EMPLOYEE ATTRITION ANALYSIS OF CHOSEN LTD. USING R

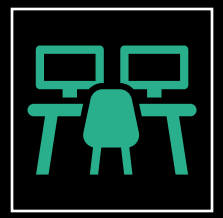
CONTENT

- Employee attrition and its effects
- Business Case
- Introduction to Data Set
- Sample of Data Set
- Key Variables
- Packages used in Business Case
- Importing Data set to R
- Percentage of Employee Attrition
- Visualization of Employee Attrition rate
- Creating a subset of the Data set
- Age group with highest attrition rate
- Dependency of travelling distance on attrition
- Department with highest Attrition rate
- Attrition with respect to Gender
- Job Role grouped by Gender
- Attrition rate with respect to Gender in each department
- Effects of overtime on Employee attrition
- Change in attrition rate with increasing tenure
- Correlation matrix of numeric variables

- Observations from correlation matrix
- Relation between total years at company and current manager
- Relation between percent salary hike and performance rating
- Dependency of monthly income and job level on total working years
- Distribution of monthly income throughout job levels
- Conclusion
- Business case suggestions
- References

EMPLOYEE ATTRITION AND ITS EFFECTS

Attrition is **the number of employees who leave an organization**, including both voluntary and involuntary separation.



Employees may leave their jobs because of ergonomic discomfort.

Another significant reason for attrition is **remuneration and career growth.**

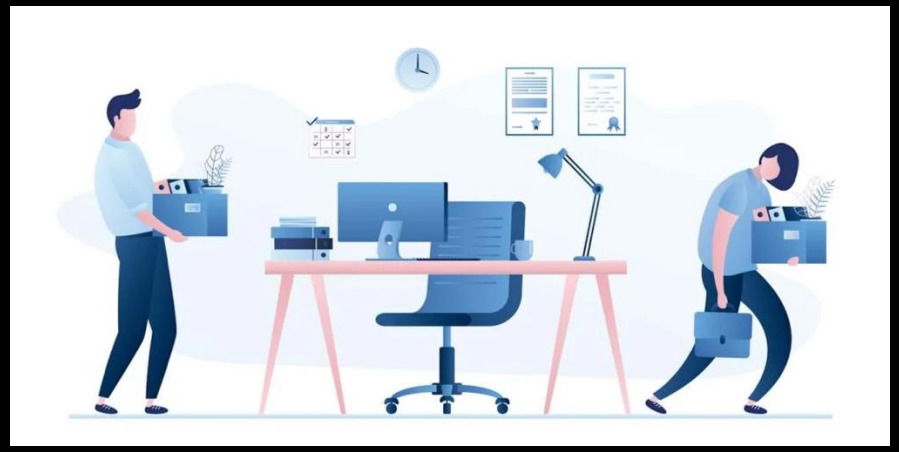
The loss of a skilled employee results in a significant **financial loss** for the firm.

Employee retention is critical for employers since **good, loyal, well-trained, and diligent employees** are necessary to run a successful firm.

BUSINESS CASE OBJECTIVES

- Determine the factors of employee attrition in a company.
- To study the employee attrition rate in the company.
- Identifying the department that has maximum employee attrition.
- Understand attrition in job levels and age.
- To provide remedial measures and suggestions to retain the employees in the organization.

INTRODUCTION TO DATA SET



- This data collection is from the website Kaggle, which has a tremendous quantity of open-source data sets on which we can work.
- The link to the data set is https://www.kaggle.com/datasets/patelprashant/employee-attrition
- The data set is spread across 1470 entries and 19 variables

SAMPLE OF THE DATA SET

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Age	Attritio	Department	DistanceFromHom	EnvironmentSatisfaction	Gender	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MonthlyIncome	OverTime	PercentSalaryHike	PerformanceRating	TotalWorkingYears	WorkLifeBalance	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager
2	18	No	Research & Development	10	4	Female	2	1	Laboratory Technician	3	1200	No	12	3	0	3	0	0	0
3	18	Yes	Research & Development	15	2	Male	3	1	Laboratory Technician	2	1878	Yes	14	3	0	3	0	0	0
4	18	Yes	Sales	13	3	Male	3	1	Sales Representative	3	1420	No	13	3	0	3	0	0	0
5	18	Yes	Research & Development	13	2	Female	3	1	Laboratory Technician	4	1569	Yes	12	3	0	4	0	0	0
6	18	No	Sales	5	2	Male	3	1	Sales Representative	4	1051	No	15	3	0	3	0	0	0
7	18	Yes	Sales	18	3	Male	3	1	Sales Representative	3	1904	No	12	3	0	3	0	0	0
8	18	No	Sales	1	4	Female	3	1	Sales Representative	4	1611	No	15	3	0	4	0	0	0
9	18	No	Sales	14	2	Female	3	1	Sales Representative	3	1514	No	16	3	0	1	0	0	0
10	19	Yes	Research & Development	22	4	Male	3	1	Laboratory Technician	3	1675	Yes	19	3	0	2	0	0	0
11	19	Yes	Research & Development	11	3	Female	1	1	Laboratory Technician	1	2325	No	21	4	1	4	0	0	0
12	19	No	Sales	3	2	Female	3	1	Sales Representative	2	1483	No	14	3	1	3	1	0	0
13	19	Yes	Sales	12	2	Male	2	1	Sales Representative	4	1102	No	22	4	1	2	1	1	0
14	19	Yes	Human Resources	12	1	Male	2	1	Talent Acquisition	4	2564	No	12	3	1	4	1	0	0
15	19	Yes	Research & Development	21	4	Male	2	1	Laboratory Technician	2	2121	Yes	13	3	1	4	1	0	0
16	19	No	Sales	9	3	Male	3	1	Sales Representative	1	2552	No	25	4	1	3	1	0	0
17	19	Yes	Sales	10	1	Female	2	1	Sales Representative	2	1859	Yes	25	4	1	4	1	0	0
18	19	No	Sales	25	2	Female	4	1	Sales Representative	4	2994	Yes	12	3	1	3	1	0	1
19	20	Yes	Sales	16	4	Female	2	1	Sales Representative	4	2926	Yes	18	3	1	3	1	1	0
20	20	Yes	Research & Development	12	3	Female	2	1	Laboratory Technician	3	2044	No	13	3	2	2	2	0	2
21	20	Yes	Research & Development	19	4	Female	3	1	Laboratory Technician	4	2323	Yes	14	3	2	3	2	0	2
22	20	No	Research & Development	21	3	Male	4	1	Laboratory Technician	4	2678	No	17	3	2	3	2	2	2
23	20	No	Sales	1	4	Female	2	1	Sales Representative	2	2836	No	13	3	1	4	1	0	0
24	20	Yes	Sales	10	4	Male	3	1	Sales Representative	3	1009	Yes	11	3	1	3	1	1	1
25	20	No	Research & Development	2	3	Female	3	1	Laboratory Technician	3	2783	No	19	3	2	3	2	2	2
26	20	No	Research & Development	9	4	Male	3	1	Laboratory Technician	1	2728	No	11	3	2	3	2	0	2
27	20	Yes	Sales	14	1	Male	3	1	Sales Representative	1	2973	No	19	3	1	3	1	0	0
28	20	Yes	Sales	11	4	Female	2	1	Sales Representative	1	2600	Yes	15	3	1	3	1	0	0
29	20	No	Sales	3	1	Male	2	1	Sales Representative	3	3033	No	12	3	2	2	2	1	2
30	21	No	Sales	15	3	Male	3	1	Sales Representative	4	1232	No	14	3	0	3	0	0	0
31	21	Yes	Research & Development	11	1	Female	2	1	Laboratory Technician	2	2174	Yes	11	3	3	3	3	1	2
32	21	No	Research & Development	9	1	Male	3	1	Laboratory Technician	4	2610	No	24	4	3	2	3	2	2
33	21	Yes	Research & Development	12	3	Female	4	1	Laboratory Technician	2	2716	No	15	3	1	3	1	0	0
34	21	No	Research & Development	22	3	Male	3	1	Laboratory Technician	3	3447	No	11	3	3	3	3	1	2
35	21	No	Sales	3	4	Male	2	1	Sales Representative	3	3230	No	17	3	3	4	3	1	0
36	21	Yes	Research & Development	17	2	Male	3	1	Laboratory Technician	2	2679	No	13	3	1	3	1	1	0
37	21	Yes	Sales	18	4	Female	3	1	Sales Representative	4	2693	No	19	3	1	2	1	0	0
38	21	No	Research & Development	5	3	Male	3	1	Laboratory Technician	1	2380	Yes	11	3	2	3	2	1	2
39	21	Yes	Sales	10	3	Female	2	1	Sales Representative	1	1416	No	13	3	1	2	1	1	0
40	21	No	Sales	1	4	Female	2	1	Sales Representative	2	2070	Yes	11	3	2	4	2	2	2
41	21	Yes	Sales	10	1	Female	2	1	Sales Representative	3	2625	No	20	4	2	1	2	2	2
42	21	No	Sales	5	3	Male	3	1	Sales Representative	4	3117	No	18	3	3	3	2	2	2
43	22	No	Sales	16	4	Male	4	1	Sales Representative	4	2935	Yes	13	3	1	2	1	0	0
44	22	No	Sales	15	2	Female	3	1	Sales Representative	4	2871	No	15	3	1	3	0	0	0
45	22	No	Sales	2	3	Male	3	1	Sales Representative	4	2523	No	14	3	3	3	2	2	1
46	22	No	Sales	19	3	Male	3	1	Sales Representative	4	2323	No	24	4	2	3	2	2	2
47	22	No	Sales	5	4	Male	4	1	Sales Representative	2	2328	Yes	16	3	4	2	4	2	2
48	22	No	Sales	11	1	Female	3	1	Sales Representative	2	2244	No	13	3	2	3	2	1	2
49	22	Yes	Sales	14	3	Male	2	1	Sales Representative	3	3894	No	16	3	4	3	2	1	2
50	22	No	Sales	6	1	Male	3	1	Sales Representative	3	2773	No	20	4	3	3	2	2	2
51	22	No	Sales	8	2	Male	1	1	Sales Representative	1	2451	No	15	3	4	2	4	1	1
52	22	Yes	Sales	13	3	Male	2	1	Sales Representative	4	2853	Yes	11	3	1	3	0	0	0

KEY VARIABLES

VARIABLE	VARIABLE TYPE	DESCRIPTION
Age	Numerical	Age of the employee between 18 and 60
Department	Categorical	Department of employment
Job Role	Categorical	Role of the employee in the organization
Job Level	Categorical	Job level of the employees from 1 to 5
Environment Satisfaction	Categorical	1 - 'Low', 2 - 'Medium', 3 - 'High', 4 - 'Very High'
Job Satisfaction	Categorical	1 - 'Low', 2 - 'Medium', 3 - 'High', 4 - 'Very High'
Performance Rating	Categorical	1 - 'Low', 2 - 'Medium', 3 - 'High', 4 - 'Very High'
Monthly Income	Numerical	The gross monthly income of the employee
Year since last promotion	Numerical	Year since the employee has received a promotion
Percent Salary Hike	Numerical	The percentage salary hike of the employee in a financial year.
Work Life Balance	Categorical	1 - 'Bad', 2 - 'Good', 3 - 'Better', 4 - 'Best'
Total years at company	Numerical	The number of years an employee has been in the organization.

