ALY6000: Executive Summary Report 3

Overview and Rationale

As a data analyst, you will need to be able to create understandable data visualizations and clearly summarize the data for executives. This assignment is intended allow you to demonstrate your skills to process data, present the data visually, and calculate basic; all with explanatory analysis.

Module Objectives

This assignment links directly to the following learning outcomes from this module:

- Calculate counts and probabilities based on categorical data
- Graph probability distributions
- Use R to manipulate datasets
- Interpret data displayed in graphs

Assignment Summary

You will write and execute an R script in order to gather the information required to complete your report. You are provided with an assignment instruction document (Module 3 Project Instructions) to guide the R-code you need to write an Executive Summary specified below.

Your Executive Summary should consist of:

Page 0: **Title Page**: name, date, title

Pages 1-3: A short introduction, methodology, key findings about the data (Content outlined the assignment below and in the Project Instructions) and your concise summary.

Page 4: **Bibliography**: This includes Youtube videos, instruction materials, google search results, texts that informed your study of statistics and R. Adhere to APA standards.

Page 5: **Appendix**: The R Script you wrote and executed.

There are three aspects of this assignment as follows:

A. Following an introduction, 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 snippets to support your observations and conclusions. Below is a sample excerpt.

Example: *A structural analysis of the inchBio data set revealed that two species dominated this study. The Bluegill and Largemouth Bass accounted for 66% of the research data.*

```
> cSpecPct

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

As a result, several Bluegill subsets were created in order to facilitate an in-depth study of one of the fish species that dominate the dataset. Below we highlight the head and tail of the Bluegill subset created to analyze the characteristics of length and weight of Bluegill with weight above 60 and with length greater than 54. Note there were 112 records in this subset.

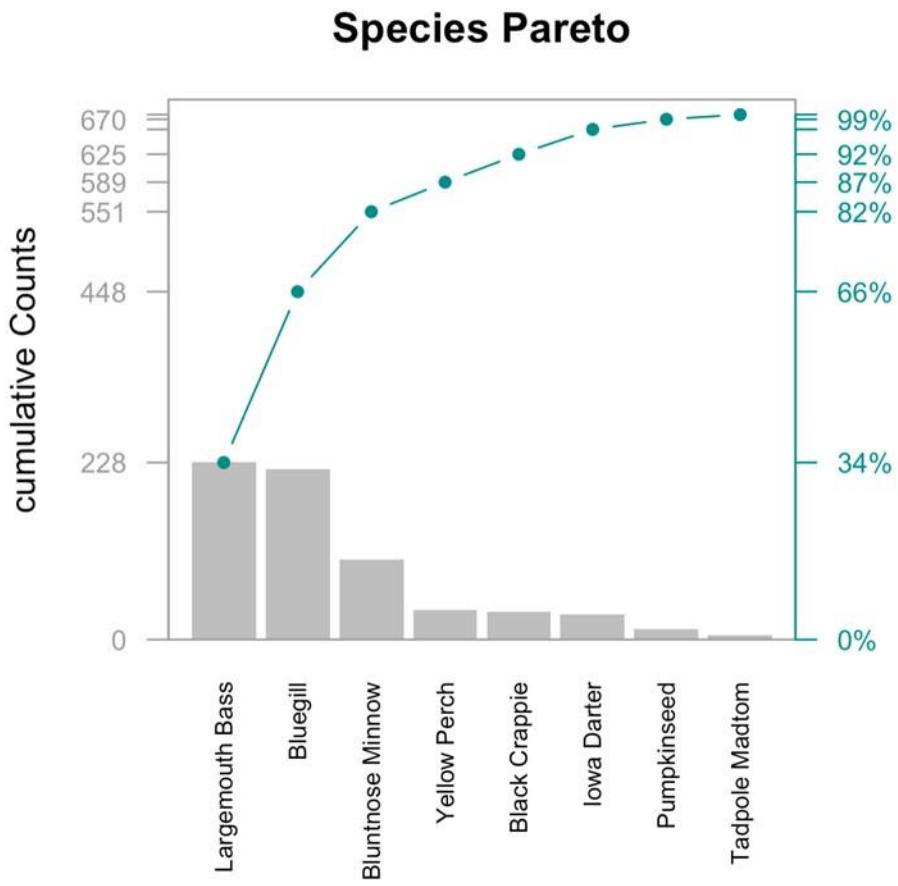
```
> headtail(tmpbgL2)
    fishID  species  tl   w
1      119 Bluegill 193 145
2      120 Bluegill 185 123
3      121 Bluegill 152  67
110    659 Bluegill 229 255
111    660 Bluegill 185 111
112    774 Bluegill 203 171
```

B. 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 barplots (bar charts). A pareto plot as illustrated below **must be** included in this part of your report. Include screen snippets 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**.**

Example: *The Pareto chart below combines the benefits of a bar plot with a cumulative distribution graph. We see how two species dominate the samples that were analyzed for this study. The pareto chart leads us to consider several new investigative avenues.*

1. *Are Bluegill and Largemouth Bass predatory species who prey on the smaller pumpkin seed or tadpole species?*
2. *Are there environment factors that are inhibiting some species from thriving in a common environment with large species.*

3. *A more detailed analysis of species length and weight is warranted to determine which species are underperforming in their development and thus present a reproductive risk or risk of increased morbidity and mortality*



C. 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.

Instructions

Using the assignment instruction document (Module 3 Project Instructions) and the dataset provided by your instructor (inchBio.csv), complete the assignment above

Supporting Files

To complete this assignment you will need to download the following files:

Module 3 Project Instructions.pdf inchBio.csv

What to Submit

Submit a single file with the following title: <LastName>_M3_Project3.pdf