# ALY 6000: Introduction to Analytics

## Module 2 - Project Assignment

## Executive Summary Report 2

Devi Somalinga Bhuvanesh

24th January 2023

## Key Findings

BullTroutRML2 dataset consists of information related to the fishes residing in two lakes, i.e., Harrison Lake and Osprey Lake. It has information of Age (in years) and Fork Length (in mm) of all the fishes. The data has been captured from two different sets of eras.

### 1. Descriptive Analysis

#### Statistical analysis for entire dataset of BullTroutRML2

```
> summary(dataset)  #to check summary of dataset which includes Min, Max, Mean, Median,
1st and 3rd Quartile
      age               fl            lake               era
 Min.   : 0.000   Min.   : 20.0   Length:96          Length:96
 1st Qu.: 4.000   1st Qu.:258.0   Class :character   Class :character
 Median : 6.000   Median :352.5   Mode  :character   Mode  :character
 Mean   : 5.771   Mean   :326.1
 3rd Qu.: 8.000   3rd Qu.:406.0
 Max.   :14.000   Max.   :688.0
> str(dataset)      #to check structure of dataset
'data.frame':   96 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: chr  "Harrison" "Harrison" "Harrison" "Harrison" ...
 $ era : chr  "1977-80" "1977-80" "1977-80" "1977-80" ...
> var(dataset$age)  #Variance of age in dataset
[1] 8.557456
> sd(dataset$age)   #Standard Deviation of age in dataset
[1] 2.925313
> var(dataset$fl)   #Variance of fl in dataset
[1] 12589.34
> sd(dataset$fl)    #Standard Deviation of fl in dataset
[1] 112.2022
> skewness(dataset$age)  #Skewness of age in dataset
[1] 0.212223
> skewness(dataset$fl)   #Skewness of fl in dataset
[1] -0.5103268
> kurtosis(dataset$age)  #Kurtosis of age in dataset
[1] 2.735144
> kurtosis(dataset$fl)   #Kurtosis of fl in dataset
[1] 3.914769
```

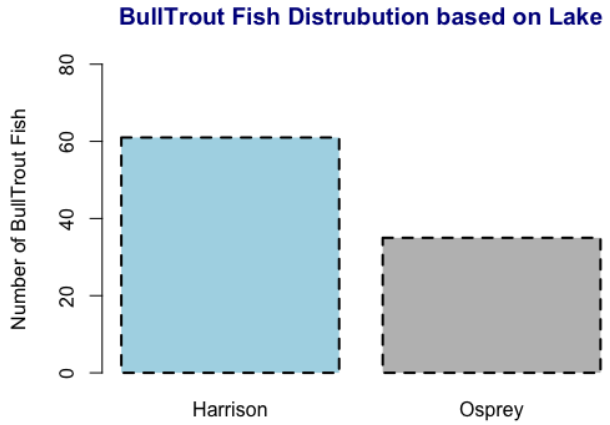
- The above statistical information conveys that BullTroutRML2 dataset consists of information related to 96 fishes in 4 variables namely, age, fl, lake, and era.
- The youngest of fish in the dataset is 0 years old with the smallest fork length of 20mm whereas the oldest fish is 14 years old with a fork length of 459mm. The highest fork length observed in the dataset is 688mm which is a clear outlier in Osprey Lake of (ref plot 3).
- The average age of these 96 fishes is 5.771 years with a standard deviation of 2.925 years and the average fork length is 326.1 mm with a standard deviation of 112.202mm.
- While looking at the shape and distribution of dataset, the age of the fishes has a skewness of 0.21 (positively skewed) which indicates that a greater number of fishes are in the age group below 5.7 years (mean age). On the other hand, the fork length of the fishes is negatively skewed (-0.51) indicating that a greater number of fishes have fork length greater than 326.1 (mean of fork length).
- Kurtosis value with more than 3 indicates that it has sharper peak compared to standard normal distribution. According to the above statistical information, the fork length distribution of the fish has a sharper peak while the age distribution of the fish has a smoother peak compared to standard normal distribution.

***Kindly Note:*** *Similar descriptive analysis has been done for Harrison and Osprey database separately which is included in the Appendix and the key findings are used to analyze the visual graph in the next section.*

## 2. Visualization

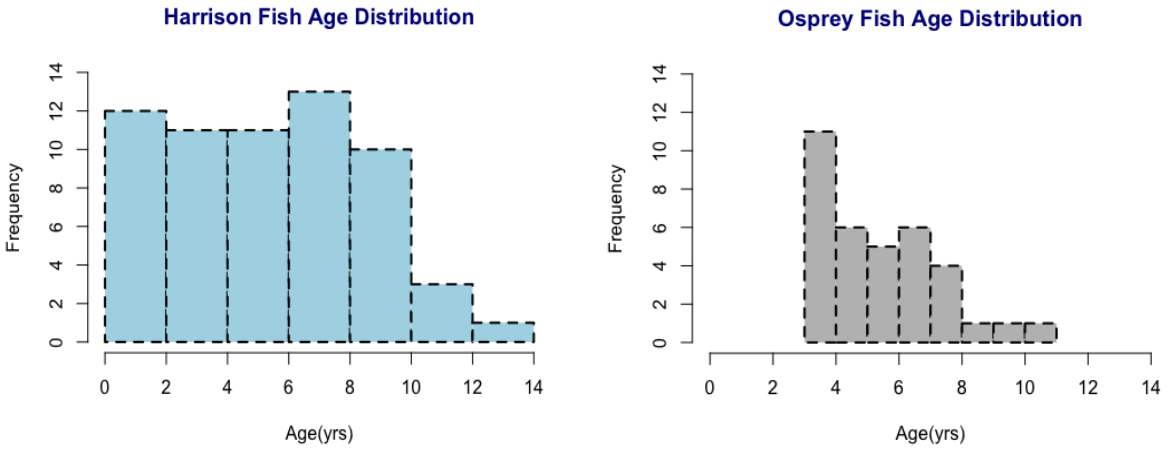
In the Visualization section, a comparative study about fishes in Harrison Lake and Osprey Lake is presented.

**Plot 1: Bar graph of BullTrout Fish Distribution based on Lake**




Out of 96 fish, the bar graph indicates that around 2/3rd of the fish, 61, are in Harrison Lake while 35 fish are residing in Osprey Lake.

**Plot 2: Comparison of Age Distribution of Fish in Harrison Lake Vs Osprey Lake**



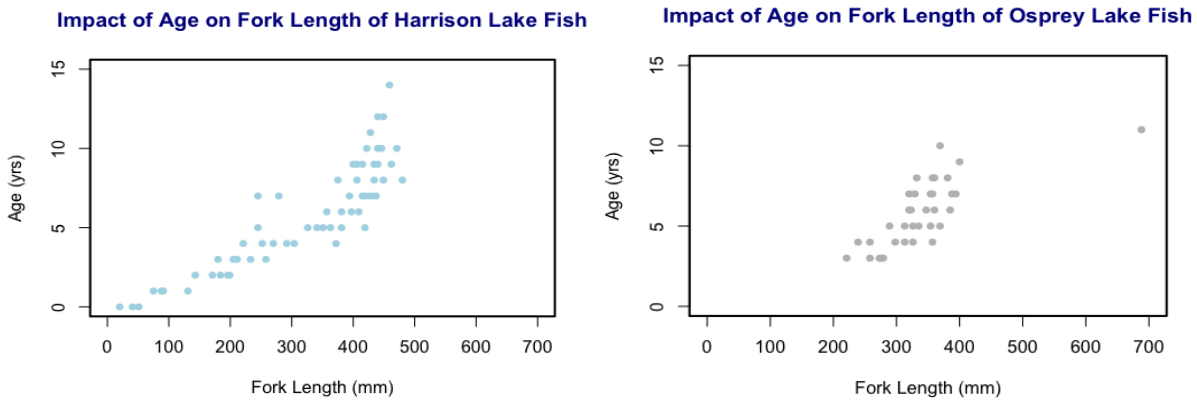
The histogram depicts the frequency distribution of age in Harrison Lake and Osprey Lake.

While looking at the ages of fish in both the lakes, it can be clearly observed that the ages range from 0 to 14 years old in Harrison Lake, which is the minimum and maximum age of the entire database respectively. Alternatively, in Osprey Lake, there are no fish that are younger than 3 years old and older than 11 years old.

In Harrison Lake, the age group of 6-8 years has the highest count of fish followed by 0-2 years. In Osprey Lake, the age group of 3-4 years has the highest count of fish.

However, it is evident that with the increase in age, the number of fish reduces in both the lakes which could be an indicator of average life expectancy of bull trout fish.

**Plot 3: Impact of Age on Fork Length of the fish in Harrison Lake Vs Osprey Lake**




The scatter plot represents how the fork length of fish varies with respect to the age of the fish in both the lakes.

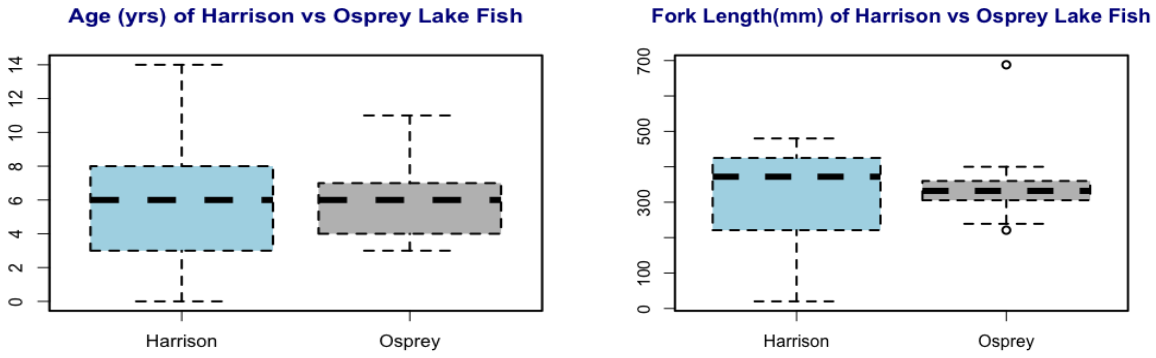
Since Harrison Lake has 61 fish ranging from 0 to 14 years old, the scatter plots cover a wide range with fork length starting from 20 to 480 mm (Refer Descriptive Analysis of Harrison Data in Appendix). In Osprey Lake, 35 fish ranging from 3 to 11 years have a fork length ranging from 221 to 688 mm (Refer Descriptive Analysis of Osprey Data in Appendix). 688 mm in Osprey Lake is an outlier in the dataset while all other fish have a fork length ranging from 221 to 400 mm.

In contrast, it can also be observed that an 8-year-old fish in Harrison Lake has the highest fork length with 480 mm which is higher than that of oldest fish in the lake. Also, it can be noted that 11-year-old fish in Osprey has 688mm of fork length which is higher than that of 14 years old fish in Harrison Lake.

At few instances, the fork length also varies for fishes in same age. For example, three 0-year-old fishes in Harrison Lake have fork lengths of 20mm, 41mm, and 51mm. Similarly, there are five fishes in Osprey Lake who are 3 years old with varying fork length ranging from 221 to 279mm.

