

# HR Analytics CASE STUDY

## SUBMISSION

Group Name:

1. Ria Nag
2. Piyush Gaur
3. Priya Chopra
4. Sanjana Rao

## Background- HR Analytics Case Study

### Background

A large company XYZ employees 4000 employees , at a given point in time has 15% attrition rate year-on-year. Such a high attrition impacts the company

- Delivery delays of associate leaving thus impacting customer experience and partners
- Large size department to be maintained to retain resources
- Frequent On Job trainings and induction sessions for new associates to be conducted

### Business Objective

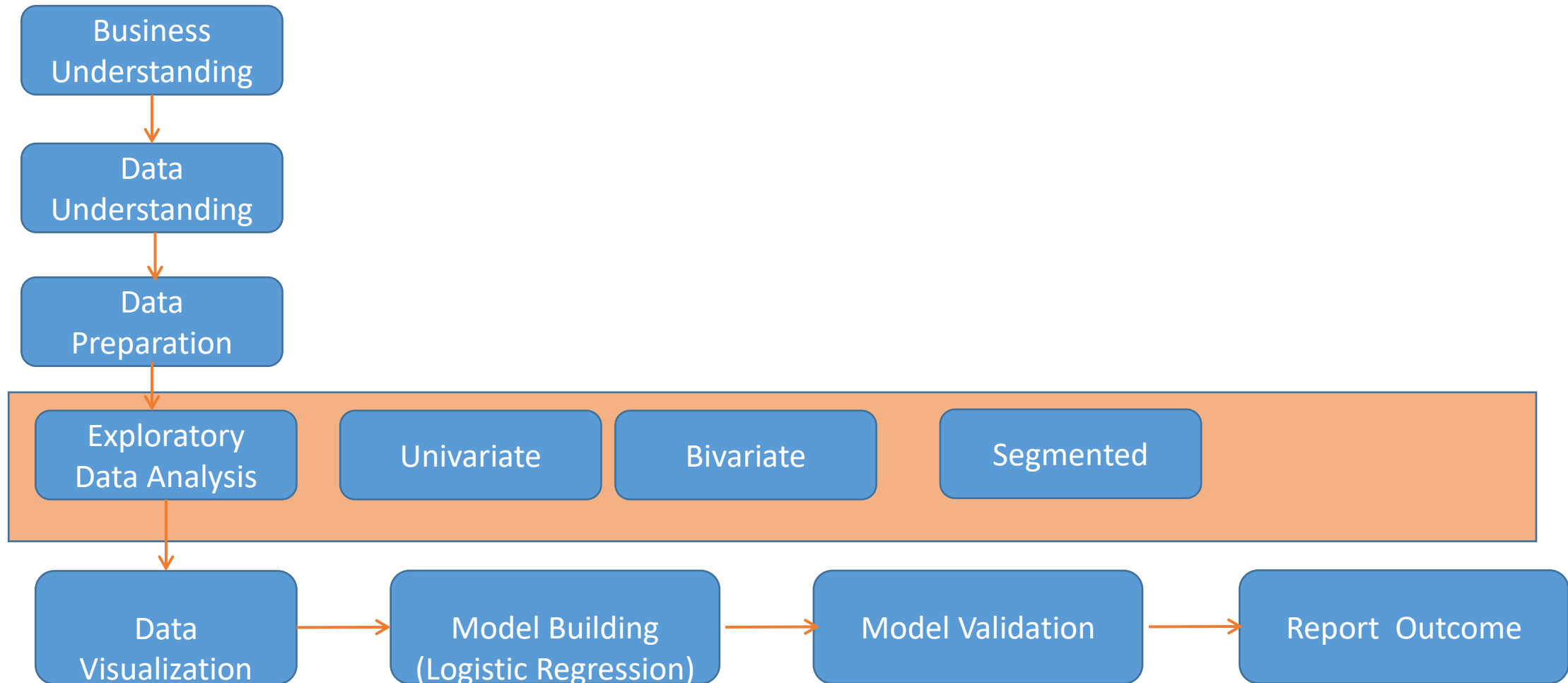
Identify factors to focus on the curb attrition i.e. required changes at workplace so that employees can be retained for long term

### GOAL

*Model the probability of Attrition using logistic regression*

*Outcome would help management assess factors impacting attrition and mitigation action can be planned*

# Problem Solving Methodology



## METHODS USED FOR DATA PREPARATION

---

- All missing values and outlier treatment was done and all data quality issues were addressed
- Continuous variables were checked for collinearity using correlation matrix
- All continuous variables were scaled to the same scale
- All character and discrete variables were first converted to factors and then converted to numerical dummy variables
- Two metrics were derived from the data:
  - Number of leaves taken by each employee in a year
  - Average time at work per day for each employee in a year
- For Bi-variate Analysis of each variable versus Attrition , the y-axis represent the total count of each employee in each segment of variable when normalized to same scale having total value equal to 1

## METHODS USED FOR DATA PREPARATION

---

- A logistic regression model was built following stepwise selection followed by backward selection
- The cost optimization was done using maximum likelihood method
- Parameters with high multicollinearity and low significance were removed from the model
- Only 7 variables were kept in final model to reduce complexity of the model

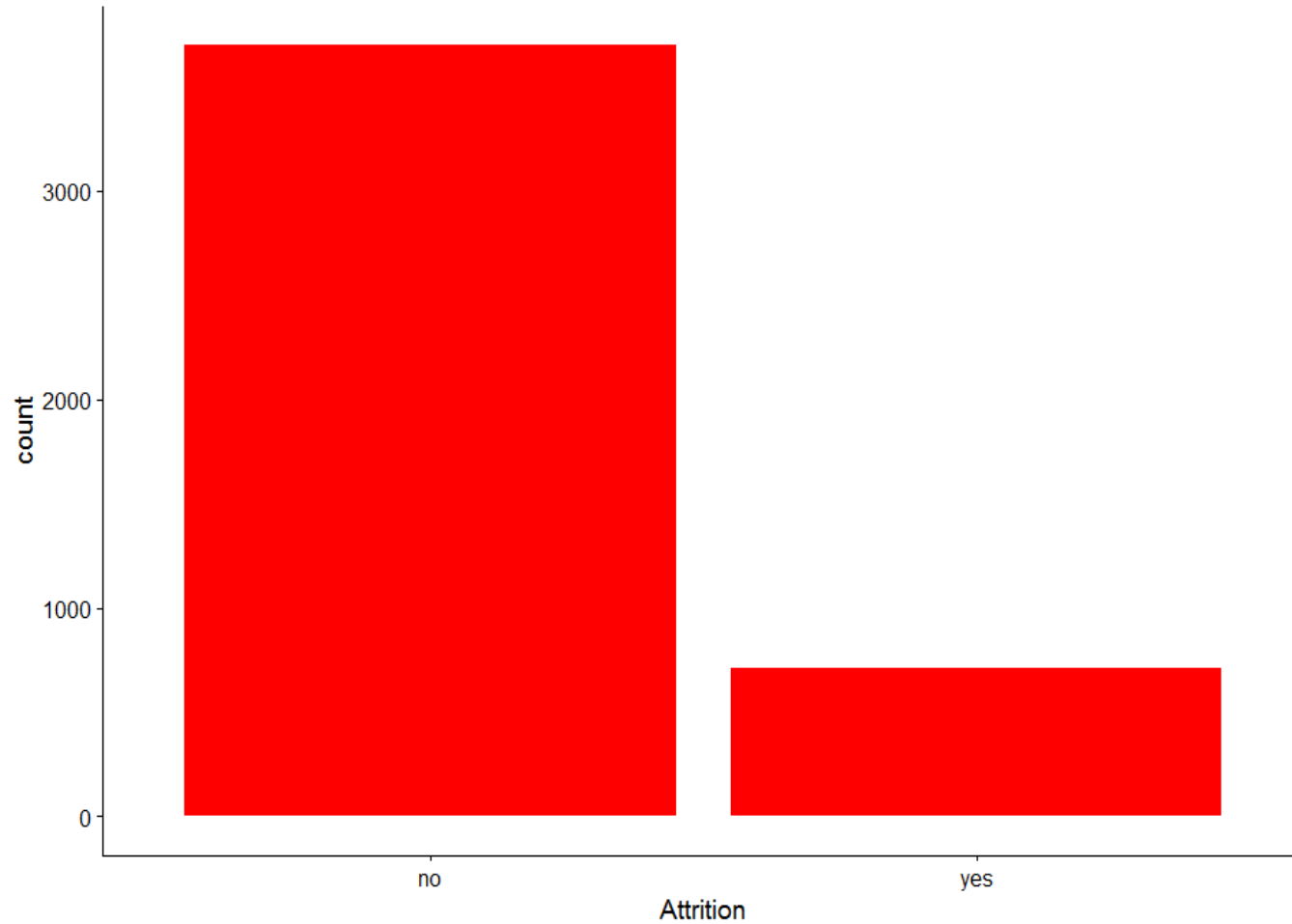
## Data Understanding – Type of Variables

---

Continuous Variables	Discrete Variable	Character Variable
Age	Environment Satisfaction	Attrition
Average Time at work	Job Satisfaction	Business Travel
Distance From Home	Work Life Balance	Department
Monthly Income	Job Involvement	Education Field
Percent Salary hike	Performance Rating	Gender
Total Working Years	Education	Job Role
Years at company	Job Level	Marital Status
Years Since Last Promotion	Num Companies Worked	
Years with Current Manager	Stock Option Level	
	Training Time Last Year	

*Character Variables*  
*Univariate, Bivariate, Segmented Analysis*

# Attrition Volume

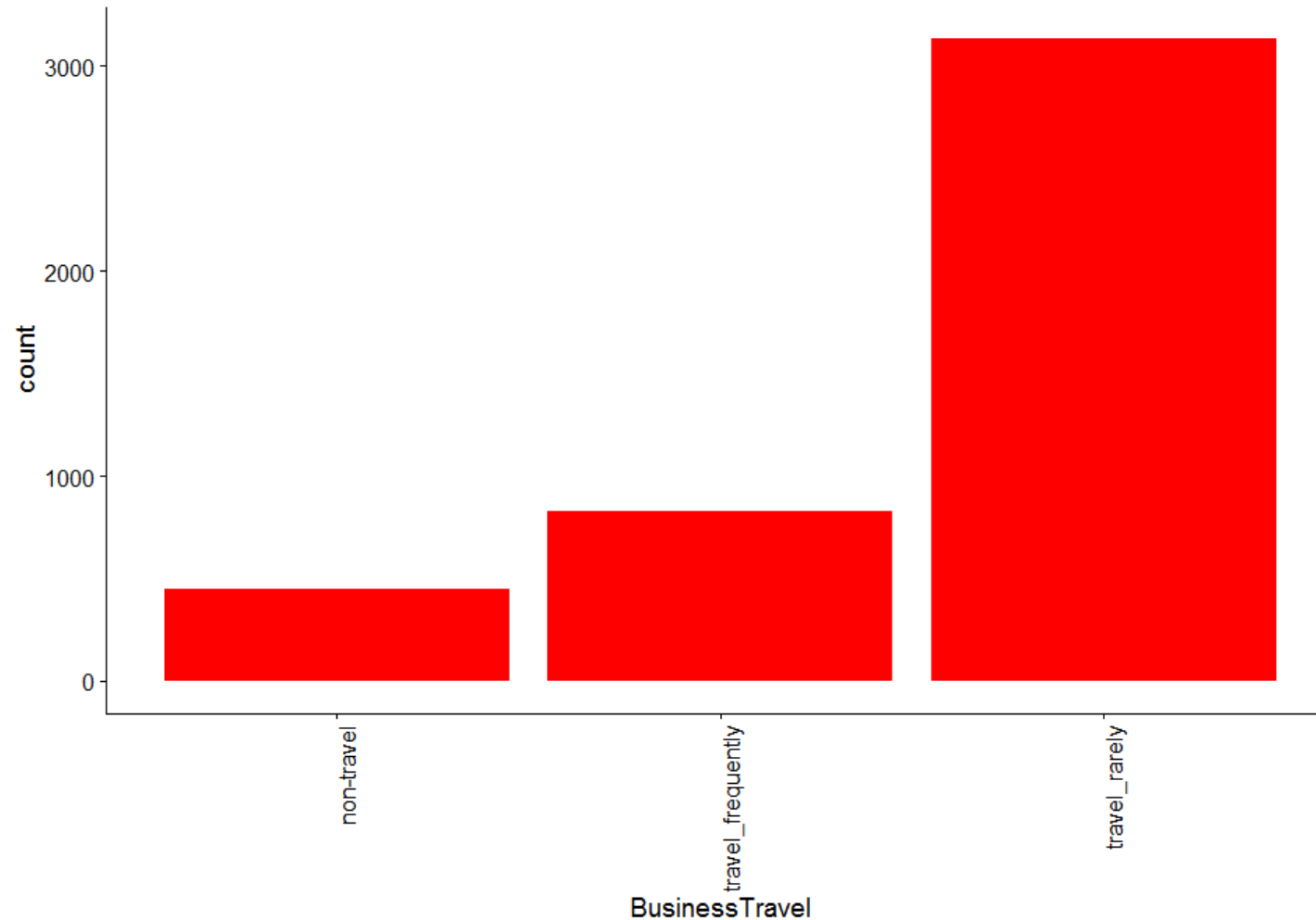


## Observation

*Out of 4000 employees approx 16% leave the company*



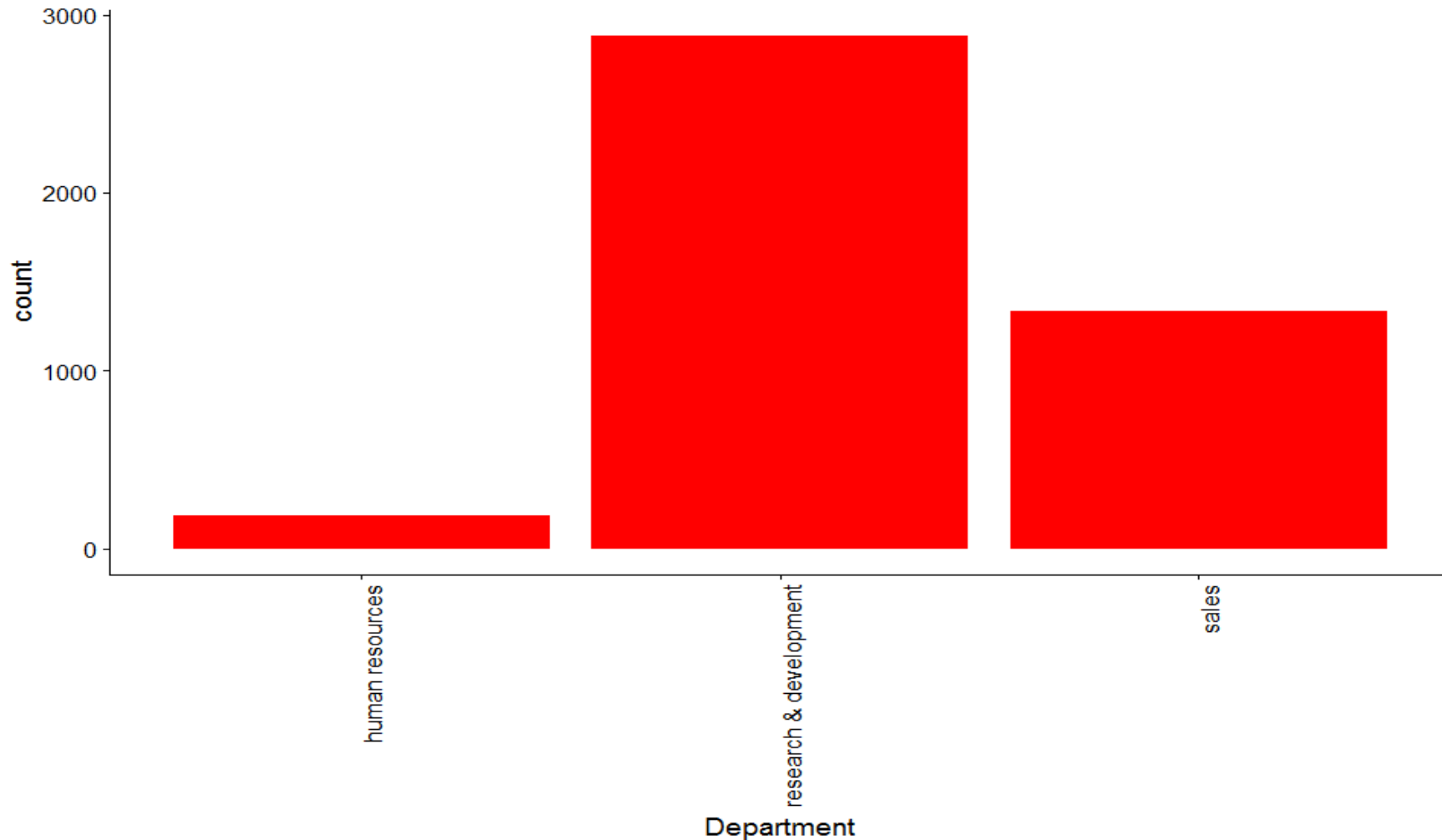
# Business Travels Volume in Company



## Observation

*Most employees travel rarely for business needs*

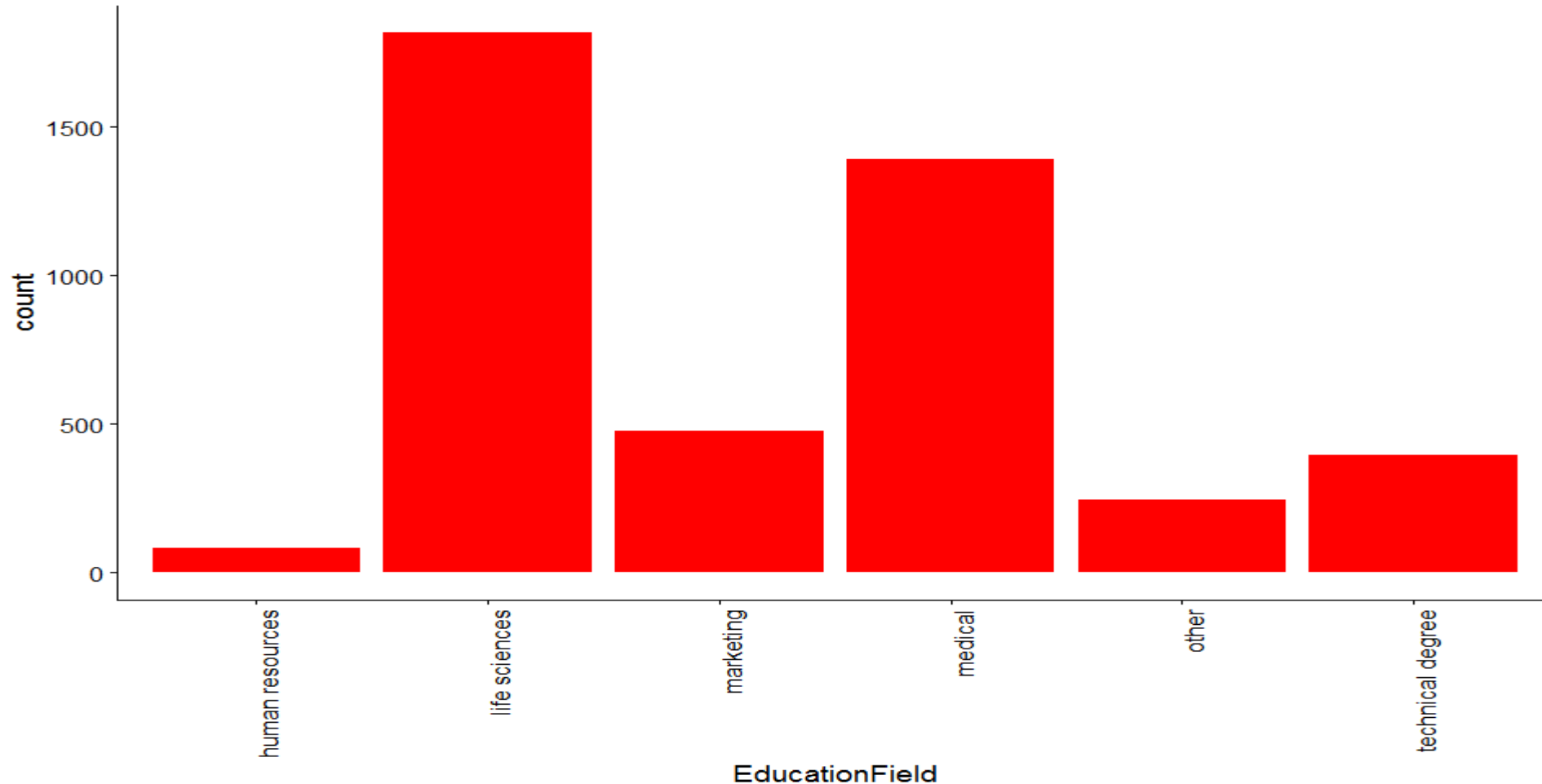
# Department Strength



## Observation

*Employee strength is higher for research and development (R&D) department followed by Sales Department*

