CaseStudy2

Dawkins

2/6/2021

Libraries

```r
library(tidyr)
library(readr)
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
```

```r
library(ggplot2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

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

```r
library(caret)
```

```
## Loading required package: lattice
```

```r
library(e1071)
```

Load Data

```r
CaseStudy2_data <- read_csv("C:/Users/OaklandHillsMansion/OneDrive - Southern Methodist University/SMU/
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   Attrition = col_character(),
##   BusinessTravel = col_character(),
##   Department = col_character(),
##   EducationField = col_character(),
##   Gender = col_character(),
##   JobRole = col_character(),
##   MaritalStatus = col_character(),
##   Over18 = col_character(),
##   OverTime = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```r
# Columns that need to be removed
removed_columns = c("ID", "EmployeeCount", "StandardHours")

# Numerical Columns that, need to be converted to factors
num_fact = c("Attrition","RelationshipSatisfaction", "StockOptionLevel", "WorkLifeBalance", "JobSatisfa


# Remove "removed_columns" from data set
# Create "Legacy" column and unite "Department" & "JobRole"
df <- as.data.frame(CaseStudy2_data) %>%
  select(-all_of(removed_columns)) %>%
  mutate(Legacy = YearsInCurrentRole*TotalWorkingYears*JobLevel*EnvironmentSatisfaction*MonthlyIncome*Ag
  unite("DeptRole",Department, JobRole)

# Convert "num_fact" cols to factors
for (i in 1:length(num_fact)){
  df[,num_fact[i]] <- as.factor(df[,num_fact[i]])
}

# Selects all "char" & "factor" columns
df_char_fac <- df %>%
  select_if(function(col) is.character(col) | is.factor(col))

# Selects all "numeric" columns and keeps the "Attrition" column
df_nums <- df %>%
  select_if(function(col) is.numeric(col))

# Summary
summary(df$Attrition)
```

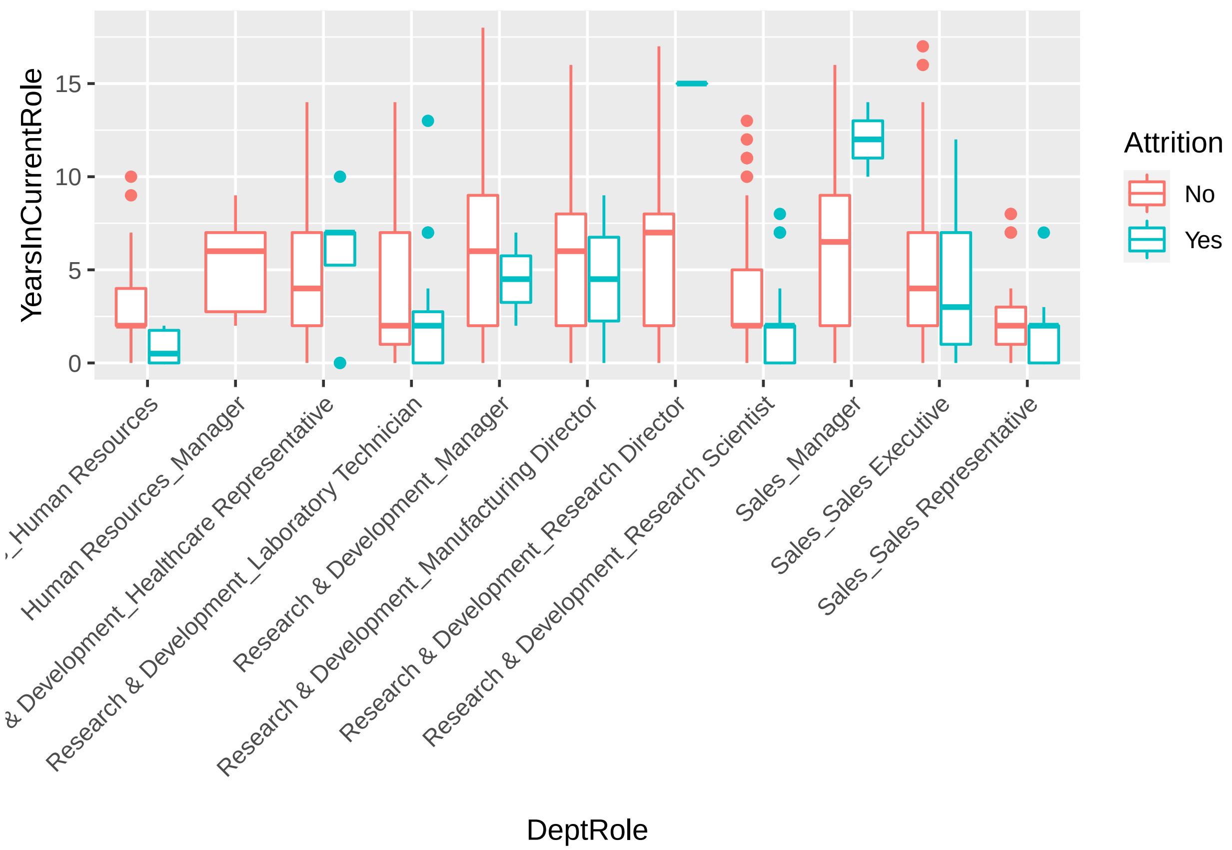
```
##  No Yes
## 730 140
```

```r
summary(df$MonthlyIncome)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1081    2840    4946    6390    8182   19999
```

Box Plot: YearsInCurrentRole vs Dept Role

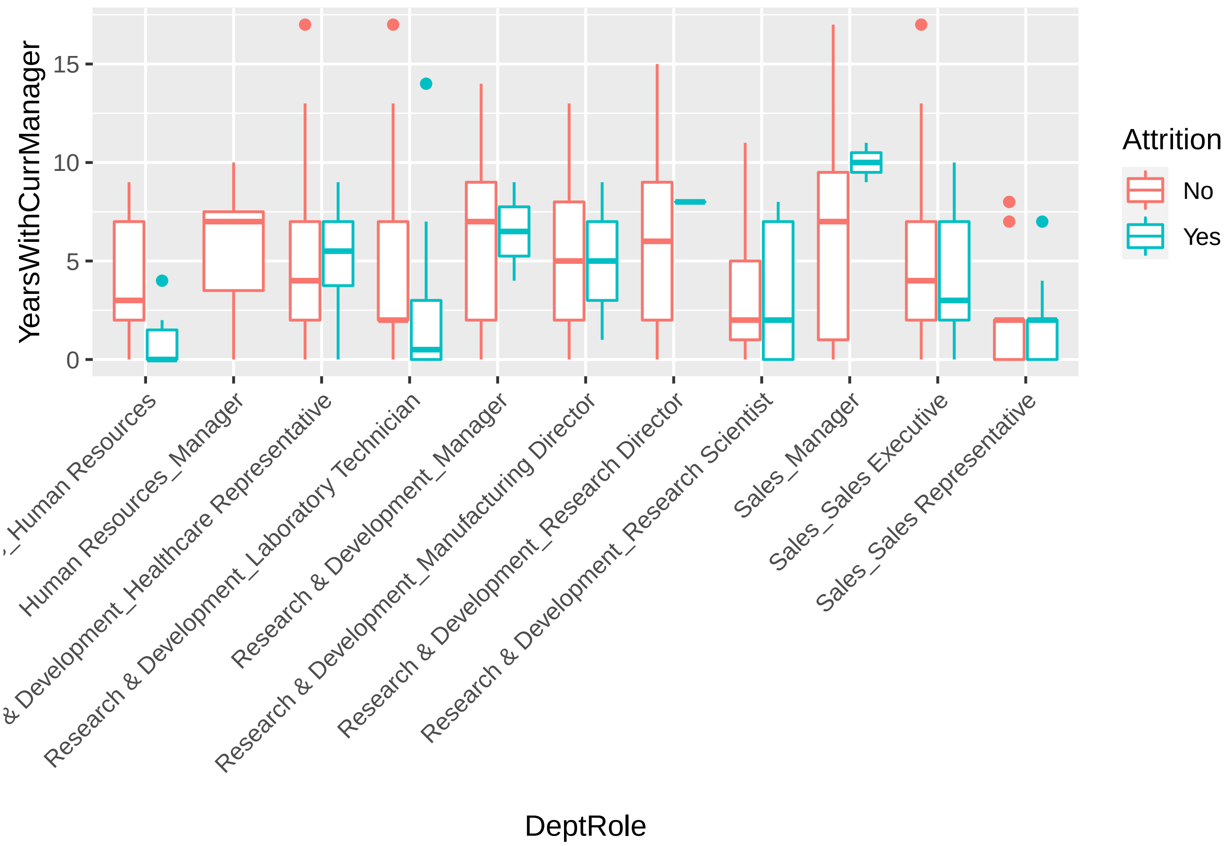
```r
# Box Plot: YearsInCurrentRole vs Dept Role
  plot_yrRole_deptRole <- df %>%
    ggplot(aes_string(y="YearsInCurrentRole",x="DeptRole", color="Attrition"))+
    geom_boxplot()+
    theme(axis.text.x=element_text(angle=45,hjust=1))

plot_yrRole_deptRole
```



Box Plot: YearsWithCurrManager vs Dept Role

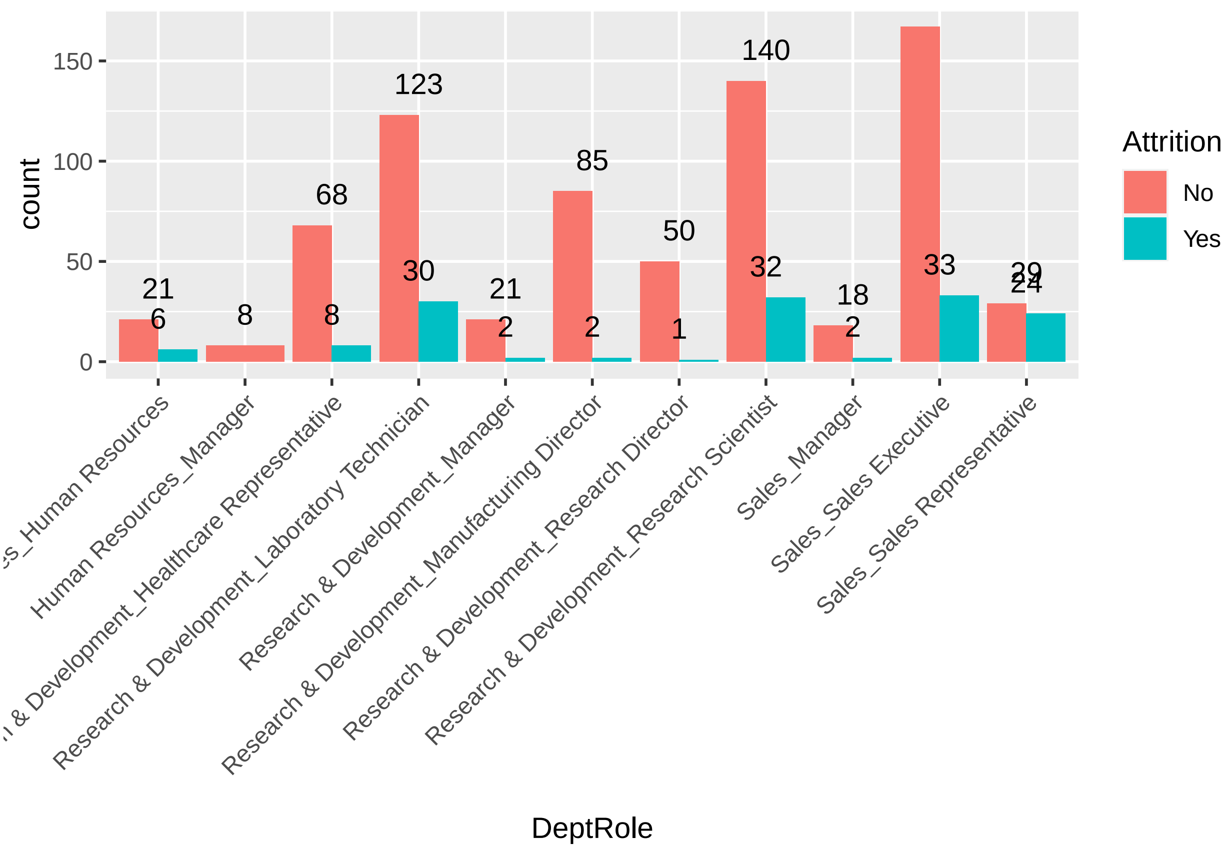
```r
# Box Plot: YearsWithCurrManager vs Dept Role
plot_yrMgr_deptRole  <- df %>%
```

```
  ggplot(aes_string(y="YearsWithCurrManager",x="DeptRole", color="Attrition"))+
  geom_boxplot()+
  theme(axis.text.x=element_text(angle=45,hjust=1))

plot_yrMgr_deptRole
```



Bar Graph: Attritionn Count by Dept Role

```
bar_attr_deptRole <- df %>%
  ggplot(aes_string(x="DeptRole", fill="Attrition"))+
  geom_bar(position = "dodge")+
  geom_text(stat='count', aes(label=..count..), vjust=-1)+
  theme(axis.text.x=element_text(angle=45,hjust=1))

bar_attr_deptRole
```



Build and Cross Validate Naive Bayes Model

```
# Features included in the model
features <- c("MaritalStatus", "OverTime", "DeptRole", "Legacy", "MonthlyIncome", "TotalWorkingYears",

iterations = 100

# Initialize Matrix
masterAcc = matrix(nrow=iterations)
masterSens = matrix(nrow=iterations)
masterSpcf = matrix(nrow=iterations)

# Set split percentage
splitPerc = .7

# Loop through model runs
set.seed(5)
for(j in 1:iterations)
{
  # Get trainIndices
  trainIndices = sample(1:dim(df)[1], round(splitPerc*dim(df)[1]))
  # Get train data frame
  train = df[trainIndices,]
```

```
  # Get test data frame
  test = df[-trainIndices,]

  # Create model using the train data frame
  model = naiveBayes(train[,features], train$Attrition)
  # Create table
  table(predict(model, test[,features]), test$Attrition)
  # Get Confusion Matrix
  CM = confusionMatrix(table(predict(model, test[,features]), test$Attrition), positive = "Yes")
  # Pull Accuracy, Sensitivity and Specification from confusion matrix
  masterAcc[j] = CM$overall[1]
  masterSens[j] = CM$byClass[1]
  masterSpcf[j] = CM$byClass[2]
}

# Get mean for pulled stats
MeanAcc = colMeans(masterAcc)
MeanSens = colMeans(masterSens)
MeanSpcf = colMeans(masterSpcf)

# Print stats
MeanAcc
```

```
## [1] 0.7806897
```

```
MeanSens
```

```
## [1] 0.6476431
```

```
MeanSpcf
```

```
## [1] 0.8062515
```

Make Prediction csv for Attrition Comp Set

```
library(readxl)
# Load Attrition Comp Set
CaseStudy2CompSet_No_Attrition <- read_csv("C:/Users/OaklandHillsMansion/Desktop/codeCAMP/SMU/CaseStudy
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   BusinessTravel = col_character(),
##   Department = col_character(),
##   EducationField = col_character(),
##   Gender = col_character(),
##   JobRole = col_character(),
##   MaritalStatus = col_character(),
```

```
##   Over18 = col_character(),
##   OverTime = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```r
# Add Legacy column and unite "Department" and "Jobrole" column
df_no_attr <- as.data.frame(CaseStudy2CompSet_No_Attrition) %>%
  mutate(Legacy = YearsInCurrentRole*Education*StockOptionLevel*JobLevel*EnvironmentSatisfaction*Relati
  unite("DeptRole",Department, JobRole)

# Create Model using "features" defined in previous code block
model = naiveBayes(df[,features], df$Attrition)
# Get and save model predictions with ID label, to table_predictions
model_predictions = predict(model, df_no_attr[,features])
```

```
## Warning in predict.naiveBayes(model, df_no_attr[, features]): Type mismatch
## between training and new data for variable 'StockOptionLevel'. Did you use
## factors with numeric labels for training, and numeric values for new data?

## Warning in predict.naiveBayes(model, df_no_attr[, features]): Type mismatch
## between training and new data for variable 'JobInvolvement'. Did you use factors
## with numeric labels for training, and numeric values for new data?
```

```r
model_predictions = as.data.frame(model_predictions)
df_no_attr_id = as.data.frame(df_no_attr$ID)

# Renam columns
colnames(model_predictions) <- "Attrition"
colnames(df_no_attr_id) <- "ID"

# Bind colummns
attr_table_perdictions = cbind(df_no_attr_id, model_predictions)

# Write table_perdictions to CSV
write.csv(attr_table_perdictions, file = "Case2PredictionsClassifyDawkins.csv", row.names=FALSE)
```

Build and Cross Validation Multiple Regression Model

```r
df_salary <- as.data.frame(CaseStudy2_data) %>%
  select(-all_of(removed_columns))

# New Cross Validation
set.seed(4)


## More Stable Measure ... Average of many MSPEs

numMSPEs = 500
RMSE = numeric(numMSPEs)
AdjRSquared = numeric(numMSPEs)
```

```r
RSquared = numeric(numMSPEs)

for (i in 1:numMSPEs)
{
  TrainObs = sample(seq(1,dim(df_salary)[1]),round(.75*dim(df_salary)[1]),replace = FALSE)
  salary_train = df_salary[TrainObs,]
  salary_test = df_salary[-TrainObs,]

  salary_model = lm(MonthlyIncome ~ TotalWorkingYears+BusinessTravel+
             JobLevel*JobRole, data = salary_train)

  salary_model_predictions = predict(salary_model, newdata = salary_test)


  RMSE[i] = sqrt(mean(salary_model$residuals^2))
  RSquared[i] = summary(salary_model)$r.squared
  AdjRSquared[i] = summary(salary_model)$adj.r.squared

}
```

```
## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading

## Warning in predict.lm(salary_model, newdata = salary_test): prediction from a
## rank-deficient fit may be misleading
```

```r
# R Squared
mean(RMSE)
```

```
## [1] 973.7647
```

```r
# Adj R Squared
mean(RSquared)
```

```
## [1] 0.9549344
```

```r
# Root Mean Square Error
mean(AdjRSquared)
```

```
## [1] 0.9535076
```

Make Prediction csv for Attrition Comp Set

```r
CaseStudy2CompSet_No_Salary <- read_excel("C:/Users/OaklandHillsMansion/Desktop/codeCAMP/SMU/CaseStudy2/

df_no_salary <- as.data.frame(CaseStudy2CompSet_No_Salary)

# Create MLR model, using data with a salary
salary_model = lm(MonthlyIncome ~ TotalWorkingYears+BusinessTravel+
           JobLevel*JobRole, data = df_salary)

# Get and save model predictions with ID label, to table_predictions
salary_model_predictions = predict(salary_model, newdata = df_no_salary)
salary_model_predictions = as.data.frame(salary_model_predictions)

# Save as Dataframe
df_no_salary_id = as.data.frame(df_no_salary$ID)

# Rename Columns
colnames(salary_model_predictions) <- "MonthlyIncome"
colnames(df_no_salary_id) <- "ID"

# Bind columns
salary_table_perdictions = cbind(df_no_salary_id, salary_model_predictions)

# Write table_perdictions to CSV
write.csv(salary_table_perdictions, file = "Case2PredictionsRegressDawkins.csv", row.names=FALSE)
```

Executive Summary:

The attrition model had an average accuracy of ~78%, an average specification of ~81% and an average sensitivity of ~65%. A Legacy attribute was derived from existing features in the data set and was applied to each employee, to aid in the predicting attrition.

The salary model had root mean square error of $962 and Adj. R-squared of 94.8%.

Other interesting trends: - There was no attrition amongst the human resource managers and 100% - Non managers, at 22%, have the second highest turnover rate

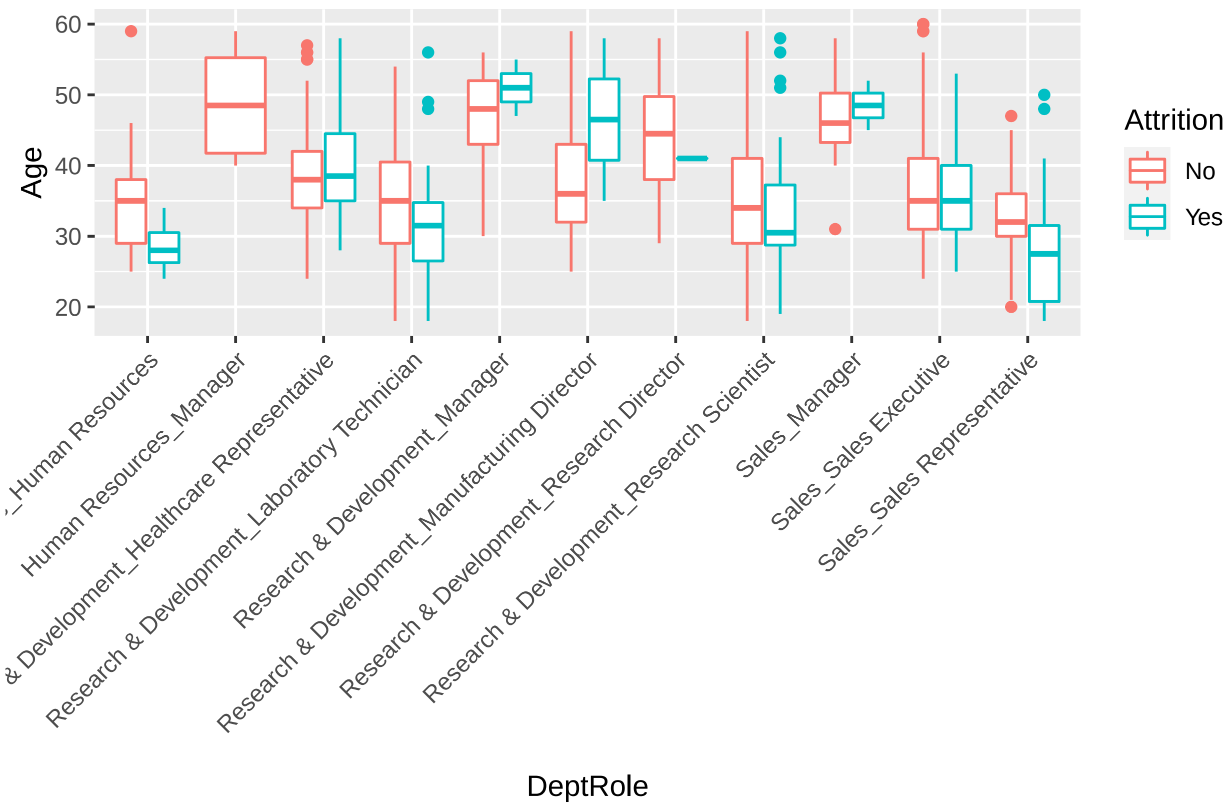
- All sales managers who left, left after 8-12 years with their current manager
- Sales Reps, at 45%, has the highest turnover rate
- R&D Research Scientist who left,left the company within first 5 years in their current role
- R&D Research Scientist ~20% and R&D Lab Techs ~19% are have the third and fourth highest turnover rates

Appendix:

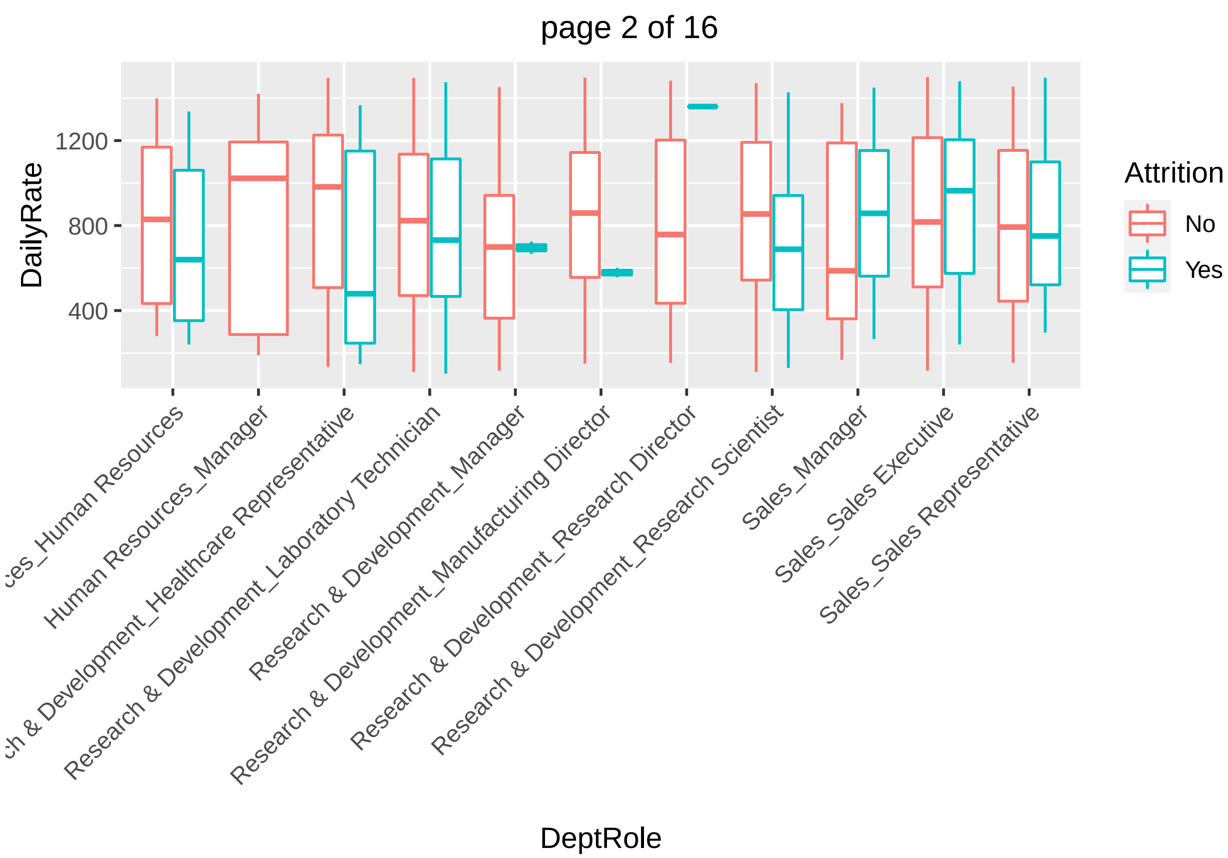
Get Box Plots for Numerical Columns