Therefore, apart from age, there are other factors which influence the fork length of the fish.

**Plot 4: Comparison of Age and Fork Length of Fish in Harrison Lake Vs Osprey Lake**




Using Box plots, the age and fork length of fish in both the lakes has been compared.

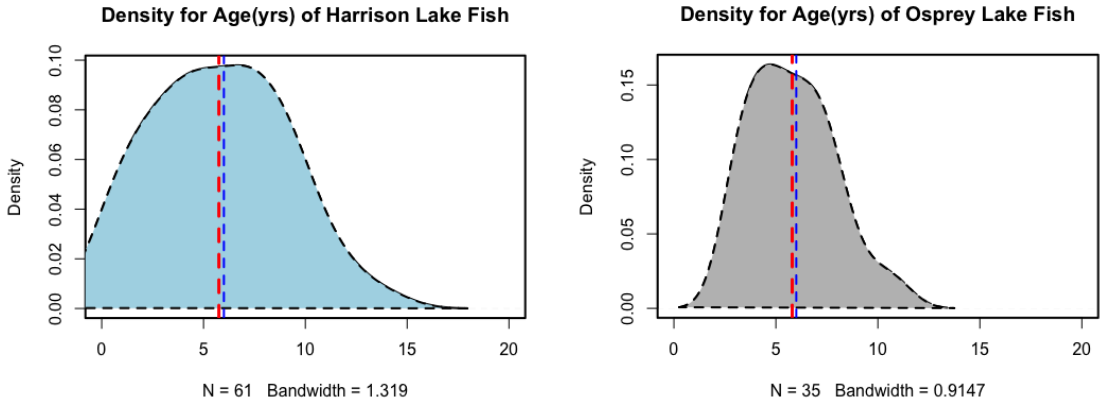
While looking at the age box plot, even though the range of age varies in both the lakes, the median age of both the fishes are same, i.e., 6 years. In Harrison Lake, the box plot indicates that 25% of the fishes are in the age group within 0-3 years (1st Quartile), 50% within 0-6 years, and 75% within 0-8 years (3rd Quartile) (Refer Descriptive Analysis of Harrison Data in Appendix). Similarly, 25% of the fish in Osprey Lake are in the age group 3-4 years, 50% in the 3-6 years, and 75% in 3-7 years old. (Refer Descriptive Analysis of Osprey Data in Appendix)

In the fork length box plot, the median fork length of fish in Harrison Lake is 372mm and Osprey Lake is 332mm which is almost near even though the range of fork length varies. 25% of the fish fork length in Harrison Lake ranges from 20-221 mm, 50% from 20-319 mm, and 75% from 20-425mm (Refer Descriptive Analysis of Harrison Data in Appendix). On the other hand, the fork length of fish in Osprey Lake greatly varies and begins from the 1st Quartile point of Harrison Lake, i.e., from 221mm (which makes sense since min age in Osprey is 3 years old). 25% of the fish in Osprey ranges have fork length from 221-305.5mmm, 50% from 221-332mm, and 75% from 221-360mm (Refer Descriptive Analysis of Osprey Data in Appendix). Osprey has the highest fork length of 688m.

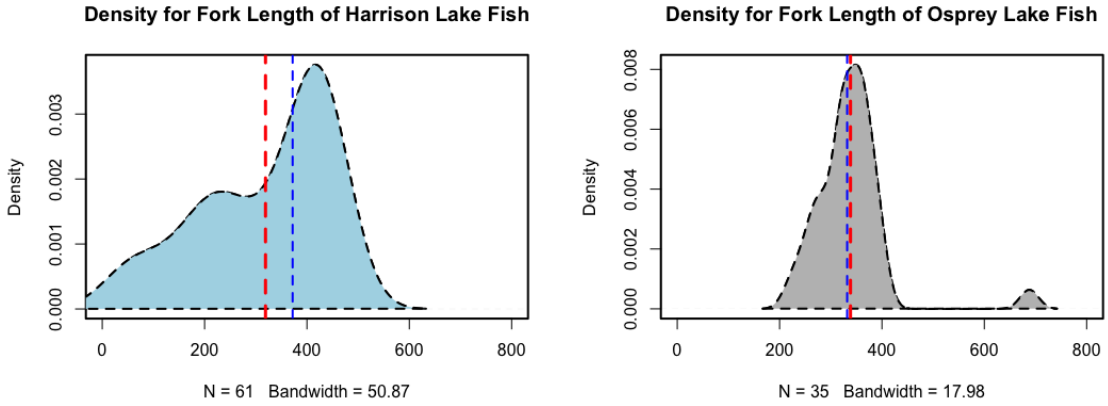
**Plot 5: Density for Age and Fork Length in Harrison vs Osprey Lake**

The density plot below for Age and Fork length shows the shape and distribution of respective data along with mean and median highlighted. Skewness and Kurtosis can also be observed in the same.

*Kindly Note: Mean is Red in color and Median is Blue in color*



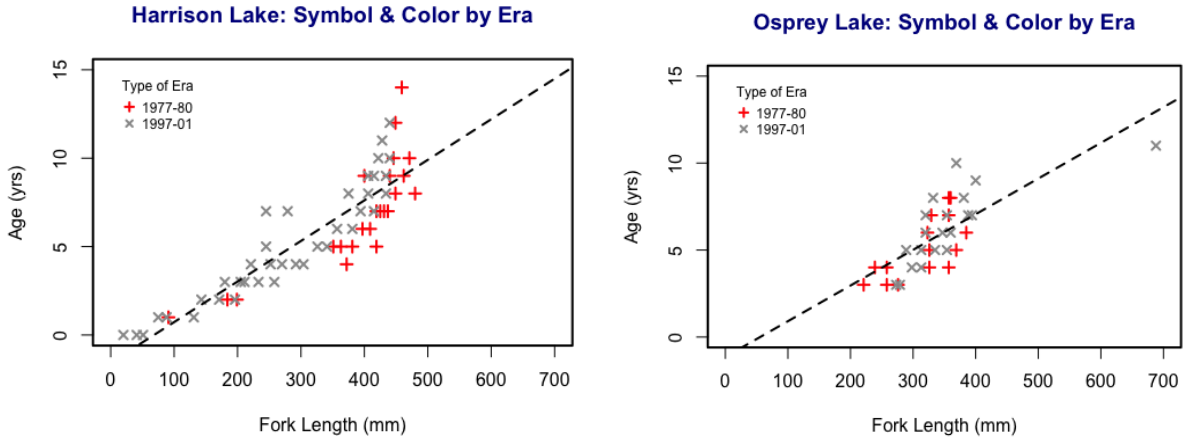

The density age plot is almost reaching a symmetric curve with nearly equal values of mean and median. Harrison Lake has positive skewness with 0.16 and 0.51 in Osprey Lake which indicates majority of the fishes fall below the average age, I.e., 5.8 years in both the lake. The kurtosis value is less than 3 in both the lakes. (Refer Descriptive Analysis of Harrison and Osprey Data in Appendix)




The graphical representation of density plot for fork length of fish greatly varies from the age density plot.

In Harrison Lake, the distribution is negatively skewed (-0.73) which indicates that there are more fish with fork length greater than average fork length, i.e., 319 mm. Here, the median (372) is greater than mean, as expected. The kurtosis is less than 3. On the other hand, the kurtosis for Osprey is 14 which aligns with sharp peak. The fork length of Osprey Lake fish is positively skewed with 2.67 which indicates more fish fall below the average fork length, i.e., 338.5mm. The median is 332 mm which is slightly less than the mean value. The outlier is clearly visible in the density plot for Osprey Lake (tiny peak between 600-800 mm). (Refer Descriptive Analysis of Harrison and Osprey Data in Appendix)

**Plot 6: Age and Frequency of fish based on Era – Harrison vs Osprey Lake**




The above scatter plot shows the nature of fish in different era, i.e., 1977-80 and 1997-01 in both the lakes.

In Harrison Lake, the youngest fish with a minimum fork length was from the period of 1997-01 era while in Osprey Lake, it is from 1977-80 era. Also, the oldest fish and longest fork length in Harrison Lake is from 1977-80 era while it is not so in Osprey Lake.

In Harrison Lake, 1977-80 era most of the fish had higher fork length approximately in the range of 350-500mm (with only 3 fish below this range). This aligns with the fact that most of the fish in this era are older than approx. 3-4 years old. 1997-01 era has distribution across all ranges in terms of both age and fork length. Unlike Harrison, in Osprey Lake, distribution of age and fork length are very similar for both the eras.

In addition, 38 and 20 fish are from 1997-01 era from Harrison Lake and Osprey Lake respectively which indicates more number of fishes were found in recent period.

**3. Summary**

There are a total of 96 fish in the dataset, with 61 from Harrison Lake and 35 from Osprey Lake. The main takeaway from the data is that apart from the age of the fish, there are also other factors which influence the fork length of the fish. The age distribution of fish in both the lakes reflects positive skewness (slightly above zero) which indicates that more younger fish were documented in the dataset compared to older ones (which could also be an indicative of the life expectancy of bull trout which is around 10-12 years). The fork length for an adult fish (4-7 years) is generally in the range of 300-500 mm with one outlier at 688 mm.

# Bibliography


- Kabacoff, R. (2015). *R in Action: Data Analysis and Graphics with R* (2nd ed.). Manning.
- ALY 6000.20926.202325. Learning to use R - Instructor Videos Lesson 6-10. https://northeastern.instructure.com/courses/131162/pages/learning-to-use-r-instructor-videos-2?module_item_id=8426810
- Coder, R. (2021, March 31). *Smooth scatter plot in R*. R CHARTS | a Collection of Charts and Graphs Made With the R Programming Language. https://r-charts.com/correlation/smooth-scatter-plot/
- *how to load csv files in google colab for R?* (2019, October 8). Stack Overflow. https://stackoverflow.com/questions/58285503/how-to-load-csv-files-in-google-colab-for-r
- GeeksforGeeks. (2021b, June 17). *Overlay Density Plots in Base R*. https://www.geeksforgeeks.org/overlay-density-plots-in-base-r/
- S. (2015, June 29). *Part II: Introduction to Base-R Graphing | Scott Withrow*. Scott Withrow | Assistant Professor|Koc University. Retrieved January 24, 2023, from https://mysite.ku.edu.tr/swithrow/2015/06/29/part-ii-introduction-to-base-r-graphing/
- F. (2021, December 18). *Import CSV Files into R Step-by-Step Guide*. R-bloggers. Retrieved January 21, 2023, from https://www.r-bloggers.com/2021/12/import-csv-files-into-r-step-by-step-guide/
- *plot legends without border and with white background*. (2012, April 11). Stack Overflow. https://stackoverflow.com/questions/10108073/plot-legends-without-border-and-with-white-background
- GeeksforGeeks. (2022, January 25). *Skewness and Kurtosis in R Programming*. https://www.geeksforgeeks.org/skewness-and-kurtosis-in-r-programming/
- Coder, R. (2020, December 12). *Barplot in R*. R CODER. https://r-coder.com/barplot-r/
- Li, D. (2020, October 23). *Chapter 19 Scatterplots and Best Fit Lines - Two Sets | Basic R Guide for NSC Statistics*. https://bookdown.org/dli/rguide/scatterplots-and-best-fit-lines-two-sets.html
- *R: Add Legends to Plots*. (n.d.). https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/legend.html
- Simplilearn. (2020, June 18). *R Tutorial For Beginners Part - 1 | R Programming For Beginners | R Language Tutorial | Simplilearn* [Video]. YouTube. https://www.youtube.com/watch?v=7NLPPFU0O3w
- *headTail function - RDocumentation*. (n.d.). https://www.rdocumentation.org/packages/psych/versions/2.2.9/topics/headTail
- *R Graphics - Plotting*. (n.d.). https://www.w3schools.com/r/r_graph_plot.asp
- For Q11, I discussed with Ms. Manvika, a Student in Winter 2023 MPS Analytics course on 23rd January 2023 over phone call.
- For Q15, I discussed with Ms. Anisha, a Student in Winter 2023 CPS on 3rd January 2023 over phone call.

