Name: Husain Rangwala

Date: 04th February 2021

Title: Module 2 Project

Executive Summary Report 2

Key findings

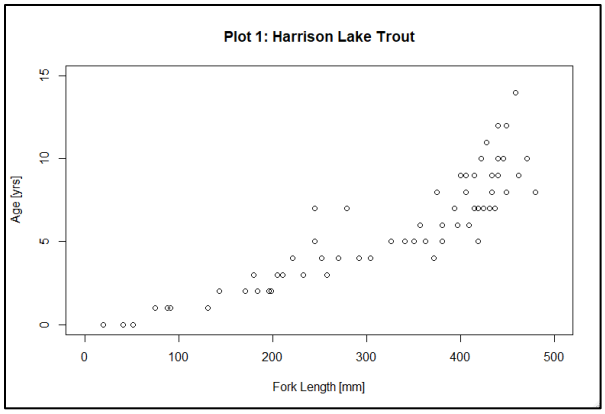
There are three aspects of this assignment as follows:

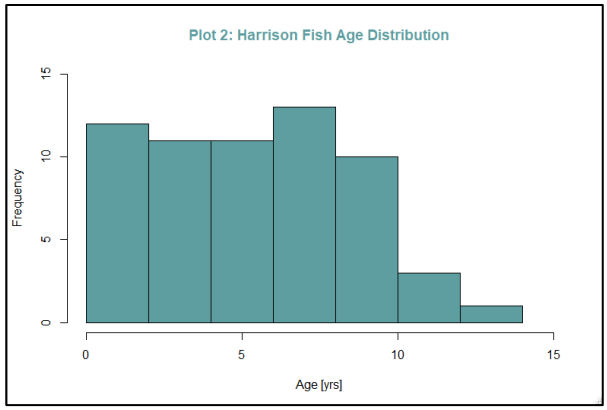
Provide an analysis of descriptive characteristics of the data set provided by your instructor. This includes pertinent statistics including mean, median, quartiles, variance, standard deviation, skew, kurtosis, outliers etc. Include R console screen shots to support your observations and conclusions.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

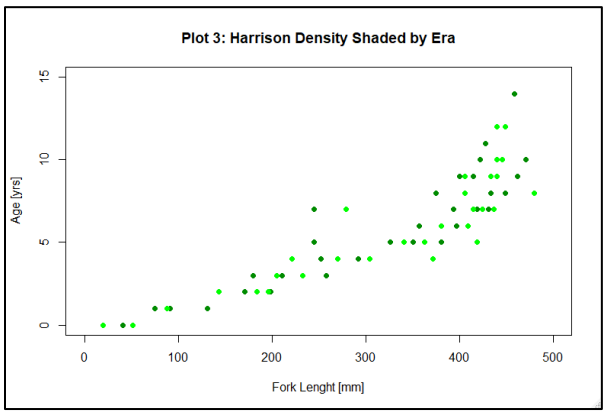
From the data present in the BullTroutRML2 dataset, the distributing factors are the age, fork length of bull trout from two lakes (Harrison and Osprey) and two different time period 1977-80 & 1997-01.

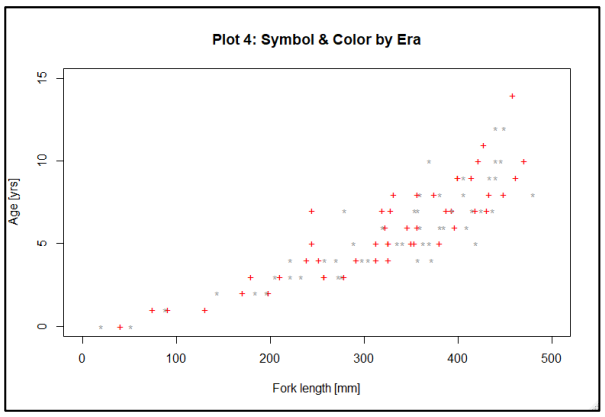
```
> headtail(BullTrout, n=5)
age  fl    lake    era
1   14 459 Harrison 1977-80
2   12 449 Harrison 1977-80
3   10 471 Harrison 1977-80
4   10 446 Harrison 1977-80
5    9 400 Harrison 1977-80
92   5 289   Osprey 1997-01
93   4 313   Osprey 1997-01
94   4 298   Osprey 1997-01
95   3 279   Osprey 1997-01
96   3 273   Osprey 1997-01
```

