

Name: Tanvi Jain

Date: 2/03/2022

Title: Module 3 Project – Executive Summary Report3

Report

The data set that is provided is a sampling of fish species. Bluegill and Largemouth bass are the two most common species in the sample. The dataset includes fish life spans, lengths, and weights, as well as a statistical table that summarizes the data. This project is built on a graphical depiction of Bluegill, Largemouth, and other fish traits.

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

```
print("Tanvi Jain")

install.packages("FSA")
install.packages("FSAdat")
install.packages("tidyr")
install.packages("plyr")
install.packages("tidyverse")
library("FSA")
library("FSAdat")

library("magrittr")
library("dplyr")
library("tidyr")
library("tidyverse")
library("plyr")
```

2. Import the inchBio.csv and name the table <bio>

```
> bio <- read.csv('/Users/tanvi/Downloads/inchBio.csv')
```

3. Display the head, tail and structure of <bio>

```
> head(bio)
```

	netID	fishID	species	tl	w	tag	scale
1	12	16	Bluegill	61	2.9		FALSE
2	12	23	Bluegill	66	4.5		FALSE
3	12	30	Bluegill	70	5.2		FALSE
4	12	44	Bluegill	38	0.5		FALSE
5	12	50	Bluegill	42	1.0		FALSE
6	12	65	Bluegill	54	2.1		FALSE

```
> 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: Factor w/ 8 levels "Black Crappie",...: 2 2 2 2 2 2 2 2 2 2 ...
 $ 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 : Factor w/ 193 levels "", "1014", "1015",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ scale : logi FALSE FALSE FALSE FALSE FALSE FALSE ...
```

```
> tail(bio)
```

	netID	fishID	species	tl	w	tag	scale
671	121	808	Black Crappie	323	509	1050	TRUE
672	121	809	Black Crappie	282	352	1700	TRUE
673	121	812	Black Crappie	142	37		TRUE
674	110	863	Black Crappie	307	415	1783	TRUE
675	129	870	Black Crappie	279	344	1789	TRUE
676	129	879	Black Crappie	302	397	1792	TRUE

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

```
> counts <- count(bio$species)
> View(counts)
```

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

```
> counts['x']
```

	x
1	Black Crappie
2	Bluegill
3	Bluntnose Minnow
4	Iowa Darter
5	Largemouth Bass
6	Pumpkinseed
7	Tadpole Madtom
8	Yellow Perch

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

```
> (tmp <- count(bio$species))
```

	x	freq
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

```
> View(tmp)
```

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

```
> tmp2<-head(tmp,5)
> View(tmp2)
> tmp2
```

	x	freq
1	Black Crappie	36
2	Bluegill	220
3	Bluntnose Minnow	103
4	Iowa Darter	32
5	Largemouth Bass	228

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

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

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

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

9. Convert <w> to a data frame named <t> and display the results

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

	Var1	Freq
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

10. Extract and display the frequency values from the <t> data frame

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

11. 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 <bio>

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

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

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

```
> (cSpecPct<- prop.table(table(bio$species)))
```

Black Crappie	Bluegill	Bluntnose Minnow	Iowa Darter	Largemouth Bass
0.05325444	0.32544379	0.15236686	0.04733728	0.33727811
Pumpkinseed	Tadpole Madtom	Yellow Perch		
0.01923077	0.00887574	0.05621302		

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

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

14. Create a barplot of <cSpec> with the following: titled Fish Count with the following specifications:

- Title: Fish Count

- Y axis is labeled “COUNTS”
- Y axis limits of 0 to 250
- Color the bars Light Blue
- Rotate X axis label to be vertical
- Set the X axis font magnification to 60% of nominal

