Northeastern University

Soni_M2_Project2

Soni Manan[a]

[a] *College of Professional Studies, Master of Professional Studies in Analytics. Toronto*

Under the guidance of

Prof. Dr. Shafiqul

Introduction

Bull Trout fishes can be seen in Canada as well as Pacific Northwest. In the historical (1977-1980) era samples were from before restrictive sportfishing regulatory regimes were implemented (in the 1990s) that led to changes in abundance and population structure of bull trout.[1] All populations of the bull trout were listed as Endangered in 1998 so the Act was implemented.[2] This executive summary is focused on samples taken from Harrison Mountain Lake of Alberta, Canada before and after sporting-regulation change. In sample, key features taken are age and fork length where age has been assigned from fish's otolith (ear stone or ear bone) which is the most commonly used structure for determining the age of fish[3] and provides essential insights.

Characteristic of Data

Data has 4 columns such age, fork length, lake and era, and has no missing/null values. The mean of age and fork length are respectively 5.74 and 319 respectively and their standard deviation are 3.34 and 128.62 respectively which demonstrat that data is more spread out. Skewness of age is 0.167 indicating normal distribution and for fork length its -0.735 which implies that most of the samples are greater than its mean.[5]

Analysis of 1977-1980 Era

Plot 1: Harrison Lake Trout

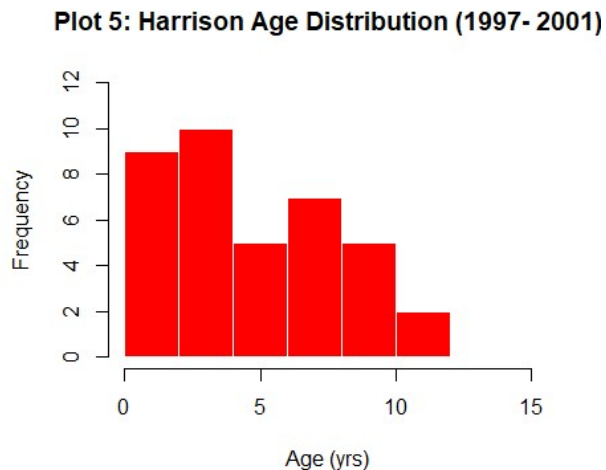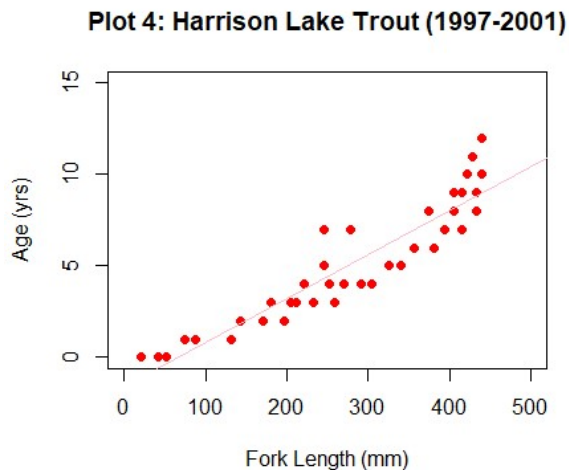Plot 2: Harrison Fish Length Distribution
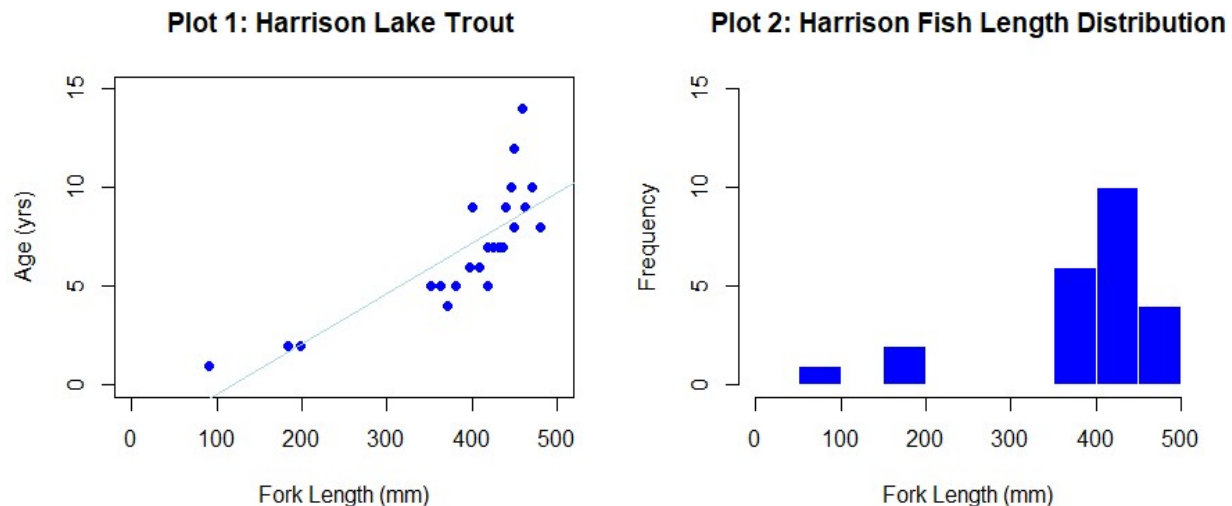
According to Harrison Lake data, in 1977-1980 era, Plot 2 describes that bull trout fishes used to have fork length of 350 mm or more . Majority of samples that were caught were found to be of 4 to 11 years in age as shown in Plot 3.


