Assignment #1
October 3, 2022

This assignment is intended to test your understanding of the analysis and visualization of one variable data. Assignments should be submitted as a digitally generated LaTeX document (questions 9-11 can be done by hand in class). The datasets can be found on the jupyter server and listed below:

- pokemon.csv: Information on all generations of Pokémon.
- lego.csv: Information on every lego set ever released
- movies.csv: Information on roughly 1000 movies
- penguins.csv: Information of penguins from several different islands

1. Visualize and describe the distribution of weights for Pokémon with a primary type of rock.
2. Interpret the variance and standard deviation for City themed lego set prices in plain language.
3. Compare the price distribution of Star Wars themed Lego sets to Friends themed lego sets using pairwise boxplots. Describe your comparison.
4. Visualize the distribution of Adelie penguin culmen lengths using a dotplot. In general when is a dotplot advantageous over a histogram?
5. Visualize the distribution for the primary types of pokémon using a barchart.
6. Determine the proportion of pokémon that are 'grass' type in the dataset.
7. Compare the height distribution of fire pokémon to water pokémon using overlayed histograms, and overlaid density plots. Describe your comparison.
8. Compare the runtime distribution of comedy movies to action movies using overlayed density plots. Describe your comparison.
9. Show the relative frequency distribution for movie genres.
10. Prove $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$
11. Assume the following list of numbers is a population that we would like to describe. Calculate each of the following:

$$\{2, 4, 6, 8, 10, 12\}$$

    (a) Arithmetic mean
    (b) Geometric mean
    (c) Median
    (d) IQR
    (e) Variance
    (f) Standard Deviation

12. In the previous question, discuss how answers change if the numbers are a sample intended to make inference on a population.