



Executive Summary Report 3

Name : Shamim Sherafati

ALY 6000 – Introduction to Analytics - Module 3 - Project

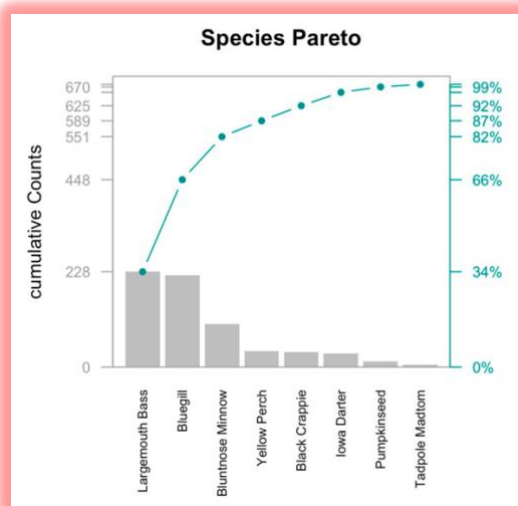
Date : 2022-10-11

Campus : Vancouver

NUID : 002742363

Abstract

The Report that follows is based on ALY 6000 Module 3. This is authentic work that Miss. Shamim Sherafati completed for the assignment ALY 6000:Introduction to Analytics, and she is submitting it on time, before the decided date and time.



KEY FINDINGS ABOUT THE DATASET

1. Print your name at the top of the script and load these libraries: FSA, FSAdat, magrittr, dplyr, tidyr plyr and tidyverse.

Describe number 1: First, We should loading the libraries and recording our names.

```
name <- ("Shamim Sherafati")
print(name)

[1] "Shamim Sherafati"

install.packages("FSA")

trying URL 'https://mirror.rcg.sfu.ca/mirror/CRAN/bin/macosx/contrib/4.2/FSA_0.9.3.tgz'
Content type 'application/octet-stream' length 1117449 bytes (1.1 MB)
=====
downloaded 1.1 MB

The downloaded binary packages are in
  /var/folders/cn/j3k0jn5d6qvctnpb78k90_rm0000gn/T//RtmpyXYgbk/downloaded_packages
install.packages("FSAdat")

trying URL 'https://mirror.rcg.sfu.ca/mirror/CRAN/bin/macosx/contrib/4.2/FSAdat_0.3.9.tgz'
Content type 'application/octet-stream' length 919196 bytes (897 KB)
=====
downloaded 897 KB

The downloaded binary packages are in
  /var/folders/cn/j3k0jn5d6qvctnpb78k90_rm0000gn/T//RtmpyXYgbk/downloaded_packages
install.packages("magrittr")
Error in install.packages : Updating loaded packages
install.packages("plyr")

trying URL 'https://mirror.rcg.sfu.ca/mirror/CRAN/bin/macosx/contrib/4.2/plyr_1.8.7.tgz'
Content type 'application/octet-stream' length 1014519 bytes (990 KB)
=====
downloaded 990 KB
```

```
The downloaded binary packages are in
  /var/folders/cn/j3k0jn5d6qvctnpb78k90_rm0000gn/T//RtmpyXYgbk/downloaded_packages
install.packages("dplyr")
Error in install.packages : Updating loaded packages
install.packages("tidyr")
trying URL 'https://mirror.rcg.sfu.ca/mirror/CRAN/bin/macosx/contrib/4.2/tidyr_1.2.1.t
gz'
Content type 'application/octet-stream' length 998893 bytes (975 KB)
=====
downloaded 975 KB

The downloaded binary packages are in
  /var/folders/cn/j3k0jn5d6qvctnpb78k90_rm0000gn/T//RtmpyXYgbk/downloaded_packages
install.packages("magrittr")
trying URL 'https://mirror.rcg.sfu.ca/mirror/CRAN/bin/macosx/contrib/4.2/magrittr_2.0.
3.tgz'
Content type 'application/octet-stream' length 227506 bytes (222 KB)
=====
downloaded 222 KB

The downloaded binary packages are in
  /var/folders/cn/j3k0jn5d6qvctnpb78k90_rm0000gn/T//RtmpyXYgbk/downloaded_packages
install.packages("dplyr")
trying URL 'https://mirror.rcg.sfu.ca/mirror/CRAN/bin/macosx/contrib/4.2/dplyr_1.0.10.
tgz'
Content type 'application/octet-stream' length 1310338 bytes (1.2 MB)
=====
downloaded 1.2 MB

The downloaded binary packages are in
  /var/folders/cn/j3k0jn5d6qvctnpb78k90_rm0000gn/T//RtmpyXYgbk/downloaded_packages
install.packages("tidyverse")
trying URL 'https://mirror.rcg.sfu.ca/mirror/CRAN/bin/macosx/contrib/4.2/tidyverse_1.3
.2.tgz'
Content type 'application/octet-stream' length 420896 bytes (411 KB)
=====
downloaded 411 KB
```

```
The downloaded binary packages are in
/var/folders/cn/j3k0jn5d6qvctnpb78k90_rm0000gn/T//RtmpyXYgbk/downloaded_packages
```

+1. Call library

```
library(FSA)
## FSA v0.9.3. See citation('FSA') if used in publication.
## Run fishR() for related website and fishR('IFAR') for related book.
library(FSAdata)
## FSAdata v0.3.9. See ?FSAdata to find data for specific fisheries analyses.
library(magrittr)
library(plyr)

Attaching package: 'plyr'

The following object is masked from 'package:FSA':

    mapvalues
library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:plyr':

    arrange, count, desc, failwith, id, mutate, rename, summarise, summarize

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
library(tidyr)

Attaching package: 'tidyr'
```

The following object is masked from 'package:magrittr':

```

      extract
library(tidyverse)
Registered S3 methods overwritten by 'dbplyr':
  method      from
  print.tbl_lazy
  print.tbl_sql

— Attaching packages —
— tidyverse 1.3.2 — ✓ ggplot2 3.3.6    ✓ purrr 0.3.4
✓ tibble 3.1.8    ✓ stringr 1.4.1
✓ readr 2.1.2    ✓ forcats 0.5.2— Conflicts —
                        tidyverse_conflicts() —

✗ dplyr::arrange() masks plyr::arrange()
✗ purrr::compact() masks plyr::compact()
✗ dplyr::count() masks plyr::count()
✗ tidyr::extract() masks magrittr::extract()
✗ dplyr::failwith() masks plyr::failwith()
✗ dplyr::filter() masks stats::filter()
✗ dplyr::id() masks plyr::id()
✗ dplyr::lag() masks stats::lag()
✗ dplyr::mutate() masks plyr::mutate()
✗ dplyr::rename() masks plyr::rename()
✗ purrr::set_names() masks magrittr::set_names()
✗ dplyr::summarise() masks plyr::summarise()
✗ dplyr::summarize() masks plyr::summarize()

```

2. Import the inchBio.csv and name the table .

Describe number 2: To begin analysing and processing the data, we should load the dataset in this stage.

```

bio <- read.table (file=~ /Documents/ALY 6000/Module 3/inchBio.csv", header= TRUE, sep
=",", stringsAsFactors = FALSE)

bio