PACKAGES USED IN THE BUSINESS CASE

#Function to install the package

install.packages("<packagename>")

#Function to call the library

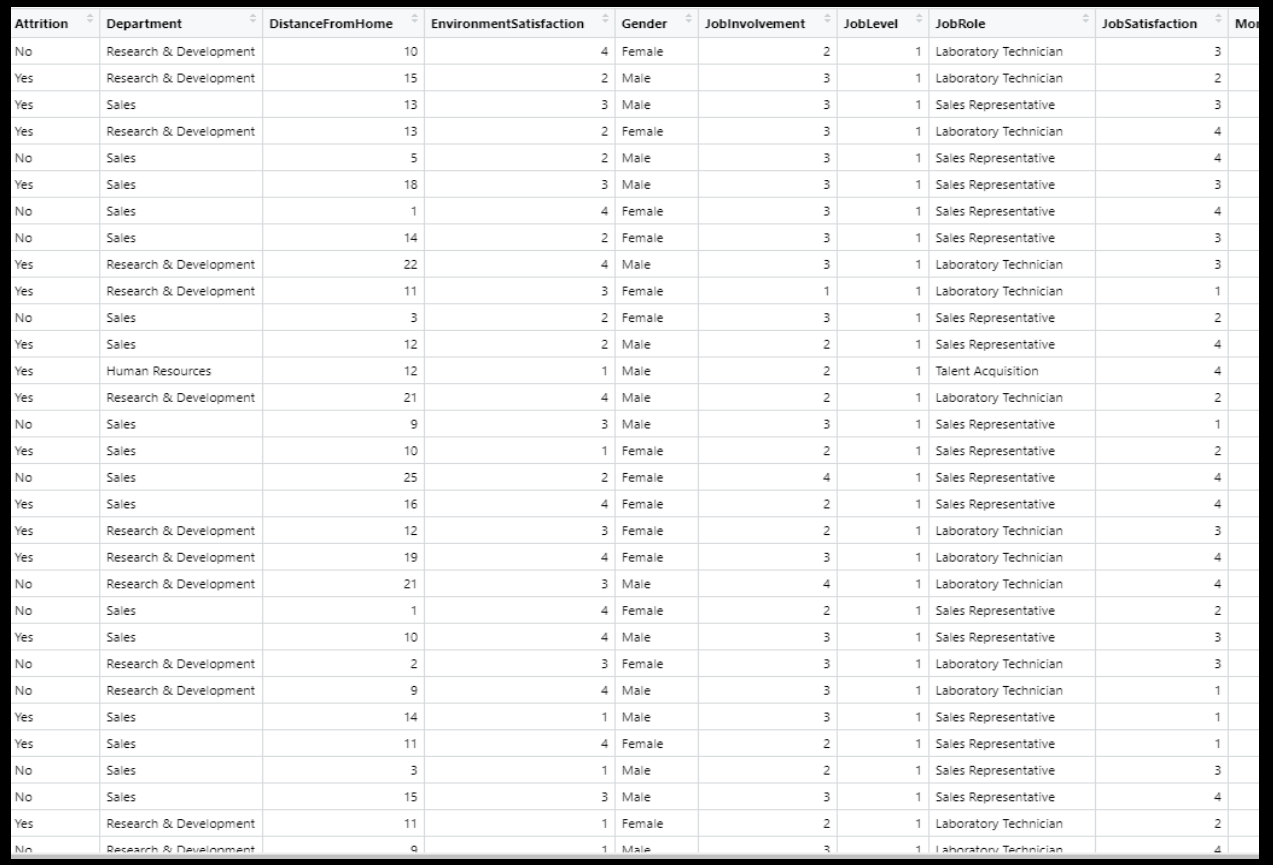
library(<libraryname>)

Packages used:

- readr
- tidyverse
- ggplot2
- ggcorrplot

IMPORTING DATA SET TO R

```
#installing the readr package
install.packages("readr")

#calling the readr library
library(readr)

#importing .csv file to R
attrition <- read_csv("attrition_original.csv")

#viewing data set in R
View(attrition)
```



Attrition	Department	DistanceFromHome	EnvironmentSatisfaction	Gender	JobInvolvement	JobLevel	JobRole	JobSatisfaction	Mo
No	Research & Development	10	4	Female	2	1	Laboratory Technician	3	
Yes	Research & Development	15	2	Male	3	1	Laboratory Technician	2	
Yes	Sales	13	3	Male	3	1	Sales Representative	3	
Yes	Research & Development	13	2	Female	3	1	Laboratory Technician	4	
No	Sales	5	2	Male	3	1	Sales Representative	4	
Yes	Sales	18	3	Male	3	1	Sales Representative	3	
No	Sales	1	4	Female	3	1	Sales Representative	4	
No	Sales	14	2	Female	3	1	Sales Representative	3	
Yes	Research & Development	22	4	Male	3	1	Laboratory Technician	3	
Yes	Research & Development	11	3	Female	1	1	Laboratory Technician	1	
No	Sales	3	2	Female	3	1	Sales Representative	2	
Yes	Sales	12	2	Male	2	1	Sales Representative	4	
Yes	Human Resources	12	1	Male	2	1	Talent Acquisition	4	
Yes	Research & Development	21	4	Male	2	1	Laboratory Technician	2	
No	Sales	9	3	Male	3	1	Sales Representative	1	
Yes	Sales	10	1	Female	2	1	Sales Representative	2	
No	Sales	25	2	Female	4	1	Sales Representative	4	
Yes	Sales	16	4	Female	2	1	Sales Representative	4	
Yes	Research & Development	12	3	Female	2	1	Laboratory Technician	3	
Yes	Research & Development	19	4	Female	3	1	Laboratory Technician	4	
No	Research & Development	21	3	Male	4	1	Laboratory Technician	4	
No	Sales	1	4	Female	2	1	Sales Representative	2	
Yes	Sales	10	4	Male	3	1	Sales Representative	3	
No	Research & Development	2	3	Female	3	1	Laboratory Technician	3	
No	Research & Development	9	4	Male	3	1	Laboratory Technician	1	
Yes	Sales	14	1	Male	3	1	Sales Representative	1	
Yes	Sales	11	4	Female	2	1	Sales Representative	1	
No	Sales	3	1	Male	2	1	Sales Representative	3	
No	Sales	15	3	Male	3	1	Sales Representative	4	
Yes	Research & Development	11	1	Female	2	1	Laboratory Technician	2	
No	Research & Development	9	1	Male	3	1	Laboratory Technician	4	

