#toyota\_r<-read.csv("/Volumes/Data/Course Content/DS content/Linear Regression/Toyoto\_Corrola.csv")

toyota\_r<-Toyota

#check if na values are there

colSums([is.na](http://is.na/)(toyota\_r))

toyota<-na.omit(toyota\_r)

colnames(toyota)

toyota<-toyota[,c(3,4,7,9,16,13,14,17,18)]

#toyota1<-subset(toyota,select = c("Price","cc") )

#Scatter Plot Matrix:

pairs(toyota)

#Correlation Matrix:

cor(toyota)

#Regression Model and Summary

model.car<-lm(Price~.,data = toyota)

summary(model.car)

#Multi-colinearity

install.packages("car")

library(car)

car::vif(model.car)

##Subset selection

#library(MASS)

#stepAIC(model.car)

#########Model Validation

#Diagnostic Plots:

#Residual Plots, QQ-Plos, Std. Residuals vs Fitted

plot(model.car)

#Residuals vs Regressors

residualPlots(model.car)

#Added Variable Plots

avPlots(model.car)

#QQ plots of studentized residuals

qqPlot(model.car)

#Deletion Diagnostics

influenceIndexPlot(model.car) # Index Plots of the influence measures

####Iteration 1

#Remove 77th observation

toyota['Age2']<-toyota$Age\_08\_04\*toyota$Age\_08\_04

toyota1<-toyota[-c(79),]

model.car1<-lm(Price~.,data = toyota1)

summary(model.car1)

plot(model.car1)

residualPlots(model.car1)

qqPlot(model.car1)

influenceIndexPlot(model.car1)

########

toyota2<-toyota[-c(79,219),]

model.car2<-lm(Price~.,data = toyota2)

summary(model.car2)

qqPlot(model.car2)

influenceIndexPlot(model.car2)

###Final

model.car2<-lm(Price~.,data = toyota[-c(950,79,600,220,219,218,520,957,217,518),-c(7)])

summary(model.car2)

qqPlot(model.car2)

influenceIndexPlot(model.car2)

##Predict for new data

testdata<-data.frame(Age\_08\_04=20,KM =2000,HP =90,Gears =5,cc=1500,

                     Quarterly\_Tax=200,Weight=1500,Age2=400)

predict(model.car2,testdata)