

Module 2 Project

Course Name: ALY6000 Introduction to Analytics

Student Name: Sanjana Sandeep Mohile

Date: 2022/01/28

Title: Executive Summary Report 2

Professor: Mohammad Shafiqul Islam

NUID: 002123793

INTRODUCTION

The following project will provide you with a thorough understanding of R and RStudio as well as some hands-on experience. The objective of the project is to learn:

- How to load and install several libraries
- How to only show a portion of the dataset's observations
- How to create a new dataset from an existing one
- How to decipher a dataset's structure and summary
- With a given set of specifications, how to generate a scatterplot, histogram, overdense plot, regression line, and legend

A. Provide an analysis of descriptive characteristics of the data set provided by your instructor. This includes pertinent statistics including mean, median, quartiles, variance, standard deviation, skew, kurtosis, outliers, etc. Include R console screenshots to support your observations and conclusions. Below is a sample excerpt of an analysis of Harrison Lake fish from the BullTroutRML2 dataset.

```
> str(harrison_data_new) #Q7
'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 ...
```

Explanation: -

- ➔ On applying a filter [lake == Harrison] to the original dataset i.e. BullTroutRML2, we get a new dataset which is named harrison_data_new in this document.
- ➔ This new dataset contains 61 observations and 4 variables: namely – age, fl, lake, era.
- ➔ Here, age and fl are of same data type i.e. int whereas lake and era belong to factor data structure.

- ➔ Lake and era, both have two elements each; namely – Harrison and Osprey; 1977-80 and 1997-01 respectively.

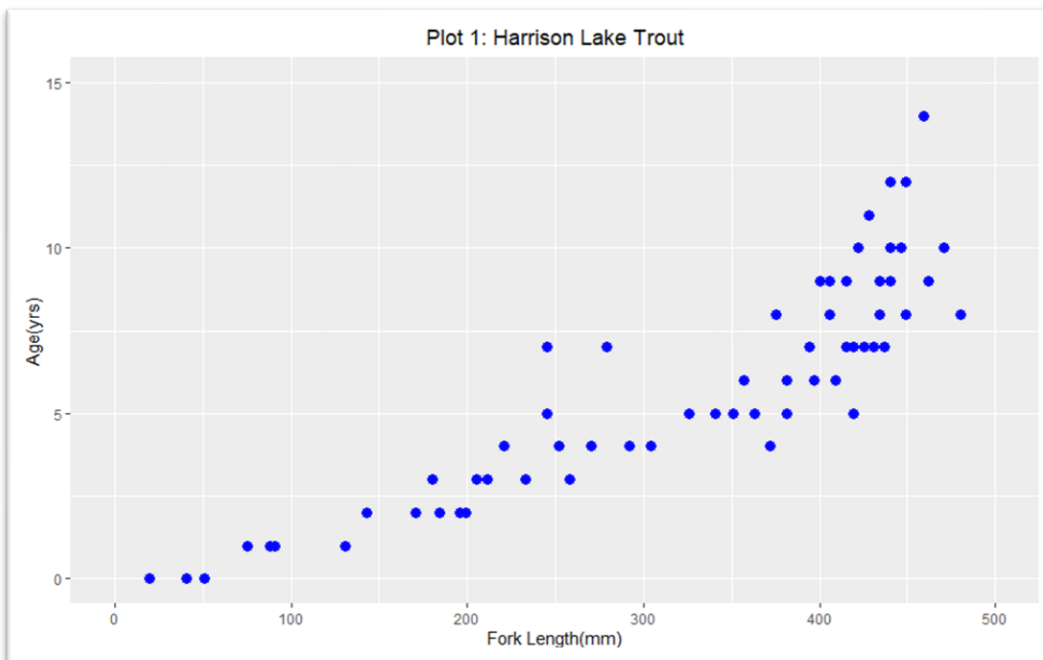
```
> summary(harrison_data_new)
      age      fl      lake      era
Min.   : 0.000  Min.   : 20  Harrison:61  1977-80:23
1st Qu.: 3.000  1st Qu.:221  Osprey   : 0  1997-01:38
Median : 6.000  Median :372
Mean   : 5.754  Mean   :319
3rd Qu.: 8.000  3rd Qu.:425
Max.   :14.000  Max.   :480
```

Explanation: -

- ➔ The following is the summary of the filtered dataset harrison_data_new.
- ➔ Harrison Lake has a minimum age of 0 years and a maximum of 14 years. The mean age of the lake is found to be 5.754 years and its median is seen to be 6.
- ➔ The fork length has a minimum of 20mm and a maximum of 480mm. Its median is seen to be 372mm and the mean is found to be 319mm.
- ➔ Since the data set is filtered with Harrison Lake's observations, only the Osprey number is zero and all the 61 observations are of Harrison Lake only.
- ➔ Era has two levels, 1977-1980 and 1997-2001 has 23 and 38 observations respectively.

