**Module 3 PROJECT (DUE End of WEEK4)**

Sunday October 11, at 11:59 pm

**Executive Report**

*Research into fish species found in inchBio.csv dataset:*

1.Submit your Executive Summary with file name LastName-M3-Project3.

2. Your report is based on analyzing the inchBio.csv dataset. You will create and execute an R script in order to gather the information required to complete your report. A script template (see: ***M3 Project R-Script Guidance***) of comments will be provided to guide you in your development of the R-code you will need to create in order to write the Executive Summary specified below.

3. **IMPORTANT**! Your **Executive Summary** will be presented using a DOC file and it should consist of:

Page 1: **Title Page**: name, date, title, class, instructor, etc.

Pages 2-5. **Key Findings** about fish species. This section should be organized on two main sections:   
**Introduction**: where you will add meaningful information as indicated by your instructors in the first three weeks: Topic of the data (fisheries), overall description of the data set, questions you want to answer from the data set, statistical and graphical tools to be utilized, etc.   
**Analysis:** where you provide your results in a well‑organized and professional order, including deep application of critical thinking to make meaningful observations of the results you obtained.   
**Conclusions:** provide a clear three, four or more sentences summary of the key points that you want the audience to walk away with regarding your global analysis. This summary should present accurate analysis and be supported by the data presented in the rest of the report.

Page 6: This page contains both, the bibliography and the appendix.   
**Bibliography**. Include any resource you used to support your work: academic references, youtube videos, instruction materials, google search results, texts that informed your study of statistics and R, etc. Adhere to APA standards.  
**Appendix:** For the appendix write: An additional file containing the R codes has been attached to this report. The name of this file is M3\_project3\_mylastname.R

4. **GUIDANCE** for the EXECUTIVE SUMMARY:

**Analysis section**

**PART A: Descriptive Statistics**

1. Provide an analysis of descriptive characteristics of the various fish species (e.g., mean, median, quartiles, variance, standard deviation, skew, kurtosis, outliers etc.).
2. Notice that this question is not included in the tasks included in the R file.
3. Include R codes in an organized manner, do not add image screen shots, add the codes (text).
4. Notice that you have three variables that can be used for descriptive statistics: Species is categorical, you can count each observation and their probabilities. Total length (tl) is numerical. Weight (w) is numerical (notice that w has some missing values, or NA, you need to remove them from your calculations, try using code: na.rm).
5. Always round the number of decimals to either 2, 3, or 4. Use code: round (x, digits =)
6. Make meaningful 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. As a result, Bluegill and Largemouth Bass subsets were created in order to facilitate an in-depth study of the fish species that dominate the dataset.*

1. Here you can mention how many records were for each species, what was the mean of each species and their differences, median, etc. You can also prepare graphs to show your results (bar plots to show the mean differences, box plots and histograms to compare data distributions).

**PART B: PLOT ANALYSIS**

1. Provide an analysis using at least 3 visualization plots. Here you can present the graphs you created using the R codes you produced using file **M3-Project3\_Guide(Students).R**, or you can create additional graphs. Use boxplots, histograms, or bar plots, which we have seen in class. Use other graphs only if you feel comfortable.
2. R codes # 18-21 produce a bar plot resembling a pareto chart. Instead of showing cumulative percentages, it shows cumulative counts. An image example is shown below, your graph should look more or less like that. What conclusions can you make from the data displayed in that graph? Remember, apply deep critical thinking on your reports.

