# Homework 3

Skylar Liu

2024-01-27

Homework #3, Stat 660, Spring 2024, Due Class #4, January 29

1. Please find the fossil data on Canvas, and download it and its documentation

2. Do a scatterplot of the data, with X = age. As seen near the end of Lecture 1, it is distinctly not linear, and hence a perfect example for semiparametric regression.

```r
# clear workspace
rm(list = ls())
# set the seed
set.seed(382957)

# import data
fossil = read.csv("~/660 - Flexible Regression/Homework/Homework3/fossil.csv")
X = fossil$Age
Y = fossil$Strontium.Ratio

# scatterplot X vs Y
plot(X, Y, type="p",pch='*',main="Fossil Age vs. Strontium Ratio",
     xlab="Age",ylab="Strontium Ratio",cex=2)
```
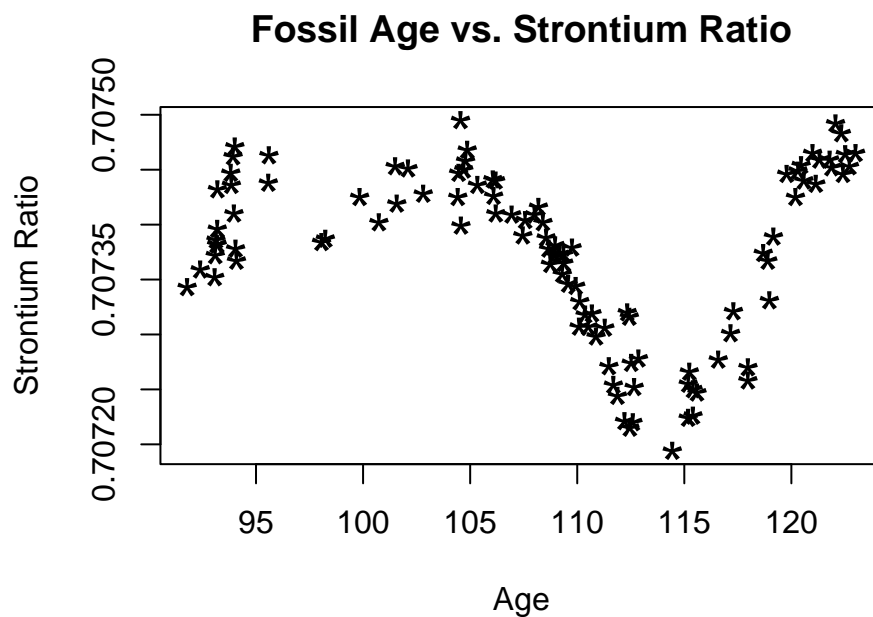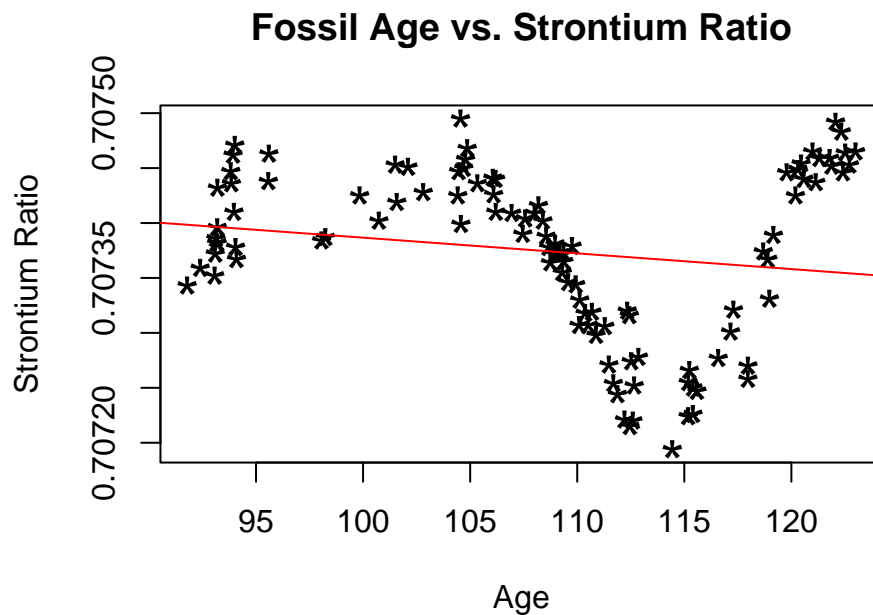
3. Fit a linear model and add the fit to the scatterplot

```r
# linear regression of Y on X
lin = lm(Y ~ X)

# scatterplot X vs Y
plot(X, Y, type="p",pch='*',main="Fossil Age vs. Strontium Ratio",
     xlab="Age",ylab="Strontium Ratio",cex=2)
abline(lin, col="red")
```


