

EMPLOYEE ATTRITION

SRIHITHA SIVANNAGARI

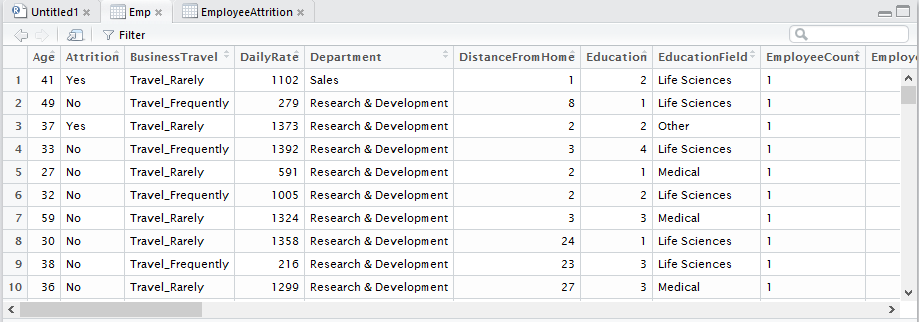
> setwd("~/HEMASIRI")

> getwd()

[1] "C:/Users/srihi/OneDrive/Documents/HEMASIRI"

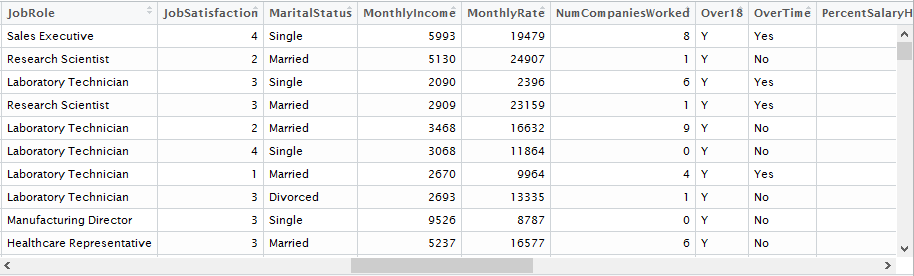
> Emp<-read.csv("Employee.csv", header = T, sep = ",")

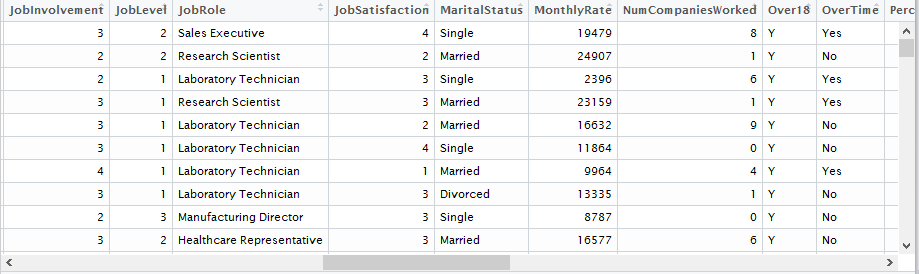
> View(Emp)



> EmployeeAttrition<-Emp[,-19]

> View(EmployeeAttrition)





REMOVING REDUNDANT DATA

|  |
| --- |
| setwd("C:/CSULA/MSIS/spring 2017/5270")  > getwd()  [1] "C:/CSULA/MSIS/spring 2017/5270"  > correct<-read.csv("incorrect.csv", header = T, sep = ",")  > View(correct)  >correct<-correct1[-7,]  >correct1<-correct[-7,]  > View(correct1) |
| BEFORE    AFTER |
| |  | | --- | |  | |

1. **How does the experience factor( number of total working years) influence performance factor.**

setwd("C:/Users/lenovo pc/Desktop/Do not touch these/BE/R")

Employee= read.csv("WA\_Fn-UseC\_-HR-Employee-Attrition (2).csv")

emp\_df<-data.frame(Employee)

performance\_rating<-emp\_df$Department

department<-emp\_df$Department

performance\_rating<-emp\_df$PerformanceRating

TotalWorkingYears<-emp\_df$TotalWorkingYears

Compare\_df<-data.frame(performance\_rating, TotalWorkingYears)

install.packages("ggplot2")

