ATTRITION ANALYSIS

Huy Hoang Nguyen

DDS Analytics

Objective

- Exploratory Data Analysis to understand main factors for Attrition problem
- Job Role Trends
- Build Models to predict Salary and Attrition
- Validation of Models for Salary and Attrition
- Create a ShinyApp for data visualization

Objective

- Exploratory Data Analysis to understand main factors for Attrition problem
- Job Role Trends
- Build Models to predict Salary and Attrition
- Validation of Models for Salary and Attrition
- Create a ShinyApp for data visualization

Objective

- Exploratory Data Analysis to understand main factors for Attrition problem
- Job Role Trends
- Build Models to predict Salary and Attrition
- Validation of Models for Salary and Attrition
- Create a ShinyApp for data visualization

Objective

- Exploratory Data Analysis to understand main factors for Attrition problem
- Job Role Trends
- Build Models to predict Salary and Attrition
- Validation of Models for Salary and Attrition
- Create a ShinyApp for data visualization

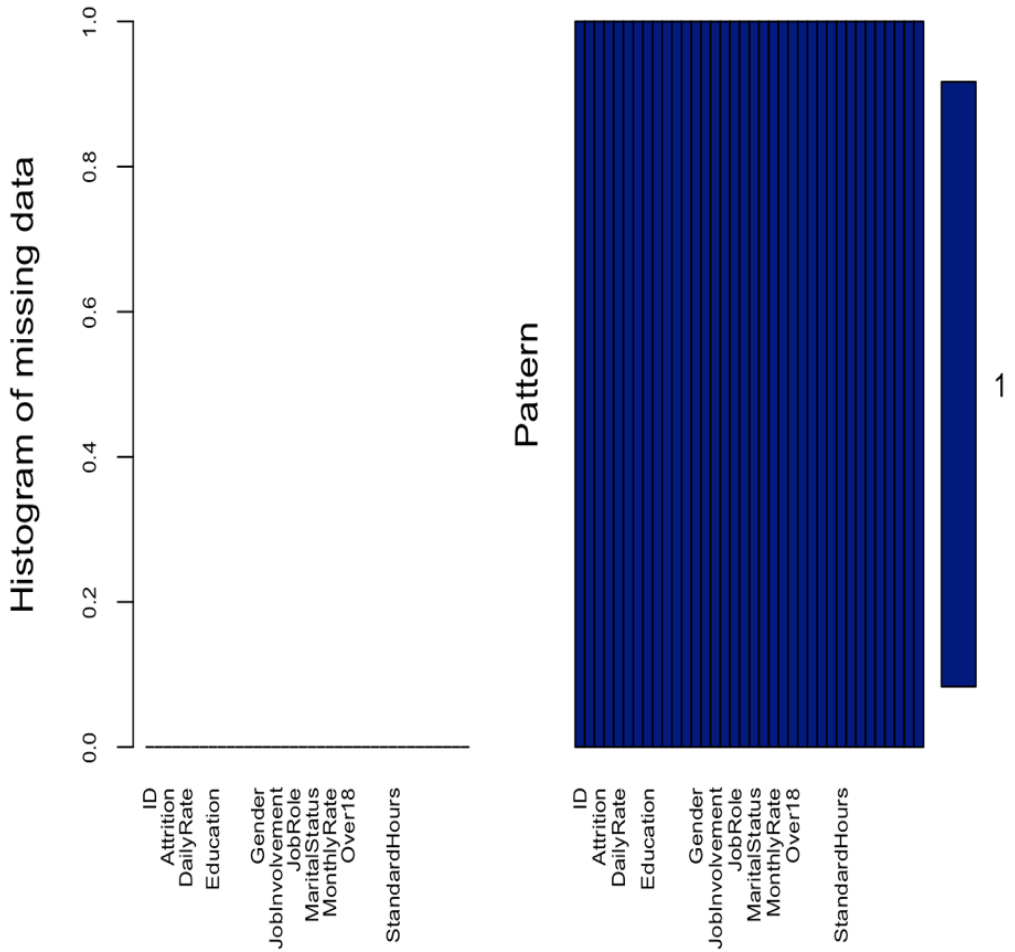
Objective

- Exploratory Data Analysis to understand main factors for Attrition problem
- Job Role Trends
- Build Models to predict Salary and Attrition
- Validation of Models for Salary and Attrition
- Create a ShinyApp for data visualization

About me:

