IBM Employee Attrition R Markdown

Scotty

2024-01-19

Attrition in an Organization:

This is my first official R Markdown document. I have utilized a public dataset of attrition rates within an organization to showcase my newly learned R proficiency to document and solve a case study question.

Case Study:

I will use R to:

- Determine causes for increased team member attrition.
- Document R coding and provide visuals to provide evidence of possible correlations.
- Lastly answer the case study question of how likely a team member of 5 years is to leave the company.

Getting Started:

I first installed and loaded all needed R packages.

I then loaded the dataset that I would be analyzing. I additionally used the colnames function to show the names of all columns found within the dataset.

```
HR_Employee_Attrition_dataset <- read.csv("~/HR_Employee_Attrition_dataset.csv")
colnames(HR_Employee_Attrition_dataset)
```

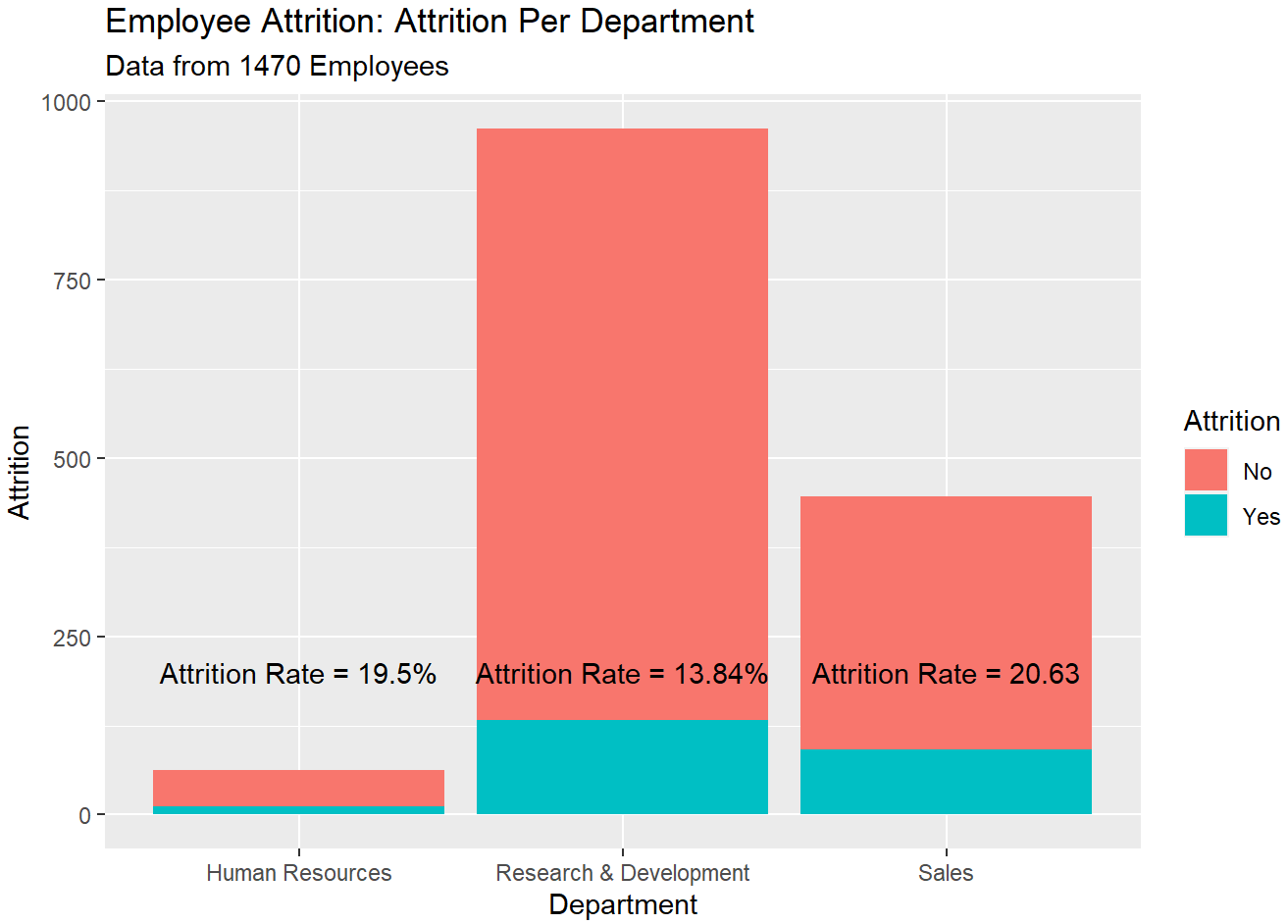
```
##  [1] "Age"                      "Attrition"
##  [3] "BusinessTravel"           "DailyRate"
##  [5] "Department"               "DistanceFromHome"
##  [7] "Education"                "EducationField"
##  [9] "EmployeeCount"            "EmployeeNumber"
## [11] "EnvironmentSatisfaction"  "Gender"
## [13] "HourlyRate"               "JobInvolvement"
## [15] "JobLevel"                 "JobRole"
## [17] "JobSatisfaction"          "MaritalStatus"
## [19] "MonthlyIncome"            "MonthlyRate"
## [21] "NumCompaniesWorked"       "Over18"
## [23] "OverTime"                 "PercentSalaryHike"
## [25] "PerformanceRating"        "RelationshipSatisfaction"
## [27] "StandardHours"            "StockOptionLevel"
## [29] "TotalWorkingYears"        "TrainingTimesLastYear"
## [31] "WorkLifeBalance"          "YearsAtCompany"
## [33] "YearsInCurrentRole"       "YearsSinceLastPromotion"
## [35] "YearsWithCurrManager"
```