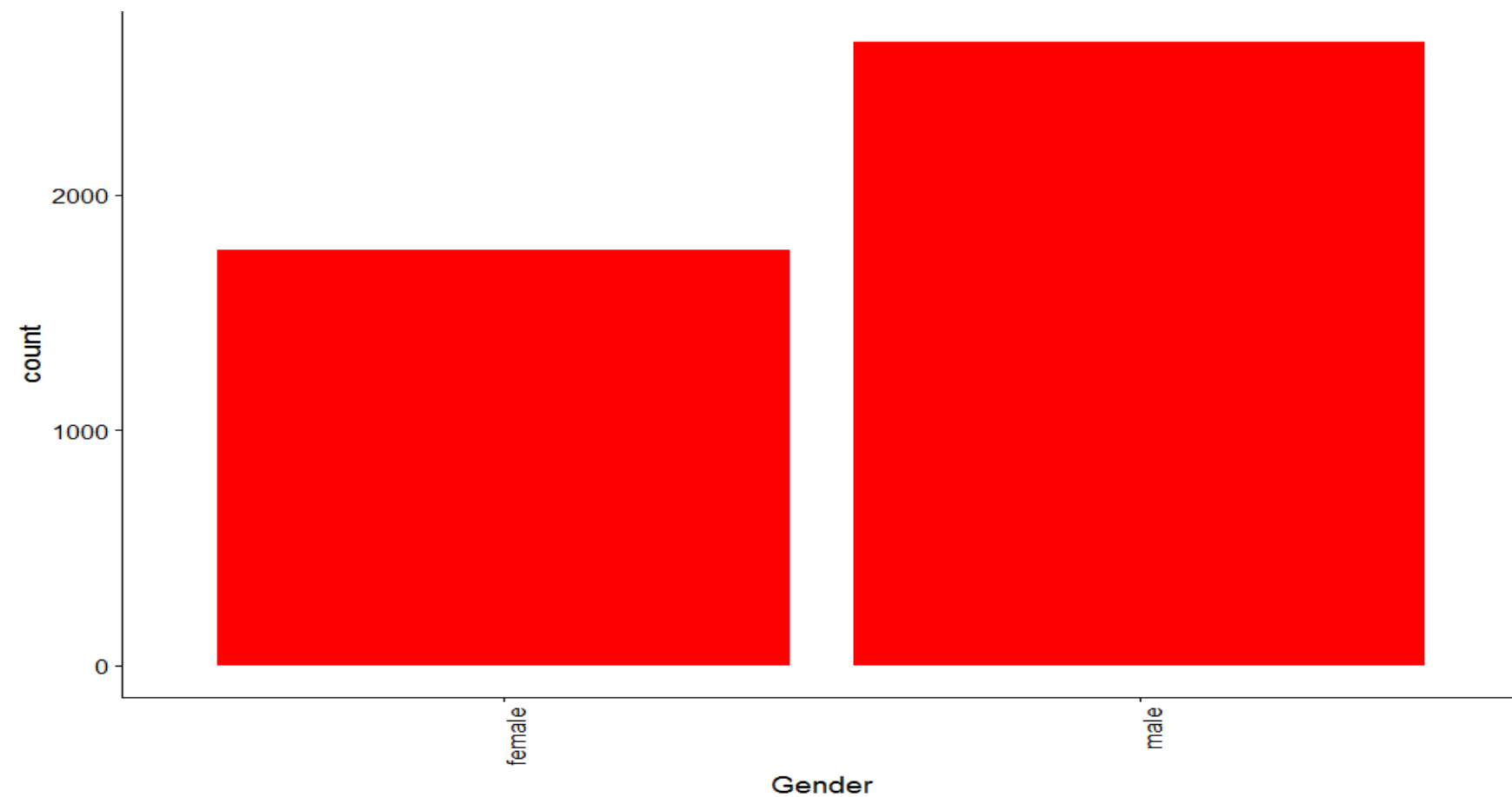
# Educational background



## Observation

Most employees  
*are from Life Sciences and medical field*

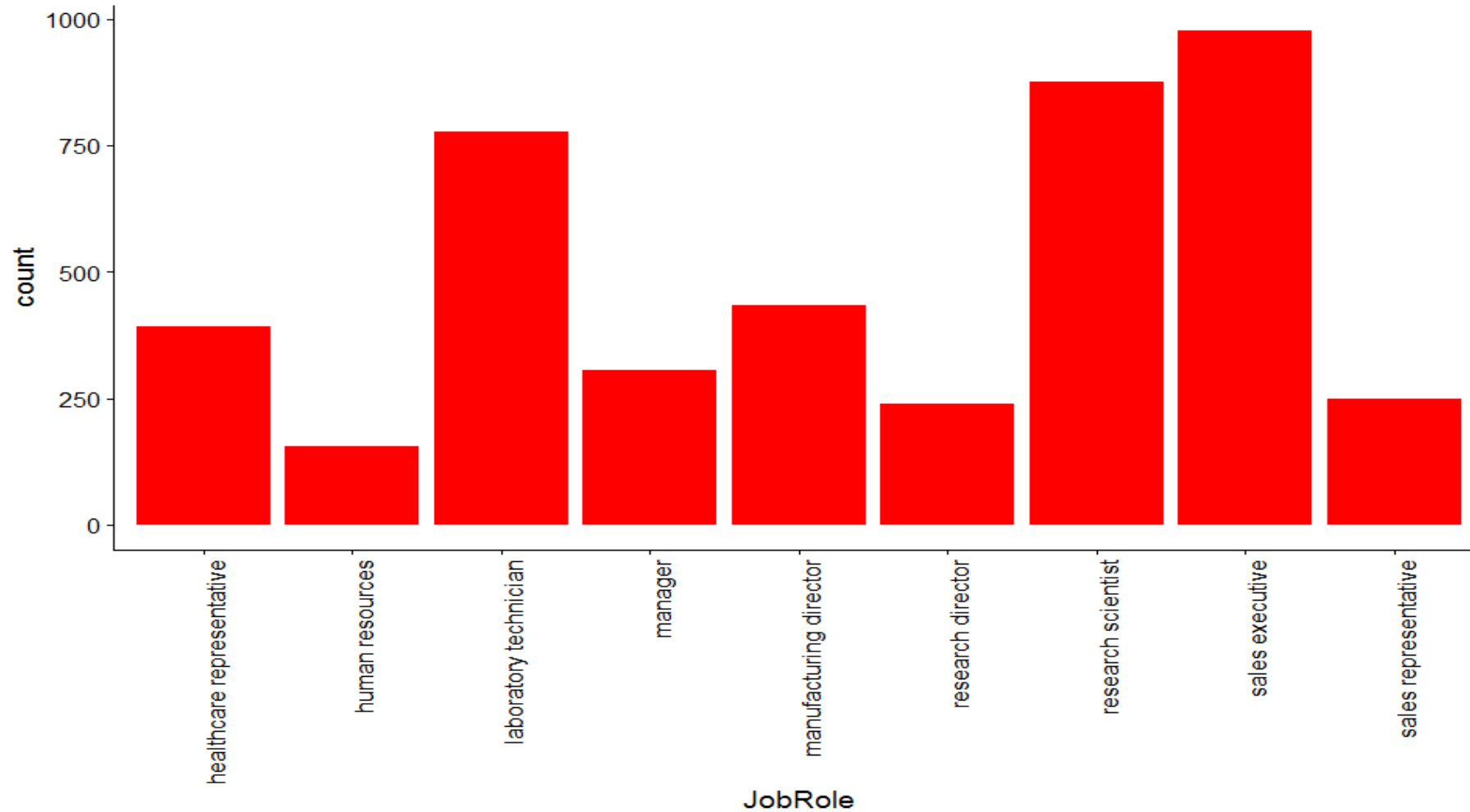
# Gender



## Observation

*Males strength is higher than female*

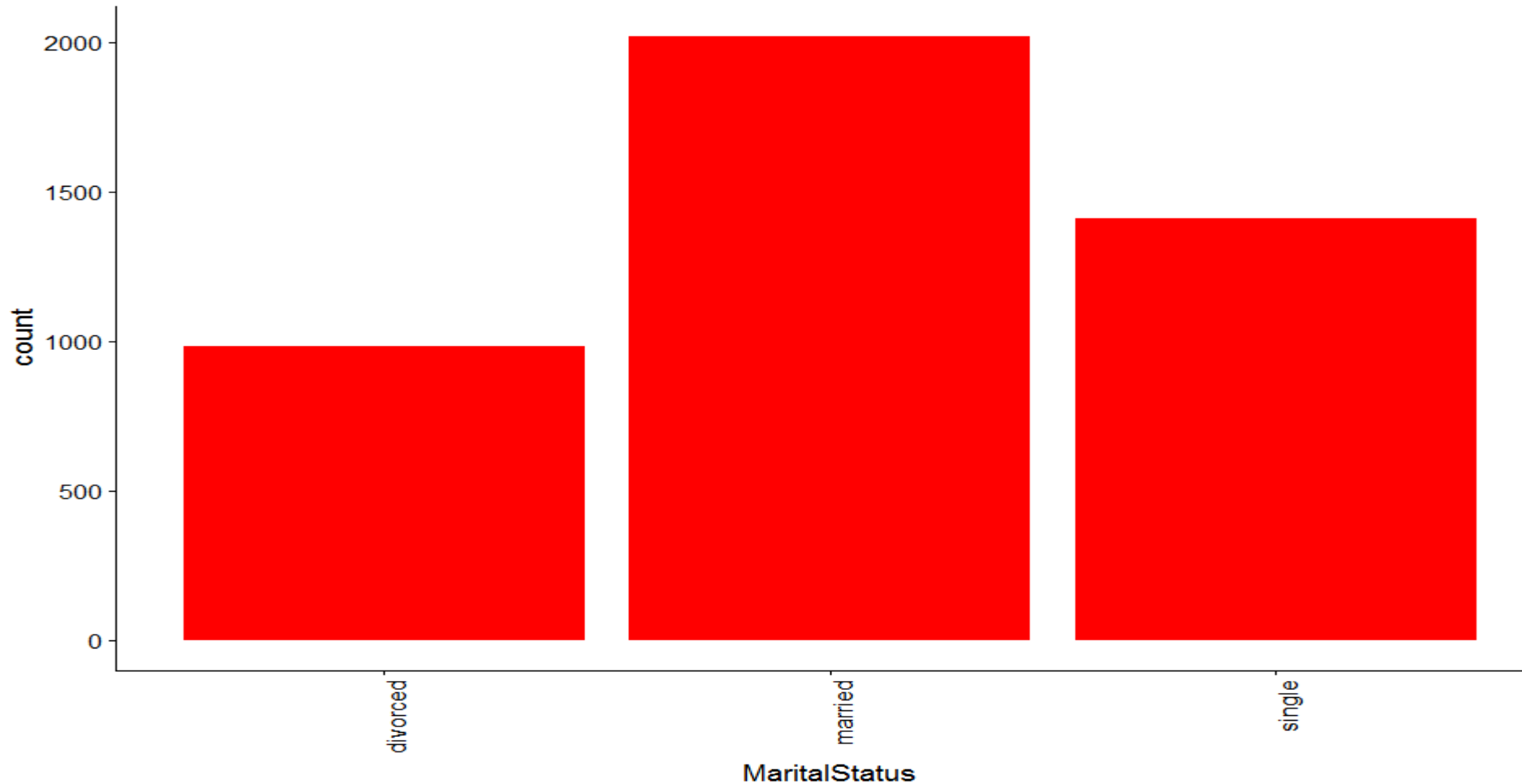
# Job Roles in company



## Observation

*Most common Job role is of Sales executive or research Analyst*

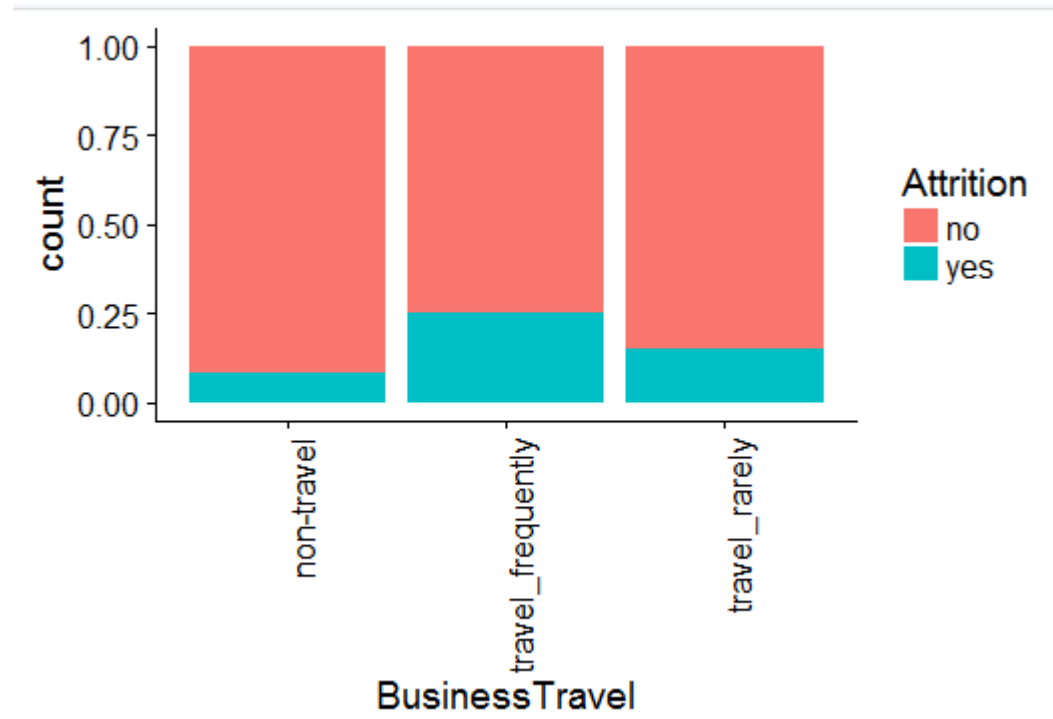
# Martial Status of Employees



## Observation

There are higher volume of Married employees compared to single and divorced are few as well

