CSC 401, Package Comparisons #1, Fall 2024 Name_____

Instructions: In these assignments the goal will be to compare the from scratch processes we looked at in class, to the functionality of built-in functions that perform similar tasks. You should compare the functionality of our custom-built examples and the ability to further customize them, to the functionality of package functions. Describe any limitations of the from-scratch versions and the built-in functions. Run an example dataset through both and compare the results.

Submission: Create a Word document that discusses the comparison. Include graphs and explanations here, which package functions you are comparing, etc. With your submission, include your R code file.

Tasks:

1. Compare your basic statistic functions that we created in class, to built-in functions such as mean(), median(), sd(), etc. (from the stats package). Look up the package documentation and consider what options are available, for example, the built-in mean function can also perform trimmed means. Can it do other types of means such as the harmonic or geometric mean as well? Analyze 4-5 statistics/metrics that we have custom code for in lecture. Apply the functions to an example.

2. Compare our distance metrics we developed in class to the dist() function (for example you can find a discussion of its functionality here: https://www.geeksforgeeks.org/how-to-use-dist-function-in-r/#, but also look at the documentation). Apply the functions to an example.

3. Compare the scale() function to our custom scaling methods. Which one(s) is(are) included in the scale() function, and which ones would have to be implemented by hand. Discuss why it is better to apply custom scaling in machine learning to the test and training sets.