This is an academic project for 'Intro to Data Science' course of Florida Atlantic University

Project #1—Exploring Palmer Penguins Data (100 pts)

Overview:
For this project, you will utilize the R programming language[1] and the *tidyverse*[2] package with *ggplot2*[3] for data visualization, to perform data summarization, preparation, and visualization on the Palmer Penguins[4] dataset. This dataset contains various measurements and attributes of different penguin species. Note that Project #1 provides more specific steps with points versus the following projects in this course to get you situated and comfortable with R and RStudio, as well as the process of thinking about data analytics. Regardless, all assignment code should run (if not points will be deducted)! Moreover, be creative and provide as much detail as you feel is needed. In the real world, data science projects tend to involve quite of bit of interpretation and explanation—which is also expected in each of your course assignments.

1. Create a R Notebook and install, at least, the following packages: *tidyverse, ggplot2, skimr*
2. Data Loading and Exploration (15 pts):
    a. Load the packages and load the Palmer Penguins dataset.
    b. Display the first few rows of the dataset to examine its structure.
    c. Provide a brief description of the dataset's variables.
3. Data Summarization (30 pts):
    a. Calculate summary statistics (mean, median, standard deviation, max, and min) for each relevant numeric variable.
    b. Create grouped summaries based on penguin species using the `group_by` and `summarize` functions.
    c. Discuss the insights gained from the summarization process. Note any interesting patterns, anomalies, missing, etc.
    d. Apply the *skimr*[5] package to the data and discuss the output. How does it compare to what you did in the previous steps? What additional information is provided? How is this useful?
4. Data Visualization (30 pts):
    a. Create at least three different types of visualizations using *ggplot2* (e.g., scatter plot, bar plot, box plot, histogram, etc.) to explore relationships between variables.
    b. Ensure appropriate labeling, coloring, and titling of the visualizations.
    c. Interpret the insights obtained from each visualization.
5. Project Report and Interpretation (25 pts):
    a. Compile a comprehensive project report either directly in the R Notebook or, if you decided not to use a notebook format, an R Script plus a Word document.
    b. Summarize overall patterns, trends, or relationships you discovered. What can you say about each penguin?
    c. Reflect on the value of using R and the *tidyverse* for doing data analysis.

[1] https://www.r-project.org/
[2] https://www.tidyverse.org/
[3] https://ggplot2.tidyverse.org/
[4] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0090081
[5] https://github.com/ropensci/skimr