

DATA SCI 415 - Homework 8

Due Saturday, April 12, 2025, 11:59 PM

1. Consider the `USArrests` data. We will now perform hierarchical clustering on the states.
 - (a) Using hierarchical clustering with complete linkage and Euclidean distance, cluster the states.
 - (b) Cut the dendrogram at a height that results in three distinct clusters. Which states belong to which clusters?
 - (c) Hierarchically cluster the states using complete linkage and Euclidean distance, *after scaling the variables to have standard deviation one*. Now which states belong to which clusters?
 - (d) What effect does scaling the variables have on the hierarchical clustering obtained? In your opinion, should the variables be scaled before the inter-observation dissimilarities are computed? Provide a justification for your answer.