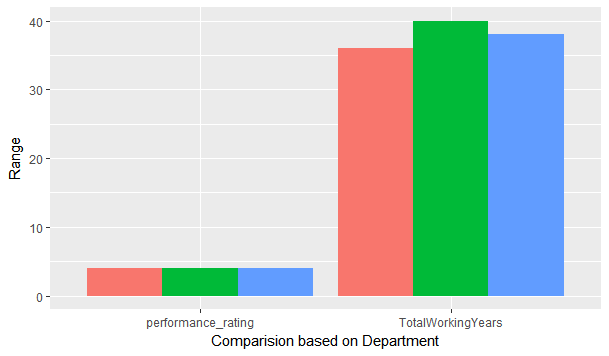
library(ggplot2)

install.packages("reshape2")

library(reshape2)

cmp\_long<-melt (Compare\_df,id.vars = "department")

ggplot(cmp\_long,aes(x=variable,y=value,fill=factor(department)))+geom\_bar(stat="identity",position="dodge")+scale\_fill\_discrete(name="Department",breaks=c(1, 2, 3),labels=c("Sales", "Human Resources", "Research & Development"))+xlab("Comparision based on Department")+ylab("Range")



The above visualization shows us the comparison of performance rating and total working year. From the visualization, we understand that the performance rating for all departments is same but the total working hours for different departments are not. For this, initially we have used a data frame to which we assigned the table to access the columns department, performance rating and to total working years. We had to install packages ggplot2 and reshape2 to use for comparison purpose of the different columns.

The above visualization illustrates that performance rating and number of total working years are independent factors.

1. **What age group does the highest over time among all the other working age groups?**

df\_1 <- data.frame(Employee$OverTime,Employee$ï..Age)

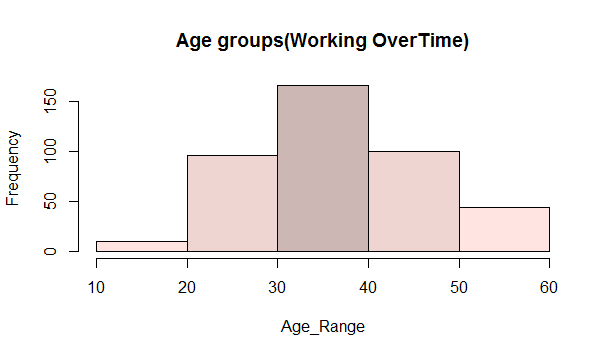
mylist <- list(split(df\_1,df\_1$Employee.OverTime))

otlist <- data.frame(sapply(mylist,'[','Yes'))

colors <- c("mistyrose", "mistyrose2", "mistyrose3", "mistyrose2", "mistyrose1")

Age\_Range<-otlist$Yes.Employee.ï..Age

hist(Age\_Range,breaks=c(10,20,30,40,50,60),main="Age groups(Working OverTime)",col=colors)



In our data set we have various age groups under working. The above visualization provides us an insight of the highest age group that does works over time. It can easily be understood that the age group which is highly involved in working overtime of all others is 30 to 40 age group in our data set.

We have used a histogram for better comprehension of the comparison. We initially assigned two columns overtime and age to a data frame and then applied “split” for the histogram.

1. **Which job role has highest raise in the salary?**

grping <- group\_by(Employee,JobRole) %>% summarise(average\_hike = round (mean (PercentSalaryHike) ,2))

d=data.frame(grping)

hike<-d[,c('JobRole','average\_hike')]

print(hike)

p <- plot\_ly(hike, labels = ~JobRole, values = ~average\_hike, type = 'pie',

textposition = 'inside',

textinfo = 'label+percent',

insidetextfont = list(color = '#FFFFFF'),

hoverinfo = 'text',

text = ~paste('$', average\_hike, 'is the average'),

marker = list(colors = colors,

line = list(color = '#FFFFFF', width = 1)),

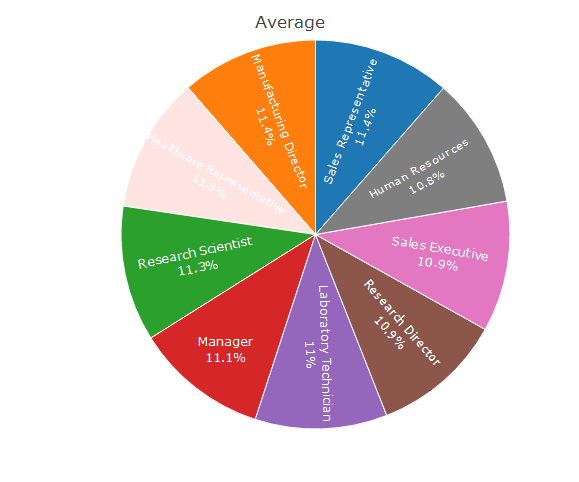
showlegend = FALSE) %>%

layout(title = 'Average',

xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),

yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

print(p)



This visualization illustrates the average raise in salary rate of employees based on their position. The pie chart conveys the percentage of average raise in salaries. We can understand that both “Manufacturing Director” and “Sales Representatives” have the highest increase in salary hikes among all the other employee positions.