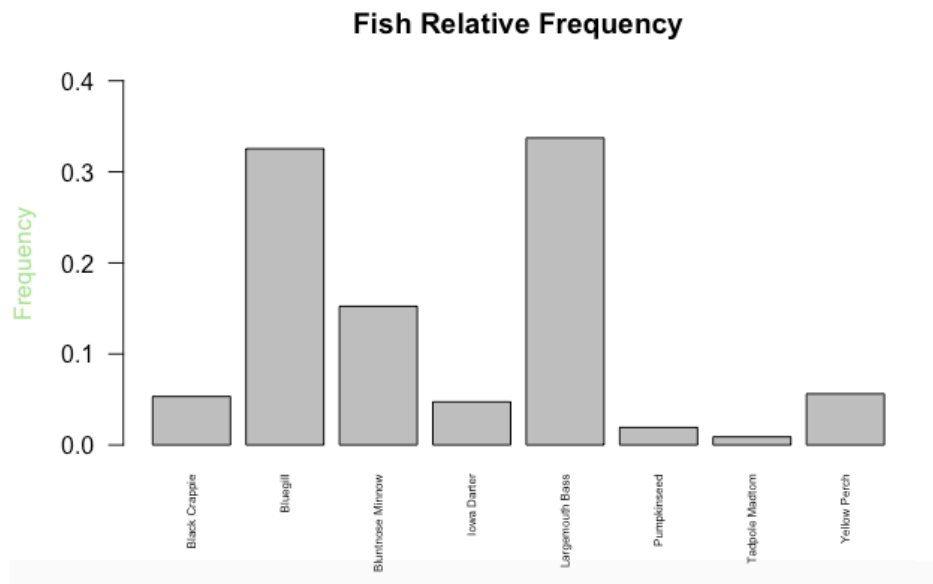
```
barplot(cSpec, main="Fish Count", ylab="Counts",
horiz=F,col="LightGreen",cex.names = 0.6, las=2)
```

```
> barplot(cSpec, main="Fish Count", ylab="Counts", horiz=F,col="LightBlue",cex.names = 0.6, las=2)
```

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

- Y axis limits of 0 to 0.4
- Color the bars Light Green
- Title of “Fish Relative Frequency”

```
barplot(cSpectPct, main="Fish Relative Frequency", ylim = c(0,0.4),
cex.names=0.5, las=2, ylab="Frequency", col.lab="LightGreen")
```



To better comprehend the determined relative frequency, I produced a bar figure. The graph depicts the species and their frequency.

Two species, "Bluegill" and "Largemouth Bass," have the greatest frequencies, while "Pumpkin seed" and "Tadpole Madlom" have the lowest frequencies of all the species listed.

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

```

      Var1      Freq
5 Largemouth Bass 0.33727811
2      Bluegill 0.32544379
3 Bluntnose Minnow 0.15236686
8      Yellow Perch 0.05621302
1      Black Crappie 0.05325444
4      Iowa Darter 0.04733728
6      Pumpkinseed 0.01923077
7      Tadpole Madtom 0.00887574

```

```

d <- u[order(-u$Freq),]
d

```


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

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

```
> (colnames(d) <- c("Species", "RelFreq"))  
[1] "Species" "RelFreq"
```

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

```
d$cumfreq <- cumsum(d$RelFreq)  
d$counts=sort(t$Freq,decreasing = T)  
d$cumcounts <- cumsum(d$counts)  
print(d)
```

```
> d$cumfreq <- cumsum(d$RelFreq)
```

```
> d$counts=sort(t$Freq,decreasing = T)  
> d$cumcounts <- cumsum(d$counts)  
> print(d)
```

	Species	RelFreq	cumfreq	counts	cumcounts
5	Largemouth Bass	0.33727811	0.3372781	228	228
2	Bluegill	0.32544379	0.6627219	220	448
3	Bluntnose Minnow	0.15236686	0.8150888	103	551
8	Yellow Perch	0.05621302	0.8713018	38	589
1	Black Crappie	0.05325444	0.9245562	36	625
4	Iowa Darter	0.04733728	0.9718935	32	657
6	Pumpkinseed	0.01923077	0.9911243	13	670
7	Tadpole Madtom	0.00887574	1.0000000	6	676

19. Create a parameter variable <def_par> to store parameter variables

```
def_par <- par()
def_par
```

```
> def_par <- par()
> def_par
$xlog
[1] FALSE

$ylog
[1] FALSE

$adj
[1] 0.5

$ann
[1] TRUE

$ask
[1] FALSE

$bg
[1] "white"

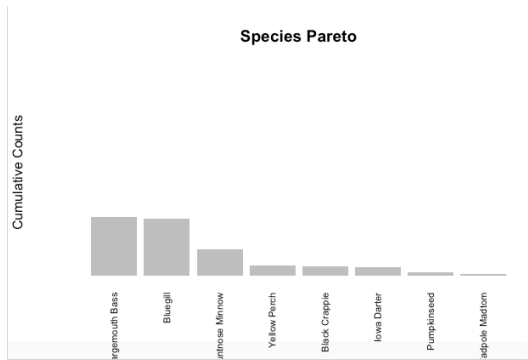
$bty
[1] "o"
```

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

- d\$counts of width 1, spacing of .15
- no boarder
- Axes: F
- Y axis 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