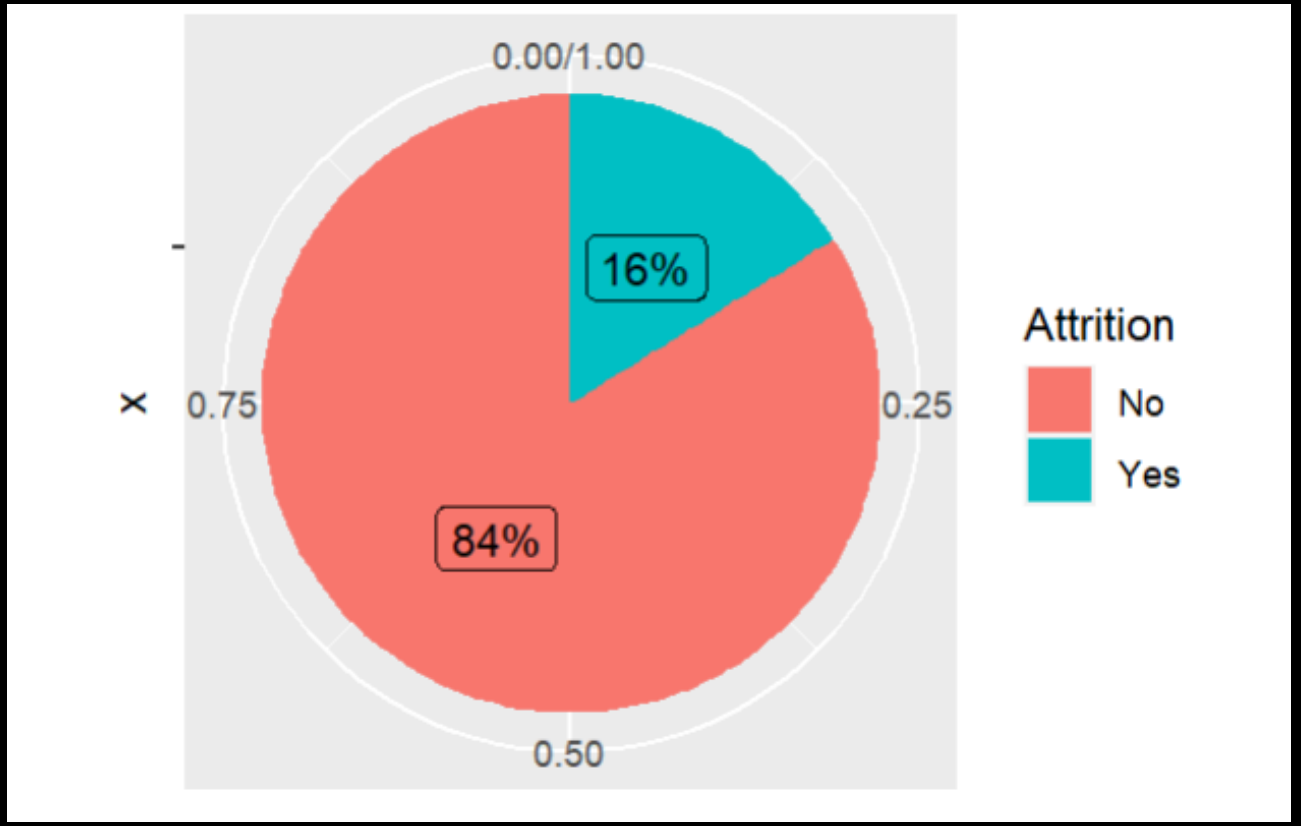
PERCENTAGE OF EMPLOYEE ATTRITION

	Attrition	n	perc	labels
1	Yes	237	0.1612245	16%
2	No	1233	0.8387755	84%

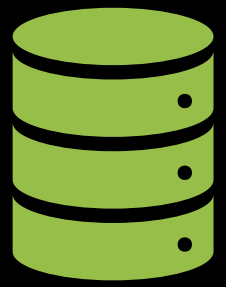
```
install.packages("tidyverse")
library(tidyverse)

attrition_rate <- attrition %>%
  group_by(Attrition) %>% # grouping Variable to be transformed
count() %>%  ungroup() %>%  mutate(perc = `n` / sum(`n`)) %>%  #calculating the
percentages
  arrange(perc) %>%  mutate(labels = scales::percent(perc))

View(attrition_rate)
```

VISUALUZATION OF EMPLOYEE ATTRITION RATE

```
ggplot(attrition_rate, aes(x = "", y = perc, fill =
Attrition)) +
  geom_col() +
  geom_label(aes(label = labels),
             position = position_stack(vjust = 0.5),
             show.legend = FALSE) +
  coord_polar(theta = "y")
```



CREATING A SUBSET OF THE DATA SET



Creating a subset of the original dataset with the data of the employees who have left the organization.

The subset will help us analyze each variable affecting employee attrition.

```
attrition_yes <- attrition_original %>% filter(Attrition == "Yes")
View(attrition_yes)
```

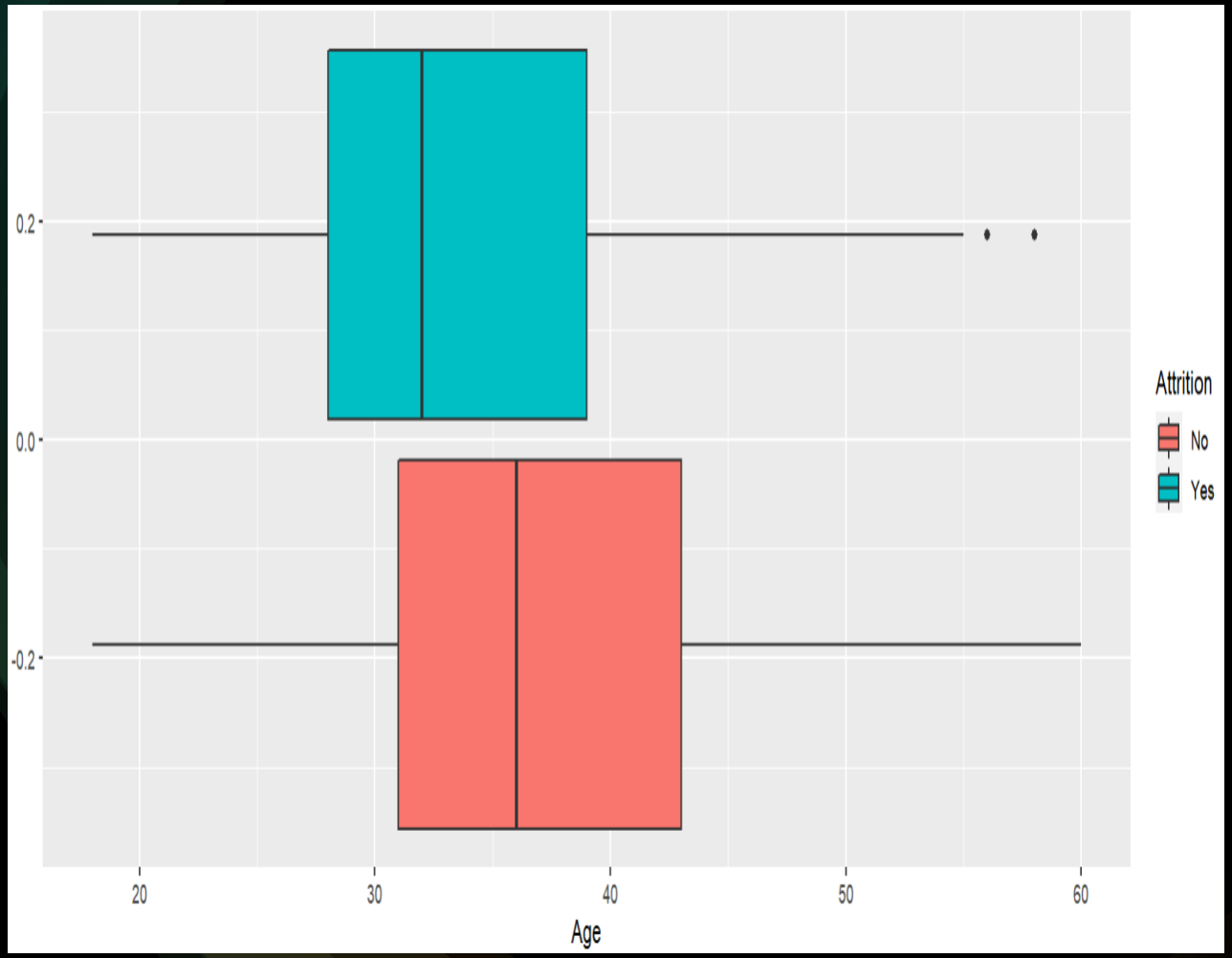
AGE GROUP WITH HIGHEST ATTRITION RATE

#visualization to identify the age group with highest attrition

```
ggplot(attrition,   aes(x = Age,   fill = Attrition)) +
 geom_boxplot()
```

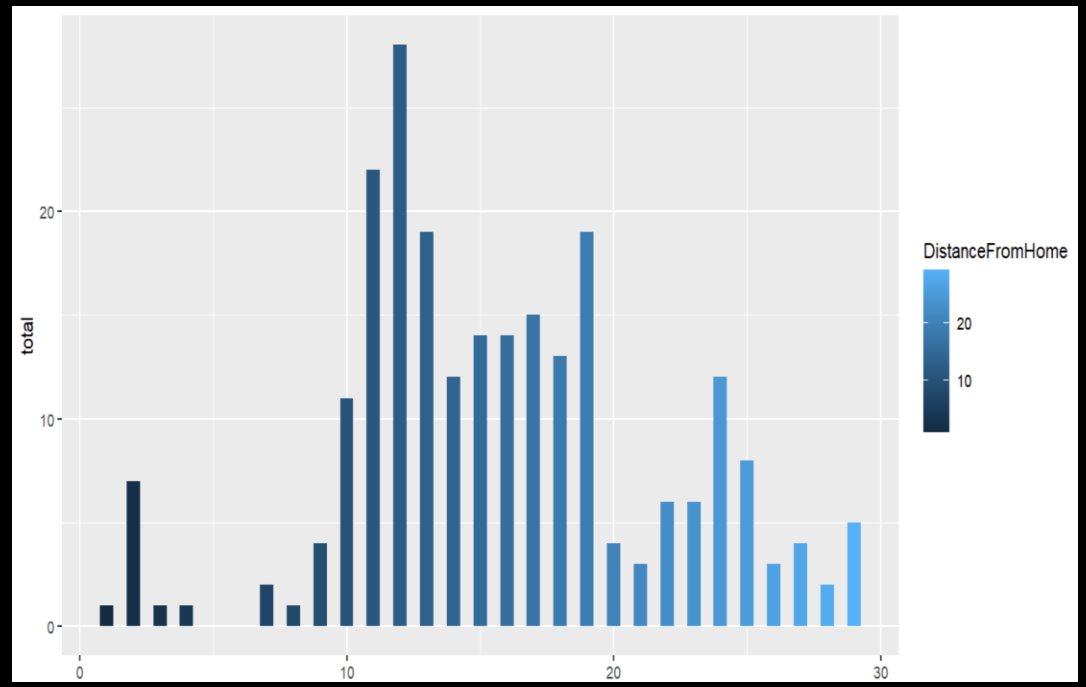
The box plot starts with the 1st quartile and ends at the 3rd quartile and the line in the box plot shows the median.

We can observe that age group with highest attrition is between 25-40.