The same query has been displayed using 3D pie chart.

grping <- group\_by(Employee,JobRole) %>% summarise(average\_hike = round(mean(PercentSalaryHike),2))

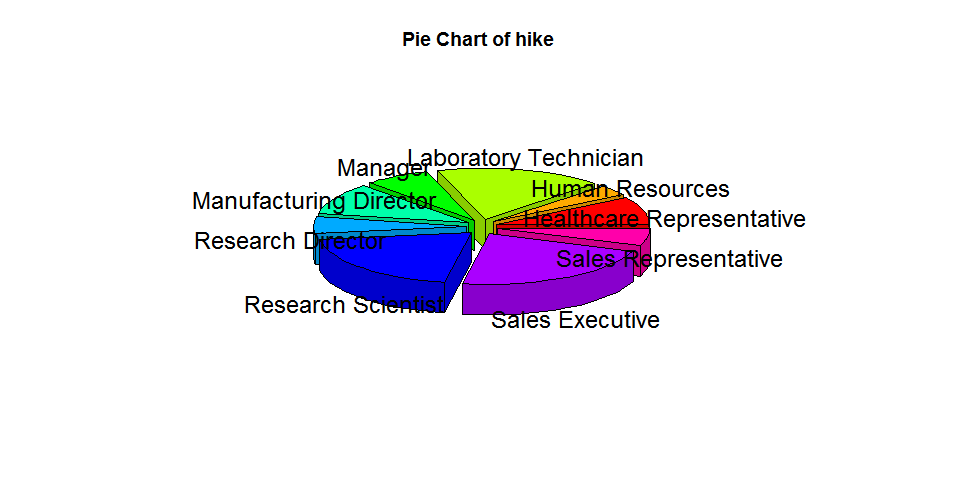
d=data.frame(grping)

hike<-d[,2]

job\_Role<-d[,1]

grping1<-table(Employee$JobRole)

pie3D(grping1,hike,labels=job\_Role,explode = 0.1, main = "Pie Chart of hike ")



1. **Which people do their jobs for many years based on their education background**

attrition<-Employee$Attrition

Educationfield<-Employee$EducationField

TotalWorkingYears<-Employee$TotalWorkingYears

attrition\_Workinghors<-data.frame(attrition,TotalWorkingYears,Educationfield)

onlyyes <- list(split(attrition\_Workinghors,attrition\_Workinghors$attrition))

yeslist <- data.frame(sapply(onlyyes,'[','Yes'))

colors<-c("yellow", "violet", "orange","blue", "pink", "cyan")

p <- plot\_ly(yeslist, x = ~Yes.attrition, y = ~Yes.Educationfield, z = ~Yes.TotalWorkingYears,

