

**Problem 1 . (B)**

It is divided into 2 clusters.

Cluster 1: This group is centered around shorter eruption times (approximately 1.5 to 3 minutes) and shorter waiting times (approximately 45 to 70 minutes).

Cluster 2: This group is centered around longer eruption times (approximately 4 to 5 minutes) and longer waiting times (around 75 to 95 minutes).

**Problem 1. (C)****Steps of Hierarchical Clustering (Bottom-Up Approach)**

1. Initialize Clusters: Begin with each data point as its own cluster, resulting in  $n$  clusters for  $n$  data points.
2. Calculate Proximity: Compute a distance (e.g., Euclidean) or similarity matrix between each pair of clusters.
3. Merge Closest Clusters: Identify and merge the two clusters that are closest to each other, reducing the total number of clusters by one.
4. Update Proximity Matrix: Recalculate the distances between the newly formed cluster and all remaining clusters.
5. Repeat Steps 3 and 4: Continue merging the closest clusters iteratively until only a single cluster remains or until reaching a predefined number of clusters.
6. Create Dendrogram: Visualize the clustering process with a dendrogram, which shows the hierarchical relationships and merging sequence of clusters.

As indicated by the scatter plot, Hierarchical clustering is suitable in the case of an Old Faithful Geyser dataset, as two unequivocal clusters are separated. It is ideal for this kind of Hierarchical clustering since, through its bottom-up approach, the method can keep merging smaller groups into the two main clusters progressively, hence showing the natural structure of the data. Besides, since there are merely 272 cases in the dataset, the computation intensity of hierarchical clustering is within the reach of a modern-day computer system. Therefore, this will be a good approach to try out and visualize the clusters using this method.