Plot 3: Harrison Age Distribution (1977- 1980)
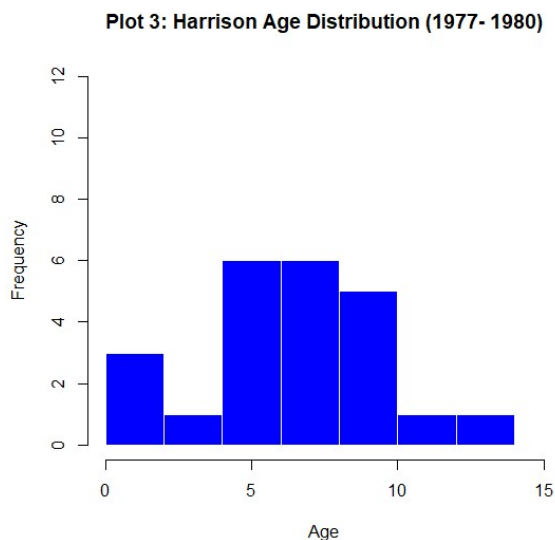
Analysis of 1997-2001 Era

In 1997-2001 era, bull trout fishes showed a linear increase in their fork length as their age advances as shown in Plot 4. According to Plot 5, observations are likely to be inclined toward fishes of age 0-4 years. As seen in Plot 6, fishes which were caught were likely to be 200 mm or above in fork length.


Plot 4: Harrison Lake Trout (1997-2001)

Plot 5: Harrison Age Distribution (1997- 2001)

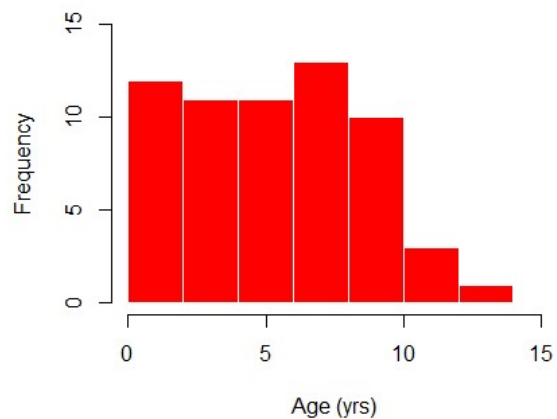Plot 6: Harrison Fork Length Distribution (1997- 2001)
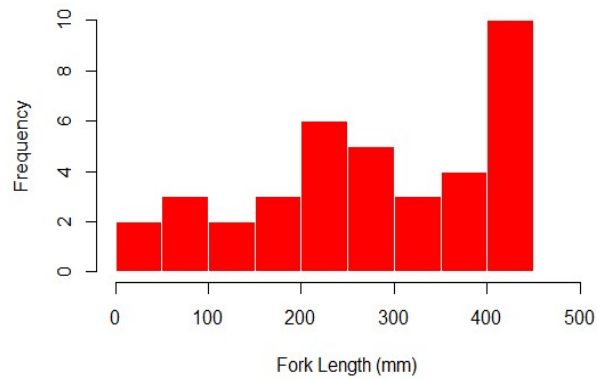
Conclusion