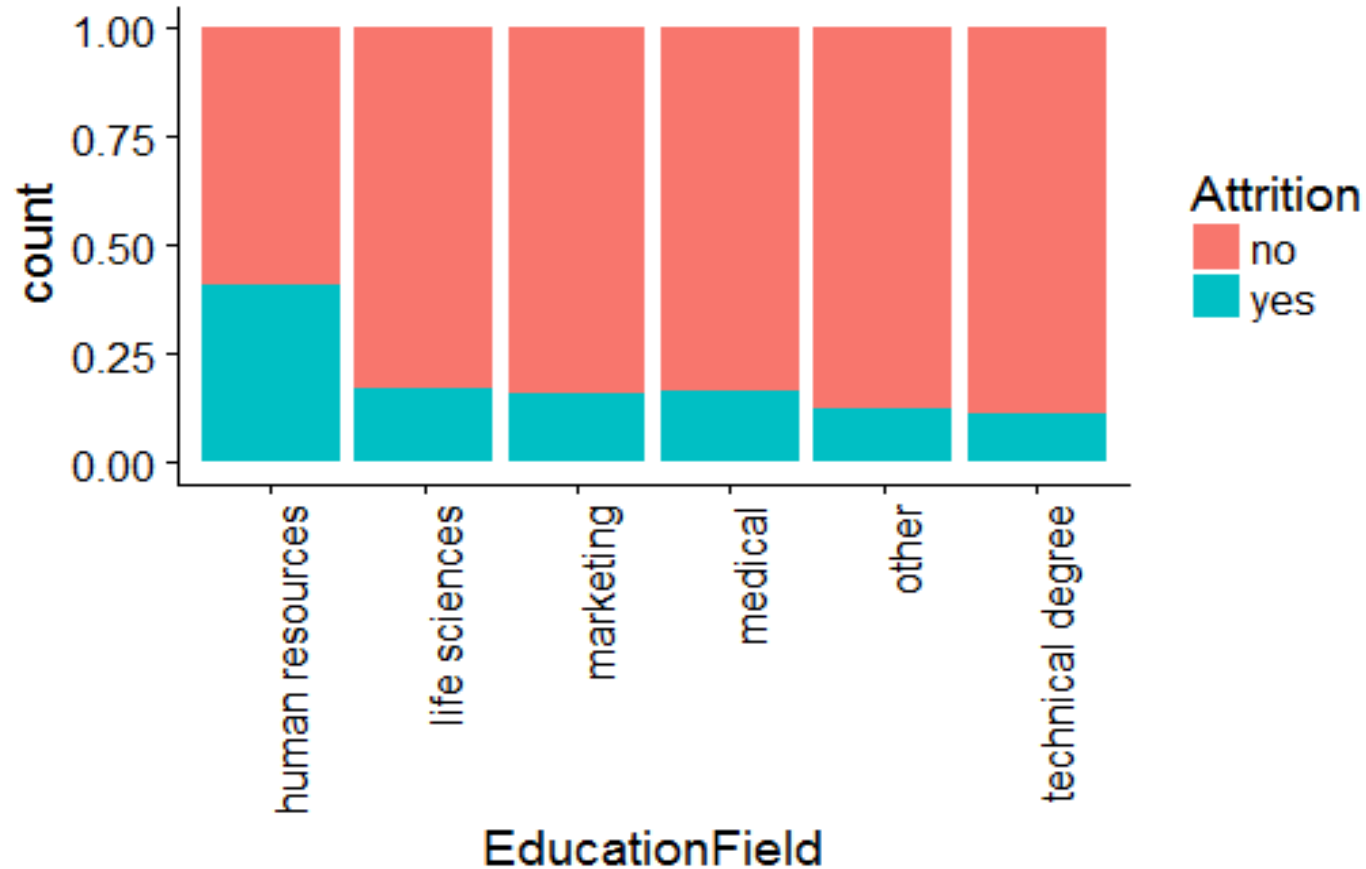
# Travel Impact on Attrition



## Observation

Attrition is higher for those who frequently travel for business needs

# Education Field vs Attrition

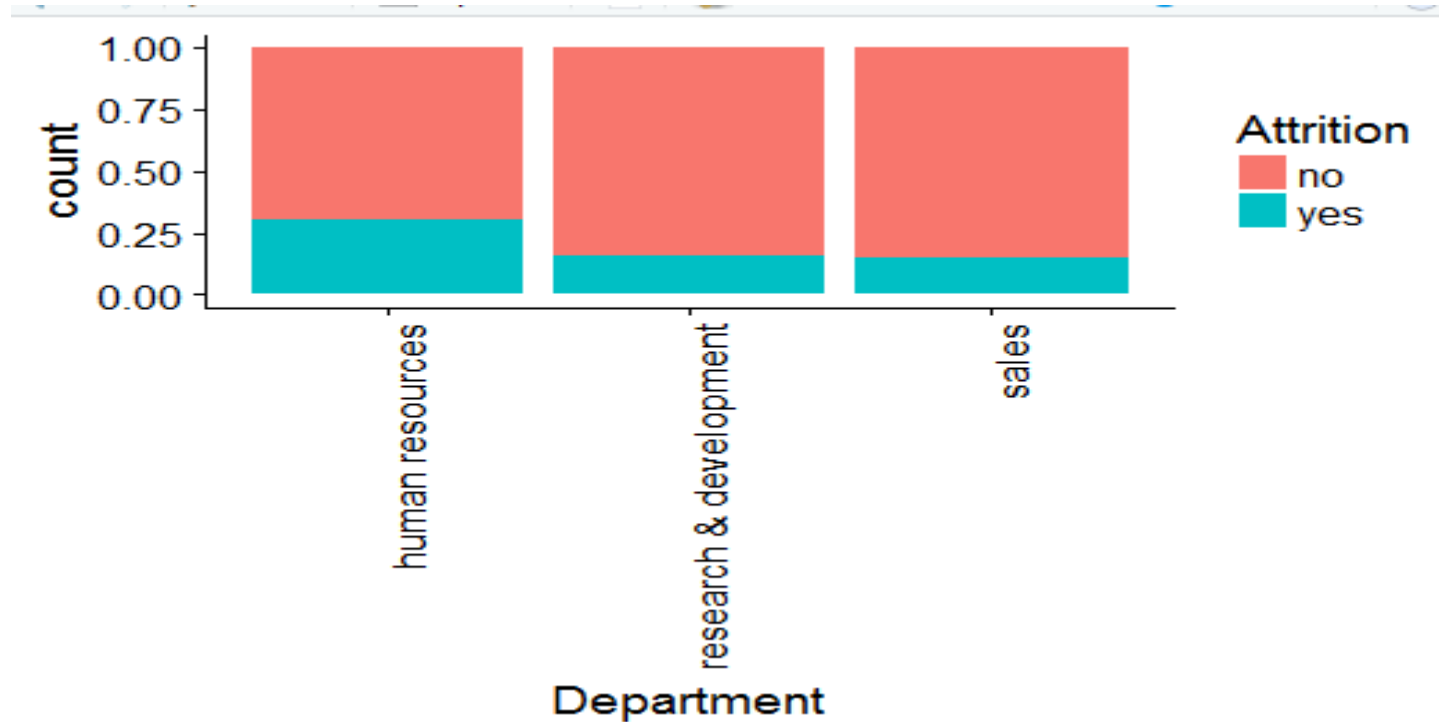


## Observation

Attrition is higher in human resources employees



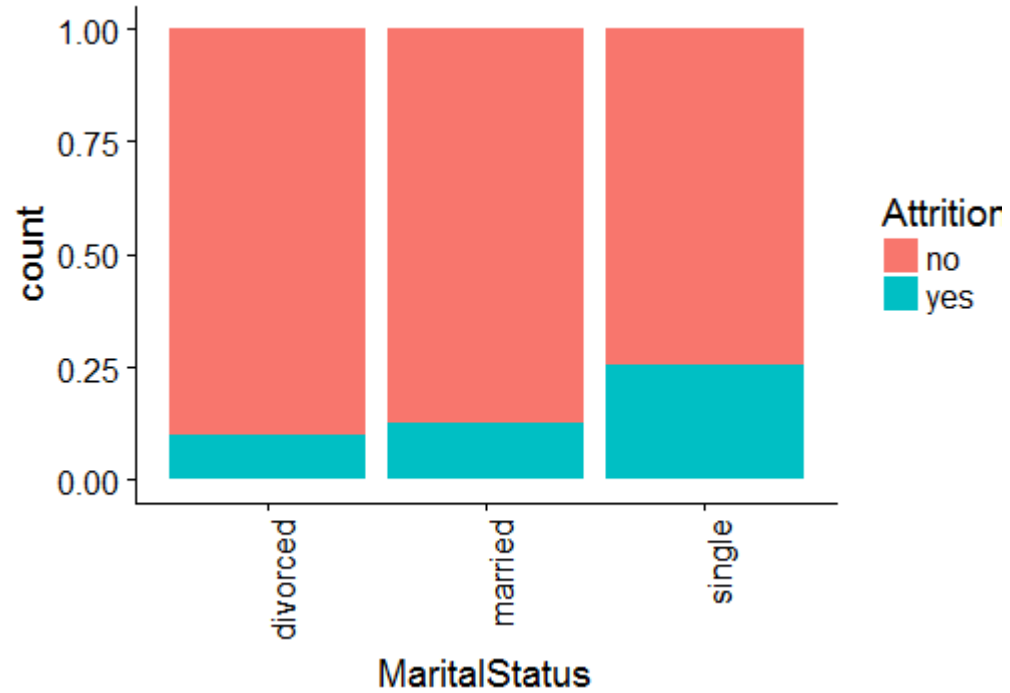
# Department prone to Attrition



## Observation

Attrition is higher for human resource department compared to rest

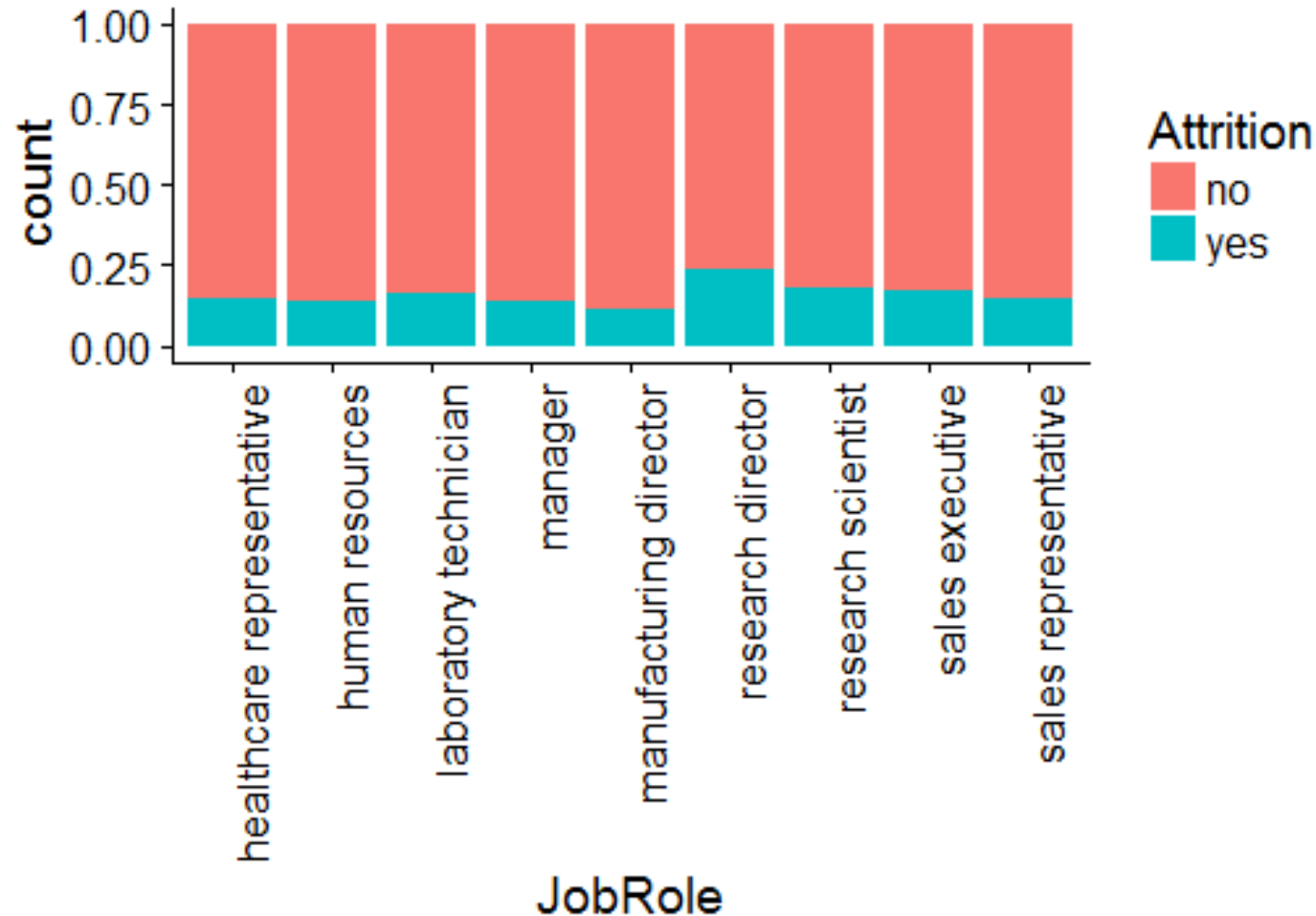
# Marital Status vs Attrition



## Observation

Attrition is higher for unmarried associates

# Job Role impact on Attrition

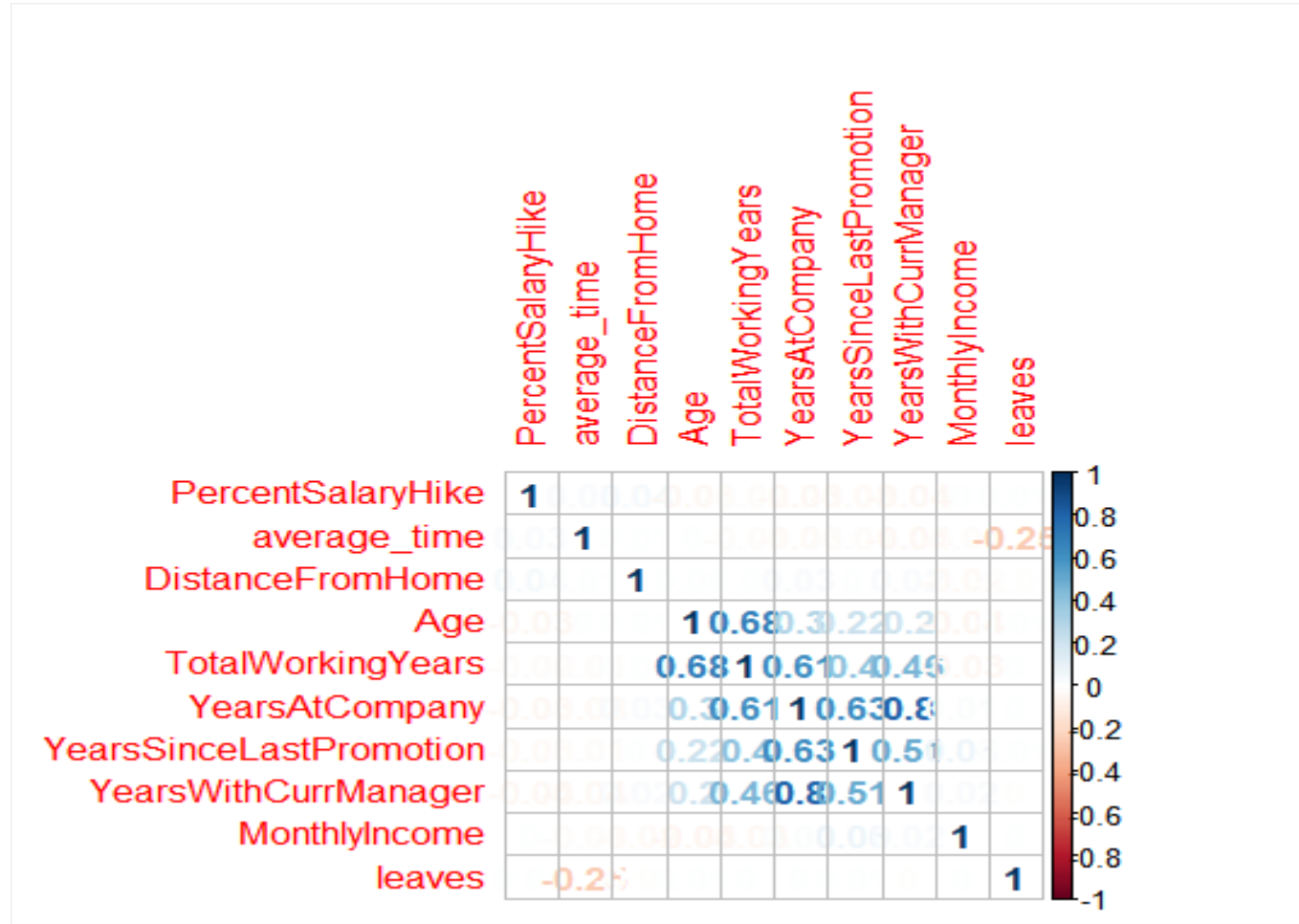


## Observation

Attrition is higher in employees with Job role as research director

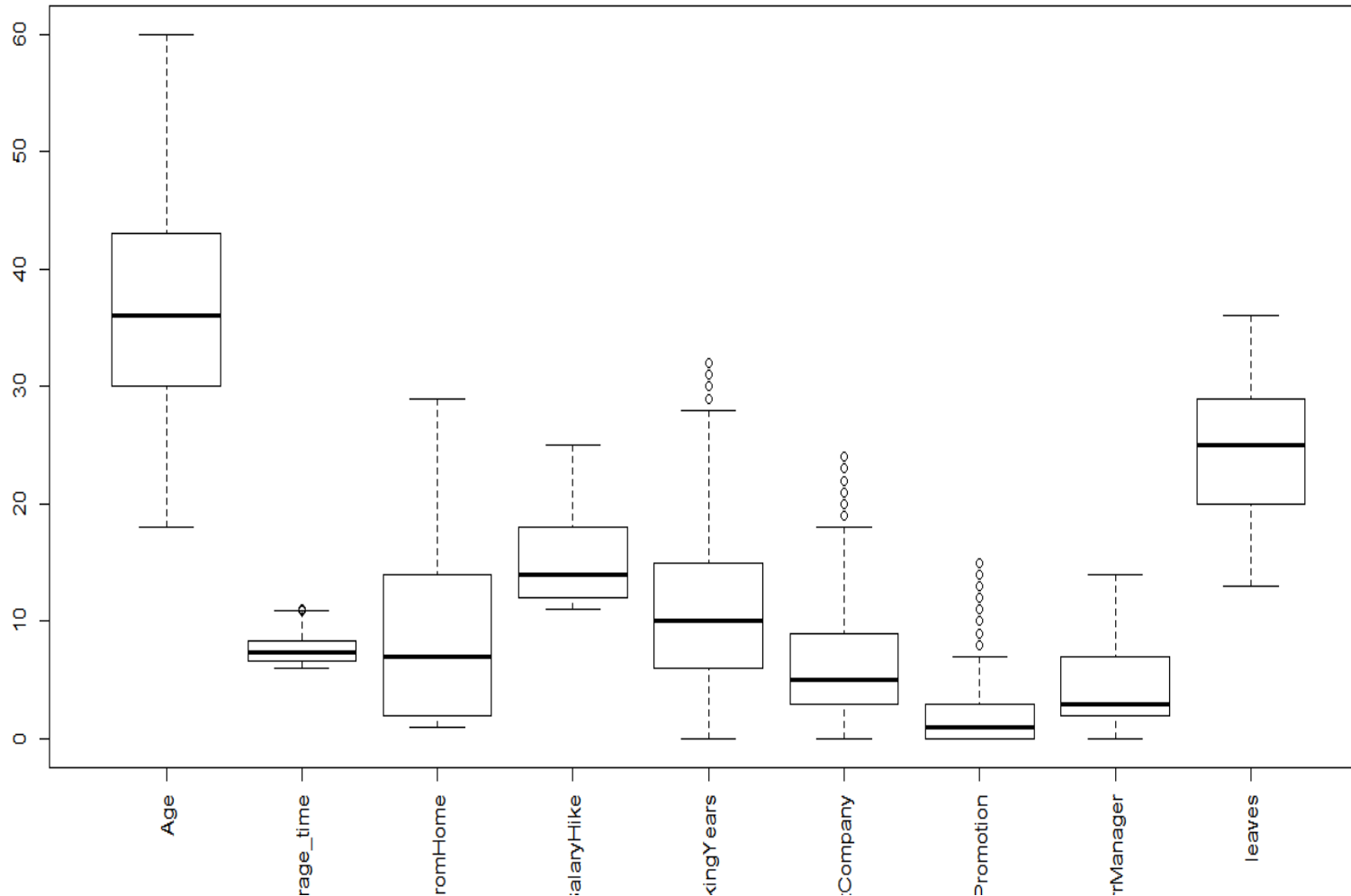
*Continuous variables*  
*Univariate, Bivariate, Segmented Analysis*

# Correlation Heat Map



Observation  
 Years at Company is in high correlation to Years with Current Manager

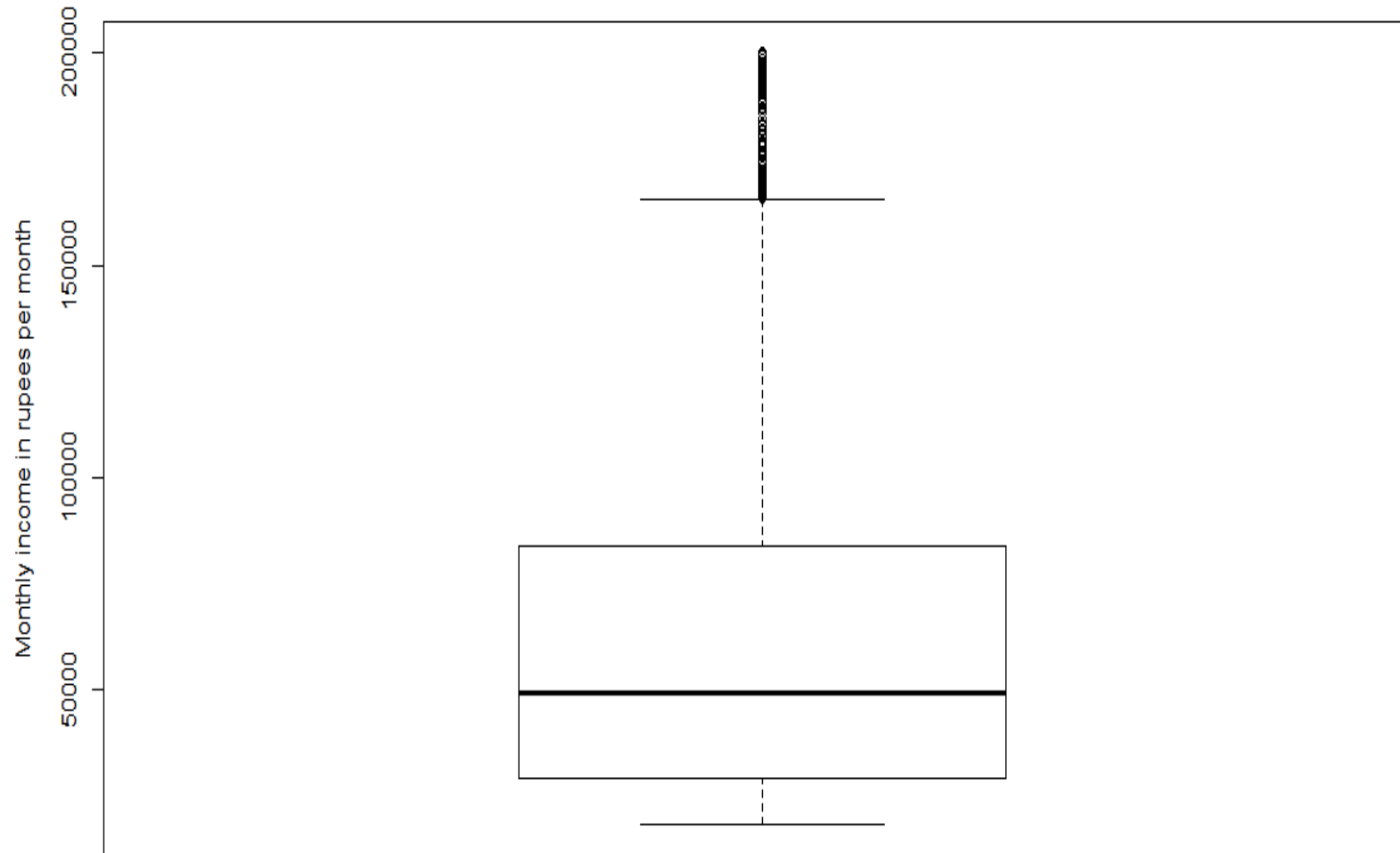
# Median Value for Continuous Variables



## Observation

- Most employees have a age around 35
- Most employees live at less than distance of 10KM from home
- Most employees take average 8 mins to reach office
- Average Percent Salary hike is 15%
- Average Working years of Employees is around 11 years
- Most employees have 6 years of working experience in company
- It has been average 2 years since last promotion for employees
- Most employees have been 4 years with current manager
- Most employees take 25 days of leaves per year outside weekend

