M2_Project2

2022-10-11

Q1 Print your name at the top of the script. Include the prefix: "Plotting Basics:"

```
print("Plotting Basics: Udaikiran")
```

```
## [1] "Plotting Basics: Udaikiran"
```

Q2 Import libraries including: plyr, FSA, FSAdata, magrittr, dplyr, plotrix, ggplot2, and moments

```
library(plyr)
library(FSA)
library(FSAdata)
library(magrittr)
library(dplyr)
library(plotrix)
library(ggplot2)
library(moments)
```

Q3 Load the BullTroutRML2 dataset (https://www.rdocumentation.org/packages/FSAdata/versions/0.3.9/topics/BullTroutRML2)

```
data("BullTroutRML2")
```

Q4 Print the first and last 3 records from the dataset

This dataset seems to be of lakes with fields including the Age of the lake, fork Length, and which era it belongs to.

```
head(BullTroutRML2,n=3)
```

```
##   age  fl     lake     era
## 1  14 459 Harrison 1977-80
## 2  12 449 Harrison 1977-80
## 3  10 471 Harrison 1977-80
```

```
tail(BullTroutRML2,n=3)
```

```
##    age  fl   lake     era
## 94   4 298 Osprey 1997-01
## 95   3 279 Osprey 1997-01
## 96   3 273 Osprey 1997-01
```

Q5 Filter out all records except those from Harrison Lake

Majority of the entries are done for Harrison Lake as the total entries are 96 and out of them, 61 entries are for Harrison lake. And it seems like most of the entries are in the era 1997-01.

```
BullTroutRML2 %<>% dplyr::filter(lake=="Harrison")
print(BullTroutRML2)
```

```
##    age  fl     lake     era
## 1   14 459 Harrison 1977-80
## 2   12 449 Harrison 1977-80
## 3   10 471 Harrison 1977-80
## 4   10 446 Harrison 1977-80
## 5    9 400 Harrison 1977-80
## 6    9 440 Harrison 1977-80
## 7    9 462 Harrison 1977-80
## 8    8 480 Harrison 1977-80
## 9    8 449 Harrison 1977-80
## 10   7 437 Harrison 1977-80
## 11   7 431 Harrison 1977-80
## 12   7 425 Harrison 1977-80
## 13   7 419 Harrison 1977-80
## 14   6 409 Harrison 1977-80
## 15   6 397 Harrison 1977-80
## 16   5 419 Harrison 1977-80
## 17   5 381 Harrison 1977-80
## 18   5 363 Harrison 1977-80
## 19   5 351 Harrison 1977-80
## 20   4 372 Harrison 1977-80
## 21   2 199 Harrison 1977-80
## 22   2 184 Harrison 1977-80
## 23   1  91 Harrison 1977-80
## 24  12 440 Harrison 1997-01
## 25  11 428 Harrison 1997-01
## 26  10 440 Harrison 1997-01
## 27  10 422 Harrison 1997-01
## 28   9 434 Harrison 1997-01
## 29   9 415 Harrison 1997-01
## 30   9 406 Harrison 1997-01
## 31   8 434 Harrison 1997-01
## 32   8 406 Harrison 1997-01
## 33   8 375 Harrison 1997-01
## 34   7 415 Harrison 1997-01
## 35   7 394 Harrison 1997-01
## 36   6 381 Harrison 1997-01
## 37   6 357 Harrison 1997-01
## 38   5 341 Harrison 1997-01
## 39   5 326 Harrison 1997-01
## 40   4 304 Harrison 1997-01
```

```
## 41   4 292 Harrison 1997-01
## 42   4 270 Harrison 1997-01
## 43   4 252 Harrison 1997-01
## 44   4 221 Harrison 1997-01
## 45   3 258 Harrison 1997-01
## 46   3 233 Harrison 1997-01
## 47   3 211 Harrison 1997-01
## 48   3 205 Harrison 1997-01
## 49   3 180 Harrison 1997-01
## 50   2 196 Harrison 1997-01
## 51   2 171 Harrison 1997-01
## 52   2 143 Harrison 1997-01
## 53   1 131 Harrison 1997-01
## 54   1  88 Harrison 1997-01
## 55   1  75 Harrison 1997-01
## 56   0  51 Harrison 1997-01
## 57   0  41 Harrison 1997-01
## 58   0  20 Harrison 1997-01
## 59   7 245 Harrison 1997-01
## 60   7 279 Harrison 1997-01
## 61   5 245 Harrison 1997-01
```