```r
# Initiate plot list
plot_list <- list()
# Loop through numerical features and plot
for (i in 1:ncol(df_nums)) {
  plot_list[[i]] <- df %>%
    ggplot(aes_string(y=colnames(df_nums)[i],x="DeptRole", color="Attrition"))+
    geom_boxplot()+
    theme(axis.text.x=element_text(angle=45,hjust=1))

}
# Plot all plots in plot_list
marrangeGrob(grobs = plot_list,
                   nrow=1, ncol=1)
```

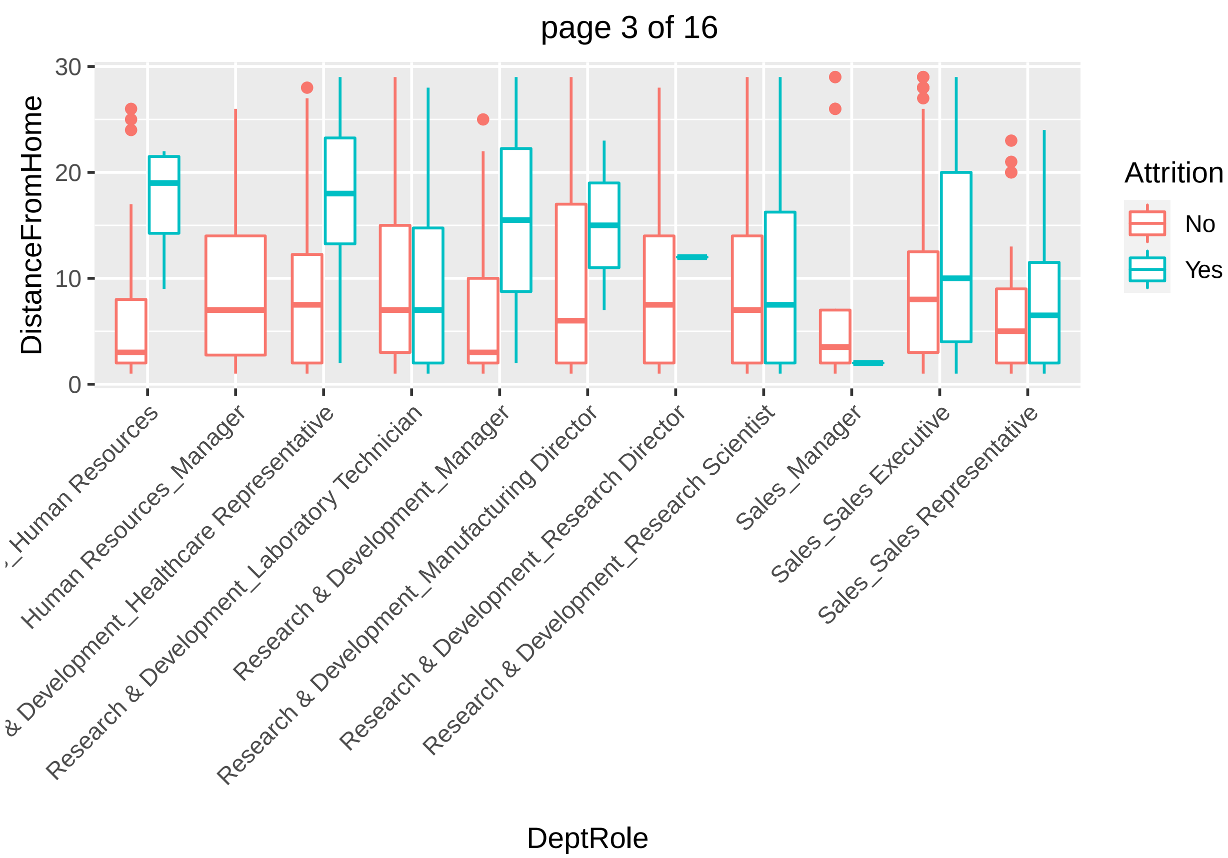
page 1 of 16



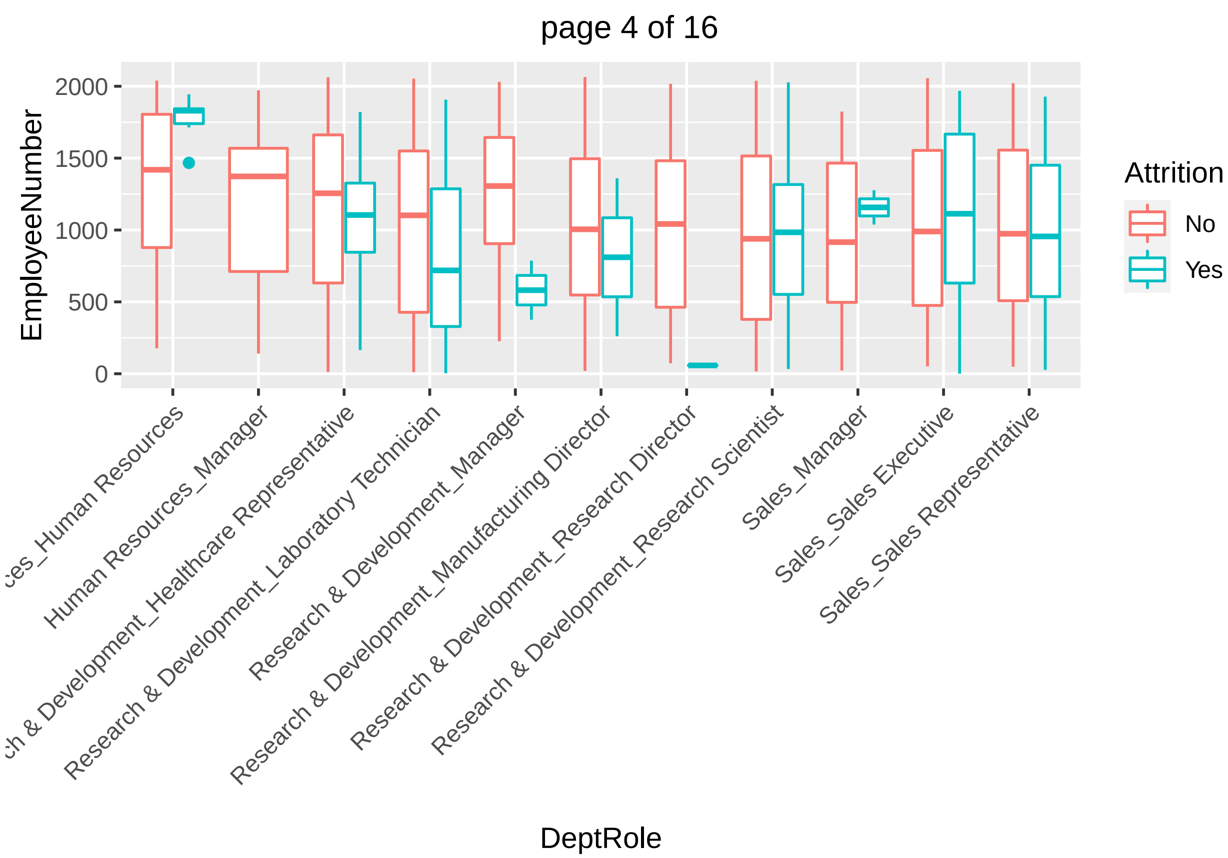
60
50
40
30
20
Age
Attrition
No
Yes
_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole



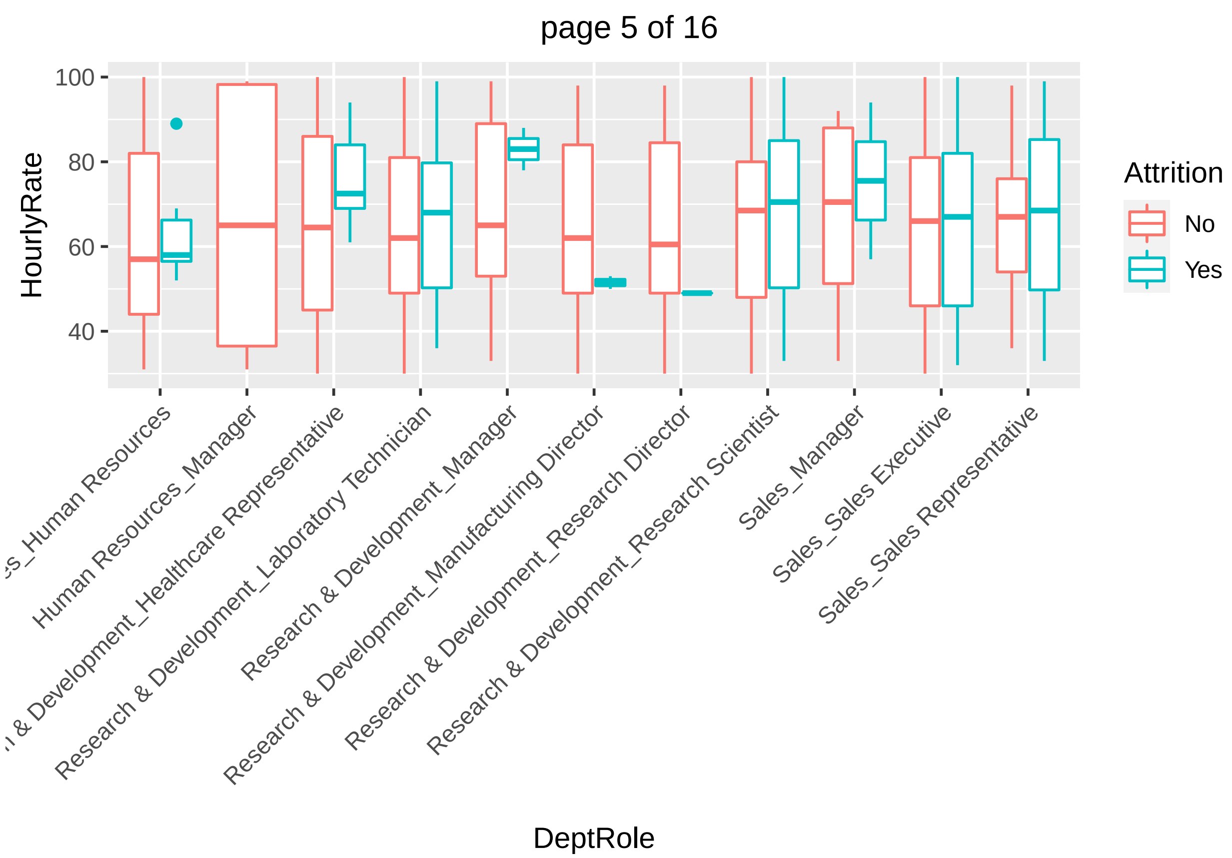
page 2 of 16
DailyRate
1200
800
400
Attrition
No
Yes
ces_Human Resources
Human Resources_Manager
ch & Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole

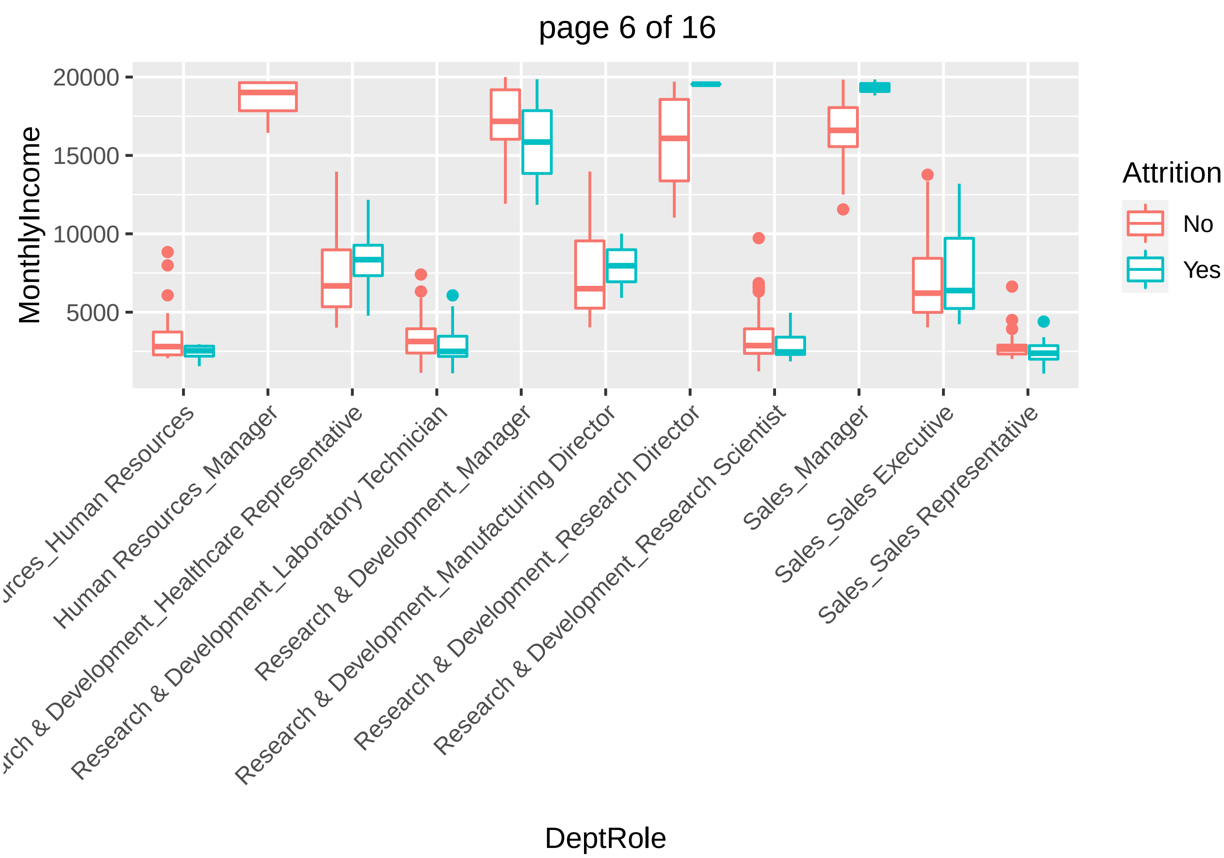


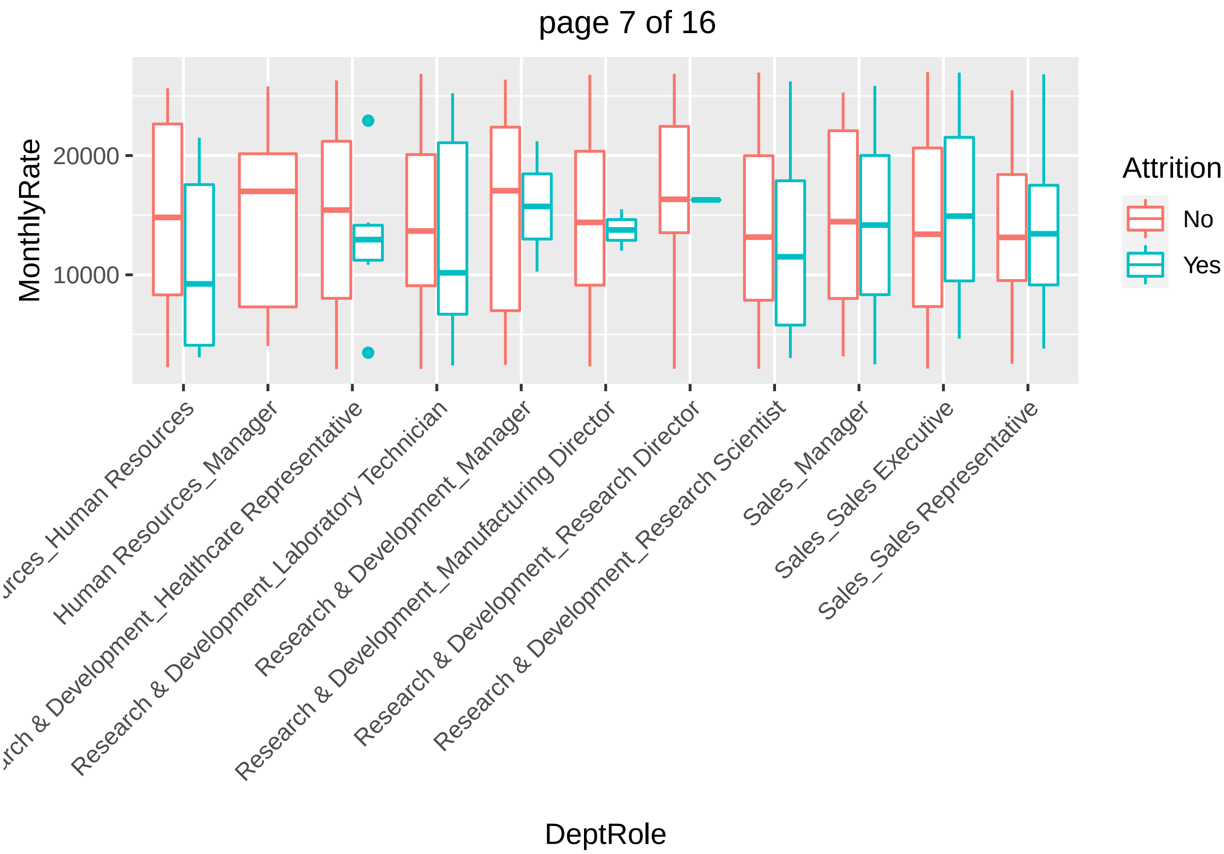
page 3 of 16
DistanceFromHome
0
10
20
30
DeptRole
_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
Attrition
No
Yes

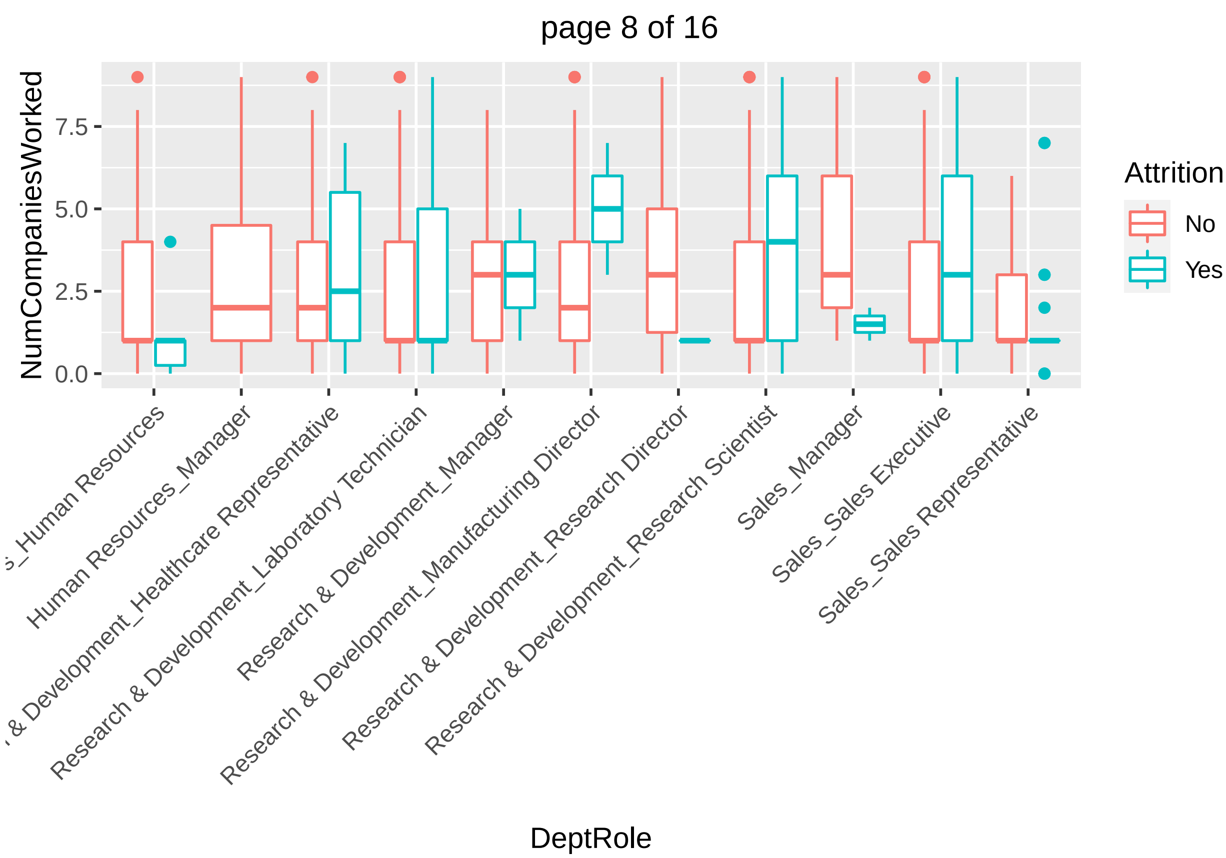


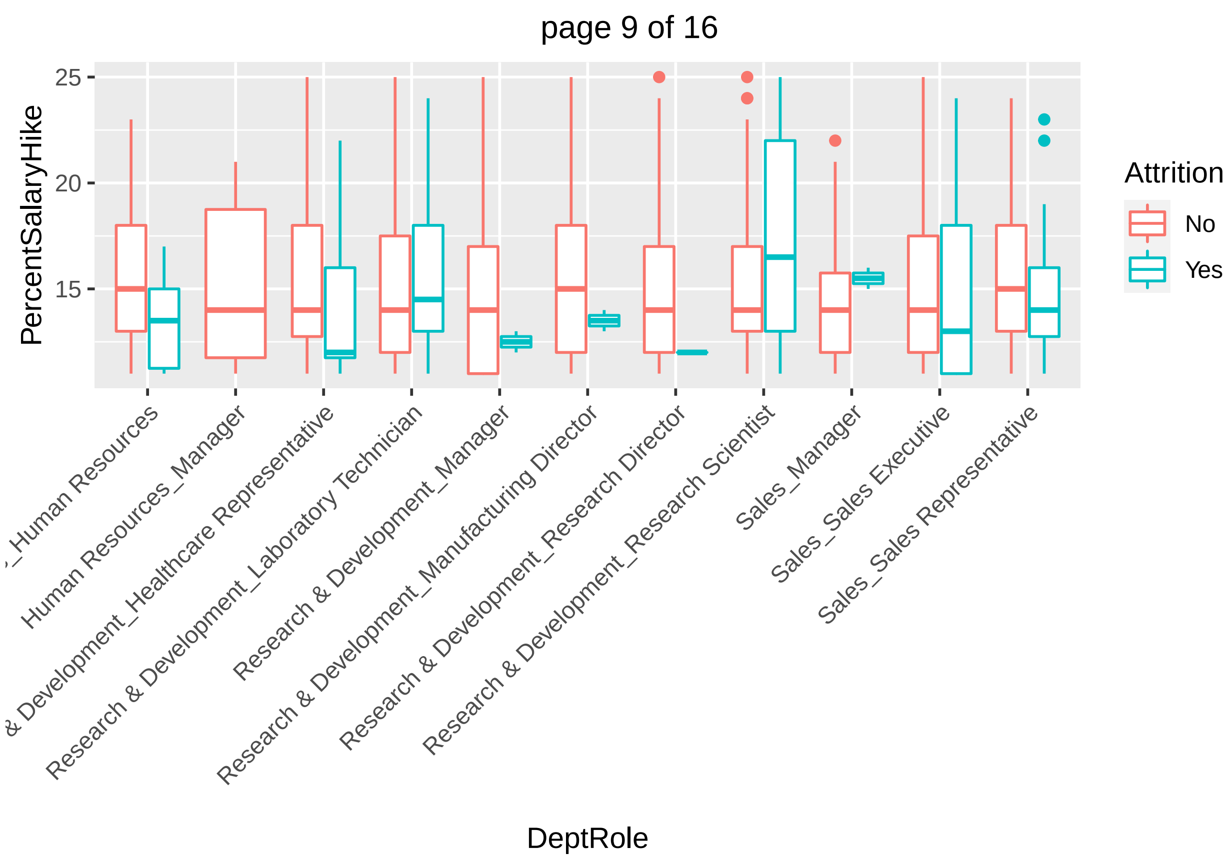
page 4 of 16
EmployeeNumber
2000
1500
1000
500
0
Attrition
No
Yes
DeptRole
Human Resources_Manager
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative


page 5 of 16
HourlyRate
100
80
60
40
Attrition
No
Yes
DeptRole
Human Resources_Manager
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist

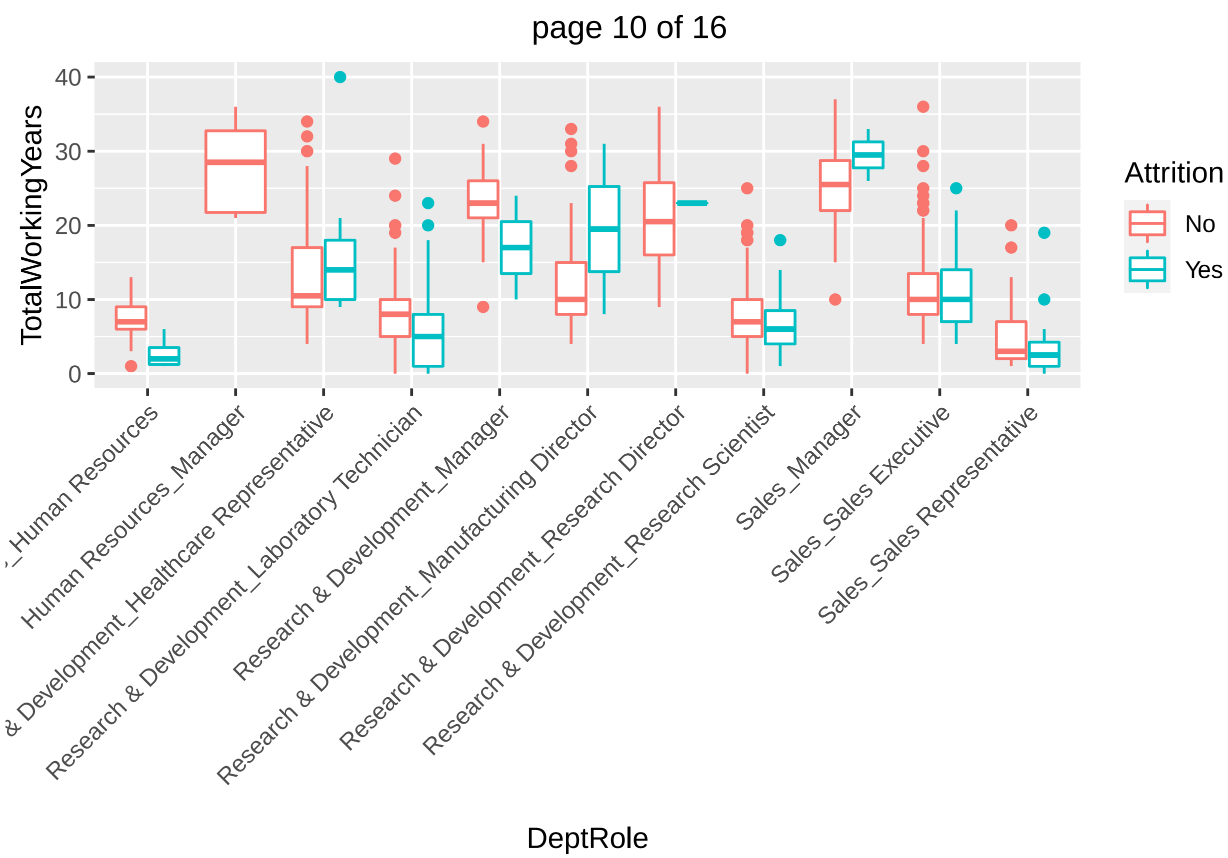

page 6 of 16
MonthlyIncome
20000
15000
10000
5000
DeptRole
Attrition
No
Yes
Human Resources_Manager
Research & Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative


page 7 of 16
MonthlyRate
20000
10000
Attrition
No
Yes
Human Resources_Manager
