Executive Summary Report 2

Name: Raj Tank

NUID: 002988601

Date: 19/01/2022

Introduction

The data used in R script is primarily focused on BullTrout's growth (Fish Species) in North America by North American Journal of Fisheries Management.

First and last record of the dataset BullTroutRML2

```
> print(first_last_data)
   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
```

Summary of BullTroutRML2

Descriptive statistics of the dataset that provides Mean, Median, Mode, and max(an overview of the dataset)

```
> summary(BullTroutRML2)
      age              fl              lake         era
 Min.   : 0.000   Min.   : 20.0   Harrison:61   1977-80:38
 1st Qu.: 4.000   1st Qu.:258.0   Osprey  :35   1997-01:58
 Median : 6.000   Median :352.5
 Mean   : 5.771   Mean   :326.1
 3rd Qu.: 8.000   3rd Qu.:406.0
 Max.   :14.000   Max.   :688.0
```

Displaying the only data where lake is Harrison(Data has 61 records but just showing 5 rows with four variables and structure of it.

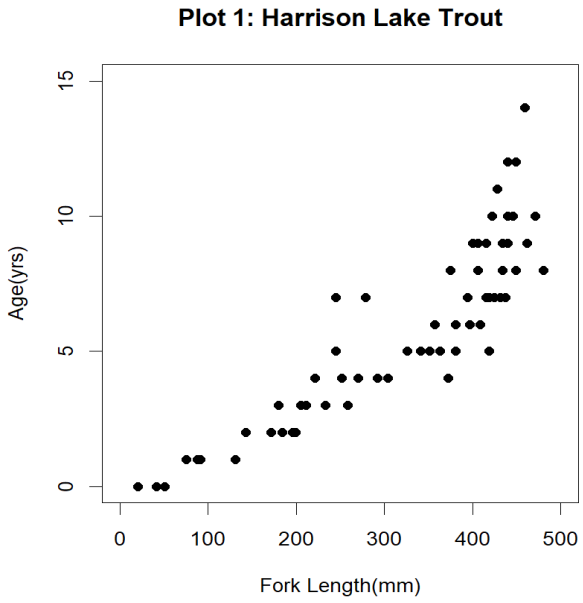
```
> filtered_data=filter(BullTroutRML2,lake=="Harrison")
> print(filtered_data)
   age  fl     lake     era
1   14 459 Harrison 1977-80
2   12 449 Harrison 1977-80
3   10 471 Harrison 1977-80
4   10 446 Harrison 1977-80
5    9 400 Harrison 1977-80
```

