DSI- 06 Homework 9:

Introduction with Statistical Learning with Applications to Python (ISLP) pg. 554-555

9. Consider the USArrests data. We will now perform hierarchical clustering on the states.

1. Using hierarchical clustering with complete linkage and Euclidean distance, cluster the states.
2. Cut the dendrogram at a height that results in three distinct clusters. Which states belong to which clusters?
3. Hierarchically cluster the states using complete linkage and Euclidean distance, *after scaling the variables to have standard deviation one*.
4. 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.