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole

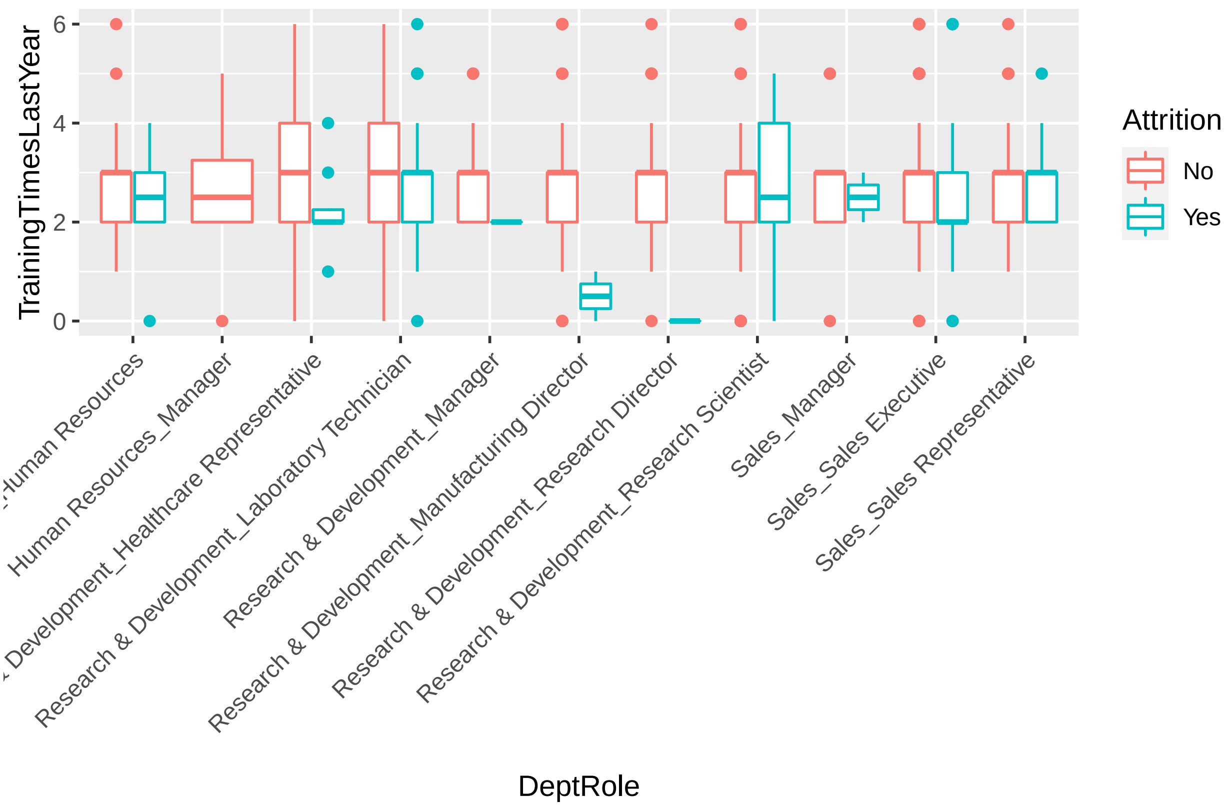

page 8 of 16
NumCompaniesWorked
7.5
5.0
2.5
0.0
Attrition
No
Yes
s_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole



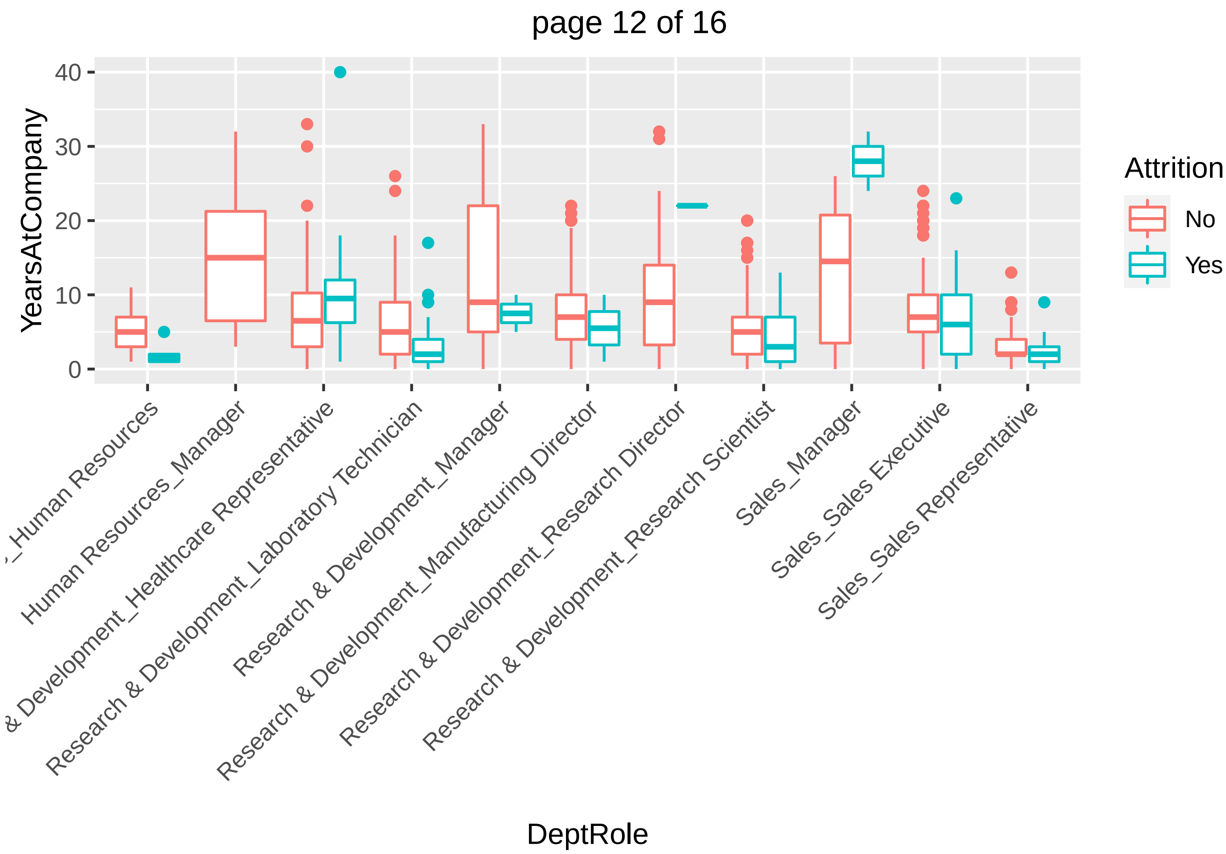
page 9 of 16
PercentSalaryHike
25
20
15
DeptRole
_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
Attrition
No
Yes

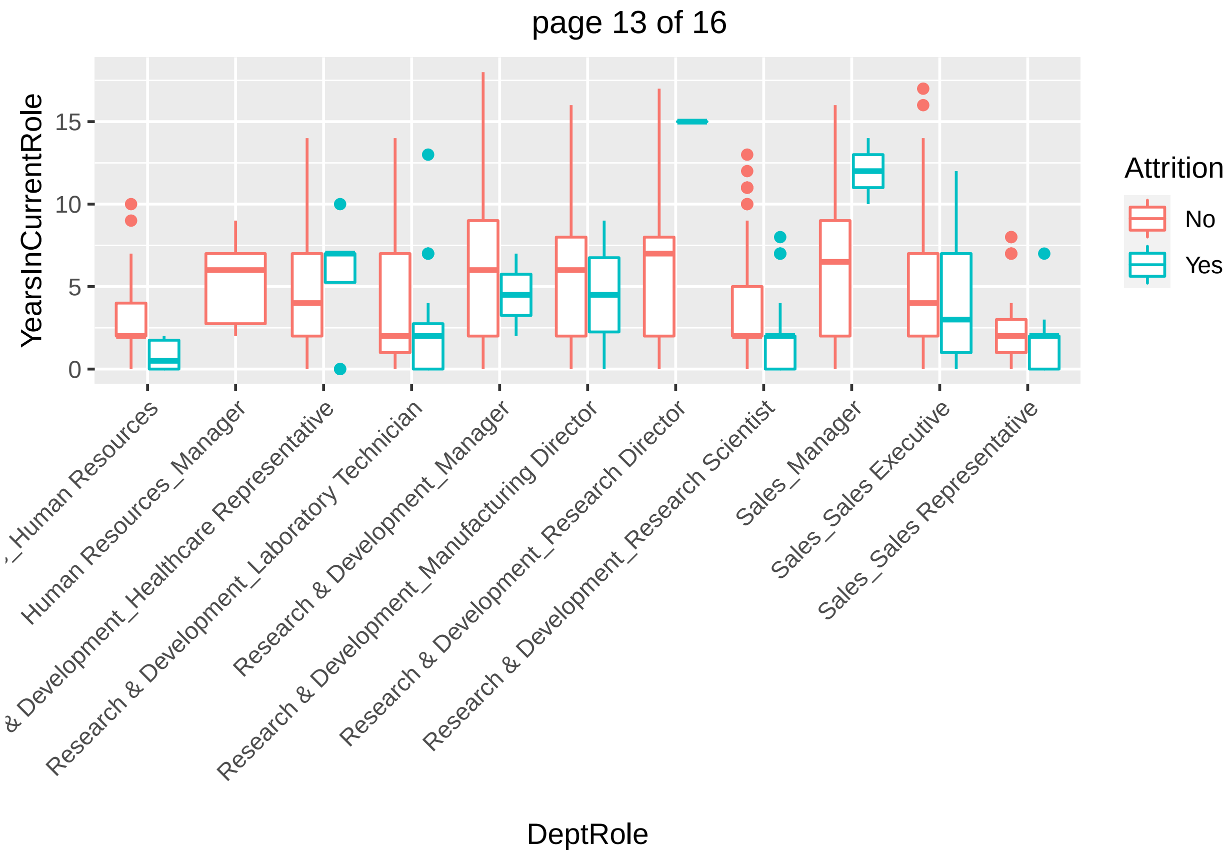

page 10 of 16
TotalWorkingYears
40
30
20
10
0
DeptRole
Attrition
No
Yes
_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative

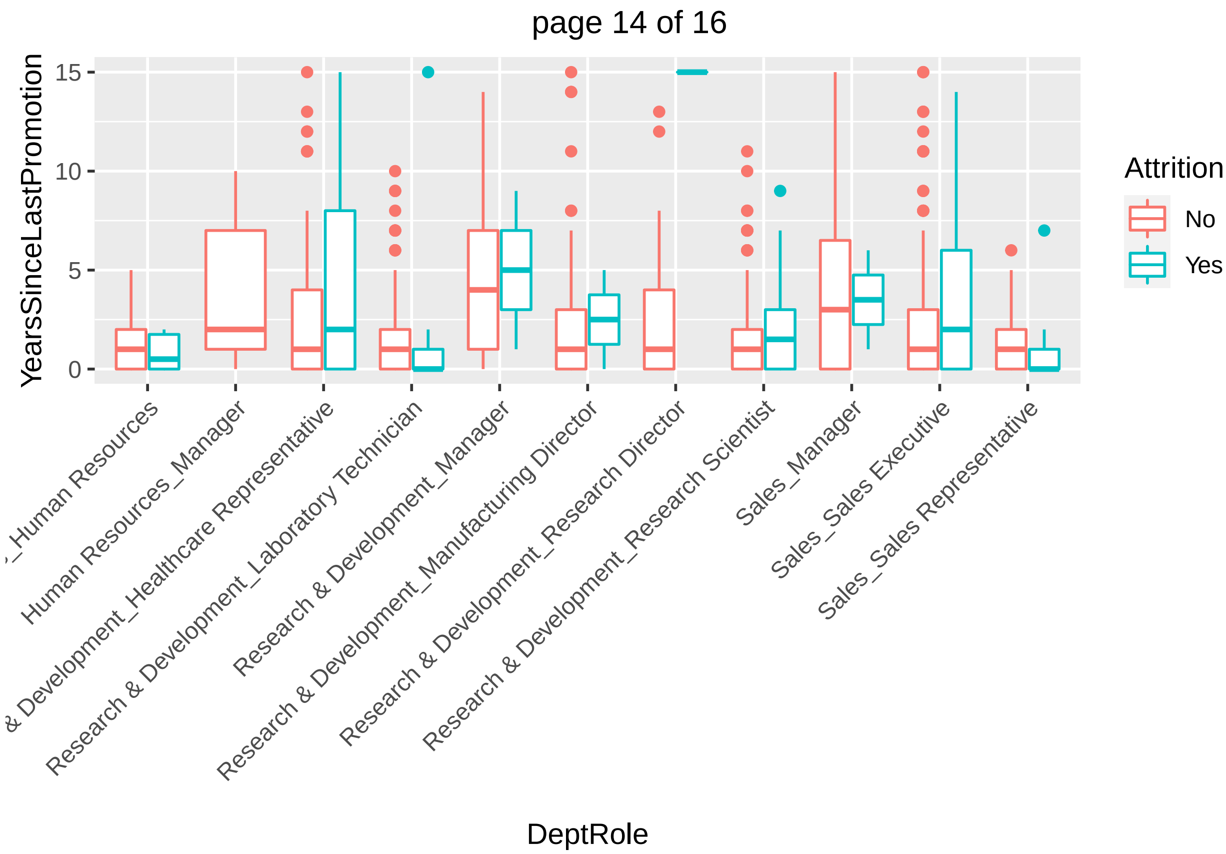
page 11 of 16

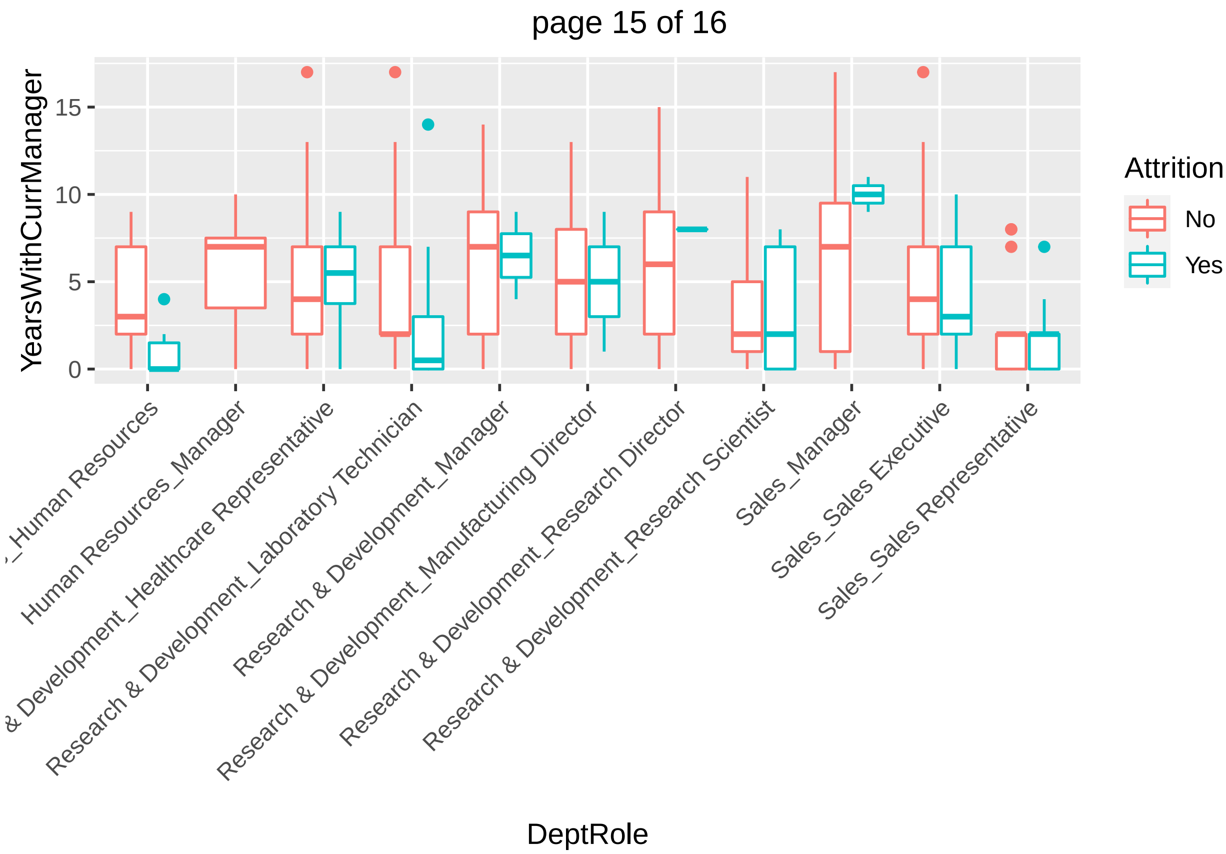


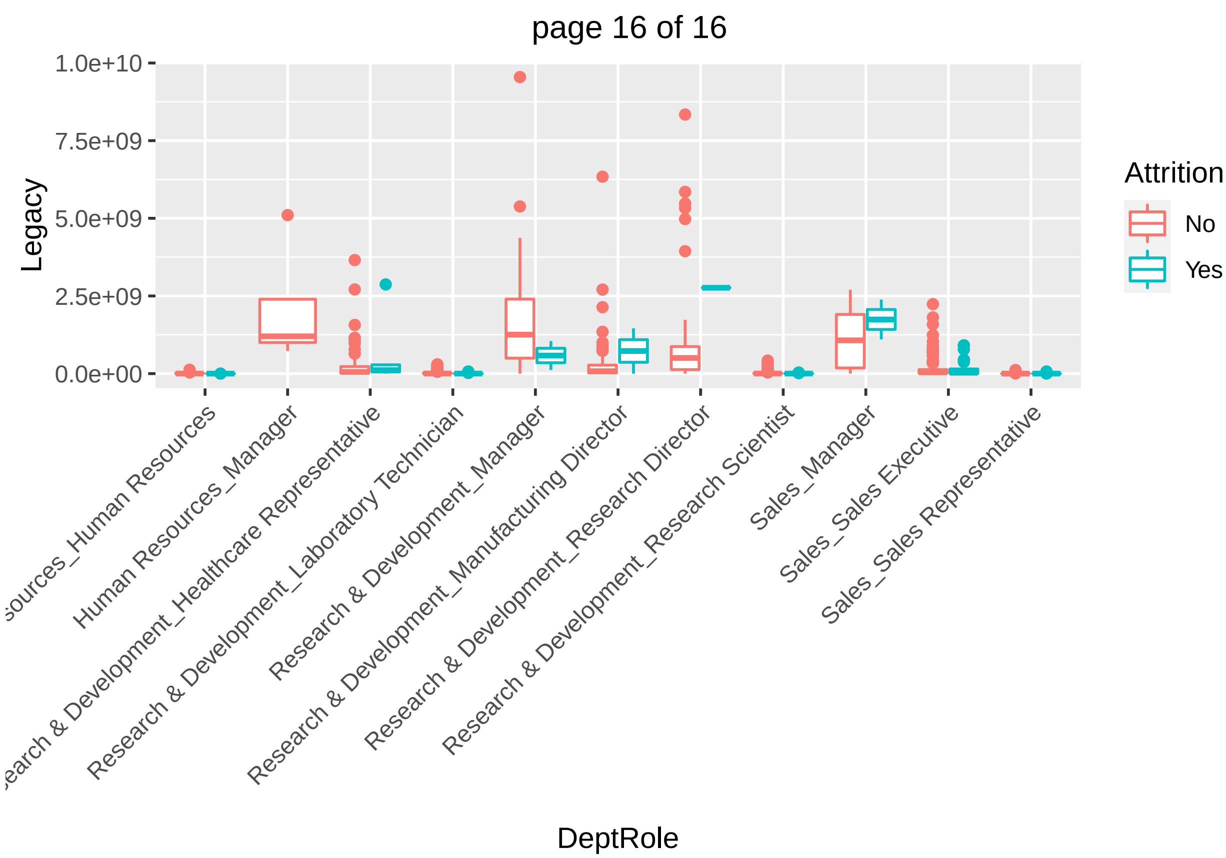
TrainingTimesLastYear
6
4
2
0
Attrition
No
Yes
Human Resources
Human Resources_Manager
Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole


page 12 of 16
YearsAtCompany
0
10
20
30
40
DeptRole
Attrition
No
Yes
Human Resources_Manager
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative


page 13 of 16
YearsInCurrentRole
15
10
5
0
_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole
Attrition
No
Yes


page 14 of 16
YearsSinceLastPromotion
0
5
10
15
Attrition
No
Yes
_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole


page 15 of 16
YearsWithCurrManager
15
10
5
0
Attrition
No
Yes
_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole



Get Scatter Plots for Numerical Columns

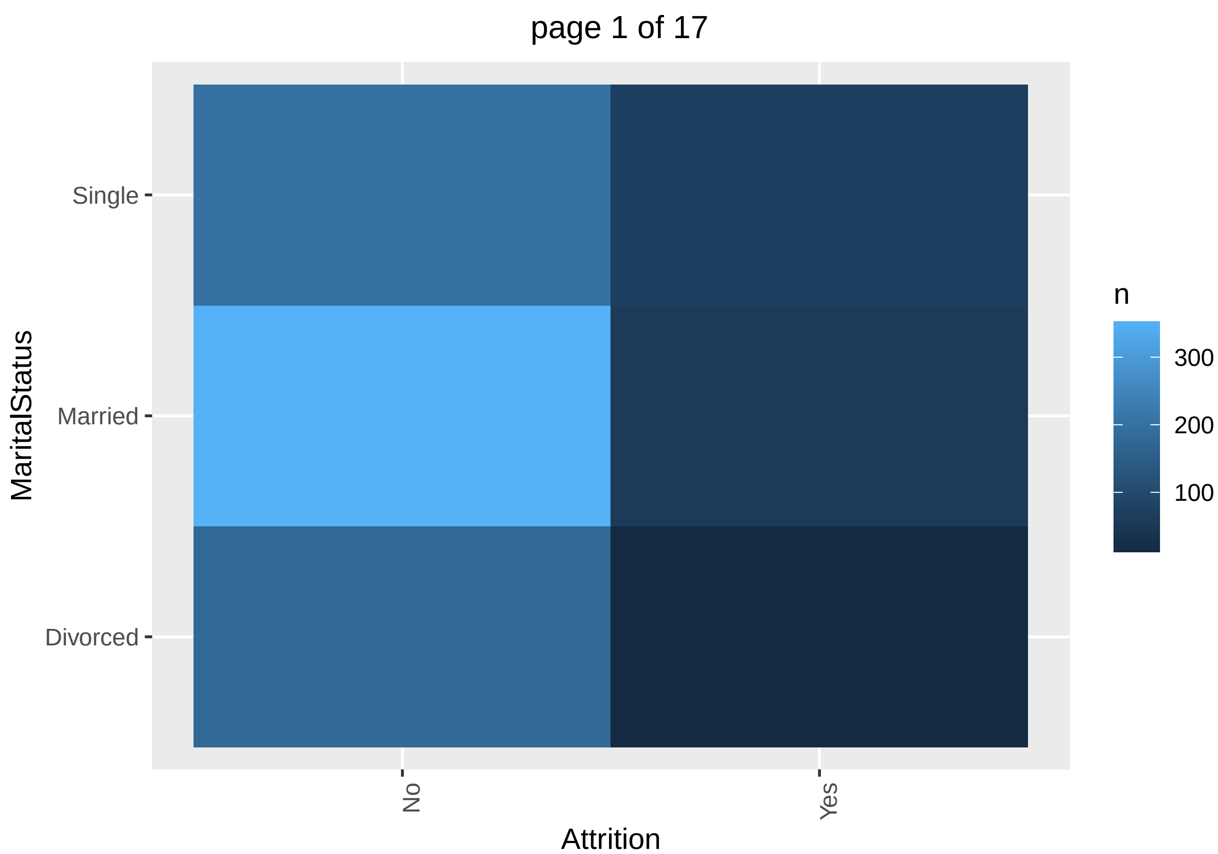
```
# Initiate plot list
plot_list <- list()
# Loop through numerical features and plot
for (i in 1:ncol(df_nums)) {
  plot_list[[i]] <- df %>%
    ggplot(aes_string(x=colnames(df_nums)[i], y="MonthlyIncome", color="Attrition"))+
    geom_point()