- Data Cruncher
- Hype Crew

Exploratory Data Analysis

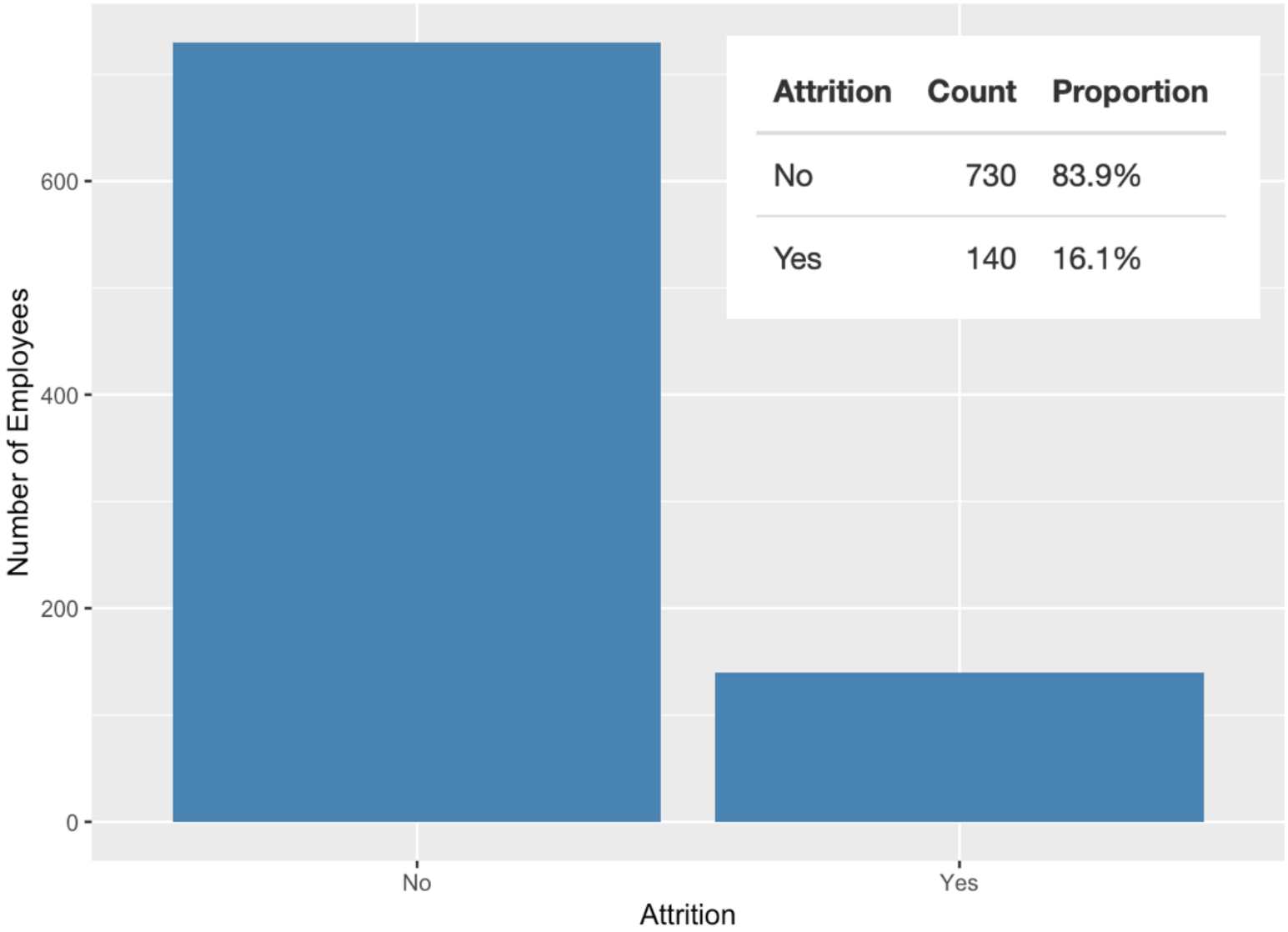


OUR DATASET

Number of rows	870
Number of columns	36
Column type frequency:	
factor	9
numeric	27