Analysis Utilizing Plots

Now it was time to use ggplot2 to start analyzing the data and discover correlations that would help me uncover which factors contributed the most to employee turnover.

Plot 1 (Attrition Per Department)

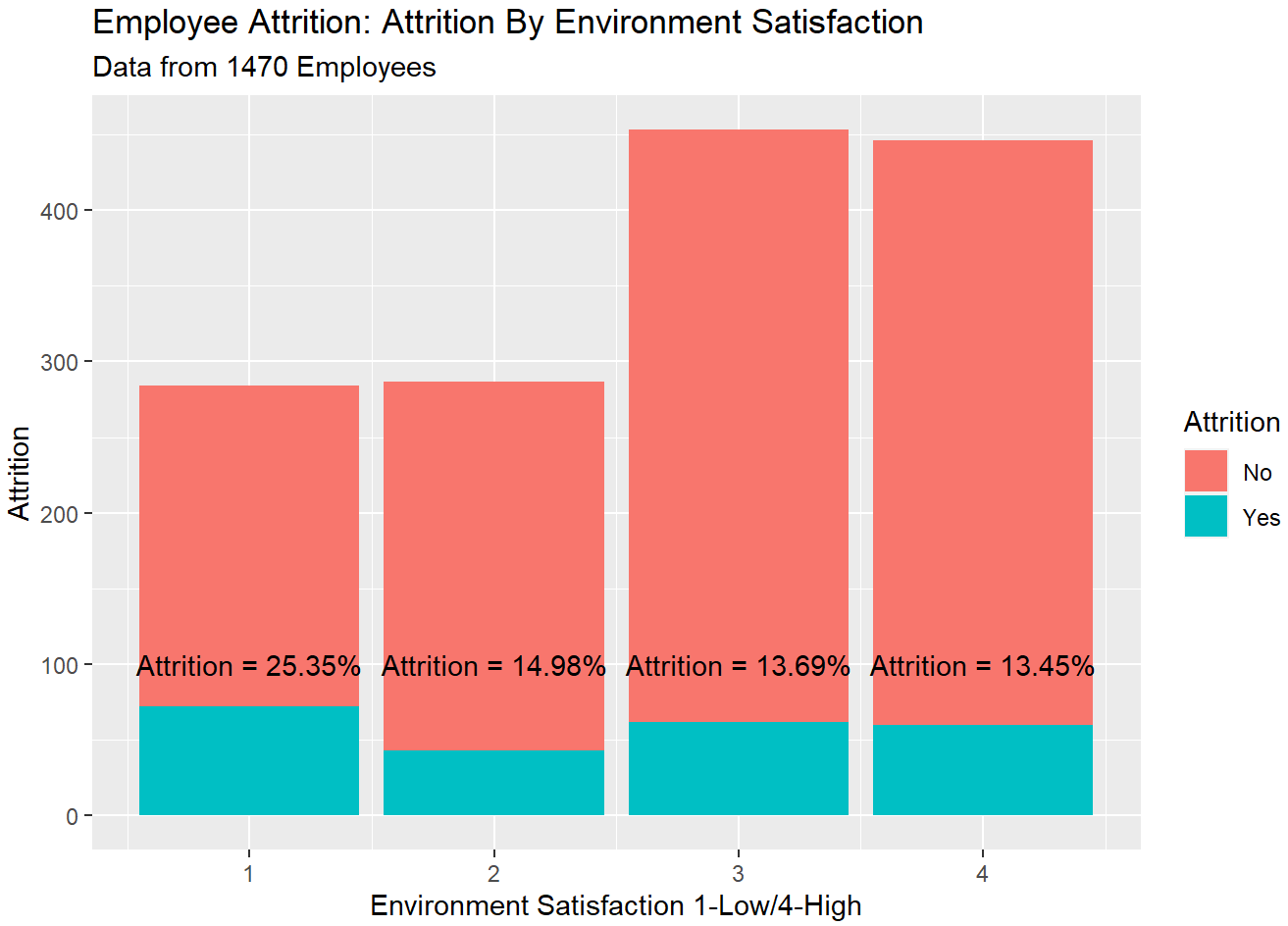
```
library(ggplot2)
ggplot(data=HR_Employee_Attrition_dataset) + geom_bar(mapping = aes(x=Department, fill=Attritio
n)) +
  labs(title="Employee Attrition: Attrition Per Department", subtitle="Data from 1470 Employee
s") +
  annotate("text", x="Human Resources", y=200,label="Attrition Rate = 19.5%") +
  annotate("text", x="Research & Development", y=200, label="Attrition Rate = 13.84%") +
  annotate("text", x="Sales", y=200, label="Attrition Rate = 20.63") + ylab("Attrition")
```



Note that attrition occurs most in the Sales Department, closely followed by the Human Resources Department.

Plot 2 (Attrition by Environment Satisfaction)