# Appendix

**> #Q1: Print name with prefix "Plotting Basics:"**

> print("Plotting Basics: Devi Somalinga Bhuvanesh") #to print the text with prefix

[1] "Plotting Basics: Devi Somalinga Bhuvanesh"

**> #Q2: Import libraries including:**

> #plyr, FSA, FSAdata, magrittr, dplyr, plotrix, ggplot2, and moments

> install.packages("plyr")  #to install new package

Error in install.packages : Updating loaded packages

> install.packages("plyr")

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/plyr_1.8.8.tgz'

Content type 'application/x-gzip' length 1015309 bytes (991 KB)

==================================================

downloaded 991 KB

The downloaded binary packages are in

/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages

> library(plyr)  #to import the new package in the library

> install.packages("FSA")

Error in install.packages : Updating loaded packages

> install.packages("FSA")

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/FSA_0.9.3.tgz'

Content type 'application/x-gzip' length 1117449 bytes (1.1 MB)

==================================================

downloaded 1.1 MB

The downloaded binary packages are in

```
/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages
> library(FSA)
> install.packages("FSAdata")
Error in install.packages : Updating loaded packages
> install.packages("FSAdata")
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/FSAdata_0.4.0.tgz'
Content type 'application/x-gzip' length 957828 bytes (935 KB)
==================================================
downloaded 935 KB


The downloaded binary packages are in
/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages
> library(FSAdata)
> install.packages("magrittr")
Error in install.packages : Updating loaded packages
> install.packages("magrittr")
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/magrittr_2.0.3.tgz'
Content type 'application/x-gzip' length 227506 bytes (222 KB)
==================================================
downloaded 222 KB


The downloaded binary packages are in
/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages
> library(magrittr)
> install.packages("dplyr")
Error in install.packages : Updating loaded packages
> install.packages("dplyr")
```

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/dplyr_1.0.10.tgz'
Content type 'application/x-gzip' length 1310338 bytes (1.2 MB)
==================================================
downloaded 1.2 MB


The downloaded binary packages are in
/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages
> library(dplyr)
> install.packages("plotrix")
Error in install.packages : Updating loaded packages
> install.packages("plotrix")
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/plotrix_3.8-2.tgz'
Content type 'application/x-gzip' length 1137125 bytes (1.1 MB)
==================================================
downloaded 1.1 MB


The downloaded binary packages are in
/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages
> library(plotrix)
> install.packages("ggplot2")
Error in install.packages : Updating loaded packages
> install.packages("ggplot2")
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ggplot2_3.4.0.tgz'
Content type 'application/x-gzip' length 4211307 bytes (4.0 MB)
==================================================
downloaded 4.0 MB
```

The downloaded binary packages are in

/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages

> library(ggplot2)

> install.packages("moments")

Error in install.packages : Updating loaded packages

> install.packages("moments")

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/moments_0.14.1.tgz'

Content type 'application/x-gzip' length 54374 bytes (53 KB)

==================================================

downloaded 53 KB

The downloaded binary packages are in

/var/folders/1y/qwg6z9nj78nfkv1gf3qrgts40000gp/T//Rtmpzww5Ll/downloaded_packages

> library(moments)

**> #Q3: Load the BullTroutRML2 dataset**

> dataset <- read.csv('/Users/devi/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 2/BullTroutRML2.csv', header=TRUE, sep=",") #to read the file from the directory

> dataset  #to view the loaded dataset for verification

| | 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

>

**> #Q4: Print the first and last 3 records from the dataset**

> #Method1 - The row number does not get changed

> dataset  #to view the dataset

  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
> head(dataset,n=3)     #to view first n records of dataset
  age  fl     lake     era
1  14 459 Harrison 1977-80
2  12 449 Harrison 1977-80
3  10 471 Harrison 1977-80
> tail(dataset,n=3)     #to view last n records of dataset
   age  fl   lake     era
94   4 298 Osprey 1997-01
95   3 279 Osprey 1997-01
96   3 273 Osprey 1997-01
> headtail(dataset,n=3) #to view first and last n records of dataset
```

```
  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
>
```

**> #Method2 - The row number is continuous**

```
> dataset               #to view the dataset
  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

> new1<-dataset[1:3, ]  #assigning the range of 1st three records of dataset to new1

> new1  #check output of new1

 age  fl  lake  era

1  14 459 Harrison 1977-80

2  12 449 Harrison 1977-80

3  10 471 Harrison 1977-80

> new2<-dataset[94:96, ]  #assigning the range of last three records of dataset to new2

> new2  #check output of new2

```
  age  fl   lake     era
94   4 298 Osprey 1997-01
95   3 279 Osprey 1997-01
96   3 273 Osprey 1997-01
> combine(new1,new2)      #combine new1 and new2
  age  fl     lake     era
1  14 459 Harrison 1977-80
2  12 449 Harrison 1977-80
3  10 471 Harrison 1977-80
4   4 298   Osprey 1997-01
5   3 279   Osprey 1997-01
6   3 273   Osprey 1997-01
```

**> #Q5:Filter out all records except those from Harrison Lake**

```
> harrisondata<-dataset[dataset$lake == "Harrison",] #From the dataset, extracting "Harrison" data from lake column
> harrisondata          #to view the extracted data in the assigned variable
   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(harrisondata,n=3) #filter out first n and last n data's from harrisondata

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**

> str(harrisondata) #to view the structure of Harrison dataset - class, variables, observations, values

'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: chr "Harrison" "Harrison" "Harrison" "Harrison" ...

$ era : chr "1977-80" "1977-80" "1977-80" "1977-80" ...

**> #Q8 Display the summary of the filtered dataset and save it as <t>**

> t<-summary(harrisondata) #assigning summary of Harrison dataset to t - Min, Max, Mean, Median, 1st and 3rd Quartile, Length, Class