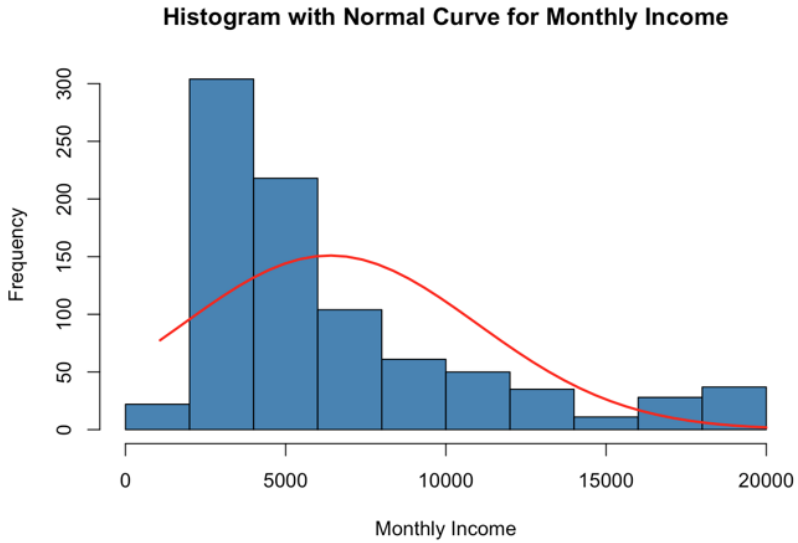
- NO MISSING DATA
- 870 OBSERVATIONS WITH 36 VARIABLES

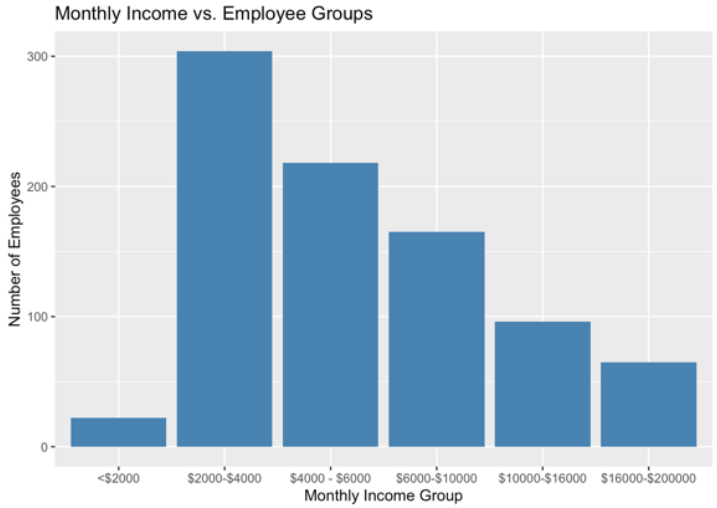
Attrition

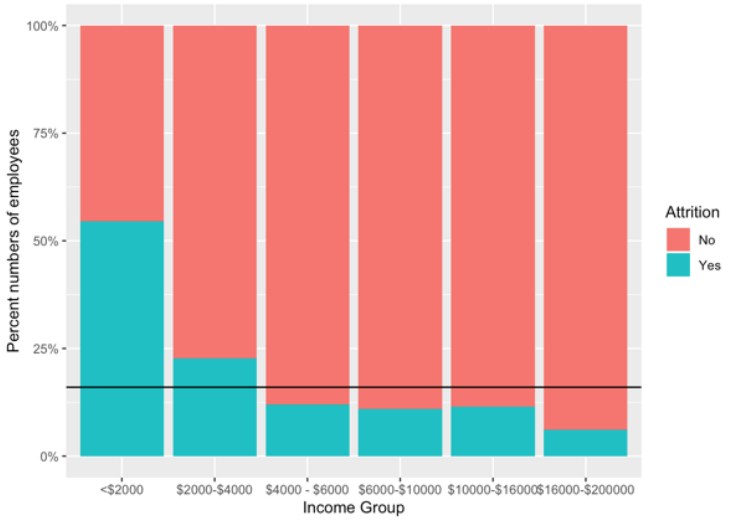


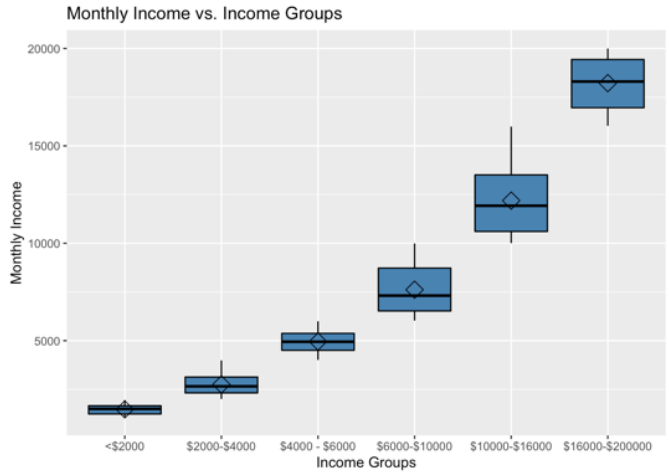
Attrition	Count	Proportion
No	730	83.9%
Yes	140	16.1%