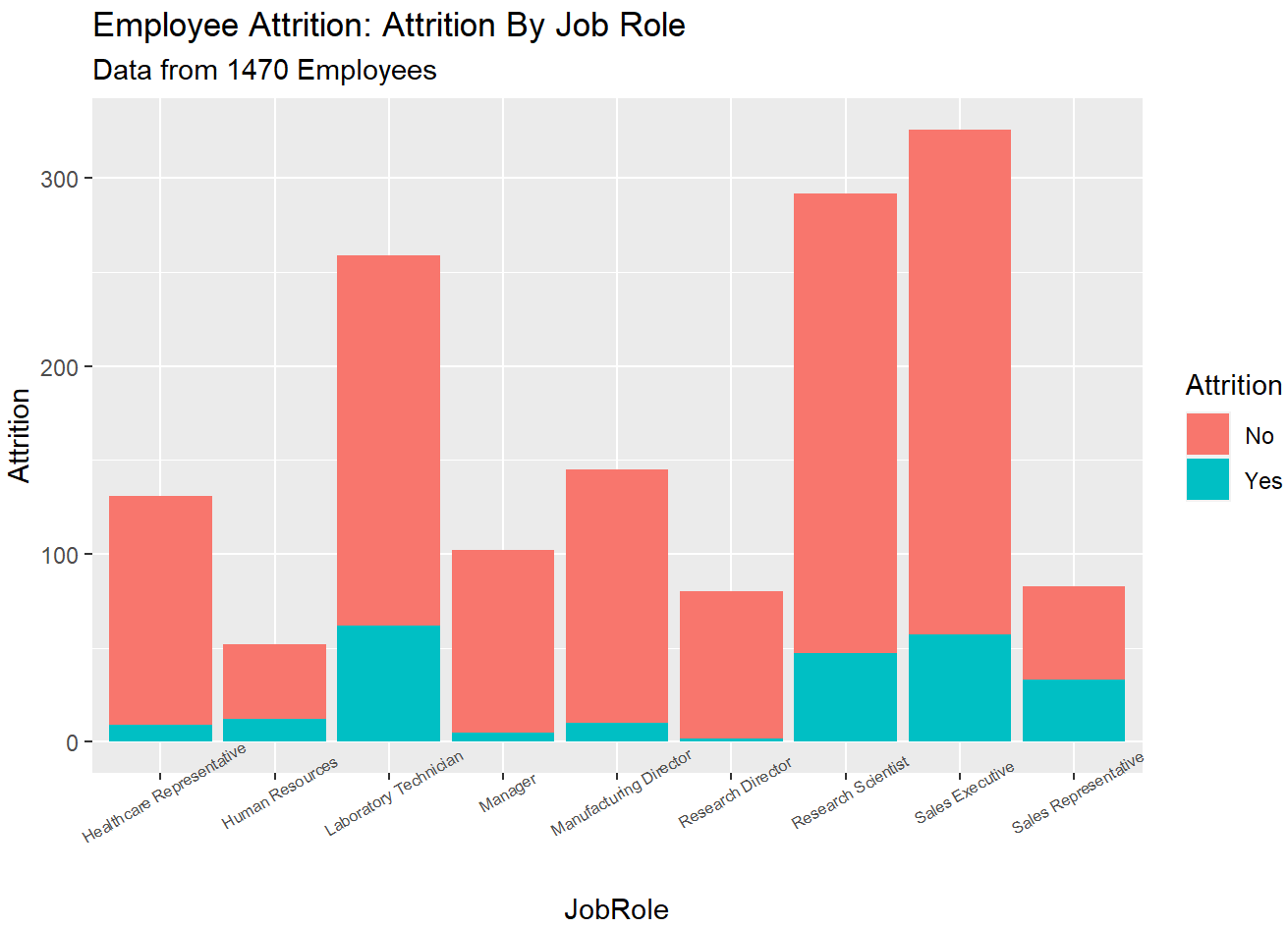
```
ggplot(data=HR_Employee_Attrition_dataset) + geom_bar(mapping=aes(x=EnvironmentSatisfaction, fil
l=Attrition)) +
  labs(title="Employee Attrition: Attrition By Environment Satisfaction", subtitle="Data from 14
70 Employees") +
  annotate("text", x=1, y=100,label="Attrition = 25.35%") +
  annotate("text", x=2, y=100, label="Attrition = 14.98%") +
  annotate("text", x=3, y=100, label="Attrition = 13.69%") +
  annotate("text", x=4, y=100, label="Attrition = 13.45%") +
  ylab("Attrition") + xlab("Environment Satisfaction 1-Low/4-High")
```



Note that unsurprisingly attrition is directly correlated to Environment Satisfaction. Employees least satisfied with their working environment were most likely to leave the company.

Plot 3 (Attrition by Job Role)

```
ggplot(data=HR_Employee_Attrition_dataset) + geom_bar(mapping=aes(x=JobRole, fill=Attrition)) +
  theme(axis.text.x = element_text(size=6, angle=30)) +
  labs(title="Employee Attrition: Attrition By Job Role", subtitle="Data from 1470 Employees") +
  ylab("Attrition")
```



Note that Sales Representatives contributed the most to employee attrition.