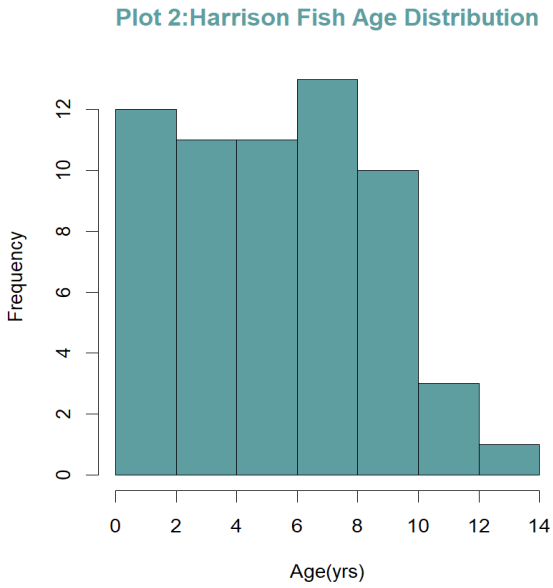
```
> str(filtered_data)
'data.frame':   61 obs. of  4 variables:
 $ age : int  14 12 10 10 9 9 9 8 8 7 ...
 $ fl  : int  459 449 471 446 400 440 462 480 449 437 ...
 $ lake: Factor w/ 2 levels "Harrison","Osprey": 1 1 1 1 1 1 1 1 1 1 ...
 $ era : Factor w/ 2 levels "1977-80","1997-01": 1 1 1 1 1 1 1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

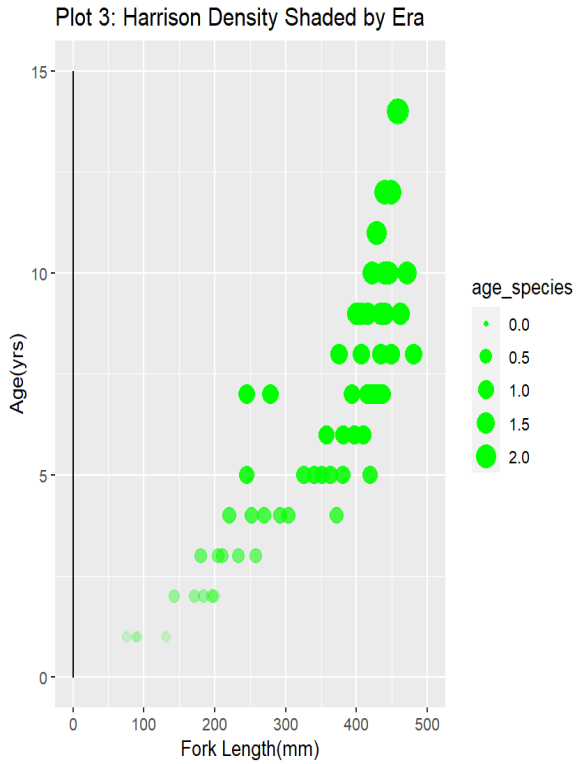
Methodology

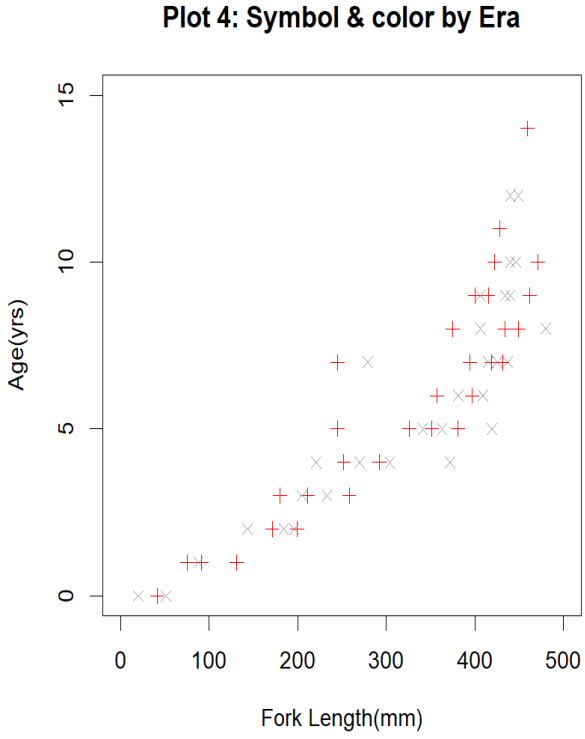
Data visualization in R with BullTroutRML2

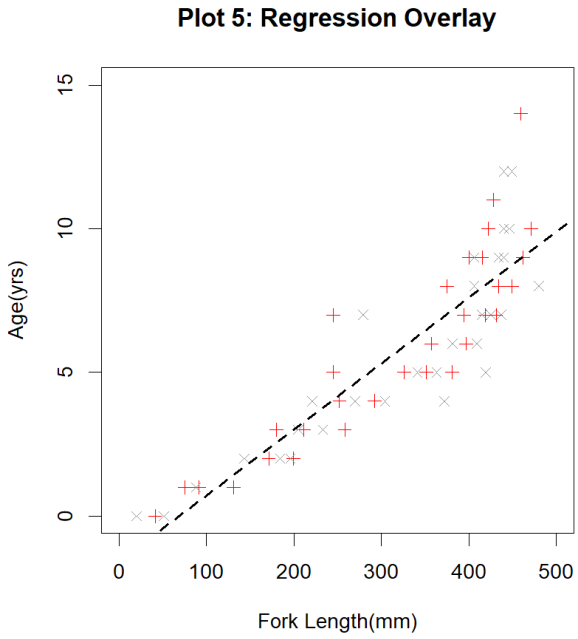
The plots has linear relationship between Age and Fork Length. In addition, I can also create linear regression model that predicts Fork length according to Age.

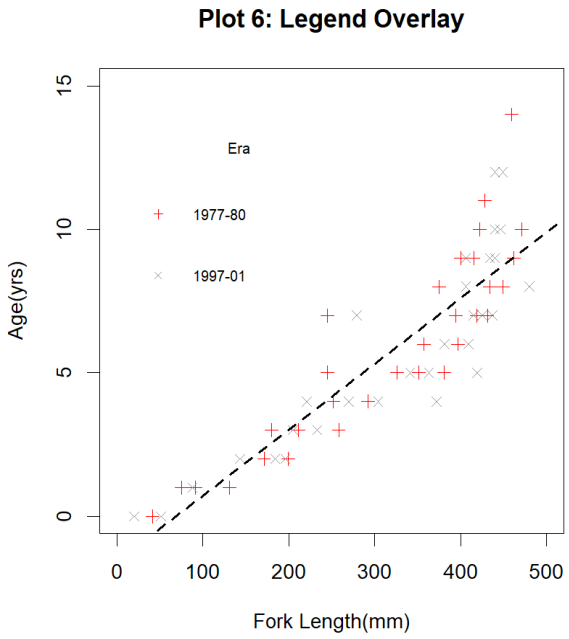












The main aim of this dataset is to see the impact before and after implementing some restriction of sportfishing.

Before analysis, the dataset needed to be checked by ensuring the right structure, shape, missing values, and its variables. Then, the data was analysed and processed by R such as (Exploratory Data Analysis, plot, filteration and descriptive statistics.)

Key findings

The data is small that possesses information regarding fish species such as Age, Fork Length, Lake, and Year.

By summarizing the data, I acquired the information of Mean, Median, Mode, and Frequency

Harrison lake has a high volume of data(61) as compared to lake Osprey that had only 30 observations.

As seen from the graph,From the age 0 to 10, analyzed frequency was the same, which was 12; however, the least frequency(less than 2) was noted by the species that lied in between 10 and 14(Age(yrs.)).

BullTroutRML2 has a few outliers that needed to be fixed and the most of the data is concentrated on at the age of 5-8.