> t

age fl lake era

Min. : 0.000 Min. : 20 Length:61 Length:61

1st Qu.: 3.000 1st Qu.:221 Class :character Class :character

Median : 6.000 Median :372 Mode :character Mode :character

Mean : 5.754 Mean :319

3rd Qu.: 8.000 3rd Qu.:425

Max. :14.000 Max. :480

**> #Q9 Create a scatterplot for "age" and "fl"**

> harrisondata #to view harrisondata

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

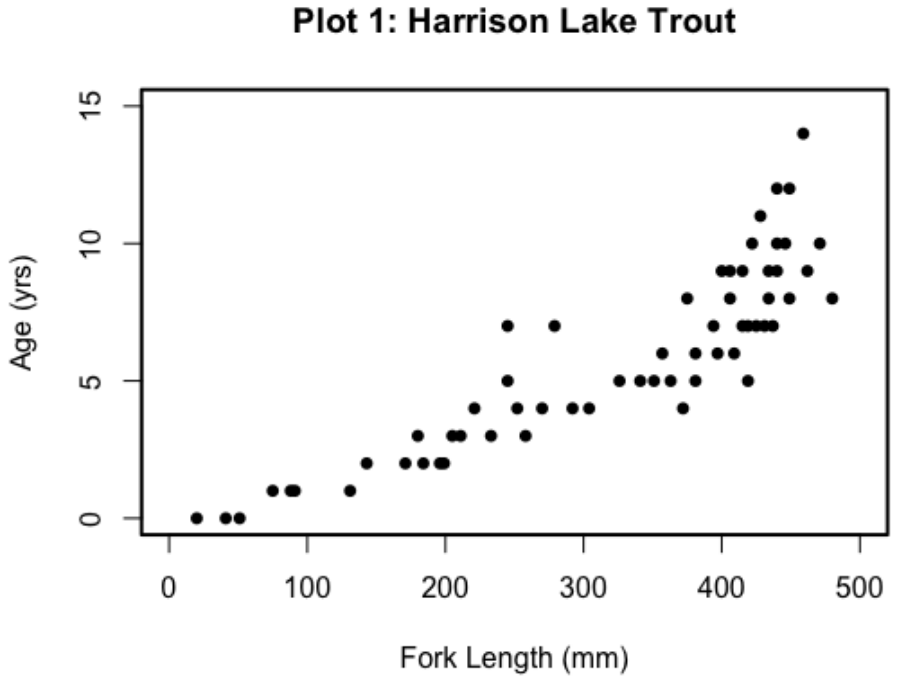
56 0 51 Harrison 1997-01

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
> plot(harrisondata$fl,harrisondata$age,
+      main="Plot 1: Harrison Lake Trout",
+      ylab="Age (yrs)", xlab="Fork Length (mm)",
+      xlim=c(0,500), ylim=c(0,15),
+      pch=20)
>
> #plotted fl in X axis and Age in Y axis from harrisondata,
> #use main to provide title,
> #use xlab and ylab too provide labels
> #use xlim and ylim to set x and y axis limits
> #pch=20 is the value to set the shape of plotted points - small filled circle
>
```



**> #Q10 Plot an "Age" histogram with the following specifications**

> harrisondata #to view harrisondata

| | 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

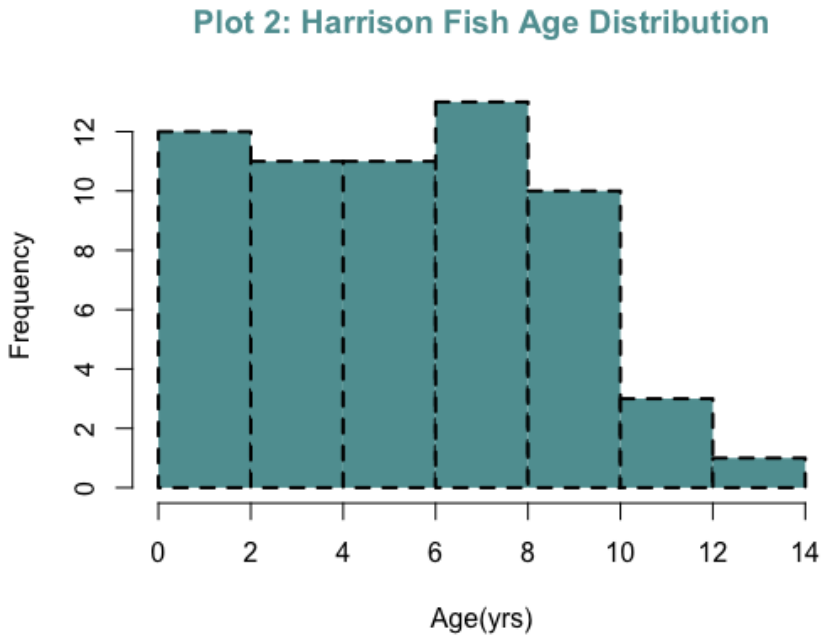
52 2 143 Harrison 1997-01

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
> hist(harrisondata$age,
+      main = "Plot 2: Harrison Fish Age Distribution",
+      col.main="cadetblue",col="cadetblue",
+      ylab="Frequency", xlab="Age(yrs)")
>
> #an histogram of age data from harrisondata is plotted
> #main is used to provide title for the graph
> #col.main is used to assign color to the title
> #col is used to assign color to the histogram
> #xlab and ylab is used to provide labels
```



```
> #Q11 Create an overdense plot
> harrisondata   #to view harrisondata
   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

> twoshade <- c("pink", "yellow") #two colors are chosen to plot in the graph

> harrisondata$era <- as.factor(harrisondata$era) #values in Era column are converted to factor to store the unique values which will help to associate with other variables

> harrisondata$era

[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

> twoshade[harrisondata$era] #the two colors are assigned to the two unique values under Era column in harrison data

[1] "pink" "pink" "pink" "pink" "pink" "pink" "pink" "pink"

[9] "pink" "pink" "pink" "pink" "pink" "pink" "pink" "pink"

[17] "pink" "pink" "pink" "pink" "pink" "pink" "pink" "yellow"

[25] "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow"

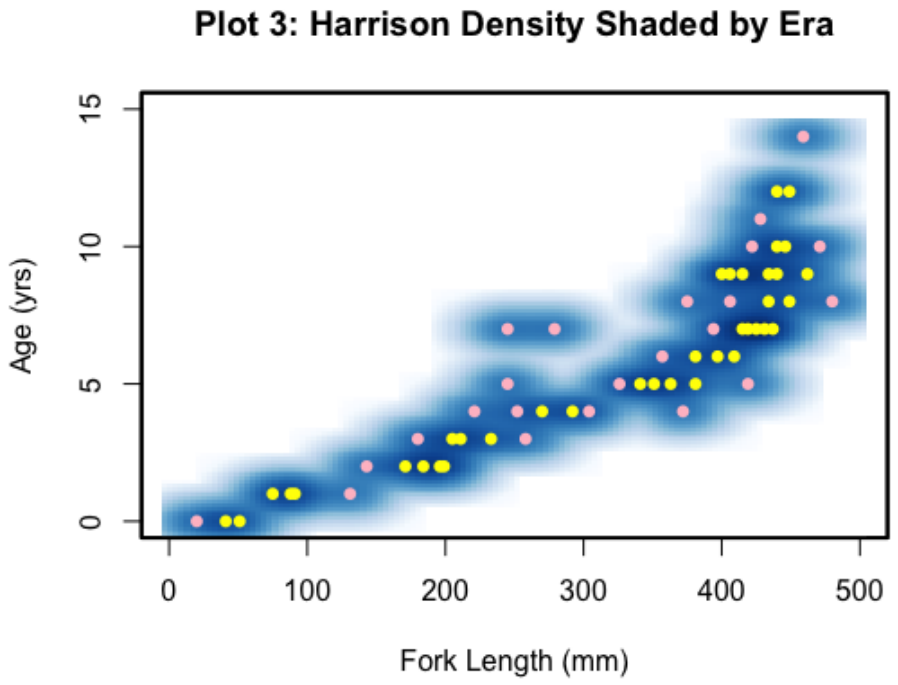
[33] "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow"

[41] "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow"

[49] "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow" "yellow"

[57] "yellow" "yellow" "yellow" "yellow" "yellow"

```
> smoothScatter(harrisondata$fl,harrisondata$age,
+             main="Plot 3: Harrison Density Shaded by Era",
+             ylab="Age (yrs)", xlab="Fork Length (mm)",
+             xlim=c(0,500), ylim=c(0,15),
+             pch=20, col=twoshade[harrisondata$era])
>
> #smoothScatter is used to show the two-dimensional kernel density estimate
> #plotted fl in X axis and Age in Y axis from harrisondata,
> #use main to provide title,
> #use xlab and ylab too provide labels
> #use xlim and ylim to set x and y axis limits
> #pch=20 is the value to set the shape of plotted points - small filled circle
> #col is used to assign two colors to the values in Era column in Harrison data
>
```



**> #Q12 New object “tmp” that includes the first 3 and last 3 records of the whole data set.**

> dataset  #to view whole dataset

| | 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

> tmp <-headtail(dataset,n=3) #extract first and last n records from dataset and assign to tmp

> tmp #to view tmp output

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

>

**> #Q13 Display the “era” column in the new “tmp” object**

> #Method1 - prints the values in Era column in column format - 4th column in tmp

> print(tmp[4])

era

1 1977-80

2 1977-80

```
3  1977-80
94 1997-01
95 1997-01
96 1997-01
>
> #Method2 - prints the values in Era column horizontally
> print(tmp[ ,"era"])
[1] "1977-80" "1977-80" "1977-80" "1997-01" "1997-01" "1997-01"
>
> #Method3 - prints the values in Era column horizontally
> print(tmp$era)
[1] "1977-80" "1977-80" "1977-80" "1997-01" "1997-01" "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)    #pch argument for + and X is represented by 3 and 4 which is assigned to pchs
> pchs              #to view output of pchs
[1] 3 4
> cols <- c("red","gray60")  #to assign two colors to cols
> cols           #to view assigned values in cols
[1] "red"    "gray60"
>

> #Q15 Convert the tmp object values to numeric values.
> #Then create a numeric numEra object from the tmp$era object
> tmp <-headtail(dataset,n=3)      #to extract first and last n data from dataset
> tmp                          #to view the extracted data which is assigned in 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

> is.numeric(tmp$age) #to check whether the values in age column in tmp is numeric

[1] TRUE

> is.numeric(tmp$fl) #to check whether the values in fl column in tmp is numeric

[1] TRUE

> tmp$lake <- as.factor(tmp$lake) #to convert lake from character to factor to create levels for unique data

> tmp$lake #to view the levels in lake column

[1] Harrison Harrison Harrison Osprey Osprey Osprey

Levels: Harrison Osprey

> tmp$lake <- as.numeric(tmp$lake) #to assign numeric values to each level in the lake column

> tmp$lake #to view the numeric values assigned in lake column

[1] 1 1 1 2 2 2

> is.numeric(tmp$lake) #to check whether the values in lake is numeric

[1] TRUE

> tmp$era <- as.factor(tmp$era) #to convert era values from date to factor to create levels for unique data

> tmp$era #to view the levels in era column

[1] 1977-80 1977-80 1977-80 1997-01 1997-01 1997-01

Levels: 1977-80 1997-01

> tmp$era <- as.numeric(tmp$era) #to assign numeric values to each level in the era column

> tmp$era #to view the numeric values assigned in era column

[1] 1 1 1 2 2 2

> is.numeric(tmp$era) #to assign numeric values to each level in the era column

```
[1] TRUE
>
```

**> #Q16 Associate the cols vector with the tmp era values**

```
>
> #To assign cols with tmp era values
> cols <- c("red","gray60")       #to assign two colors to cols
> cols                          #to view assigned values in cols
[1] "red"    "gray60"
> tmp$era <- as.factor(tmp$era)    #to convert era values from date to factor to create levels for unique data
> tmp$era                       #to view the levels/unique data in era column
[1] 1 1 1 2 2 2
Levels: 1 2
> cols[tmp$era]                  #to assign cols value to each level under era column
[1] "red"    "red"    "red"    "gray60" "gray60" "gray60"
>
> #Similarly, to ossign pchs symbols to tmp era values
> pchs <- c(3,4)                 #pch argument for + and X is represented by 3 and 4 which is assigned to pchs
> pchs                          #to view output of pchs
[1] 3 4
> pchs[tmp$era]                  #to assign pchs symbol to each level under era column
[1] 3 3 3 4 4 4
```

**> #Q17 Create a plot of “Age (yrs)” (y variable) versus “Fork Length (mm)” (x variable)**

```
> harrisondata                   #to view harrisondata
  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
> harrisondata$era <- as.factor(harrisondata$era)    #to convert the values in Era from Date to Factor to create unique levels
> harrisondata$era                #to view the output of Era column in Harrisondata as factors
[1] 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80
[10] 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80
[19] 1977-80 1977-80 1977-80 1977-80 1977-80 1997-01 1997-01 1997-01 1997-01
[28] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
[37] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
[46] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
[55] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
Levels: 1977-80 1997-01
> cols[harrisondata$era]          #assigning the colours to the unique values in Era column of Harrison data
[1] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "red"
[9] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "red"
[17] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "gray60"
[25] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[33] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[41] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[49] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[57] "gray60" "gray60" "gray60" "gray60" "gray60"
> plot(harrisondata$fl,harrisondata$age,
+     main="Plot 4: Symbol & Color by Era",
+     ylab="Age (yrs)", xlab="Fork Length (mm)",
+     xlim=c(0,500), ylim=c(0,15),
+     pch=pchs[harrisondata$era],col=cols[harrisondata$era])
>
> #Plotting the graph of fl (x-axis) and Age (y-axis) from harrisondata
```

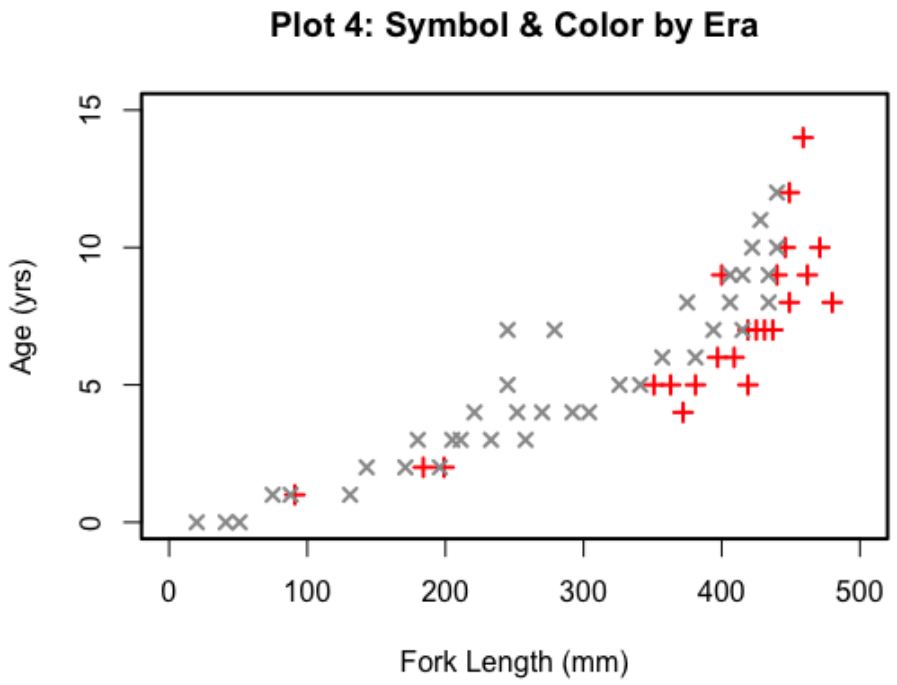
> #Providing the title using main function