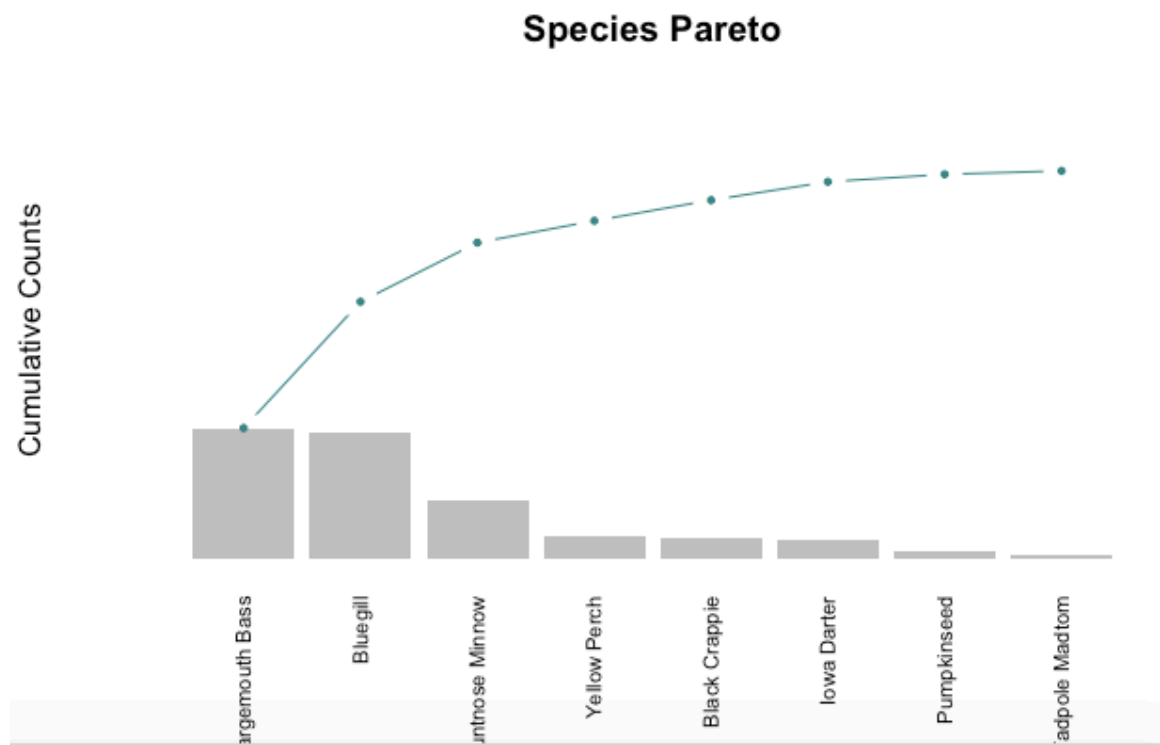
```
> pc <- barplot(d$count, width = 1, space = 0.15, border = NA, axes = F, ylim = c(0, 3.5 * max(d$count, na.rm = T)), ylab = "Cumulative Counts", cex.name = 0.7, cex.axis = 0.7, names.arg = d$Species, main = "Species Pareto", las = 2)
```



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

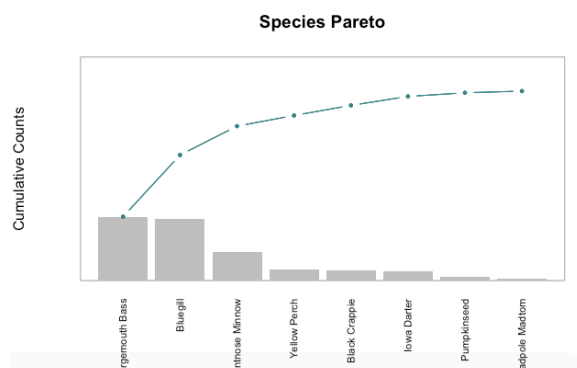
- Spec line type is b
- Scale plotting text at 70%
- Data values are solid circles with color cyan4
 - The below given is a Pareto plot that represents that environment factors play a major role in the growth of species. Some fish species are predators which feed upon smaller fishes or insects which led to the lesser growth of other fish species. This is represented by the bar plot.

```
> lines(pc, d$cumcounts, type="b", cex=0.7, pch=20, col="cyan4")
```



- 22. Place a grey box around the pareto plot (hint: <https://www.statmethods.net/advgraphs/parameters.html>)

