#linear regression

x<-c(14,15,16,32)

y<-c(34,65,32,67)

relation<-lm(y~x)

print(relation)

print(summary(relation))

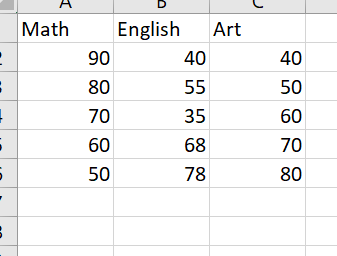
a<-data.frame(x=20)

result<-predict(relation,a)

print(result)

plot(x,y,col="blue",main="Dataset",abline(lm(x~y)),pch=16,cex=1.3,xlab="",ylab="")

#pca

Library(factoextra)

**install.packages("FactoMineR") library(FactoMineR)**

x=read.csv("C:/Users/Akki/OneDrive/Desktop/JEENAL/students.csv")

x

cov\_mat=cov(x)

cov\_mat

ex=eigen(cov\_mat)

ex

data=PCA(X,ncp=3,graph=TRUE)

datapca$eig

datapca$var

datapca$var$coord

fviz\_screeplot(datapca,addlabels=TRUE,ylim=c(0,50))

**#timeseries**

install.packages(“timeSeries”)

install.packages(“forecast”)

**library(forecast)**

**data1=tavle(AirPassengers)**

**data1**

**View(data1)**

**frequency(AirPassengers)**

**tsdata=ts(AirPassengers,frequency=12)**

**tsdata**

**plot(tsdata)**

**d=decompose(tsdata,”multiplicative”)**

**plot(d)**

**plot(d$trend)**

**plot(d$seasonal)**

**boxplot(AirPassengers~cycle(AirPassengers,xlab=”date”,ylab=”passengers count in 1000,main=”monthly box plot”))**

**mymodel<-arima(AirPassengers)**

**mymodel**

**#hypothesis**

**Data<-c(“”)**

**t.test(data, mu=30)**

**sample<-c()**

**sample2<-c()**

**t.test(sample,sample2)**

**before<-c()**

**after<-c()**

**t.test(before,after,paired=TRUE)**

**group1<-c()**

**group2<-c()**

**group3<-c()**

**cg=data.frame(cbind,(group1,group2,group3))**

**cg**

**boxplot(cg)**

**stacked\_g=stack(cg)**

**stack\_g**

**av=aov(values~ind,data=stacked\_g)**

**summary(av)**

**#data preprocessing**

**data= read.csv("titanic\_train.csv")**

**data[complete.cases(data$Age), ]**

**head(data$Age)**

**```{r setup, 1}**

**clean= read.csv("titanic\_train.csv")**

**clean$Age=replace(clean$Age, is.na(clean$Age), 0)**

**head(clean$Age)**

**{r setup, 2}**

**library(Hmisc)**

**var2= read.csv("titanic\_train.csv")**

**var2$Age=impute(var2$Age, mean)**

**head(var2$Age)**

**{r setup, 3}**

**categories= read.csv("titanic\_train.csv")**

**categories$Sex=factor(categories$Sex)**

**head(categories$Sex)**

**{r setup, 4}**

**varr1= read.csv("titanic\_train.csv")**

**boxplot(varr1$Fare, main="Boxplot of Fare", ylab="Fare", col="orange", horizontal=TRUE)**

**```{r}**

**try= read.csv("titanic\_train.csv")**

**boxplot(try$Age)**

**outliers=boxplot.stats(try$Age)$out**

**try$Age[is.na(try$Age)] <- median\_age**

**#clustering**

**df=read.csv()**

**df**

**plot(df)**

**boxplot(df)**

**set.seed(20)**

**c1=kmeans(df[,1:2],3)**

**c1**

**iris**

**View(iris)**

**head(iris)**

**summary(iris)**

**plot(iris)**

**plot(df[,3:4])**

**kmeansc1=kmeans(iris[,3:4],3)**

**kmeansc1**

**table(kmeansc14cluster,iris$Species)**

**boxplot(iris)**