Executive Summary Report 3

Aditi Rajmane

ALY6000: Introduction to Data Analytics

Prof. Tom Breur

October 13, 2022

Head of the data

	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

Tail of the data

	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

Structure of the data

```
> 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 ...
```

Summary of the data

```
> summary(bio)
     netID            fishID          species               tl       
 Min.   :  1.00   Min.   :  7.0   Length:676         Min.   : 27.0  
 1st Qu.: 13.00   1st Qu.:175.8   Class :character   1st Qu.: 66.0  
 Median : 37.00   Median :345.5   Mode  :character   Median :189.5  
 Mean   : 67.65   Mean   :434.2                      Mean   :186.5  
 3rd Qu.:109.00   3rd Qu.:695.5                      3rd Qu.:295.0  
 Max.   :206.00   Max.   :915.0                      Max.   :429.0  

       w              tag              scale          
 Min.   :   0.2   Length:676         Mode :logical  
 1st Qu.:   2.0   Class :character   FALSE:213      
 Median :  54.5   Mode  :character   TRUE :463      
 Mean   : 126.8                                     
 3rd Qu.: 190.5                                     
 Max.   :1070.0                                     
 NA's   :165                                        
```

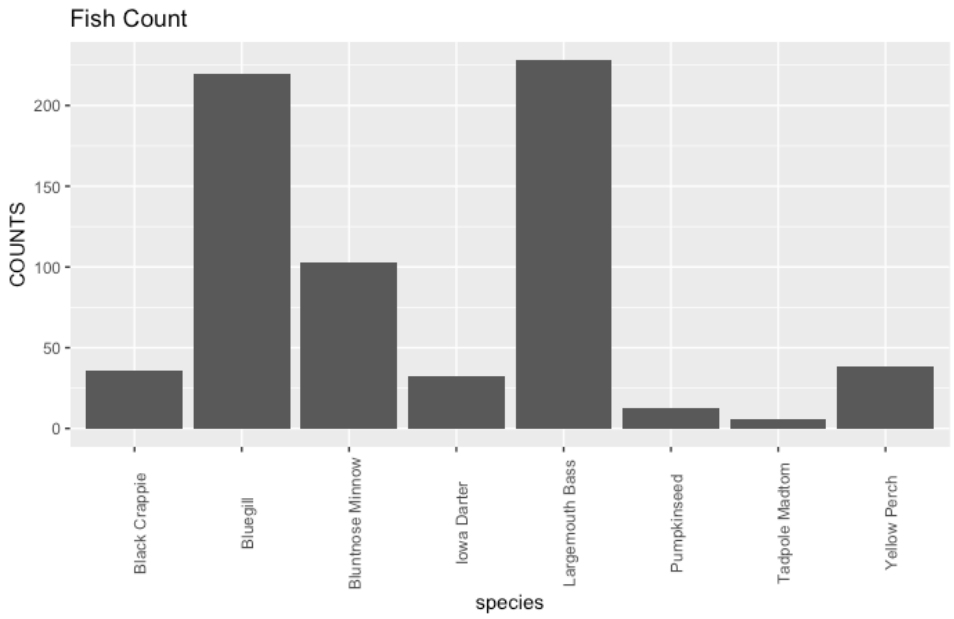
```
> describe(bio$tl)
   vars   n  mean     sd median trimmed    mad min max range skew
X1    1 676 186.5 109.59  189.5  183.53 171.24  27 429   402  0.1
   kurtosis   se
X1    -1.34 4.22
```

```
> describe(bio$w, na.rm = TRUE)
   vars   n  mean     sd median trimmed   mad min  max  range skew
X1    1 511 126.8 167.16   54.5   94.47 79.32 0.2 1070 1069.8 1.69
   kurtosis   se
X1      3.2 7.39
```