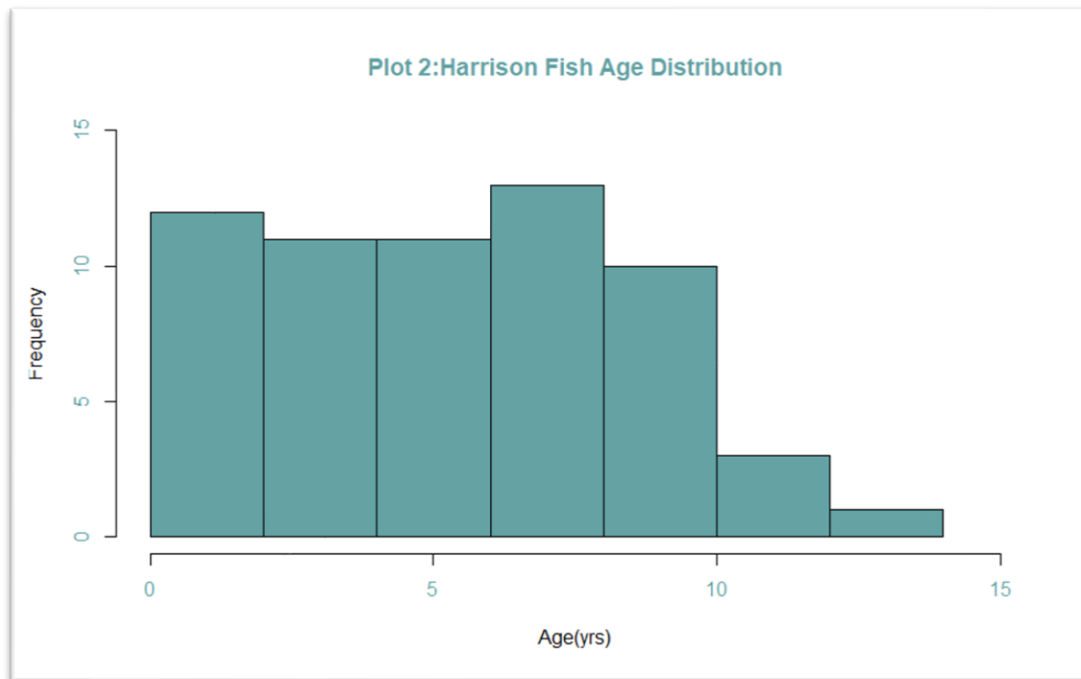
B. Provide the executive with visualizations (at least 6) in that help them see the key characteristics you want to highlight. They can be boxplots, histograms, frequency and probability distributions, barplots (bar charts) or pareto. Not only is the goal to present your visual results, but also to explain the significance of what the visuals are displaying.

1]



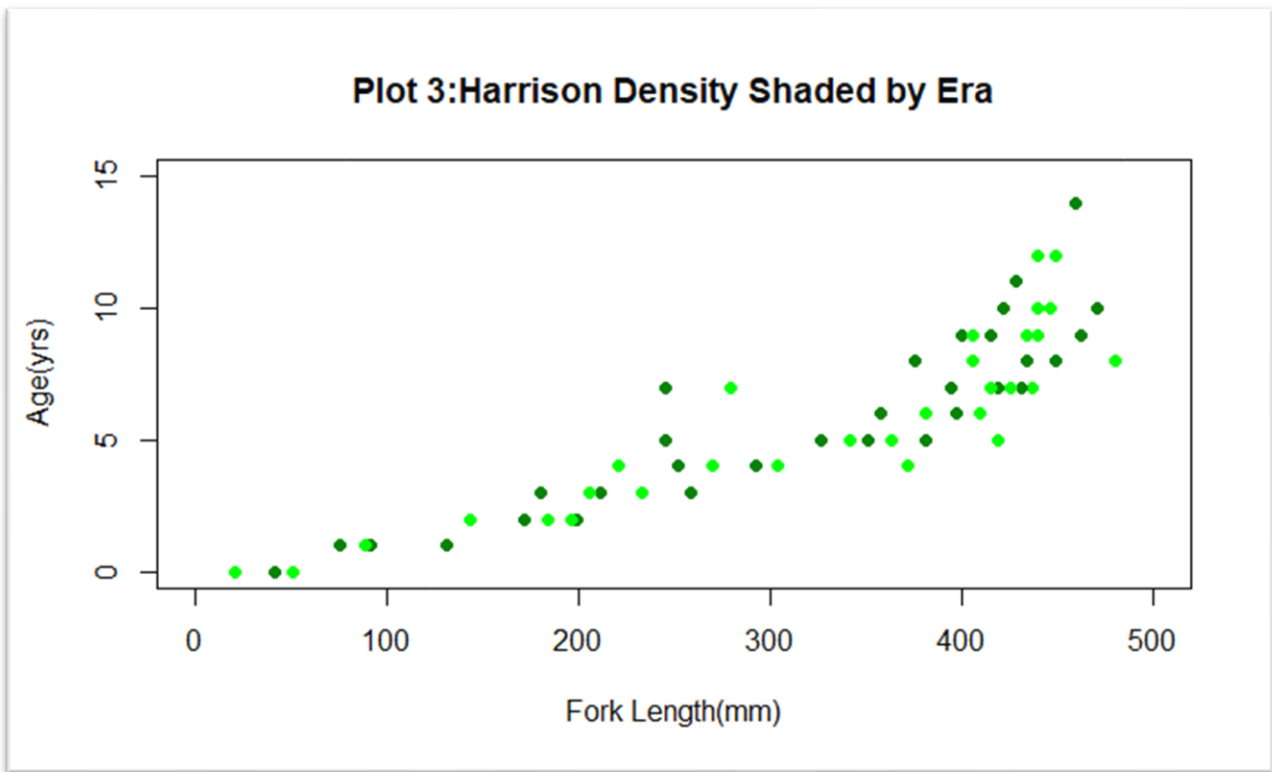
- ➔ The Harrison Lake Trout Scatter plot is plotted using the `harrison_data_new` dataset.
- ➔ On the x-axis, fork length is measured in mm whereas on the y-axis age is measured in years.
- ➔ By observing, one thing becomes very clear and that is the length of the lake has increased gradually over time.
- ➔ Over the age of 15 years, the lake rose from 20mm (which is the minimum length we observed above, on giving the command `summary(harrison_data_new)`) to 480mm(which was again observed from the `summary(harrison_data_new)`).