}
# Plot all plots in plot_list
marrangeGrob(grobs = plot_list,
                   nrow=3, ncol=2)
```


page 1 of 3
MonthlyIncome
Age
EmployeeNumber
DailyRate
HourlyRate
DistanceFromHome
MonthlyIncome
Attrition
No
Yes

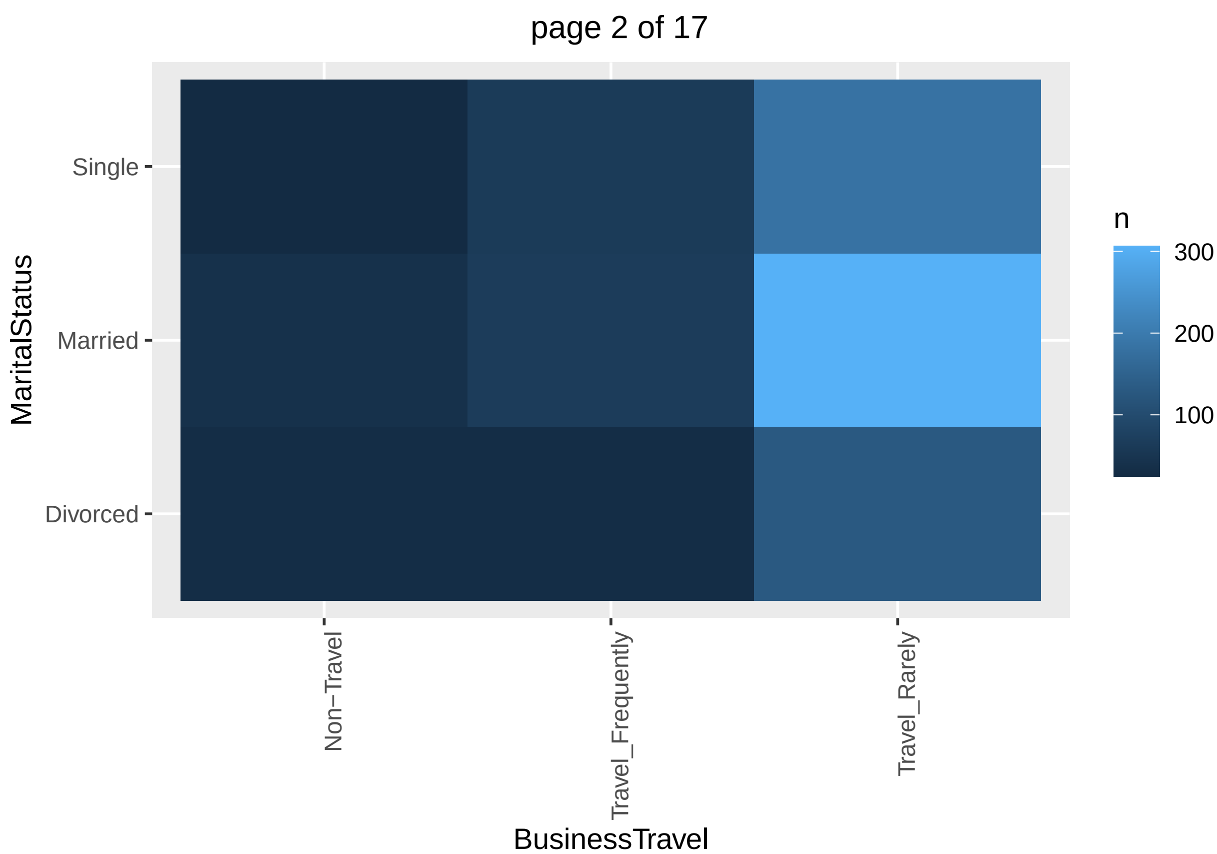

page 2 of 3
MonthlyIncome
20000
15000
10000
5000
MonthlyRate
10000
20000
TotalWorkingYears
0
10
20
30
40
NumCompaniesWorked
0.0
2.5
5.0
7.5
TrainingTimesLastYear
0
2
4
6
PercentSalaryHike
15
20
25
YearsAtCompany
0
10
20
30
40
Attrition
No
Yes

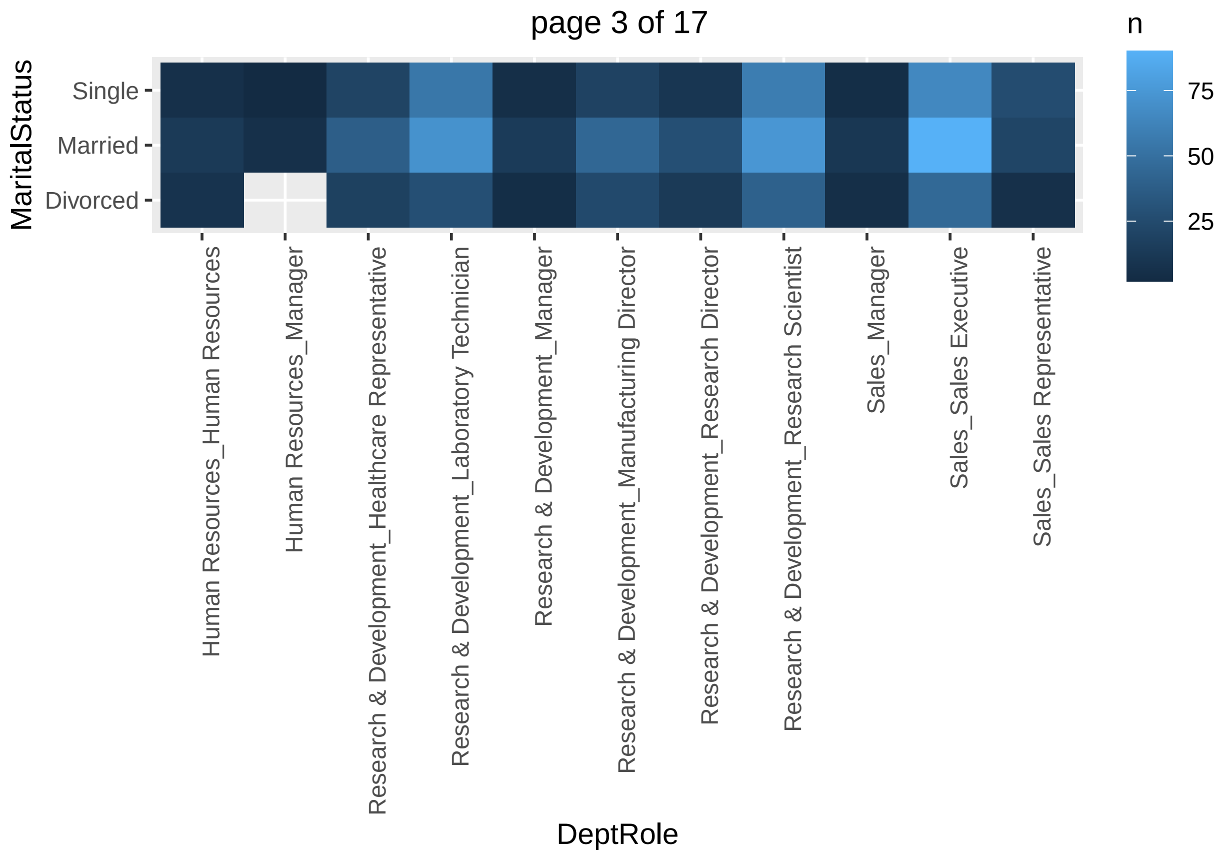
Get Tile Plots for Char_Fac Columns

```
# Initiate plot list
plot_list <- list()
# Constant Variable
const_var <- "MaritalStatus"
# Loop through character factors and plot
for (i in 1:ncol(df_char_fac)) {
  plot_list[[i]] <- df %>%
    dplyr::count(!!as.name(colnames(df_char_fac)[i]), !!as.name(const_var)) %>%
    ggplot(aes_string(x=colnames(df_char_fac)[i], y=const_var)) +
    geom_tile(mapping=aes(fill=n))+
    theme(axis.text.x=element_text(angle=90,hjust=1))

