

Name: Husain Rangwala

Date: 12th February 2021

Title: Module 3 Project

Executive Summary Report 3

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 counts, cumulative counts, and frequency, percentages, etc. Include R console screen shots to support your observations and conclusions. Below is a sample excerpt.

A systemic review of the data collection of *inchBio* disclosed that the research was dominated by two major type of species that are; Bluegill and largemouth bass

	Species	RelFreq
1	Black Crappie	36
2	Bluegill	220
3	Bluntnose Minnow	103
4	Iowa Darter	32
5	Largemouth Bass	228
6	Pumpkinseed	13
7	Tadpole Madtom	6
8	Yellow Perch	38

These two species accounts to almost 2/3th of the total count from the dataset.

On that reaction, many smaller groups of largemouth bass were generated in order to the study various aspect of this fish species.

> headtail(temp)

	netID	fishID	species	tl	w	tag	scale
356	14	85	Largemouth Bass	68	3.5		TRUE
357	3	89	Largemouth Bass	338	NA	g0996	TRUE
358	1	90	Largemouth Bass	356	NA	y06073	TRUE
581	131	913	Largemouth Bass	251	180.0		TRUE
582	131	914	Largemouth Bass	378	646.0	1709	TRUE
583	131	915	Largemouth Bass	325	377.0	1785	FALSE

From the above head and tail values for the Largemouth Bass subset was developed to examine the length characteristics Largemouth Bass with a weight ranging from 3.5 to 377 and a length greater than 60. Notice that in this subset, there were 227 documents.

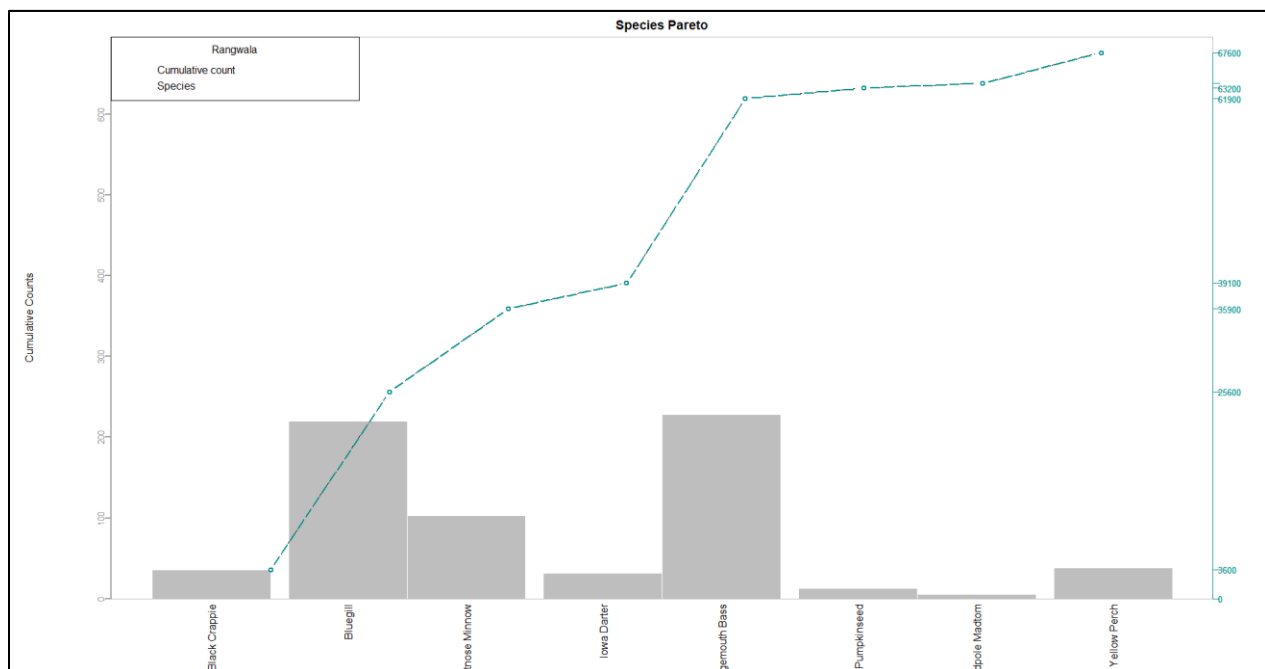
- Provide the executive with visualizations (at least 3) in that help them see the key characteristics you want to highlight. They can be boxplots, histograms, frequency and probability distributions, or bar plots (bar charts). A pareto plot as illustrated below must be included in this part of your report. Include screen shots of your plots to support your findings and conclusions. The goal is not only to present your visual results, but also to explain the significance of them.