2]



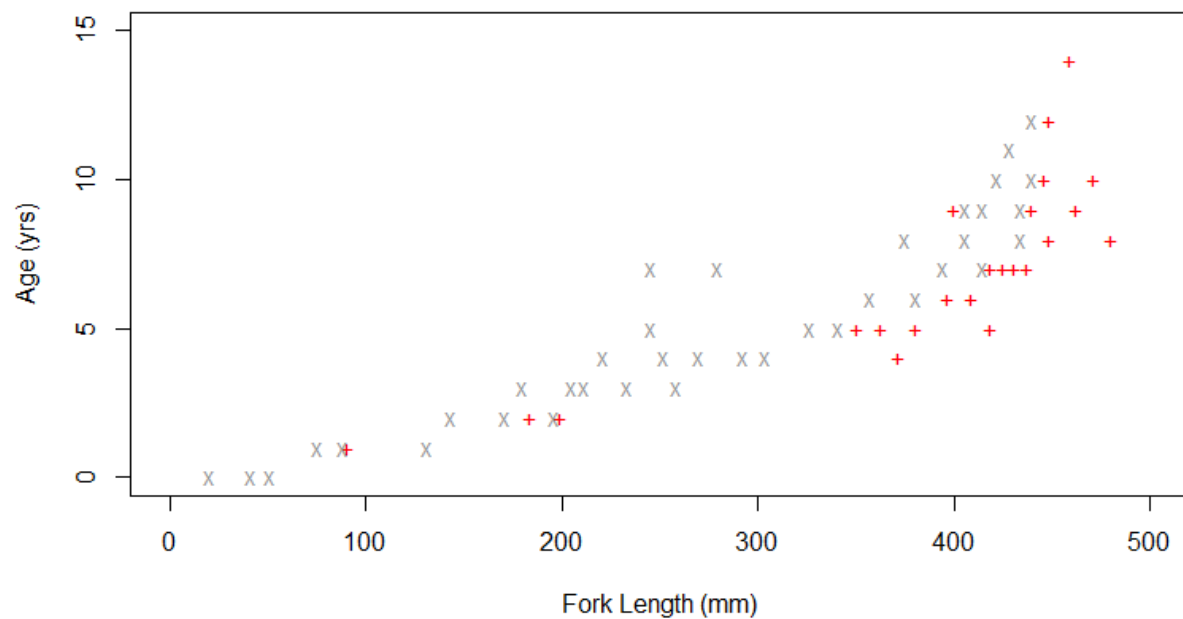
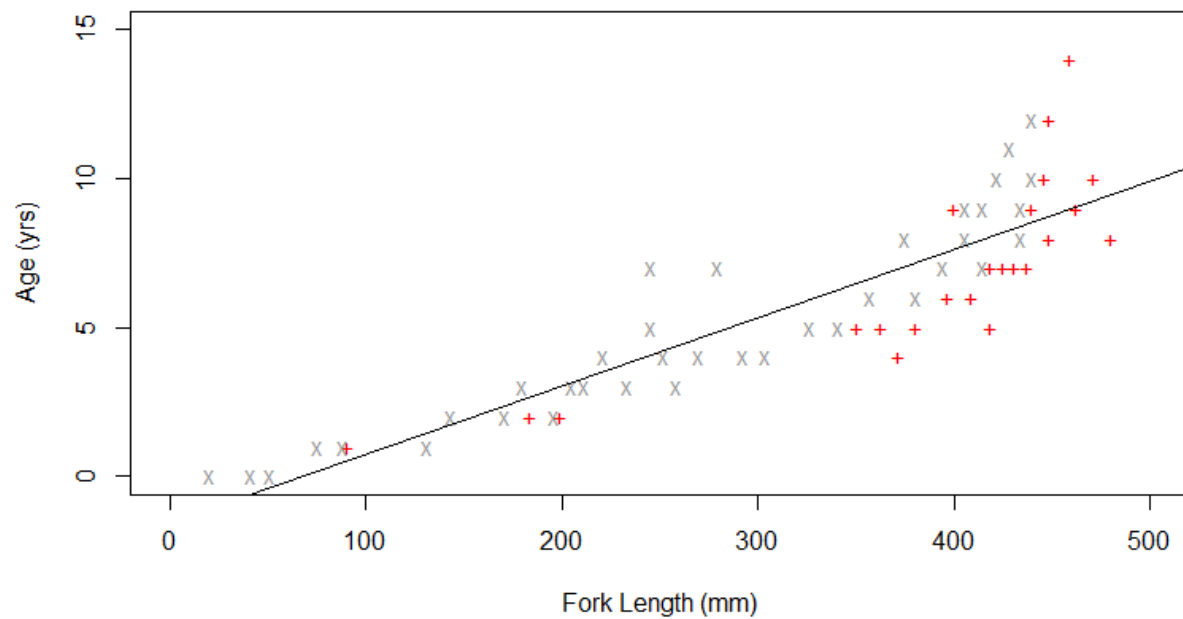
- ➔ The above histogram depicts the frequency of fishes in Harrison lake on the basis of their age.
- ➔ On observing the graph we see that the maximum age of fishes is 13-14 years approximately.
- ➔ There are at least 12 fishes aged from 0 to 2 years.
- ➔ Approximately, there are 11 fish those range from 3 to 6 years.
- ➔ Their maximum frequency is seen to be approximately 13 at the age of 6-8 years.
- ➔ After crossing 8 years, their frequency is decreased from 13 to 10.
- ➔ There is a sudden decrease in the frequency of fishes at the age of 10 years which was to be recorded as 2-3 by the age of 12 years and 1-2 by the age off 14 years.