```

3. Display the head, tail and structure of .

Describe number 3: Since it has been requested that we utilise the variable “bio” for the dataset, we will proceed as follows. Additionally, by using the head and tail function, we may examine the dataset’s beginning and last values in the following ways. Also, with structure function, we can find the data.frame values like: netID, fishID, species, tl and etc. and also it gives information about their numbers, too.

```
head(bio)
tail(bio)
str(bio)
'data.frame':  676 obs. of  7 variables:
 $ netID  : int  12 12 12 12 12 12 12 13 13 13 ...
 $ fishID : int  16 23 30 44 50 65 66 68 69 70 ...
 $ species: chr  "Bluegill" "Bluegill" "Bluegill" "Bluegill" ...
 $ tl     : int  61 66 70 38 42 54 27 36 59 39 ...
 $ w      : num  2.9 4.5 5.2 0.5 1 2.1 NA 0.5 2 0.5 ...
 $ tag    : chr  "" "" "" "" ...
 $ scale  : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

4. Create an object, , that counts and lists all the species records .

Describe number 4 : This demonstrates and confirms that there are 676 values in the Species column, which might be extremely helpful for the analysis.

```
counts <- c(bio$species)
counts
[1] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[6] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[11] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[16] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[21] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[26] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[31] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[36] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
[41] "Bluegill"      "Bluegill"      "Bluegill"      "Bluegill"      "Blu
egill"
```

[46] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[51] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[56] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[61] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[66] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[71] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[76] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[81] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[86] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[91] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[96] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[101] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[106] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[111] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[116] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[121] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[126] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[131] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[136] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[141] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[146] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[151] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu
[156] "Bluegill" egill"	"Bluegill"	"Bluegill"	"Bluegill"	"Blu

[illegible]


```

[276] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[281] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[286] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[291] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[296] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[301] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[306] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[311] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[316] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Blu
ntnose Minnow"

[321] "Bluntnose Minnow" "Bluntnose Minnow" "Bluntnose Minnow" "Iowa Darter" "Iow
a Darter"

[326] "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iow
a Darter"

[331] "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iow
a Darter"

[336] "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iow
a Darter"

[341] "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iow
a Darter"

[346] "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iow
a Darter"

[351] "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iowa Darter" "Iow
a Darter"

[356] "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Lar
gemouth Bass"

[361] "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Lar
gemouth Bass"

[366] "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Lar
gemouth Bass"

[371] "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Lar
gemouth Bass"

[376] "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Lar
gemouth Bass"

[381] "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Lar
gemouth Bass"

[386] "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Largemouth Bass" "Lar
gemouth Bass"

```

[illegible]

[506]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[511]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[516]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[521]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[526]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[531]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[536]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[541]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[546]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[551]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[556]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[561]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[566]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[571]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[576]	"Largemouth Bass" gemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Largemouth Bass"	"Lar
[581]	"Largemouth Bass" pkinseed"	"Largemouth Bass"	"Largemouth Bass"	"Pumpkinseed"	"Pum
[586]	"Pumpkinseed" pkinseed"	"Pumpkinseed"	"Pumpkinseed"	"Pumpkinseed"	"Pum
[591]	"Pumpkinseed" pkinseed"	"Pumpkinseed"	"Pumpkinseed"	"Pumpkinseed"	"Pum
[596]	"Pumpkinseed" pole Madtom"	"Tadpole Madtom"	"Tadpole Madtom"	"Tadpole Madtom"	"Tad
[601]	"Tadpole Madtom" low Perch"	"Tadpole Madtom"	"Yellow Perch"	"Yellow Perch"	"Yel
[606]	"Yellow Perch" low Perch"	"Yellow Perch"	"Yellow Perch"	"Yellow Perch"	"Yel
[611]	"Yellow Perch" low Perch"	"Yellow Perch"	"Yellow Perch"	"Yellow Perch"	"Yel
[616]	"Yellow Perch" low Perch"	"Yellow Perch"	"Yellow Perch"	"Yellow Perch"	"Yel

```
[621] "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yel
low Perch"

[626] "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yel
low Perch"

[631] "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yel
low Perch"

[636] "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yellow Perch"      "Yel
low Perch"

[641] "Black Crappie"      "Black Crappie"      "Black Crappie"      "Black Crappie"      "Bla
ck Crappie"

[646] "Black Crappie"      "Black Crappie"      "Black Crappie"      "Black Crappie"      "Bla
ck Crappie"

[651] "Black Crappie"      "Black Crappie"      "Black Crappie"      "Black Crappie"      "Bla
ck Crappie"

[656] "Black Crappie"      "Black Crappie"      "Black Crappie"      "Black Crappie"      "Bla
ck Crappie"

[661] "Black Crappie"      "Black Crappie"      "Black Crappie"      "Black Crappie"      "Bla
ck Crappie"

[666] "Black Crappie"      "Black Crappie"      "Black Crappie"      "Black Crappie"      "Bla
ck Crappie"

[671] "Black Crappie"      "Black Crappie"      "Black Crappie"      "Black Crappie"      "Bla
ck Crappie"

[676] "Black Crappie"
```

5. Display just the 8 levels (names) of the species .

Describe number 5 : The Species column contains 8 different categories, as seen by the code below. Let's check the unique values in the species column. We can discover that from the following.

```
Level8th <- unique(bio$species)

Level8th

[1] "Bluegill"      "Bluntnose Minnow" "Iowa Darter"      "Largemouth Bass"  "Pumpk
inseed"      "Tadpole Madtom"

[7] "Yellow Perch"      "Black Crappie"
```

6. Create a object that displays the different species and the number of record of each species in the dataset. Include this information in your report.

Describe number 6 : We can use the table() method to generate results in order to get the count of all the various rows in the 8 distinct types of Species variables. As a result, they are 8 different species with their number of records which can be seen in the following result. For example, Tadpole Madtom with 6 records has the lowest number of score. While, Largemouth Bass and Bluegill have the highest number of records with 228 and 220 respectively records.

```
tmp <- table (bio$species)
```

```
tmp
```

	Black Crappie Pumpkinseed	Bluegill Bluntnose Minnow	Iowa Darter	Largemouth Bass
13	36	220	103	32
				228
	Tadpole Madtom	Yellow Perch		
	6	38		

7. Create a subset, , of just the species variable and display the first five records .

Describe number 7 : In this stage, I created which will show the the species variable and with using the head function, I show the first five records of dataset.

```
tmp2 <- subset(bio, select = species)
tmp2

head(tmp2, 5)
```

8. Create a table, , of the species variable. Display the class of w .

Describe number 8 : Next, I created a table of those species variable and used the class function to be sure that its table.

```
w <- table (bio$species)
w
```

	Black Crappie Pumpkinseed	Bluegill Bluntnose Minnow	Iowa Darter	Largemouth Bass
13	36	220	103	32
				228
	Tadpole Madtom	Yellow Perch		
	6	38		

```
class(w)
[1] "table"
```

9. Convert to a data frame named and display the results .

Describe number 9: After that, I want to know the frequency of the values, and to achieve so, I'll convert it into a Data Frame and called it.