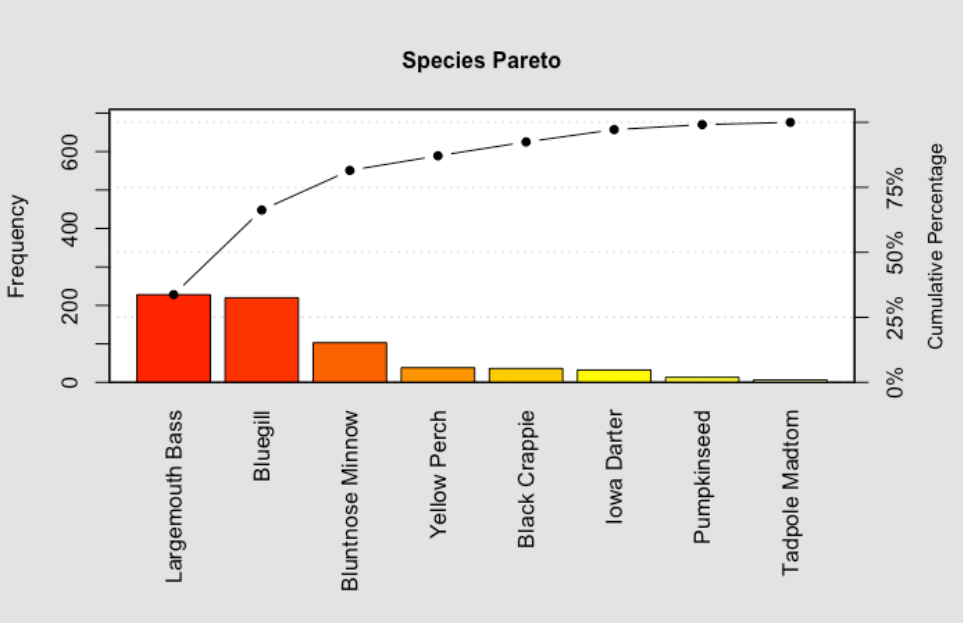
```
> count_species
           species   n
1  Largemouth Bass 228
2         Bluegill 220
3 Bluntnose Minnow 103
4     Yellow Perch  38
5    Black Crappie  36
6      Iowa Darter  32
7      Pumpkinseed  13
8   Tadpole Madtom   6
```

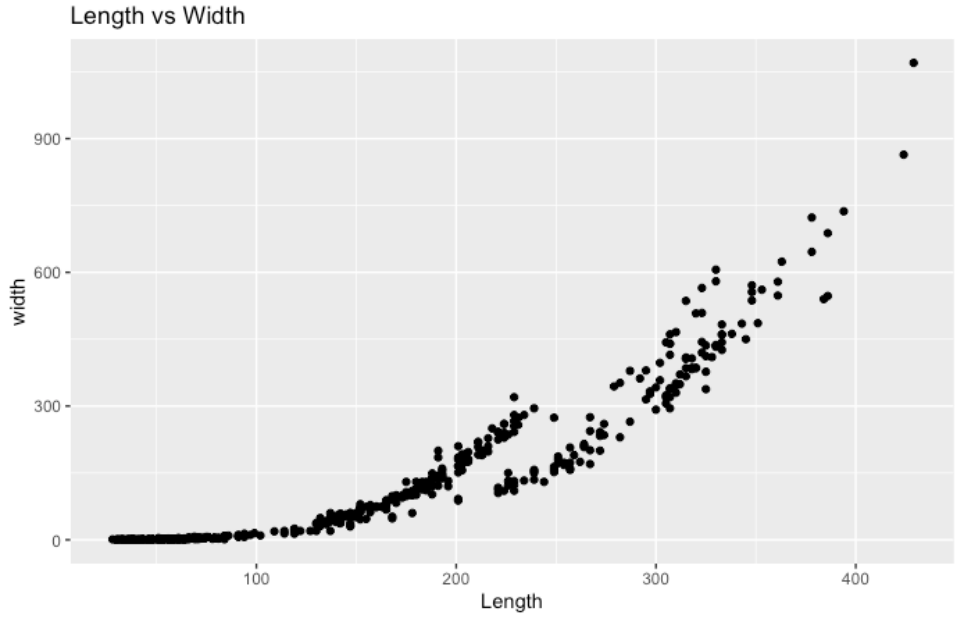
Proportions of each type of species

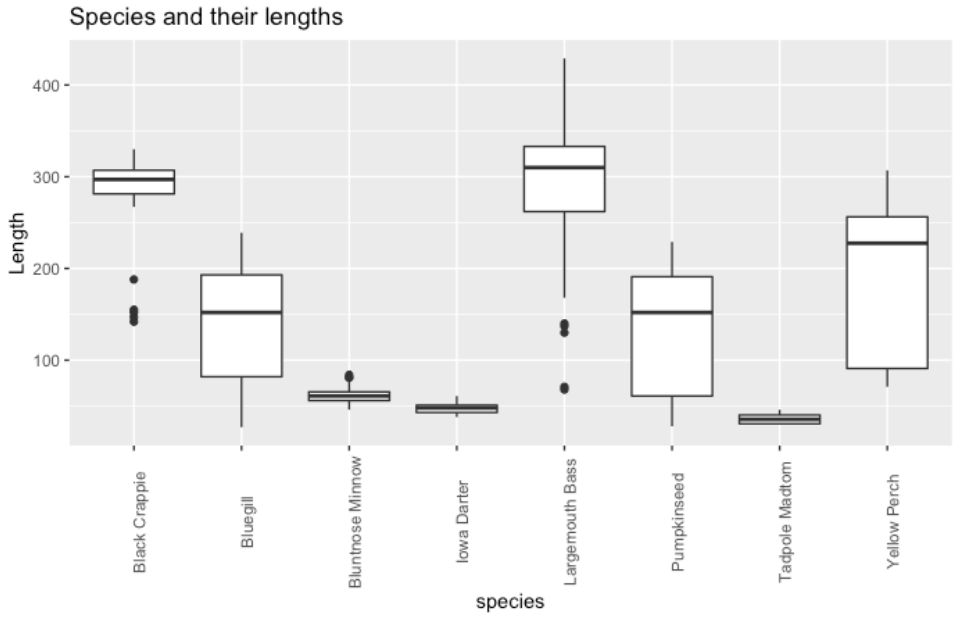
Black Crappie	Bluegill	Bluntnose Minnow	Iowa Darter
5.33	32.54	15.24	4.73
Largemouth Bass	Pumpkinseed	Tadpole Madtom	Yellow Perch
33.73	1.92	0.89	5.62