DEPENDENCY OF TRAVELLING DISTANCE ON ATTRITION

```
att_distance <- attrition_yes %>% select(Department,
DistanceFromHome) %>%
  group_by(DistanceFromHome) %>%  #variable to be
transformed
  summarise(total = n()) %>% mutate(Percent =
paste0(round(100*total / sum(total), 0), "%")) %>%
  arrange(desc(total)) #arranging percentage in decreasing
order
View(att_distance)
```



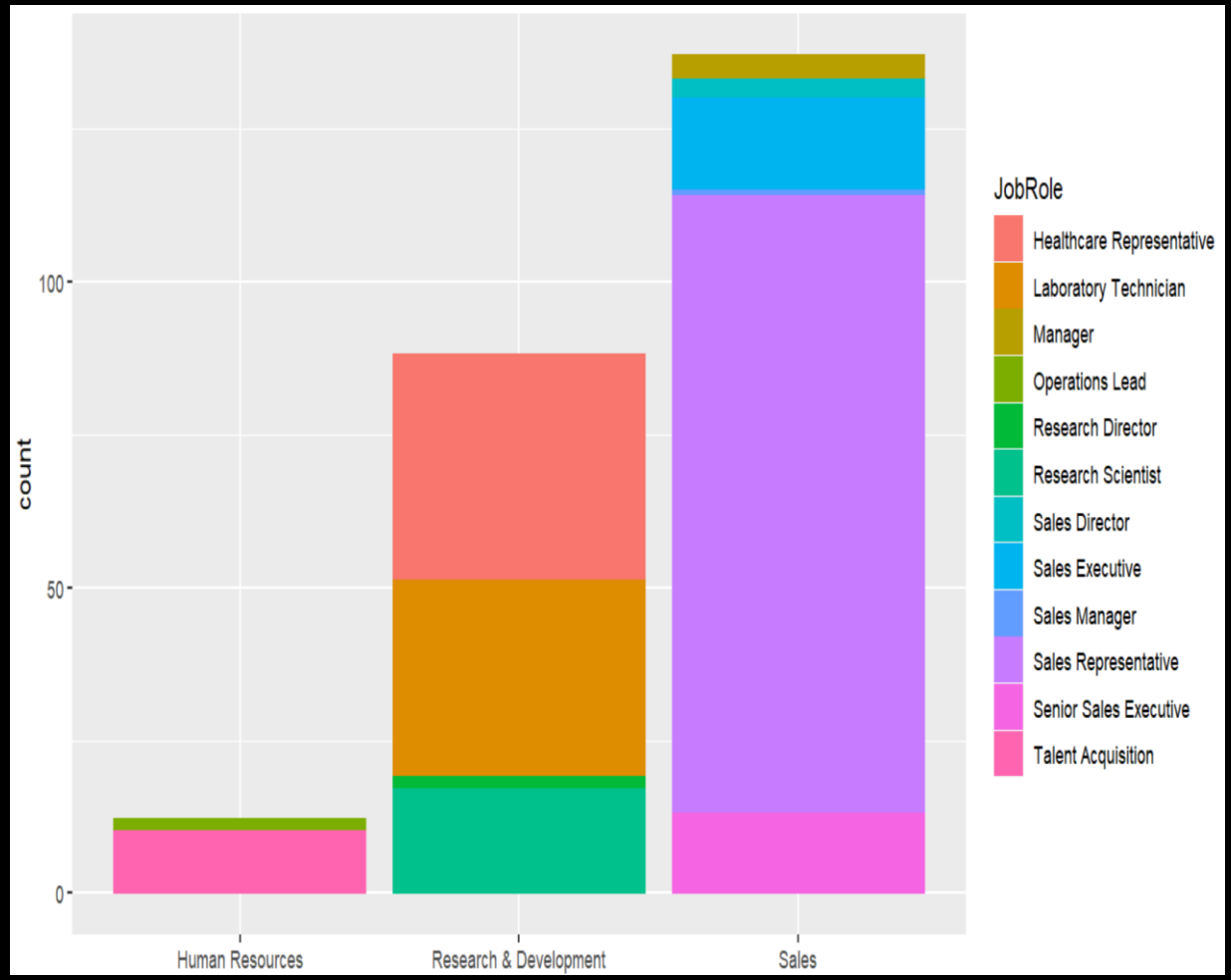

```
# A tibble: 27 × 3
   DistanceFromHome total Percent
              <dbl> <int> <chr>
 1               12    28 12%
 2               11    22 9%
 3               13    19 8%
 4               19    19 8%
 5               17    15 6%
 6               15    14 6%
 7               16    14 6%
 8               18    13 5%
 9               14    12 5%
10               24    12 5%
# … with 17 more rows
```

```
ggplot(att_distance, aes(x = DistanceFromHome , y = total)) +
  geom_col(aes(fill = DistanceFromHome),
           width = 0.5,
           position = position_dodge(width = 0.5))
```

DEPARTMENT WITH HIGHEST ATTRITION RATE

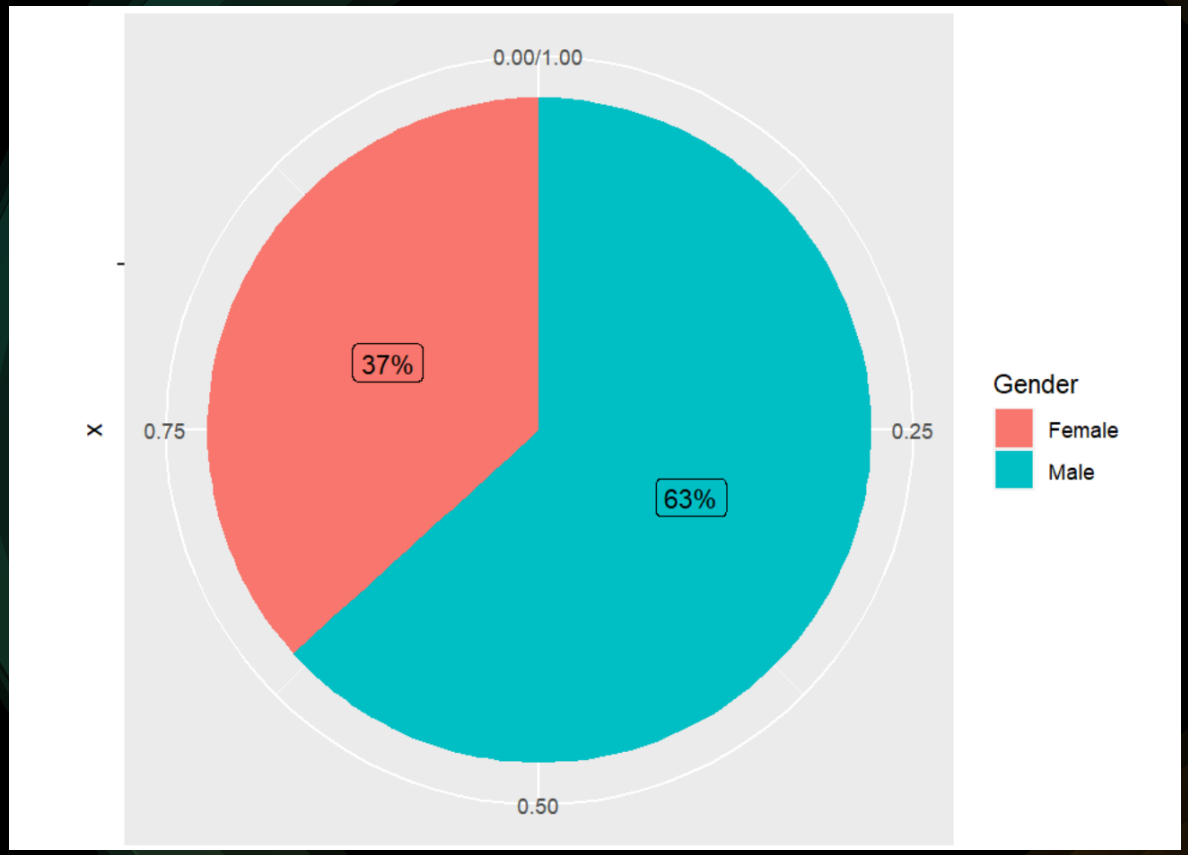
```
#visualization of the department grouped by job roles
ggplot(attrition_yes, aes(Department)) +
 geom_bar(aes(color =  JobRole, fill = JobRole),
position = "stack")
```

Job Role	Job Level
Sales Representative	1
Healthcare Representative	2
Laboratory Technician	1



ATTRITION WITH RESPECT TO GENDER

```
attrition_yes %>% group_by(Gender) %>% count() %>%
ungroup() %>% mutate(perc = `n` / sum(`n`)) %>%
  arrange(perc) %>%
  mutate(labels = scales::percent(perc)) %>%
  ggplot(aes(x = "", y = perc, fill = Gender)) +
  geom_col() +
  geom_label(aes(label = labels), position =
position_stack(vjust = 0.5), show.legend = FALSE) +
  coord_polar(theta = "y")
```