Q6 Display the first and last 3 records from the filtered dataset

```
headtail(BullTroutRML2,n=3)
```

```
##    age  fl     lake     era
## 1   14 459 Harrison 1977-80
## 2   12 449 Harrison 1977-80
## 3   10 471 Harrison 1977-80
## 59   7 245 Harrison 1997-01
## 60   7 279 Harrison 1997-01
## 61   5 245 Harrison 1997-01
```

Q7 Display the structure of the filtered dataset

The data has both integer values and repeating values for lake and era.

```
str(BullTroutRML2)
```

```
## 'data.frame':    61 obs. of  4 variables:
##  $ age : int  14 12 10 10 9 9 9 8 8 7 ...
##  $ fl  : int  459 449 471 446 400 440 462 480 449 437 ...
##  $ lake: Factor w/ 2 levels "Harrison","Osprey": 1 1 1 1 1 1 1 1 1 1 ...
##  $ era : Factor w/ 2 levels "1977-80","1997-01": 1 1 1 1 1 1 1 1 1 1 ...
```

Q8 Display the summary of the filtered dataset and save it as

```
summary(BullTroutRML2)
```

```
##       age               fl            lake         era
##  Min.   : 0.000   Min.   : 20   Harrison:61   1977-80:23
##  1st Qu.: 3.000   1st Qu.:221   Osprey  : 0   1997-01:38
```

```
##  Median : 6.000   Median :372
##  Mean   : 5.754   Mean   :319
##  3rd Qu.: 8.000   3rd Qu.:425
##  Max.   :14.000   Max.   :480

t<-summary(BullTroutRML2)
```

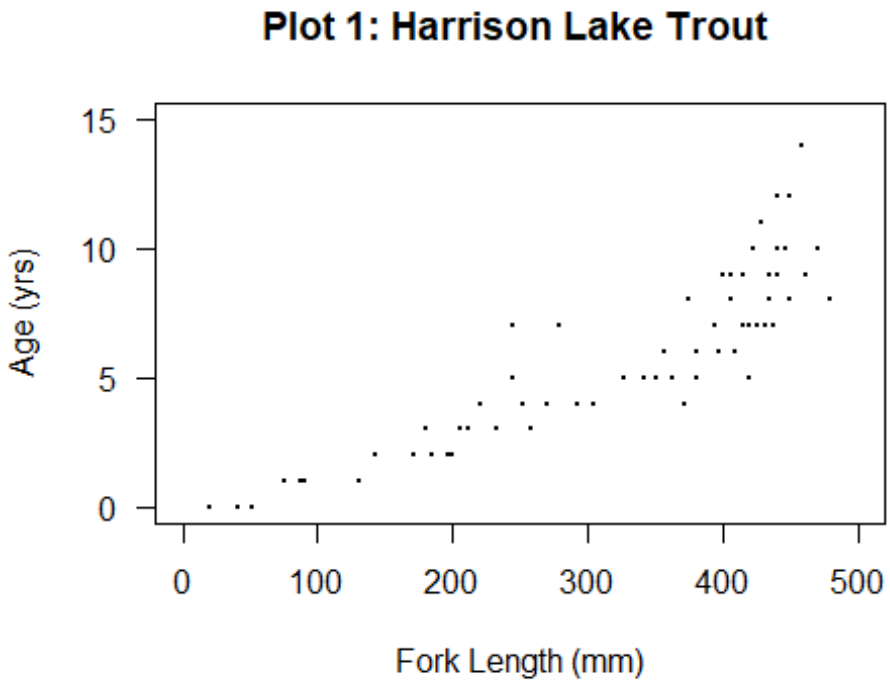
Q9 Create a scatterplot for “age” (y variable) and “fl” (x variable) with the following

specifications:

- **Limit of x axis is (0,500)**
- **Limit of y axis is (0,15)**
- **Title of graph is “Plot 1: Harrison Lake Trout**
- **X axis label is “Age (yrs)”**
- **Y axis label is “Fork Length (mm)”**
- **Use a small filled circle for the plotted data points**

With the following graph, we can conclude that the lakes with greater age have greater Fork Lengths. And more than 30% of the lakes have more than 400 Fork Lengths.

```
x_q9<- BullTroutRML2$fl
y_q9<-BullTroutRML2$age
plot(x_q9,y_q9, ylim=c(0, 15), xlim=c(0,500), ylab="Age (yrs)",
     xlab="Fork Length (mm)", pch=20, las=1, cex=.5, main="Plot 1: Harrison
Lake Trout ")
```



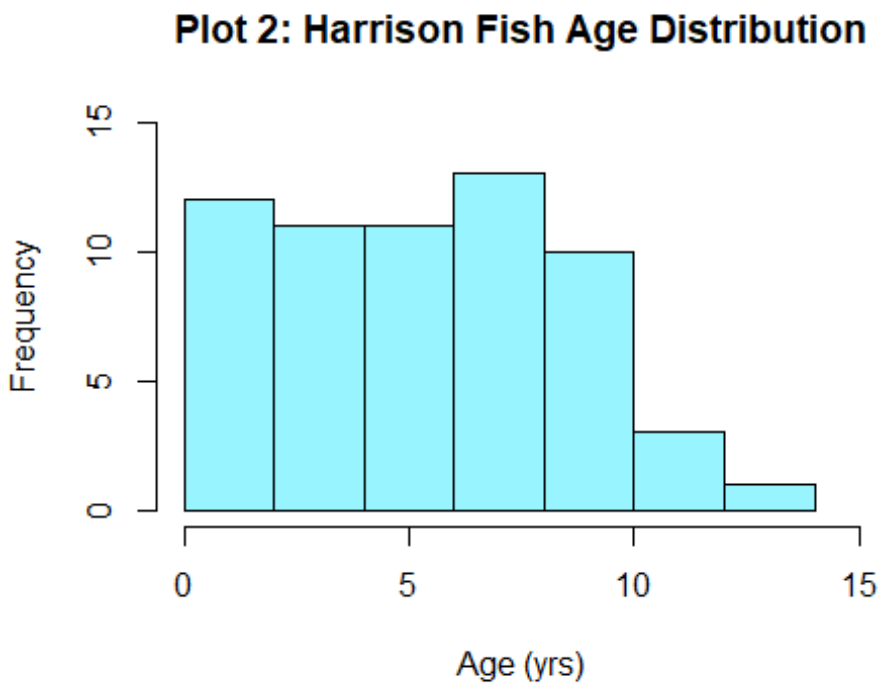
Q10 Plot an "Age" histogram with the following specifications

specifications:

- **Y axis label is "Frequency"**
- **X axis label is "Age (yrs)"**
- **Title of the histogram is "Plot 2: Harrison Fish Age Distribution"**
- **The color of the frequency plots is "cadetblue"**
- **The color of the Title is "cadetblue"**

Among all the lakes More number of lakes are having ages of 6.5-7.5 yrs. And very few are of age more than 12. There are an equal number of lakes of age between 2-5.5 yrs.

```
hist(BullTroutRML2$age, ylim=c(0,15), xlim=c(0,15), ylab="Frequency",
     xlab="Age (yrs)", main="Plot 2: Harrison Fish Age Distribution",
col="cadetblue1")
```