```
t <- as.data.frame(w)
```

```
t
class(t)
[1] "data.frame"
```

10. Extract and display the frequency values from the data frame .

Describe number 10: In this step, by using `t$Freq` show the frequency value which I convert it into a Data Frame which called .

```
t$Freq
[1] 36 220 103 32 228 13 6 38
```

11. Create a table named from the bio species attribute (variable) and confirm that you created a table which displays the number of species in the dataset .

Describe number 11: Next, if we want to figure out what percentage does every Specie hold in the dataset then we can do it using the percentage function. Since there are 676 species in total, the dataset will provide us with an estimate of the number of species, therefore we create a table called `cSpec`.

```
cSpec <- table(bio$species)
cSpec
```

Black Crappie Pumpkinseed	Bluegill	Bluntnose Minnow	Iowa Darter	Largemouth Bass
36	220	103	32	228
13				
Tadpole Madtom	Yellow Perch			
6	38			

```
class(cSpec)
[1] "table"
```

12. Create a table named that displays the species and percentage of records for each species. Confirm you created a table class.

Describe number 12: By dividing each species by 676 and multiply 100, we can obtain an estimation and percentage of the species proportion.

```
cSpecPct <- table(bio$species)/676* 100
cSpecPct
```

Black Crappie Pumpkinseed	Bluegill	Bluntnose Minnow	Iowa Darter	Largemouth Bass

```

      5.325444      32.544379      15.236686      4.733728      33.727811
1.923077
      Tadpole Madtom      Yellow Perch
      0.887574      5.621302
class(cSpecPct)
[1] "table"

```

13. Convert the table, , to a data frame named and confirm that is a data frame.

Describe number 13: In this dataset, we can use the following way to view and analyze the relative frequencies of the various values.

```

u <- as.data.frame(cSpecPct)
u
class(u)
[1] "data.frame"

```

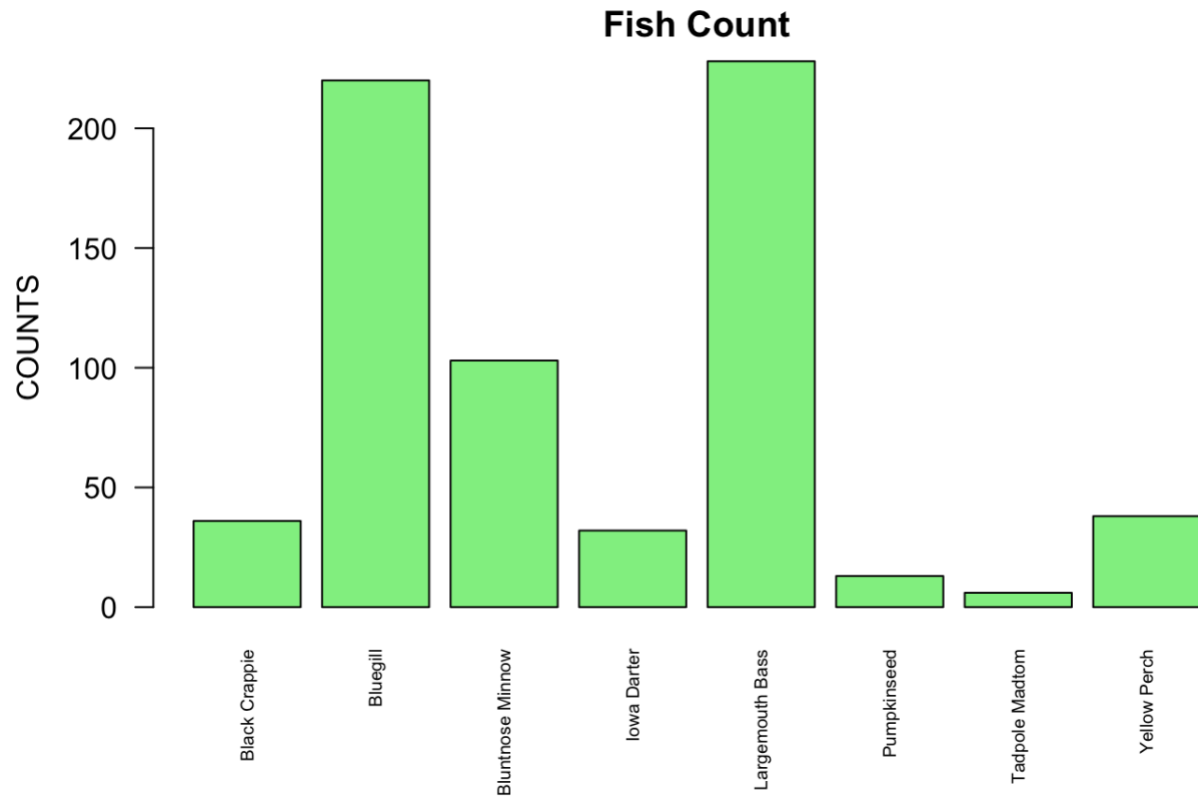
14. Create a barplot of with the following: titled Fish Count with the following specifications: • Title: Fish Count • Y axis is labeled “COUNTS” • Color the bars Light Green • Rotate Y axis to be horizontal • Set the X axis font magnification to 60% of nominal

Describe number 14: The Bar Plot above illustrate the fish count of each fish available. The species of the fish are labelled on the X-axis and the count (number) of the fish are labelled on the Y axis, respectively. The Y axis has a range of 0 to 200. This Barplot's title is "Fish Count." with Light Green colour which are shown and also, Font magnification has been maintained at 60% of the normal font size. Moreover, the following graph shows that there are more Largemouth bass than any other species of fish. Similarly, Bluegill has the second-highest fish count overall. The number of fish for both largemouth bass and bluegill is over 200. With a combined total of 676 fish and more than 200 fish each, the largemouth bass and bluegill take up around 66% of the whole lot. The average count of the Bluntnose Winnow stands between the species with the greatest and lowest lot. In comparison to largemouth bass and bluegill, the proportions of the species Black Crappie, Iowa darter, Pumpkinseed, Tadpole Madtom, and Yellow Perch are quite low. Compared to all other species, Tadpole Madtom has the fewest fish.