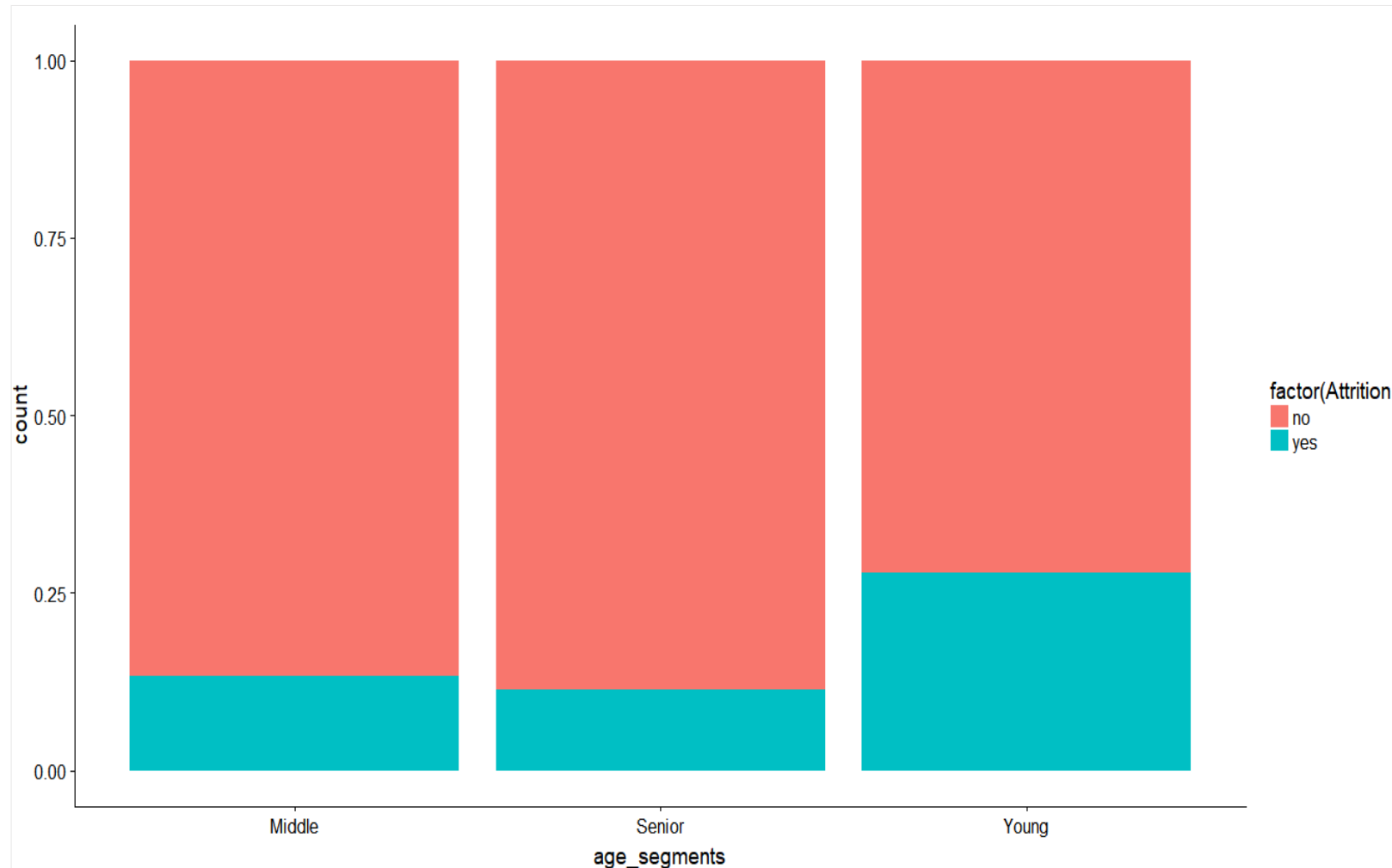
# Median Monthly Income



Observation

*Median  
monthly  
income of  
employees is  
Rs.50000*

# Age Segments vs Attrition

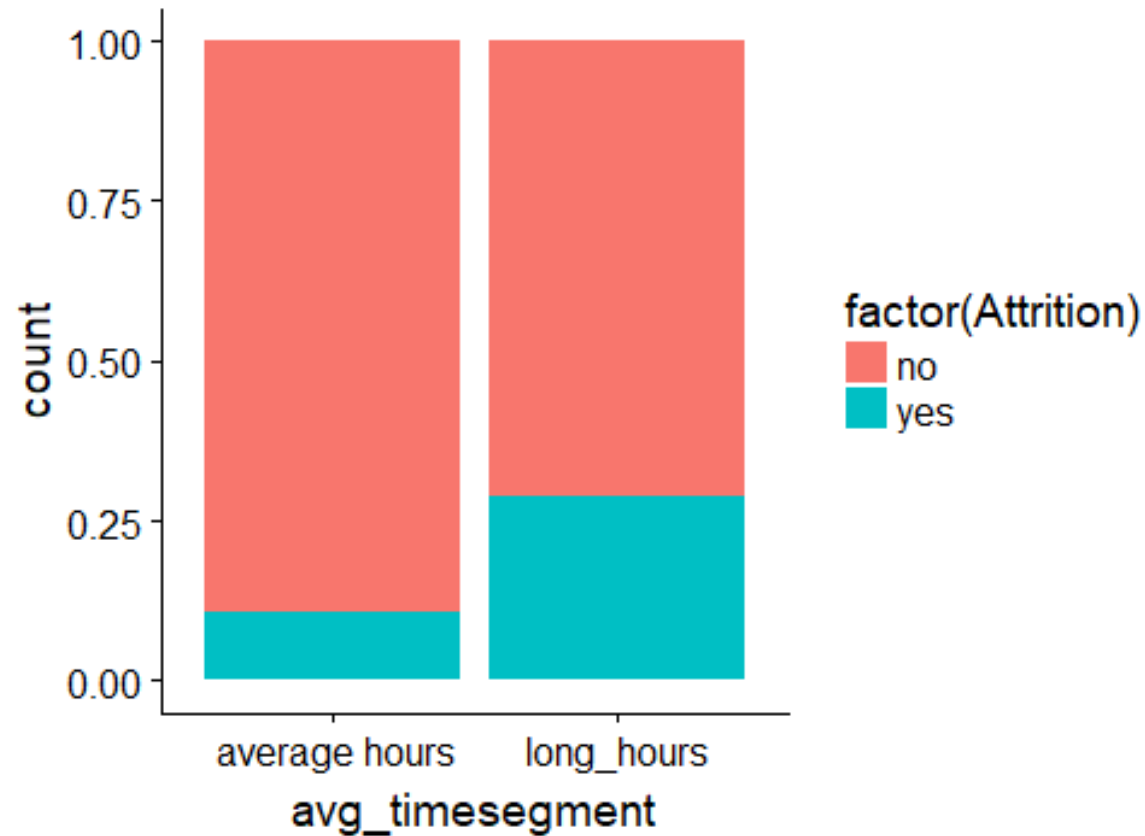


## Observation

*There is higher level of attrition among young people age group.*



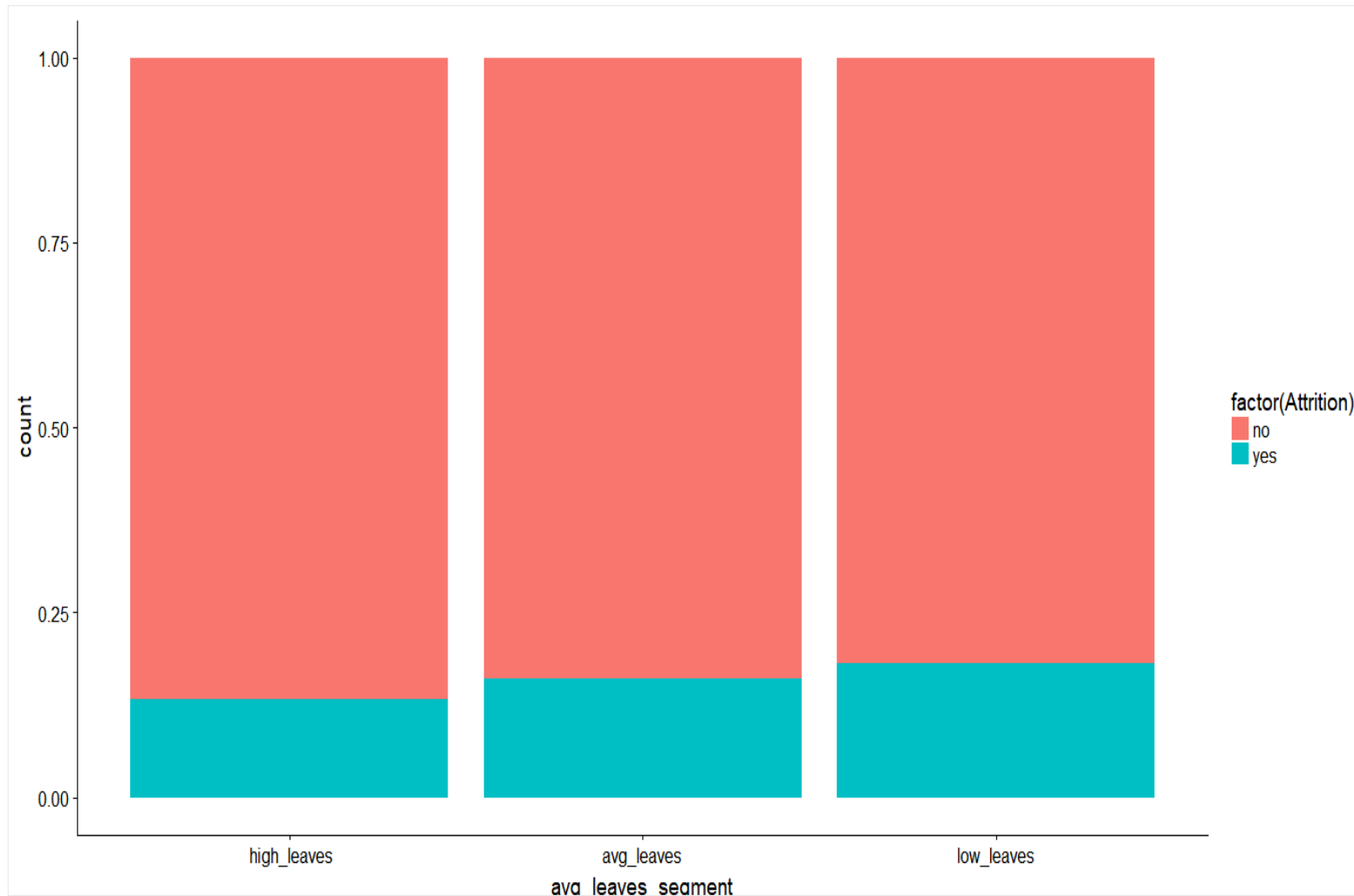
# Average Time Segment vs Attrition



## Observation

*Attrition is higher for people working more than 8 hours per day on an average*

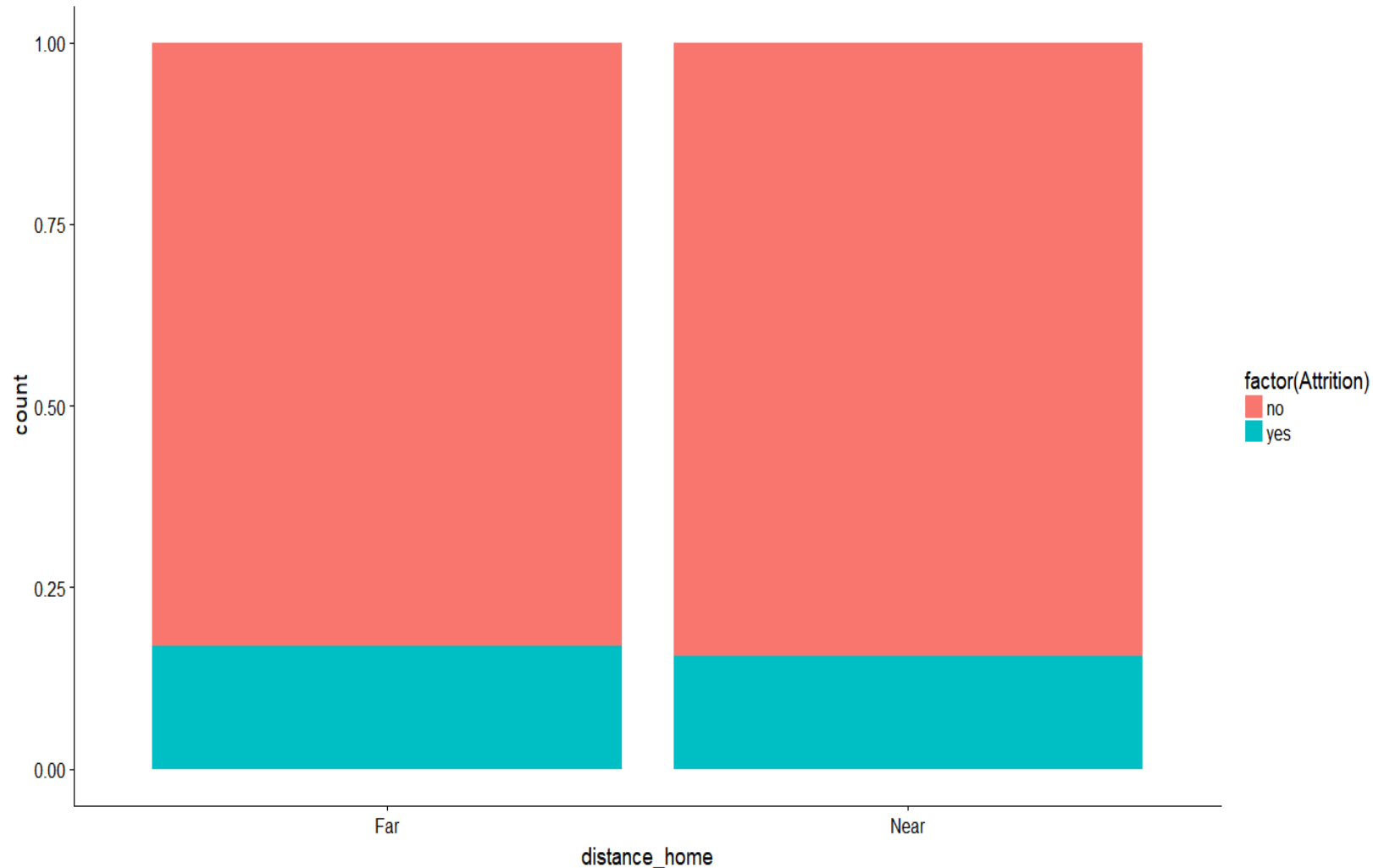
# Average Leaves vs Attrition



## Observation

*Attrition seems to be marginally decreasing as number of leaves taken by employee increases*

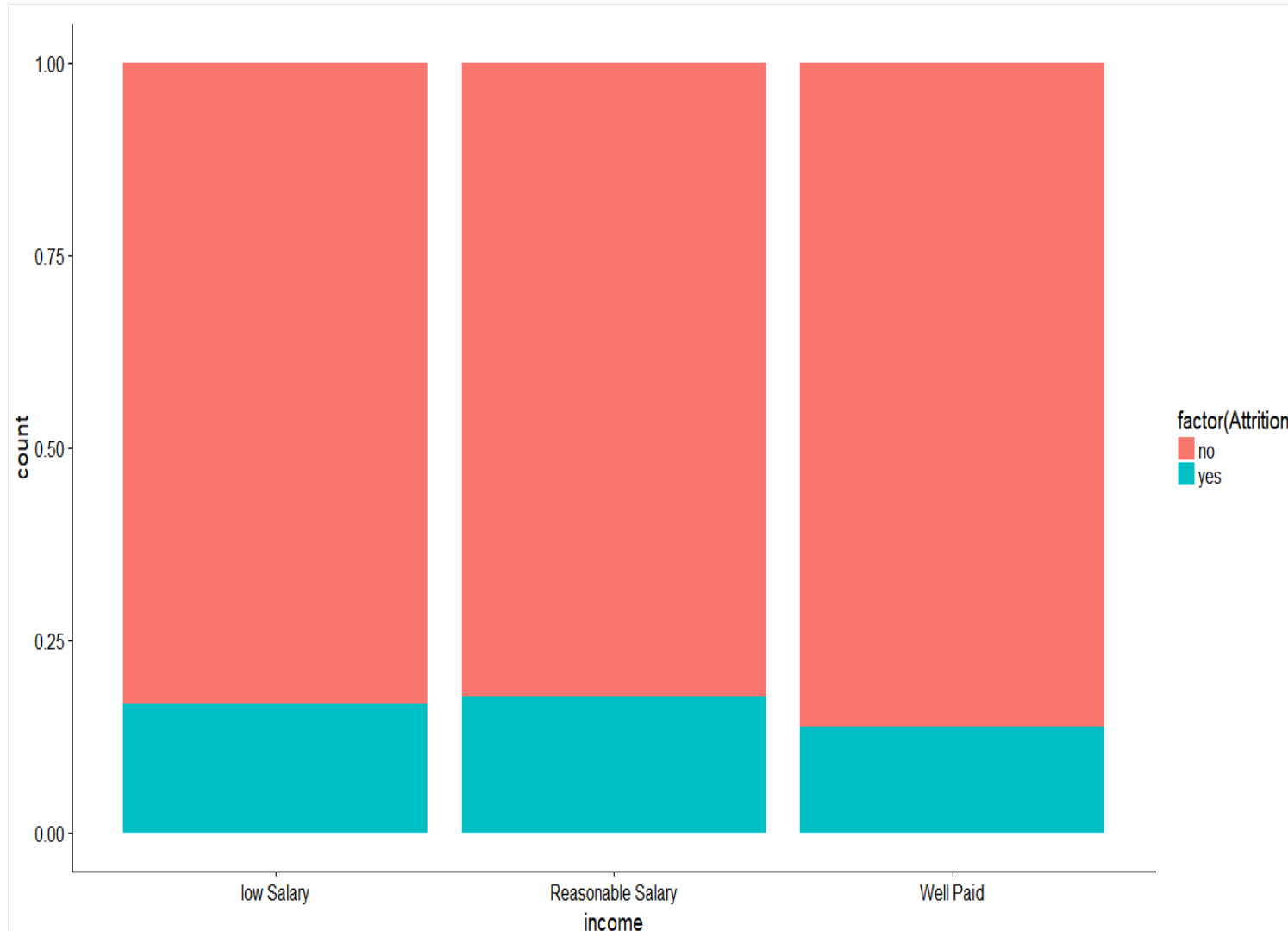
# Distance From Home vs Attrition



## Observation

*There is pretty low difference in attrition for people living less than 10 km or more*