> #Labels foor x and y axis are assigned using xlab and ylab

> #X and Y axis limits are set through xlim and ylim

> #Unique values under Era column are linked with the pch symbols and colors

>




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

> harrisondata #to view harrisondata

| | 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
> plot(harrisondata$fl,harrisondata$age,
+      main="Plot 4: Symbol & Color by Era",
```

```
+     ylab="Age (yrs)", xlab="Fork Length (mm)",
+     xlim=c(0,500), ylim=c(0,15),
+     pch=pchs[harrisondata$era],col=cols[harrisondata$era])
> par(lty=2, lwd=2, col="cadetblue") #to set the parameters lty= Type of line, lwd= Width of line, col=Colour
> agevsfl <- data.frame(harrisondata$age, harrisondata$fl)  #creating data frame to plot age with fl of harrison data
> agevsfl                                  #to view the dataframe of age and fl of harrison data
   harrisondata.age harrisondata.fl
1                14             459
2                12             449
3                10             471
4                10             446
5                 9             400
6                 9             440
7                 9             462
8                 8             480
9                 8             449
10                7             437
11                7             431
12                7             425
13                7             419
14                6             409
15                6             397
16                5             419
17                5             381
18                5             363
19                5             351
20                4             372
```

| | | |
|---|---|---|
| 21 | 2 | 199 |
| 22 | 2 | 184 |
| 23 | 1 | 91 |
| 24 | 12 | 440 |
| 25 | 11 | 428 |
| 26 | 10 | 440 |
| 27 | 10 | 422 |
| 28 | 9 | 434 |
| 29 | 9 | 415 |
| 30 | 9 | 406 |
| 31 | 8 | 434 |
| 32 | 8 | 406 |
| 33 | 8 | 375 |
| 34 | 7 | 415 |
| 35 | 7 | 394 |
| 36 | 6 | 381 |
| 37 | 6 | 357 |
| 38 | 5 | 341 |
| 39 | 5 | 326 |
| 40 | 4 | 304 |
| 41 | 4 | 292 |
| 42 | 4 | 270 |
| 43 | 4 | 252 |
| 44 | 4 | 221 |
| 45 | 3 | 258 |
| 46 | 3 | 233 |
| 47 | 3 | 211 |
| 48 | 3 | 205 |
| 49 | 3 | 180 |

| | | |
|---|---|---|
| 50 | 2 | 196 |
| 51 | 2 | 171 |
| 52 | 2 | 143 |
| 53 | 1 | 131 |
| 54 | 1 | 88 |
| 55 | 1 | 75 |
| 56 | 0 | 51 |
| 57 | 0 | 41 |
| 58 | 0 | 20 |
| 59 | 7 | 245 |
| 60 | 7 | 279 |
| 61 | 5 | 245 |

> abline(lm(harrisondata$age~harrisondata$fl,data=agevsfl)) #to create a regression linear model for age and fl

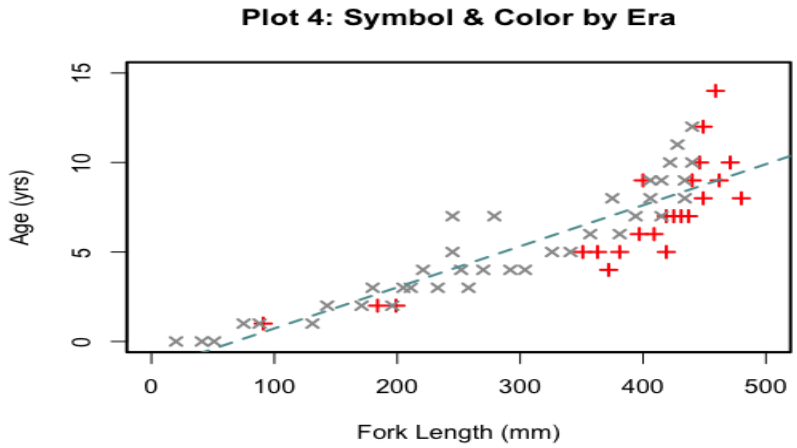
> lm(harrisondata$age~harrisondata$fl,data=agevsfl) #to know the slope and y-intercept

Call:

lm(formula = harrisondata$age ~ harrisondata$fl, data = agevsfl)

Coefficients:

| (Intercept) | harrisondata$fl |
|---|---|
| -1.56505 | 0.02294 |




>

**> #Q19 Place a legend of levels by era with pchs symbols in the top left of the plot**

> # Inset of 0.05

> # No box around the legend

> # Font size: 75% of nominal

>

> harrisondata #to view harrisondata

| | 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

> harrisondata$era <- as.factor(harrisondata$era) #to convert the values in Era from Date to Factor to create unique levels

> harrisondata$era #to view the output of Era column in Harrisondata as factors

[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

> cols[harrisondata$era] #assigning the colours to the unique values in Era column of Harrison data

[1] "red" "red" "red" "red" "red" "red" "red" "red"

[9] "red" "red" "red" "red" "red" "red" "red" "red"

[17] "red" "red" "red" "red" "red" "red" "red" "gray60"

[25] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"

[33] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"

[41] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"

```
[49] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[57] "gray60" "gray60" "gray60" "gray60" "gray60"
> plot(harrisondata$fl,harrisondata$age,
+     main="Plot 4: Symbol & Color by Era",
+     ylab="Age (yrs)", xlab="Fork Length (mm)",
+     xlim=c(0,500), ylim=c(0,15),
+     pch=pchs[harrisondata$era],col=cols[harrisondata$era])
> par(lty=2, lwd=2, col="cadetblue")
> agevsfl <- data.frame(harrisondata$age, harrisondata$fl)
> agevsfl
   harrisondata.age harrisondata.fl
1                14             459
2                12             449
3                10             471
4                10             446
5                 9             400
6                 9             440
7                 9             462
8                 8             480
9                 8             449
10                7             437
11                7             431
12                7             425
13                7             419
14                6             409
15                6             397
16                5             419
17                5             381
18                5             363
```

| | | |
|---|---|---|
| 19 | 5 | 351 |
| 20 | 4 | 372 |
| 21 | 2 | 199 |
| 22 | 2 | 184 |
| 23 | 1 | 91 |
| 24 | 12 | 440 |
| 25 | 11 | 428 |
| 26 | 10 | 440 |
| 27 | 10 | 422 |
| 28 | 9 | 434 |
| 29 | 9 | 415 |
| 30 | 9 | 406 |
| 31 | 8 | 434 |
| 32 | 8 | 406 |
| 33 | 8 | 375 |
| 34 | 7 | 415 |
| 35 | 7 | 394 |
| 36 | 6 | 381 |
| 37 | 6 | 357 |
| 38 | 5 | 341 |
| 39 | 5 | 326 |
| 40 | 4 | 304 |
| 41 | 4 | 292 |
| 42 | 4 | 270 |
| 43 | 4 | 252 |
| 44 | 4 | 221 |
| 45 | 3 | 258 |
| 46 | 3 | 233 |
| 47 | 3 | 211 |

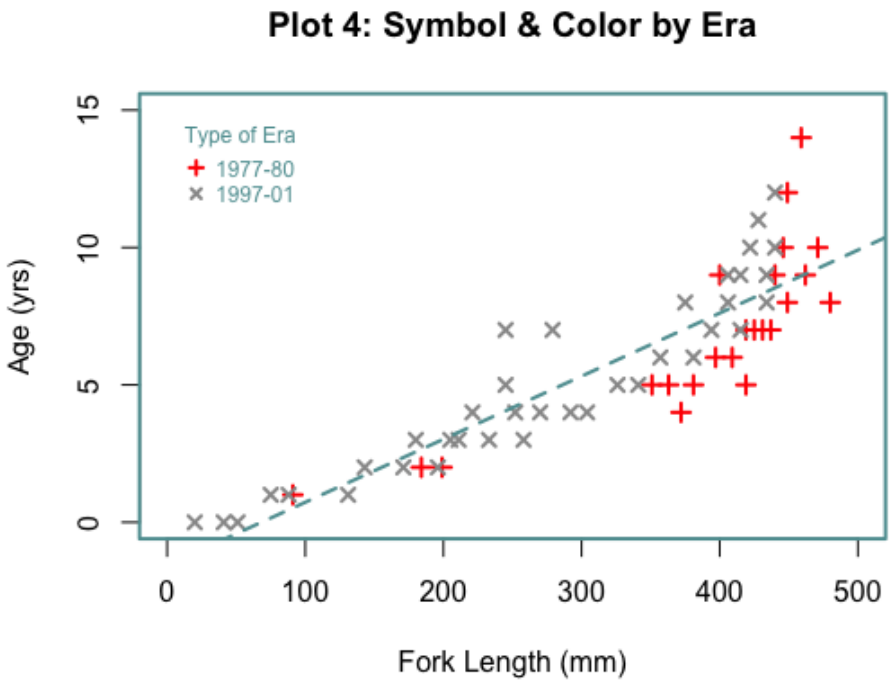
```
48          3          205
49          3          180
50          2          196
51          2          171
52          2          143
53          1          131
54          1           88
55          1           75
56          0           51
57          0           41
58          0           20
59          7          245
60          7          279
61          5          245
> abline(lm(harrisondata$age~harrisondata$fl,data=agevsfl))
> lm(harrisondata$age~harrisondata$fl,data=agevsfl)

Call:
lm(formula = harrisondata$age ~ harrisondata$fl, data = agevsfl)