Monthly Income

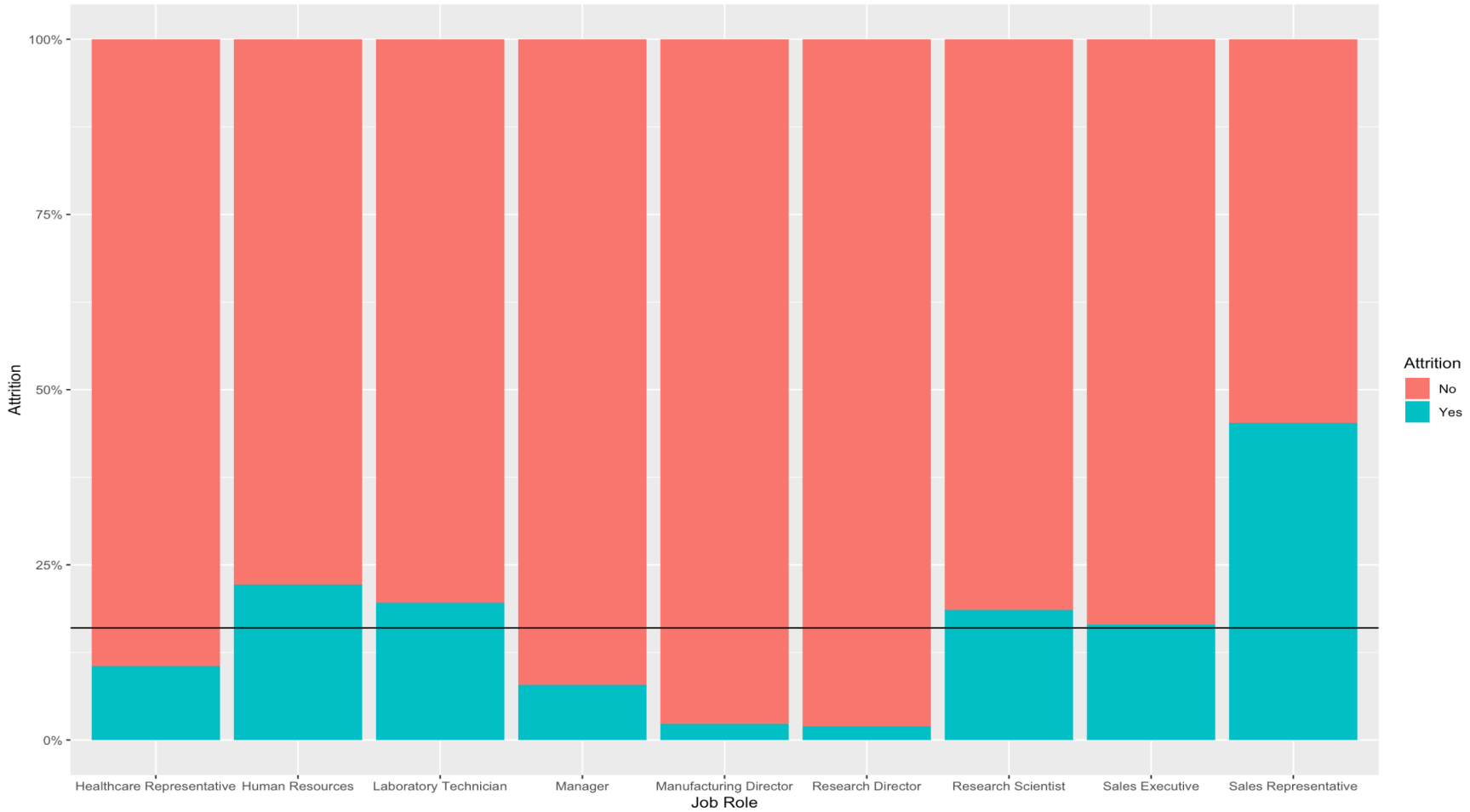




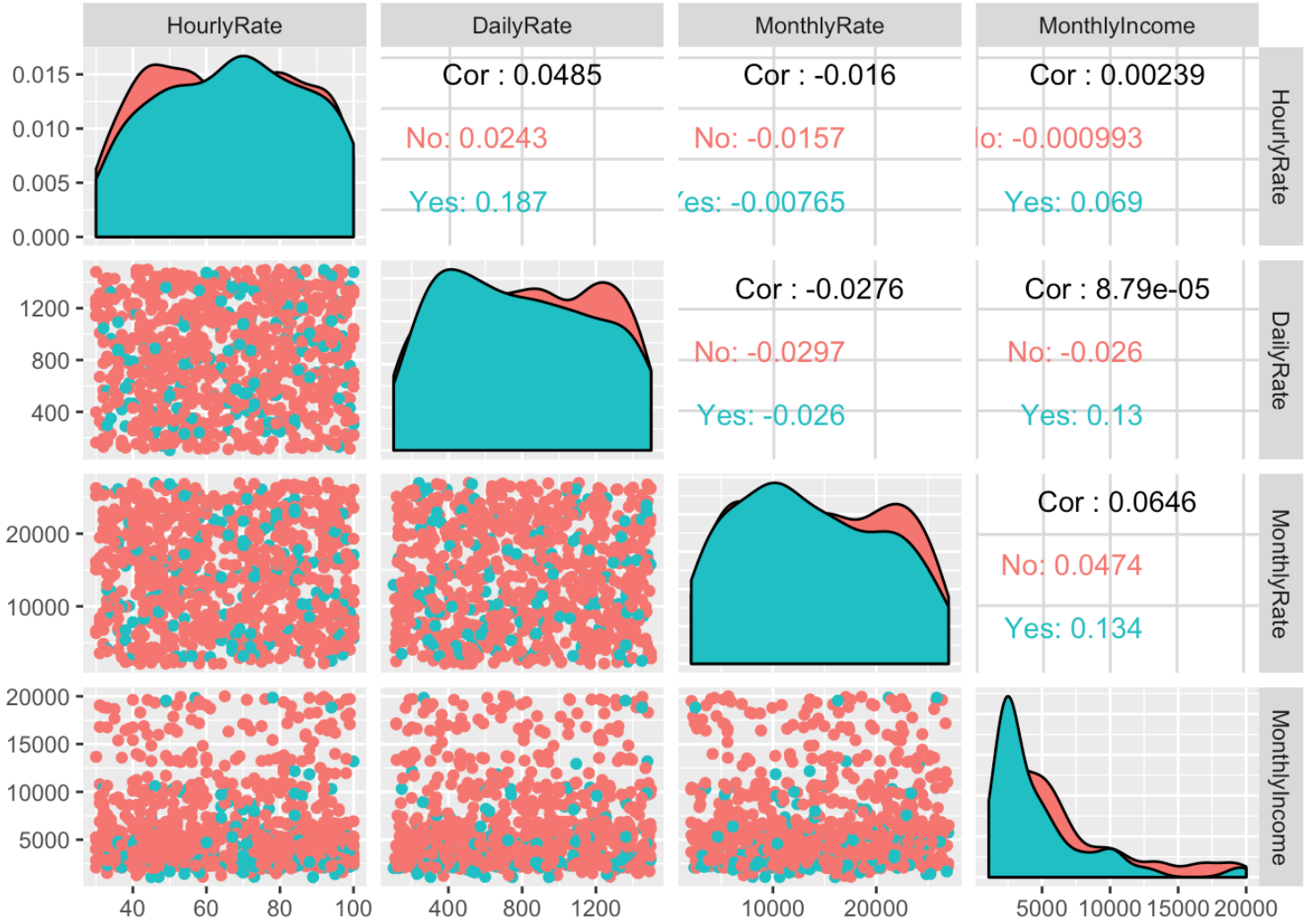




Job Role vs. Attrition rate



Hourly Rate vs Daily Rate vs Monthly Rate vs Monthly Income



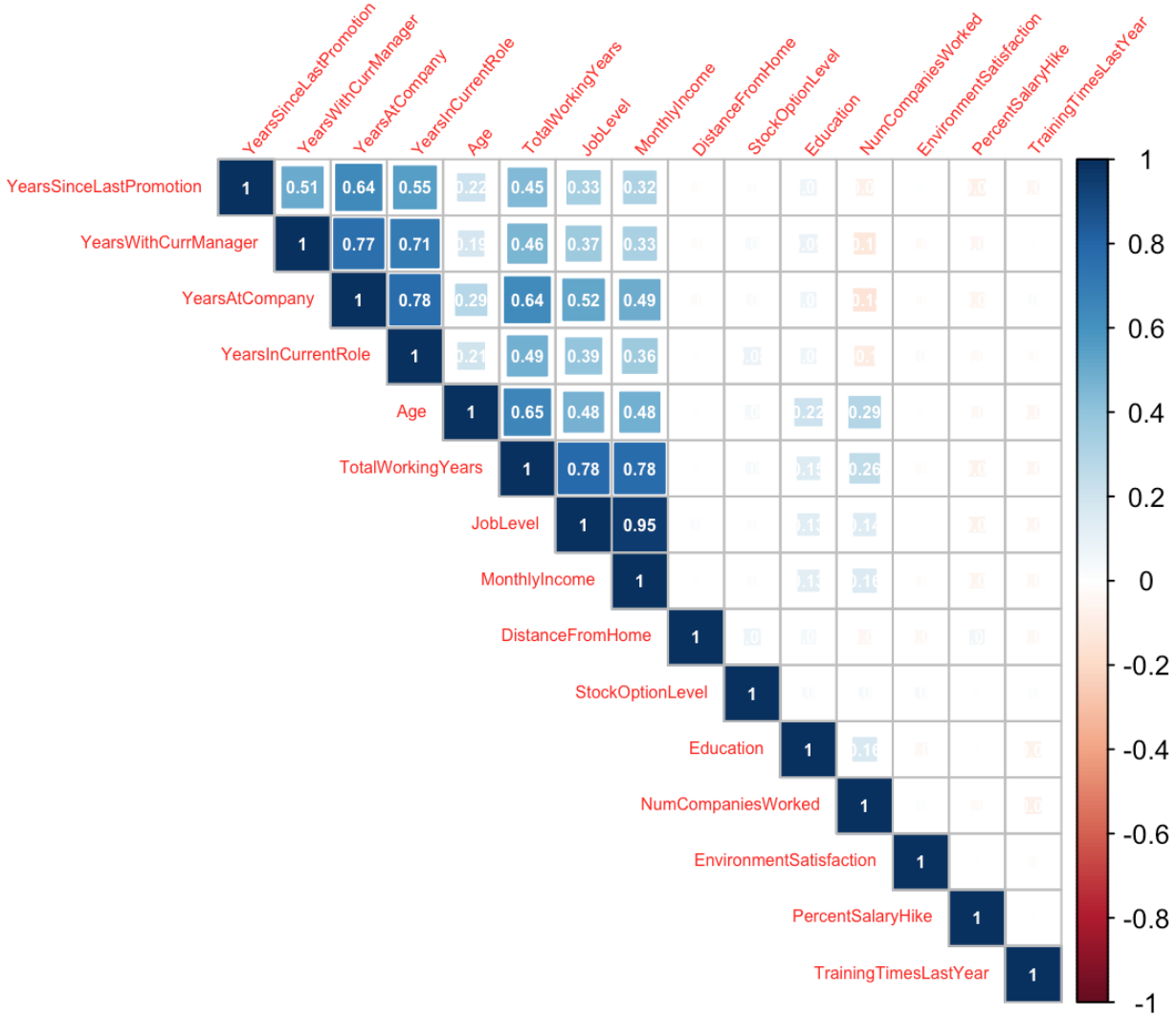
T-TEST ON ATTRITION

Variable	P-Value
Age	0.0000505
Monthly Income	0.0000002
Education	0.1421319
Distance From Home	0.0164052
Environment Satisfaction	0.0339727
Job Level	0.0000004
Number of Companies Worked	0.0978823
Percent Salary Hike	0.6692297
Stock Option Level	0.0000386
Total Working Years	0.0000007
Training Times Last Year	0.0594832
Years At Company	0.0002563
Years In Current Role	0.0000015
Years Since Last Promotion	0.8983165
Years With Current Manager	0.0000051

