wine=read.csv("wine.csv")

str(wine)

summary(wine)

model1=lm(Price~ AGST,data=wine)

model1

names(model1)

summary(model1)

model1$coefficients

#SSE

SSE=sum(model1$residuals^2)

SSE

#addd new variable

model2=lm(Price~AGST+HarvestRain,data=wine)

summary(model2)

SSE2=sum(model2$residuals^2)

SSE2

#use all variables

#year is excluded because It does not make sense. We already have Age

#and age is equivalent to having the year

model=lm(Price~.-Year,data=wine)

summary(model)

SSE3=sum(model$residuals^2)

SSE3

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

final\_model=lm(Price~.-FrancePop-Year,data=wine)

summary(final\_model)

#plot linear regression

par(mfrow=c(2,2))

plot(final\_model)

plot(wine$Age,wine$FrancePop)

pairs(wine)

#library(car)

cor(wine)

cor(wine)[2,]

scatterplotMatrix(wine)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

winetest=read.csv("wine\_test.csv")

predictTest=predict(final\_model,newdata=winetest)

predictTest

SSE=sum((winetest$Price-predictTest)^2)

SST=sum((winetest$Price-mean(wine$Price))^2)

testR2=1-(SSE/SST)

testR2