####

####Fossil Data based on Geological Society of America Bulletin by Bralwer T.J et. al.

####We will be doing a simple and polynomial regression of this dataset

####The explanatory variable, X, in this dataset is Age

####The response variable is, Y, is the Strontium Ratio.

####

####

###################Answers to Question 1###################

##We first set the working directory of the data set. It is this directory where the required files are place.

setwd("F:/OneDrive/Skeena")

##Clean the workspace .

rm(list = ls())

## Read the data into a frame called fossil

fossil = read.csv("fossil(1).csv")[1:106 ,]

##Attach the dataset so that the column names can directly be used in modelling or plotting. This may be dangerous

attach (fossil)

###################Answers to Question 2###################

## Do some plotting to visualize the relation ship between Age and Sr Ratio.

#### We see from the first plot that the relation is polynomail atleast and not linear

dev.off()

plot(age , strontium.ratio , pch = "\*" , cex = 1 , col = 1 , main = "Fossil Data plot")

###################Answers to Question 3###################

####

#### Fit a simple linear model to the data and overlay it on the plot. We see that the linear model does not model the data correctly.

####

par(mfrow = c(2,2))

plot(age , strontium.ratio , pch = "\*" , cex = 1 , col = 1 , main = "Fossil Data plot - Linear")

fit.1 = lm(strontium.ratio ~ age )

lines(age , fit.1$fitted.values , type = "p" , pch = "\*", col = 2)

legend("bottomleft",

legend = c("Linear fit" ),

col = 2,

cex = 1.1,

pch = "\*"

)

###################Answers to Question 4###################

####

#### Fit a Quadratic model to the data and overlay it on the plot. We see that the Quadratic model does not model the data correctly.

####

plot(age , strontium.ratio , pch = "\*" , cex = 1 , col = 1 , main = "Fossil Data plot - Quad")

fit.2 = lm(strontium.ratio ~ poly(age , 2))

lines(age , fit.2$fitted.values , type = "p" , col = 3)

legend("bottomleft",

legend = c("Quadratic fit" ),

col = 3,

cex = 1.1,

pch = "\*"

)

###################Answers to Question 5##################

####

#### Fit a Cubic model to the data and overlay it on the plot. We see that the Cubic model does model the data reasonably

#### close but still not correctly.

####

plot(age , strontium.ratio , pch = "\*" , cex = 1 , col = 1 , main = "Fossil Data plot - Cubic")

fit.3 = lm(strontium.ratio ~ poly(age , 3))

lines(age , fit.3$fitted.values , type = "p" , col = 4)

legend("bottomleft",

legend = c( "Cubic" ),

col = 4,

cex = 1.1,

pch = "\*"

)

###################Answers to Question 6##################

####

#### Fit a Quartic model to the data and overlay it on the plot. We see that the Quartic model does model the data reasonably

#### close but still not correctly. The fit is very similar to cubic fit.

####

plot(age , strontium.ratio , pch = "\*" , cex = 1 , col = 1 , main = "Fossil Data plot - Quartic")

fit.4 = lm(strontium.ratio ~ poly(age , 4))

lines(age , fit.4$fitted.values , type = "p" , col = 5)

legend("bottomleft",

legend = c("Quartic"),

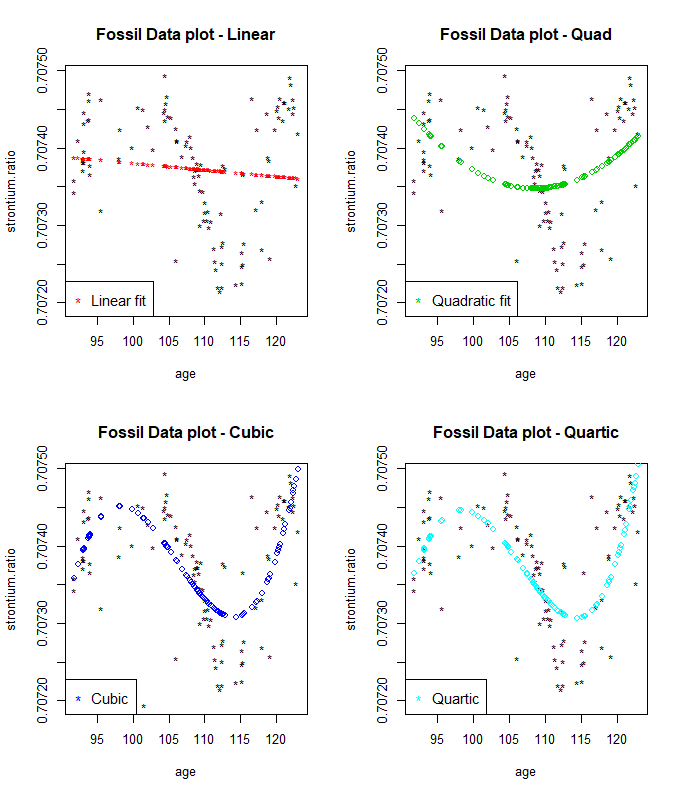
col = 5,

cex = 1.1,

pch = "\*"

)

All Plots are shown next.



###### All overlayed

dev.off()

plot(age , strontium.ratio , pch = "\*" , cex = 1 , col = 1 , main = "Fossil Data plot - Comparing fits")

lines(age , fit.1$fitted.values , type = "p" ,pch = "\*" , col = 2)

lines(age , fit.2$fitted.values , type = "p" , pch = "\*" , col = 3)

lines(age , fit.3$fitted.values , type = "p" ,pch = "\*" , col = 4)

lines(age , fit.4$fitted.values , type = "p" ,pch = "o" , col = 5 )

legend("bottomleft",

legend = c("Linear" , "Quad", "Cubic" , "Quartic"),

col = 2:5,

cex = 1.1,

pch = "\*"

)