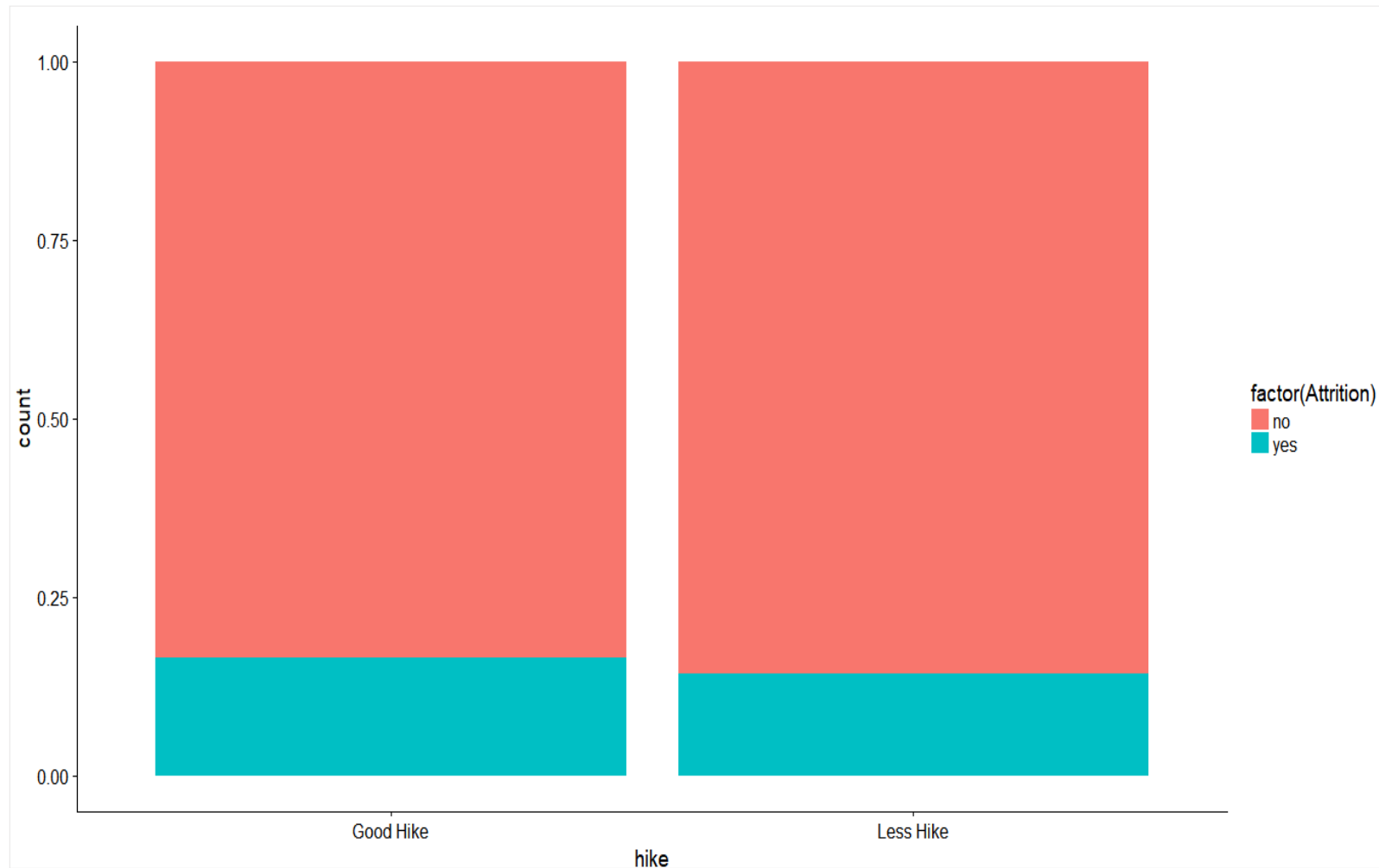
# Monthly Income vs Attrition



## Observation

*There is marginally low difference in attrition based on salary*

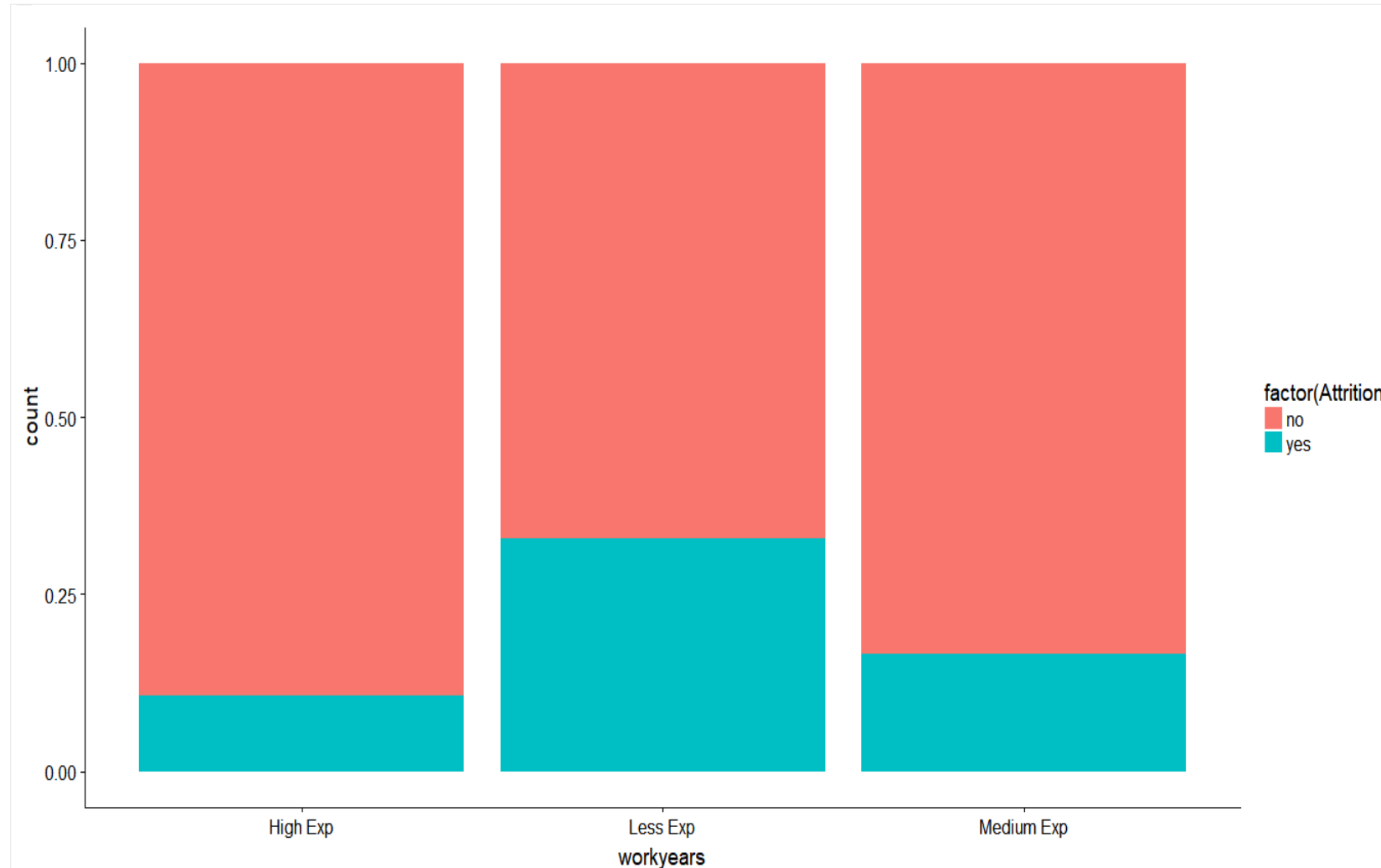
# Salary Hike vs Attrition



## Observation

*There is marginally any difference in attrition due to salary hike*

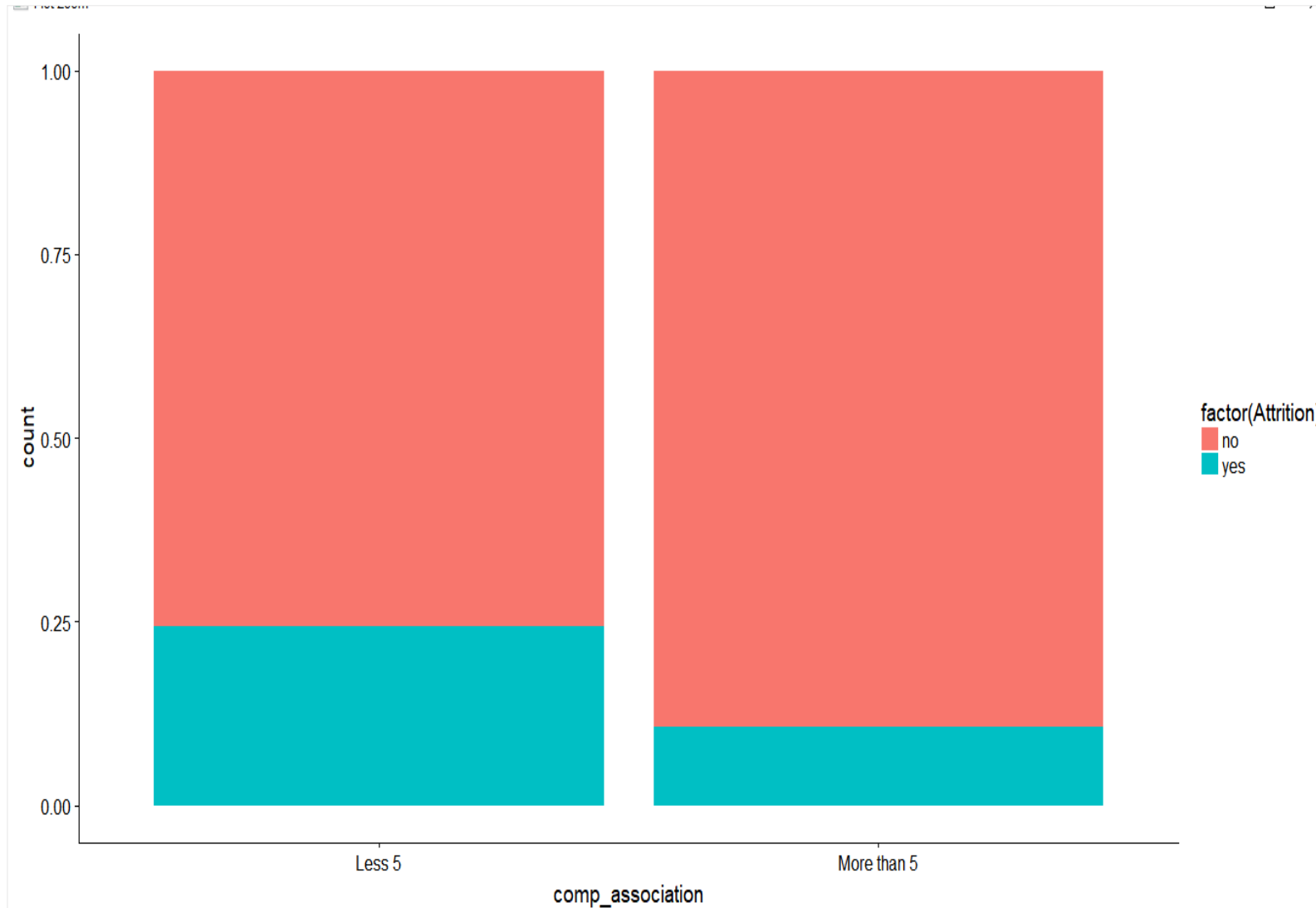
# Impact due to Work Years



## Observation

*Attrition seems to be higher for low working years i.e. less than 5*

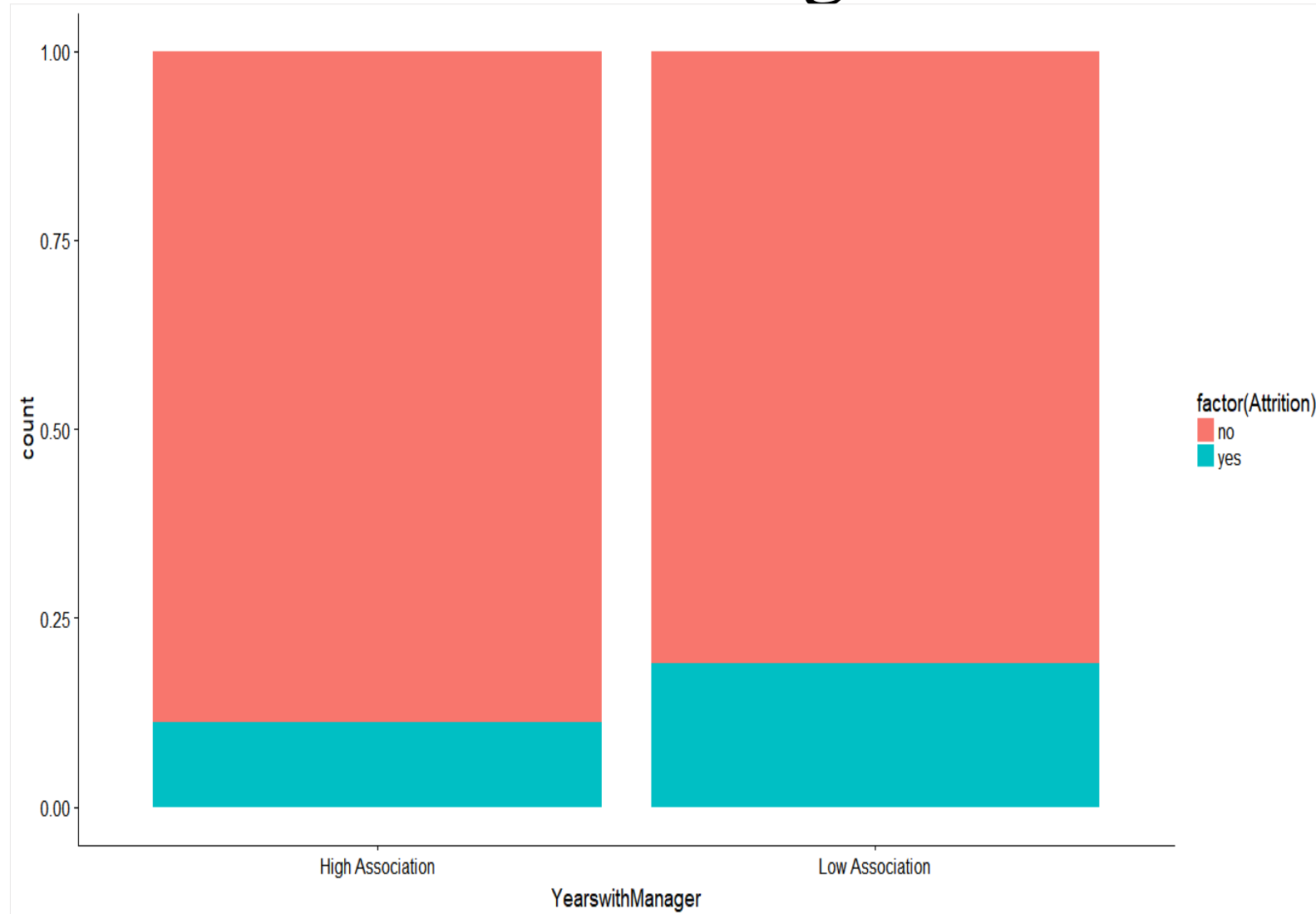
# Association with Company



## Observation

*There is a higher attrition among people who are with company for less than 5 years*

# Years With Manager vs Attrition

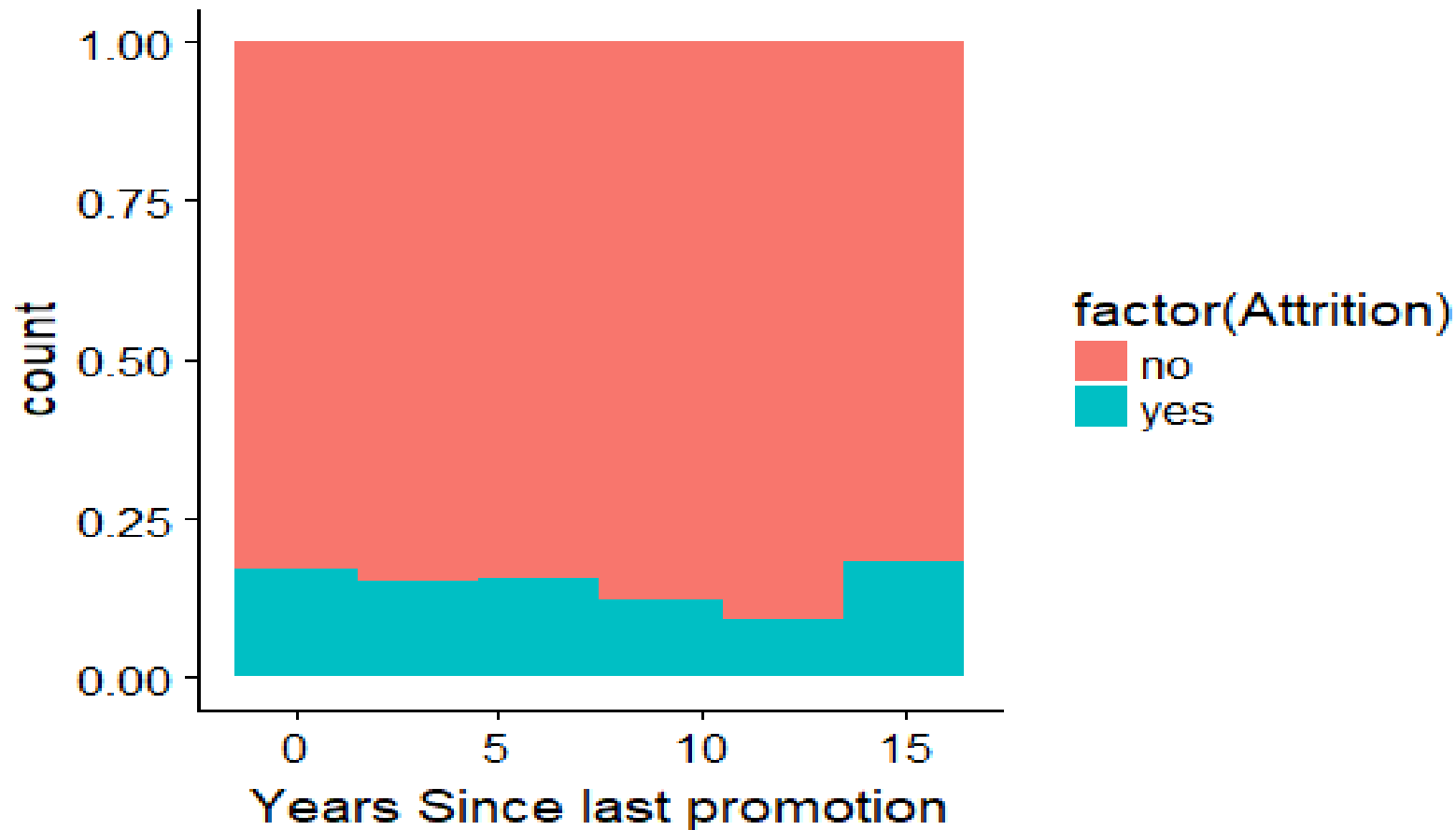


## Observation

*There is a higher attrition for low association with manager <5 years*



# Years Since Last Promotion vs Attrition

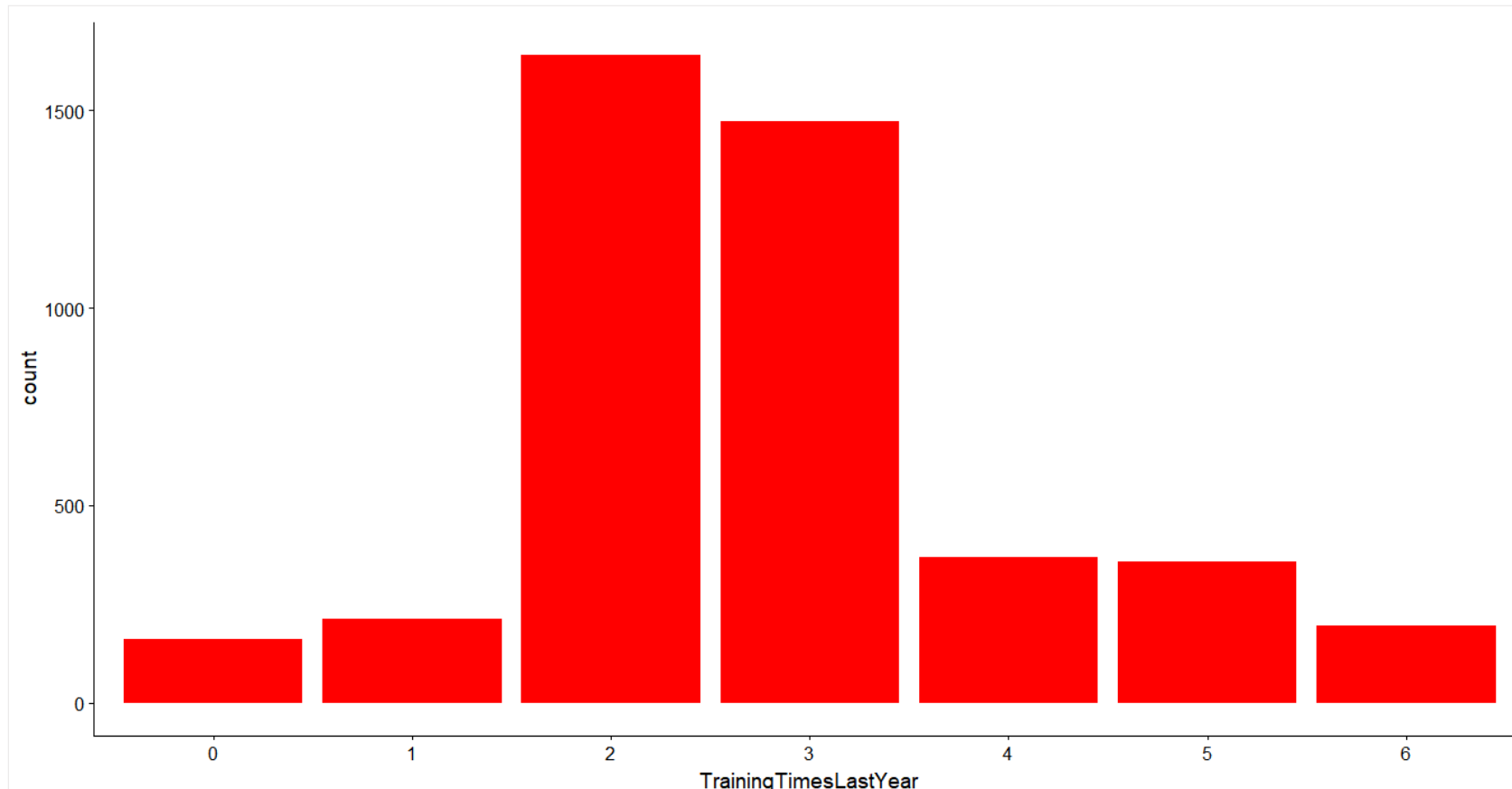


## Observation

*Attrition slightly increases after a dip with increase in YearsSince last promotion after 10 years*