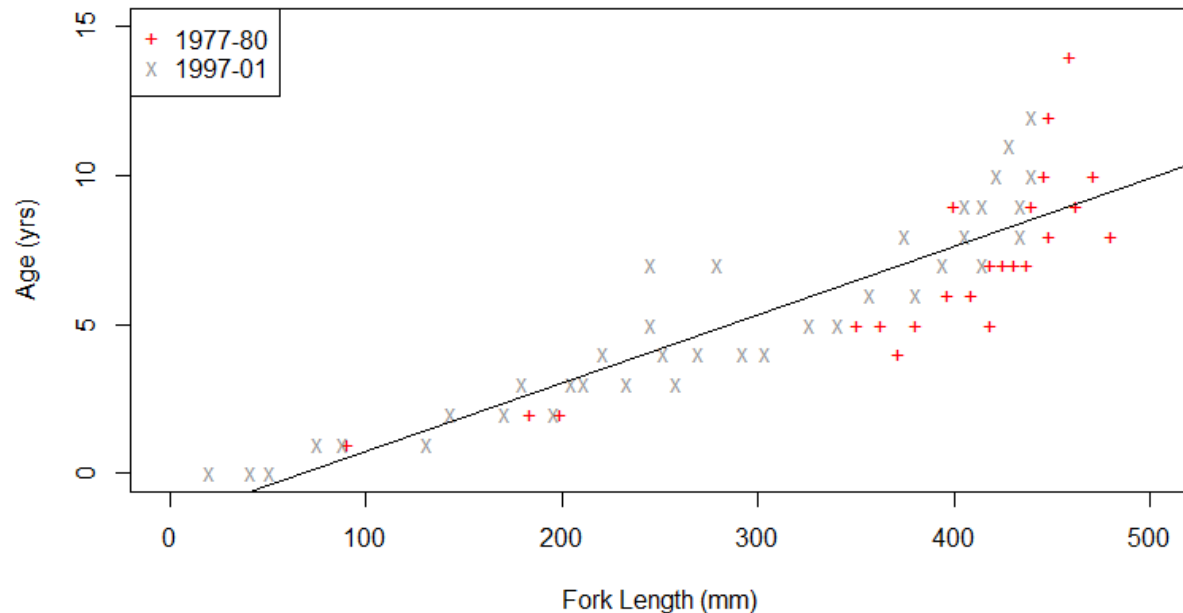
Q3]



- ➔ The above is a overdense plot of harrison_data_new dataset, in which fork length is measured vs the age of the lake.
- ➔ Since there are two levels of era, we have two different coloured points.
- ➔ The minimum fork length is measured to be approximately 20mm and the maximum is seen to be 480mm.

Q4]

Plot 4: Symbol & Color by Era**Plot 5: Regression Overlay**

Plot 6: Legend Overlay

- ➔ The above plots 4,5 and 6 are fork length vs age graph of the filtered data set i.e. harrison_new_data. Fork length is measured in mm and age is measured in years.
- ➔ In all the three graphs, the eras are of two levels. One represented by the red cross and another represented by grey X's.
- ➔ Plot 4 shows the two different levels of eras whereas plot 5 has a regression line as well. Plot 6 includes both the attributes of plots 4 and 5. In addition to that, it also has a legend which makes it easy for an individual to understand the graph.
- ➔ The red cross represents the first era i.e. of 1977-80 and the grey X represent the second era 1997-01.
- ➔ We can also observe that, not all the point lay on the regression line. Some points are seen to be below the regression line, whereas some are seen to be above it.

C. Summary

The dataset which was used in to compile this report was harrison_data_new. After applying a filter to the original dataset, BullTroutRML2, this dataset was created. The original dataset contains information on Harrison Lake and Osprey Lake, but only Harrison Lake data is included in harrison_data_new. Further, this new dataset has 61 observations and four variables: age, fl, lake, and era. The age of the lake and the length of its fork are found to be directly proportional. The data was visually represented in such a way that it was evident that two eras were involved. When we created a regression line, we observed that not all of the points were on it. Some are found both below and above the regression line. Takeaways of every graph are explained in detail in the above pages.

References

1. **Filter Data by multiple conditions in R using dplyr.** Geeksforgeeks. Retrieved January 25th,2022,from <https://www.geeksforgeeks.org/filter-data-by-multiple-conditions-in-r-using-dplyr>
2. **How to check data structure of an object in R.** Retrieved November 23rd,2021,from [str in R: How to Check Data Structure of Object in R \(r-lang.com\)](https://www.r-lang.com/str-in-r-how-to-check-data-structure-of-object-in-r/)
3. **Scatter Plot in R using ggplot2.** Retrieved December 25th,2021, from, <https://www.guru99.com/r-scatter-plot-ggplot2.html>
4. **ggplot title, subtitle, and caption.** Retrieved November 11th,2021, from [GGPlot Title, Subtitle and Caption : The Ultimate Guide - Datanovia](https://datanovia.com/ggplot-title-subtitle-and-caption-the-ultimate-guide/)
5. **Histograms in R Language.** Retrieved December 09th,2021, from <https://www.geeksforgeeks.org/histograms-in-r-language/>
6. **PCH In R Best Tips.** Retrieved December 25th,2021, from [PCH in R Best Tips - Datanovia](https://datanovia.com/pch-in-r-best-tips/)

