CSC 401, Weekly Coding Assignment #10, Fall 2024   Name_____

Instructions: For these weekly coding assignments, you will be asked to extend the examples from class to create custom code to answer the questions below. You will create an R code file that uses built-in datasets as the data sources. You will write the code, and an example showing that the code works. Be sure to include any packages in the code that are required for the functions to run (you may want to clear the environment in RStudio before your final check to make sure nothing is missing). The instructor will run the file to ensure that it works with no errors. Clearly label your code so it's clear which question/task is being responded to.

Submission:
A word document with any explanations (if needed), and a clearly labeled R code file.

Tasks/Questions:

1. Use the agglomerative example from lecture and swap out the distance metric with cosine similarity. Plot the resulting dendrogram.

2. Use the code examples from lecture on agglomerative clustering to swap out the linkage method with a) median linkage, b) centroid linkage. Compare the resulting dendrograms.

3. Perform a factor analysis, using packages as needed, on the abalone dataset from the AppliedPredictiveModeling library. First remove any variables that are categorical unless you are using a package that can handle mixed data. Create appropriate visualizations.

4. Apply PCA and PCR to the cars dataset to predict Price (remove Price from the PCA analysis and then set as the target in the PCR). Produce appropriate graphs. Perform variable selection on the principal components until all the remaining components are significant.