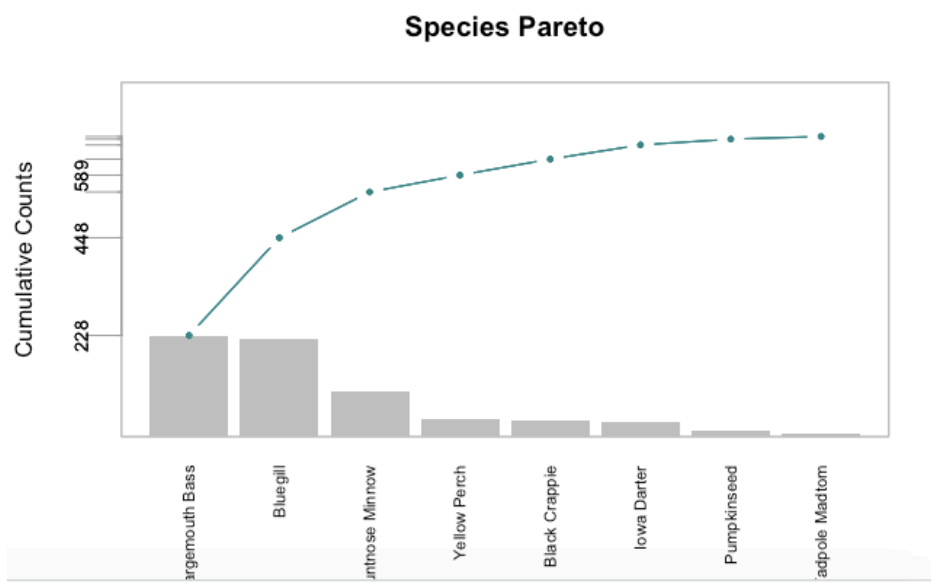
```
> lines(pc, d$cumcounts, type="b", cex=0.7, pch=20, col="cyan4", box(col="grey", pch=19))
```



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 (hint:
<https://www.statmethods.net/advgraphs/axes.html>)

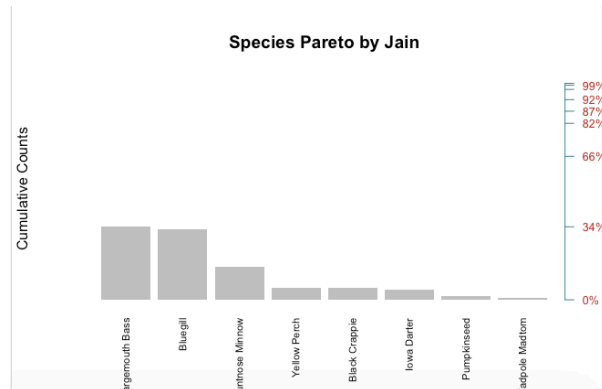
```
> axis(at = d$cumcounts, side = 2, col = "grey62", cex.axis = 0.8, tck = -0.1 )
```



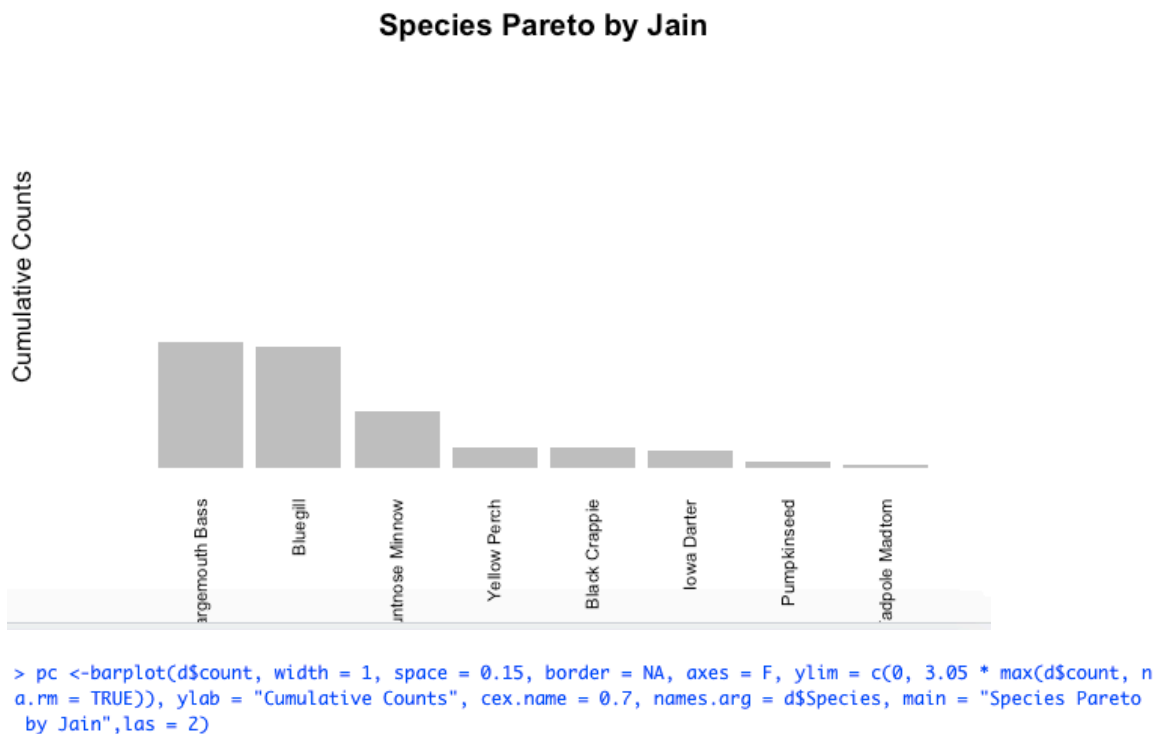
24. Add axis details on right side of box with the specifications:

- 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



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



Conclusion:

Fish species data set consists of various fish species with their relative

frequency and counts. Fishes like Bluegill, Largemouth are the predators which fed upon small insects and fishes which reduce the counts of other species. Also, there are environment factors due to which the relative frequency of the other fishes differs a lot.

All the information is being gathered and presented by the tabular representation and the graphical representation.

Bibliography:

1. R In Action by Robert I Kabacof
2. Datacamp – Learn R Programming:
https://www.youtube.com/watch?v=SWxoJqTqo08&list=PLjgj6kdf_snYBkIsWQYcYtUZiDpam7ygg

Appendix

#1 Printing my name `print("Tanvi Jain ")`

#2 Installing the packages


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

#3 importing a dataset (bio <-
read.csv("/Users/tanvi/Downloads/inchBio.csv"))

#4 Head, Tail and Structure

head(bio)

tail(bio)

str(bio)

#5 Object Count

counts <- count(bio$species)

#6 Displays the 8 levels

(displays <- counts %>% select(x))
```

#7 Display the Different Species

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

#8 Create a Table W

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

#9 Convert Dataframe to Table

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

#10 To extract and display the frequency values

```
t$Freq
```

#11 Create a table named

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

#12

```
<cSpecPct> (cSpectPct<- prop.table(table(bio$species)))
```

#13 <cSpecPct> to data frame (u <-data.frame(cSpectPct)) class(u)

#14 Create a barplot of <cSpec> font size

```
remaining barplot(cSpec, main="Fish Count", ylab="Counts",  
horiz=F,col="LightBlue",cex.names = 0.6, las=2)
```

#15 Create a barplot of <cSpecPct> barplot(cSpectPct, main="Fish
Relative Frequency", ylim = c(0,0.4), cex.names=0.5, las=2,
ylab="Frequency", col.lab="LightBlue")

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

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

```
d
```

#17 Rename the <d> columns Var 1 to Species, and Freq to RelFreq18.

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

#18 Add new variables to <d>

```
d$cumfreq <- cumsum(d$RelFreq)
d$counts=sort(t$Freq,decreasing = T) d$cumcounts <-
cumsum(d$counts)
```

```
print(d)
```

#19 Create a parameter variable

```
def_par <- par() def_par
```

#20 Create a barplot,

```
<pc> pc <-barplot(d$count, width = 1, space = 0.15, border = NA,
axes = F, ylim = c(0,3.5*max(d$counts,na.rm = T)), ylab =
"Cumulative Counts", cex.name = 0.5, cex.axis=0.7, names.arg=
d$Species, main = "Species Pareto",las = 2)
```

#21 Add a cumulative counts line to the <pc> plot lines(pc, d\$cumcounts, type="b",cex=0.7,pch=20,col="cyan4")

#22 Place a grey box around the pareto plot lines(pc, d\$cumcounts, type="b",cex=0.7,pch=20,col="cyan4", box(col="grey", pch=19))

#23 Add a left side axis axis(at = d\$cumcounts, side = 2, col = "grey62", cex.axis = 0.8,tck = -.01)

#24 Add axis details on right side of box axis(side = 4, at = c(0, d\$cumcounts), col = "cyan4", cex.axis = 0.8, las = 2, tick = TRUE,

```
line = NA, col.axis = "#CD0000", 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.name = 0.7, names.arg = d$Species,  
main = "Species Pareto by Jain", las = 2)
```