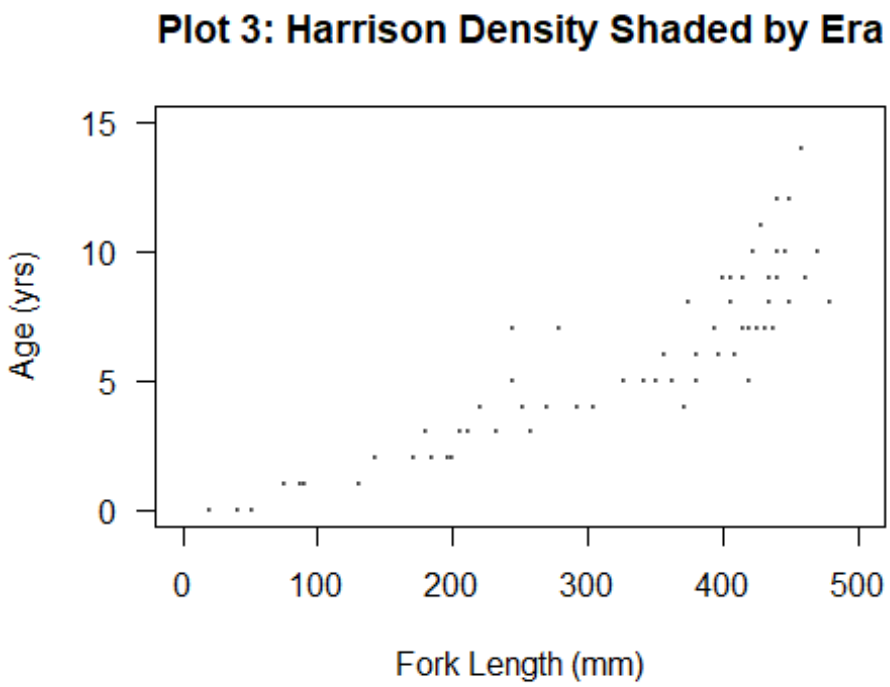


Q11 Create a overdense for "age" (y variable) and "fl" (x variable) with the following

specifications:

- **Limit of x axis is (0,500)**
- **Limit of y axis is (0,15)**
- **Title of graph is "Plot 3: Harrison Density Shaded by Era"**
- **X axis label is "Age (yrs)"**
- **Y axis label is "Fork Length (mm)"**
- **Use a small filled circle for the plotted data points**
- **include two levels of shading for the "black" data points**

```
x_q11<- BullTroutRML2$fl
y_q11<-BullTroutRML2$age
plot(x_q11,y_q11, ylim=c(0, 15), xlim=c(0,500), ylab="Age (yrs)",
     xlab="Fork Length (mm)", pch=20, col=rgb(0,0,0,1/2), las=1, cex=.5,
main="Plot 3: Harrison Density Shaded by Era")
```



Q12 Create a new object called "tmp" that includes the first 3 and last 3 records of the whole data set.

```
tmp <- BullTroutRML2[c(1:3,31:33),]
tmp
```

```
##    age  fl     lake     era
## 1   14 459 Harrison 1977-80
## 2   12 449 Harrison 1977-80
## 3   10 471 Harrison 1977-80
## 31   8 434 Harrison 1997-01
## 32   8 406 Harrison 1997-01
## 33   8 375 Harrison 1997-01
```

Q13 Display the "era" column in the new "tmp" object

```
tmp$era
```

```
## [1] 1977-80 1977-80 1977-80 1997-01 1997-01 1997-01
## Levels: 1977-80 1997-01
```

Q14 Create a pchs vector with the argument values for + and x. Then create a cols vector with the two elements "red" and "gray60"

```
pchs <- c(3,4)

cols <-c("red","gray60")
```

Q15 Convert the tmp object values to numeric values. Then create a numeric numEra object from the tmp$era object

```
as.numeric(tmp$era)

## [1] 1 1 1 2 2 2

numEra <- as.numeric(tmp$era)
```