Coefficients:
   (Intercept)  harrisondata$fl
      -1.56505         0.02294

> legend("topleft", inset=.05, title="Type of Era", c("1977-80","1997-01"),
+       pch=pchs, col=c("red", "gray60"), cex = 0.75, bty="n")
> #To include legend - placement of the legent, inset- space to be given from borders,
> #title of legend
> #c is values in the legend
```

> #pch is the symbols used in the plot

> #col is the colors used in the plot

> #cex is the font size

> #bty is the type of box and n is no box



**> #Key Findings**

**>**

**> #1. Descriptive analysis**

**>**

**> #1.1 Descriptive analysis of entire dataset of BullTroutRML2**

> dataset <- read.csv('/Users/devi/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 2/BullTroutRML2.csv', header=TRUE, sep=",") #to read the file from the directory

> dataset  #to view the loaded dataset for verification

| | 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

> summary(dataset) #to check summary of dataset which includes Min, Max, Mean, Median, 1st and 3rd Quartile

age fl lake era

Min. : 0.000 Min. : 20.0 Length:96 Length:96

1st Qu.: 4.000 1st Qu.:258.0 Class :character Class :character

Median : 6.000 Median :352.5 Mode :character Mode :character

Mean : 5.771 Mean :326.1

3rd Qu.: 8.000 3rd Qu.:406.0

Max. :14.000 Max. :688.0

> str(dataset) #to check structure of dataset

'data.frame': 96 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: chr  "Harrison" "Harrison" "Harrison" "Harrison" ...
$ era : chr  "1977-80" "1977-80" "1977-80" "1977-80" ...
> var(dataset$age)  #Variance of age in dataset
[1] 8.557456
> sd(dataset$age)   #Standard Deviation of age in dataset
[1] 2.925313
> var(dataset$fl)   #Variance of fl in dataset
[1] 12589.34
> sd(dataset$fl)    #Standard Deviation of fl in dataset
[1] 112.2022
 > skewness(dataset$age)  #Skewness of age in dataset
[1] 0.212223
> skewness(dataset$fl)   #Skewness of fl in dataset
[1] -0.5103268
> kurtosis(dataset$age)  #Kurtosis of age in dataset
[1] 2.735144
> kurtosis(dataset$fl)   #Kurtosis of fl in dataset
[1] 3.914769
>
```

**> #1.2 Descriptive analysis of only Harrison Lake Data**

```
> harrisondata<-dataset[dataset$lake == "Harrison",] #Extracting only Harrison Lake data from Dataset
> harrisondata      #To view Harrison lake data
  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

> summary(harrisondata)

```
     age             fl         lake             era
Min.   : 0.000   Min.   : 20   Length:61         Length:61
1st Qu.: 3.000   1st Qu.:221   Class :character   Class :character
Median : 6.000   Median :372   Mode  :character   Mode  :character
Mean   : 5.754   Mean   :319
3rd Qu.: 8.000   3rd Qu.:425
Max.   :14.000   Max.   :480
> str(harrisondata)
'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: chr  "Harrison" "Harrison" "Harrison" "Harrison" ...
$ era : chr  "1977-80" "1977-80" "1977-80" "1977-80" ...
> var(harrisondata$age)
[1] 11.12186
> sd(harrisondata$age)
[1] 3.334945
> var(harrisondata$fl)
[1] 16542.43
> sd(harrisondata$fl)
[1] 128.6174
> skewness(harrisondata$age,na.rm = FALSE)
[1] 0.1677228
> skewness(harrisondata$fl,na.rm = FALSE)
[1] -0.7353215
> kurtosis(harrisondata$age,na.rm = FALSE)
[1] 2.357344
> kurtosis(harrisondata$fl,na.rm = FALSE)
[1] 2.345694
```

```
> summary(harrisondata)
      age               fl          lake               era
 Min.   : 0.000   Min.   : 20   Length:61          Length:61
 1st Qu.: 3.000   1st Qu.:221   Class :character   Class :character
 Median : 6.000   Median :372   Mode  :character   Mode  :character
 Mean   : 5.754   Mean   :319
 3rd Qu.: 8.000   3rd Qu.:425
 Max.   :14.000   Max.   :480
> str(harrisondata)
'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: chr  "Harrison" "Harrison" "Harrison" "Harrison" ...
 $ era : chr  "1977-80" "1977-80" "1977-80" "1977-80" ...
> var(harrisondata$age)
[1] 11.12186
> sd(harrisondata$age)
[1] 3.334945
> var(harrisondata$fl)
[1] 16542.43
> sd(harrisondata$fl)
[1] 128.6174
> skewness(harrisondata$age,na.rm = FALSE)
[1] 0.1677228
> skewness(harrisondata$fl,na.rm = FALSE)
[1] -0.7353215
> kurtosis(harrisondata$age,na.rm = FALSE)
[1] 2.357344
> kurtosis(harrisondata$fl,na.rm = FALSE)
[1] 2.345694
```

**> #1.3 Descriptive analysis of only Osprey Lake Data**

> ospreydata<-dataset[dataset$lake == "Osprey",]  #Extracting only Osprey lake data from Dataset

> ospreydata  #To view Osprey Lake data

| | age | fl | lake | era |
|---|---|---|---|---|
| 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

> summary(ospreydata)

age fl lake era

Min. : 3.0 Min. :221.0 Length:35 Length:35

1st Qu.: 4.0 1st Qu.:305.5 Class :character Class :character

```
Median : 6.0   Median :332.0   Mode  :character   Mode  :character
Mean   : 5.8   Mean   :338.5
3rd Qu.: 7.0   3rd Qu.:360.0
Max.   :11.0   Max.   :688.0
> str(ospreydata)
'data.frame':      35 obs. of  4 variables:
$ age : int  8 8 7 7 6 6 5 5 4 4 ...
$ fl  : int  360 357 357 329 385 323 369 326 357 326 ...
$ lake: chr  "Osprey" "Osprey" "Osprey" "Osprey" ...
$ era : chr  "1977-80" "1977-80" "1977-80" "1977-80" ...
> var(ospreydata$age)
[1] 4.282353
> sd(ospreydata$age)
[1] 2.069385
> var(ospreydata$fl)
[1] 5734.492
> sd(ospreydata$fl)
[1] 75.72643
> skewness(ospreydata$age,na.rm = FALSE)
[1] 0.5139967
> skewness(ospreydata$fl,na.rm = FALSE)
[1] 2.670413
> kurtosis(ospreydata$age,na.rm = FALSE)
[1] 2.6847
> kurtosis(ospreydata$fl,na.rm = FALSE)
[1] 14.16737
>
```

