STAT313 Assignment 4

2023-05-22

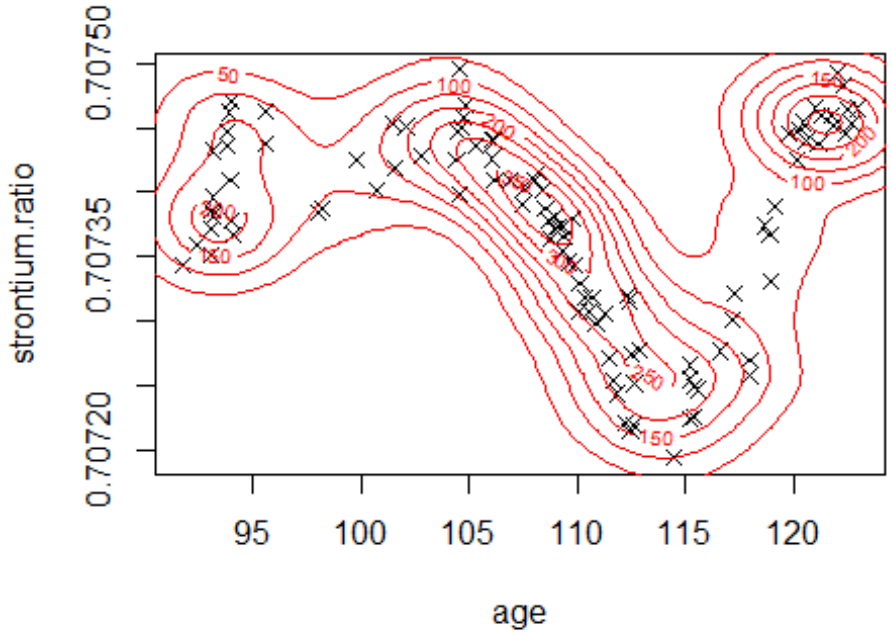
Question 1:

a)

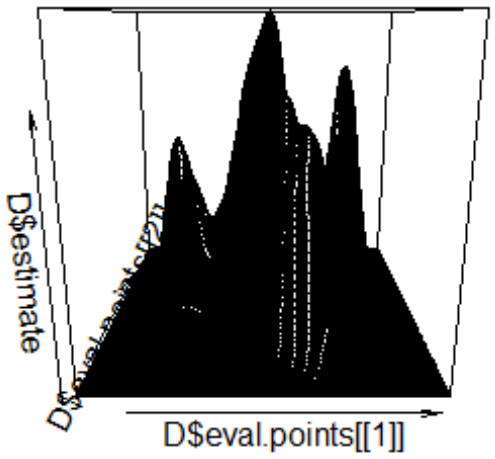
```
library(ks)
fosil <- read.table("fosil.txt", header = TRUE)

D <- kde(fosil)
```

b)

```
plot(fosil, pch=4)
contour(D$eval.points[[1]], D$eval.points[[2]], D$estimate, col='red',
add=TRUE)
```



```
persp(D$eval.points[[1]], D$eval.points[[2]], D$estimate)
```



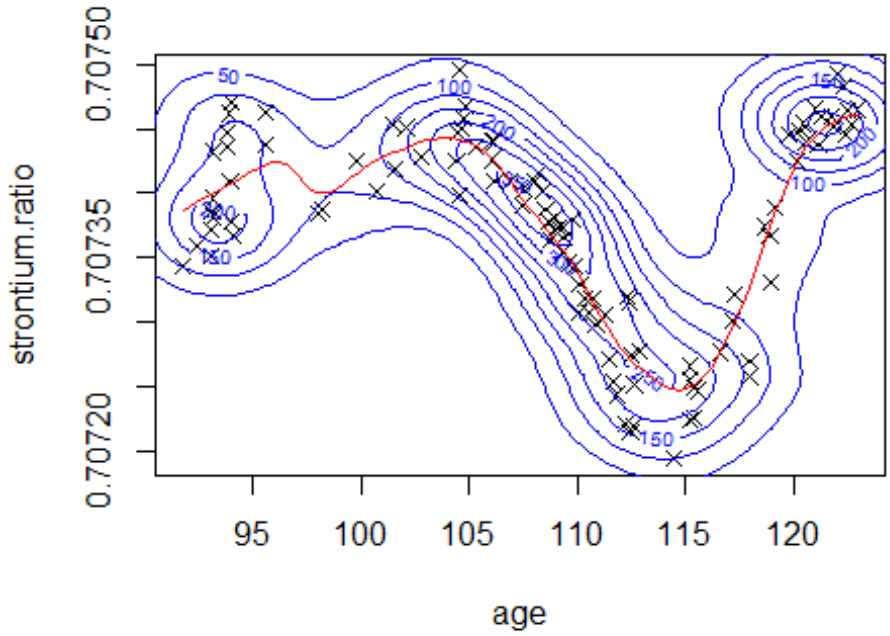
c)

```
mod1 <- ksmooth(fosil$age, fosil$strontium.ratio, kernel = "normal",
bandwidth = 3.5)
```

