INTRODUCTION TO ANALYTICS



ALY6000, WINTER 2022

Module 2 Project - Executive Summary Report 2

Submitted by: Abhinav Jain

NUID: 002938209

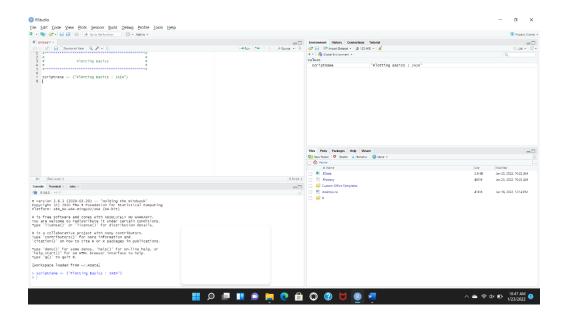
Submitted To: Richard Zhi

Date: 01/25/2022

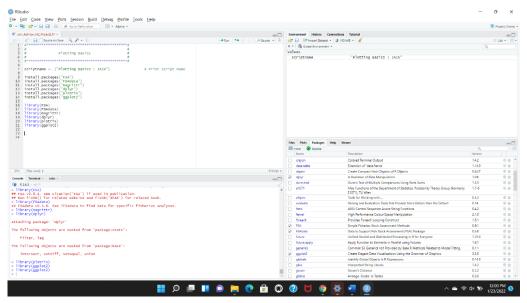
Introduction:

R is a Programming language that helps in analyzing data through various functions which are inbuilt in the library. It is platform-independent and provides a vast community to enhance the hands-on experience in the field of data analytics. In this report understand the basics of R while plotting scatterplot, histogram, and many other functions. This document includes a screenshot of the solution to each problem.

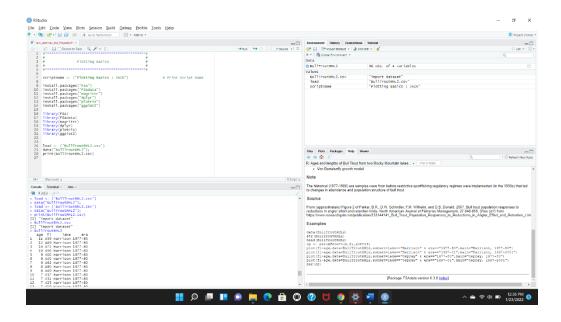
Task 1. Initially, printed my last name with scriptname <- ("Plotting Basics: Jain") [1]



Task 2. As per the requirement for this report, it is required to uninstall the latest version 4.1.2 and reinstalled older version 3.6.3, and Installed FSA, FSA data, magrittr, dplyr, plotrix, ggplot2 packages which help to get the result of the below solution.

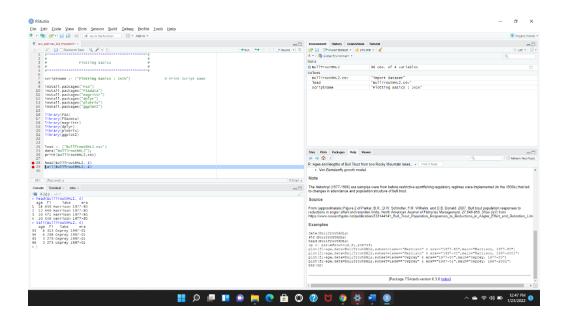


Task 3. In this report after the installation of FSA and FSA data libraries, loaded the inbuilt dataset the BullTroutRML2 dataset (BullTroutRML2.csv) [3]

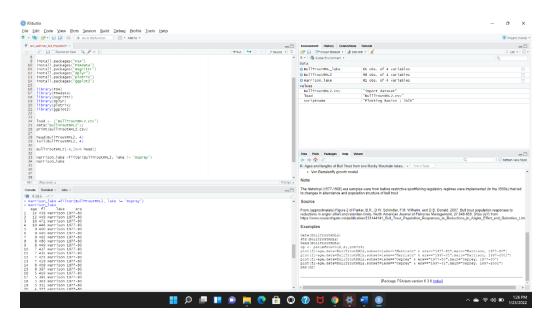


Task 4. To fetch the first and last 4 records from the BullTroutRML2 dataset, using head and tail function. [4]