```

barplot(cSpec, main = "Fish Count",
ylab = "COUNTS",
col = "LightGreen",
las=2,
cex.names = 0.6,
cex.lab=1)

```

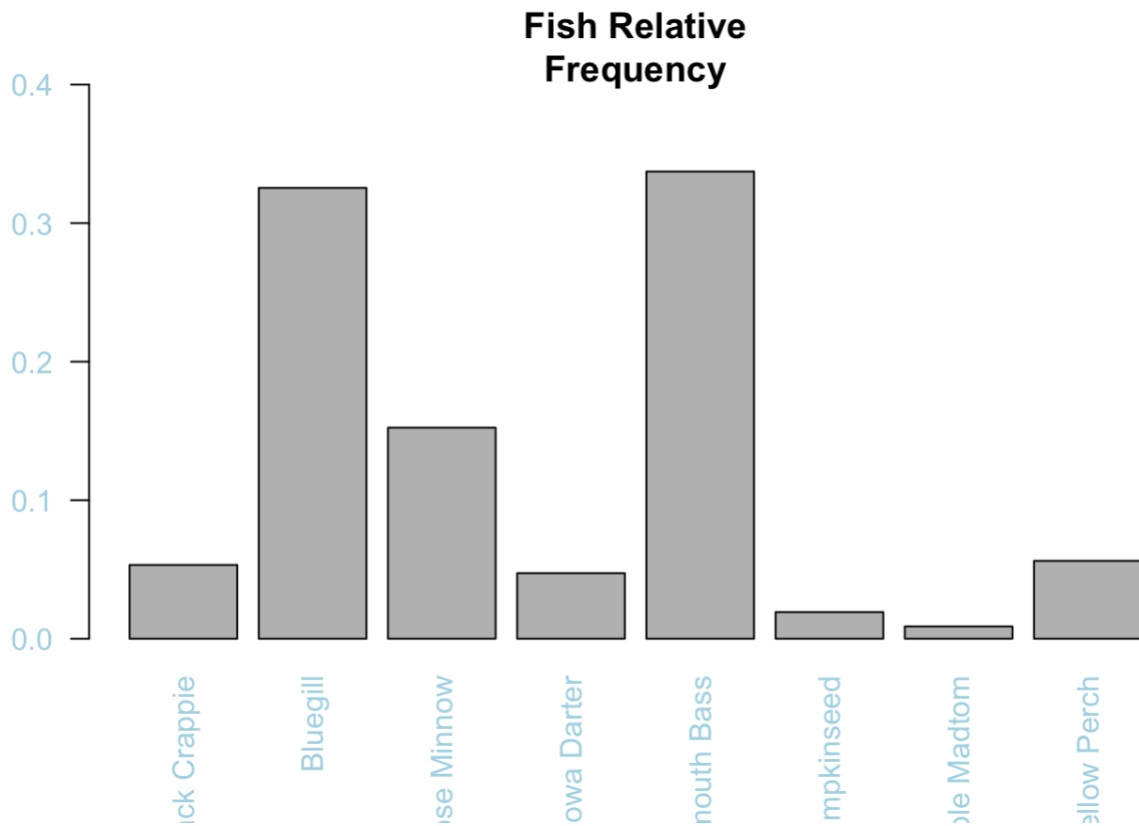


15. Create a barplot of, with the following specifications:

- Y axis limits of 0 to 4
- Y axis label color of Light Blue
- Title of "Fish Relative Frequency"

Describe number 15: The Bar Plot above shows the Fish Relative frequency of each fish available. This fish relative frequency is calculated by dividing each frequency of the fish by the total number of fish available as sample. As can be seen from the plot, The Y-axis depicts the 8 different species of fish available. The x-axis depicts the relative frequency scale which has a limit of 0 to 4 and as can be seen which it shows that this scale is from 0.0, 0.1, 0.2, 0.3 and 0.4. As we can observe Bluegill and Largemouth Bass are the most abundant fish found which are above 0.3. After these two fishes, Bluntnose Minnow came next with 0.15. Tadpole Madtom and Pumpkinseed have the lowest fish frequency among the lot which are slightly above 0.0. This bar plot is actually the same as the Fish Count plot with a difference in x-axis that shows the relative frequency scale and compare them in frequency scale.

```
barplot((cSpecPct/100),ylim =c(0,0.4), col.axis = "LightBlue", las=2, main = "Fish Relative Frequency")
```

16. Rearrange the cSpec Pct data frame in descending order of relative frequency. Save the rearranged data frame as the object .

Describe number 16: The following code can be used to arranging the relative frequency.

```
d <- u[with(u, order(-Freq)),]
d
```

17. Rename the columns Var1 to Species, and Freq to RelFreq .

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

18. Add new variables to and call them cumfreq, counts, and cumcounts.

```
d <- transform(d, cumfreq = cumsum(RelFreq))
d

dcount <- as.data.frame(t[order(-t$Freq),])
```

```
d <- transform(d, counts=RelFreq*676)
d

d <- transform(d, cumcounts=cumsum(counts))
d
```

19. Create a parameter variable to store parameter variables .

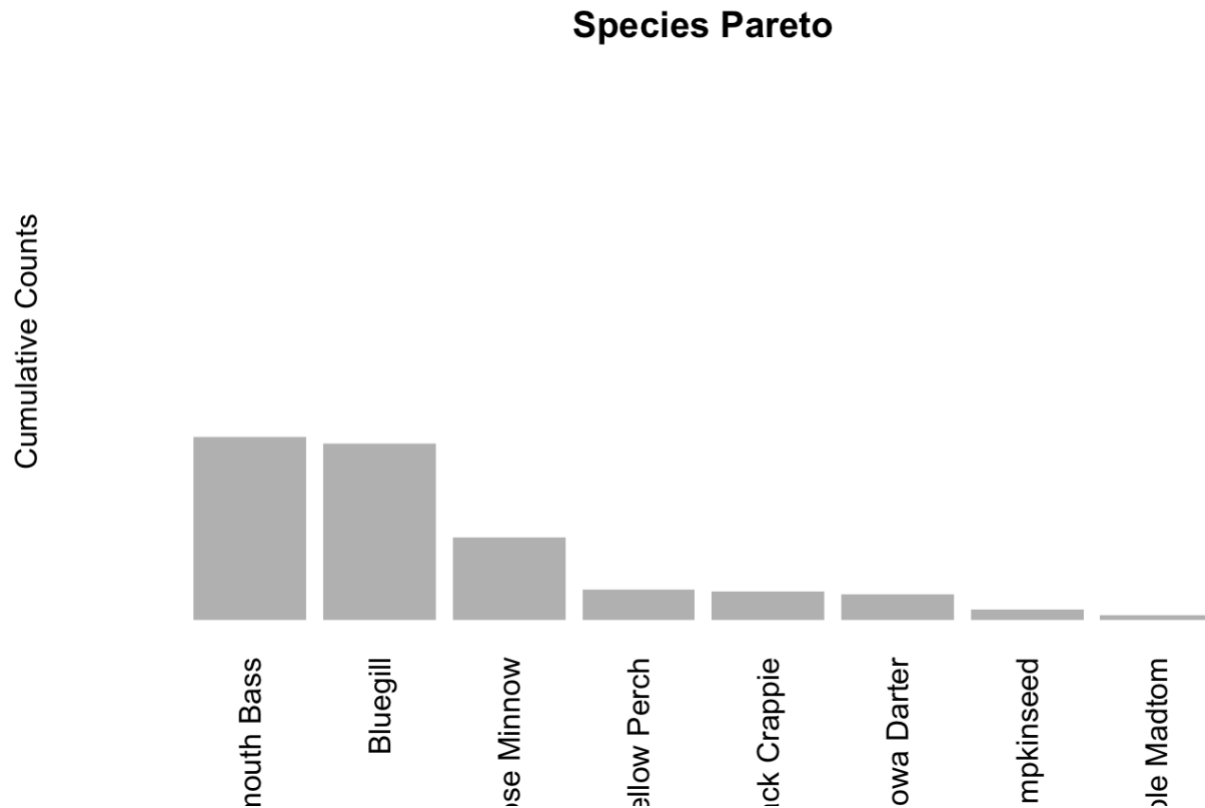
```
def_par <- par(mar = c(10,5,5,8))
def_par
$mar
[1] 5.1 4.1 2.1 2.1
```

20. Create a barplot, , 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)

Describe number 20: The bar chart below shows 8 different species fishes. As can be seen from the chart, the Y-axis shows Cumulative Counts and the X-axis shows these 8 different species fishes which as it is obvious Largemouth Bass and Bluegill experience the significant highest proportion while, Pumpkinseed and Tadpole Madtom has the lowest one.