4. Fit a quadratic model and add the fit to the scatterplot of the raw data. Does this help? Why or why not?
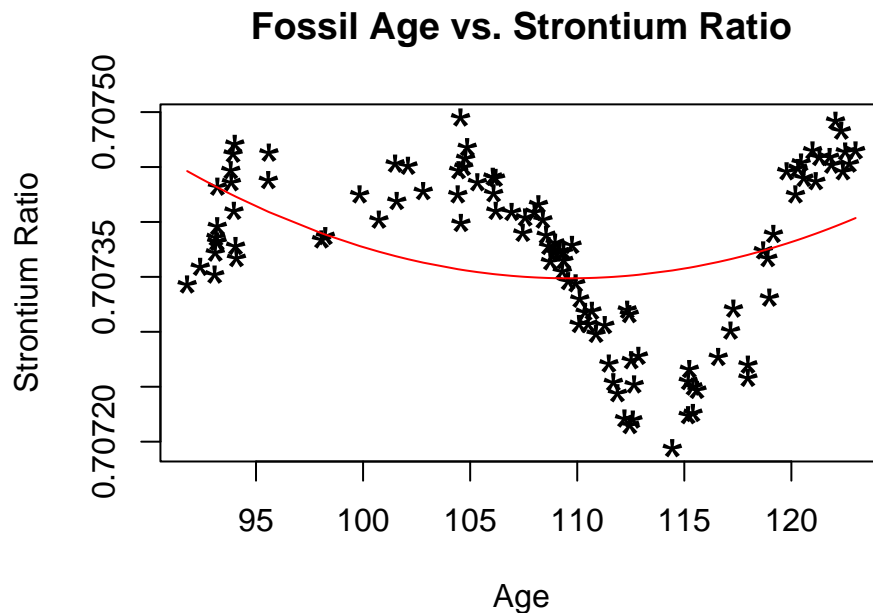
```r
#create a new variable for age^2
X2 = X^2

#fit quadratic regression model
quad = lm(Y ~ X + X2)

#create list of predicted happines levels using quadratic model
strontPredict = predict(quad)
ord = sort(X, index.return = T)$ix

#scatterplot of X and Y
plot(X, Y, type="p",pch='*',main="Fossil Age vs. Strontium Ratio",
     xlab="Age",ylab="Strontium Ratio",cex=2)

#add predicted lines based on quadratic regression model
lines(X[ord], strontPredict[ord], col='red')
```

Fossil Age vs. Strontium Ratio

A quadratic fit does not help since a quadratic line does not fit the curved path of the scatterplot. A quadratic fit is bimodal (one knot) while this graph appears to be multi-modal (two knots).