The bandwidth of 3.5 is good because it captures all the important feautures of the model without overfitting the data.

d)

```
plot(fosil, pch=4)
contour(D$eval.points[[1]], D$eval.points[[2]], D$estimate, col='blue',
add=TRUE)
lines(mod1, col='red')
```



e)

```
new_fosil <- fosil[c("strontium.ratio", "age")]
D2 <- kde(new_fosil)
```

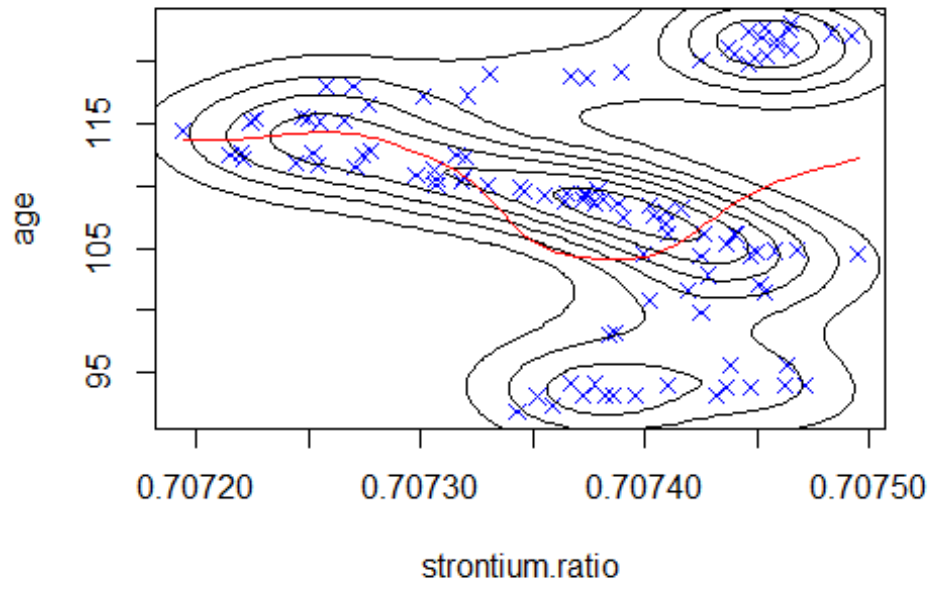
f)

```
mod2 <- ksmooth(new_fosil$strontium.ratio, new_fosil$age, kernel = 'normal',
bandwidth = 0.000055)
```

The bandwidth of 0.000055 is good because it captures the main features of the data without over fitting.

g)

```
plot(new_fosil, pch=4, col='blue')
contour(D2$eval.points[[1]], D2$eval.points[[2]], D2$estimate, add=TRUE)
lines(mod2, col='red')
```



Question 2:

a) Both graphs have 4 distinct peaks. The density with strontium as the response has a clear trend where strontium decreases around age 105 then begins to increase after 115.When age is the resposne there is no clear trend. You can get one density from the other simply by swapping the x and y axis.

b) Both regression functions form a horizontal 'S' shape. The regression function is relatively flat when age is the response variable. in comparison when strontium ratio is the response there is a clear trend where strontium ratio decreases then increases as age increases. We can not get one regression function from the other. This is because the regression functions are not rotations of one another.

c) model1 would work well to predict strontium ratio. There seems to be a relatively low resiuduals on average and the data follows a clear trend. The regression function struggles to fit the data as it exits the contour lines. This means age could be a useful predictor for the strontium ratio of a fosil.

d) model2 would NOT work well to predict age. The residuals get very large as age the strontium ratio increases. The data follows no clear trend. A high strontium ratio could mean a fosil is relatively old or relatively young. This means strontium ratio is not a good predictor for age.