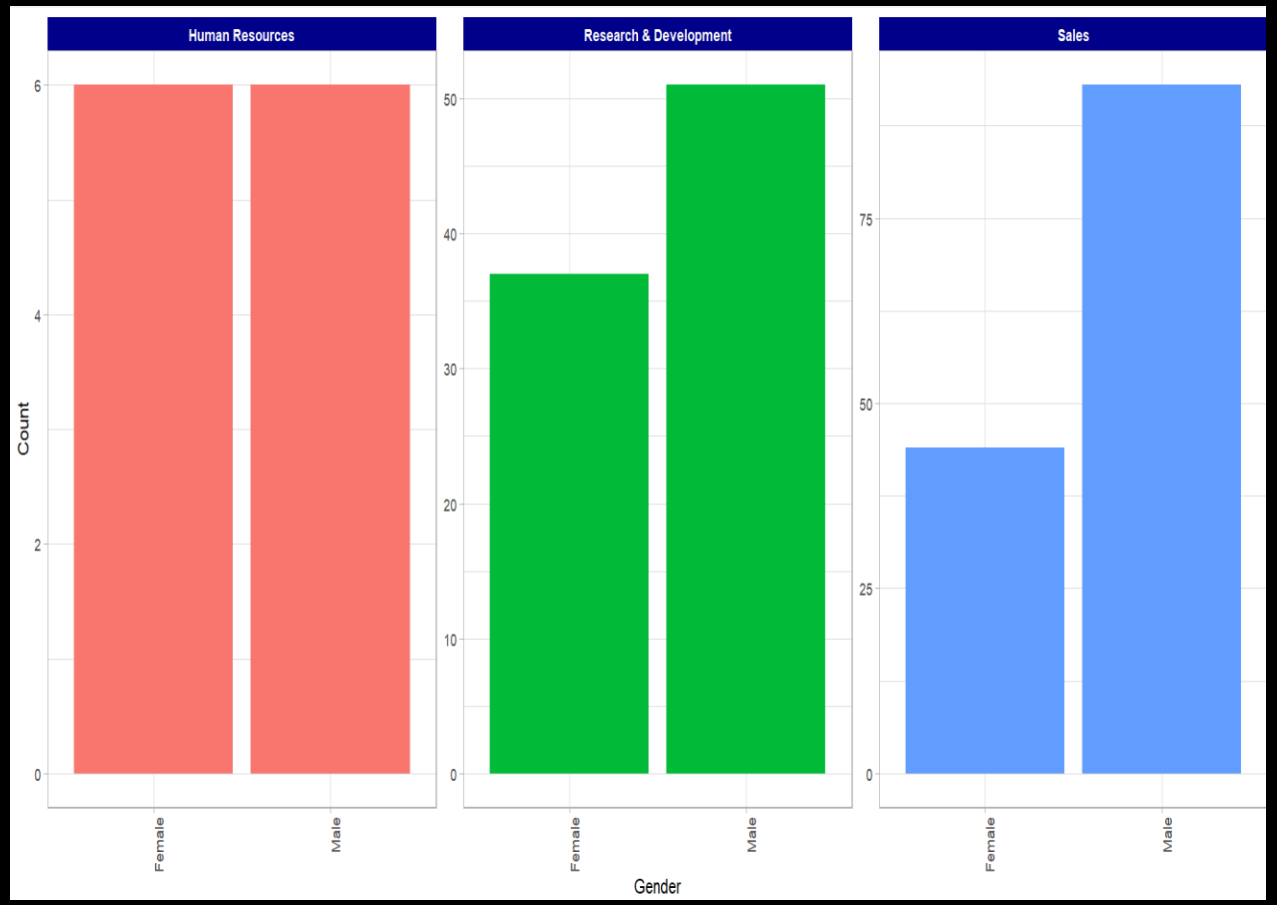
JOB ROLE GROUPED BY GENDER

```
attrition_yes %>% select(Gender, JobRole, Department) %>%
  group_by(Gender, JobRole, Department) %>%
  summarise(Count = n(), .groups = "drop") %>%
  ungroup() %>%
  arrange(desc(Count))
```

```
# A tibble: 22 × 4
   Gender JobRole                  Department              Count
   <chr>  <chr>                    <chr>                   <int>
 1 Male   Sales Representative     Sales                      69
 2 Female Sales Representative     Sales                      32
 3 Male   Healthcare Representative Research & Development    23
 4 Female Laboratory Technician    Research & Development     16
 5 Male   Laboratory Technician    Research & Development     16
 6 Female Healthcare Representative Research & Development   14
 7 Male   Research Scientist       Research & Development     12
 8 Male   Sales Executive          Sales                      11
 9 Male   Senior Sales Executive   Sales                       7
10 Female Senior Sales Executive   Sales                       6
# … with 12 more rows
```

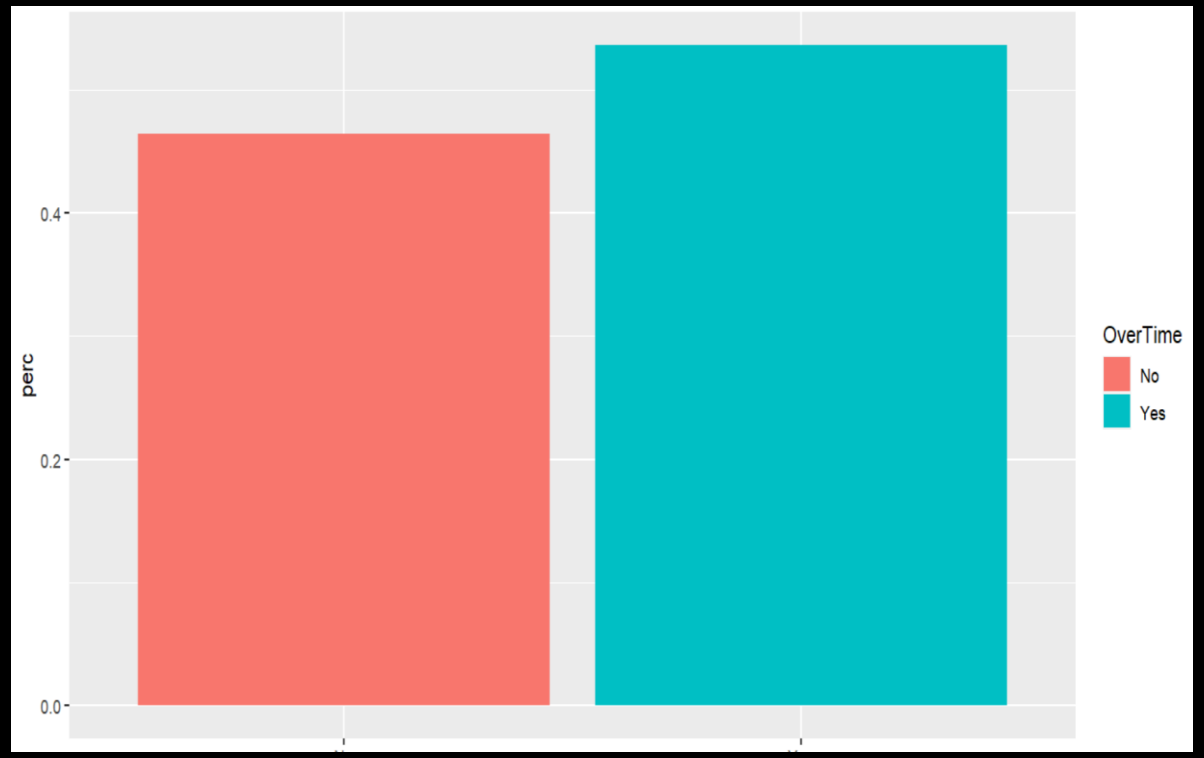
ATTRITION WITH RESPECT TO GENDER IN EACH DEPARTMENT

```
attrition_yes %>% select(Gender, JobRole, Department) %>%
  group_by(Gender, JobRole, Department) %>%
  summarise(Count = n(), .groups = "drop") %>%
  ungroup() %>%
  arrange(desc(Count)) %>%
  ggplot(aes(x= Gender, y = Count, fill = Department)) +
  geom_col() +
  theme_light() +
  facet_wrap(~ Department, scales = "free") +
  theme(strip.background = element_rect(fill = "darkblue")) +
  theme(strip.text = element_text(colour = "white", face = "bold")) +
  theme(legend.position = "none") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



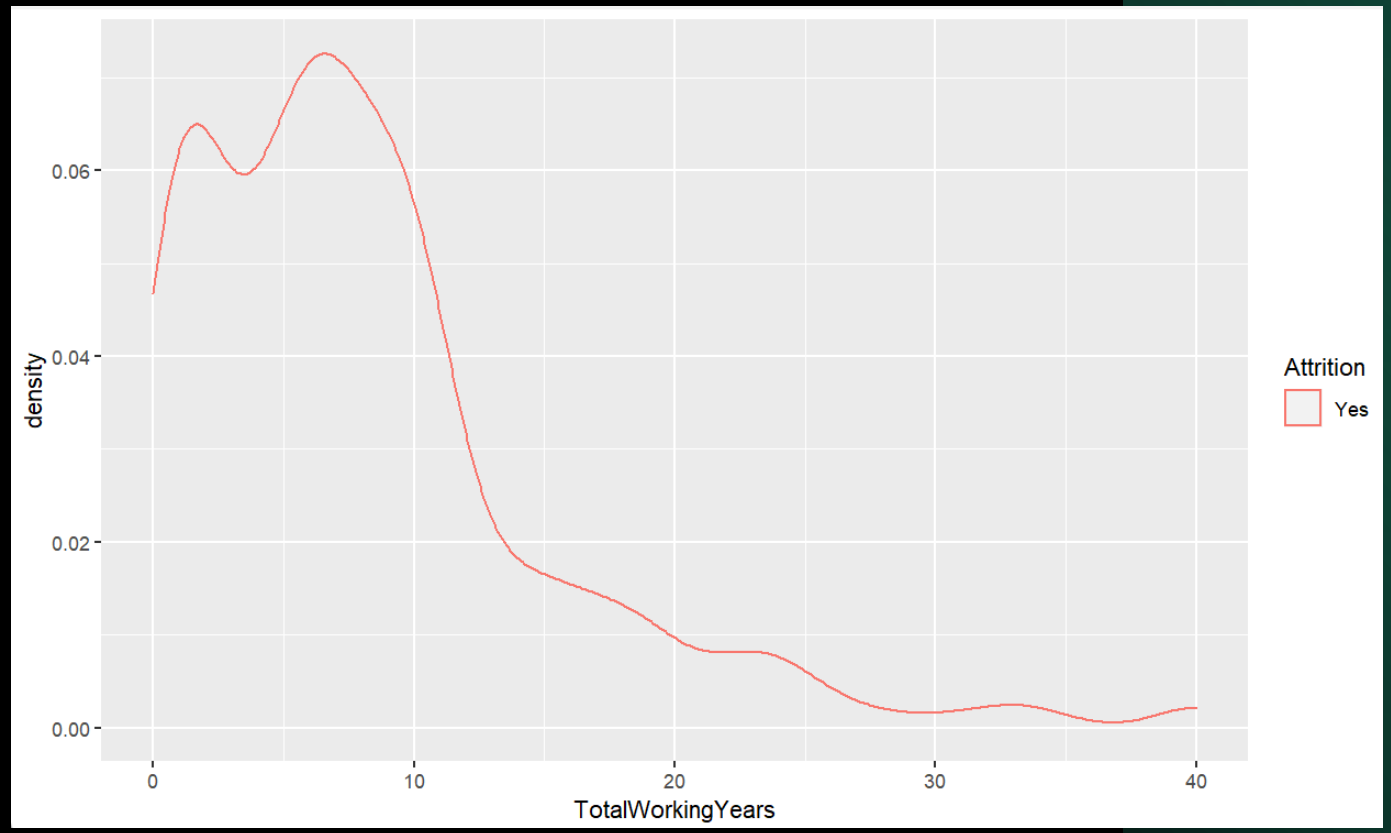
EFFECTS OF OVERTIME ON EMPLOYEE ATTRITION