*Discrete variables*  
*Univariate, Bivariate ,Segmented Analysis*

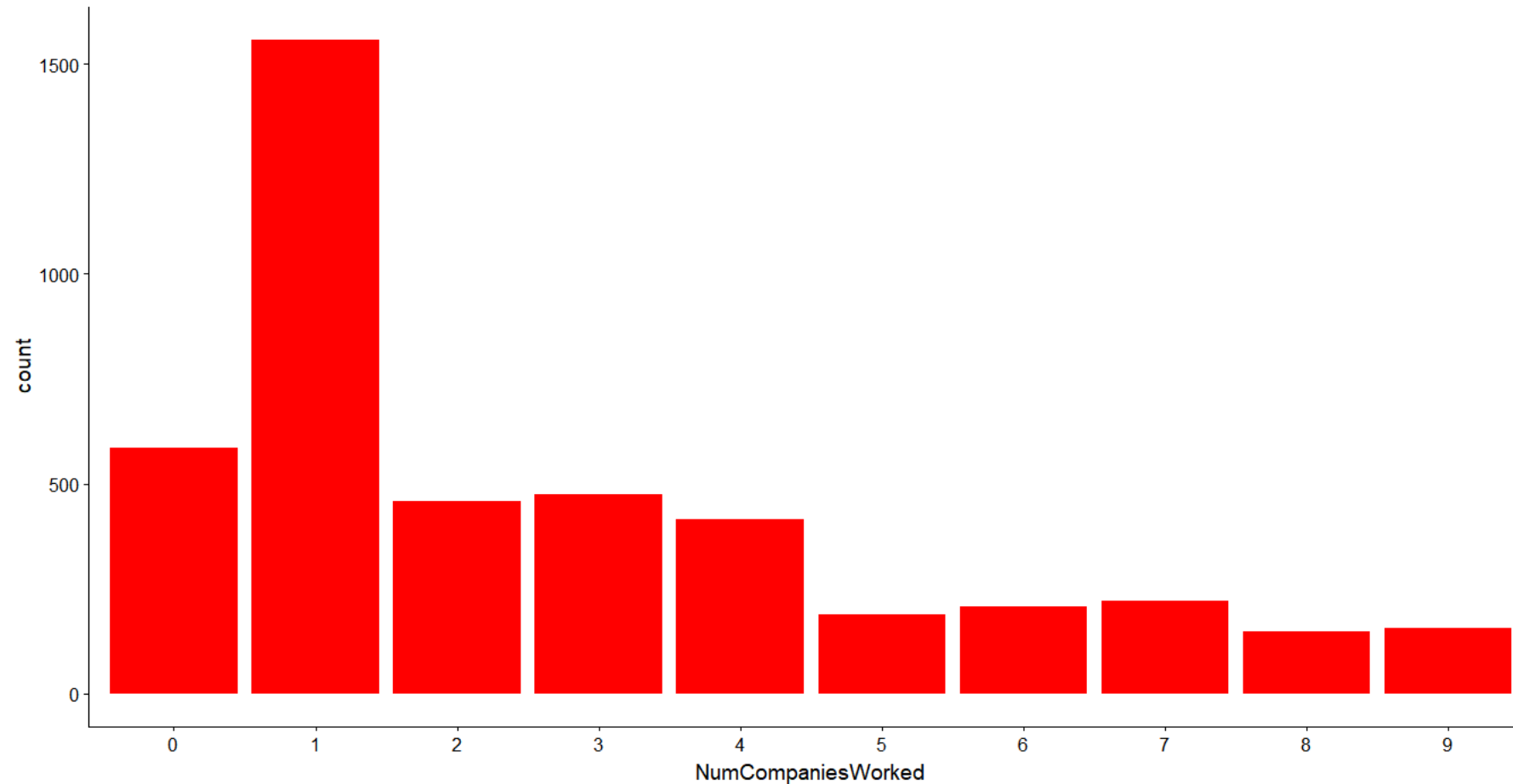
# Training Time Last Year



Observation

*Most employees have taken 2 to 3 training last year.*

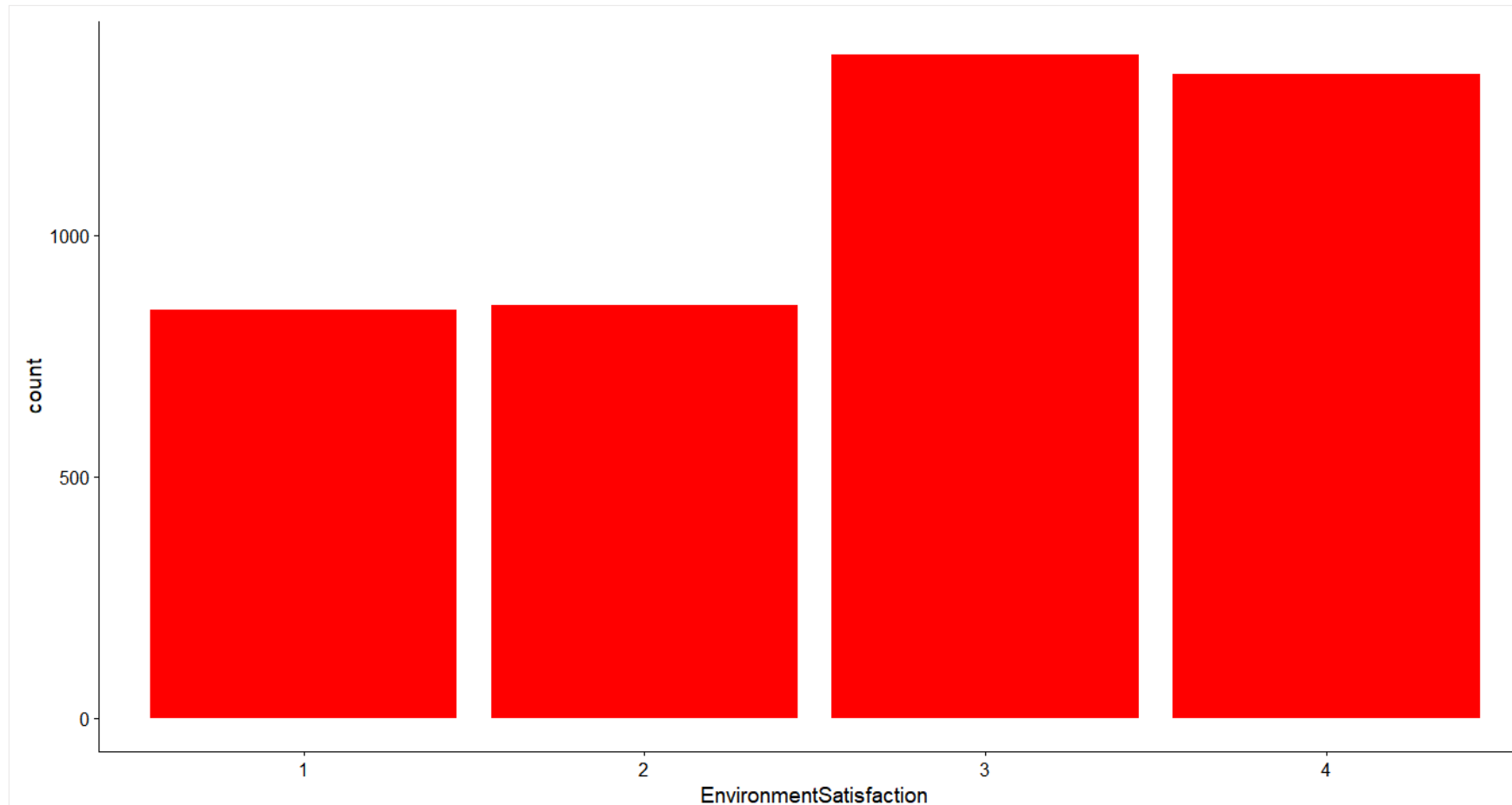
# NumCompanies Worked



Observation

*Most employees have worked for 1 company previously..*

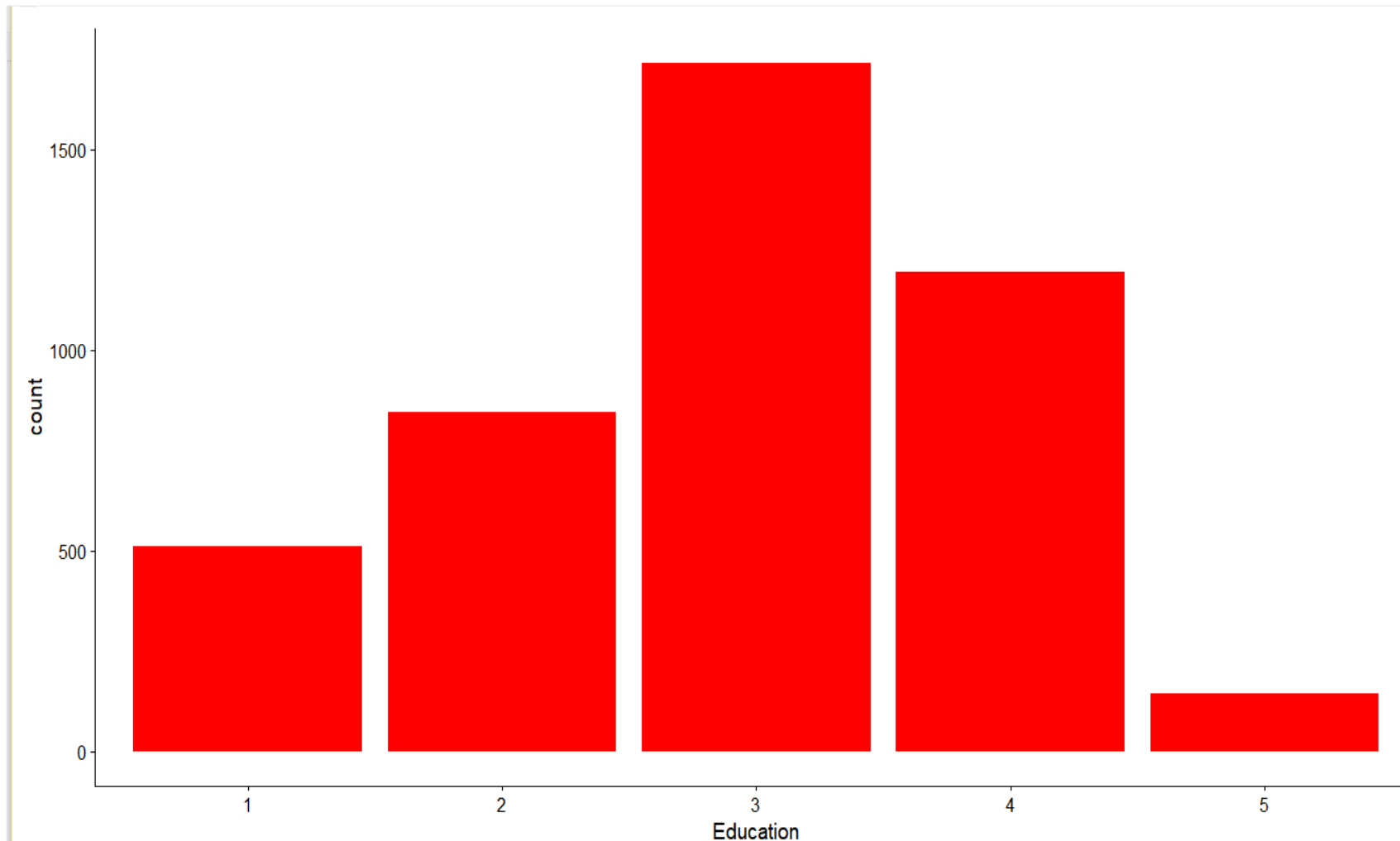
# Environment Satisfaction



## Observation

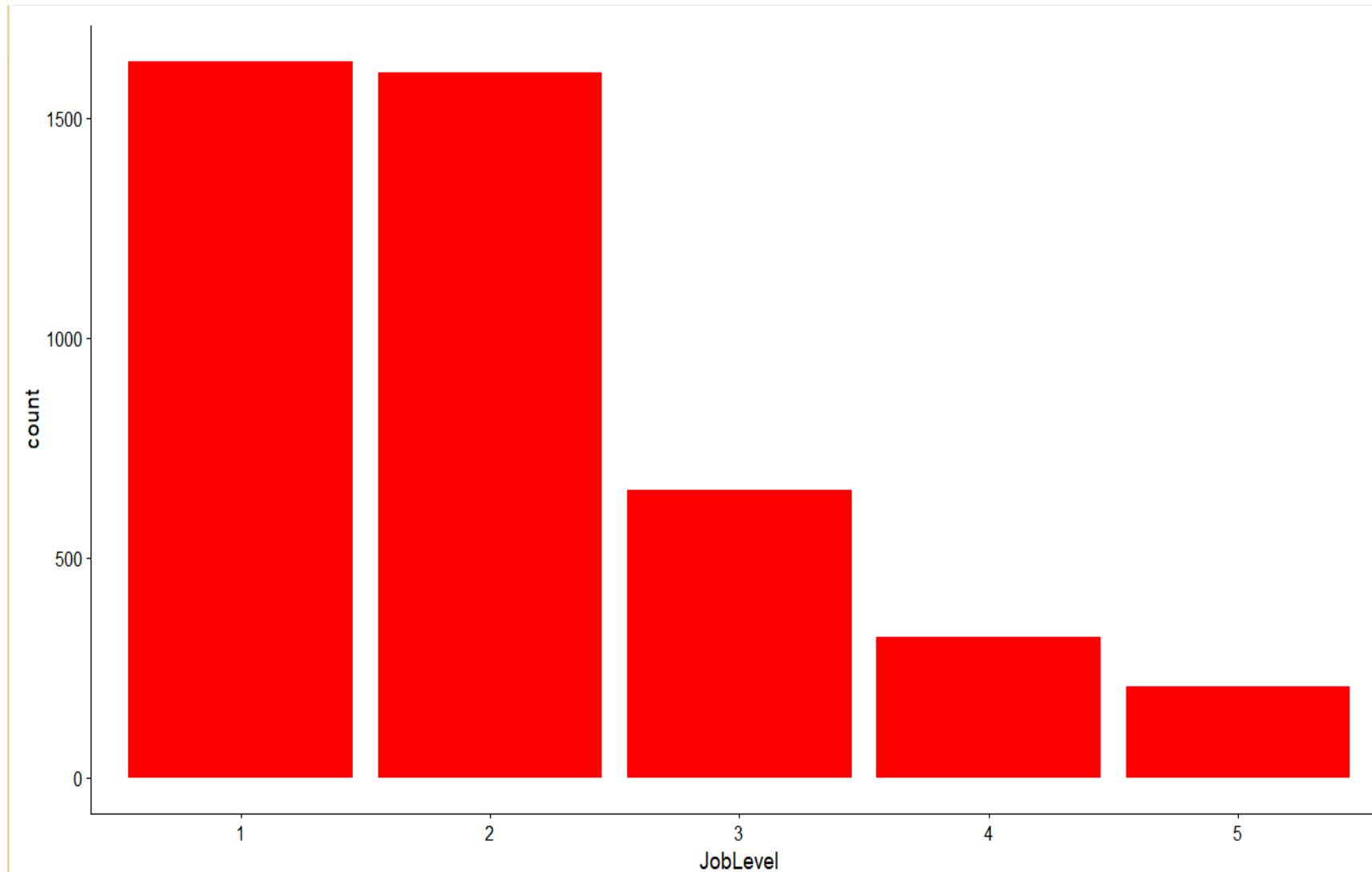
*Higher number of employees have environment satisfaction at the level 3 or 4 than 1 or 2.*

# Education



## Observation

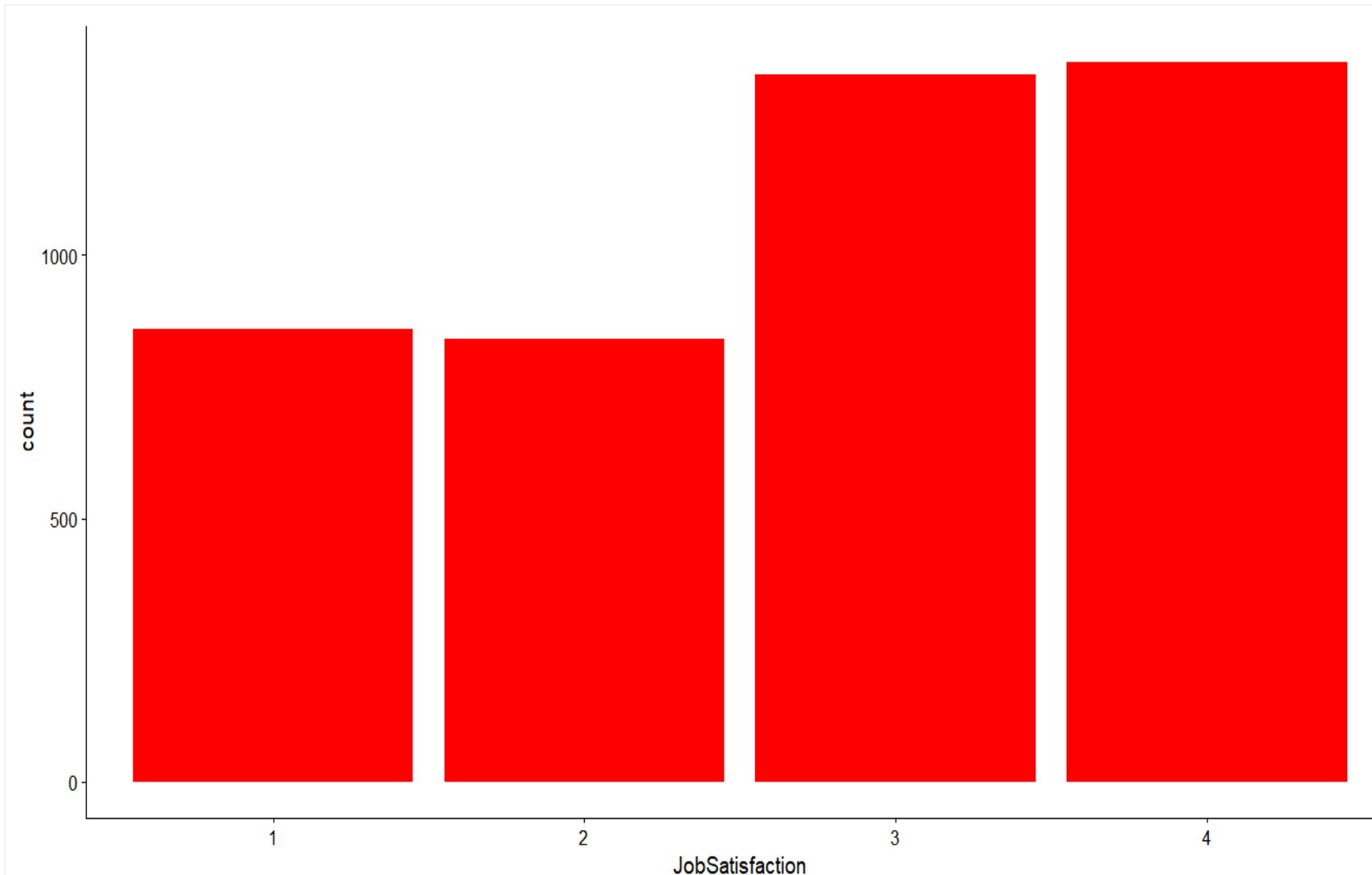
*Highest number of employees have education of the level 3 that is bachelor's level.*



Observation

*Most employees have job level 1 or 2*

# Job Satisfaction

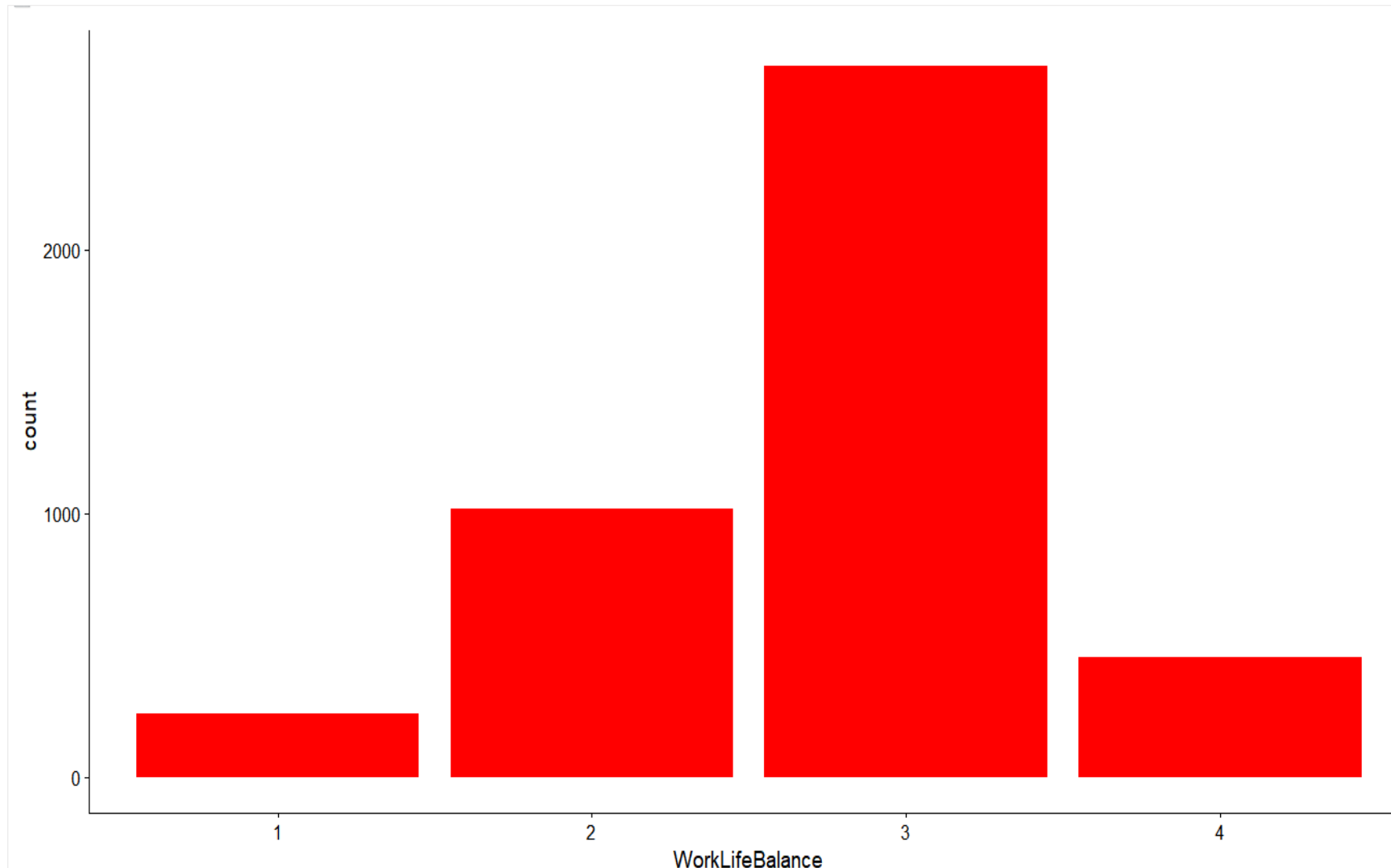


## Observation

*Most employees have expressed job satisfaction of level 3 or 4 that is high level of job satisfaction.*