Appendix

Module-2-Project.R

mohil

2022-01-27

```
r = getOption("repos")
r["CRAN"]="http://cran.us.r-project.org"
options(repos=r)

#1. Print "Plotting Basics: Lastname"
print("Plotting Basics: Mohile")

## [1] "Plotting Basics: Mohile"

#2.installing required packages
# install.packages('FSA')
# install.packages('FSAdata')
# install.packages('magrittr')
# install.packages('dplyr')
# install.packages('plotrix')
# install.packages('ggplot2')
# install.packages('moments')
#.importing required libraries
install.packages('FSA')

## Installing package into 'C:/Users/mohil/OneDrive/Documents/R/win-library/4
.1'
## (as 'lib' is unspecified)

## package 'FSA' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mohil\AppData\Local\Temp\RtmpCmmfjT\downloaded_packages

install.packages('FSAdata')

## Installing package into 'C:/Users/mohil/OneDrive/Documents/R/win-library/4
.1'
## (as 'lib' is unspecified)

## package 'FSAdata' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mohil\AppData\Local\Temp\RtmpCmmfjT\downloaded_packages

install.packages('magrittr')
```

```

## Installing package into 'C:/Users/mohil/OneDrive/Documents/R/win-library/4
.1'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##       binary source needs_compilation
## magrittr 2.0.1 2.0.2                TRUE
##
##   Binaries will be installed
## package 'magrittr' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'magrittr'

## Warning in file.copy(savedcopy, lib, recursive = TRUE):
## problem copying C:\Users\mohil\OneDrive\Documents\R\win-
## library\4.1\00LOCK\magrittr\libs\x64\magrittr.dll
## to C:\Users\mohil\OneDrive\Documents\R\win-
## library\4.1\magrittr\libs\x64\magrittr.dll: Permission denied

## Warning: restored 'magrittr'

##
## The downloaded binary packages are in
## C:\Users\mohil\AppData\Local\Temp\RtmpCmmfjT\downloaded_packages

install.packages('dplyr')

## Installing package into 'C:/Users/mohil/OneDrive/Documents/R/win-library/4
.1'
## (as 'lib' is unspecified)

## package 'dplyr' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'dplyr'

## Warning in file.copy(savedcopy, lib, recursive = TRUE):
## problem copying C:\Users\mohil\OneDrive\Documents\R\win-
## library\4.1\00LOCK\dplyr\libs\x64\dplyr.dll to C:
## \Users\mohil\OneDrive\Documents\R\win-library\4.1\dplyr\libs\x64\dplyr.dll
## :
## Permission denied

## Warning: restored 'dplyr'

##
## The downloaded binary packages are in
## C:\Users\mohil\AppData\Local\Temp\RtmpCmmfjT\downloaded_packages

install.packages('plotrix')

```

```
## Installing package into 'C:/Users/mohil/OneDrive/Documents/R/win-library/4
.1'
## (as 'lib' is unspecified)

## package 'plotrix' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mohil\AppData\Local\Temp\RtmpCmmfjT\downloaded_packages

install.packages('ggplot2')

## Installing package into 'C:/Users/mohil/OneDrive/Documents/R/win-library/4
.1'
## (as 'lib' is unspecified)

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mohil\AppData\Local\Temp\RtmpCmmfjT\downloaded_packages

install.packages('moments')

## Installing package into 'C:/Users/mohil/OneDrive/Documents/R/win-library/4
.1'
## (as 'lib' is unspecified)

## package 'moments' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mohil\AppData\Local\Temp\RtmpCmmfjT\downloaded_packages

library('FSA')

## ## FSA v0.9.1. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.

library('FSAdat')

## ## FSAdat v0.3.8. See ?FSAdat to find data for specific fisheries analys
es.

library('magrittr')
library('dplyr')

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library('plotrix')
```

```
library('ggplot2')
```

```
library('moments')
```

#3.Loading the data

```
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
## 62 8 360 Osprey 1977-80
## 63 8 357 Osprey 1977-80
## 64 7 357 Osprey 1977-80
## 65 7 329 Osprey 1977-80
## 66 6 385 Osprey 1977-80
## 67 6 323 Osprey 1977-80
## 68 5 369 Osprey 1977-80
## 69 5 326 Osprey 1977-80
## 70 4 357 Osprey 1977-80
## 71 4 326 Osprey 1977-80
## 72 4 258 Osprey 1977-80
## 73 4 239 Osprey 1977-80
## 74 3 221 Osprey 1977-80
## 75 3 258 Osprey 1977-80
## 76 3 276 Osprey 1977-80
## 77 11 688 Osprey 1997-01
## 78 10 369 Osprey 1997-01
## 79 9 400 Osprey 1997-01
## 80 8 381 Osprey 1997-01
## 81 8 332 Osprey 1997-01
## 82 7 394 Osprey 1997-01
## 83 7 388 Osprey 1997-01
## 84 7 354 Osprey 1997-01
## 85 7 320 Osprey 1997-01
## 86 6 320 Osprey 1997-01
## 87 6 347 Osprey 1997-01

```

```
## 88  6 360  Osprey 1997-01
## 89  5 354  Osprey 1997-01
## 90  5 335  Osprey 1997-01
## 91  5 313  Osprey 1997-01
## 92  5 289  Osprey 1997-01
## 93  4 313  Osprey 1997-01
## 94  4 298  Osprey 1997-01
## 95  3 279  Osprey 1997-01
## 96  3 273  Osprey 1997-01
```

#4. Printing the first and last 3 records from the BullTroutRMS2 dataset
`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
```

#5. Remove all records except those from Harrison Lake
`harrison_data_new <- filter(BullTroutRML2, lake=='Harrison')`
`harrison_data_new`

```
##   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
```

#6. Display the first and last 5 records from the filtered BullTroutRML2 data set

```
head(harrison_data_new, n=5)
```

```
##   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
```

```
tail(harrison_data_new,n=5)
```

```
##   age  fl    lake    era
## 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
```

#7. Display the structure of the filtered BullTroutRML2dataset