We can conclude that after prohibition of sportfishing of bull trout, young fishes started to appear in Harrison Lake as comparing Plot 4 with Plot 1. In addition, if we observe not a single fish crossed the fork length mark of 500 mm and age of 15 years in both the eras. Thus, the lifespan of Bull trout is up to 15 years with majority of them living for less than 11 years as shown in plot 8. The one limitation of the data is number of samples for age 5 year and less are very few compared to others as shown in Plot 7.

Plot 7: Harrison Density Shaded by Era

6
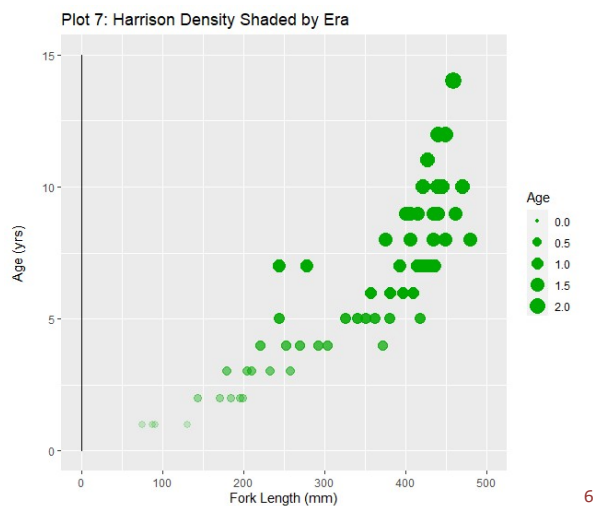
Plot 8: Harrison Age Distribution

Bibliography

[1] C.C., D. H.-H. (2019, May 18). *Package 'FSAdata'*. Retrieved from FSAdata Documentation: https://cran.r-project.org/web/packages/FSAdata/FSAdata.pdf

[1] *U.S. Fish and Wildlife Fish and Aquatic Conservation*. (n.d.). Retrieved from U.S. Fish and Wildlife Service: https://www.fws.gov/fisheries/freshwater-fish-of-america/bull_trout.html

[1] *Exploratory Data Analysis with R*. (n.d.). Retrieved from Bookdown.org: https://bookdown.org/rdpeng/exdata/exploratory-graphs.html

[4] *U.S. Fish and Wildlife Fish and Aquatic Conservation*. (n.d.). Retrieved from U.S. Fish and Wildlife Service: https://www.fws.gov/fisheries/freshwater-fish-of-america/bull_trout.html

[5] @utkarsh_kumar, U. K. (2022, Jan 25). *Geeks for geeks*. Retrieved from Skewness and Kurtosis in R Programming: https://www.geeksforgeeks.org/skewness-and-kurtosis-in-r-programming/

[6] Hayden, L. (2018, Mar 30). *Graph density with scatterplot ggplot2 in R*. Retrieved from Stackover Flow: https://stackoverflow.com/questions/49573013/graph-density-with-scatterplot-ggplot2-in-r

[7] Hadley, et al. "Subset rows using column values." *the Grammar of Data Manipulation • dplyr*, RStudio, https://dplyr.tidyverse.org/index.html.