```
pc <- barplot(d$count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d$count, na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 0.7, names.arg = d$Species, main = 'Species Pareto', las = 2)
```



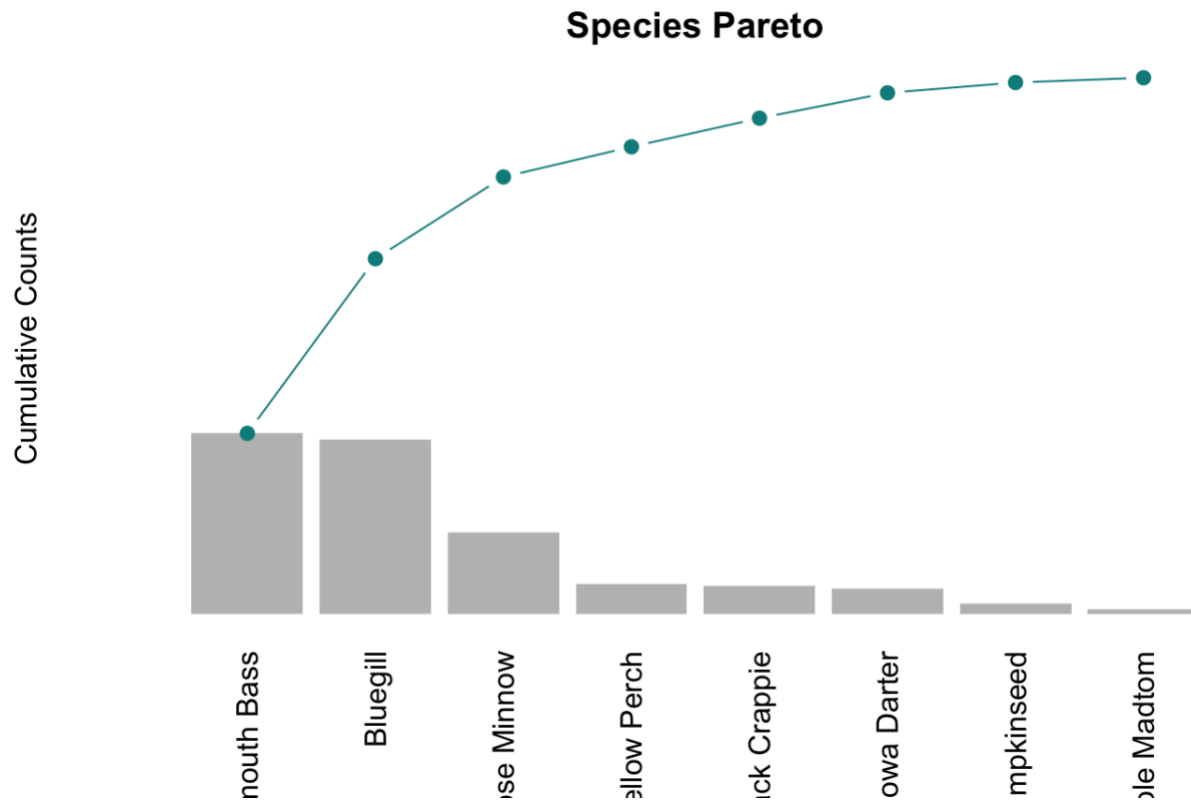
21. Add a cumulative counts line to the plot with the following:

- Spec line type is b
- Scale plotting text at 70%
- Data values are solid circles with color cyan4

Describe number 21: In this bar chart which is just the same as the bar plot in question 20 (plot “pc”) with one difference; we added a cumulative counts line which shows the sum of all the percentage values up to that category and as the Largemouth Bass to Bluntnose Minnow has the same amount, so this cumulative counts increased twice at first.

```
pc <- barplot(d$count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d$count, na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 0.7, names.arg = d$Species, main = 'Species Pareto', las = 2)

lines.default (pc, d$cumcounts, type = 'b', pch = 19, col = "cyan4")
```



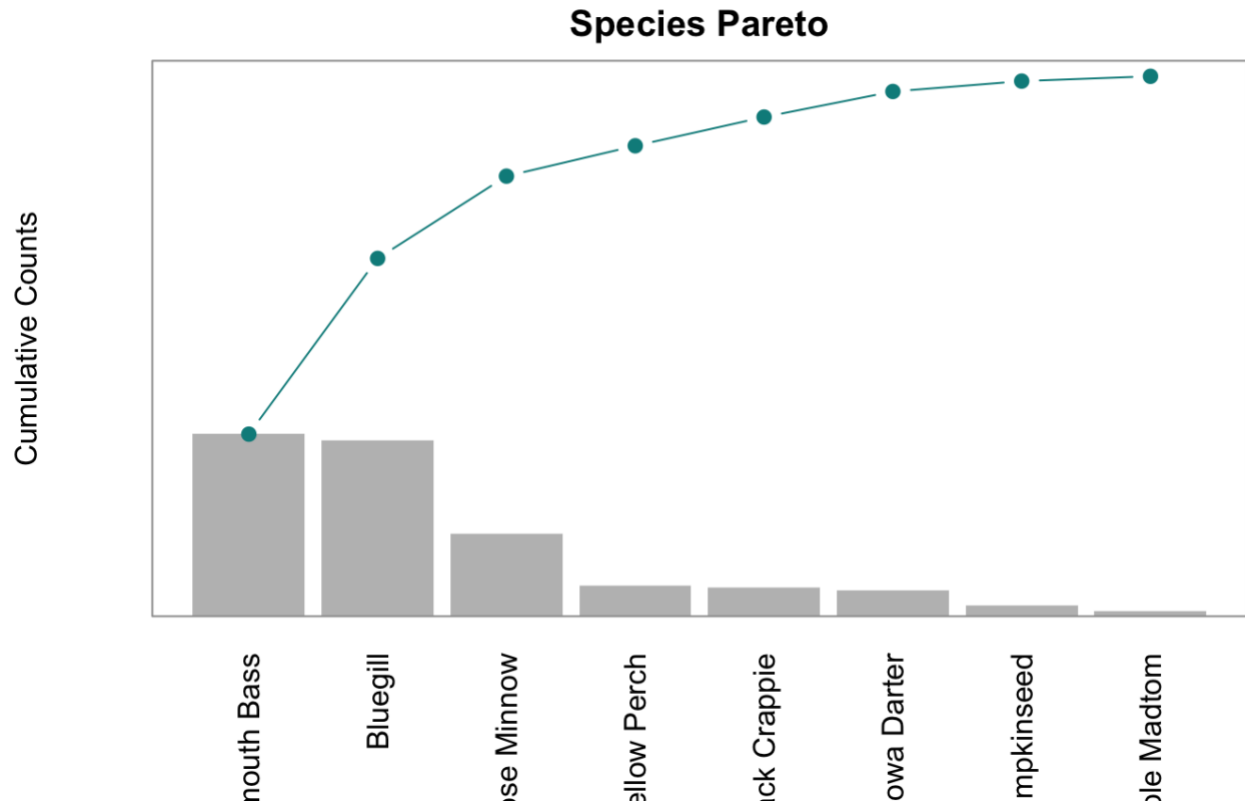
22. Place a grey box around the pareto plot .

Describe number 22: In this bar plot, just like the previous plot which has the same trend with only difference which we added a box into it.