```
> summary(ospreydata)
      age              fl            lake               era
 Min.   : 3.0   Min.   :221.0   Length:35          Length:35
 1st Qu.: 4.0   1st Qu.:305.5   Class :character   Class :character
 Median : 6.0   Median :332.0   Mode  :character   Mode  :character
 Mean   : 5.8   Mean   :338.5
 3rd Qu.: 7.0   3rd Qu.:360.0
 Max.   :11.0   Max.   :688.0
> str(ospreydata)
'data.frame':   35 obs. of  4 variables:
 $ age : int  8 8 7 7 6 6 5 5 4 4 ...
 $ fl  : int  360 357 357 329 385 323 369 326 357 326 ...
 $ lake: chr  "Osprey" "Osprey" "Osprey" "Osprey" ...
 $ era : chr  "1977-80" "1977-80" "1977-80" "1977-80" ...
> var(ospreydata$age)
[1] 4.282353
> sd(ospreydata$age)
[1] 2.069385
> var(ospreydata$fl)
[1] 5734.492
> sd(ospreydata$fl)
[1] 75.72643
> skewness(ospreydata$age,na.rm = FALSE)
[1] 0.5139967
> skewness(ospreydata$fl,na.rm = FALSE)
[1] 2.670413
> kurtosis(ospreydata$age,na.rm = FALSE)
[1] 2.6847
> kurtosis(ospreydata$fl,na.rm = FALSE)
[1] 14.16737
```

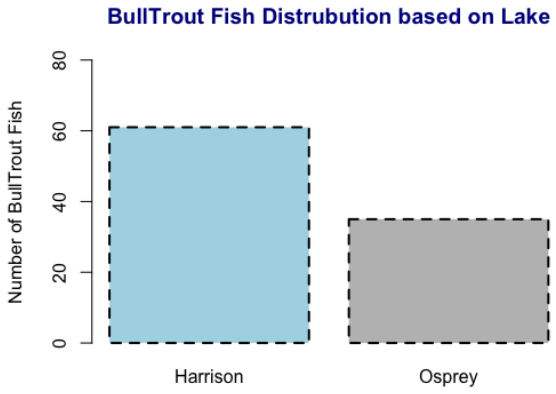
**> #2 Visualization**

**>**

**> #Visualization 1 - Bar Plot - Harrison lake vs Osprey lake**

> plot(factor(dataset$lake), col = c("lightblue","gray"), ylim=c(0,80), ylab="Number of BullTrout Fish", main="BullTrout Fish Distrubution based on Lake", col.main="darkblue")

>



**> #Visualization 2 - Histogram - Fish Age Distribution**

**> # 2.1 Harrison Lake**

> harrisondata

| | 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

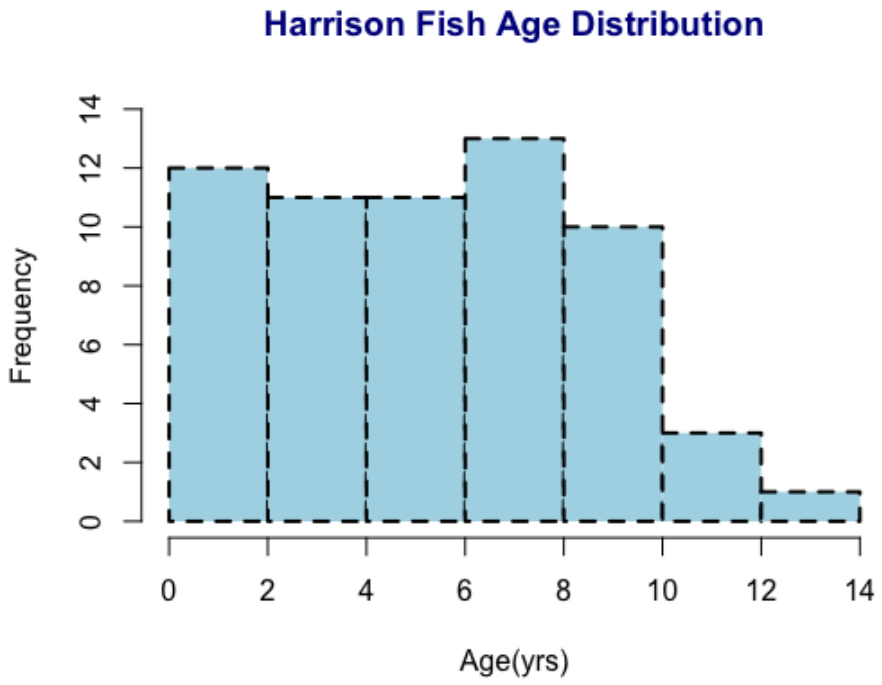
58 0 20 Harrison 1997-01

59 7 245 Harrison 1997-01

60 7 279 Harrison 1997-01

61 5 245 Harrison 1997-01

> hist(harrisondata$age,main = "Harrison Fish Age Distribution",col.main="darkblue",col="lightblue", ylab="Frequency", xlab="Age(yrs)", ylim=c(0,14),xlim=c(0,14))




**> #2.2 Osprey Lake**

> ospreydata

age fl lake era

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

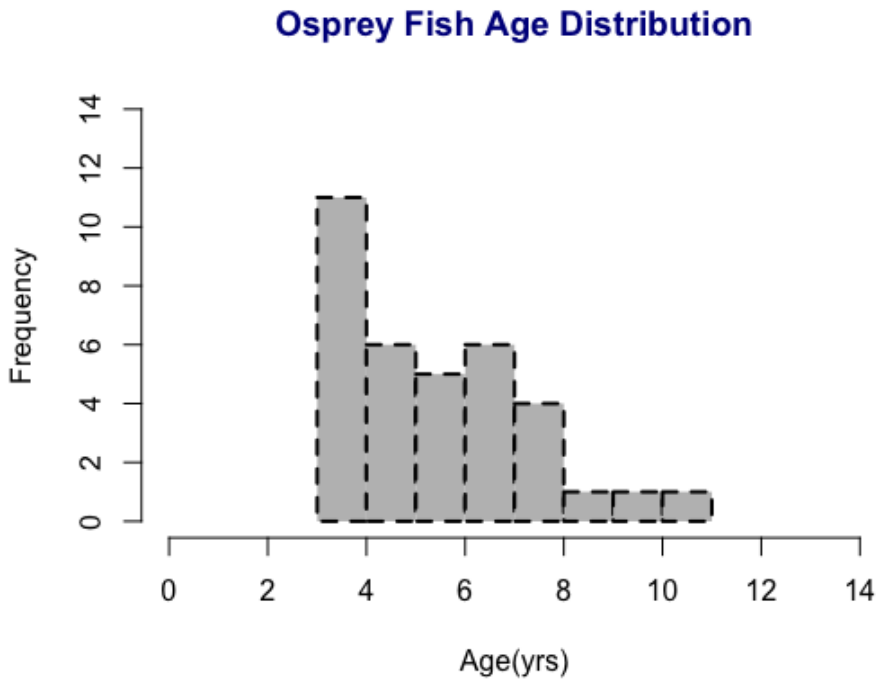
94 4 298 Osprey 1997-01

95 3 279 Osprey 1997-01

96 3 273 Osprey 1997-01

> hist(ospreydata$age,main = "Osprey Fish Age Distribution",col.main="darkblue",col="gray", ylab="Frequency", xlab="Age(yrs)", ylim=c(0,14),xlim=c(0,14))

>




**> #Visualization 3- Scatterplot - Age Versus Fork Length of Fish**

**> #3.1 Harrison Lake**

> harrisondata

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

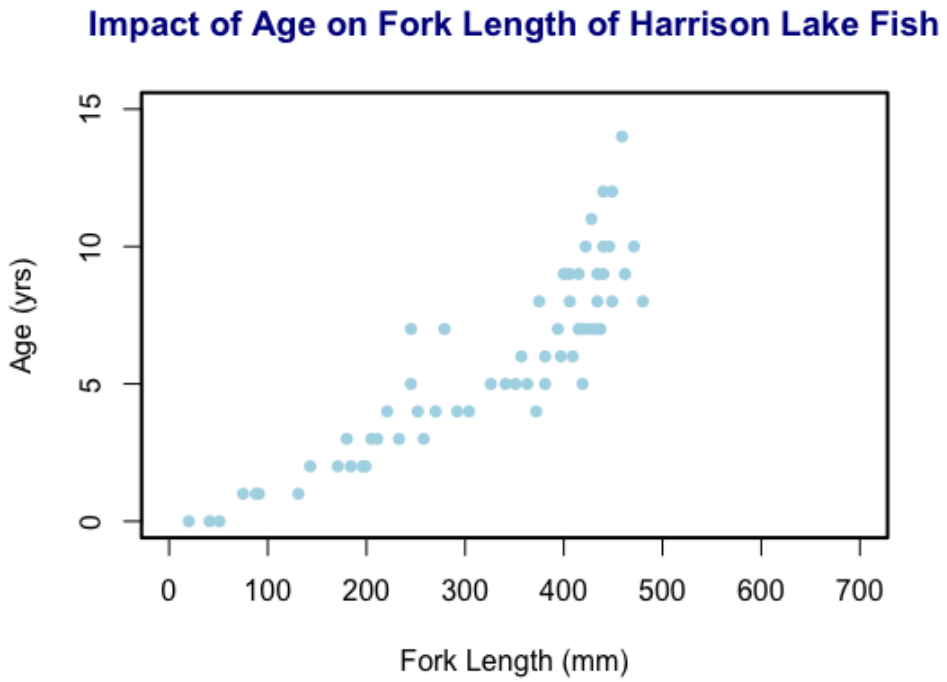
58 0 20 Harrison 1997-01

59 7 245 Harrison 1997-01

60 7 279 Harrison 1997-01

61 5 245 Harrison 1997-01

> plot(harrisondata$fl,harrisondata$age, main="Impact of Age on Fork Length of Harrison Lake Fish", col.main="darkblue", ylab="Age (yrs)", xlab="Fork Length (mm)", xlim=c(0,700), ylim=c(0,15),pch=20, col="lightblue")




**> #3.2 Osprey Lake**

> ospreydata

| | age | fl | lake | era |
|---|---|---|---|---|
| 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

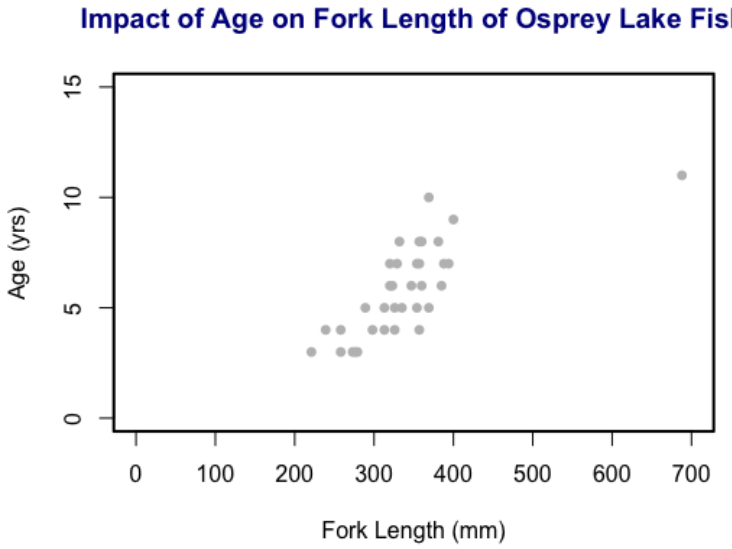
93 4 313 Osprey 1997-01

94 4 298 Osprey 1997-01

95 3 279 Osprey 1997-01

96 3 273 Osprey 1997-01

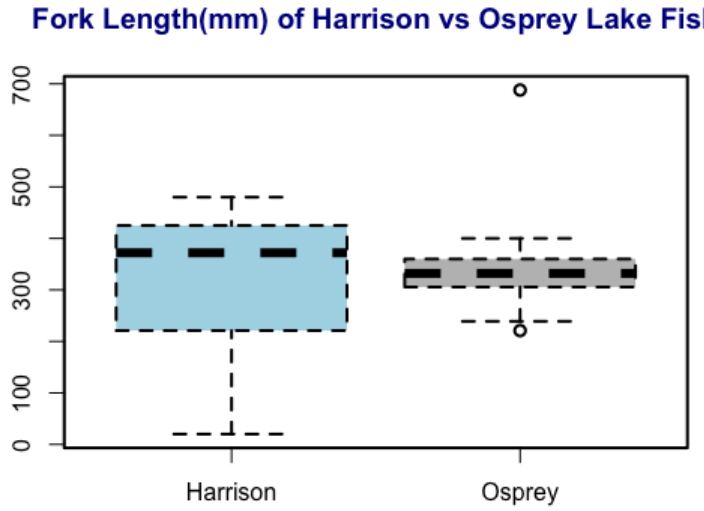
```
> plot(ospreydata$fl,ospreydata$age, main="Impact of Age on Fork Length of Osprey Lake Fish",
col.main="darkblue", ylab="Age (yrs)", xlab="Fork Length (mm)", xlim=c(0,700), ylim=c(0,15),pch=20,
col="gray")
```




**> #Visualization 4- Boxplot**

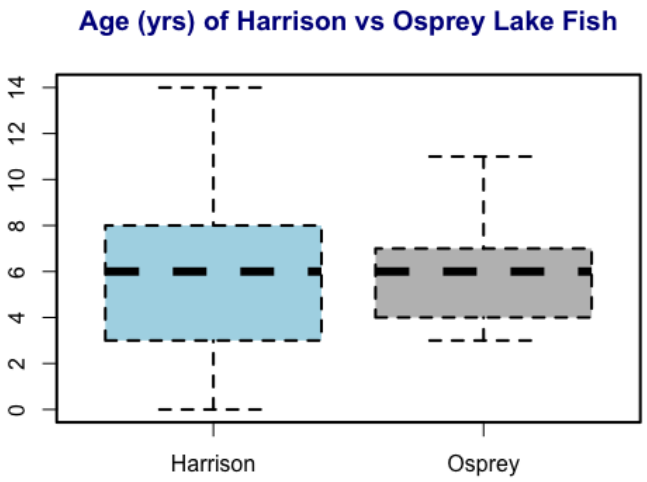
**> #.4.1 - Comparison of Fork Length of fishes in both the lakes**

> boxplot(harrisondata$fl,ospreydata$fl, main="Fork Length(mm) of Harrison vs Osprey Lake Fish", col = c("lightblue","gray"), col.main="darkblue", names=c("Harrison","Osprey"))



**> #4.2 Comparison of Age of fishes in both the lakes**

> boxplot(harrisondata$age,ospreydata$age, main="Age (yrs) of Harrison vs Osprey Lake Fish", col = c("lightblue","gray"), col.main="darkblue", names=c("Harrison","Osprey"))




>

>

**> #Visualization 5 - Density Plot**

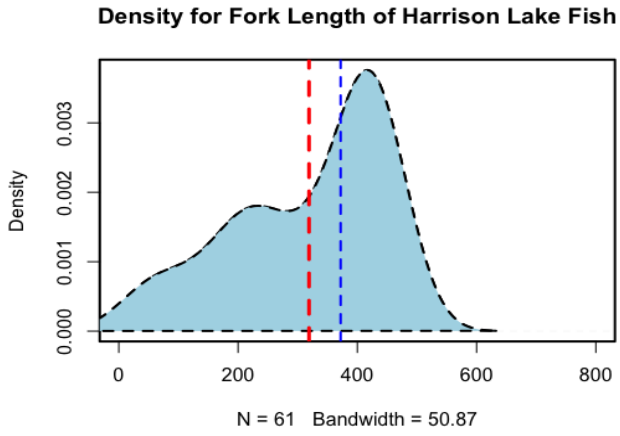
**> #5.1 Density for Fork Length(mm) of Harrison Lake Fish**

> plot(density(harrisondata$fl),xlim=c(0,800), main="Density for Fork Length of Harrison Lake Fish")

> polygon(density(harrisondata$fl), col="lightblue", border="black")

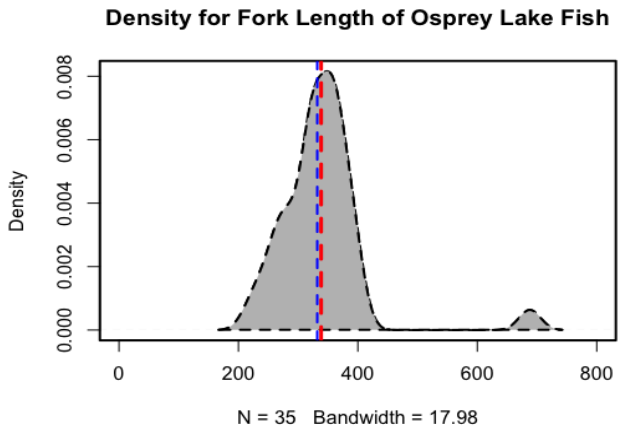
> abline(v=mean(harrisondata$fl),col="red", lwd = 3)

> abline(v=median(harrisondata$fl),col="blue", lwd=2)



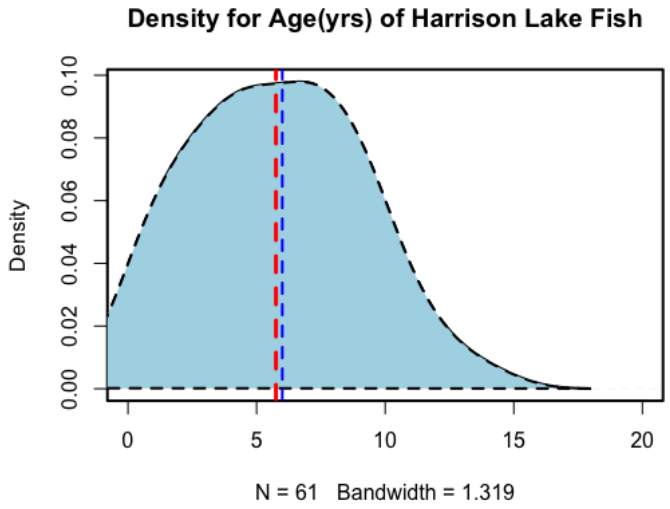
**> #5.2 Density for Fork Length(mm) of Osprey Lake Fish**