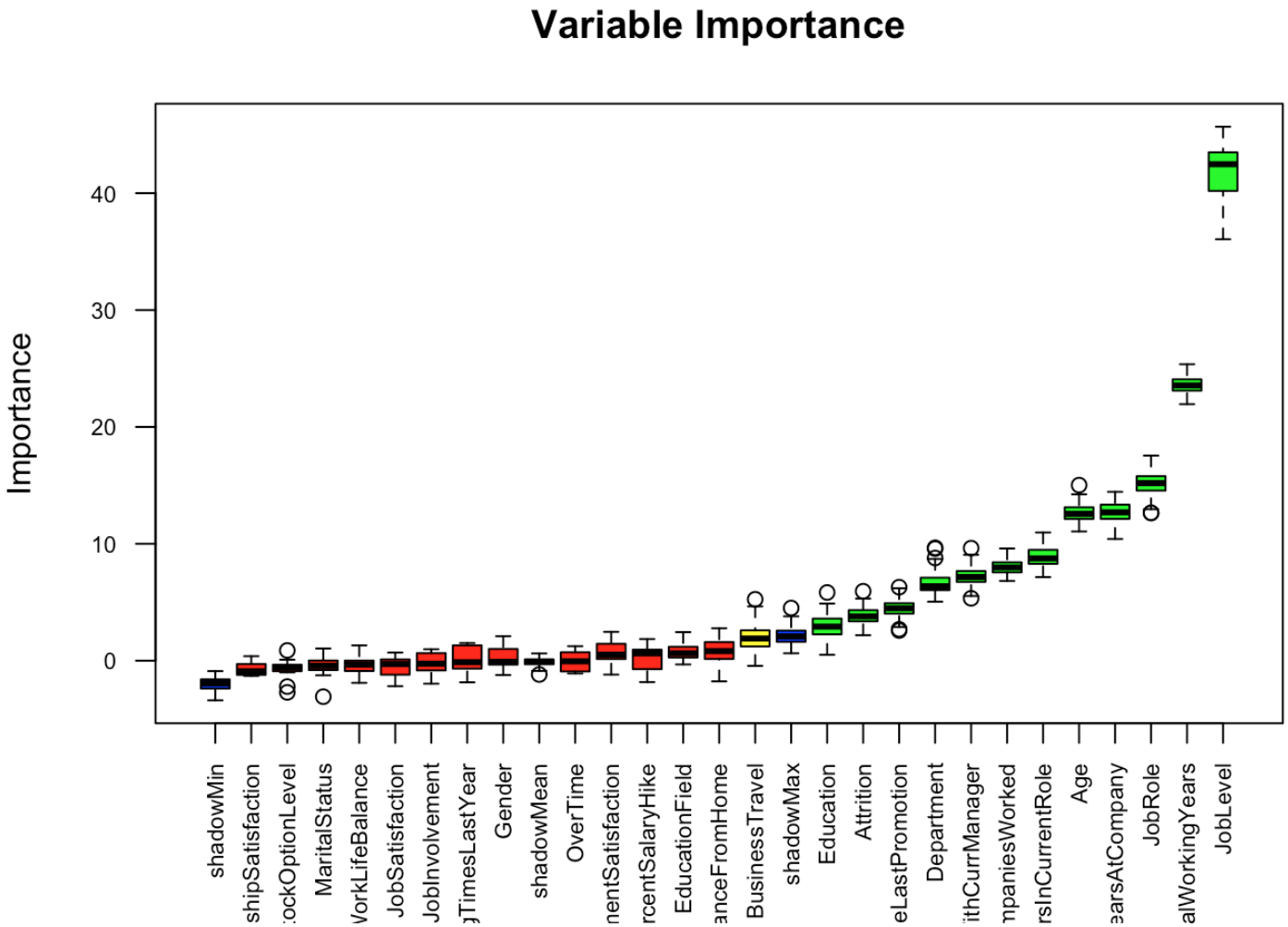
T-TEST ON GENDER

Variable	P-Value
Healthcare Representative	0.8080436
Human Resources	0.8080436
Laboratory Technician	0.8080436
Manager	0.8080436
Manufacturing Director	0.8080436
Research Director	0.8080436
Research Scientist	0.8080436
Sales Executive	0.8080436
Sales Representative	0.8080436