}

# Plot all plots in plot_list
marrangeGrob(grobs = plot_list,
                   nrow=1, ncol=1)
```

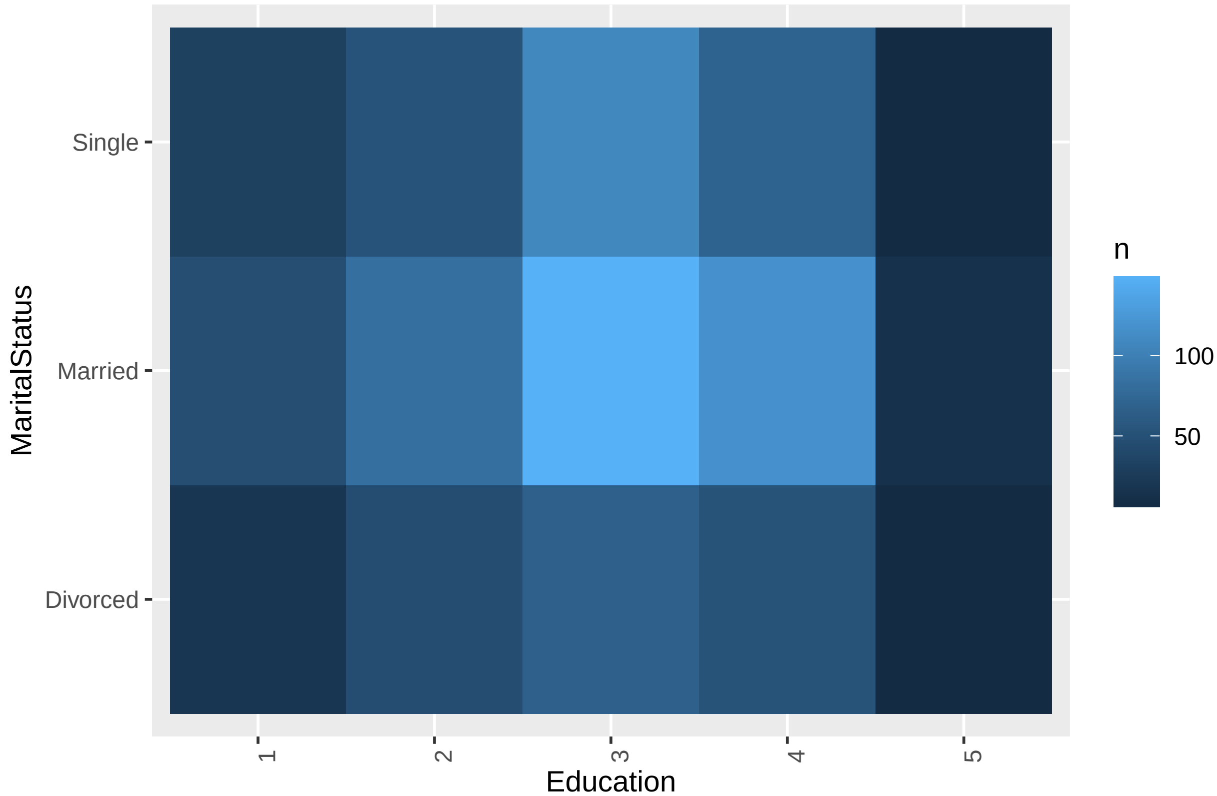


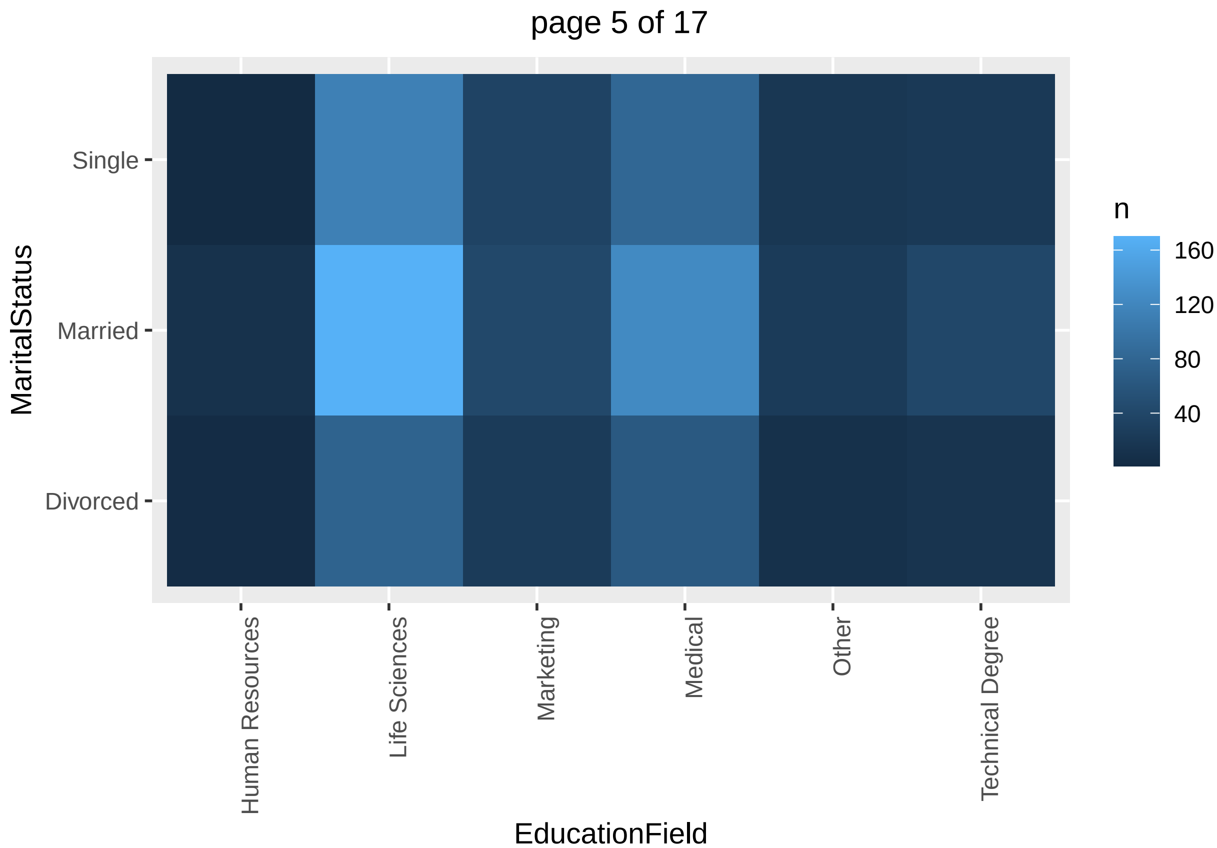
page 1 of 17
MaritalStatus
Single
Married
Divorced
Attrition
No
Yes
n
300
200
100


page 2 of 17
MaritalStatus
Single
Married
Divorced
Non−Travel
Travel_Frequently
Travel_Rarely
BusinessTravel
n
300
200
100

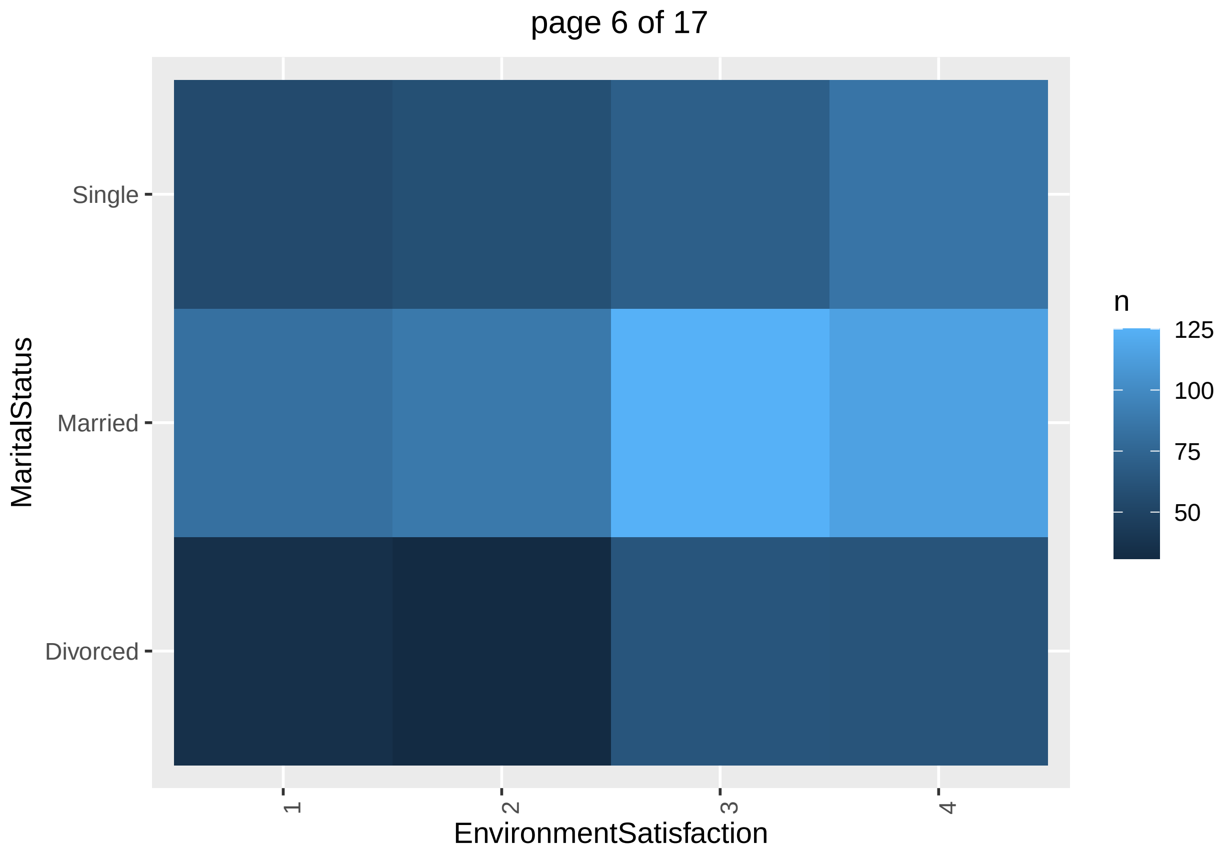

page 3 of 17
n
75
50
25
MaritalStatus
Single
Married
Divorced
Human Resources_Human Resources
Human Resources_Manager
Research & Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole

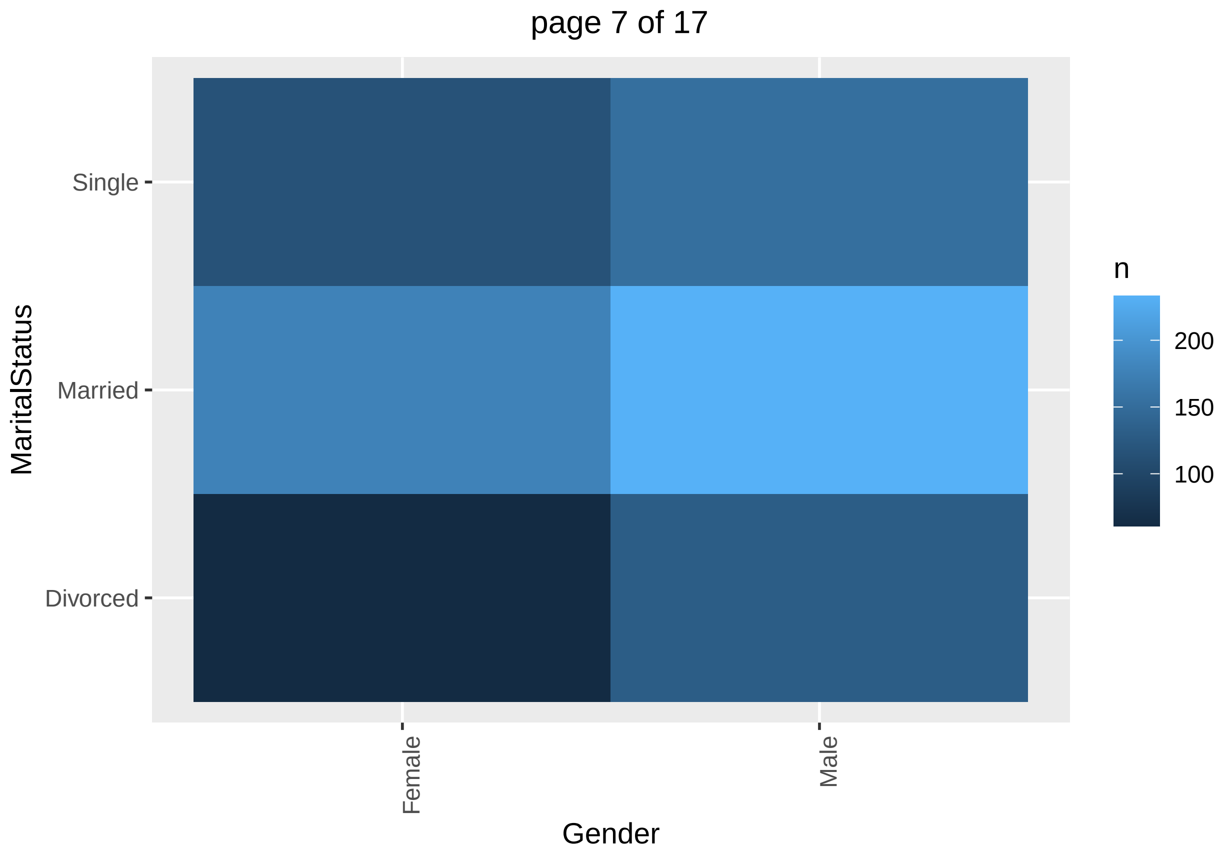
page 4 of 17

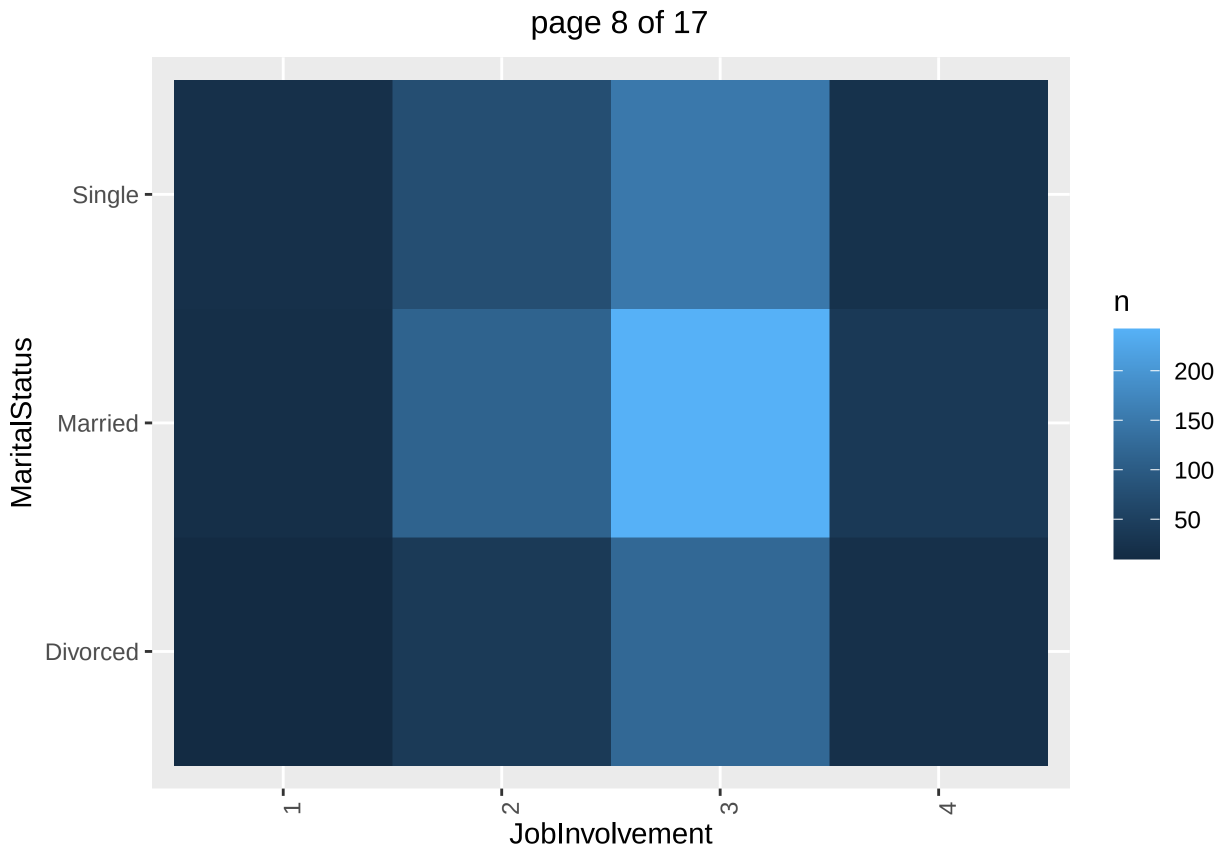

Single
Married
Divorced
MaritalStatus
1
2
3
4
5
Education
n
100
50

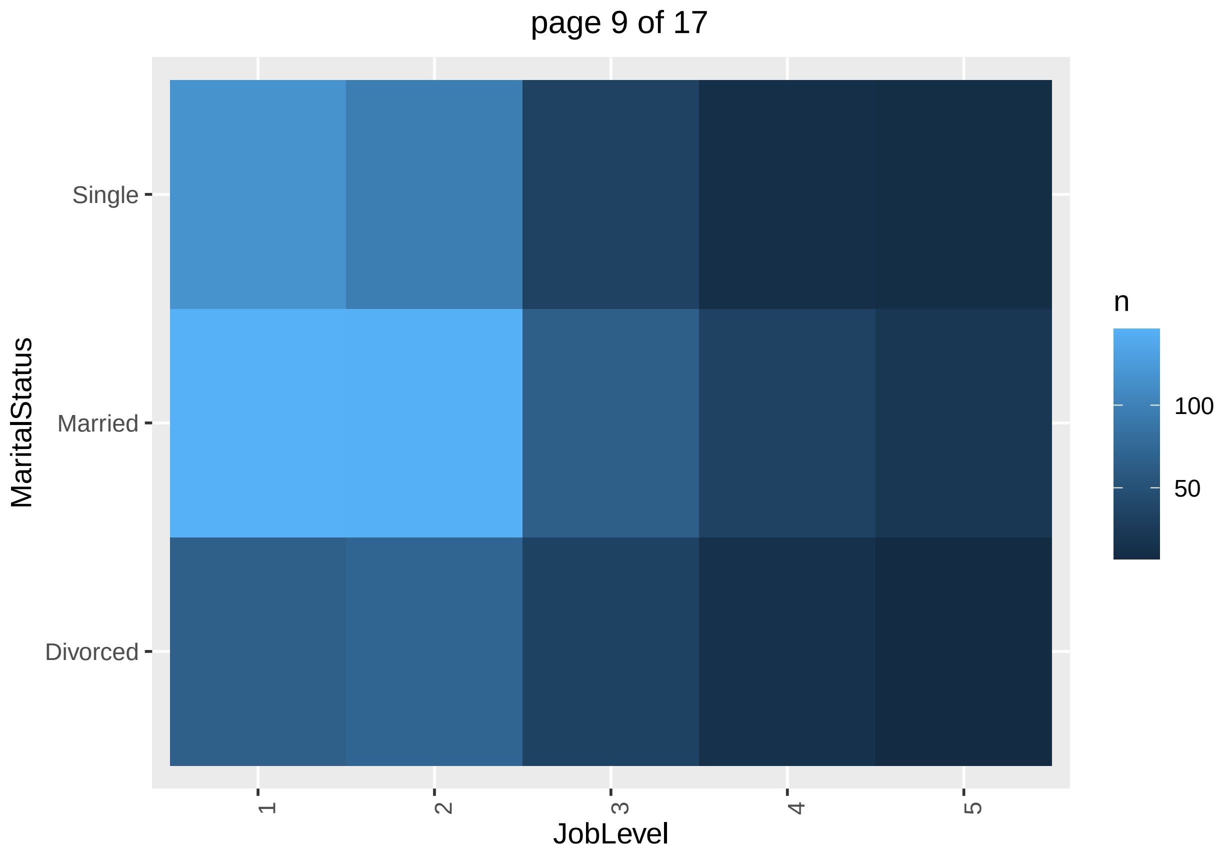


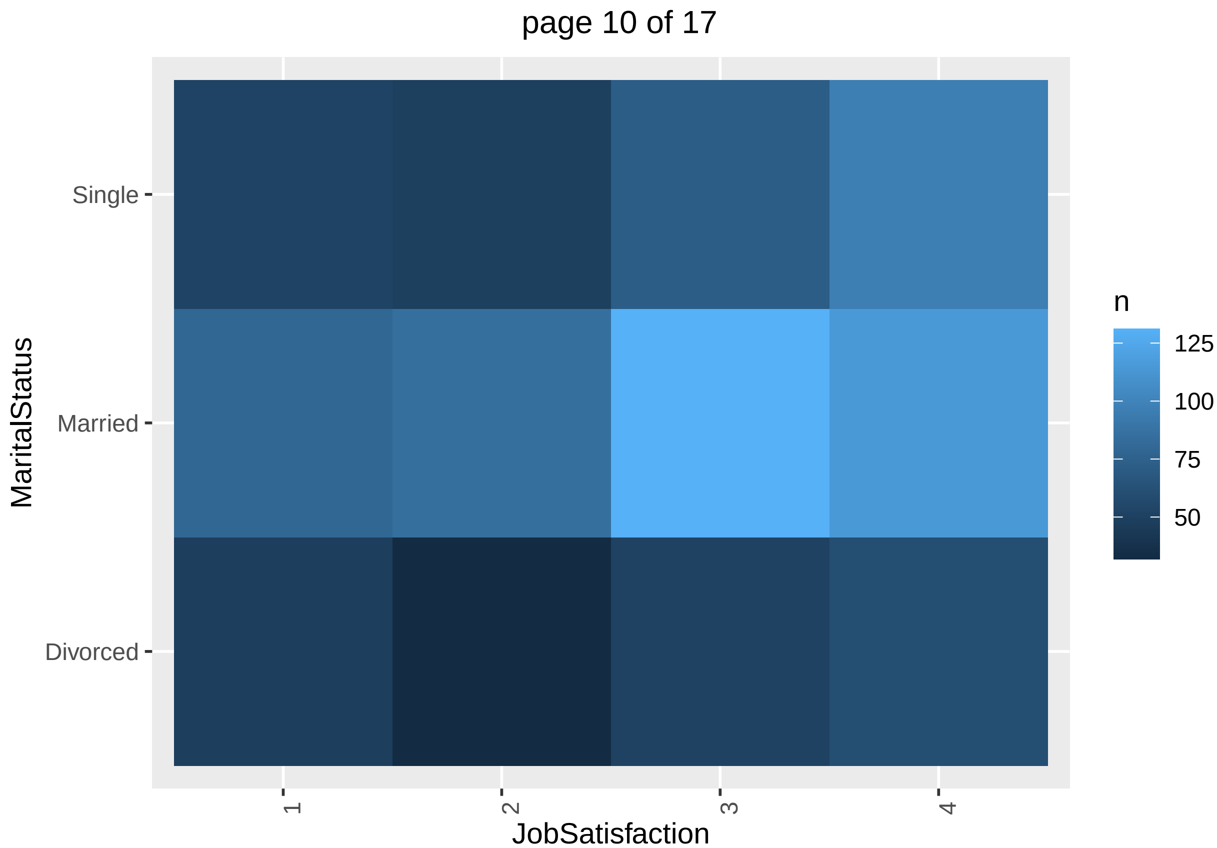
page 5 of 17
Single
Married
Divorced
MaritalStatus
Human Resources
Life Sciences
Marketing
Medical
Other
Technical Degree
EducationField
n
160
120
80
40

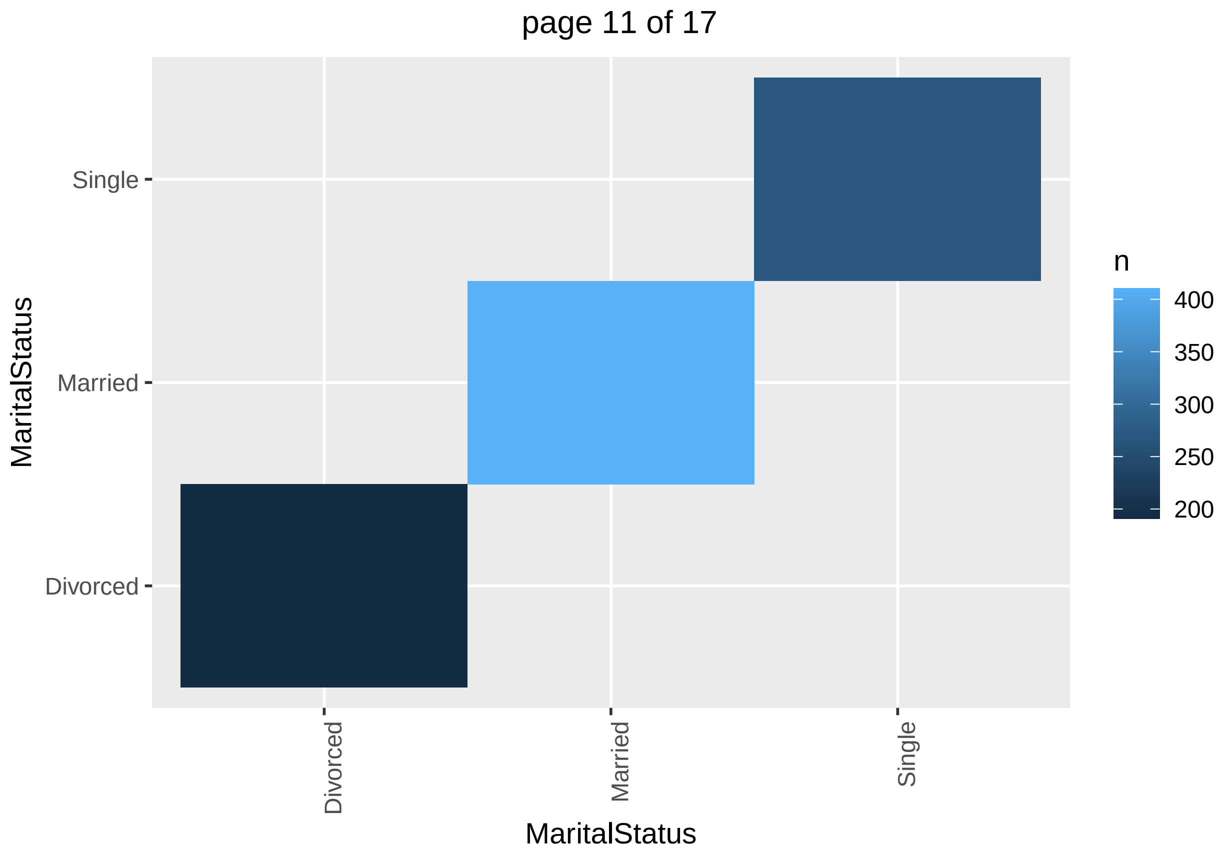

page 6 of 17
Single
Married
Divorced
MaritalStatus
1
2
3
4
EnvironmentSatisfaction
n
125
100
75
50

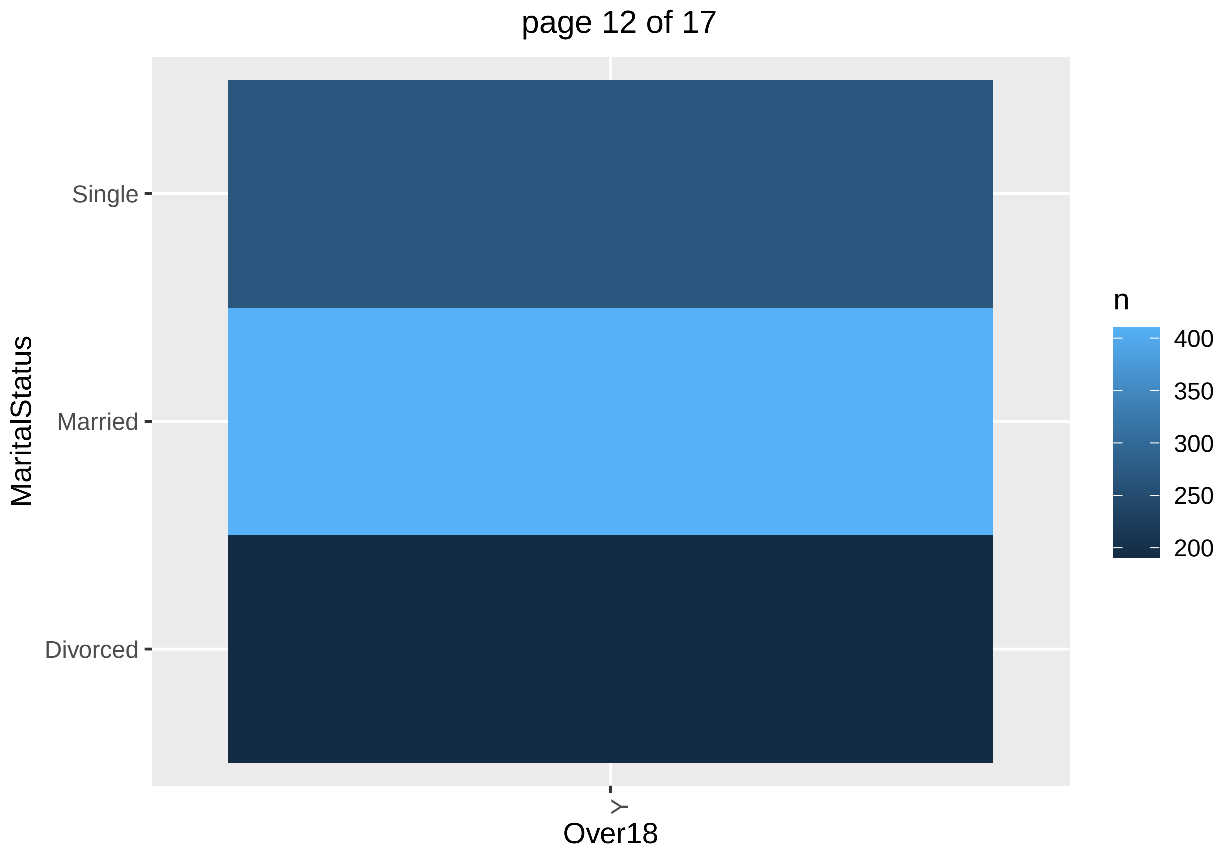

page 7 of 17
Single
Married
Divorced
MaritalStatus
Female
Male
Gender
n
200
150
100


page 8 of 17
Single