The Pareto graph blends the advantages of a bar plot with a cumulative distribution graph.

From the sample, it could be concluded that the Bluegill and largemouth bass occupies the most space among the fish species.

The Pareto map helps explore a variety of new areas of inquiry.

1. Are the bluegill and largemouth bass fish predatory species that exploit smaller tadpole madtom?
2. Does the environment play a key role for some of these fishes?
3. By further analysis we could determine if species are underperforming and what may be the cause and risk in the mortality.



- 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.

From the inchBio dataset, I could summarize that Bluegill and largemouth bass occupy almost 2/3th of the data. Some of the fewer fish species may be under mortality risk in conjunction with the bluegill and largemouth bass. A study can be carried out by tagging the fewer fish species and creating a suitable environment for them to flourish.

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
6. <https://www.statmethods.net/advgraphs/parameters.html>

Appendix

```
#-----#
```

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

```
#-----#
```

```
#. Print your name at the top of the script and load these libraries: FSA, FSAdata, magrittr,
```

```
#dplyr, tidyr plyr and tidyverse#
```

```
install.packages(c("FSA","FSAdata","magrittr","dplyr","tidyr","plyr","tidyverse"))
```

```
help(package="FSA")
```

```
help(package="FSAdata")
```

```
help(package="dplyr")
```

```
help(package="magrittr")
```

```
help(package="tidyr")
```

```
help(package="plyr")
```

```
help(package="tidyverse")
```

```
library(FSA)
```

```
library(FSAdata)
```

```
library(dplyr)
```

```
library(magrittr)
```

```
library(tidyr)
```

```
library(plyr)
```

```
library(tidyverse)
```

```
#Import the inchBio.csv and name the table#
```

```
inchBio=read.table("inchBio.csv", header = TRUE, sep = ",")
```

```
inchBio
```

```
#Display the head, tail and structure of table#
```

```
head(inchBio, n=1)
```

```
tail(inchBio, n=1)
```

```
str(inchBio)
```

```
#Create an object, <counts>, that counts and lists all the species records#
```

```
counts=unique(inchBio$species)
```

```
#Display just the 8 levels (names) of the species#
```

```
counts
```

```
#Create a <tmp> object that displays the different species and the number of record of
```

```
#each species in the dataset.#
```

```
tmpvariable=count(inchBio, vars = "species")
```

```
tmpvariable
```

```
#Create a subset, <tmp2>, of just the species variable and display the first five records#
```

```
tmpvariable2=subset(inchBio, select = c(species))
```

```
head(tmpvariable2, 5)
```

```
#Create a table, <w>, of the species variable. Display the class of w#
```

```
w= table(tmpvariable)
```

```
class(w)
```

```
#Convert <w> to a data frame named <t> and display the results#
```

```
t=as.data.frame(w)
```

```
t
```

```
#Extract and display the frequency values from the <t> data frame#
```

```
t[,2]
```

```
#Create a table named <cSpec> from the bio species attribute (variable) and confirm that#
```

```
#you created a table which displays the number of species in the dataset#
```

```
Spec= table(inchBio$species)
```

```
Spec
```

```
#Create a table named <cSpecPct> that displays the species and percentage of records for#
```

```
#each species. Confirm you created a table class#
```

```
SpecPct = table(Spec)
```

```
SpecPct = cbind(Spec/margin.table(Spec)*100)
```

```
colnames(SpecPct)= c("Percentage")
```

```
SpecPct
```

```
#Convert the table, <cSpecPct>, to a data frame named <u> and confirm that <u> is a data frame#
```

```
u = as.data.frame(SpecPct)
```

```
u
```

```
is.data.frame(u)
```

```
#Create a barplot of <cSpec> with the following: titled Fish Count with the following
```

```
#specifications:
```

```
#Title: Fish Count
```

```
#Y axis is labeled "COUNTS"
```

```
#Color the bars Light Green
```

```
#Rotate Y axis to be horizontal
```

```
#Set the X axis font magnification to 60% of nominal#
```

```
barplot(Spec,main = "Fish Count",ylab = "COUNTS", col="lightgreen", horiz = TRUE, cex.names = 0.6)
```


#Create a barplot of <cSpecPct>, with the following specifications:

#Y axis limits of 0 to 4

#Y axis label color of Light Blue

#Title of "Fish Relative Frequency" #

```
barplot(SpecPct,ylim = c(0,4), col = "lightblue", main = "Fish Relative Frequency")
```

#Rearrange the <u> cSpec Pct data frame in descending order of relative frequency. #

#Save the rearranged data frame as the object#

```
uarrange = sort(SpecPct, decreasing = TRUE)
```

```
uarrange <- as.data.frame(uarrange)
```

```
uarrange
```

#Rename the <d> columns Var 1 to Species, and Freq to RelFreq#

```
d <- as.data.frame(Spec)
```

```
colnames(d)<- c("Species","RelFreq")
```

```
d
```

#Add new variables to <d> and call them cumfreq, counts, and cumcounts#

```
d
```

```
d[, "cumfreq"] <- cumsum(d$RelFreq)
```

```
d[, "counts"] <- d$RelFreq
```

```
d[, "cumcounts"] <- cumsum(d$counts)
```

```
d
```

#Create a parameter variable <def_par> to store parameter variables#

```
#def_par=inchBio$tl+inchBio$w
```

```
def_par
```

#Create a barplot, <pc>, with the following specifications:

#d\$counts of width 1, spacing of .15

#no boarder

#Axes: F

#Yaxis limit 0,3.05*max

#d\$counts na.rm is true

#y label is Cumulative Counts

#scale x axis to 70%

#names.arg: d\$Species

#Title of the barplot is "Species Pareto"

#las: 2)#

par(mar=c(8,8,2,5), mgp=c(6,0.4,0))

barplot(d\$counts, width = 1, space = c(0, .15), border = F,

ylim = c(0,3.05*max(d\$counts, na.rm=T)),

ylab = "Cumulative Counts", names.arg = d\$Species,

cex.axis = 0.7, main = "Species Pareto", las=2)

#Add a cumulative counts line to the <pc> plot with the following:

#Spec line type is b

#Scale plotting text at 70%

#Data values are solid circles with color cyan4#

px = lines(d\$cumfreq ,type="b",lty=5,col="cyan4",lwd=2)

#Place a grey box around the pareto plot#

```

box(lty = "solid", col = "grey")

#Add a left side axis with the following specifications
#Horizontal values at tick marks at cumcounts on side 2
#Tickmark color of grey62
#Color of axis is grey62
#Axis scaled to 80% of normal#
par(mar=c(8,8,2,5), mgp=c(6,0.4,0))
barplot(d$counts, width = 1, space = c(0, .15), border = F,
        ylim = c(0,3.05*max(d$counts, na.rm=T)),
        ylab = "Cumulative Counts", names.arg = d$Species,
        cex.axis = 0.7, main = "Species Pareto", las=2, yaxt="n")
box(lty = "solid", col = "grey")
px=lines(d$cumfreq ,type="b",lty=5,col="cyan4",lwd=2)
axis(2, at=pretty(range(d$cumfreq)), lwd=2, col = "grey62", cex.axis= 0.8 , col.axis="grey62")
px

#Add axis details on right side of box with the specifications:
#Spec: Side 4
#Tickmarks at cumcounts with labels from 0 to cumfreq with %,
#Axis color of cyan5 and label color of cyan4
#Axis font scaled to 80% of nominal#

axis(4, at= c(0,d$cumfreq),labels = paste(c(0, round(d$cumfreq * 100))),"%",
      las=1, col = "cyan4",cex.axis=0.8, col.axis="cyan4")

#Display the finished Species Pareto Plot (without the star watermarks). Have your last name on the
plot#

```

```
par(mar=c(8,8,2,5), mgp=c(6,0.4,0))  
barplot(d$counts, width = 1, space = c(0, .15), border = F,  
        ylim = c(0,3.05*max(d$counts, na.rm=T)),  
        ylab = "Cumulative Counts", names.arg = d$Species,  
        cex.axis = 0.7, main = "Species Pareto", las=2, yaxt="n")  
box(lty = "solid", col = "grey")  
lines(d$cumfreq ,type="b",lty=5,col="cyan4",lwd=2)  
axis(2, at=pretty(range(d$cumfreq)), lwd=2, col = "grey62", cex.axis= 0.8 , col.axis="grey62")  
axis(4, at= c(0,d$cumfreq),labels = paste(c(0, round(d$cumfreq * 100))),"%",  
        las=1, col = "cyan4",cex.axis=0.8, col.axis="cyan4")  
legend("topleft",title = "Rangwala",legend = c("Cumulative count","Species"))
```