```
str(harrison_data_new)
```

```
## '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 ...
```

#8. Display the summary of the filtered BullTroutRML2dataset

```
summary(harrison_data_new)
```

```
##      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
```

#9. 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"
# Y axis label is "Age (yrs)"
# X axis label is "Fork Length (mm)"
# Use a small filled circle for the plotted data points
install.packages('ggplot2')
```

```
## Warning: package 'ggplot2' is in use and will not be installed
```

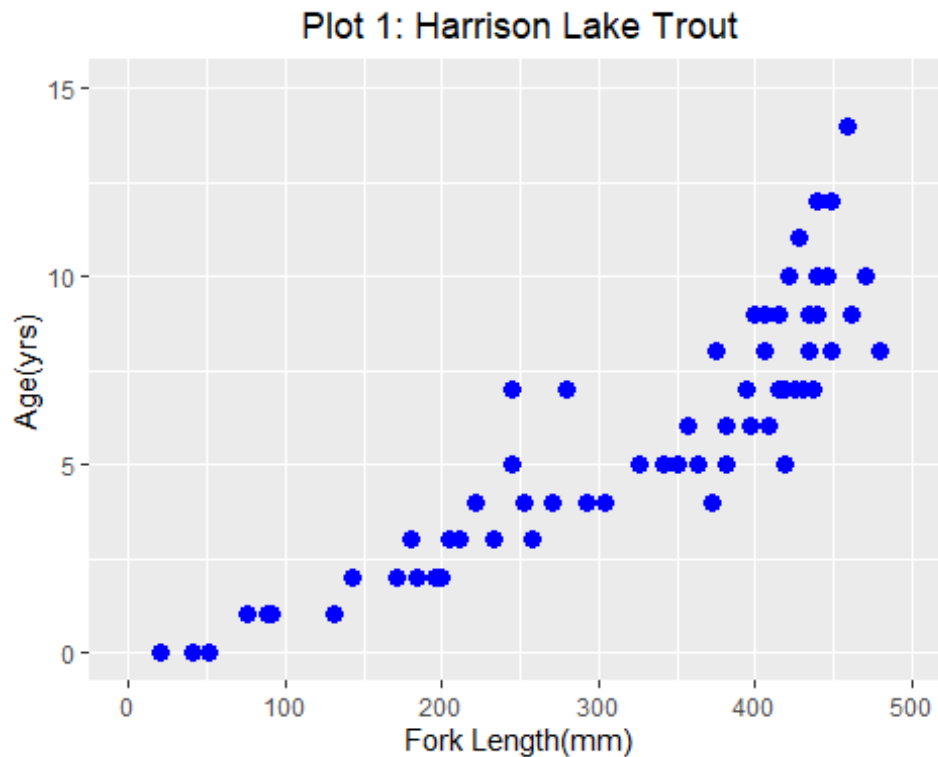
```
library('ggplot2')
```

```
?ggplot
```

```
## starting httpd help server ...
```

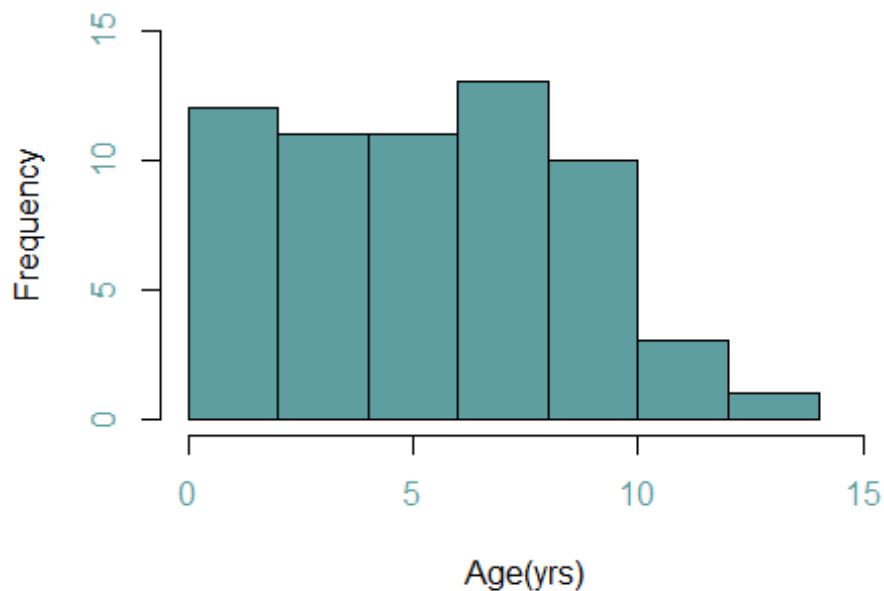
```
## done
```

```
scatter_plot <- ggplot(harrison_data_new, aes(x=fl,y=age)) +
  geom_point(colour = "blue",size = 3) +
  xlim(0,500) + ylim(0,15) + labs(x="Fork Length(mm)",
  y="Age(yrs)",title="Plot 1: Harrison Lake Trout")+
  theme(plot.title = element_text(hjust = 0.5))
scatter_plot
```



```
#10. Plot an "Age" histogram with the following specifications
#Y axis label is "Frequency"
#X axis label is "Age (yrs)"
#Title of the histogram is "Plot 2: Harrison Fish Age Distribution"
#X and Y axis limits is 0, 15
#The color of the frequency plots is "cadetblue"
#The color of the Title is "cadetblue"
#to plot the scatter plot graph
hist(harrison_data_new$age,xlab = "Age(yrs)", ylab = "Frequency",
      main = "Plot 2:Harrison Fish Age Distribution", xlim = c(0,15),
      ylim = c(0,15),col = 'cadetblue',col.axis='cadetblue',col.main = 'c
      adetblue')
```