```
> plot(density(ospreydata$fl),xlim=c(0,800), main="Density for Fork Length of Osprey Lake Fish")
> polygon(density(ospreydata$fl), col="gray", border="black")
> abline(v=mean(ospreydata$fl),col="red", lwd=3)
> abline(v=median(ospreydata$fl),col="blue", lwd=2)
>
```




**> #5.3 Density for Age(yrs) of Harrison Lake Fish**

```
> plot(density(harrisondata$age),xlim=c(0,20), main="Density for Age(yrs) of Harrison Lake Fish")
> polygon(density(harrisondata$age), col="lightblue", border="black")
> abline(v=mean(harrisondata$age),col="red", lwd = 3)
> abline(v=median(harrisondata$age),col="blue", lwd=2)
```




**> #5.4 Density for Age(yrs) of Osprey Lake Fish**

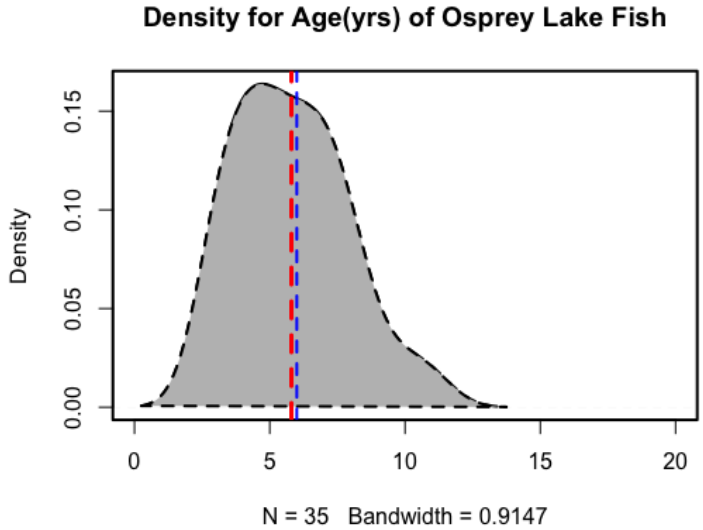
> plot(density(ospreydata$age),xlim=c(0,20), main="Density for Age(yrs) of Osprey Lake Fish")

> polygon(density(ospreydata$age), col="gray", border="black")

> abline(v=mean(ospreydata$age),col="red", lwd=3)

> abline(v=median(ospreydata$age),col="blue", lwd=2)

>



**> #Visualization 6 - Period**

**> #6.1 Harrison Data by Period**

> harrisondata

| | 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

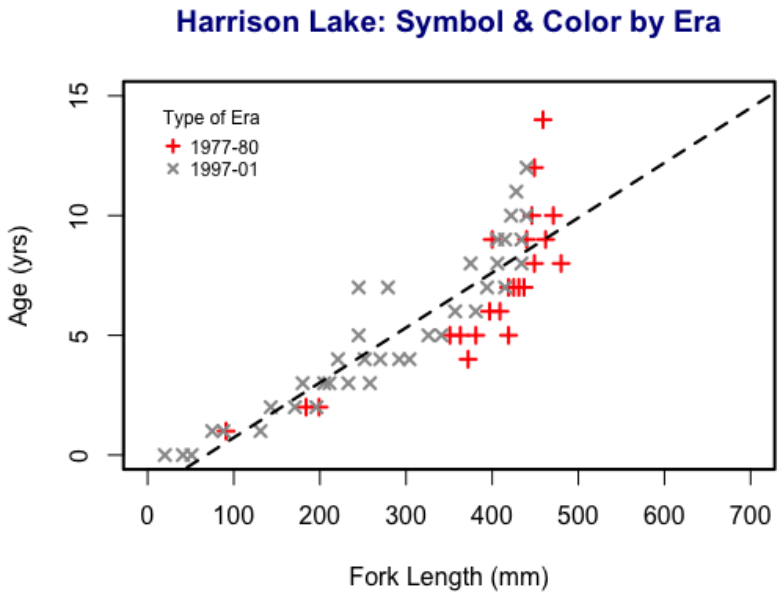
53 1 131 Harrison 1997-01

```
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
> harrisondata$era <- as.factor(harrisondata$era)
> harrisondata$era
[1] 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80
[10] 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80 1977-80
[19] 1977-80 1977-80 1977-80 1977-80 1977-80 1997-01 1997-01 1997-01 1997-01
[28] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
[37] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
[46] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
[55] 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01 1997-01
Levels: 1977-80 1997-01
> cols[harrisondata$era]
[1] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "red"
[9] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "red"
[17] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "gray60"
[25] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[33] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[41] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[49] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[57] "gray60" "gray60" "gray60" "gray60" "gray60"
> plot(harrisondata$fl,harrisondata$age,
+      main="Harrison Lake: Symbol & Color by Era", col.main="darkblue",
```

```
+     ylab="Age (yrs)", xlab="Fork Length (mm)",
+     xlim=c(0,700), ylim=c(0,15),
+     pch=pchs[harrisondata$era],col=cols[harrisondata$era])
> par(lty=2, lwd=2, col="black")
> agevsfl <- data.frame(harrisondata$age, harrisondata$fl)
> agevsfl
   harrisondata.age harrisondata.fl
1                14             459
2                12             449
3                10             471
4                10             446
5                 9             400
6                 9             440
7                 9             462
8                 8             480
9                 8             449
10                7             437
11                7             431
12                7             425
13                7             419
14                6             409
15                6             397
16                5             419
17                5             381
18                5             363
19                5             351
20                4             372
21                2             199
22                2             184
```

| | | |
|---|---|---|
| 23 | 1 | 91 |
| 24 | 12 | 440 |
| 25 | 11 | 428 |
| 26 | 10 | 440 |
| 27 | 10 | 422 |
| 28 | 9 | 434 |
| 29 | 9 | 415 |
| 30 | 9 | 406 |
| 31 | 8 | 434 |
| 32 | 8 | 406 |
| 33 | 8 | 375 |
| 34 | 7 | 415 |
| 35 | 7 | 394 |
| 36 | 6 | 381 |
| 37 | 6 | 357 |
| 38 | 5 | 341 |
| 39 | 5 | 326 |
| 40 | 4 | 304 |
| 41 | 4 | 292 |
| 42 | 4 | 270 |
| 43 | 4 | 252 |
| 44 | 4 | 221 |
| 45 | 3 | 258 |
| 46 | 3 | 233 |
| 47 | 3 | 211 |
| 48 | 3 | 205 |
| 49 | 3 | 180 |
| 50 | 2 | 196 |
| 51 | 2 | 171 |

```
52          2          143
53          1          131
54          1          88
55          1          75
56          0          51
57          0          41
58          0          20
59          7          245
60          7          279
61          5          245
> abline(lm(harrisondata$age~harrisondata$fl,data=agevsfl))
> lm(harrisondata$age~harrisondata$fl,data=agevsfl)

Call:
lm(formula = harrisondata$age ~ harrisondata$fl, data = agevsfl)

Coefficients:
   (Intercept)  harrisondata$fl
      -1.56505         0.02294

> legend("topleft", inset=.05, title="Type of Era", c("1977-80","1997-01"),
+       pch=pchs, col=c("red", "gray60"), cex = 0.75, bty="n")
```




**> #6.2 Osprey Data by Period**

> ospreydata

| | age | fl | lake | era |
|---|---|---|---|---|
| 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

> ospreydata$era <- as.factor(ospreydata$era)

> ospreydata$era

[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 1997-01 1997-01 1997-01

[19] 1997-01 1997-01 1997-01 1997-01 1997-01 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

Levels: 1977-80 1997-01

```
> cols[ospreydata$era]
[1] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "red"
[9] "red"    "red"    "red"    "red"    "red"    "red"    "red"    "gray60"
[17] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[25] "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60" "gray60"
[33] "gray60" "gray60" "gray60"
> plot(ospreydata$fl,ospreydata$age,
+      main="Osprey Lake: Symbol & Color by Era", col.main="darkblue",
+      ylab="Age (yrs)", xlab="Fork Length (mm)",
+      xlim=c(0,700), ylim=c(0,15),
+      pch=pchs[ospreydata$era],col=cols[ospreydata$era])
> par(lty=2, lwd=2, col="black")
> agevsfl <- data.frame(ospreydata$age, ospreydata$fl)
> agevsfl
   ospreydata.age ospreydata.fl
1               8           360
2               8           357
3               7           357
4               7           329
5               6           385
6               6           323
7               5           369
8               5           326
9               4           357
10              4           326
11              4           258
12              4           239
13              3           221
14              3           258
```

15 3 276

16 11 688

17 10 369

18 9 400

19 8 381

20 8 332

21 7 394

22 7 388

23 7 354

24 7 320

25 6 320

26 6 347

27 6 360

28 5 354

29 5 335

30 5 313

31 5 289

32 4 313

33 4 298

34 3 279

35 3 273

> abline(lm(ospreydata$age~ospreydata$fl,data=agevsfl))

> lm(ospreydata$age~ospreydata$fl,data=agevsfl)

Call:

lm(formula = ospreydata$age ~ ospreydata$fl, data = agevsfl)

Coefficients:

(Intercept) ospreydata$fl

-1.14242 0.02051

```
> legend("topleft", inset=.05, title="Type of Era", c("1977-80","1997-01"),
+       pch=pchs, col=c("red", "gray60"), cex = 0.75, bty="n")
>
```