marker = list(color = ~mpg, colorscale = colors, showscale = TRUE)) %>%

add\_markers() %>%

layout(scene = list(xaxis = list(title = 'Attrition'),

yaxis = list(title = 'Education'),

zaxis = list(title = 'Years Worked')),

annotations = list(

x = 1.13,

y = 1.05,

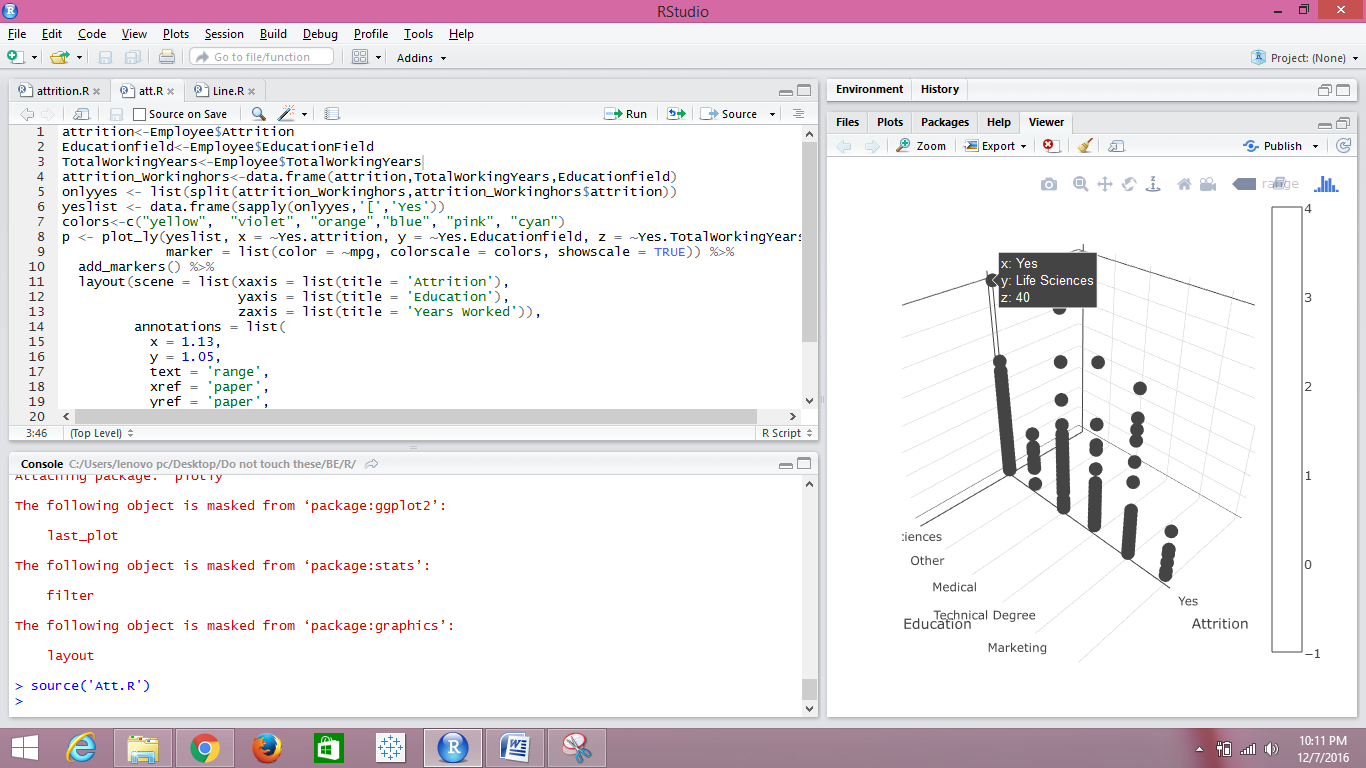
text = 'range',

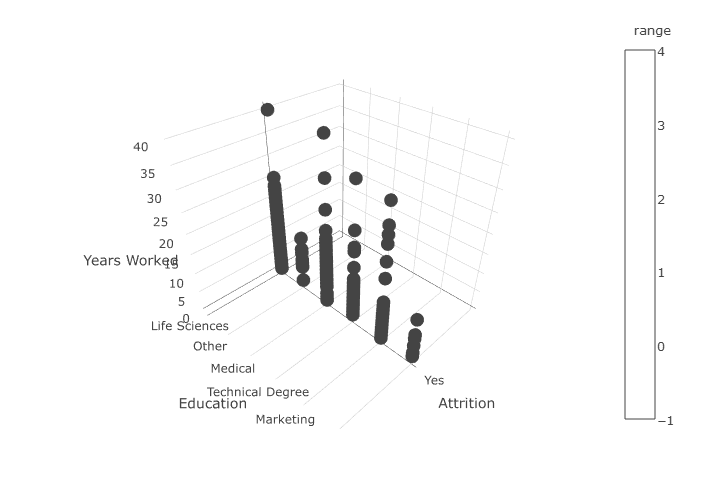
xref = 'paper',

yref = 'paper',

showarrow = FALSE

))

print(p) 



Our data set consists different streams of education for people with different employee positions. We tried to analyze using a scatter plot that what count of people are sticking to a job role which is relevant to their educational background. From the above visualization, which is a scatter plot conveys that people with “Life Science” as educational background are highest in number who do not work for other fields comparatively.

1. Compare Monthly income and Monthly rate of every department.

monthlyrate<-Employee$MonthlyRate

Department<-Employee$Department

avg\_grp <- group\_by(Employee,Department) %>% summarise(average\_monthlyincome = round(mean(MonthlyIncome)))

percentages<-data.frame(avg\_grp)

avg\_grp1 <- group\_by(Employee,Department) %>% summarise(average\_monthlyrate = round(mean(monthlyrate)))

percentages1<-data.frame(avg\_grp1)