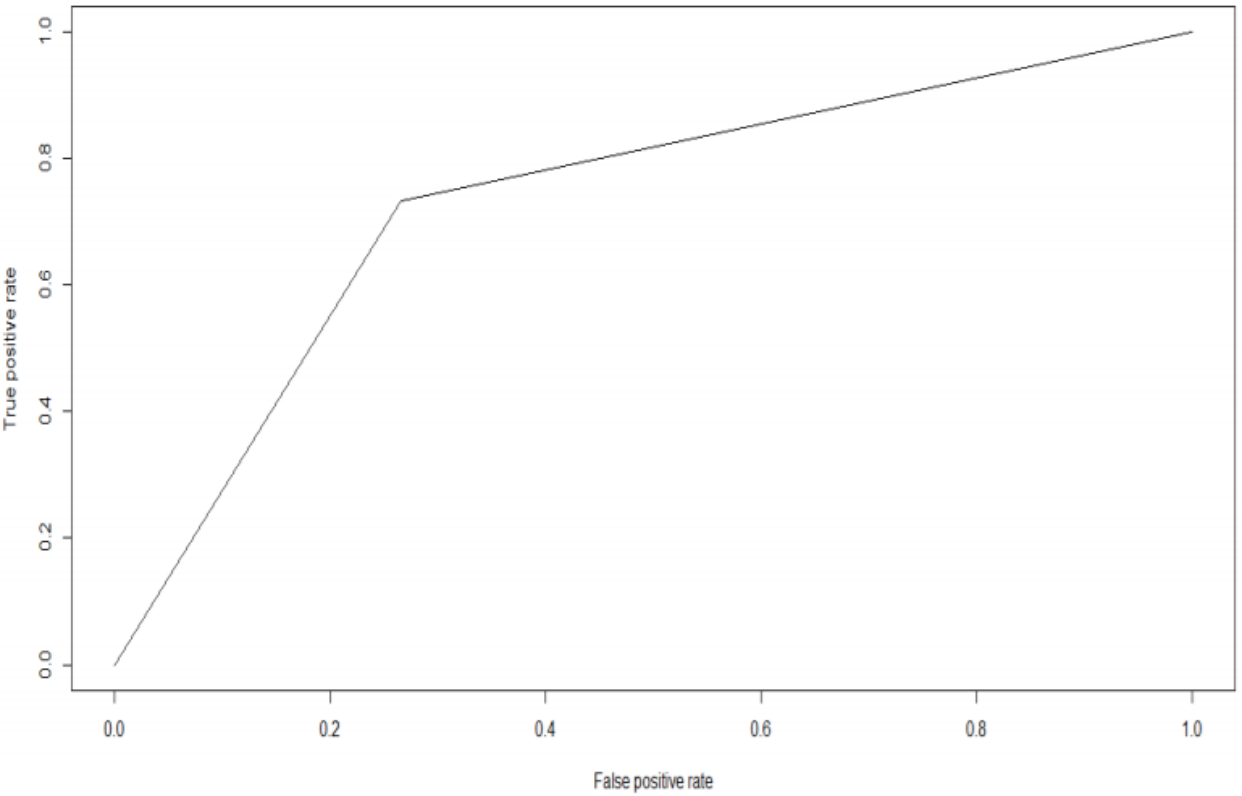
Logistic Regression Model

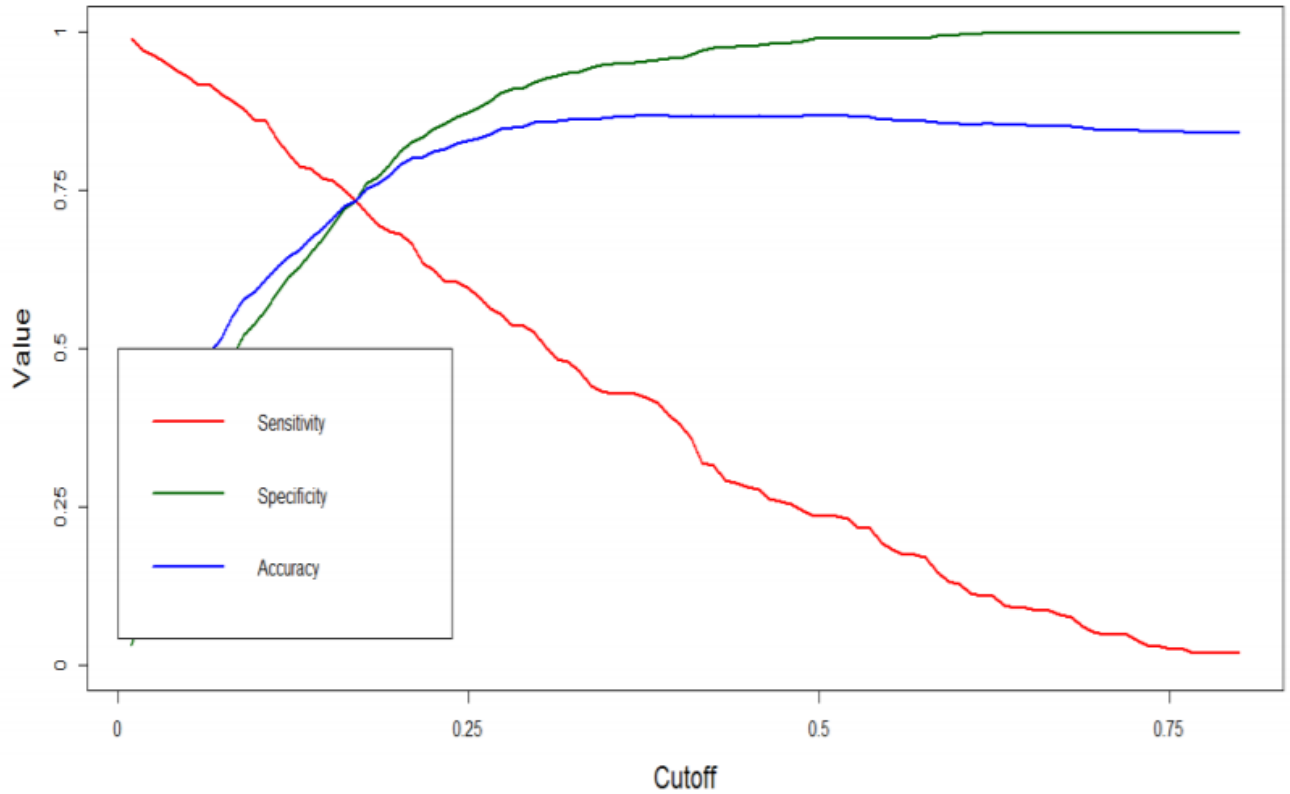
- 70% data was sampled as train and remaining as test
- The first model was built with 53 variables
- stepAIC function removed insignificant variables
- Based on the multicollinearity and significance of the variables and after many iterations, final model was built with 10 variables
- The final model had the following variables
- Age, NumCompaniesWorked, YearsSinceLastPromotion
- YearsWithCurrManager, EnvironmentSatisfaction, JobSatisfaction
- Avg_working_hours, BusinessTravelTravelFrequently
- JobRoleManufacturing.Director, MaritalStatusSingle

Model Evaluation

- Cut off value - 0.1695

Accuracy	Sensitivity	Specificity
0.733938	0.7323944	0.7342342





Conclusions

- The following factors affect the attrition of the employees
- Age, number of companies worked, years since last promotion
- Years with current manager, environment satisfaction, job satisfaction
- Average working hours, business travel frequently
- Manufacturing director job role, marital status single
- The employees seem to be well compensated, so it was not a driving factor
- Employees without environment satisfaction, job satisfaction and working long hours tend to leave the company
- Single employees are more prone to job switch
- Employees not promoted for a long time tend to leave
- Frequently travelling employees seem to enjoy the trips, they're less likely to leave