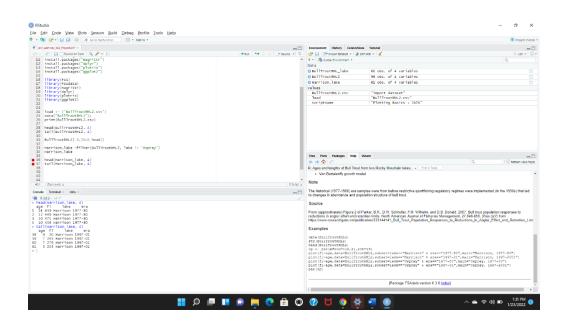
head() & tail()



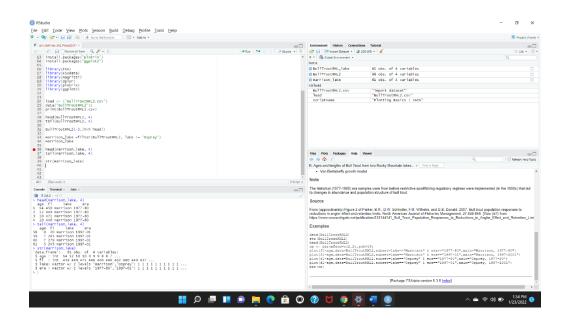
Task 5. To filter the data a new dataset was created and rename the original dataset (BullTroutRML2) Harrison_Lake and removed all records except Harrison Lake. Earlier there was 96 observation in the dataset after filtration, now 61 observation. [5]



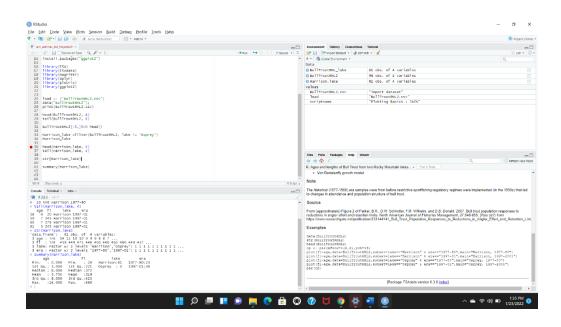
Task 6. This task displayed the first and last 4 records of the filtered dataset which is available with the name Harrison_Lake [4]



Task 7. This solution displayed the structure of the filtered data frame Harrision_Lake which has 61 observations and 4 variables

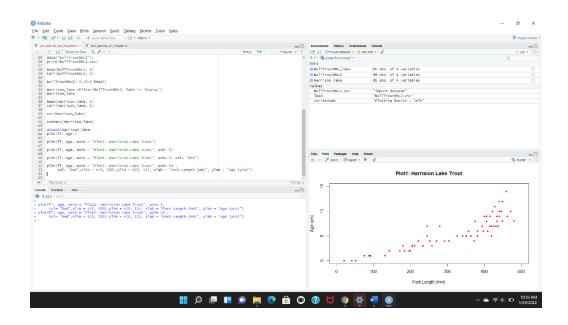


Task 8. This solution displayed the summary of the new filtered Harrision_Lake which shows the minimum, maximum, mean, and median value of the dataset.

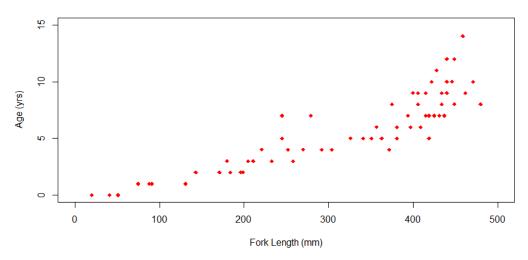


Task 9. As per the requirement in this solution created the scatterplot for the 'age' (y variable) and 'fl' (x variable) with the following specification:

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

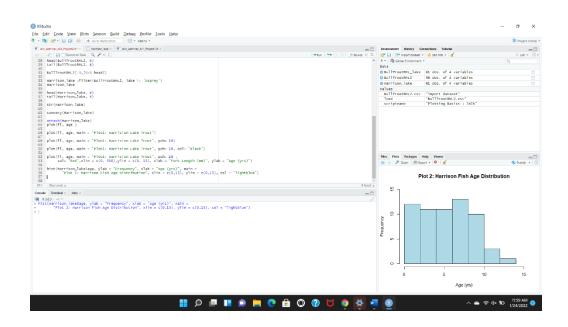


Plot1: Harrision Lake Trout

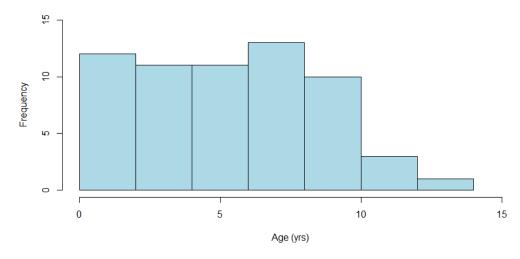


Task 10. In this solution, plotted "Age" histogram with the following specifications [10]

- Y-axis label is "Frequency"
- X-axis label is "Age (yrs)"
- Title of the histogram is "Plot 2: Harrison Fish Age Distribution"
- •X and Y-axis limits are 0, 15
- The color of the frequency plots is "light blue"

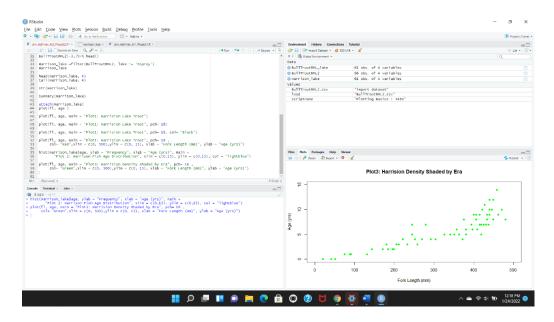


Plot 2: Harrison Fish Age Distribution

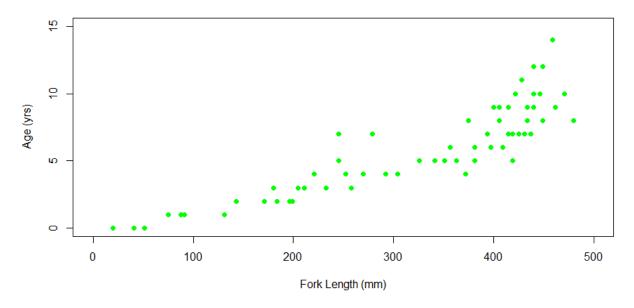


Task 11. This solution created an overdense plot using the below parameters with the previous scatterplot

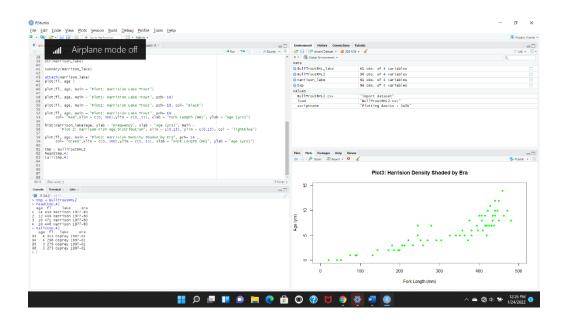
- Title the plot "Plot 3: Harrison Density Shaded by Era"
- Y-axis label is "Age (yrs)"
- Y-axis limits are 0 to 15
- X-axis label is "Fork Length (mm)"
- X-axis limits are 0 to 500
- include two levels of shading for the "green" data points.
- Plot solid circles as data points