```
#examining the number of employees who clocked extra hours
attrition_yes %>%
  group_by(OverTime) %>% # Variable to be transformed
  count() %>%
  ungroup() %>%
  mutate(perc = `n` / sum(`n`)) %>%
  arrange(perc) %>%
  mutate(labels = scales::percent(perc)) %>%
  ggplot(aes(x = OverTime, y = perc, fill = OverTime)) +
  geom_col()
```



OverTime	n	perc	labels
<chr>	<int>	<dbl>	<chr>
1 No	110	0.464	46.4%
2 Yes	127	0.536	53.6%

CHANGE IN ATTRITION RATE WITH INCREASING TENURE



#density plot to show how attrition reduces with tenure

ggplot(attrition_yes, aes(x=TotalWorkingYears, col=Attrition)) + geom_density()

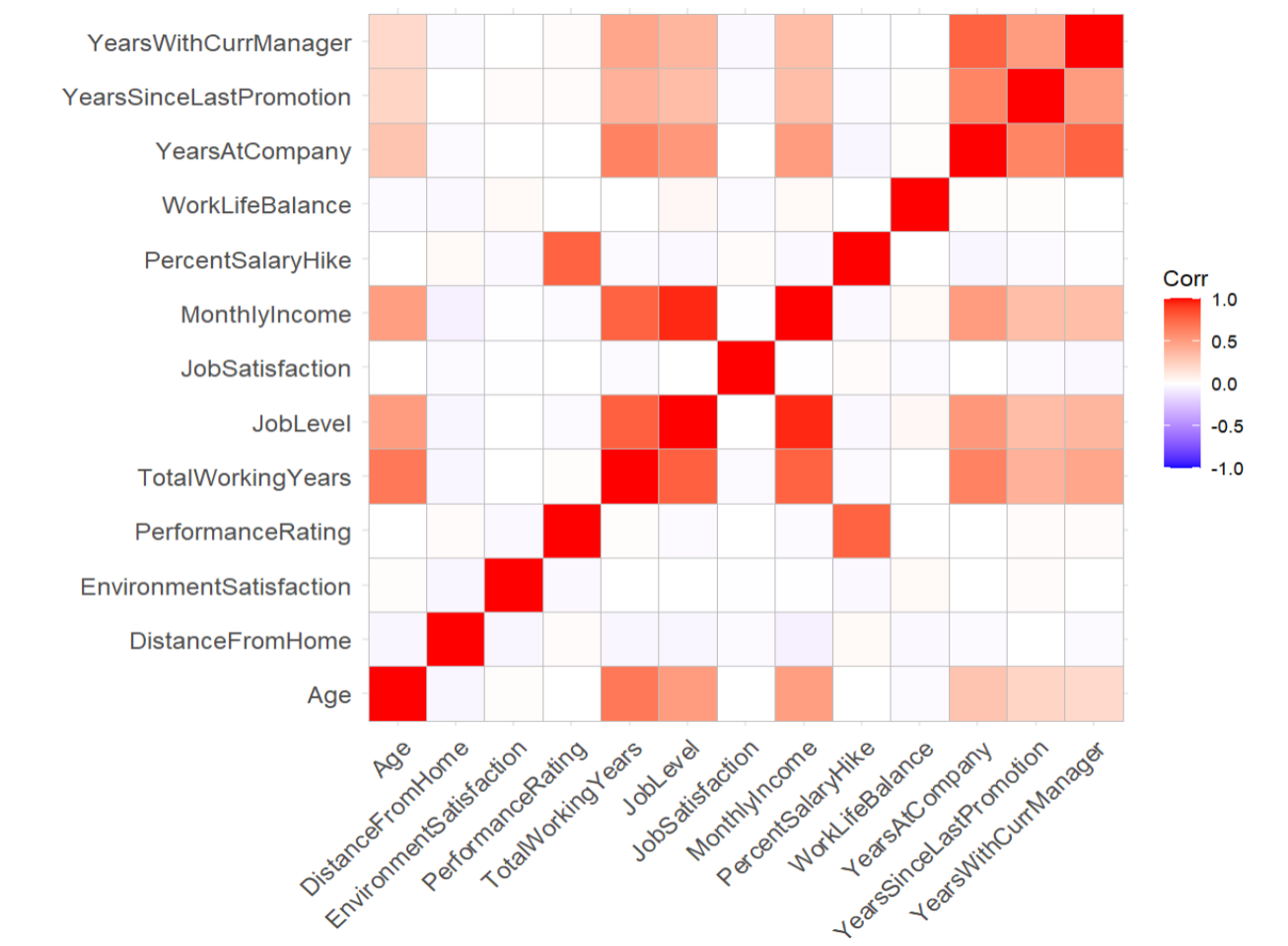
We can see from the density plot that as work experience increases people think less about leaving the company.

CORRELATION MATRIX OF NUMERICAL VARIABLES

```
#installing ggcorrplot package
install.packages("ggcorrplot")
library(ggcorrplot)
library(ggplot2)
#creating a correlation matrix of all numeric variables
df <- attrition[, c("Age", "DistanceFromHome", "EnvironmentSatisfaction", "PerformanceRating", "TotalWorkingYears",
"JobLevel", "JobSatisfaction", "MonthlyIncome", "PercentSalaryHike", "WorkLifeBalance", "YearsAtCompany",
"YearsSinceLastPromotion", "YearsWithCurrManager")]
correlation_matrix <- cor(df)
View(correlation_matrix)
```

	Age	DistanceFromHome	EnvironmentSatisfaction	PerformanceRating	TotalWorkingYears	JobLevel	JobSatisfaction	MonthlyIncome	PercentSalaryHike	WorkLifeBalance	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager
Age	1.000000000	-0.044429697	0.010146428	0.001903896	0.680380536	0.509604228	-0.004891877	0.497854567	0.003633585	-0.021490028	0.311308770	0.216513368	0.202088602
DistanceFromHome	-0.044429697	1.000000000	-0.044926757	0.017806278	-0.037982891	-0.042881421	-0.022503932	-0.060348054	0.027148588	-0.029025116	-0.019468324	-0.001847878	-0.019926544
EnvironmentSatisfaction	0.010146428	-0.044926757	1.000000000	-0.029547952	-0.002693070	0.001211699	-0.006784353	-0.006259088	-0.031701195	0.027627295	0.001457549	0.016193606	-0.004998723
PerformanceRating	0.001903896	0.017806278	-0.029547952	1.000000000	0.006743668	-0.021222082	0.002297197	-0.017120138	0.773549996	0.002572361	0.003435126	0.017896066	0.022827169
TotalWorkingYears	0.680380536	-0.037982891	-0.002693070	0.006743668	1.000000000	0.782207805	-0.020185073	0.772893246	-0.020608488	0.001007646	0.628133155	0.404857759	0.459188397
JobLevel	0.509604228	-0.042881421	0.001211699	-0.021222082	0.782207805	1.000000000	-0.001943708	0.950299913	-0.034730492	0.037817746	0.534738687	0.353885347	0.375280608
JobSatisfaction	-0.004891877	-0.022503932	-0.006784353	0.002297197	-0.020185073	-0.001943708	1.000000000	-0.007156742	0.020002039	-0.019458710	-0.003802628	-0.018213568	-0.027656214
MonthlyIncome	0.497854567	-0.060348054	-0.006259088	-0.017120138	0.772893246	0.950299913	-0.007156742	1.000000000	-0.027268586	0.030683082	0.514284826	0.344977638	0.344078883
PercentSalaryHike	0.003633585	0.027148588	-0.031701195	0.773549996	-0.020608488	-0.034730492	0.020002039	-0.027268586	1.000000000	-0.003179636	-0.035991262	-0.022154313	-0.011985248
WorkLifeBalance	-0.021490028	-0.029025116	0.027627295	0.002572361	0.001007646	0.037817746	-0.019458710	0.030683082	-0.003179636	1.000000000	0.012089185	0.008941249	0.002759440
YearsAtCompany	0.311308770	-0.019468324	0.001457549	0.003435126	0.628133155	0.534738687	-0.003802628	0.514284826	-0.035991262	0.012089185	1.000000000	0.618408865	0.768212425
YearsSinceLastPromotion	0.216513368	-0.001847878	0.016193606	0.017896066	0.404857759	0.353885347	-0.018213568	0.344977638	-0.022154313	0.008941249	0.618408865	1.000000000	0.510223636
YearsWithCurrManager	0.202088602	-0.019926544	-0.004998723	0.022827169	0.459188397	0.375280608	-0.027656214	0.344078883	-0.011985248	0.002759440	0.768212425	0.510223636	1.000000000

OBSERVATIONS FROM CORRELATION MATRIX



- Strong relationships observed in the correlation matrix:

1. Years with current manager and years at company (0.76).

1. Percent salary hike and performance rating (0.77).

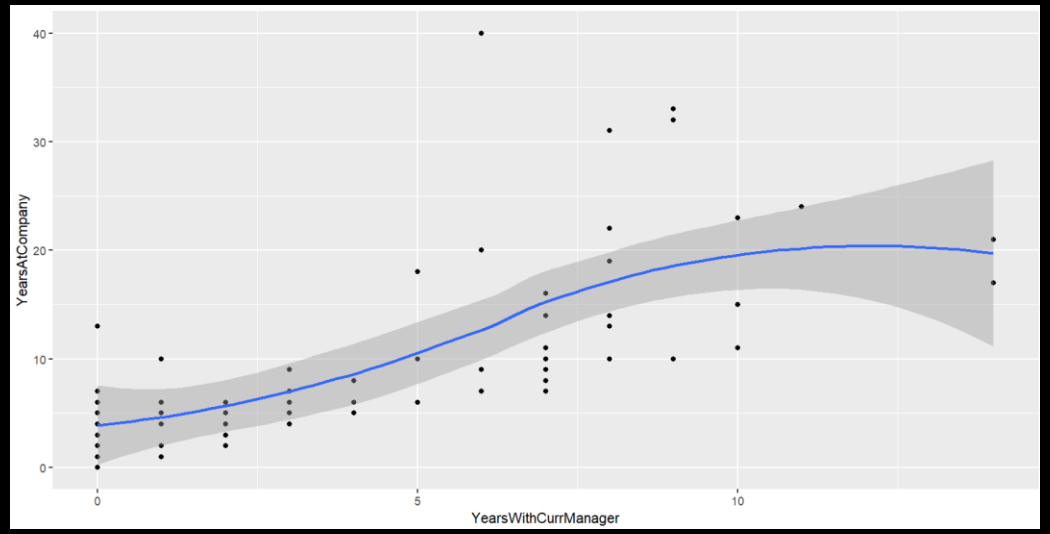
2. Job level and total working years (0.78)

3. Job Level and Monthly Income (0.95)

RELATION BETWEEN THE TOTAL YEARS AT COMPANY AND CURRENT MANAGER