Q16 Associate the cols vector with the tmp era values

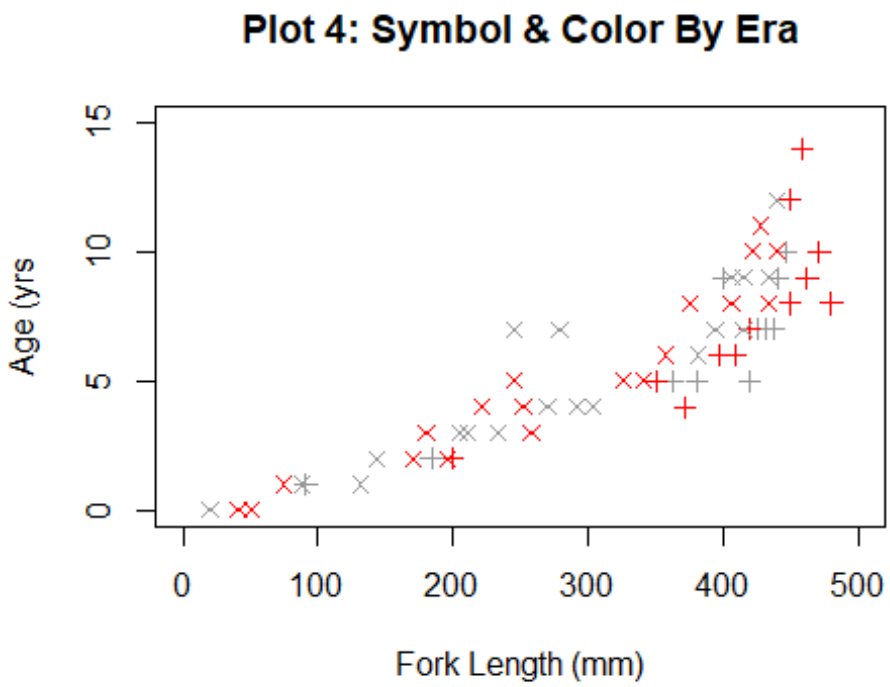
```
cols[tmp$era]

## [1] "red"    "red"    "red"    "gray60" "gray60" "gray60"
```

Q17 Create a plot of “Age (yrs)” (y variable) versus “Fork Length (mm)” (x variable) with the

following specifications:

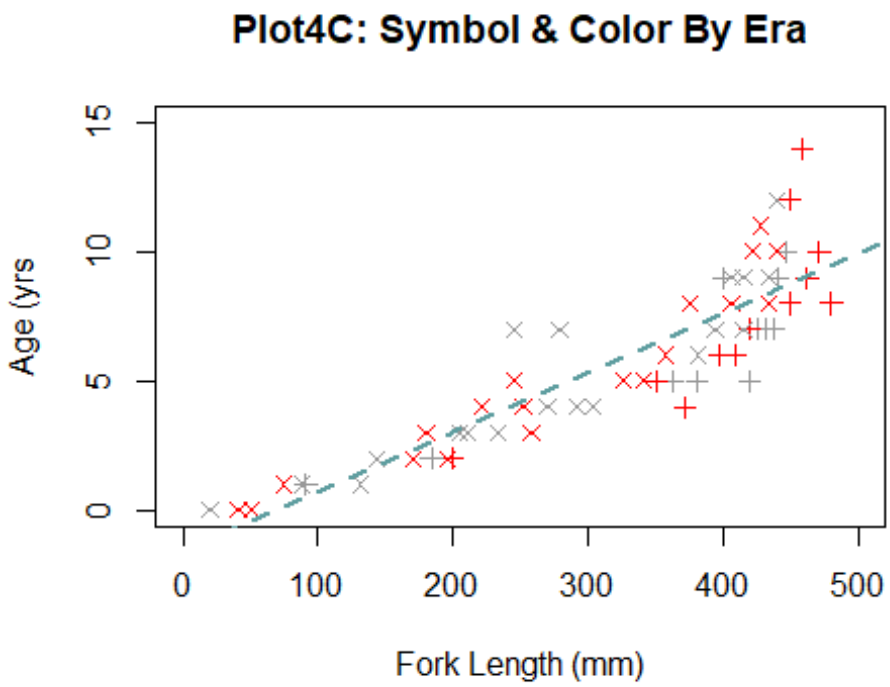
- **Limit of x axis is (0,500)**
- **Limit of y axis is (0,15)**
- **Title of graph is “Plot 4: Symbol & Color by Era”**
- **X axis label is “Age (yrs)”**
- **Y axis label is “Fork Length (mm)”**
- **Set pch equal to pchs era values**
- **Set col equal to cols era values.**

```
plot(age~fl,data=BullTroutRML2, ylim=c(0, 15), xlim=c(0,500), ylab="Age
(yrs",
     xlab="Fork Length (mm)", pch=pchs[era], col=cols[tmp$era], main="Plot 4:
Symbol & Color By Era")
```