```
pc <- barplot(d$count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.0
5 * max(d$count, na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 0.7, names.arg
= d$Species, main = 'Species Pareto', las = 2)

lines.default (pc, d$cumcounts, type = 'b', pch = 19, col = "cyan4")

box(col = 'grey62')
```



23. 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

Describe number 23: In this question, we continue the previous trend of “pc” bar chart, however we added Horizontal values in left side of the box. So now as can be observed from the chart, we can analyse it more accurately; Largemouth Bass which it is 22800 and Bluegill which is approximately 22800 have the highest amount in comare with other type of fishes. But, Tadpole Madtom and Pumkinseed have the lowest amount which are just above 0.

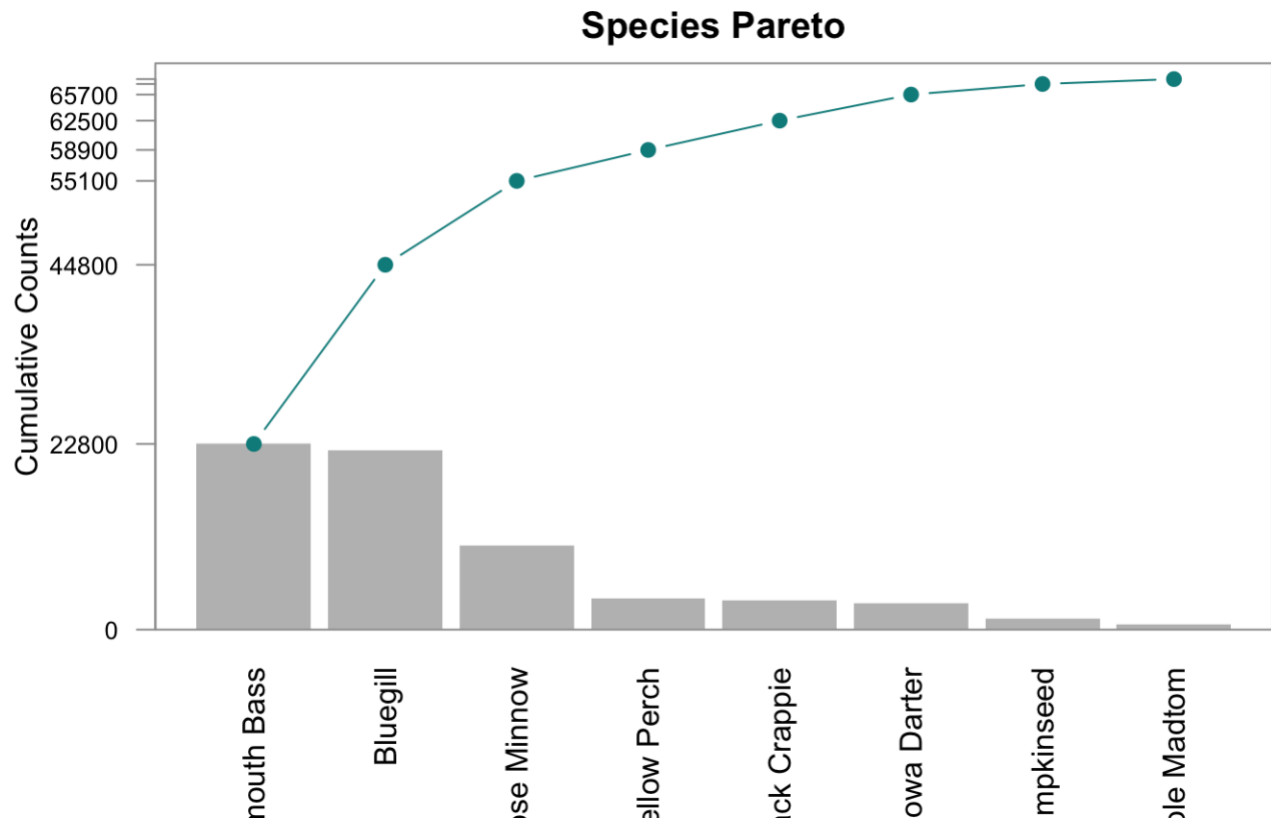
```
pc <- barplot(d$count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d$count, na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 0.7, names.arg = d$Species, main = 'Species Pareto', las = 2)

lines.default (pc, d$cumcounts, type = 'b', pch = 19, col = "cyan4")

box(col = 'grey62')

axis(side = 2, las = 2, at = c(0, d$cumcounts), tick = TRUE, line = NA, col.ticks = "grey62", col =