```
#grouping employees by department, number of years at the company and with current manager
attrition_yes %>% select(Department, YearsWithCurrManager, YearsAtCompany) %>%
  group_by(Department, YearsWithCurrManager, YearsAtCompany) %>%
  summarise(Count = n(), .groups = "drop") %>%
  arrange(desc(Count))
```

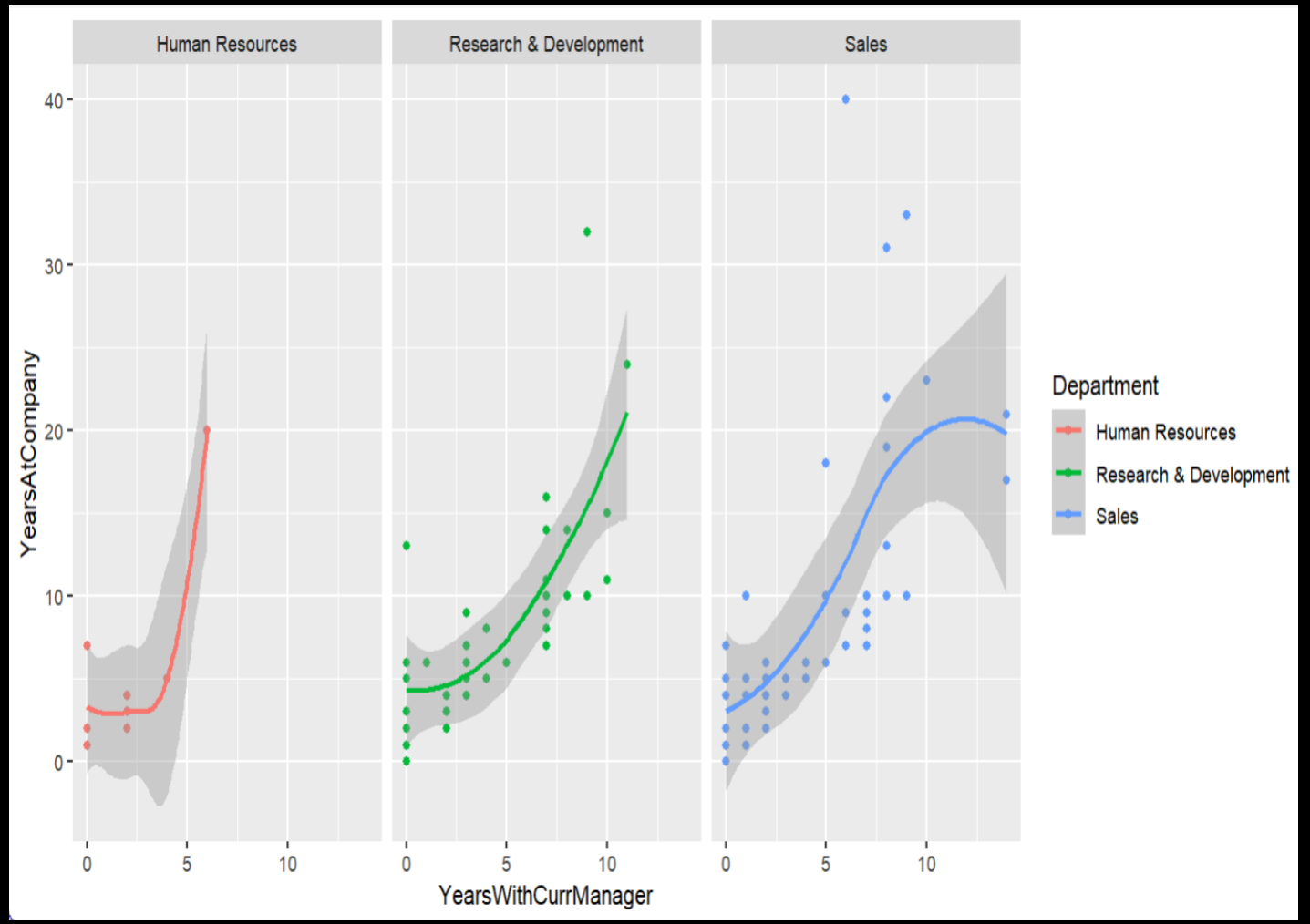
```
# A tibble: 81 × 4
   Department              YearsWithCurrManager YearsAtCompany Count
   <chr>                                  <dbl>          <dbl> <int>
 1 Sales                                      0              1    35
 2 Research & Development                     0              1    16
 3 Research & Development                     2              2    10
 4 Sales                                      2              2    10
 5 Research & Development                     2              3     9
 6 Sales                                      0              0     9
 7 Sales                                      2              3     8
 8 Sales                                      3              4     8
 9 Research & Development                     0              0     7
10 Sales                                      4              5     6
# … with 71 more rows
```




```
#we can see a positive correlation between the two variables
attrition_yes %>% select(YearsWithCurrManager, YearsAtCompany) %>%
  group_by(YearsWithCurrManager, YearsAtCompany) %>%
  summarise(Count = n(), .groups = "drop") %>%
  arrange(desc(Count)) %>%
  ggplot(aes(x= YearsWithCurrManager, y= YearsAtCompany)) + geom_point() +
  geom_smooth(span = 1)
```