Conclusion

Considering all the points graphs, and details generated above, it can be reiterated that:

After putting restriction on Sportfishing, around half population of Bull Trout species increased as compared to old times, where there was no restriction.

The plot depicts correlation between Fork Length and Age.

In addition, the histogram of age is Right skewed with different values of Mean, Median, and Mode; nonetheless, Median is 0.25 % higher than Mean. Also, the maximum age was 14 in the filtered data.

Bibliography

1. Add legends to plots in R software *: the easiest way! - Easy Guides - Wiki - STHDA*. (n.d.). Statistical Tools for High-Throughput Data Analysis. http://www.sthda.com/english/wiki/add-legends-to-plots-in-r-software-the-easiest-way
2. Adepoju, F., Mwaura, W., & Markan, Z. (2020, November 24). *Pushing a project to GitHub*. CircleCI. https://circleci.com/blog/pushing-a-project-to-github/?utm_source=google&utm_medium=sem&utm_campaign=sem-google-dg--uscan-en-dsa-maxConv-auth-nb&utm_term=g_-_c__dsa_&utm_content=&gclid=Cj0KCQiAosmPBhCPARIsAHOen-NcW6TMyHrEbl3cxDYIk9YE3TH6atUVFWZoB20kLFkp7v4H_vEKHk4aAhWNEALw_wcB
3. Holtz, Y. (n.d.). *Data visualization with R and ggplot2 | the R Graph Gallery*. R-Graph-Gallery. https://www.r-graph-gallery.com/ggplot2-package.html
4. Introduction to Plotting in R. (2012, November 22). YouTube. https://www.youtube.com/watch?v=SjcUlHh3UJg
5. Plot a legend outside of the plotting area in base graphics? (2010, October 14). Stack Overflow. https://stackoverflow.com/questions/3932038/plot-a-legend-outside-of-the-plotting-area-in-base-graphics
6. R. (2022, January 27). *GitHub - rajtank/Week_2_Rproject*. GitHub. Retrieved January 27, 2022, from https://github.com/rajtank/Week_2_Rproject

Appendix