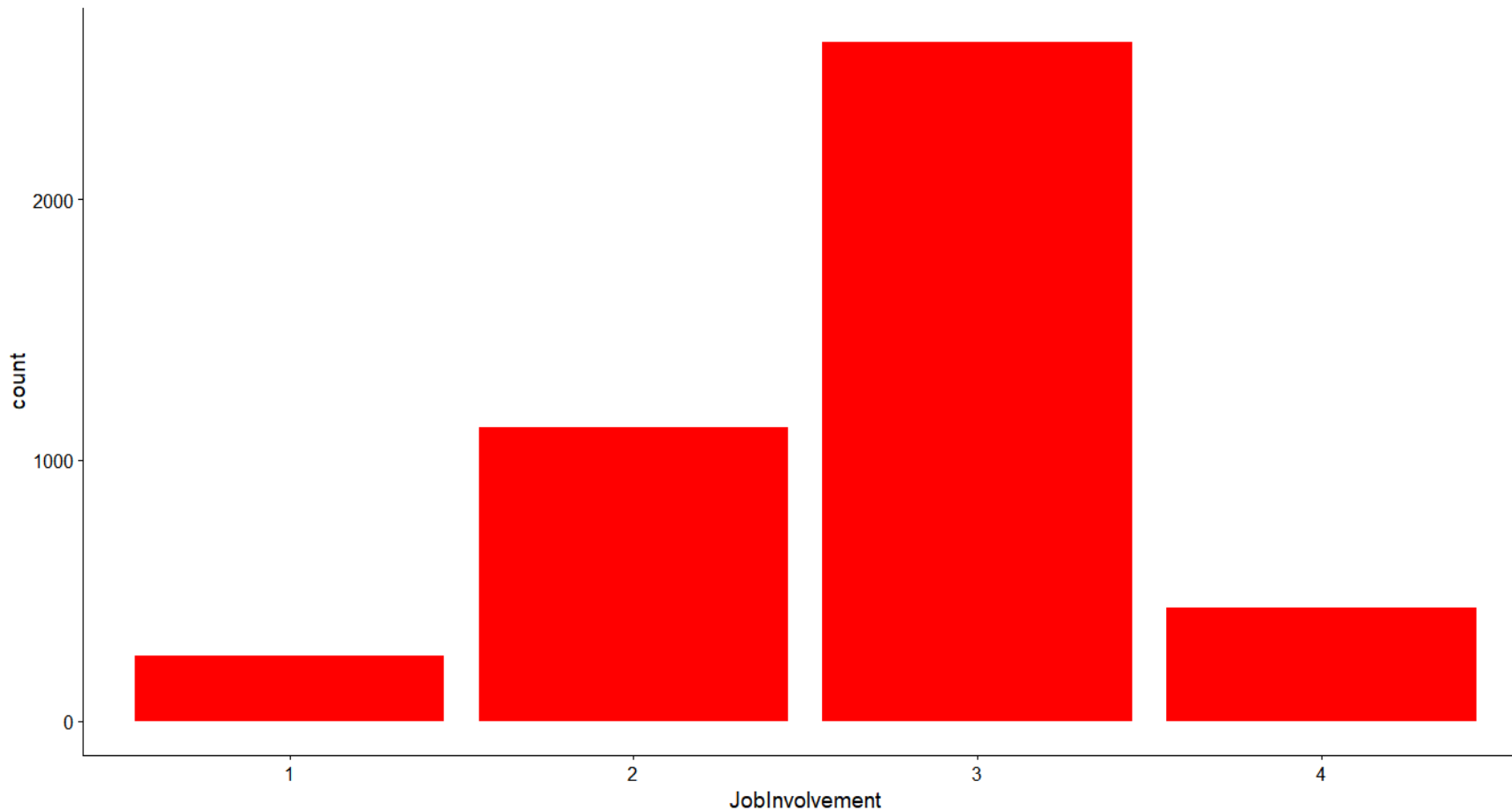
# Work Life Balance



## Observation

*Most employees have expressed better work life balance of level 3*

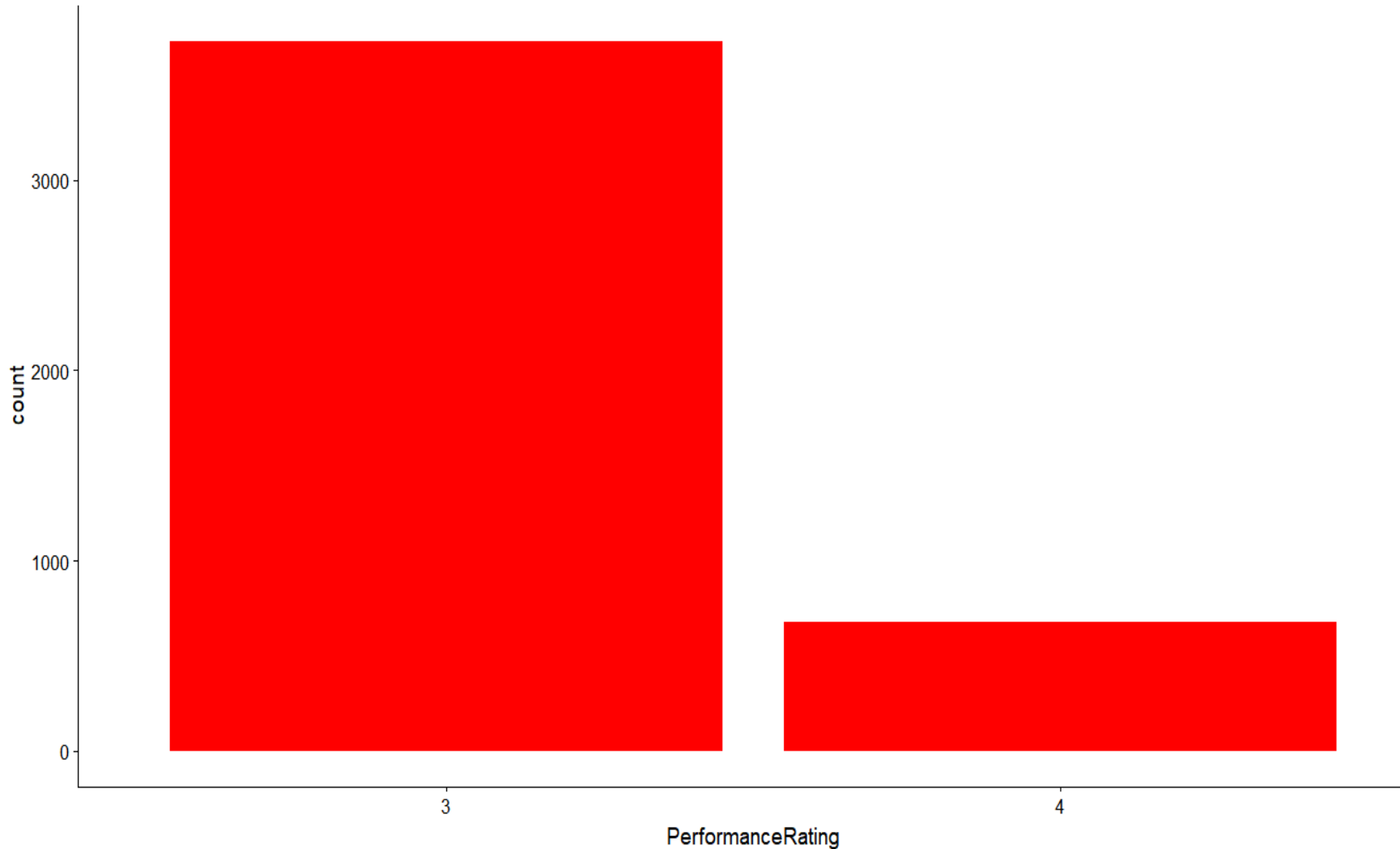
# Job Involvement



Observation

*Most employees  
have high job  
involvement of  
level 3*

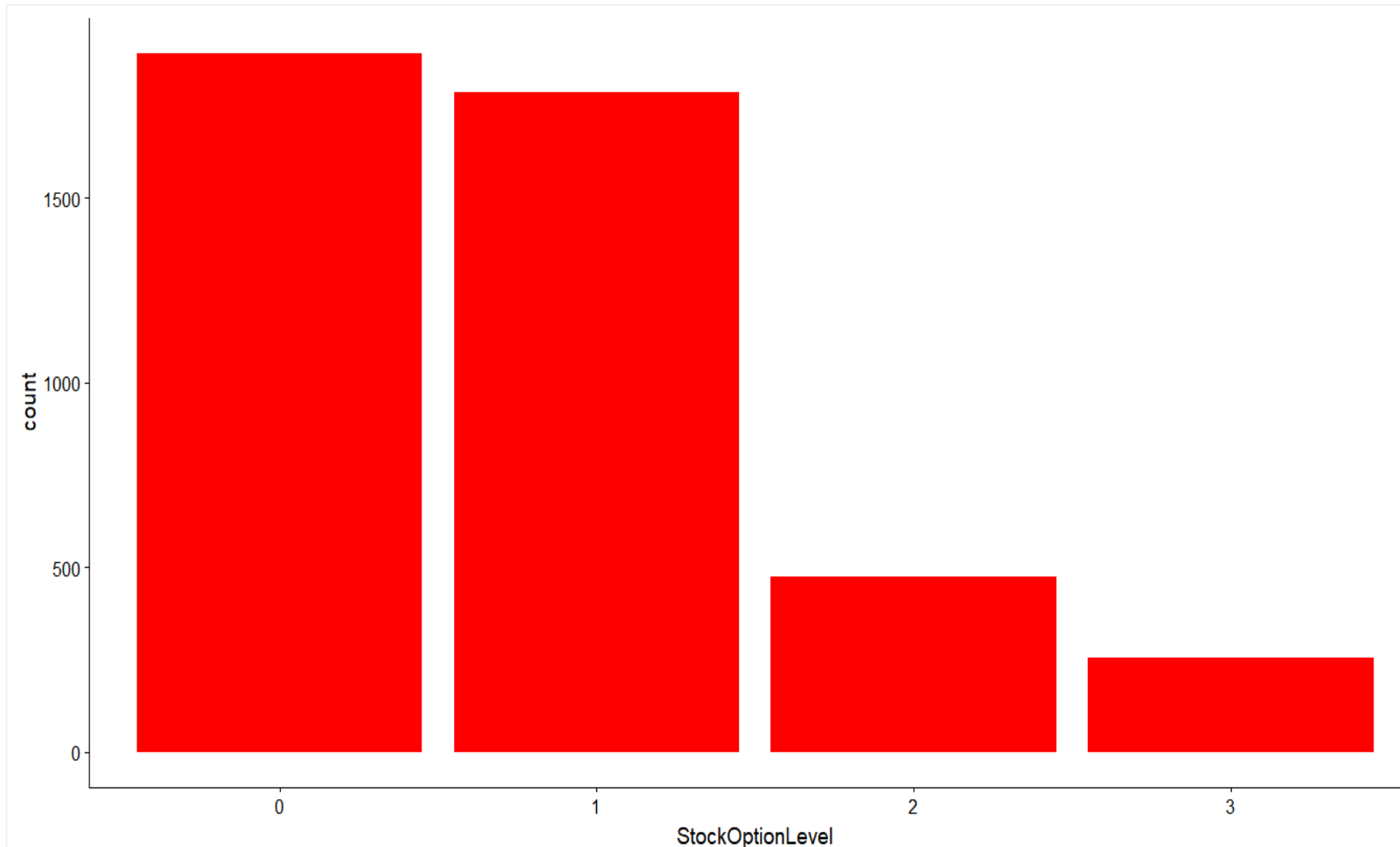
# Performance Rating



## Observation

*Most employees have received a performance rating of 3 instead of 4*

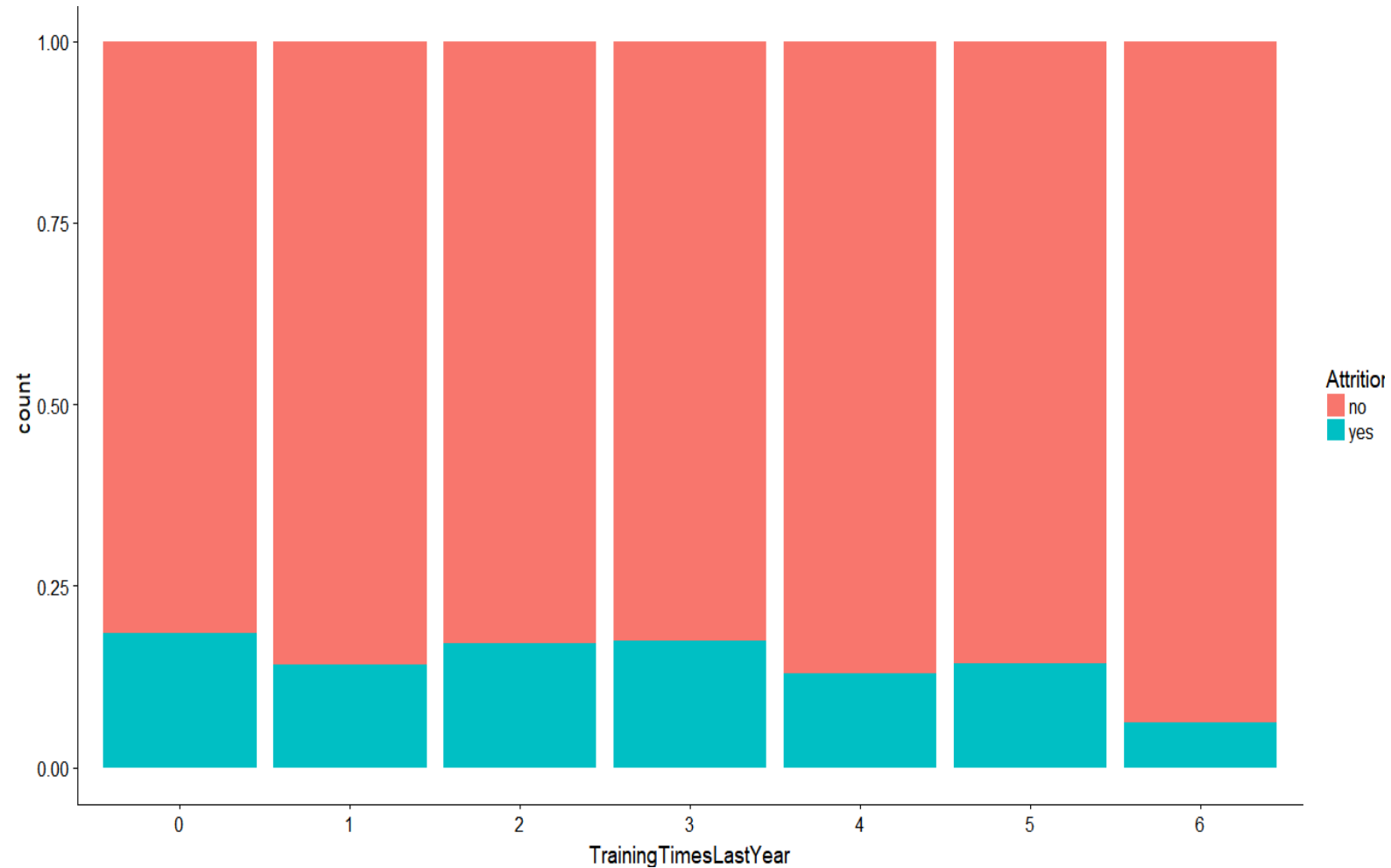
# StockOptionLevel



Observation

*Most employees have stock option level of 0 or 1.*

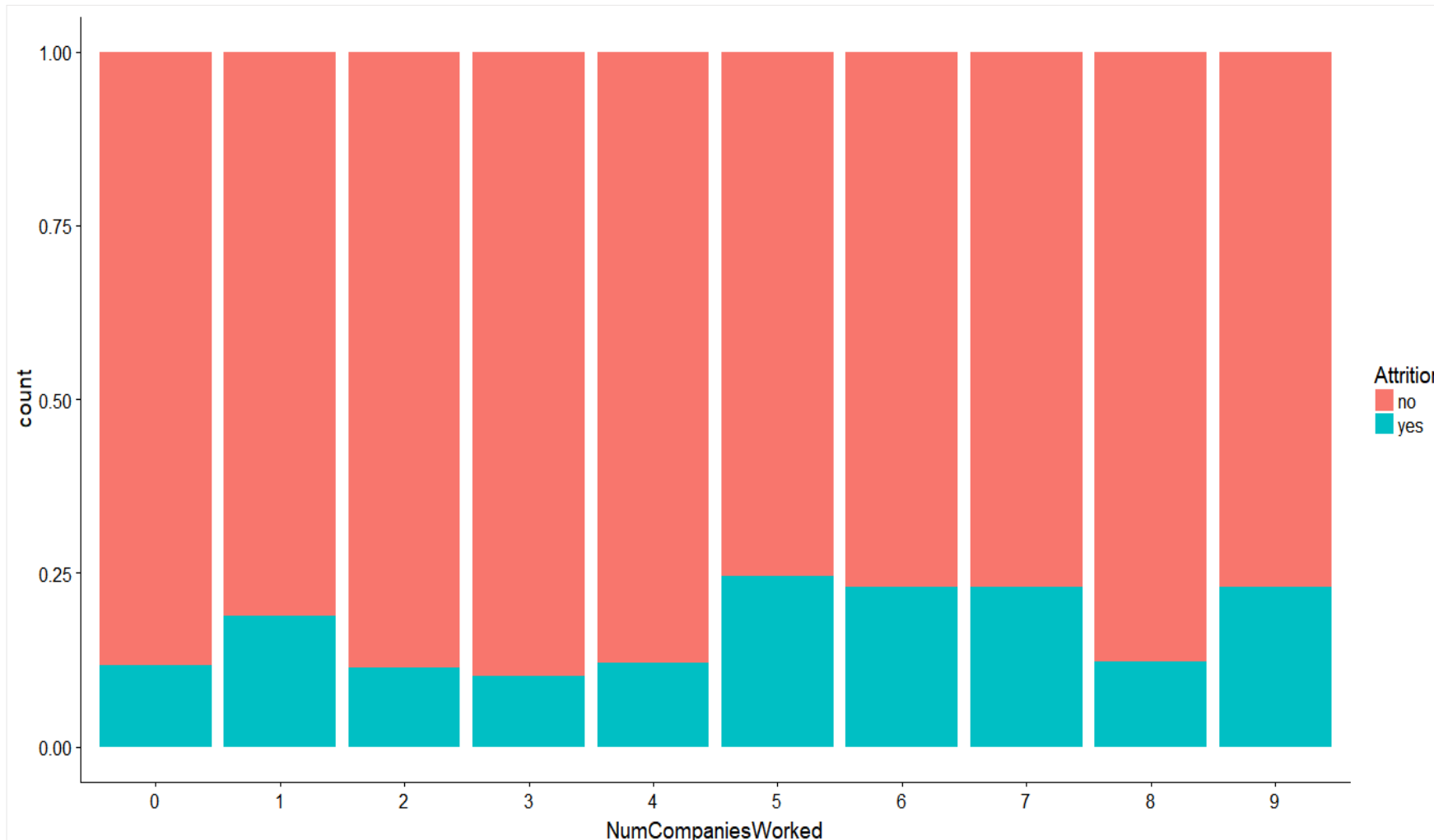
# Training Time Last Year vs Attrition



## Observation

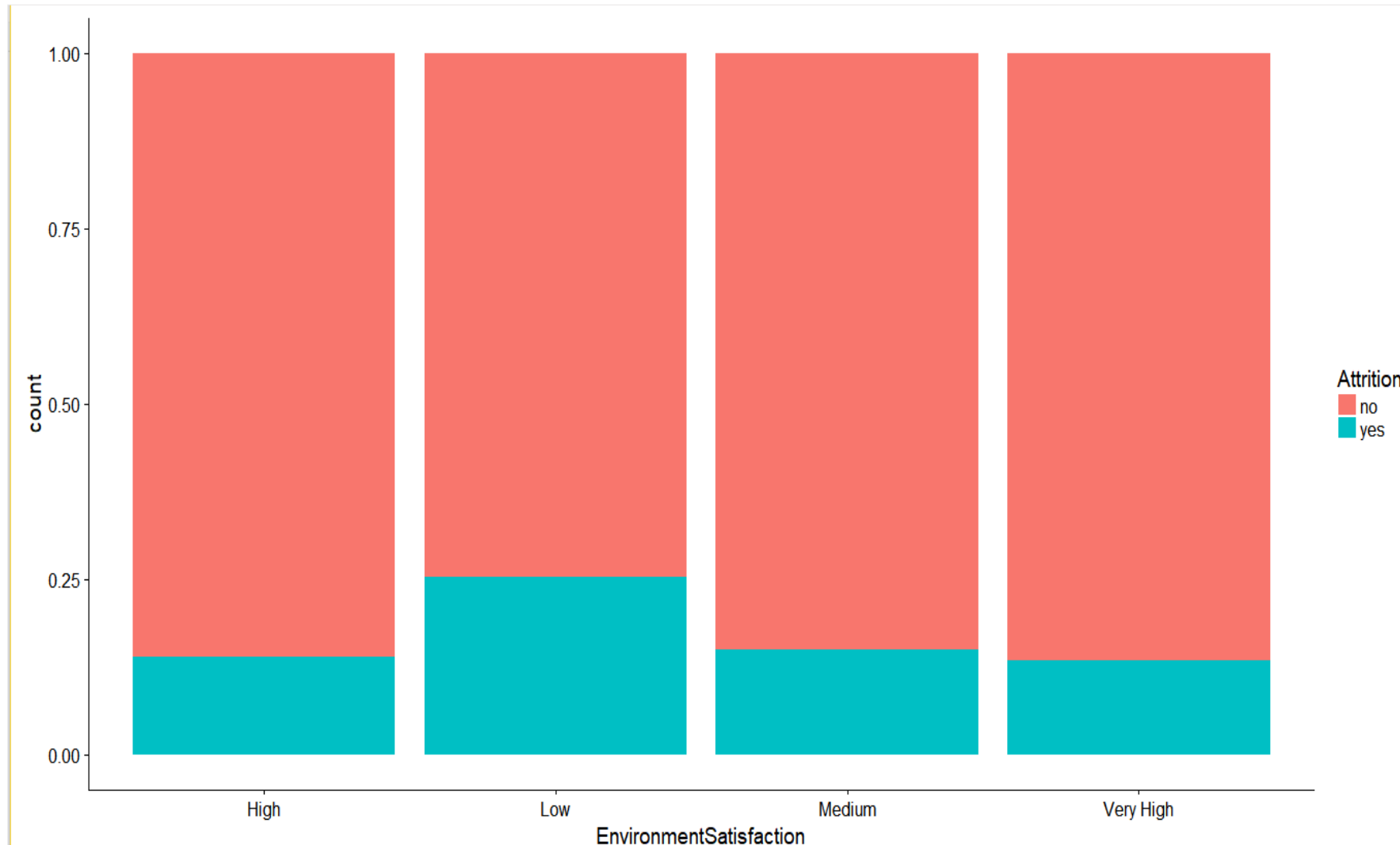
*It shows there is very low percentage of attrition with people taking 6 TrainingTimesLast Year*

# NumCompaniesWorked vs Attrition



Observation  
 Attrition is higher for employees with total number of companies that they have worked for is greater than 4 upto 9 except when it is 8.