```
#visualizing the relationship for each department
attrition_yes %>% select(Department, YearsWithCurrManager, YearsAtCompany) %>%
  group_by(Department, YearsWithCurrManager, YearsAtCompany) %>%
  summarise(Count = n(), .groups = "drop") %>%
  arrange(desc(Count)) %>%
  ggplot(aes(x= YearsWithCurrManager, y= YearsAtCompany , color= Department)) +
geom_point() + facet_grid(~ Department) +
  geom_smooth(span = 1)
```

Based on the plots, we can assume that a positive relationship with the management will aid in employee retention.

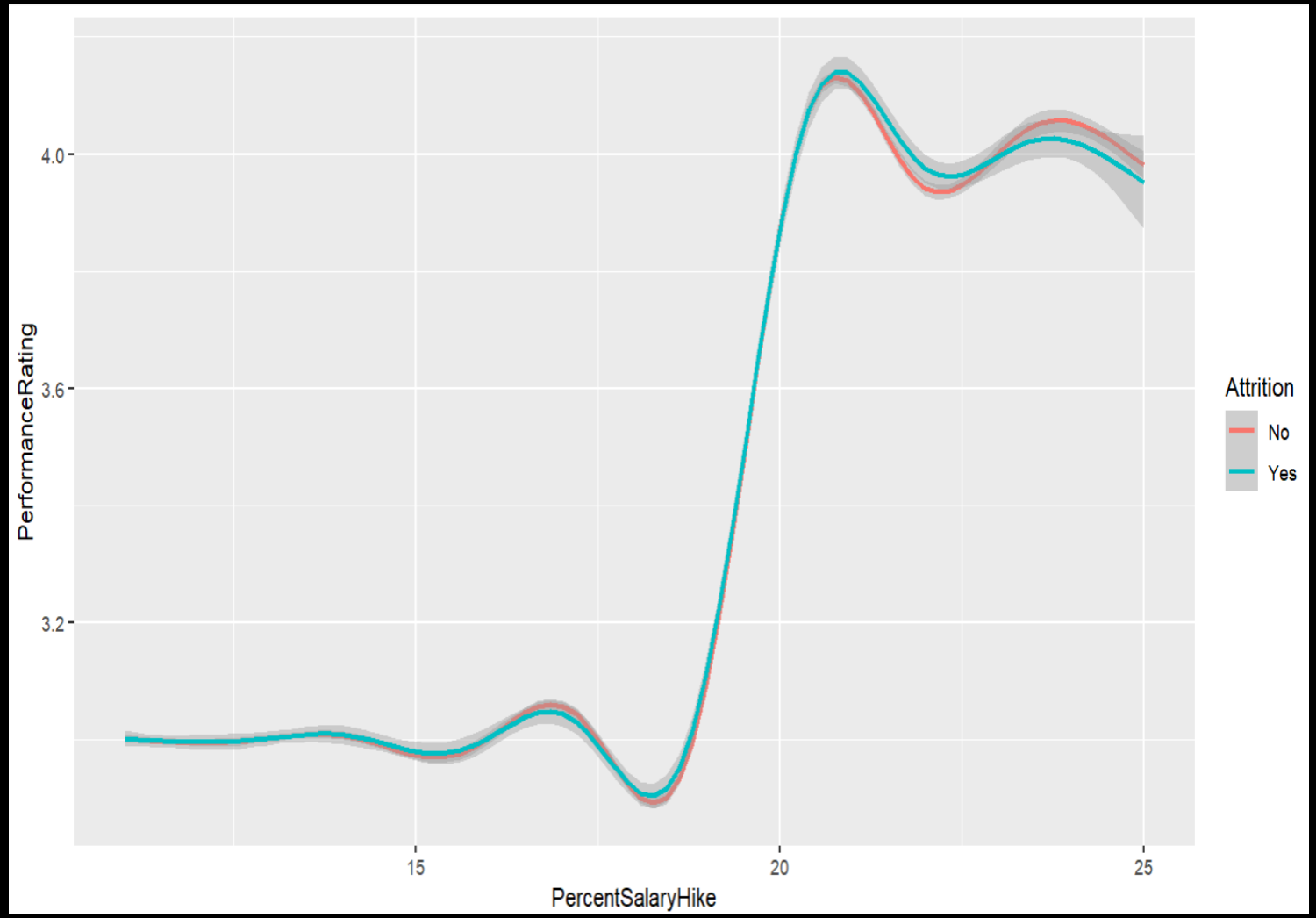


RELATION BETWEEN PERCENT SALARY HIKE AND PERFORMANCE RATING

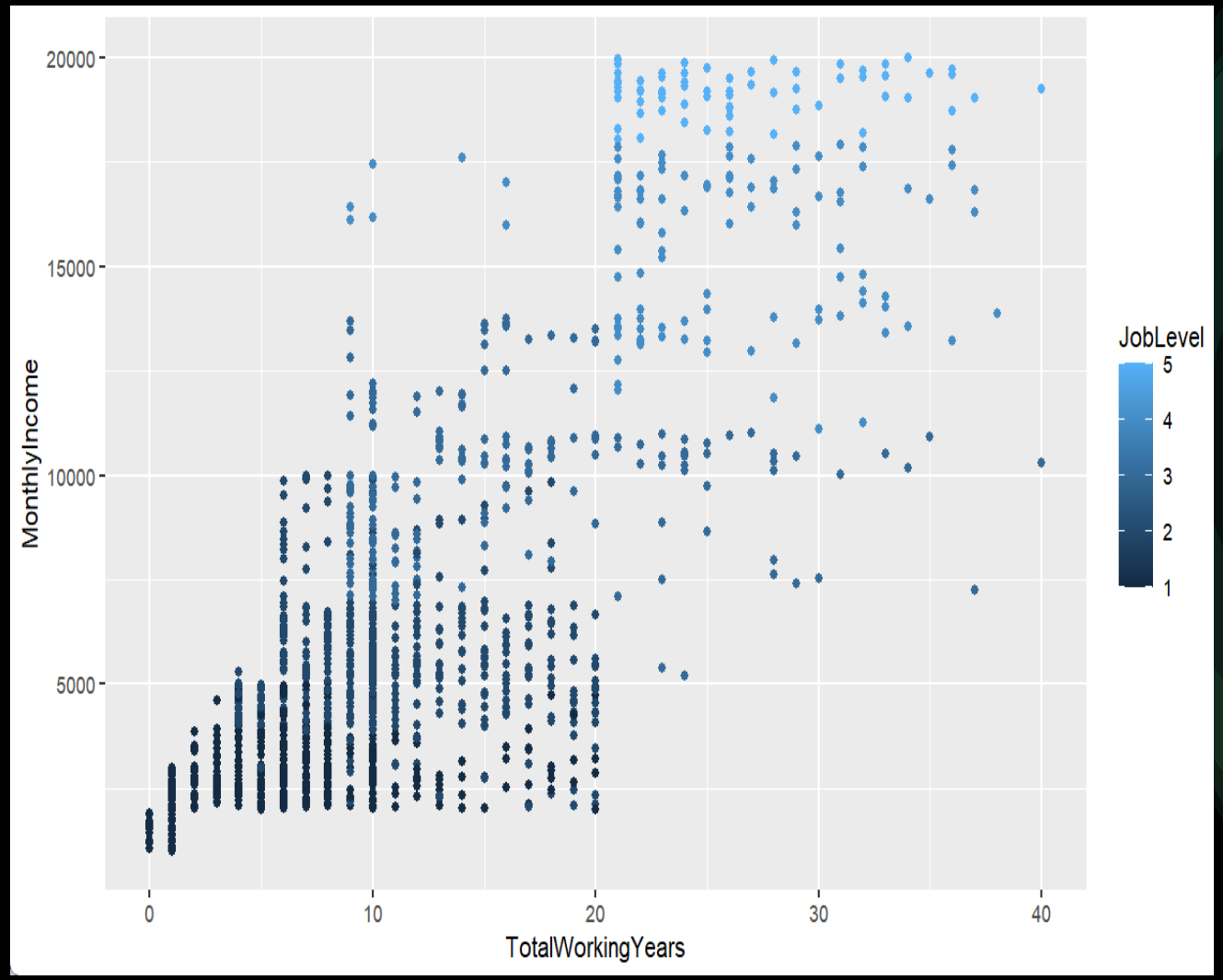
```
ggplot(attrition, aes(x=PercentSalaryHike,
y=PerformanceRating, color = Attrition)) +
geom_smooth(span = 0.5)
```

The plot shows that the percent salary hike of employees who left the company is less than that of people who stayed.

We can conclude that giving better percentage salary hikes will help in retaining employees.



DEPENDENCY OF MONTHLY INCOME AND JOB LEVEL ON TOTAL WORKING YEARS

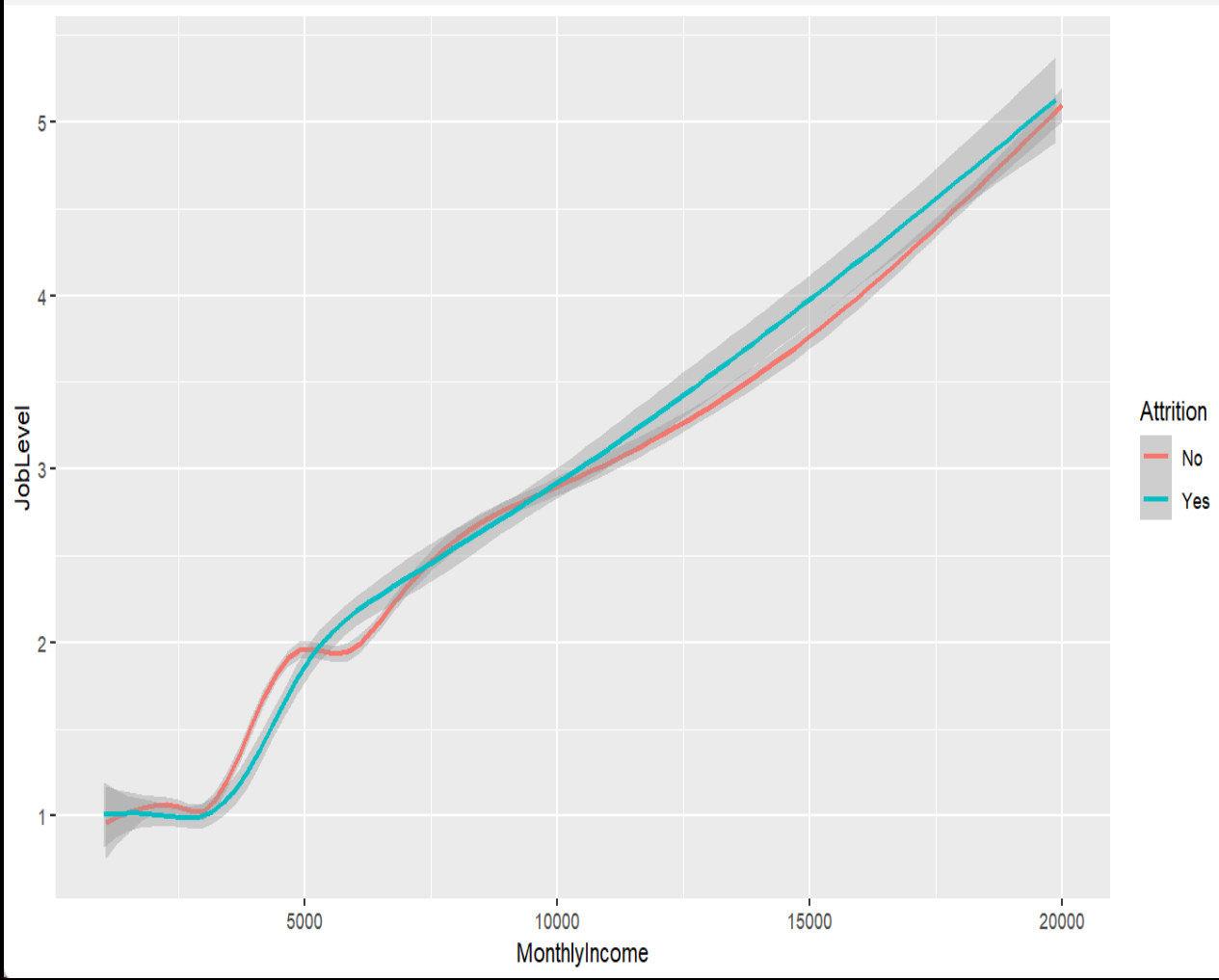


#plotting a point plot between total working years and monthly income and grouping it with Job level

```
ggplot(attrition, aes(x= TotalWorkingYears, y=MonthlyIncome, color = JobLevel)) + geom_point()
```

We may conclude from the plot that as tenure increases, so does monthly income and job level.

DISTRIBUTION OF MONTHLY INCOME THROUGHOUT JOB LEVELS



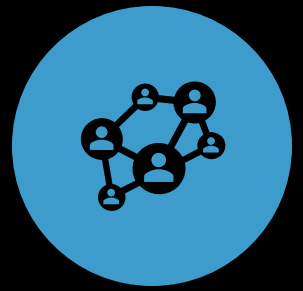
#Visualizing distribution of monthly income throughout Job levels

```
ggplot(attrition, aes(x=MonthlyIncome, y=JobLevel, color = Attrition)) +
geom_smooth()
```

We can see that as job level increases the monthly income also increases.

CONCLUSION

Employee attrition is mostly influenced by percent salary increase, promotion, work level, and monthly income.



A positive relationship with management will aid in lowering employee attrition.

Focusing on the requirements of employees at the ground level will aid in employee retention.

Extra benefits such as transportation and seasonal rewards keep the employee motivated to stay with the company.

BUSINESS CASE SUGGESTIONS

Emphasis on the needs of entry-level employees, with extra training sessions provided as needed.

Provide benefits such as transportation and seasonal rewards as a recognition for their work.

Arrange wellness sessions to increase employee engagement.

Overtime should be voluntary rather than mandatory.

Monthly one-on-one meetings to boost interactions and aid in employee career development.

Provide equal percentage compensation increases for employees with the same performance grade.

REFERENCES

- The link to the data set is https://www.kaggle.com/datasets/patelprashant/employee-attrition

- https://www.kaggle.com/code/otasra/eda-xgboost-randomforest-performance-tuning

THANK YOU