```
#1 plotting last name that includes "Plotting Basics : Lastname"
print("Plotting Basics: Raj Tank")

#2 installing some neccesary libraries
install.packages("FSA")
install.packages("FSAdata")
install.packages("magrittr")
install.packages("dplyr")
install.packages("plotrix")
install.packages("ggplot2")
install.packages("moments")
library(FSA)
library(FSAdata)
library(magrittr)
library(dplyr)
library(plotrix)
library(ggplot2)
library(moments)

#3 loading BullTroutRML2 dataset
data("BullTroutRML2")

#4 printing the first and last 3 records from BullTroutRML2 dataset
first_last_data=headtail(BullTroutRML2,n = 3)
print(first_last_data)
```

```
#5 removing all the records except Harrison Lake
filtered_data=filter(BullTroutRML2,lake=="Harrison")
print(filtered_data)

#6 displaying the first and last recors from filteres data of BullTroutRML5
records=headtail(filtered_data,n=5)
print(records)

#7 displaying the sturcture of filtered BullTroutRML2
str(filtered_data)

#8 displaying the summary of BullTroutRML2
summary(BullTroutRML2)

#9 ploting scatterplot accordig to the provided details
plot(age~fl,data=filtered_data,xlim=c(0,500),ylim=c(0,15),
    main="Plot 1: Harrison Lake Trout",ylab="Age(yrs)",
    xlab="Fork Length(mm)",pch=16)

#10 plotting histogram of age
hist(filtered_data$age,ylab="Frequency",xlab="Age(yrs)",
    main = "Plot 2:Harrison Fish Age Distribution",
    col="cadetblue",col.main="cadetblue")

#11 plotting overdense plot
age_species =filtered_data$age/mean(filtered_data$age)
ggplot(data=filtered_data, aes(y=age)) +geom_density()+
```

```
  geom_point(data=filtered_data, aes(y=age, x=fl,size=age_species),alpha
=age_species,
  color="green") +lims(x = c(0,500),y=c(0,15))+
  labs(title = "Plot 3: Harrison Density Shaded by Era"
  ,x = "Fork Length(mm)",y = "Age(yrs)")

#12 creating object "tmp" stores first and last records of BullTroutRML2
tmp<-headtail(filtered_data,n = 3)
print(tmp)

#13 dislaying Era column in new tmp object
tmp$era

#14 creating pchs vector with argument values for + and x
pchs<-c(3,4)

#15  creating vector with two element "red" and "gray60"
cols<-c("red","gray60")

#16 converting tmp era values to numeric
num=as.numeric(tmp$era)
print(num)

#17 intializing the col vector with tmp era values
cols[tmp$era]

#18 creating plot between age and fork lengh
```

```
plot(age~fl,data=filtered_data,main="Plot 4: Symbol & color by Era",
    xlim=c(0,500),ylim=c(0,15),ylab="Age(yrs)",
    xlab="Fork Length(mm)",pch=pchs,col=cols)

#19 plotting a regression line ovea the plot 4
plot(age~fl,data=filtered_data,main="Plot 5: Regression Overlay",
    xlim=c(0,500),ylim=c(0,15),ylab="Age(yrs)",
    xlab="Fork Length(mm)",pch=pchs,col=cols)
regression_line=lm(age~fl,data=filtered_data)
abline(regression_line,lty=2,lwd=2)

#20 placing a legend on plot 5
plot(age~fl,data=filtered_data,main="Plot 6: Legend Overlay",
    xlim=c(0,500),ylim=c(0,15),ylab="Age(yrs)",
    xlab="Fork Length(mm)",pch=pchs,col=cols)
regression_line=lm(age~fl,data=filtered_data)
abline(regression_line,lty=2,lwd=2)
#placing legend at the top of the graph
legend("topleft", inset=c(0.05), legend=levels(filtered_data$era), pch=pchs,
      col=cols,bty="n",title="Era",cex = 0.70)
```