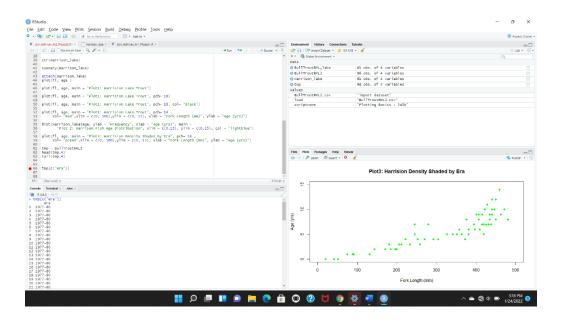
Plot3: Harrision Density Shaded by Era



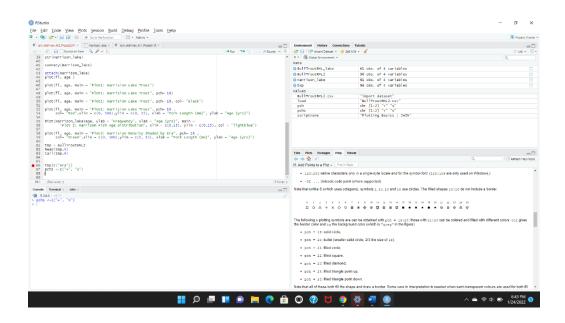
Task 12. This solution created a new object called "tmp" that includes the first 3 and last 3 records of the theHarrison_Lake data set. [12]



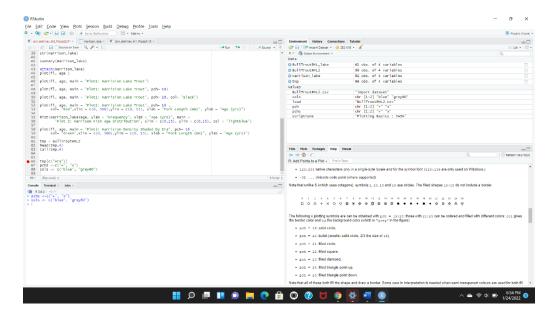
Task 13. This solution displayed the "era" column (variable) in the new "tmp" object



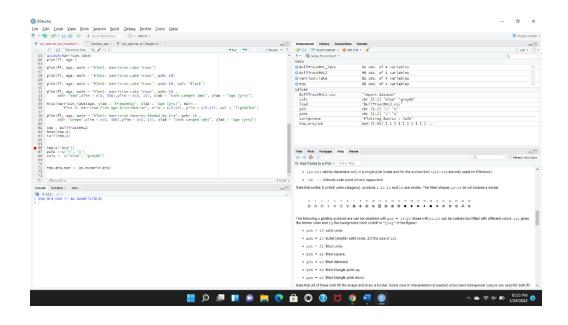
Task 14. This solution created a pchs vector with the argument values for + and x.



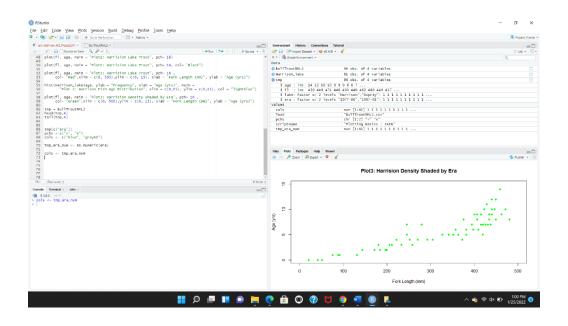
Task 15. This solution created a cols vector with the two elements "blue" and "gray60"



Task 16. To convert the tmp era values from factor to numeric values,[16] tmp_era_num<- as.numeric(era)