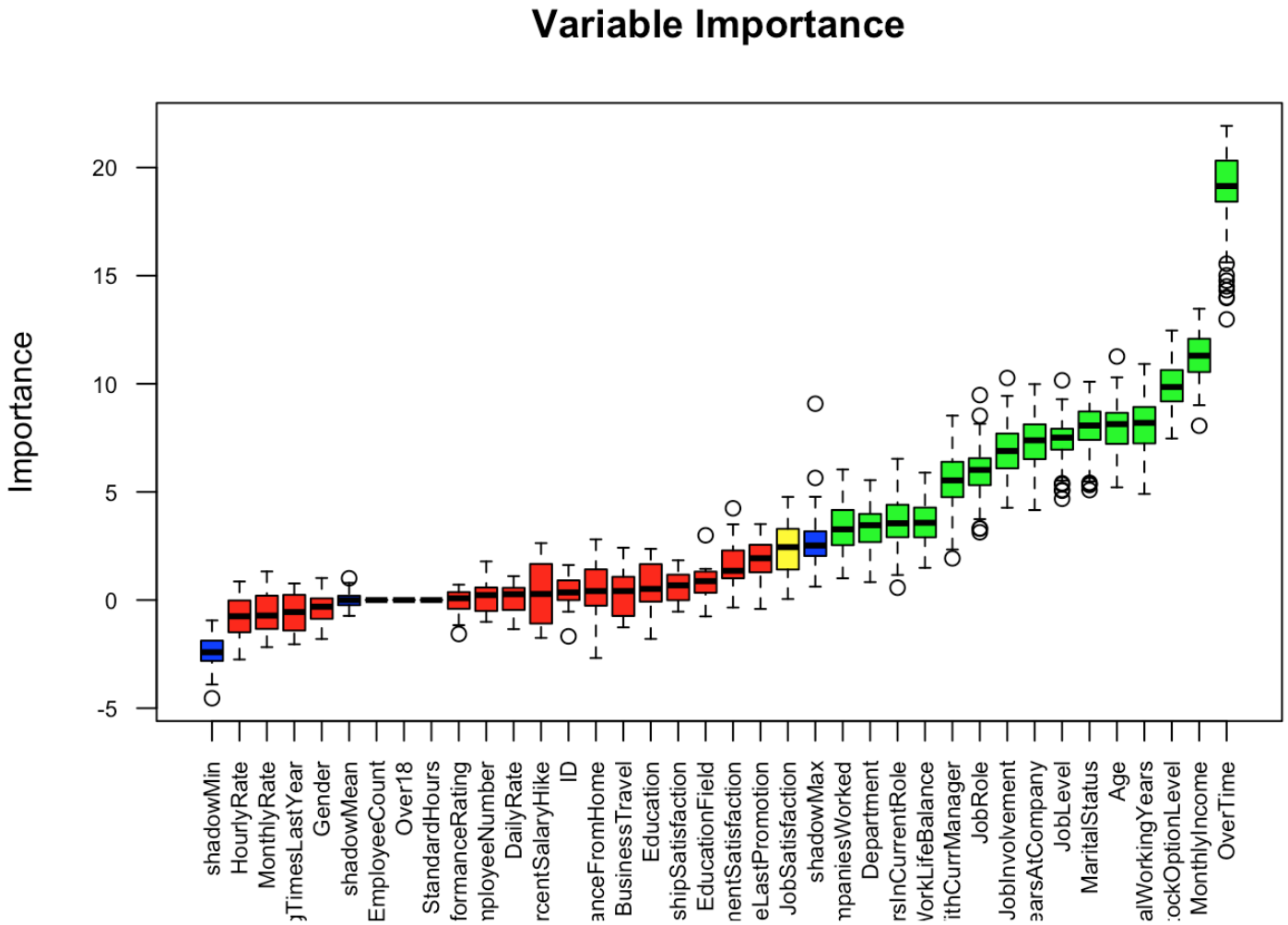
Numeric variables correlation



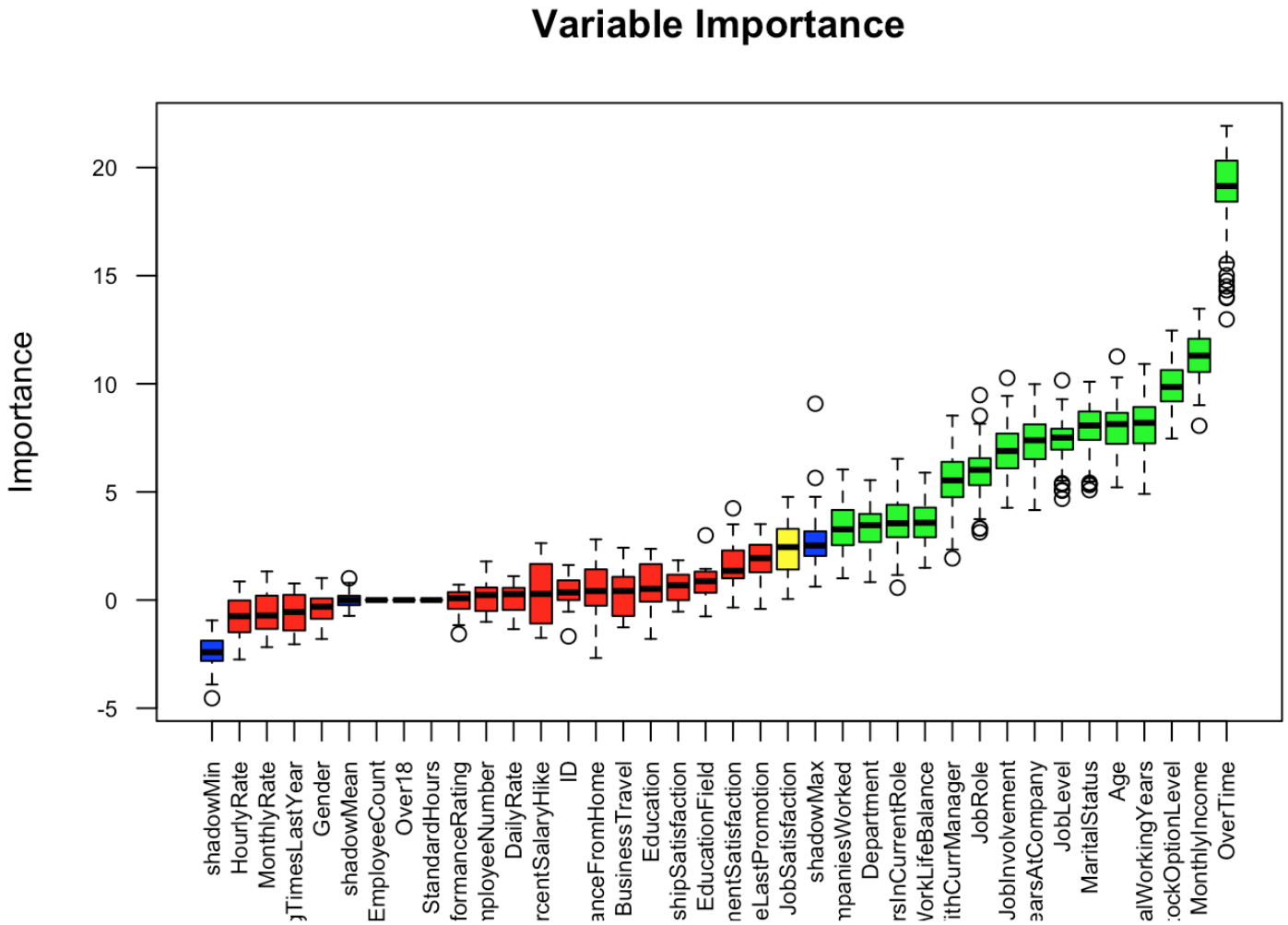
Automated EDA - Monthly Income



Automated EDA - Attrition



Automated EDA - Attrition



Multiple Linear Regression and Validation for Salary

- Build a model by using multiple linear regression on Monthly Income related to 14 important variables: Age, Attrition, BusinessTravel, Department, Education, JobLevel, JobRole, NumCompaniesWorked, TotalWorkingYears, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager.
- RMSE is around 1031.816 USD.
- Validation for Salary.

Naive Bayes classifiers and Validation for Attrition

- Build a model by using Naive Bayes classifiers on Attrition related to 17 important variables: Age, Department, EnvironmentSatisfaction, JobInvolvement, JobLevel, JobRole, JobSatisfaction, MaritalStatus, MonthlyIncome, NumCompaniesWorked, OverTime, StockOptionLevel, TotalWorkingYears, WorkLifeBalance, YearsAtCompany, YearsInCurrentRole, YearsWithCurrManager.

```
                Accuracy : 0.8046
                  95% CI : (0.7378, 0.8607)
    No Information Rate : 0.8678
    P-Value [Acc > NIR] : 0.992836

                   Kappa : 0.3604

 Mcnemar's Test P-Value : 0.003551