"grey62", cex.axis=0.8 )
```



24. 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

Describe number 24: An example of a Pareto analysis is shown in the above bar graph. The findings are shown in descending order on the species Pareto chart, with the biggest bar starting on the left and the smallest bar being farthest on the right. We learn about the number of fish on the Y-axis and the eight different species on the X-axis in this plot that we want to analyse. The cumulative frequency of the species is represented by the blue dots and the cumulative frequency in percentage is shown by the right axis. This demonstrates the low cumulative frequency and low population size of the species on the right. The first blue dot is displayed at 33–34% because the largemouth bass takes up around 33% of the entire lot. The number of fish keeps growing as we move to the right, and the blue dots will update to show you a new proportion of the overall fish population. By the time the last specie is complete, we will have practically all of the fishes present, making the total number of fish 676. Also, analysing them with the measures of cumulative counts depicts that Largemouth Bass and Bluegill has approximately the same amount which is 22800 and Tadpole Madton with almost above 0 has the less amount as a Cumulative Counts.

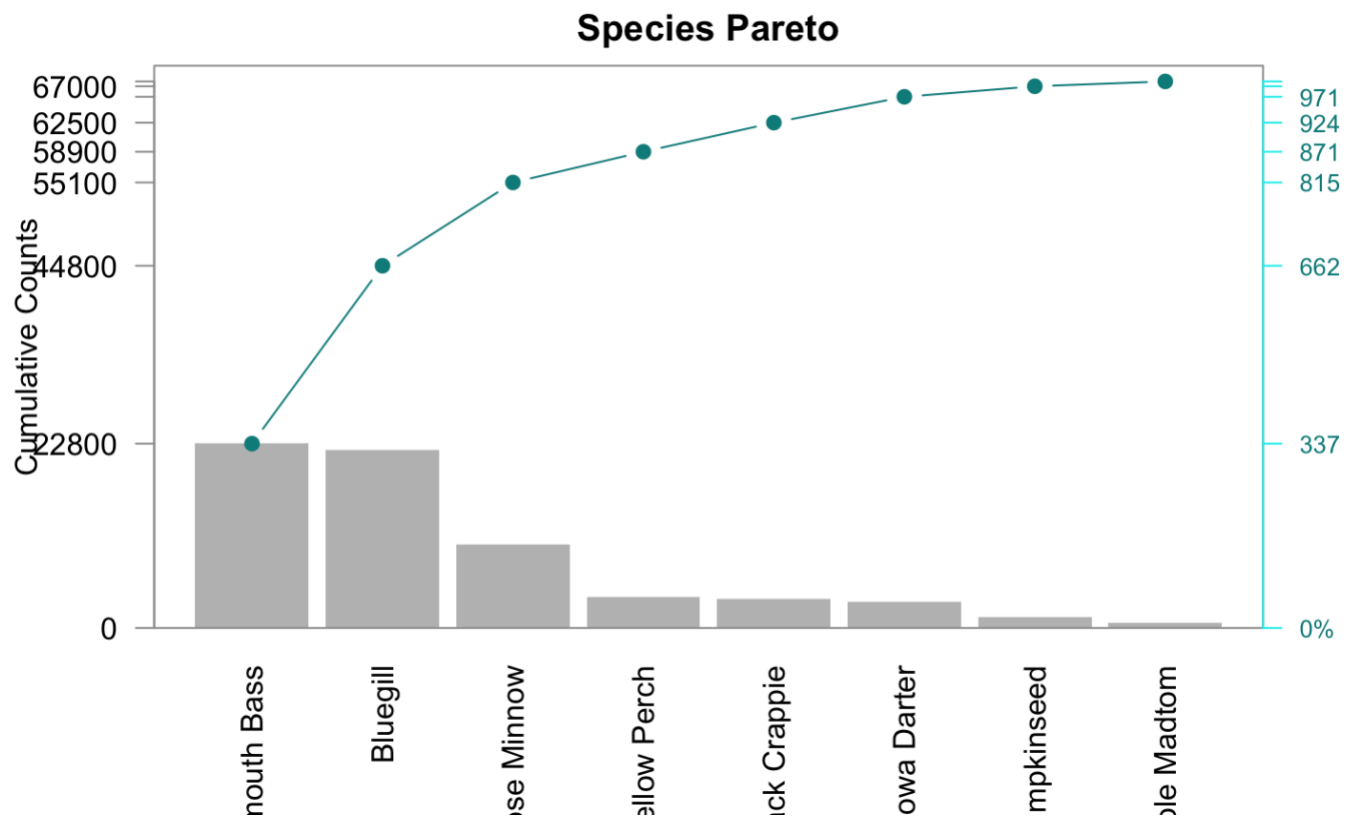
```
pc <- barplot(d$count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.0
5 * max(d$count, na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 0.7, names.arg
= d$Species, main = 'Species Pareto', las = 2)

lines.default(pc, d$cumcounts, type = 'b', pch = 19, col = "cyan4")
```

```
box(col = 'grey62')

axis(side = 2, las = 2, at = c(0, d$cumcounts), tick = TRUE, line = NA, col.ticks = "grey62", col = "grey62")

axis(side = 4, at = c(0, d$cumcounts), col = "cyan2", cex.axis = 0.8, las = 2, tick = TRUE, line = NA, col.axis = "dark cyan", labels = paste0(round( c(0,d$cumfreq) * 100,digits = 0),'%'))
```



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

```
pc <- barplot(d$count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d$count, na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 1.7, names.arg = d$Species, main = 'Sherafati', las = 2)

lines.default (pc, d$cumcounts, type = 'b', pch = 19, col = "cyan4")

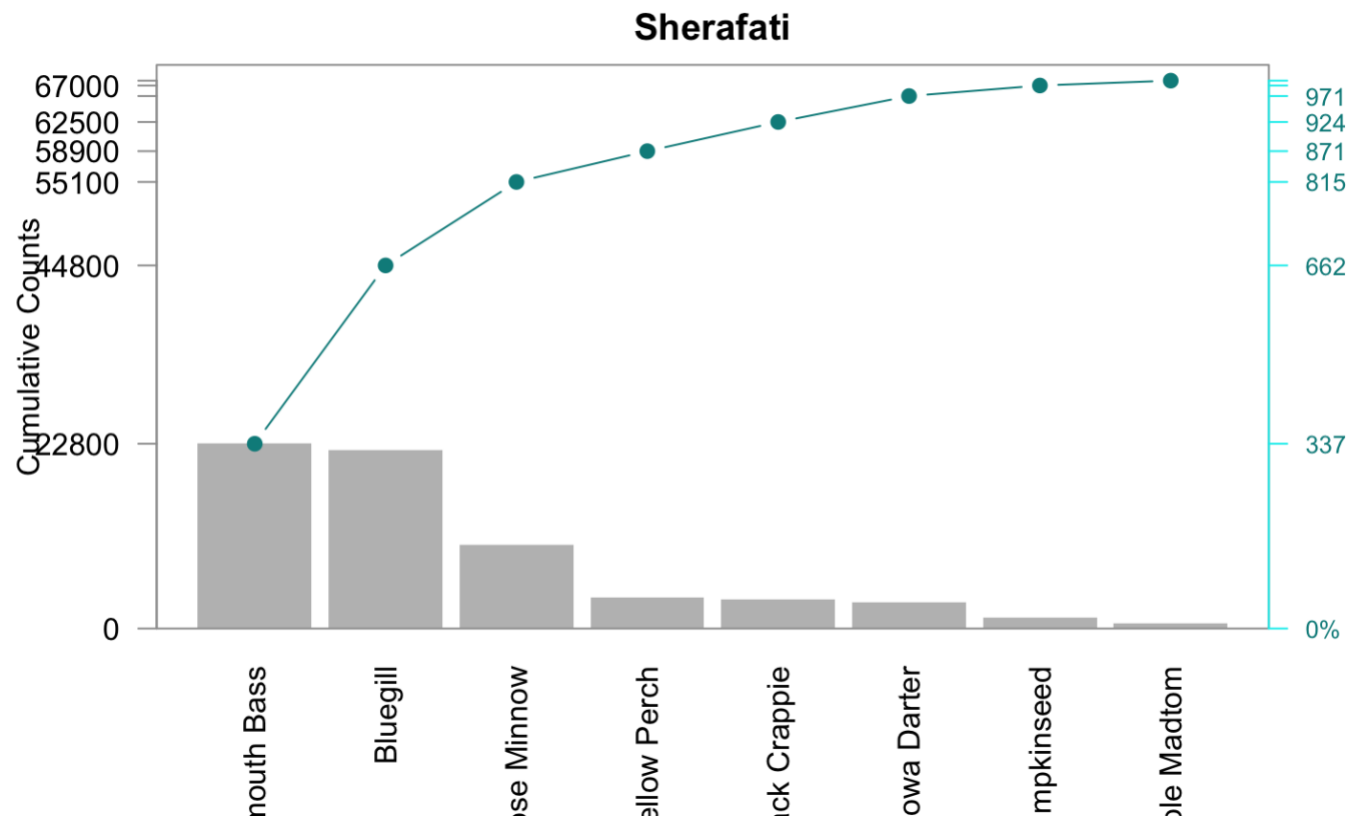
box(col = 'grey62')