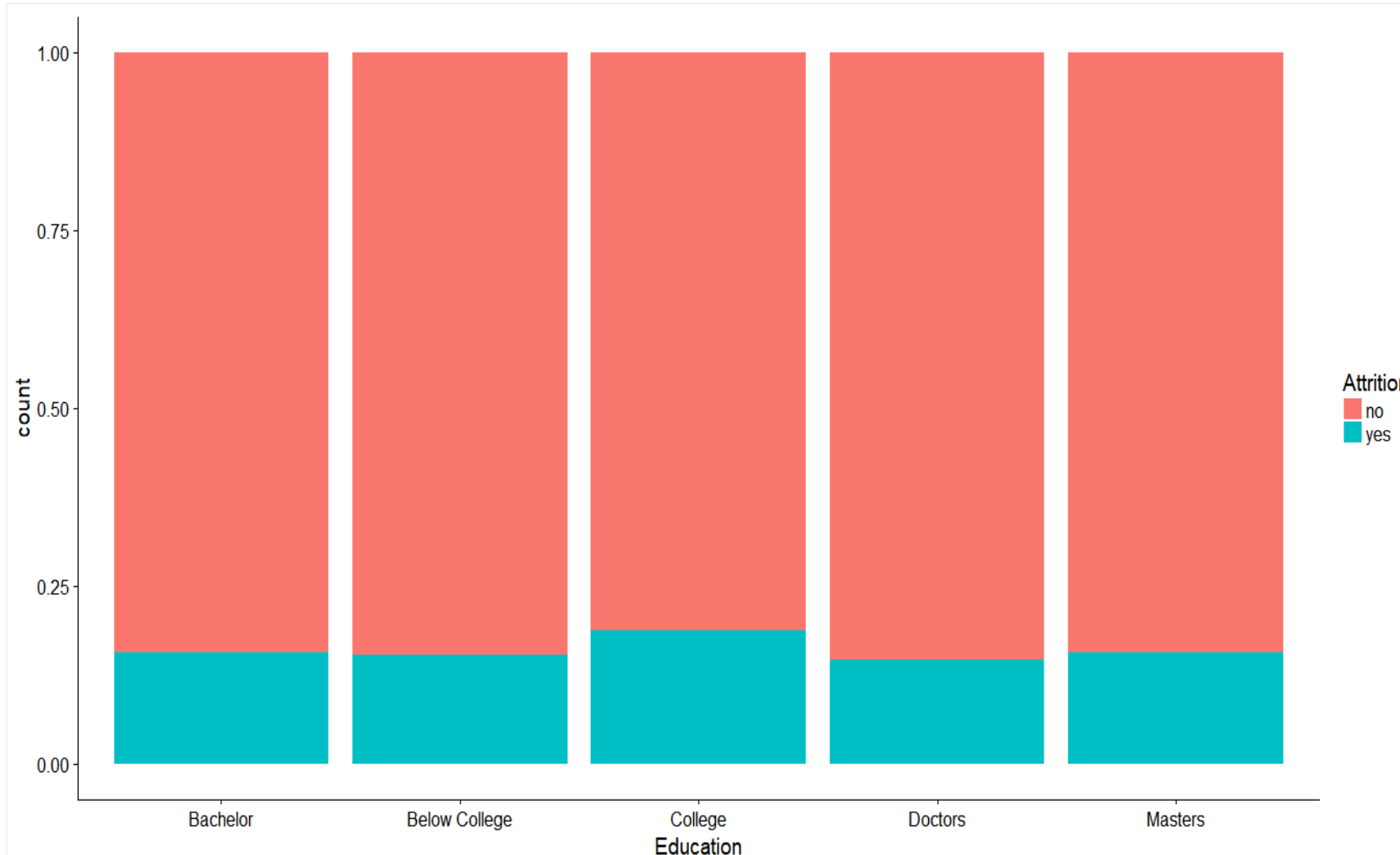
# Environment Satisfaction vs Attrition



## Observation

*Attritions are little higher when Environment satisfactions are low*

# Education vs Attrition

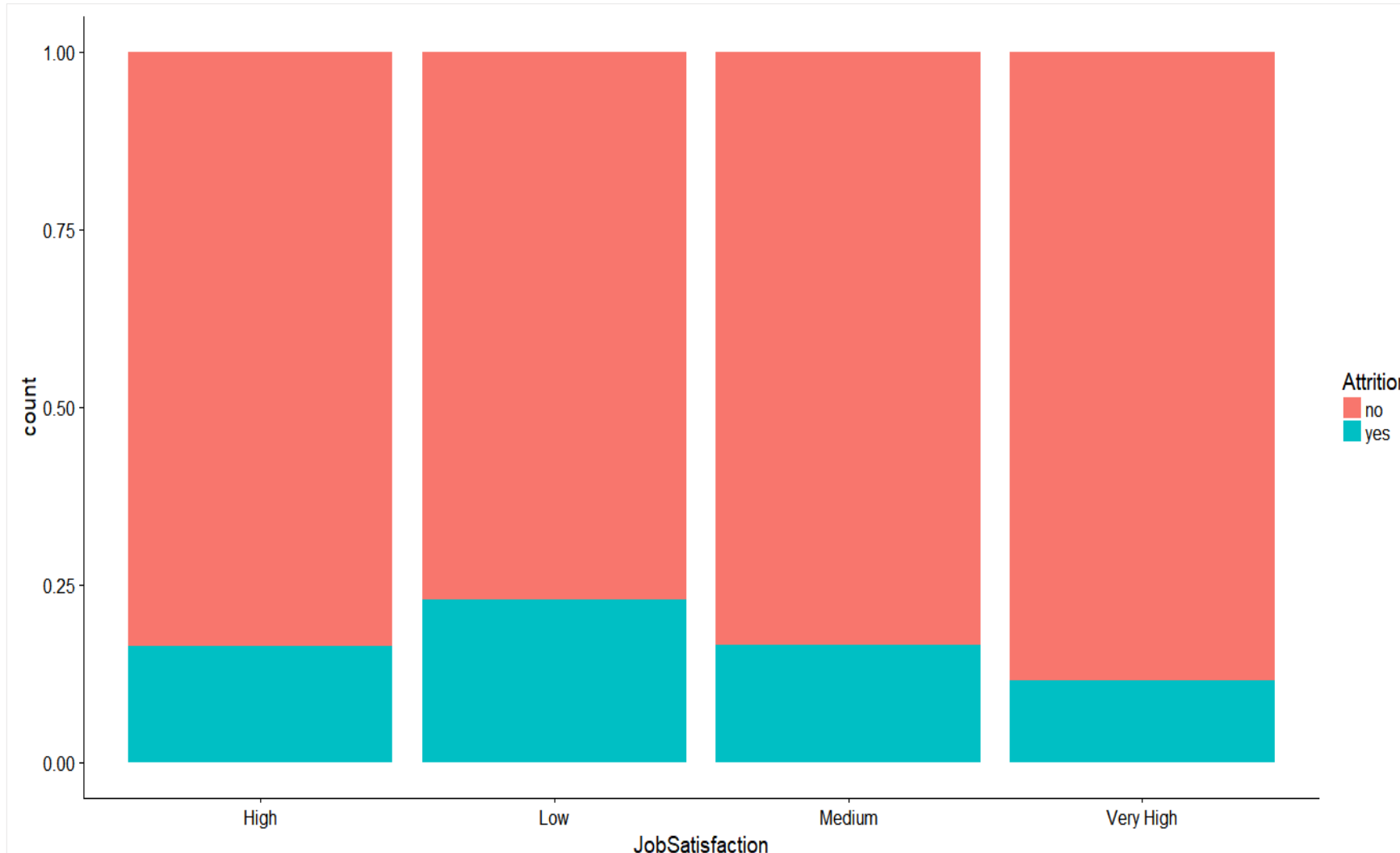


## Observation

*Attrition is little higher for people who have college education*



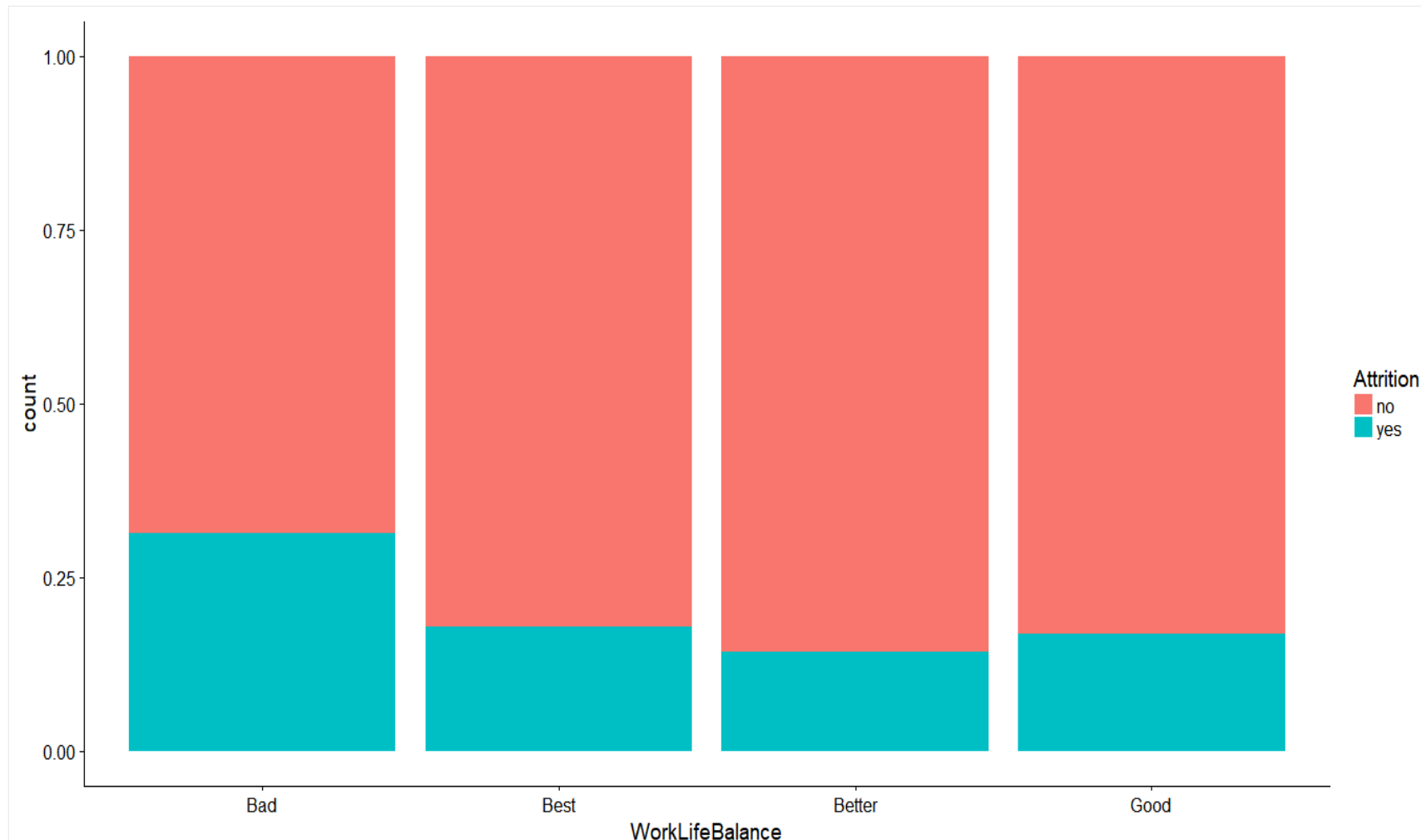
# Job Satisfaction vs Attrition



## Observation

*Attrition is little higher for employees with low Job Satisfaction*

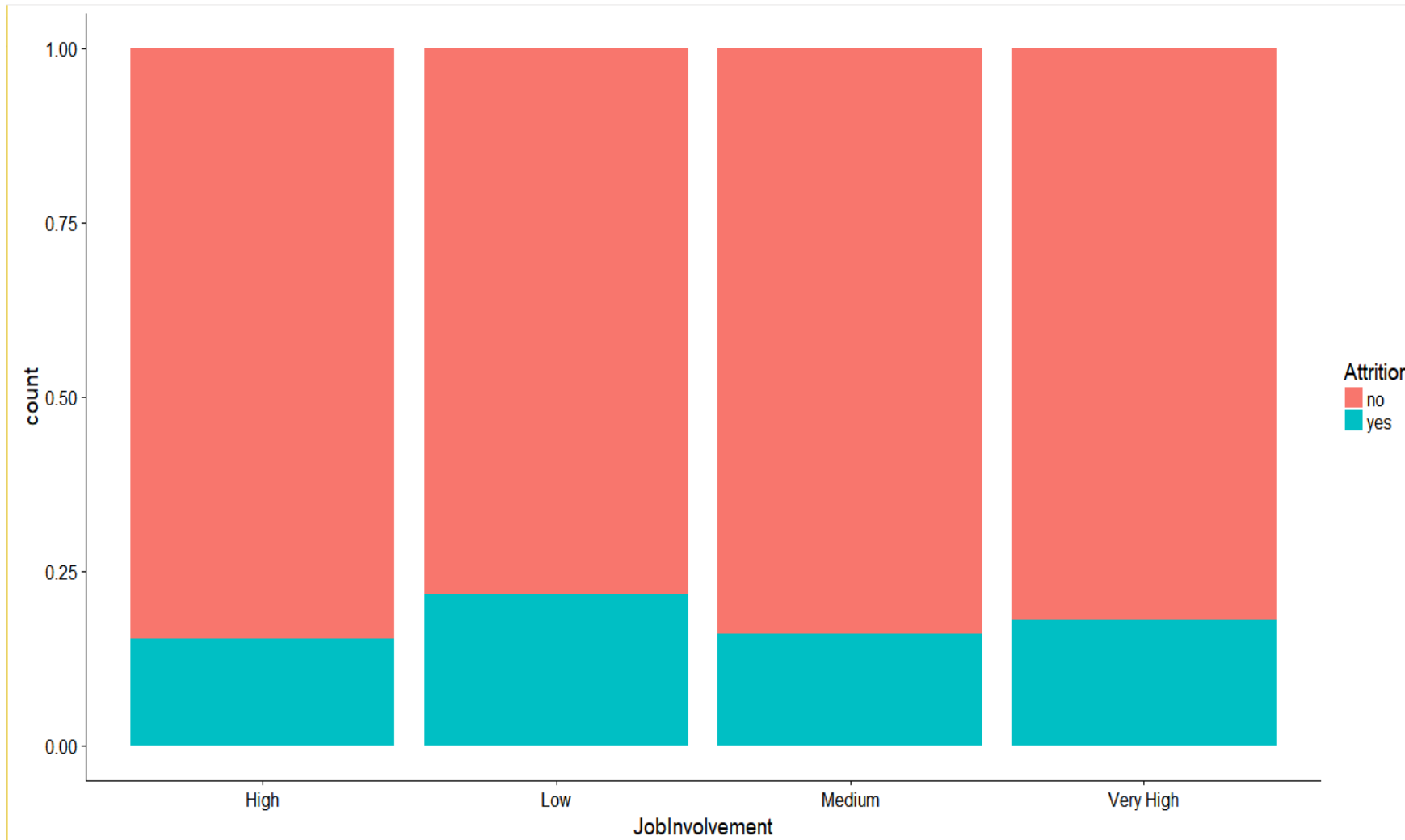
# Work Life Balance vs Attrition



## Observation

*Attrition is little higher where WorkLife Balance is bad.*

# Job Involvement vs Attrition



Observation

*Attrition is marginally higher for low job involvement.*

## MODEL OUTCOME

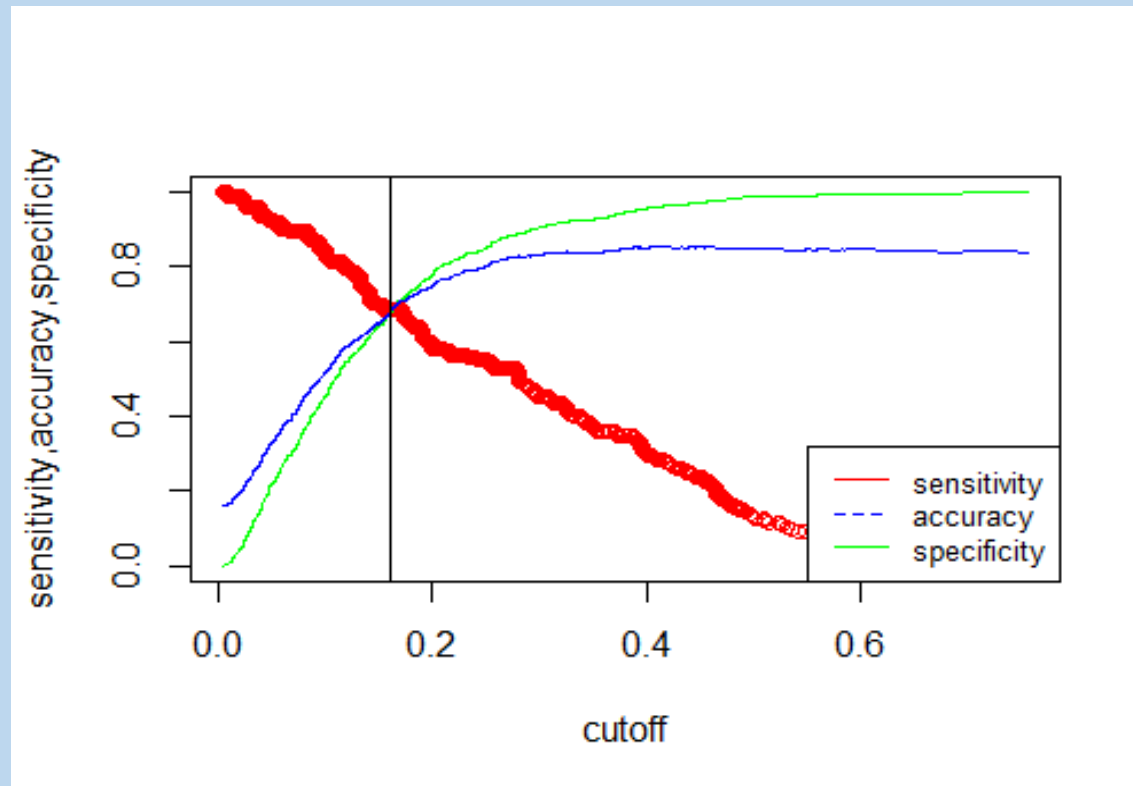
The final model selected contains the following predictors 7 with their beta coefficients given

Predictors	Coefficients	Std Errors	z value	p values
(Intercept)	-1.92751	0.08489	-22.705	< 2e-16 ***
EnvironmentSatisfaction.x4	-0.51440	0.12247	-4.200	2.67e-05 ***
JobSatisfaction.x4	-0.78326	0.12541	-6.245	4.23e-10 ***
MaritalStatus.xsingle	1.01215	0.10755	9.411	< 2e-16 ***
Age	-0.41136	0.05827	-7.059	1.67e-12 ***
average_time	0.50042	0.04990	10.029	< 2e-16 ***
YearsSinceLastPromotion	0.31267	0.06730	4.646	3.39e-06 ***
YearsWithCurrManager	-0.59299	0.07336	-8.083	6.33e-16 ***

All 7 variables have p values less than 0.001 and vif values less than 2

# SENSITIVITY, SPECIFICITY, ACCURACY OF MODEL

Sensitivity, specificity accuracy of the model for different cutoff values for probability of attrition



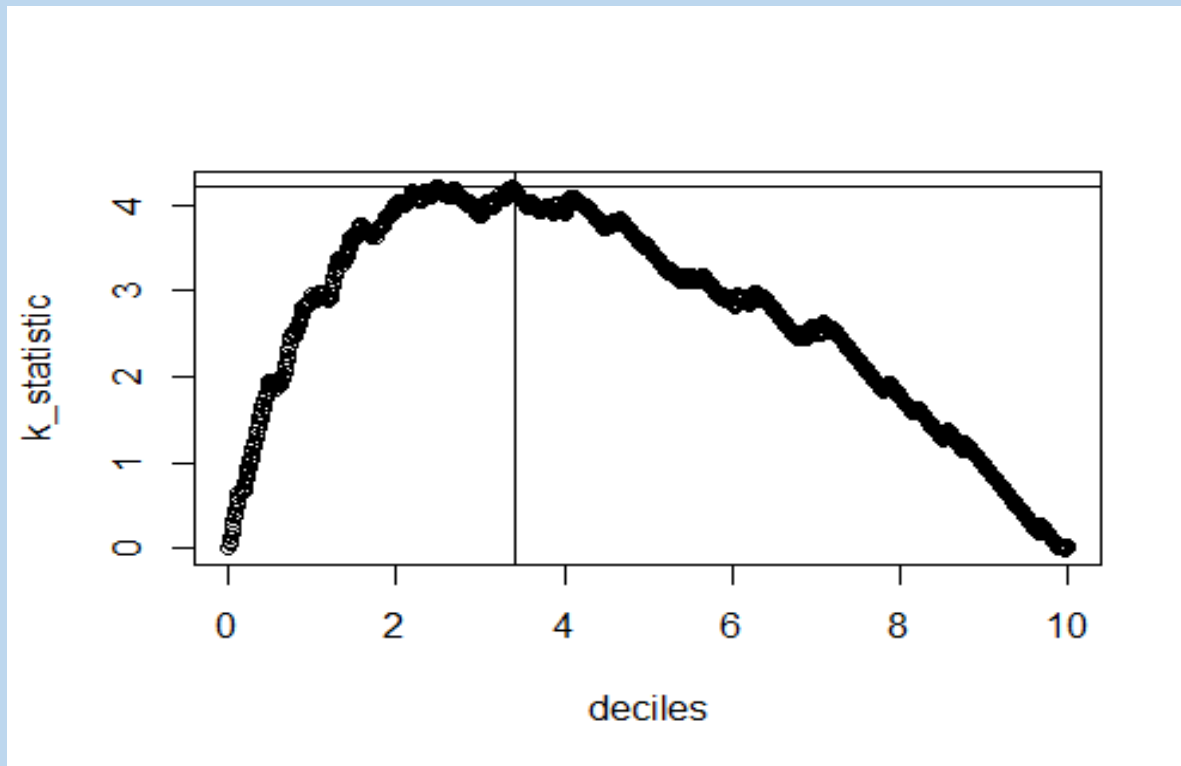
## SENSITIVITY, SPECIFICITY, ACCURACY OF MODEL

Sensitivity, specificity, accuracy of the model for 3 different sets of testing and training data (bootstrapping method)

training and testing dataset iteration number	sensitivity	specificity	accuracy	cutoff	K statistic
1	0.7	0.7	0.7	0.16	0.42
2	0.72	0.72	0.72	0.18	0.45
3	0.69	0.69	0.69	0.16	0.41

- Sensitivity, specificity, accuracy of the model for 3 different sets of testing and training data are very similar .
- This makes us confident of our model.

# K S CHART



K statistic value for the model is 0.42 and lies within 4th decile

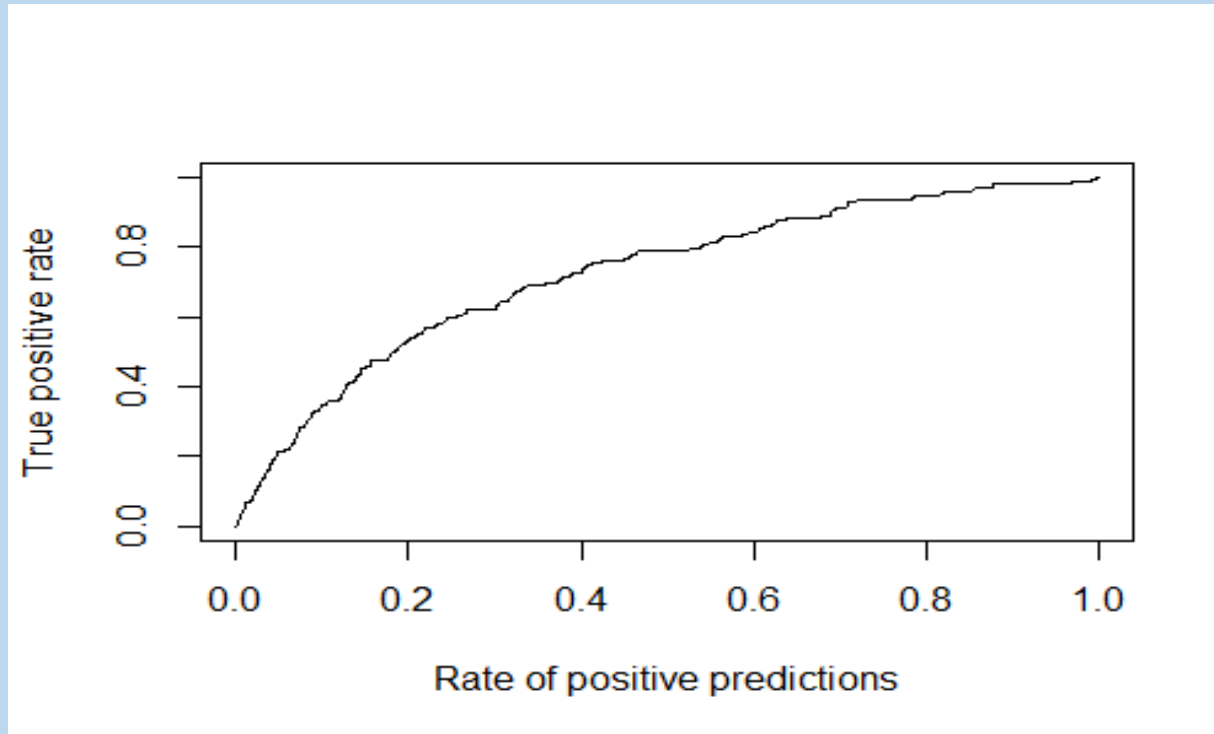
## CONFUSION MATRIX OF THE MODEL

---

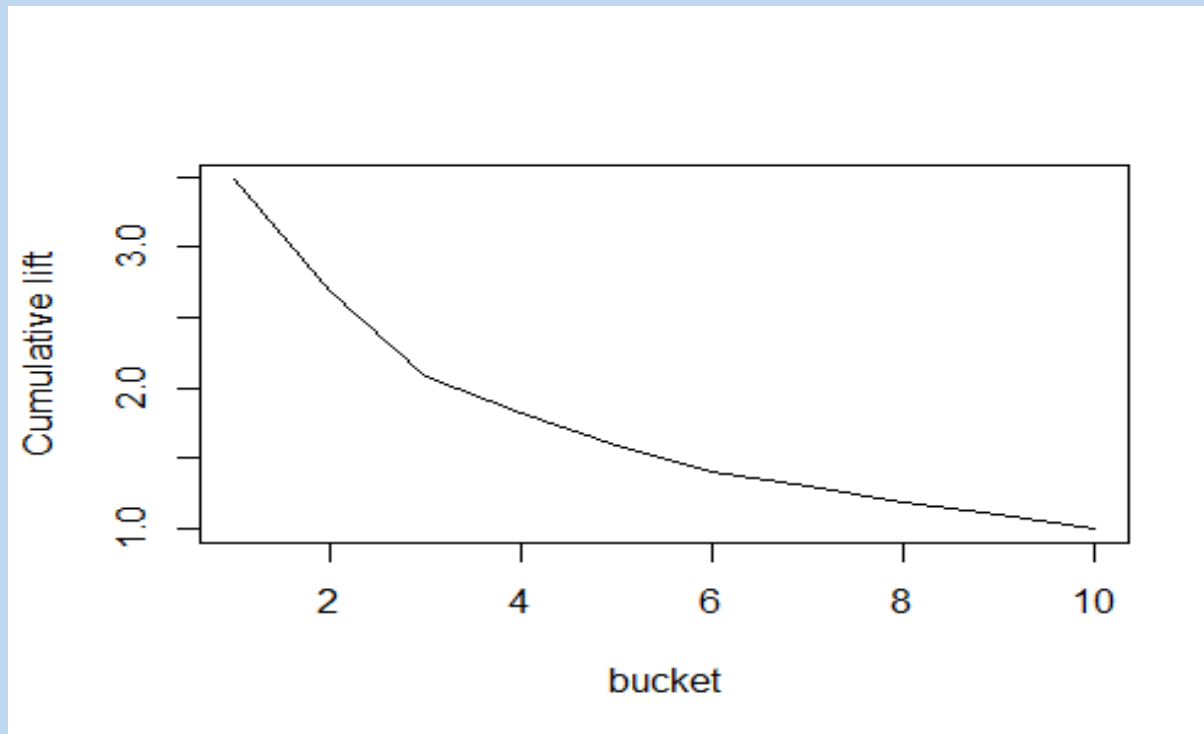
	Reference	
Prediction	No	Yes
No	777	64
Yes	333	149



# GAIN CHART

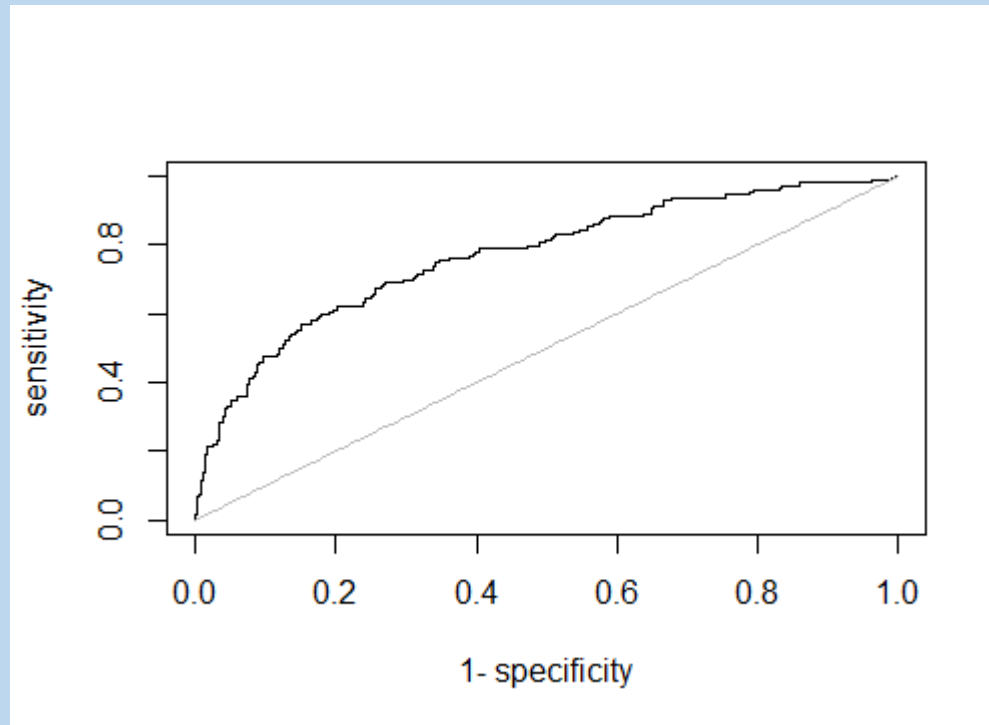


# LIFT CHART



CUMULATIVE LIFT IS MORE THAN 3 WITHIN FIRST DECILE

## ROC CURVE FOR THE MODEL



Area under the ROC curve is 0.77

## CONCLUSIONS

---

- Our model shows all model evaluation parameters like accuracy, sensitivity, specificity, k statistic and area under ROC curve within acceptable range.

EDA and our model shows:

- 4 variables are negatively correlated to attrition:
  - Environment satisfaction
  - Job satisfaction
  - Age
  - Years of association with current manager
- 3 variables are positively correlated to attrition:
  - Single marital status
  - Years since last promotion
  - Average number of hours spent at work per day.

## RECOMMENDATIONS

---

XYZ company is recommended to take feedback from employees who have low environment satisfaction and low job satisfaction especially from those who are young, single, have low number of association with their current managers , have spent average longer hours at work and have not been promoted .

Such category of employees are more prone to leave organization and increase attrition rate. They might need right mentoring, career guidance, skill enhancement, etc inorder to turn their jobs as their career progression platform.

THANK YOU