            Sensitivity : 0.8278
            Specificity : 0.6522
```

- Validation for Attrition

Shiny App

- https://hnguye01.shinyapps.io/DDSAnalyticsApp/

DDSAnalytics Visualization

Huy Hoang Nguyen

Attrition Analysis

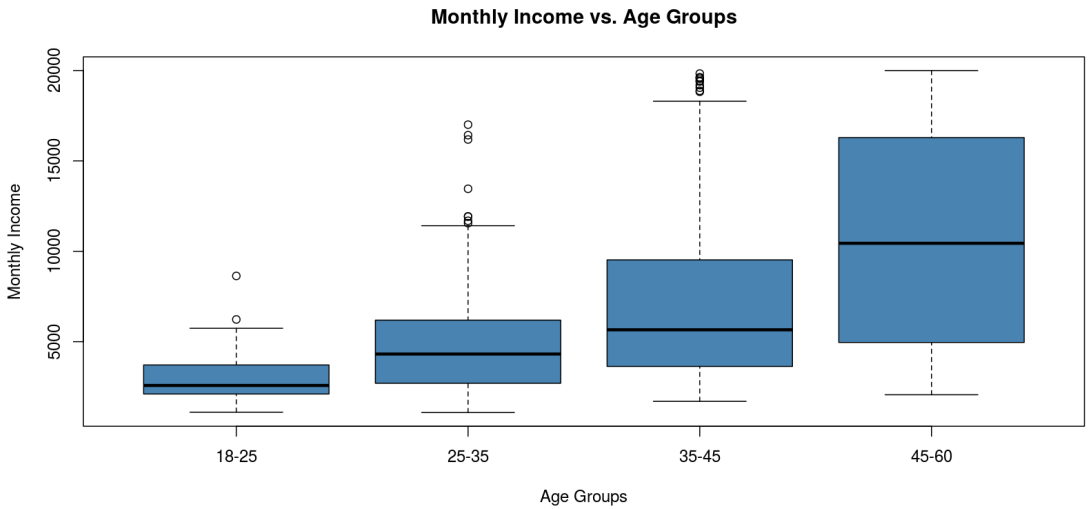
(Visualization between Monthly Income or Attrition and Feature Variables)

Features:

Age

Please choose:

- (•) Monthly Income
- ( ) Attrition



Created by HHN - hoangnguyen@smu.edu (2019)

Conclusion

- EDA to get information about important variables to build models
- Sales Representatives with highest rate and Manufacturing and Research Directors with the lowest rates in Attrition
- Github link: https://github.com/hnguye01/6306two

Conclusion

- EDA to get information about important variables to build models
- Sales Representatives with highest rate and Manufacturing and Research Directors with the lowest rates in Attrition
- Github link: https://github.com/hnguye01/6306two

Conclusion

- EDA to get information about important variables to build models
- Sales Representatives with highest rate and Manufacturing and Research Directors with the lowest rates in Attrition
- Github link: https://github.com/hnguye01/6306two

Thank you for your attention