axis(side = 2, las = 2, at = c(0, d$cumcounts), tick = TRUE, line = NA, col.ticks = "grey62", col = "grey62")

axis(side = 4, at = c(0, d$cumcounts), col = "cyan2", cex.axis = 0.8, las = 2, tick = TRUE, line = NA, col.axis = "dark cyan", labels = paste0(round( c(0,d$cumfreq) * 100,digits = 0),'%'))
```

```
axis(side = 4, at = c(0, d$cumcounts), col = "cyan2", cex.axis = 0.8, las = 2, tick = TRUE,

line = NA, col.axis = "dark cyan", labels = paste0(round( c(0,d$cumfreq) * 100,digits = 0), '%'))
```



SUMMARY

To sum up, We can see from the data file inchBio.csv that there are a total of 7 data variables with 676 data objects, 8 levels of distinct species on which we are doing our descriptive analysis, and a total of 7 data variables. Largemouth bass, with a count strength of 288, is the species with the highest count strength, while Tadpole Madtom, with a count strength of 6, has the lowest. The highest for tadpole has a cumulative count of 676, whereas the largest largemouth bass has a cumulative count of 288, and the smallest tadpole has a relative frequency of 0.00. However, Tadpole Madtom's cumulative frequency, which is 100, may be the greatest. Additionally, I learned how to calculate the descriptive statistics of the dataset and how to utilise the results to infer information and make assumptions and decisions about it. The Pareto and Barplot charts made the analysis super interesting and offered a lot of visual insights into the dataset.

BIBLIOGRAPHY

geeksforgeeks.com

Tutorialspoint.com

Youtube.com

Stackoverflow.com

R in Action by R. Kabacoff

APPENDIX

```
name <- ("Shamim Sherafati")
```

```
print(name)
```

```
install.packages("FSA")
```

```
install.packages("FSAdata")
```

```
install.packages("magrittr")
```

```
install.packages("dplyr")
```

```
install.packages("tidyr")
```

```
install.packages("plyr")
```

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

bio <- read.table (file=~ /Documents/ALY 6000/Module 3/inchBio.csv", header= TRUE, sep="," ,
stringsAsFactors = FALSE)

bio
head(bio)
tail(bio)
str(bio)
counts <- c(bio$species)
counts
Level8th <- unique(bio$species)
Level8th
tmp <- table (bio$species)
tmp
tmp2 <- subset(bio, select = species)
tmp2
head(tmp2, 5)
w <- table (bio$species)
w
class(w)
t <- as.data.frame(w)
t
class(t)
t$Freq
```

```
cSpec <- table(bio$species)
cSpec
class(cSpec)
cSpecPct <- table(bio$species)/676* 100
cSpecPct
class(cSpecPct)
u <- as.data.frame(cSpecPct)
u
class(u)
barplot(cSpec, main = "Fish Count", ylab = "COUNTS", col = "LightGreen", las=2, cex.names = 0.6,
cex.lab=1)
barplot((cSpecPct/100),ylim =c(0,0.4), col.axis = "LightBlue", las=2, main = "Fish Relative Frequency")
d <- u[with(u, order(-Freq)),]
d
names(d) <- c("Species", "RelFreq")
d
d <- transform(d, cumfreq = cumsum(RelFreq))
d
dcount <- as.data.frame(t[order(-t$Freq),])
d <- transform(d, counts=RelFreq*676)
d
d <- transform(d, cumcounts=cumsum(counts))
d
def_par <- par(mar = c(10,5,5,8))
def_par
pc <- barplot(d\count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d\count,
na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 1.7, names.arg = d$Species, main = 'Species
Pareto', las = 2)
pc <- barplot(d\count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d\count,
na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 0.7, names.arg = d\Species, main = 'Species
Pareto', las = 2) lines.default (pc, d\cumcounts, type = 'b', pch = 19, col = "cyan4")
```

```
pc <- barplot(d\count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d\count,
na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 1.7, names.arg = d\Species, main = 'Species
Pareto', las = 2) lines.default(pc, d\cumcounts, type = 'b', pch = 19, col = "cyan4")
```

```
box(col = 'grey62')
```

```
pc <- barplot(d\count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d\count,
na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 1.7, names.arg = d\Species, main = 'Species
Pareto', las = 2) lines.default(pc, d\cumcounts, type = 'b', pch = 19, col = "cyan4") box(col = 'grey62')
```

```
axis(side = 2, las = 2, at = c(0, d$cumcounts), tick = TRUE, line = NA, col.ticks = "grey62", col =
"grey62", cex.axis=0.8 )
```

```
pc <- barplot(d\count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d\count,
na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 1.7, names.arg = d\Species, main = 'Species
Pareto', las = 2) lines.default(pc, d\cumcounts, type = 'b', pch = 19, col = "cyan4") box(col = 'grey62')
axis(side = 2, las = 2, at = c(0, d$cumcounts), tick = TRUE, line = NA, col.ticks = "grey62", col =
"grey62")
```

```
axis(side = 4, at = c(0, d\cumcounts), col = "cyan2", cex.axis = 0.8, las = 2, tick = TRUE, line = NA,
col.axis = "dark cyan", labels = paste0(round( c(0,d\cumfreq) * 100,digits = 0),'%'))
```

```
pc <- barplot(d\count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.05 * max(d\count,
na.rm = TRUE)), ylab = 'Cumulative Counts', cex.axis = 1.7, names.arg = d\Species, main = 'Sherafati',
las = 2) lines.default(pc, d\cumcounts, type = 'b', pch = 19, col = "cyan4") box(col = 'grey62') axis(side
= 2, las = 2, at = c(0, d$cumcounts), tick = TRUE, line = NA, col.ticks = "grey62", col = "grey62")
```

```
axis(side = 4, at = c(0, d\cumcounts), col = "cyan2", cex.axis = 0.8, las = 2, tick = TRUE, line = NA,
col.axis = "dark cyan", labels = paste0(round( c(0,d\cumfreq) * 100,digits = 0),'%'))
```

```
.
```