Q18 Plot a regression line of the previous plot with a dashed line with width 2 and color "cadetblue"

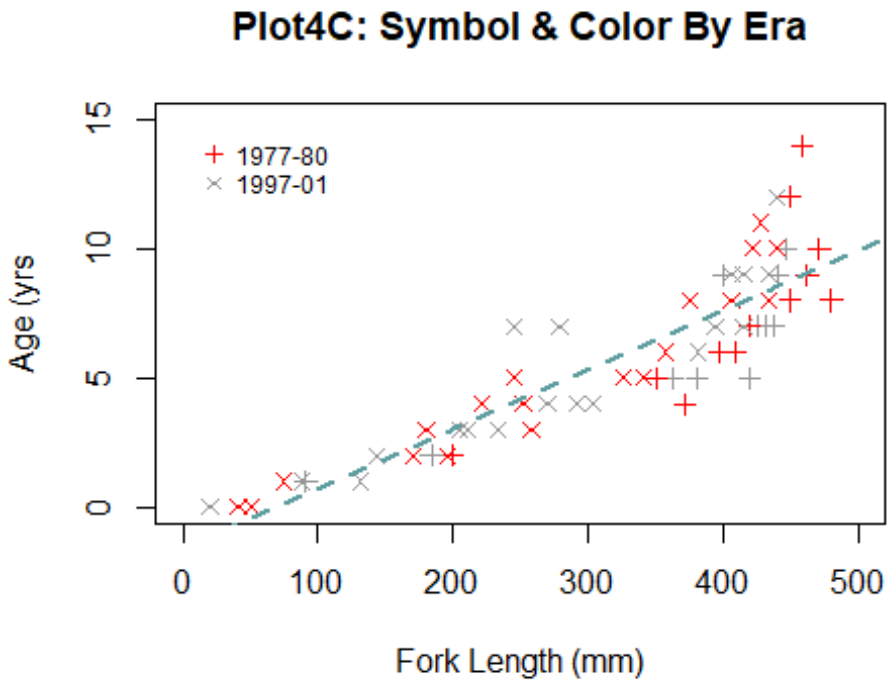
```
cols[numEra]

## [1] "red"    "red"    "red"    "gray60" "gray60" "gray60"

plot(age~fl,data=BullTroutRML2, ylim=c(0, 15), xlim=c(0,500), ylab="Age
(yrs",
     xlab="Fork Length (mm)", pch=pchs[era], col=cols[numEra], main="Plot4C:
Symbol & Color By Era")
regLine <-lm(age~fl, data=BullTroutRML2)
abline(regLine, col="cadetblue", lty=2, lwd=2)
```



Q19 Place a legend of levels by era with pchs symbols in the top left of the plot with the following ##specifications:

- **Inset of 0.05**
- **No box around the legend**
- **Font size: 75% of nominal**

```
plot(age~fl,data=BullTroutRML2, ylim=c(0, 15), xlim=c(0,500), ylab="Age
(yrs",
     xlab="Fork Length (mm)", pch=pchs[era], col=cols[numEra], main="Plot4C:
Symbol & Color By Era")
regLine <-lm(age~fl, data=BullTroutRML2)
abline(regLine, col="cadetblue", lty=2, lwd=2)
legend("topleft", inset=0.05, legend=levels(BullTroutRML2$era),
       pch=pchs, col=cols, bty="n", cex=0.75)
```


Plot4C: Symbol & Color By Era
1977-80
1997-01
Age (yrs
0
5
10
15
0
100
200
300
400
500
Fork Length (mm)