5. Fit a cubic model and add the fit to the scatterplot of the raw data. Does this help? Why or why not?
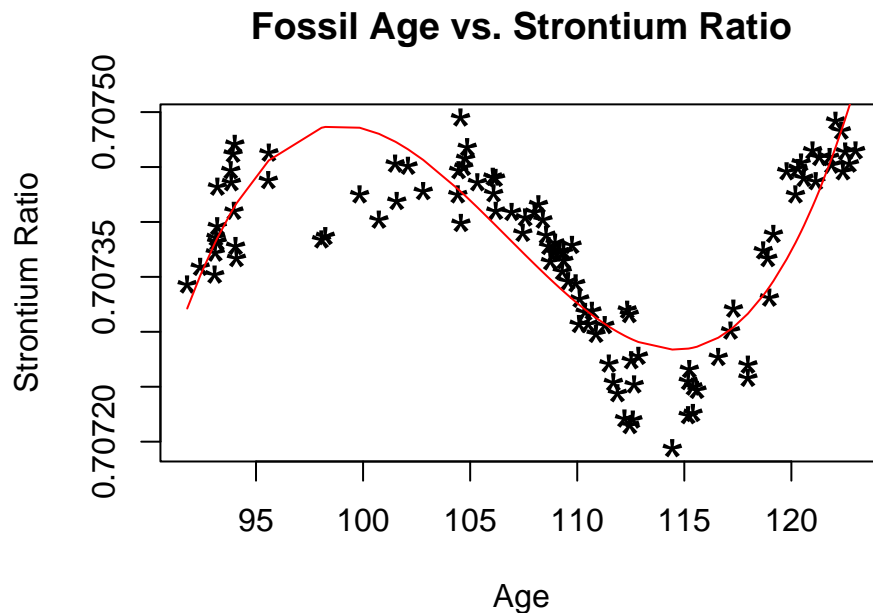
```
#create a new variable for age^3
X3 = X^3

#fit quadratic regression model
cub = lm(Y ~ X + X2 + X3)

#create list of predicted happines levels using quadratic model
strontPredict = predict(cub)

#scatterplot of X and Y
plot(X, Y, type="p",pch='*',main="Fossil Age vs. Strontium Ratio",
     xlab="Age",ylab="Strontium Ratio",cex=2)

#add predicted lines based on quadratic regression model
lines(X[ord], strontPredict[ord], col='red')
```

Fossil Age vs. Strontium Ratio

6. Fit a quartic $(x+x^{2+x}3+x\char`^4)$ model and add the fit to the scatterplot of the raw data. Does this help? Why or why not?

```
#create a new variable for age^4
X4 = X^4

#fit quadratic regression model
quart = lm(Y ~ X + X2 + X3 + X4)

#create list of predicted happines levels using quadratic model
strontPredict = predict(quart)

#scatterplot of X and Y
plot(X, Y, type="p",pch='*',main="Fossil Age vs. Strontium Ratio",
     xlab="Age",ylab="Strontium Ratio",cex=2)

#add predicted lines based on quadratic regression model
lines(X[ord], strontPredict[ord], col='red')
```

Fossil Age vs. Strontium Ratio
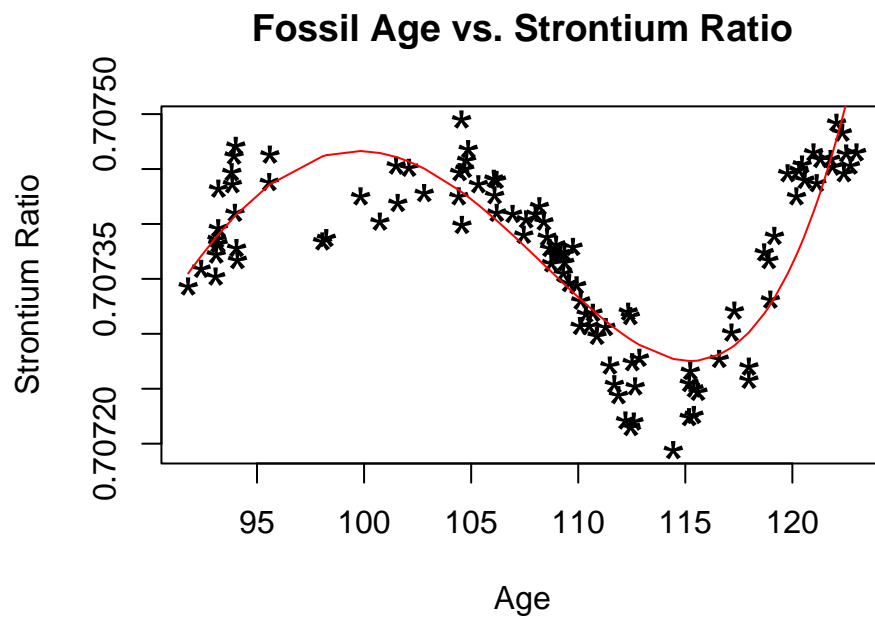
This model does not really help as it's very similar to the cubic regression model. Since the cubic model follows the curves with the right amount of knots, I would prefer that model.