Married
Divorced
MaritalStatus
1
2
3
4
JobInvolvement
n
200
150
100
50

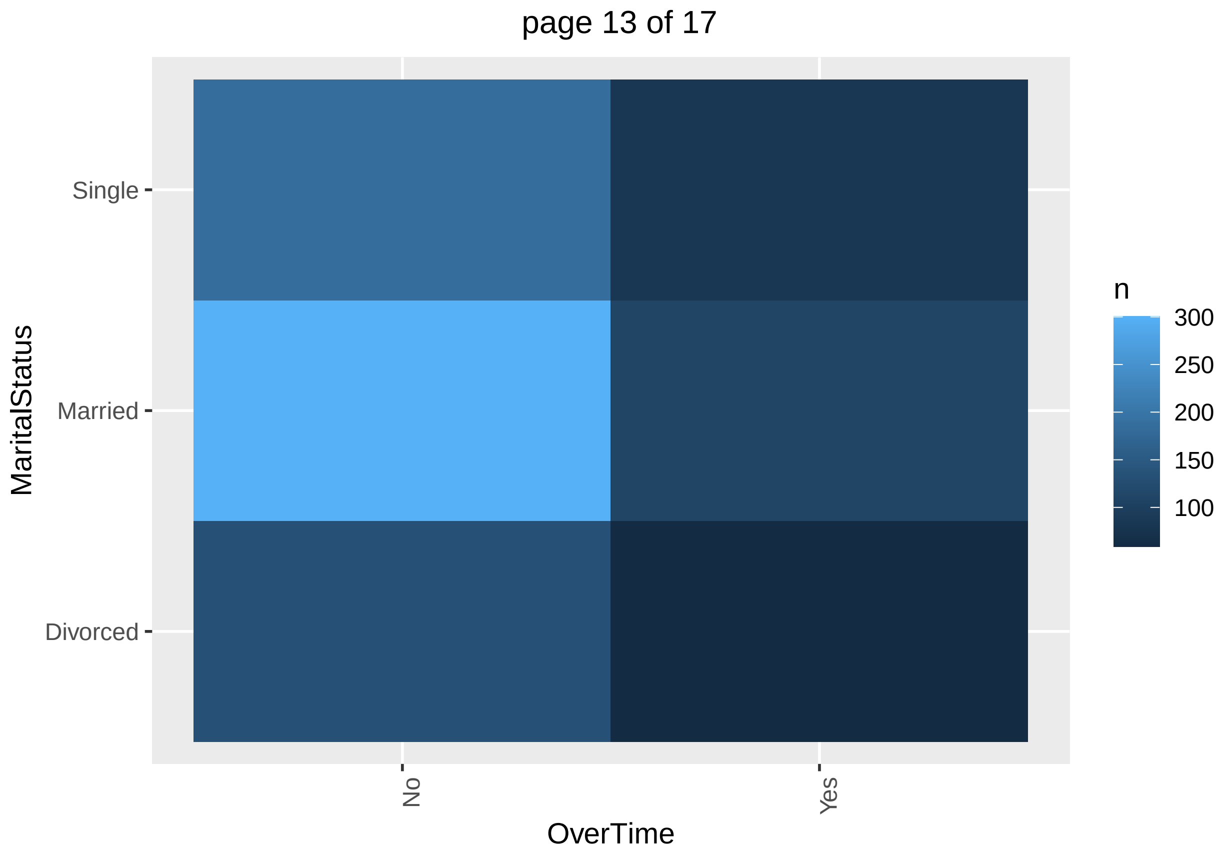

page 9 of 17
Single
Married
Divorced
MaritalStatus
1
2
3
4
5
JobLevel
n
100
50

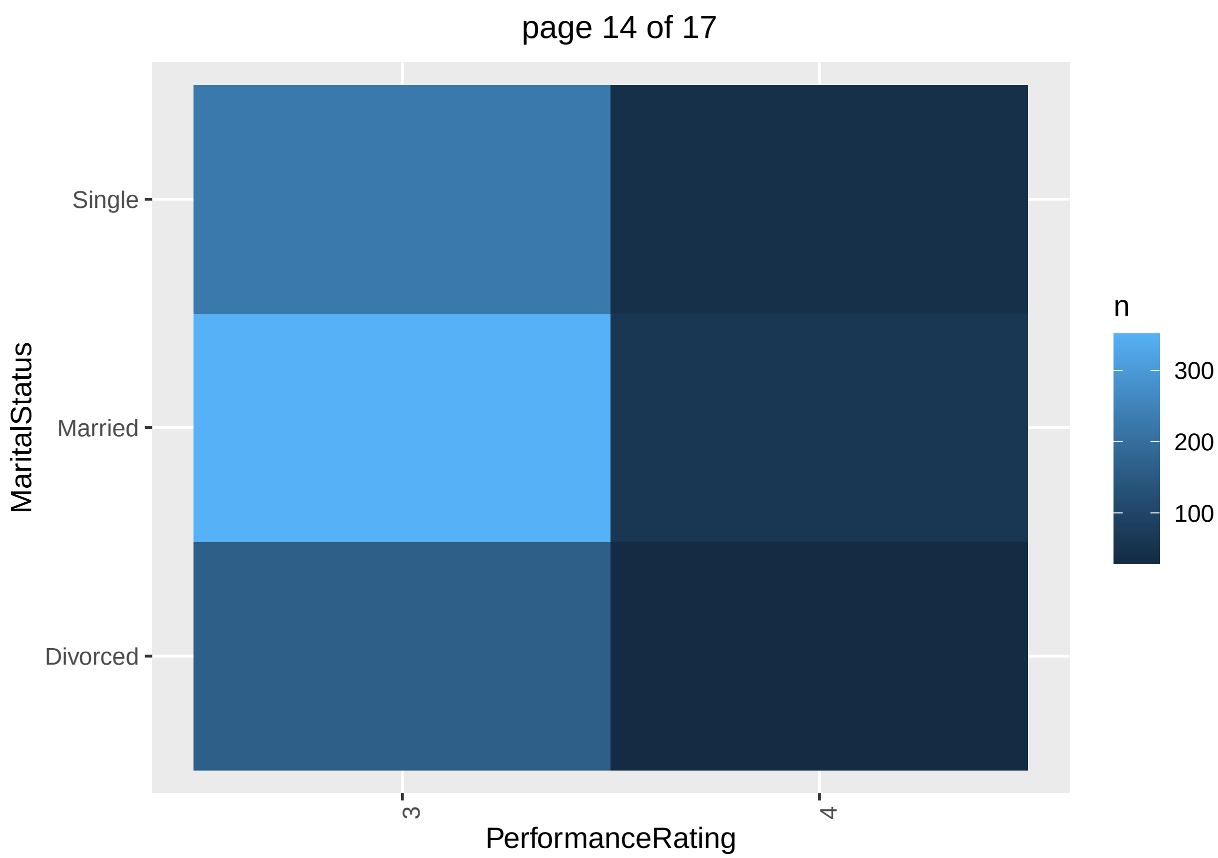

page 10 of 17
Single
Married
Divorced
MaritalStatus
1
2
3
4
JobSatisfaction
n
125
100
75
50

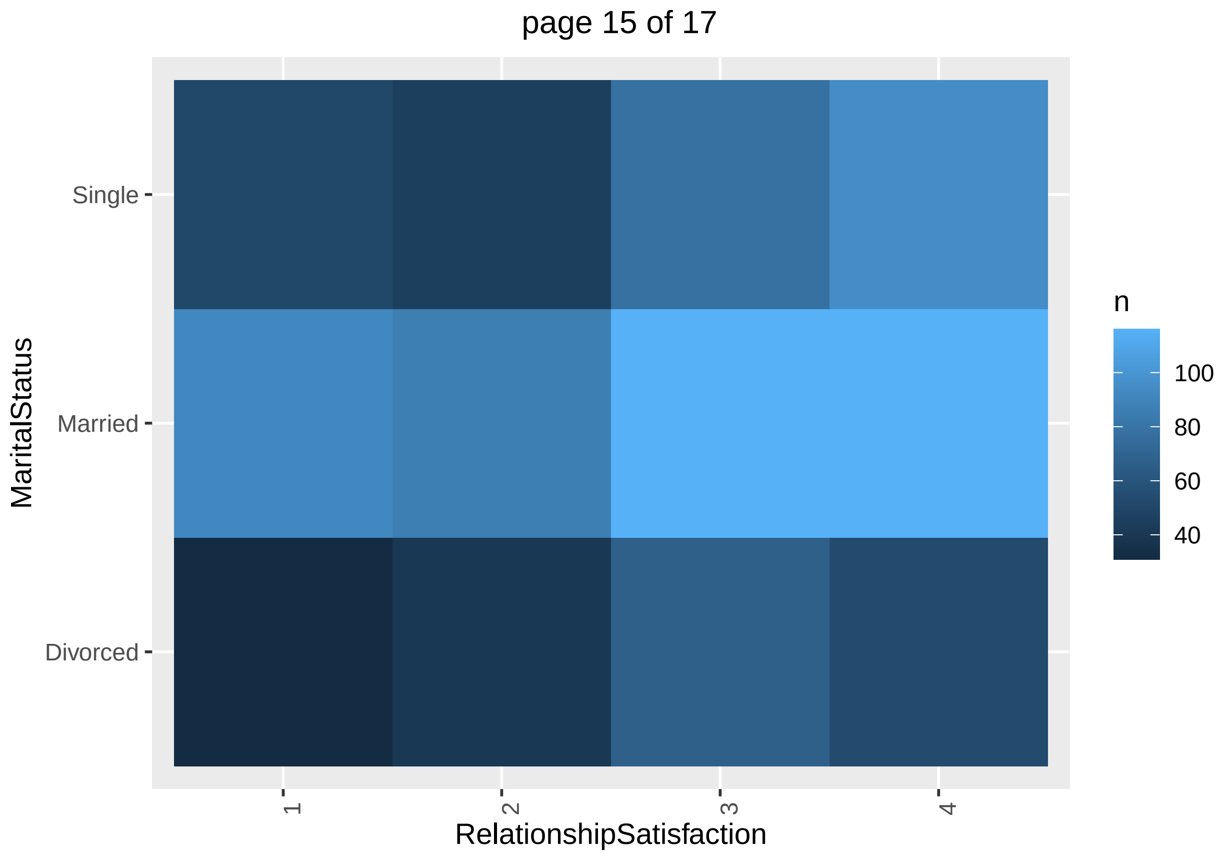

page 11 of 17
MaritalStatus
Single
Married
Divorced
Divorced
Married
Single
MaritalStatus
n
400
350
300
250
200

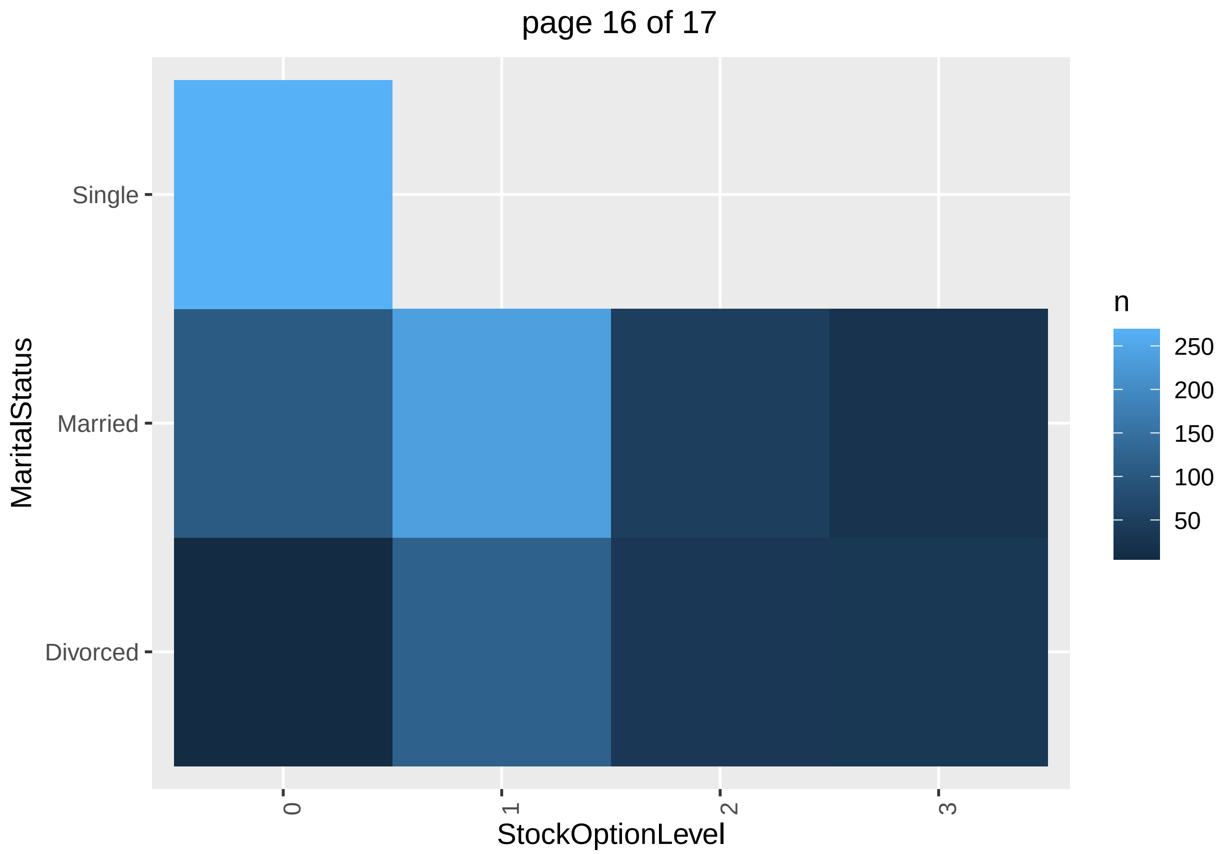


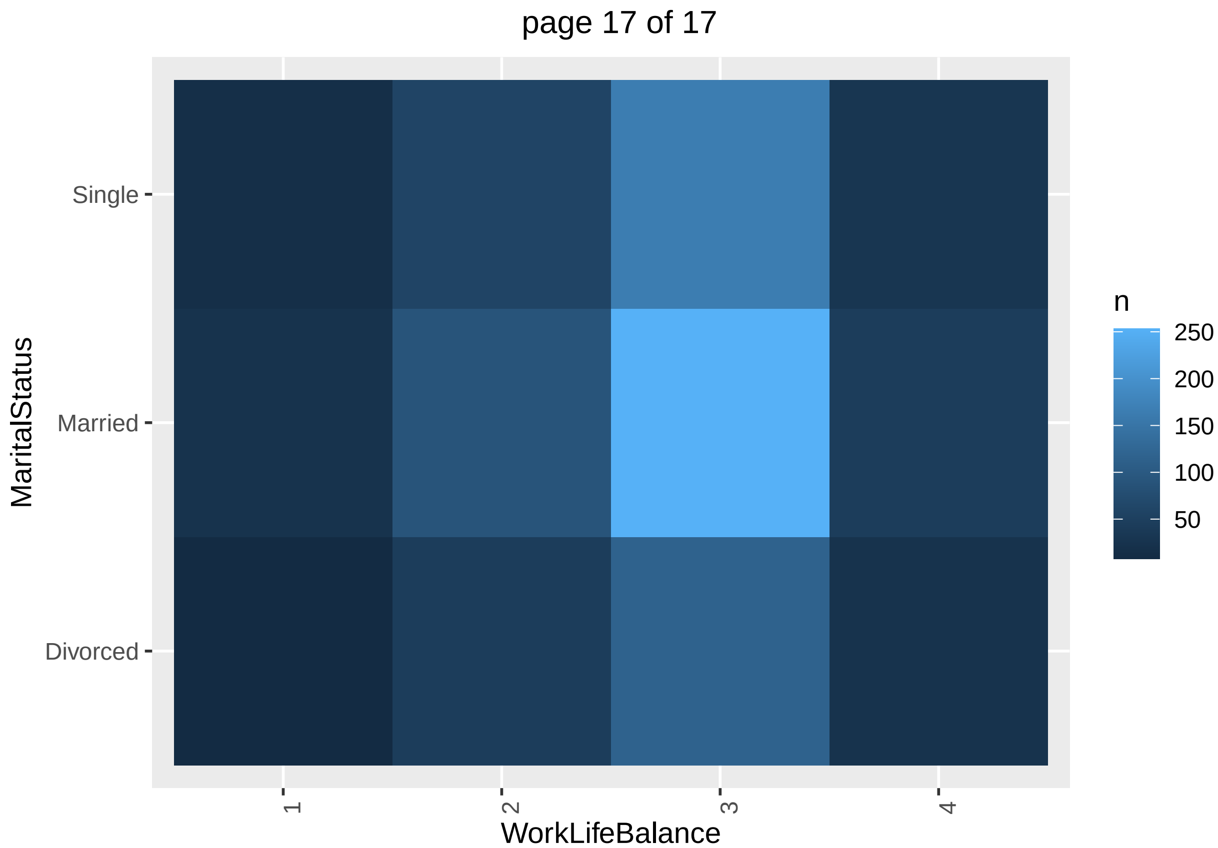
page 12 of 17
MaritalStatus
Single
Married
Divorced
Over18
Y
n
400
350
300
250
200


page 13 of 17
MaritalStatus
Single
Married
Divorced
OverTime
No
Yes
n
300
250
200
150
100


page 14 of 17
Single
Married
Divorced
MaritalStatus
3
4
PerformanceRating
n
300
200
100

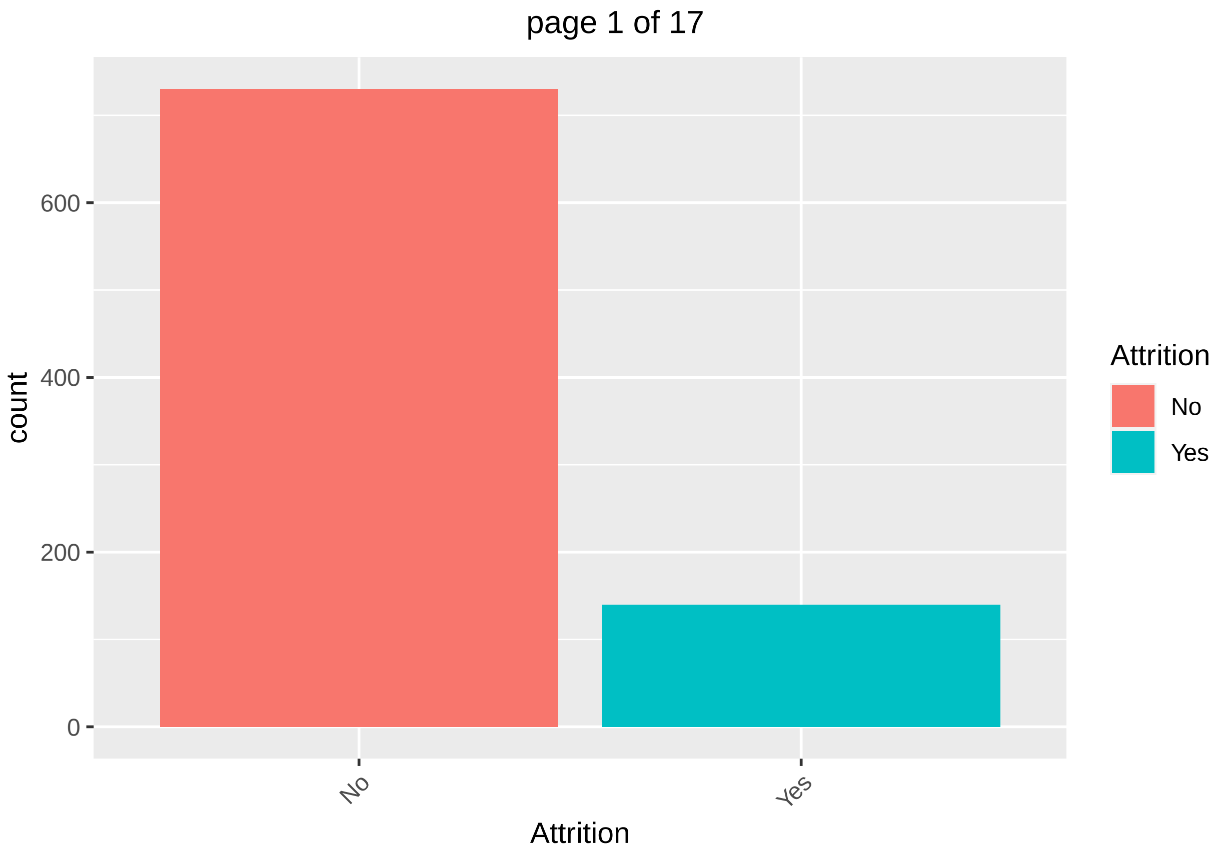

page 15 of 17
Single
Married
Divorced
MaritalStatus
1
2
3
4
RelationshipSatisfaction
n
100
80
60
40

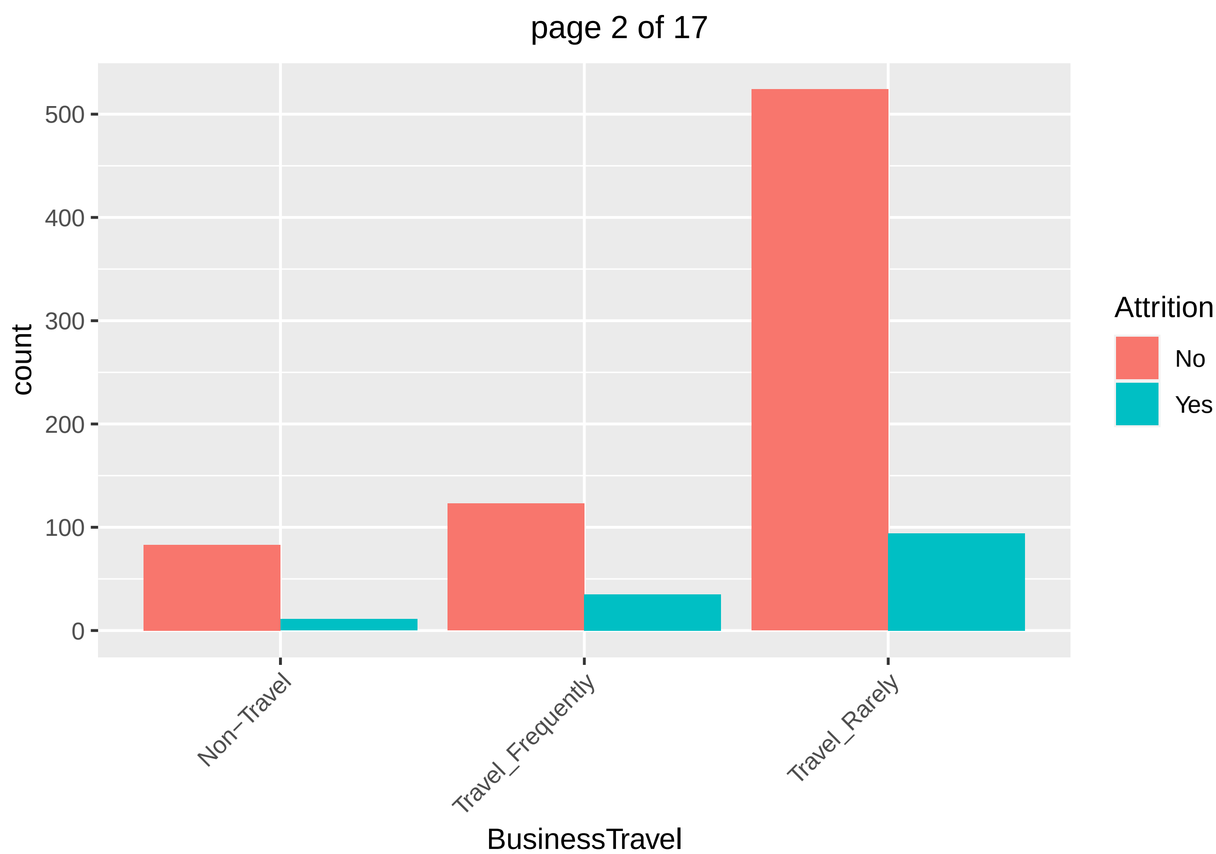

page 16 of 17
Single
Married
Divorced
MaritalStatus
0
1
2
3
StockOptionLevel
n
250
200
150
100
50

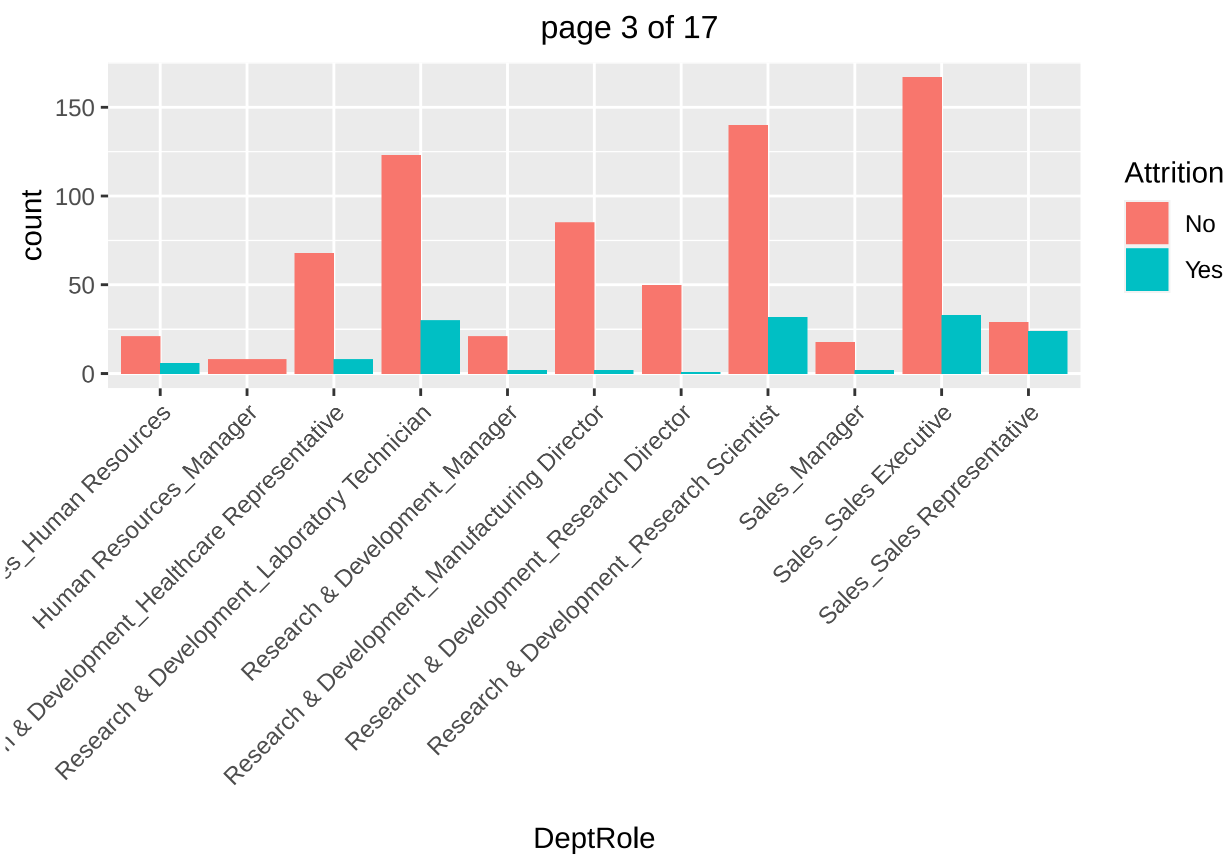


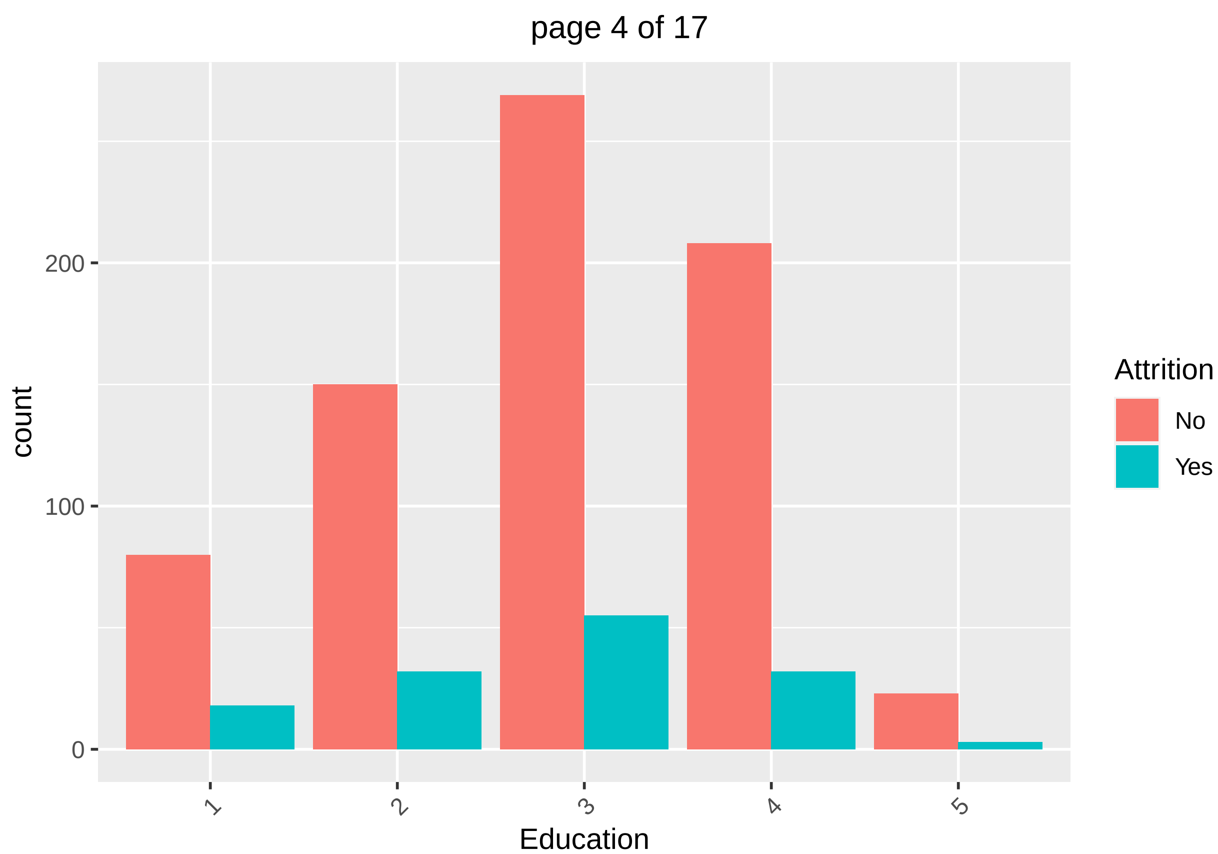
Get Bar Graphs fod Categorical

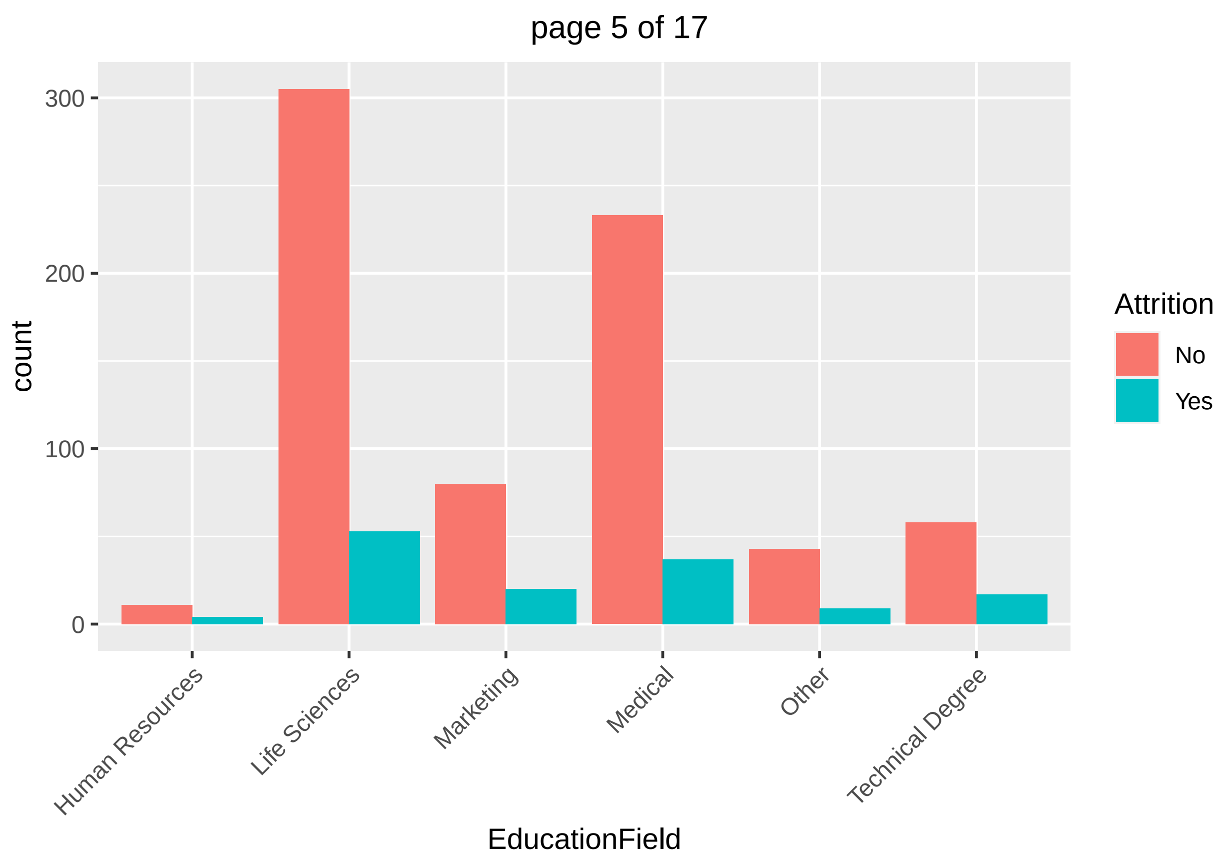
```
# Initiate plot list
plot_list <- list()