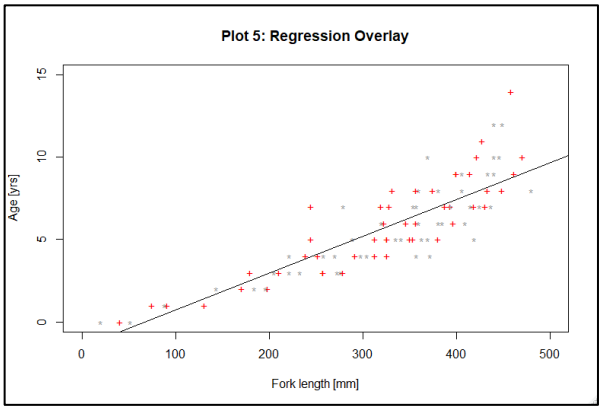
Provide the executive with visualizations (at least 6) in that help them see the key characteristics you want to highlight. They can be boxplots, histograms, frequency and probability distributions, bar plots (bar charts) or pareto. Not only is the goal to present your visual results, but also to explain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

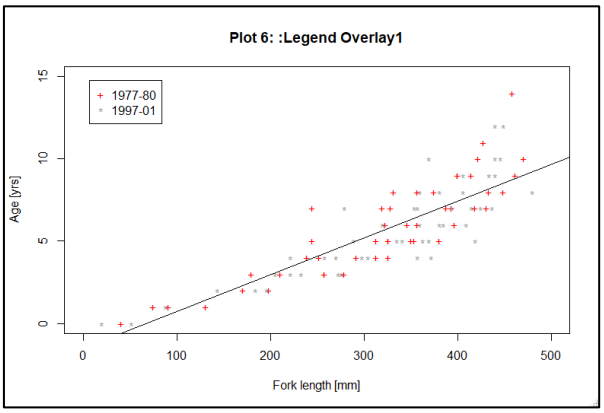












Finally, provide a clear two to three sentence paragraph summary of the key points that you want the audience to walk away with regarding your analysis. This summary should present accurate analysis and be supported by the data presented in the rest of the report.

By separating the Dataset on the bases on the origin of the lake, we could provide more analysis on the dataset. By comparing the Harrison lake data, we could study the length and age of the Bull Trout on the

basis of particular year. From the data, we could examine the life cycle, the fork length and the year then compare to get the optimal age to length to year value. By plot the age to fork length graph, it could incur that the as the trout ages increases past in mean the fork length of the trout reaches to 500. Thus we could theorize that there is a linear relationship between the age and the fork length of the Bull Trout.

We could use the data from the above analysis and then predict the Bull Trout length and age for the Osprey lake.

> headtail(Newdata, n=5)

	age	fl	lake	era
1	14	459	Harrison	1977-80
2	12	449	Harrison	1977-80
3	10	471	Harrison	1977-80
4	10	446	Harrison	1977-80
5	9	400	Harrison	1977-80
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

Bibliography

Below link were used to help solving any blocker in R scripting.

1. https://www.geeksforgeeks.org/adding-elements-in-a-vector-in-r-programming-append-method/
2. https://discuss.analyticsvidhya.com/t/how-to-remove-value-from-a-vector-in-r/2975
3. https://www.dummies.com/programming/r/how-to-work-with-variable-names-in-r/
4. https://www.youtube.com/watch?v=rgAvJmvfA2c
5. R in Action: Data Analysis and Graphics with R By Robert Kabacoff