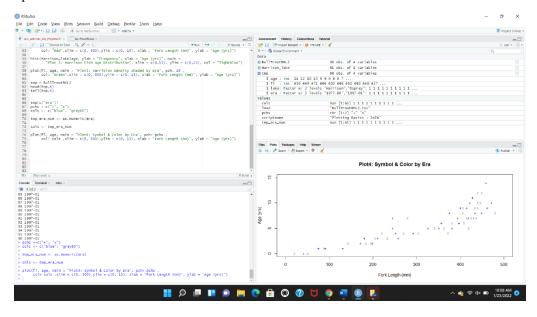


Task 17. Initialize the cols vector with the tmp era values

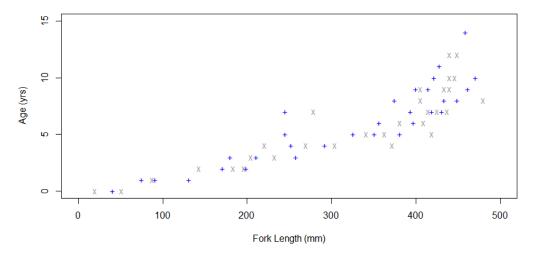


Task 18. This solution created a plot of "Age (yrs)" (y variable) versus "Fork Length (mm)" (x variable) with the instructions below:

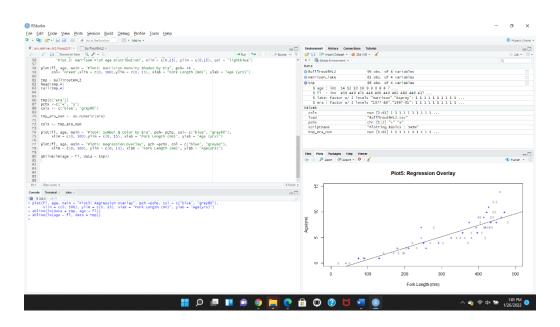
- Title of the graph is "Plot 4: Symbol & Color by Era"
- Limit of the x-axis is (0,500)
- Limit of the y axis is (0,15)
- Y-axis label is "Age (yrs)"
- X-axis label is "Fork Length (mm)"
- Set pch equal to pcs era values
- Set col equal to cols era values



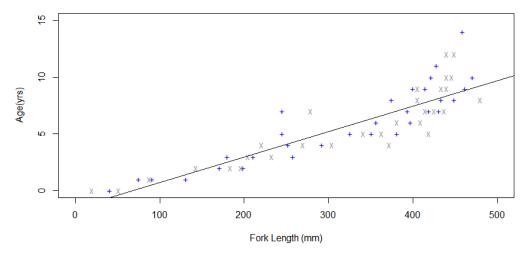
Plot4: Symbol & Color by Era



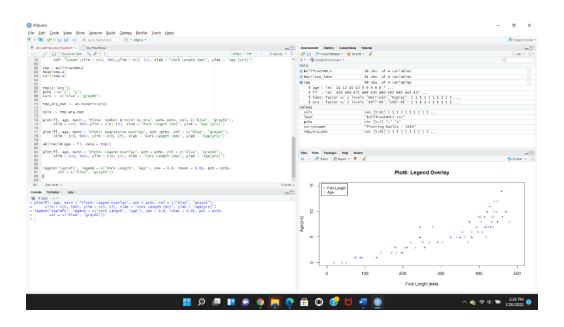
Task19 In this Task plotted regression line while comparing two variables Fork Length (mm) and Age(mm),



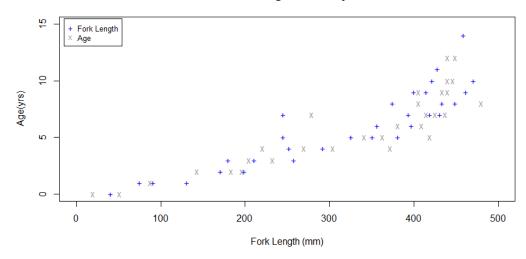
Plot5: Regression Overlay



Task 20. In this task, created the Legend for the data points in the graph with "+" Fork Length and "x" Age [20]



Plot6: Legend Overlay



Summary:

In this report, understanding the basics of R by importing libraries, loading a dataset, filtering the datapoints with different variables, renaming the dataset, removing records, and displaying the filtered dataset. Moreover, learned about the various plot like scatterplot, histogram, regression lines while making changes with the title, axis, label, and by limiting the coordinates.

These are the keys take away from this report.