# Loop through character factors and plot
for (i in 1:ncol(df_char_fac)) {
  plot_list[[i]] <- df %>%
    ggplot(aes_string(x=colnames(df_char_fac)[i], fill="Attrition"))+
    geom_bar(position = "dodge")+
    theme(axis.text.x=element_text(angle=45,hjust=1))

}
# Plot all plots in plot_list
marrangeGrob(grobs = plot_list,
                   nrow=1, ncol=1)
```

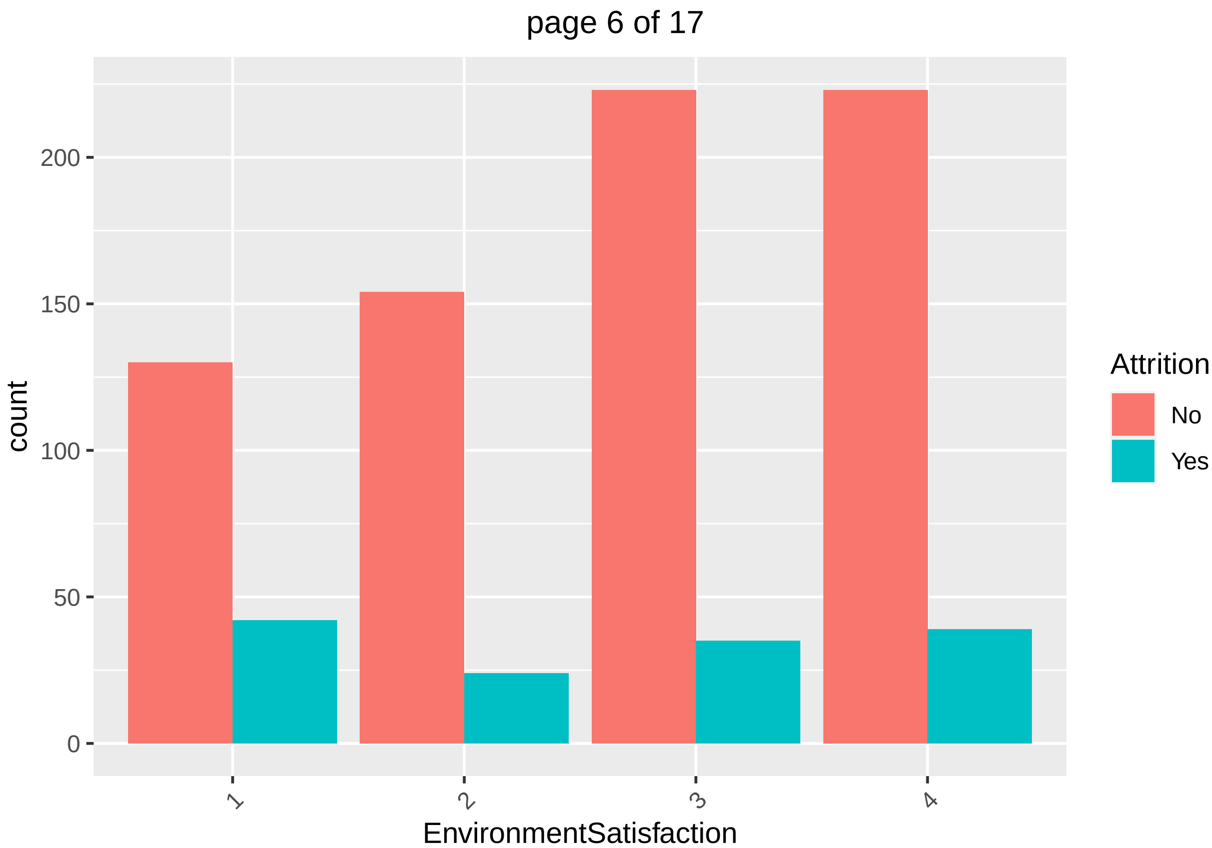

page 1 of 17
count
600
400
200
0
No
Yes
Attrition
Attrition
No
Yes

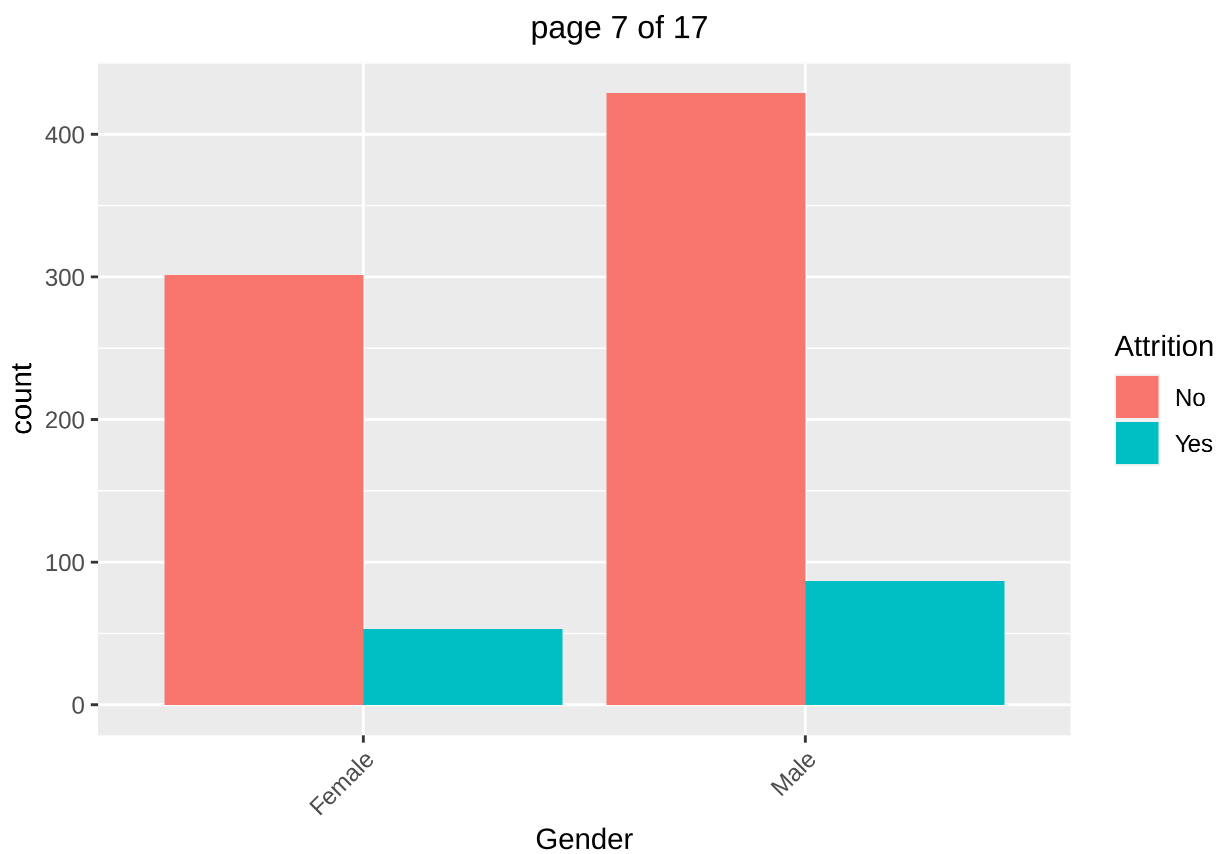

page 2 of 17
count
500
400
300
200
100
0
Attrition
No
Yes
Non−Travel
Travel_Frequently
Travel_Rarely
BusinessTravel


page 3 of 17
count
150
100
50
0
Attrition
No
Yes
s_Human Resources
Human Resources_Manager
& Development_Healthcare Representative
Research & Development_Laboratory Technician
Research & Development_Manager
Research & Development_Manufacturing Director
Research & Development_Research Director
Research & Development_Research Scientist
Sales_Manager
Sales_Sales Executive
Sales_Sales Representative
DeptRole

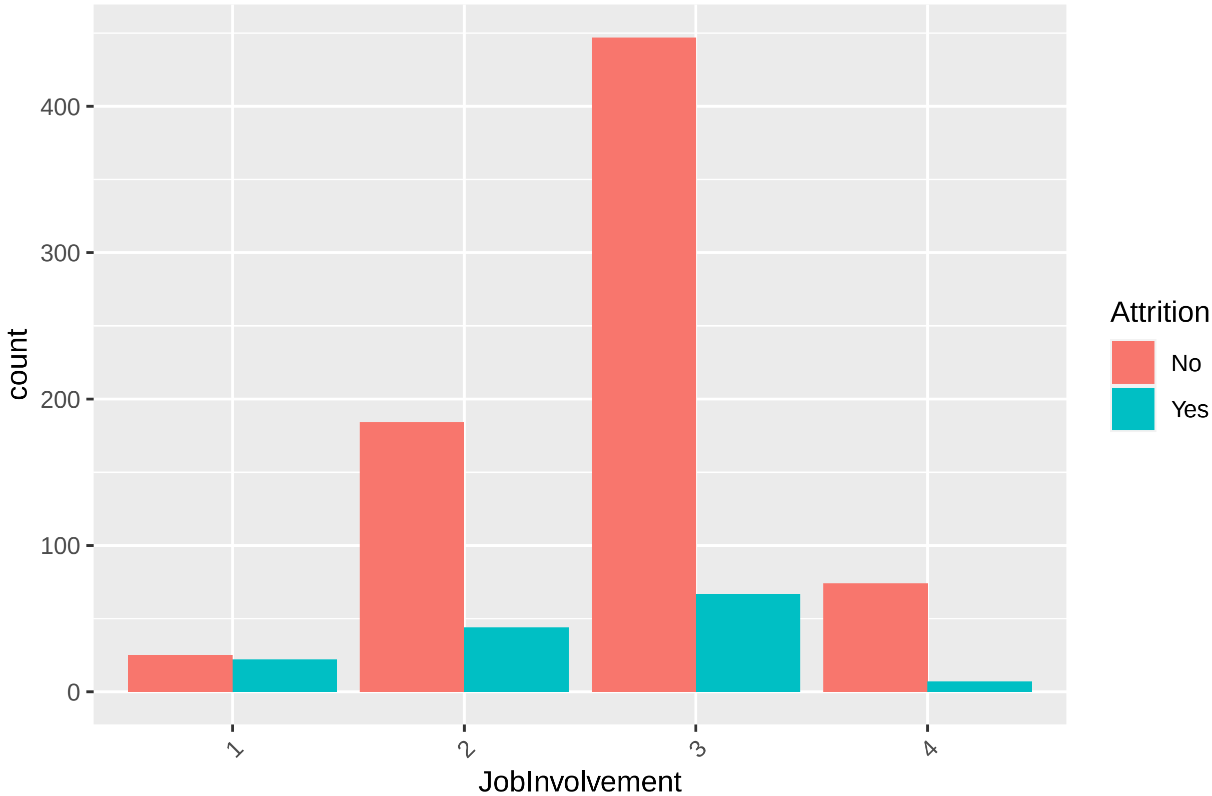

page 4 of 17
count
200
100
0
1
2
3
4
5
Education
Attrition
No
Yes


page 5 of 17
300
200
100
0
count
Human Resources
Life Sciences
Marketing
Medical
Other
Technical Degree
EducationField