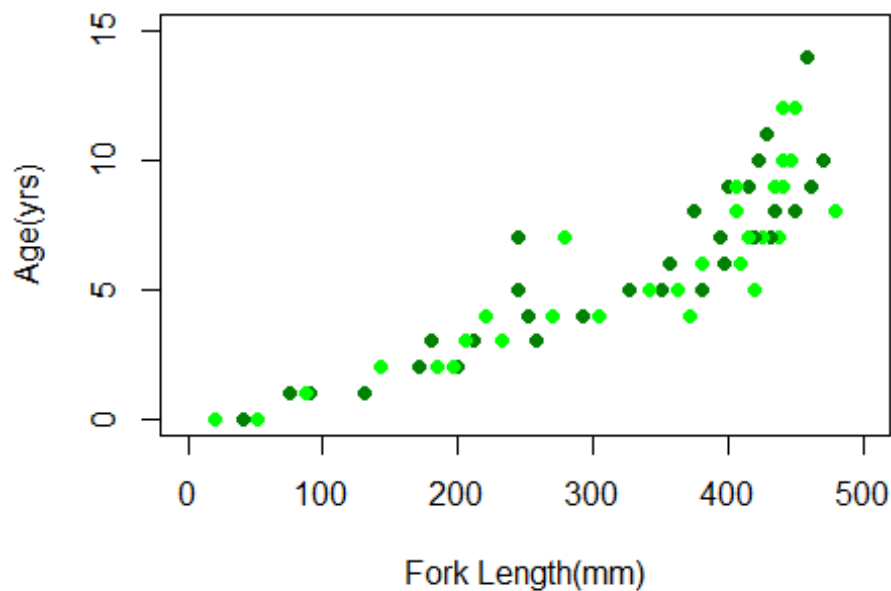
Plot 2:Harrison Fish Age Distribution



```
f1 <- harrison_data_new$f1
age <- harrison_data_new$age
#11. Create an overdense plot using the same specifications as the previous s
catterplot. But,
#Title the plot "Plot 3: Harrison Density Shaded by Era"
#Y axis label is "Age (yrs)"
#Y axis limits are 0 to 15
#X axis label is "Fork Length (mm)"
#X axis limits are 0 to 500
#include two levels of shading for the "green" data points.
#Plot solid circles as data points
```

```
overdense_plot <- plot(age ~ f1,
  main = "Plot 3:Harrison Density Shaded by Era",
  xlab = "Fork Length(mm)",
  ylab = "Age(yrs)",
  xlim = c(0,500),
  ylim = c(0,15),
  pch = 16,
  col = rgb(0,(1:2)/2,0))
```

Plot 3: Harrison Density Shaded by Era



#12. Create a new object called "tmp" that includes the first 3 and last 3 records of the BullTroutRML2 data set.

```
tmp <- headtail(BullTroutRML2,n=3)
tmp
```

```
##   age  fl   lake   era
## 1  14 459 Harrison 1977-80
## 2  12 449 Harrison 1977-80
## 3  10 471 Harrison 1977-80
## 94  4 298  Osprey 1997-01
## 95  3 279  Osprey 1997-01
## 96  3 273  Osprey 1997-01
```

#13. Display the "era" column (variable) 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
```

#14. Create a pchs vector with the argument values for + and x.

```
pchs <- c('+','x')
pchs
```

```
## [1] "+" "x"
```

#15. Create a cols vector with the two elements "red" and "gray60"

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

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

#16. Convert the tmp era values to numeric values.
conversion <- as.numeric(tmp$era)
conversion

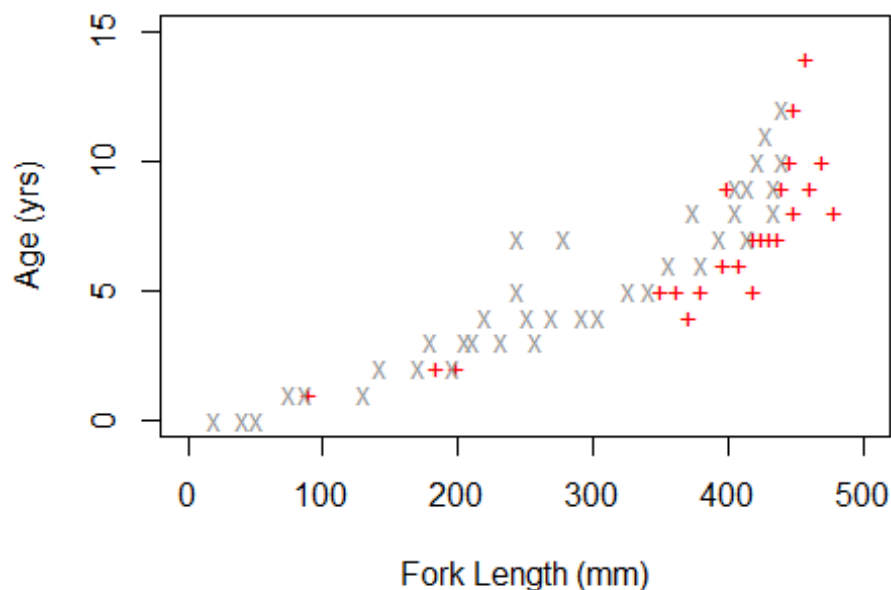
## [1] 1 1 1 2 2 2

#17. Initialize the cols vector with the tmp era values\
tmp$era <- cols
tmp$era

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

#18. Create a plot of "Age (yrs)" (y variable) versus "Fork Length (mm)" (x variable) with the following specifications:
#Title of graph is "Plot 4: Symbol & Color by Era"
#Limit of x axis is (0,500)
#Limit of y axis is (0,15)
#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= harrison_data_new, main = "Plot 4: Symbol & Color by Era",
      ylab = "Age (yrs)", ylim = c(0,15), xlab = "Fork Length (mm)", xlim = c(0,
500),
      pch=pchs[harrison_data_new$era] , col=cols[harrison_data_new$era])
```