- Installation of Packages and libraries
- Load() Import the data from the library
- Print first and last records from the dataset with the function head(), tail()
- Filter() by removing records from the dataset
- Display records from the dataset
- Fetch the structure and summary of the dataset
- Plotted the scatterplot, histogram with a variable defined in the dataset (the age and fork length) with setting a limit on-axis.
- Created and converted values of the vector
- Plotted regression line by using abline(), through which we can plot the regression line
- Placed the legend () on the top-left of the graph,

Hence, It gives me practical experience while running code in the R programming language.

Bibliography

[1] How to get the file name of the R script currently being executed (for easy automatic email of results)

WetlabStudentWetlabStudent 2 et al.

https://stackoverflow.com/questions/18264683/how-to-get-the-file-name-of-the-r-script-currently-being-executed-for-easy-auto

[3] How to use inbuilt datasets available in R

https://www.youtube.com/watch?v=X4rrrhwfGrM

[4] Print First or Last Rows of Data Frame in R (4 Examples) | head() & tail() Functions | Top & Bottom

https://www.youtube.com/watch?v=QDww8Vj2bjQ

[5] How to Filter Data in R

https://www.youtube.com/watch?v=sDoGnfL0vzg

[4] Print First or Last Rows of Data Frame in R (4 Examples) | head() & tail() Functions | Top & Bottom

https://www.youtube.com/watch?v=QDww8Vj2bjQ

[10]How To Make a Histogram in R

https://www.youtube.com/watch?v=tp_BG5wDeVU

[12]How to create an object in R Studio

https://www.youtube.com/watch?v=1_d6W2vmRGU

[16]Data Type Conversion in R (2 Examples ... - youtube.com

How to convert characters and factors to numeric in R. More details in these tutorials:

https://statisticsglobe.com/convert-character-to-numeric-in-r/https:/...

https://www.youtube.com/watch?v= rPde-SPDVE

[20] How To... Add a Legend to a Histogram in R #36 - YouTube

Learn how to plot a histogram with headings and labels in R with @Eugene O'Loughlin. The R script (36_How_To_Code.R) and data file (36_Data_File.csv) for this...

https://www.youtube.com/watch?v=CR4q U9qTEs

Appendix: RStudio File Edit Code View Plots Session Build Debug Profile Iools Help - o × O → O Go to file/function O Jain_Abhinav_M2_Project2.R × BullTroutRML2 × ♣ Project: (None) * or to fletunction sometimes M2_Project2R × BuffleutRM2 × scriptname <- ("Plotting Basics : JAIN") # Print Script Name | striptname <- (Proteing basis | finstall.packages("FSA") | finstall.packages("FSAdata") | finstall.packages("magrittr") | dinstall.packages("dplyr") | dinstall.packages("gpplotrix") | finstall.packages("ggplot2") plot(f1, age, main = "Plot1: Harrision Lake Trout", pch= 18 , col= "Red",xlim = c(0, 500),ylim = c(0, 15), xlab = "Fork Length (mm)", ylab = "Age (yrs)") hist(Harrison_lakeiage, ylab = "Frequency", xlab = "Age (yrs)", main = "Plot 2: Harrison Fish Age Distribution", xlim = c(0,15), ylim = c(0,15), col = "lightblue") へ 🧠 奈 ゆ) 🖆 5:02 PM 🕕 1/26/2022 🕕 RStudio File Edit Code View Blots Session Build Debug Profile Iools Help • The American Micropath of Session Build Debug Profile Iools Help • Ins American Micropath of Session Build Debug Profile Iools Help • Ins American Micropath of Session Build Debug Profile Iools Help • Ins American Micropath of Session Build Debug Profile Iools • Institute Tourilate, • RStudio - o × Project (None) • 63 tal(ten,4) 64 tal(ten,4) 65 tal(c(rear)) 66 tal(c(rear)) 66 tal(c(rear)) 67 pchs <-("h," "s", "grey60") 68 cols <- (C'blue", "grey60") 70 tal(p,ers_an,um) 71 cols <- tal("h," tal("

へ 🥱 🛜 中) 🖆 5:04 PM 🕕 1/26/2022 🕕