Appendix

```
#-----------------------------------#
#  Plotting Basics: Rangwala   #
#-----------------------------------#


#Import libraries including: FSA, FSAdata, magrittr, dplyr, plotrix, ggplot2, and moments#
install.packages(c("FSA","FSAdata","dplyr","plotrix","ggplot2","moments"))
help(package="FSA")
help(package="FSAdata")
help(package="dplyr")
help(package="plotrix")
help(package="ggplot2")
help(package="moments")


library(FSA)
library(FSAdata)
library(dplyr)
library(plotrix)
library(ggplot2)
library(moments)



#Load the BullTroutRML2 dataset (BullTroutRML2.csv)#
BullTrout <- as_tibble(BullTroutRML2)
BullTrout
class(BullTrout)
#Print the first and last 3 records from the BullTroutRMS2 dataset#
BullTrout1 <- headtail(BullTrout, n=3)
BullTrout1
```

```
#Remove all records except those from Harrison Lake #
Newdata <- subset(BullTrout, lake == "Harrison")
Newdata


#Display the first and last 5 records from the filtered BullTroutRML2 dataset#
Newdata1 <- headtail(Newdata, n=5)
Newdata1


#Display the structure of the filtered BullTroutRML2dataset#
str(Newdata)
#Display the summary of the filtered BullTroutRML2dataset#
summary(Newdata)


#Create a scatterplot for “age” (y variable) and “fl” (x variable) with the following
#specifications:
#Limit of x axis is (0,500)
#Limit of y axis is (0,15)
#Title of graph is “Plot 1: Harrison Lake Trout
#Y axis label is “Age (yrs)”
#X axis label is “Fork Length (mm)”
#Use a small filled circle for the plotted data points#
plot(Newdata$fl,Newdata$age,
    xlim=c(0,500),ylim=c(0,15)
    ,ylab = "Age [yrs]",xlab = "Fork Length [mm]",
    title("Plot 1: Harrison Lake Trout"))
```

```
#. Plot an “Age” histogram with the following specifications
#Y axis label is “Frequency”
#X axis label is “Age (yrs)”
#Title of the histogram is “Plot 2: Harrison Fish Age Distribution”
#X and Y axis limits is 0, 15
#The color of the frequency plots is “cadetblue”
#The color of the Title is “cadetblue”


hist(Newdata$age,main = "Plot 2: Harrison Fish Age Distribution",
    xlim = c(0,15),ylim = c(0,15),
    xlab = "Age [yrs]", ylab = "Frequency",
    col.main="cadetblue", col = "cadetblue")


#Create an overdense plot using the same specifications as the previous scatterplot. But,
# Title the plot “Plot 3: Harrison Density Shaded by Era”
# Y axis label is “Age (yrs)”
# Y axis limits are 0 to 15
# X axis label is “Fork Length (mm)”
# X axis limits are 0 to 500
# include two levels of shading for the “green” data points.
# Plot solid circles as data points
#
plot(Newdata$fl,Newdata$age,
   main = "Plot 3: Harrison Density Shaded by Era",
  ylab = "Fork Lenght [mm]", xlab = "Age [yrs]",
  xlim = c(0,500), ylim = c(0,15),
  col=c("green4", "green1"),pch=19)
```

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

tmp <- headtail(Newdata1, n=3)

tmp


#Display the “era” column (variable) in the new “tmp” object#

Eravalue <- tmp$era

Eravalue


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

pchs <- c("+", "*")

pchs


# Create a cols vector with the two elements “red” and “gray60”#

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

cols


# Convert the tmp era values to numeric values#

tmp$eranew <- as.numeric(tmp$era)

tmp


#Initialize the cols vector with the tmp era values#

tmp$cols <- cols

tmp
```

```
#Create a plot of “Age (yrs)” (y variable) versus “Fork Length (mm)” (x variable) with the
#following specifications:
#Title of graph is “Plot 4: Symbol & Color by Era”
#Limit of x axis is (0,500)
#Limit of y axis is (0,15)
#X axis label is “Age (yrs)”
#Y axis label is “Fork Length (mm)”
#Set pch equal to pchs era values
#Set col equal to cols era values


plot(BullTrout$fl,BullTrout$age,
    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)


#Plot a regression line overlay on Plot 4 and title the new graph “Plot 5: Regression Overlay”.#
plot(BullTrout$fl,BullTrout$age,
    main = "Plot 5: Regression Overlay",
    xlim = c(0,500), ylim = c(0,15),
    ylab = "Age [yrs]", xlab = "Fork length [mm]",
    pch = pchs, col = cols)
abline(lm(age ~ fl, data = BullTrout))
```

```
#Place a legend of on Plot 5 and call the new graph “Plot 6: :Legend Overlay#
plot(BullTrout$fl,BullTrout$age,
    main = "Plot 6: :Legend Overlay1",
    xlim = c(0,500), ylim = c(0,15),
    ylab = "Age [yrs]", xlab = "Fork length [mm]",
    pch = pchs, col = cols)
abline(lm(age ~ fl, data = BullTrout))
legend("topleft",legend = c("1977-80", "1997-01"),
       col = cols, pch=pchs, inset = 0.05)
```