Attrition
No
Yes

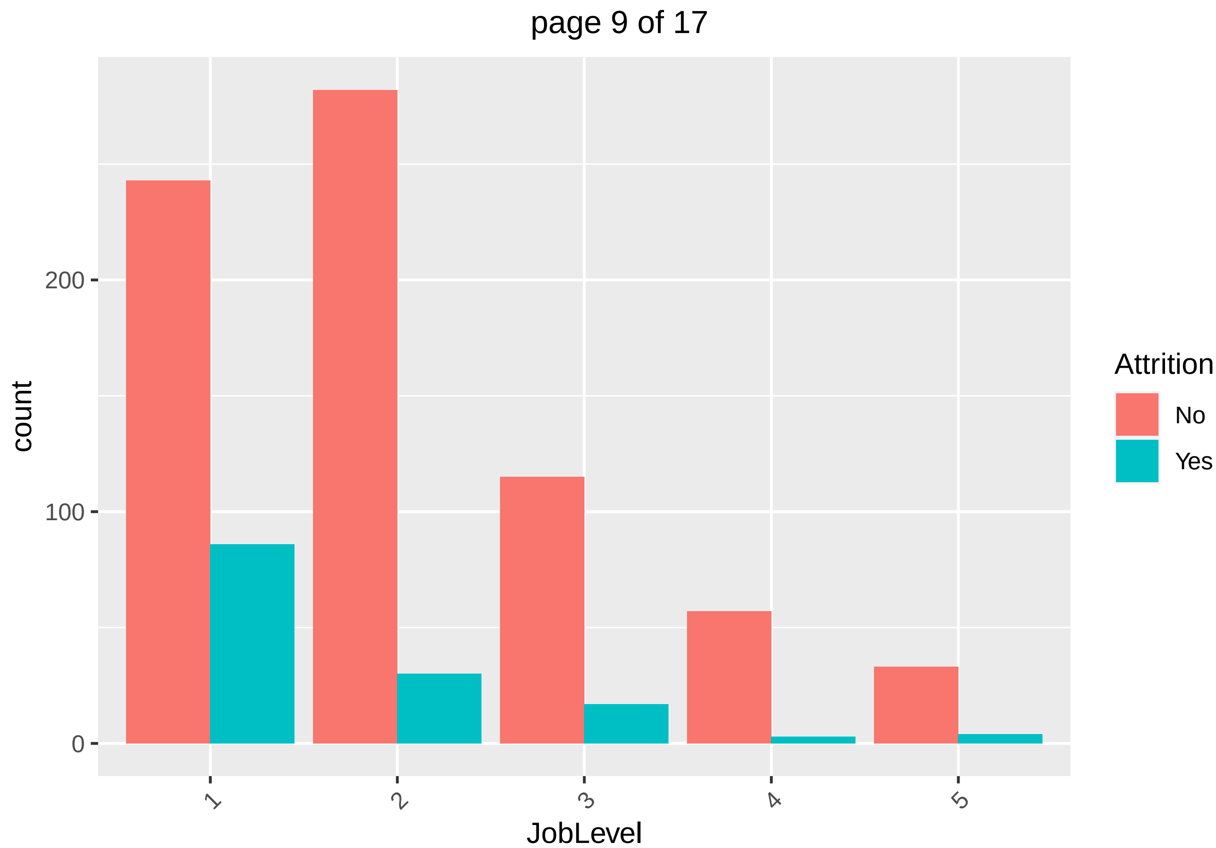

page 6 of 17
200
150
100
50
0
count
1
2
3
4
EnvironmentSatisfaction
Attrition
No
Yes

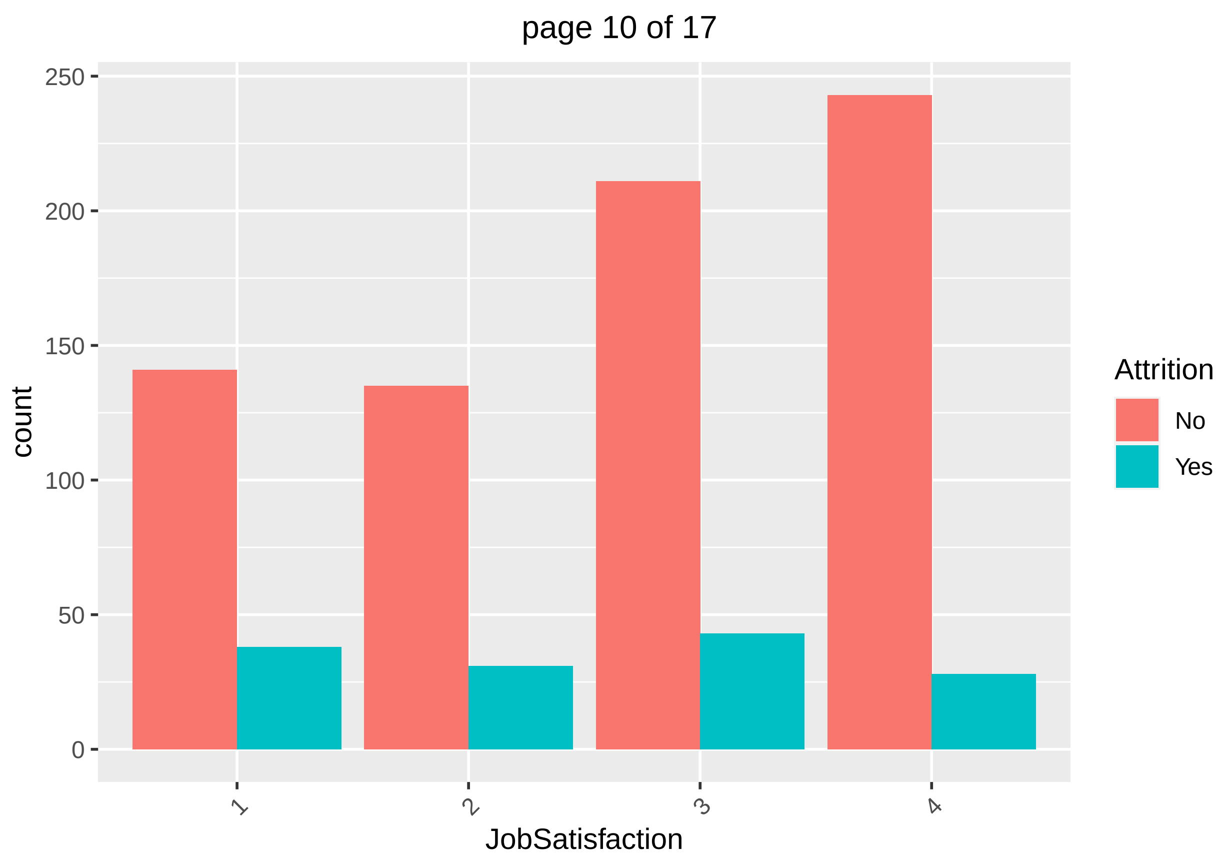

page 7 of 17
400
300
200
100
0
count
Female
Male
Gender
Attrition
No
Yes

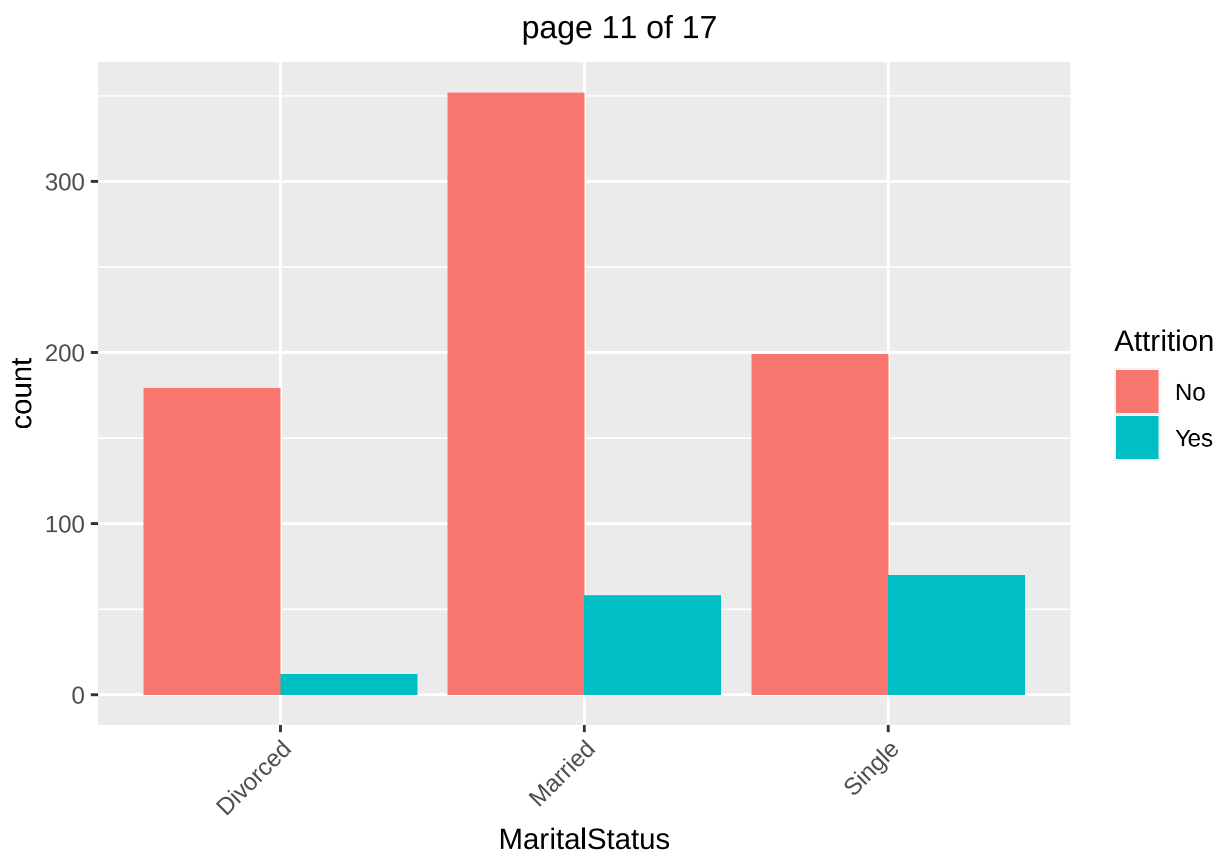
page 8 of 17

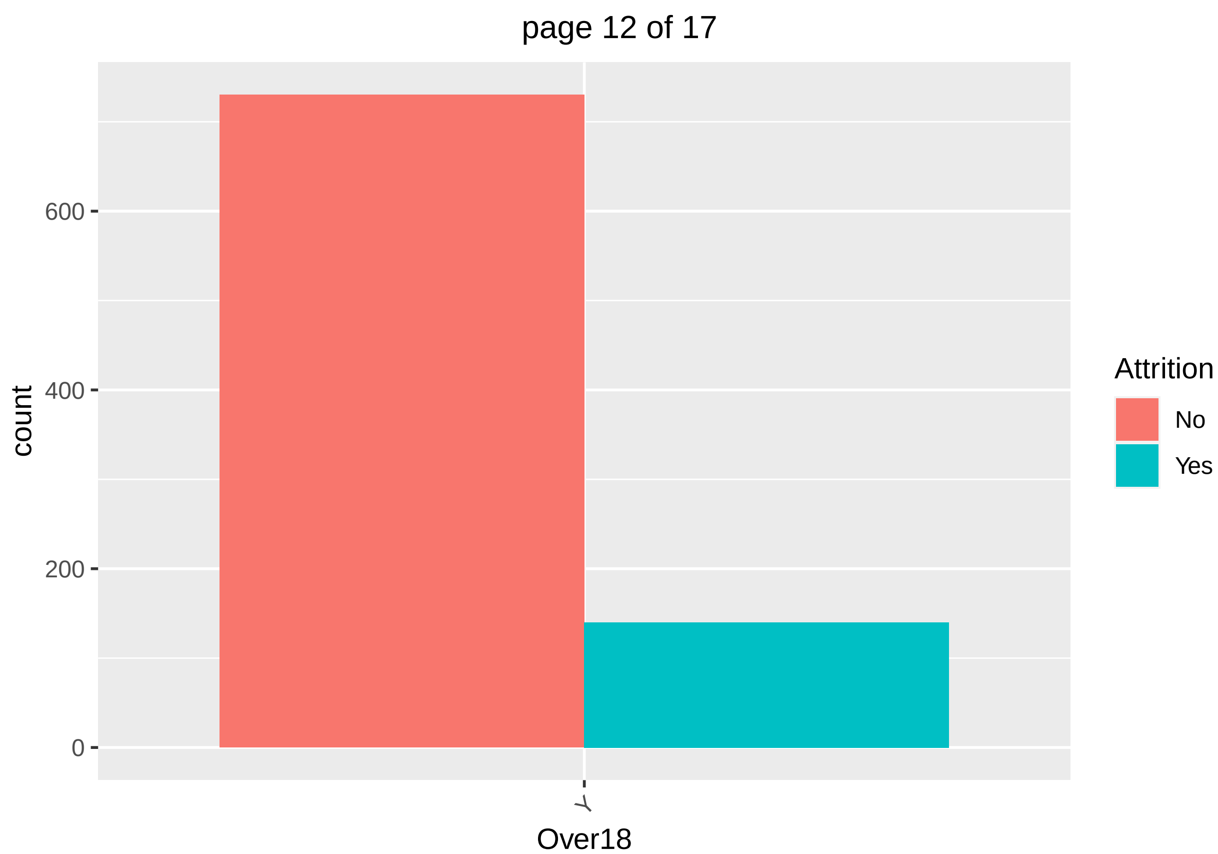


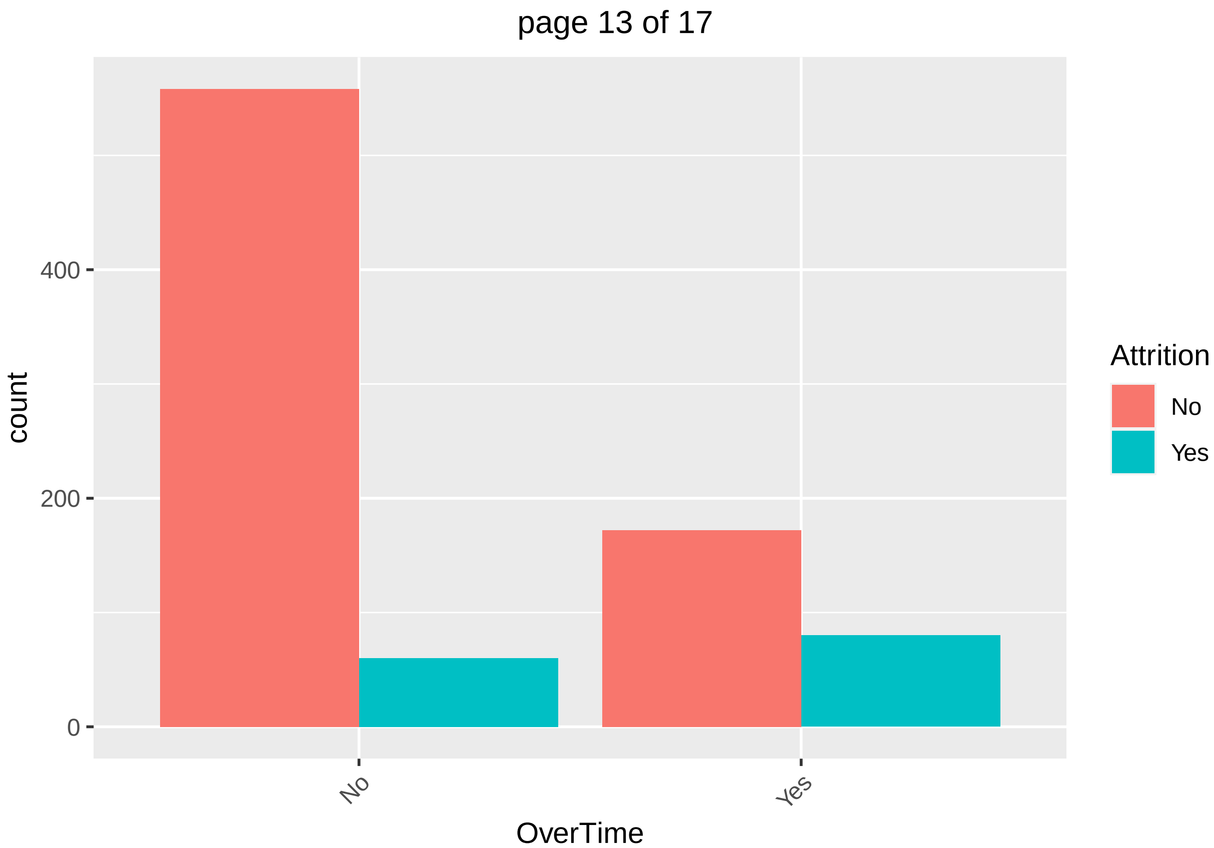
count
400
300
200
100
0
1
2
3
4
JobInvolvement
Attrition
No
Yes


page 9 of 17
count
200
100
0
1
2
3
4
5
JobLevel
Attrition
No
Yes

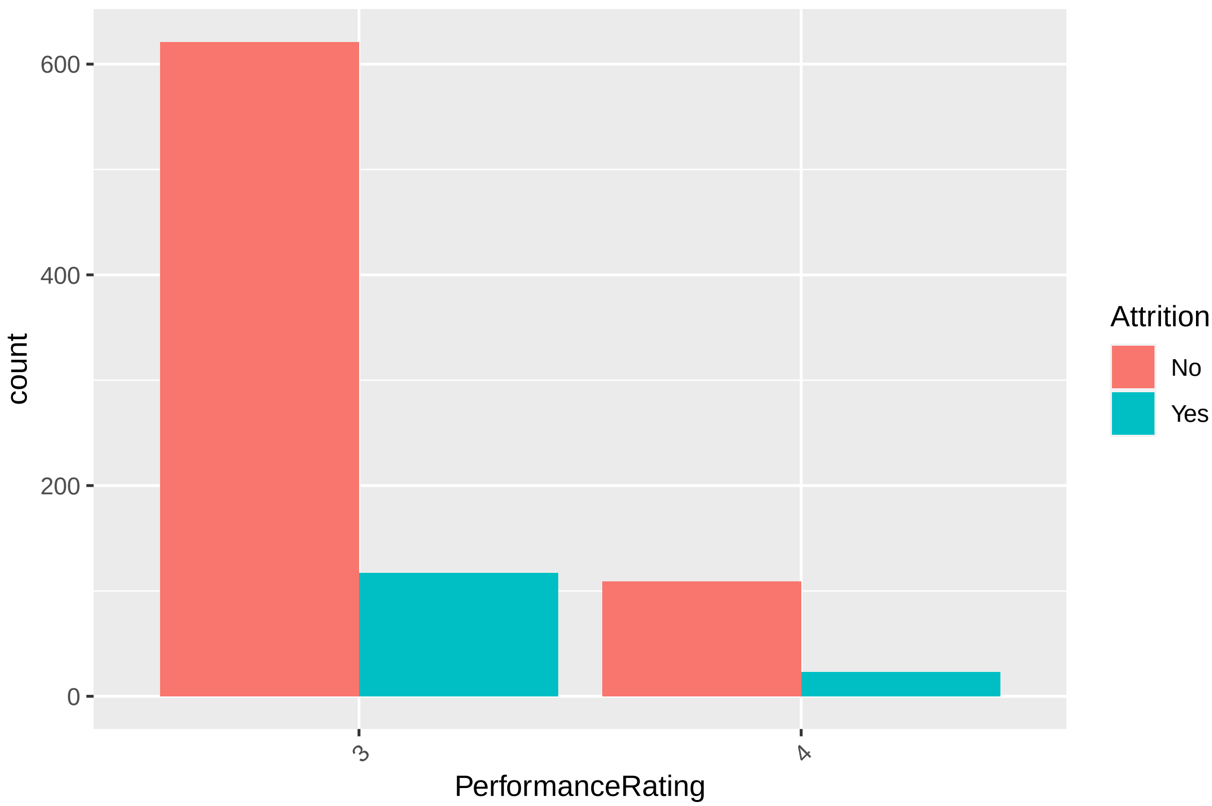

page 10 of 17
250
200
150
100
50
0
count
1
2
3
4
JobSatisfaction
Attrition
No
Yes

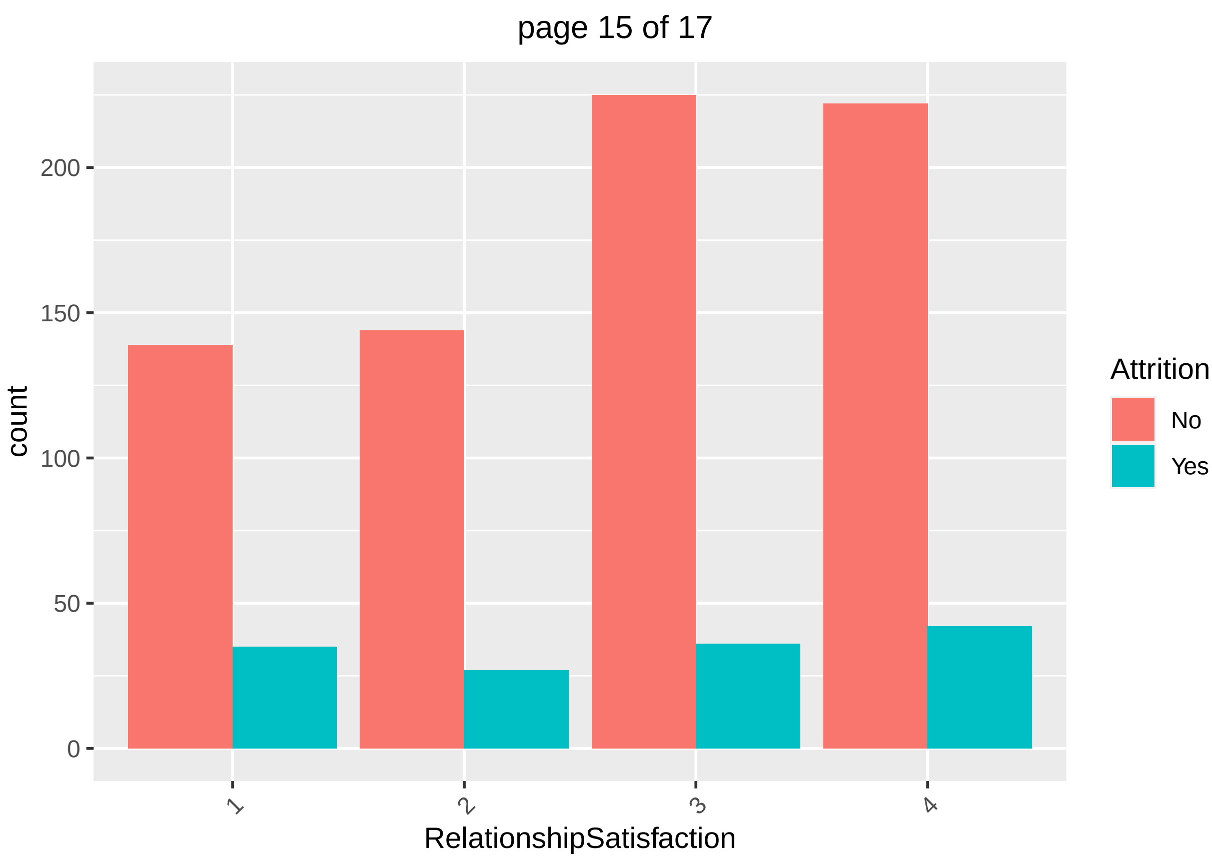

page 11 of 17
count
300
200
100
0
Divorced
Married
Single
MaritalStatus
Attrition
No
Yes

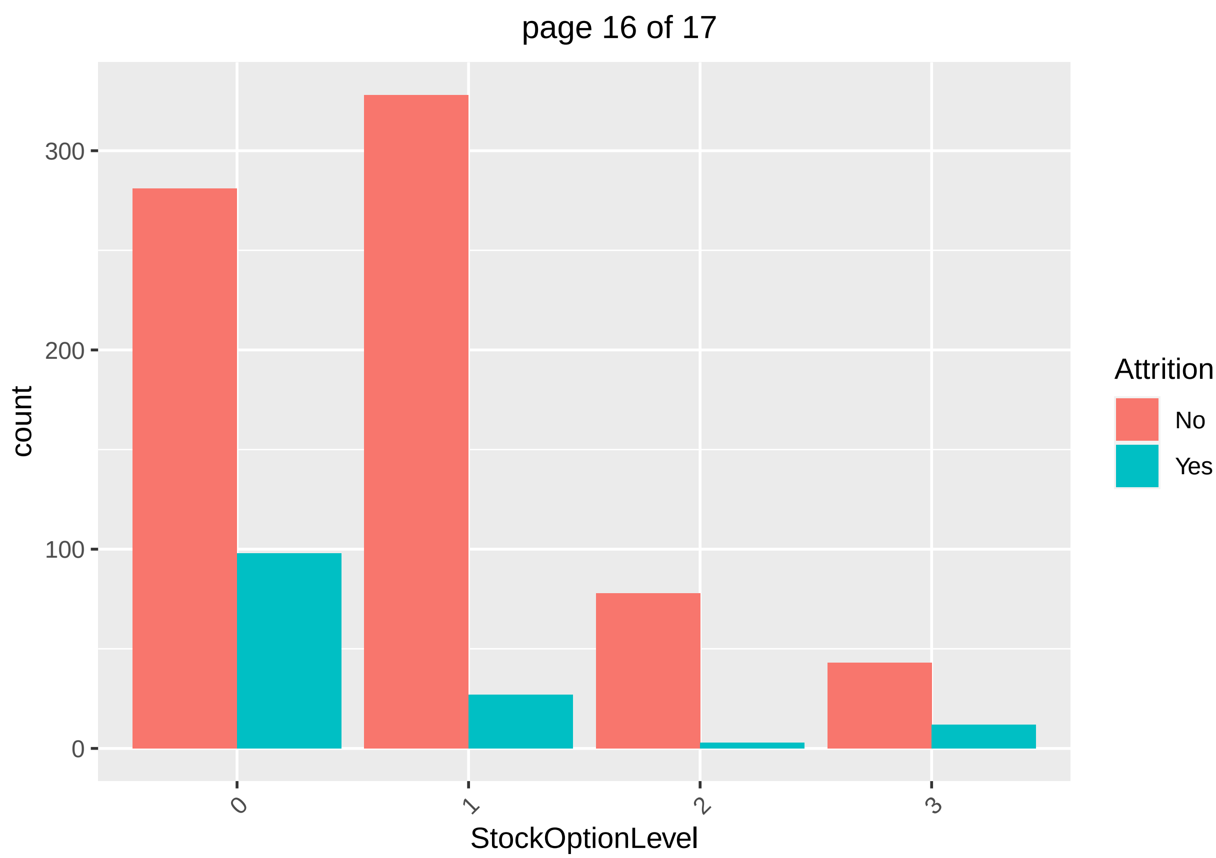

page 12 of 17
count
600
400
200
0
Y
Over18
Attrition
No
Yes

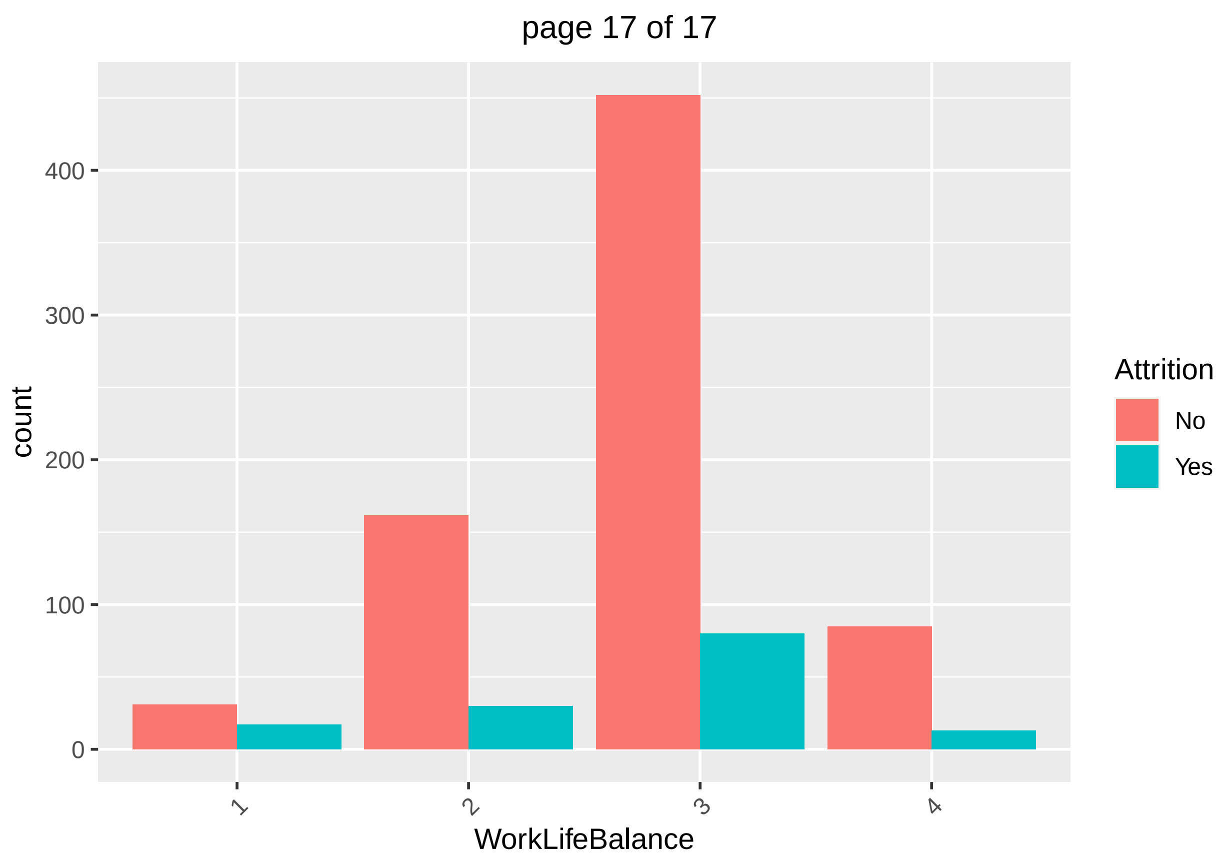

page 13 of 17
count
400
200
0
No
Yes
OverTime
Attrition
No
Yes

page 14 of 17




page 15 of 17
count
200
150
100
50
0
1
2
3
4
RelationshipSatisfaction
Attrition
No
Yes


page 16 of 17
count
300
200
100
0
0
1
2
3
StockOptionLevel
Attrition
No
Yes


page 17 of 17
count
400
300
200
100
0
1
2
3
4
WorkLifeBalance
Attrition
No
Yes