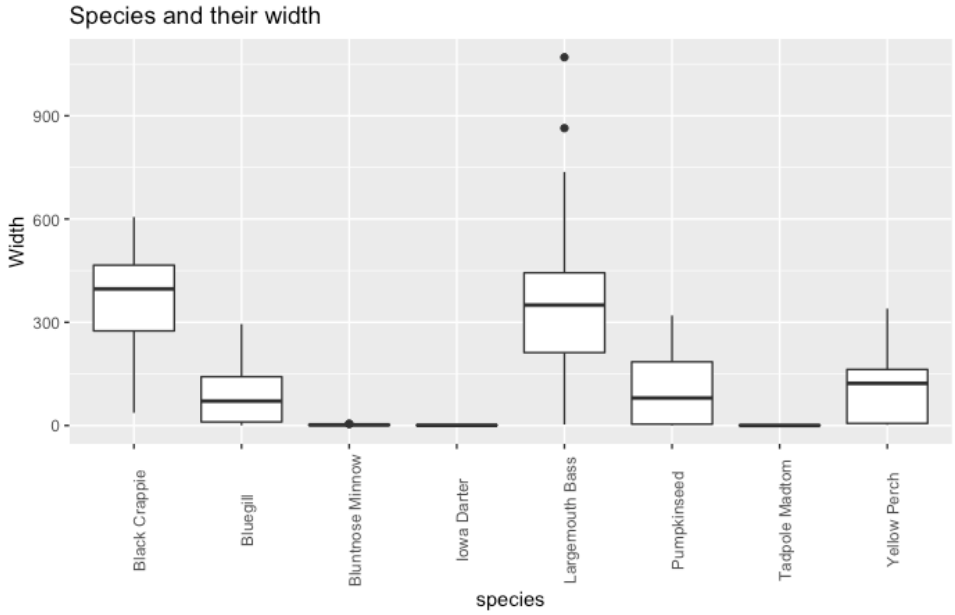


	species	n	cumulative_frequency
1	Largemouth Bass	228	228
2	Bluegill	220	448
3	Bluntnose Minnow	103	551
4	Yellow Perch	38	589
5	Black Crappie	36	625
6	Iowa Darter	32	657
7	Pumpkinseed	13	670
8	Tadpole Madtom	6	676


Species Pareto
Frequency
0
200
400
600
0%
25%
50%
75%
Cumulative Percentage
Largemouth Bass
Bluegill
Bluntnose Minnow
Yellow Perch
Black Crappie
Iowa Darter
Pumpkinseed
Tadpole Madtom


Length vs Width
width
0
300
600
900
100
200
300
400
Length






```
> #number of fishes scaled
> sum(bio$scale)
[1] 463
```

```
> #number of unique netIDs
> length(unique(bio$netID))
[1] 58
```

Summary

There are eight different types of species mentioned in the data. There are 228 largemouth bass, which contribute to the highest number of species, about 33.73%. The next largest category is bluegill, with 220 entries, followed by Bluntnose Minnow (103) and Yellow Perch (38). The Tadpole Madtom has the least number of entries, merely 6. About 66% of the fish belong to the largemouth bass and bluegill species together. As the length of the fish increases, the width of the fish also increases. Yellow perch fish have the longest lengths, and Iowa Darter fish are the shortest. Black crappie fish are wide as compared to all other species. There are about 463 fish that were weighed, and 58 fish with unique netIds.

Reference

Author	Geeksforgeeks
Date	30 May, 2021
Title	**Count Unique Values in R**
Source	https://www.geeksforgeeks.org/count-unique-values-in-r/

Author	Zach
Date	17 February, 2021
Title	**How to Create a Pareto Chart in R (Step-by-Step)**
Source	https://www.statology.org/pareto-chart-in-r/

Author	Stikpet
Date	5 June, 2022
Title	R - Pareto Chart
Source	https://www.youtube.com/watch?v=j0HUC4rhE9Q

Author	Datacamp
Date	NA
Title	**Descriptive Statistics**
Source	https://www.statmethods.net/stats/descriptives.html