install.packages('ggplot2')

install.packages("RColorBrewer")

library("ggplot")

#library("reshape2")

library("RColorBrewer")

rm(list=ls()) #make your space Empty

mydata = read.csv("humanresource.csv", header = TRUE)

summary(mydata)

table(mydata$perf)

hist(mydata$perf,col = "aquamarine3", main = "Distribution of performance ratings",

xlab = "Performance rating",ylab = "# Employees")

#performanceratingvsproportionleaving

agg\_sex=aggregate(vol\_leave~sex,data = mydata,mean)

print(agg\_sex)

ggplot(agg\_sex, aes(x=sex,y=vol\_leave)) + geom\_bar(stat = "identity",fill = "aquamarine3", colour = "yellow")+ ggtitle("Voluntary Termination rate by sex")+

labs(y="Proportion Leaving",x="Performance rating")

#businessareavsproportionleaving

agg\_area=aggregate(vol\_leave~area,data = mydata,mean)

print(agg\_area)

ggplot(agg\_area, aes(x=area,y=vol\_leave)) + geom\_bar(stat = "identity",fill = "aquamarine3", colour = "yellow")+ggtitle("Voluntary Termination rate by Business area")+

labs(y="Proportion Leaving",x="Business Area")

#areaandsexvsproportionleaving

agg\_as=aggregate(vol\_leave~area+sex,data = mydata,mean)

print(agg\_as)

ggplot(agg\_as,aes(x=area,sex,y=vol\_leave))+geom\_bar(stat = "identity",fill = "aquamarine3", colour = "red")+ggtitle(" Voluntary Termination rate by Business area")+

labs(y="Proportion Leaving",x="Business Area")

#analyzing the age of employees

hist(mydata$age,breaks=100,col="aquamarine3",main="Age Distribution",border=F,xlab="Age")

#ANALYZING THE ROLES OF THE EMPLOYEES TO GET A BETTER IDEA OF THE AGE DISTRIBUTION

mydata$role<-factor(mydata$role,levels=c("Ind","Manager","Director"))

boxplot(age~role,data=mydata,col="aquamarine3")

#ANALYZING VOLUNTARY TERMINATION BY AGE AND ROLE

agg\_age=aggregate(age~role,data = mydata,mean)

print(agg\_age)

names(agg\_age)<-c("Age","Probability")

ggplot(agg\_age,aes(x=Age,y=Probability))+geom\_bar(stat="identity",fill="aquamarine3",width=.7)

+ggtitle("Voluntary Termination Rate by Role and age")+

labs(y="Proportion Leaving",x="Age Cut")

#analyzing the salary pattern

summary(mydata$salary)

hist(mydata$salary,breaks=50,col="aquamarine3",main="Salary Distribution",xlab="Salary")

#FINDING THE SALARY DISTRIBUTION BASED ON QUANTILE

quantile(mydata$salary,probs=seq(0,1,.2))

#PLOTTING THE BOXPLOT TO SHOW SALARY DISTRIBUTION BASED ON ROLES

boxplot(salary~role,data=mydata,col="aquamarine3",main="Salary")