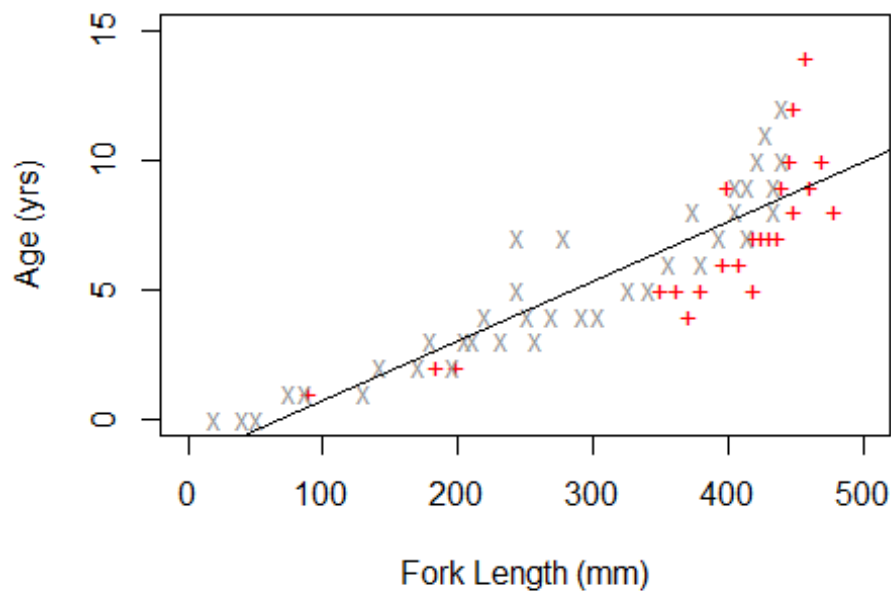
Plot 4: Symbol & Color by Era



#19. Plot a regression line overlay on Plot 4 and title the new graph "Plot 5: Regression Overlay".

```
plot(age~fl, data= harrison_data_new, main = "Plot 5: Regression Overlay",
      ylab = "Age (yrs)", ylim = c(0,15), xlab = "Fork Length (mm)", xlim = c(0,
500),
      pch=pchs[harrison_data_new$era] , col=cols[harrison_data_new$era])
r1 <- lm(age~fl,data = harrison_data_new)
abline(r1)
```

Plot 5: Regression Overlay



#20. Place a Legend of on Plot 5 and call the new graph "Plot 6: :Legend Overlay"

```
plot(age~fl, data= harrison_data_new, main = "Plot 6: Legend Overlay",
      ylab = "Age (yrs)", ylim = c(0,15), xlab = "Fork Length (mm)", xlim = c(0,
500),
      pch=pchs[harrison_data_new$era] , col=cols[harrison_data_new$era])
r1 <- lm(age~fl,data = harrison_data_new)
abline(r1)
?legend
harrison_data_new$era #to find the levels.
```

```
## [1] 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-
80
## [10] 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-
80
## [19] 1977-80 1977-80 1977-80 1977-80 1977-80 1997-01 1997-01 1997-01 1997-
01
## [28] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-
```

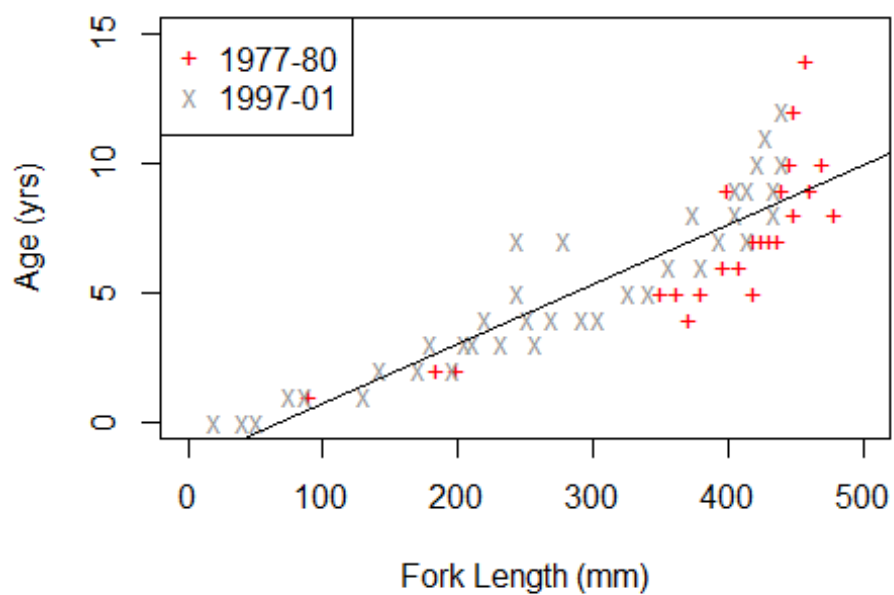
```

01
## [37] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-
01
## [46] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-
01
## [55] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
## Levels: 1977-80 1997-01

legend("topleft", legend = c("1977-80", "1997-01"), pch = pchs, col = cols)

```

Plot 6: Legend Overlay



#21. Commit your code into your github/gitlab repo.

https://github.com/SanjanaMohile/ALY6000_Module2.git