Appendix

```r
# Ploting Basics: Soni

# 1. Install one or more packages
# install.packages(c('FSAdata','FSA','dplyr','plotrix','moments','ggplot2'))

# 2. Importing
library('FSAdata')
library('magrittr')
library('FSA')
library('plotrix')
library('moments')
library('dplyr')
# 3. Loading Dataset into <dataset>
try(data("BullTroutRML2"))
# 4. Displaying last and first 3 rows
headtail(BullTroutRML2, n=3)

# Taking out unique from lake column
uniques = unique(BullTroutRML2$lake)

# 5. Filtering with 1st unique value
harrisonLakeData = BullTroutRML2 %>%  filter(BullTroutRML2$lake == uniques[1])

# For seperate observations on both Era
# data of 1977
dfOf1977 = harrisonLakeData %>%  filter(harrisonLakeData$era == "1977-80")
# data of 1997
dfOf1997 = harrisonLakeData %>%  filter(harrisonLakeData$era == "1997-01")

# 6. Displaying last and first 5 rows
headtail(harrisonLakeData,n=5)

# 7. Displaying Structure of Dataset
str(data.frame(harrisonLakeData))

# 8. Displaying summary of Dataset
summary(data.frame(harrisonLakeData))

# 9. Plotting Scatter Plot
plot(harrisonLakeData$age  ~  harrisonLakeData$fl,
    ylab="Age (yrs)",
    xlab="Fork Length (mm)",
    pch=19,
    ylim=c(0,15),
    xlim=c(0,500),
    col = "blue",
    cex= 1,
    main = "Plot 1: Harrison Lake Trout")
# 10. Plotting Harrison Lake Trout
hist(x= harrisonLakeData$age,col='red',border='white',breaks=8,ylim= c(0,15), xlim = c(0,15),x
lab="Age (yrs)",main="Plot 2: Harrison Age Distribution")

# 11. Density Plot in green shade
library('ggplot2')
Age =harrisonLakeData$age/mean(harrisonLakeData$age)
```

```r
ggplot(data=harrisonLakeData, aes(y=age)) +geom_density()+
  geom_point(data=harrisonLakeData, aes(y=age, x=fl,size=Age),alpha = Age,colour = "#00aa00")
+
  labs(
    title = "Plot 7: Harrison Density Shaded by Era", # adds title
    x = "Fork Length (mm)", # x-axis label
    y = "Age (yrs)\n", # y-axis label
    color = "Sales" # color legend
  ) + lims(x = c(0,500),y=c(0,15))
# 12.creating tmp and displaying headtail where n =3
tmp = headtail(harrisonLakeData,n=3)
print(tmp)

# 13. Era column of <tmp>
print(tmp$era)

# 14. Creating <pchs> for ploting characters + and x
pchs = c(3,4)

# 15. creating <cols> for coloring in red and gray60
cols = c('red','gray60')

# 16. converting <tmp> era values to 1,2,3 by its value using as.numeric
tmp$era = as.numeric(tmp$era)

# 17. Initializing <cols> to <tmp> era value
cols = tmp$era

# 18. Plotting graph of <tmp> by different era color
#par(mfrow=c(2,2),bg="#ffffff") # for colxrow layout
plot(tmp$age ~ tmp$fl,ylim=c(0,15), xlim=c(0,500),xlab="Fork Lenght (mm)",ylab="Age (yrs)",col
=cols,pch=pchs)
title('Plot 4: Symbol & Color by Era')

# 19. Drawing regression line on Plot:4
plot(tmp$age ~ tmp$fl,ylim=c(0,15), xlim=c(0,500),xlab="Fork Lenght (mm)",ylab="Age (yrs)",col
=cols,pch=pchs)
abline(lm(tmp$age ~ tmp$fl),)
title('Plot 5: Regression Overlay')

# 20. Plotting Legends of Plot 4, Plot 5
plot(tmp$age ~ tmp$fl,ylim=c(0,15), xlim=c(0,500),xlab="Fork Lenght (mm)",ylab="Age (yrs)",col
=cols,pch=pchs)
abline(lm(tmp$age ~ tmp$fl),)
legend(1,15, legend=c("in 1977-80", "in 1997-01"),
       col=cols, pch=pchs, cex=0.8)
title('Plot 6: Legend Overlay')

# 21. Github Repo
# https://github.com/Mxnxn/Intro-w-R/tree/master/week_2_ALY6000
```