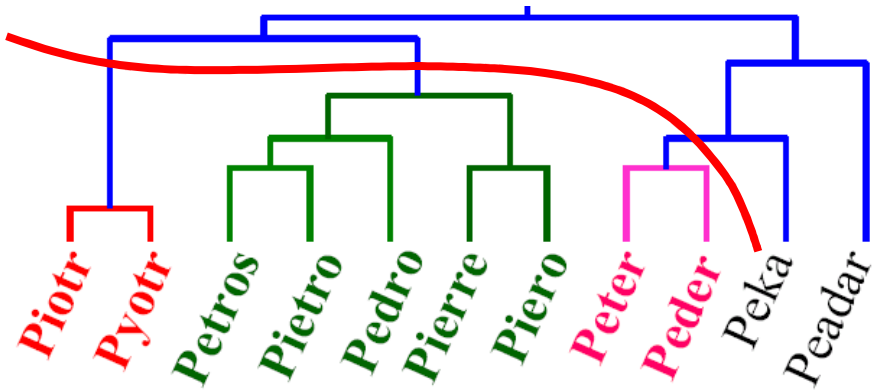
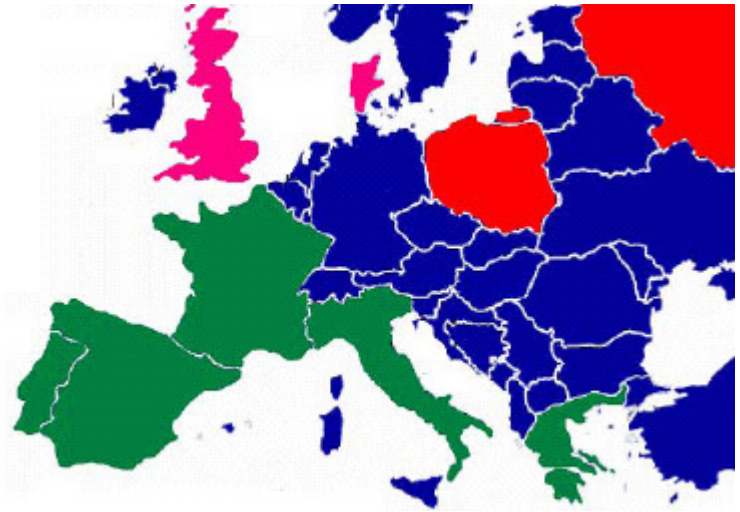
# Hierarchical clustering

- Organize data in a hierarchical fashion (dendrogram)

- Clustering obtained by cutting the dendrogram at a desired level: each connected component forms a cluster.



- How to do it?

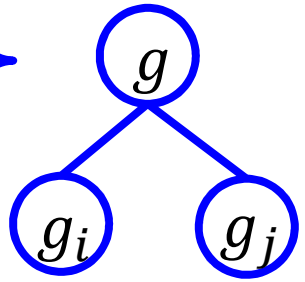
# Bottom up hierarchical clustering

- Assign each data point to its own cluster, $g_1 = \{x_1\}, g_2 = \{x_2\}, \dots, g_m = \{x_m\}$, and let $G = \{g_1, g_2, \dots, g_m\}$

$$D(g_i, g_j) = \min_{x \in g_i, y \in g_j} d(x, y)$$

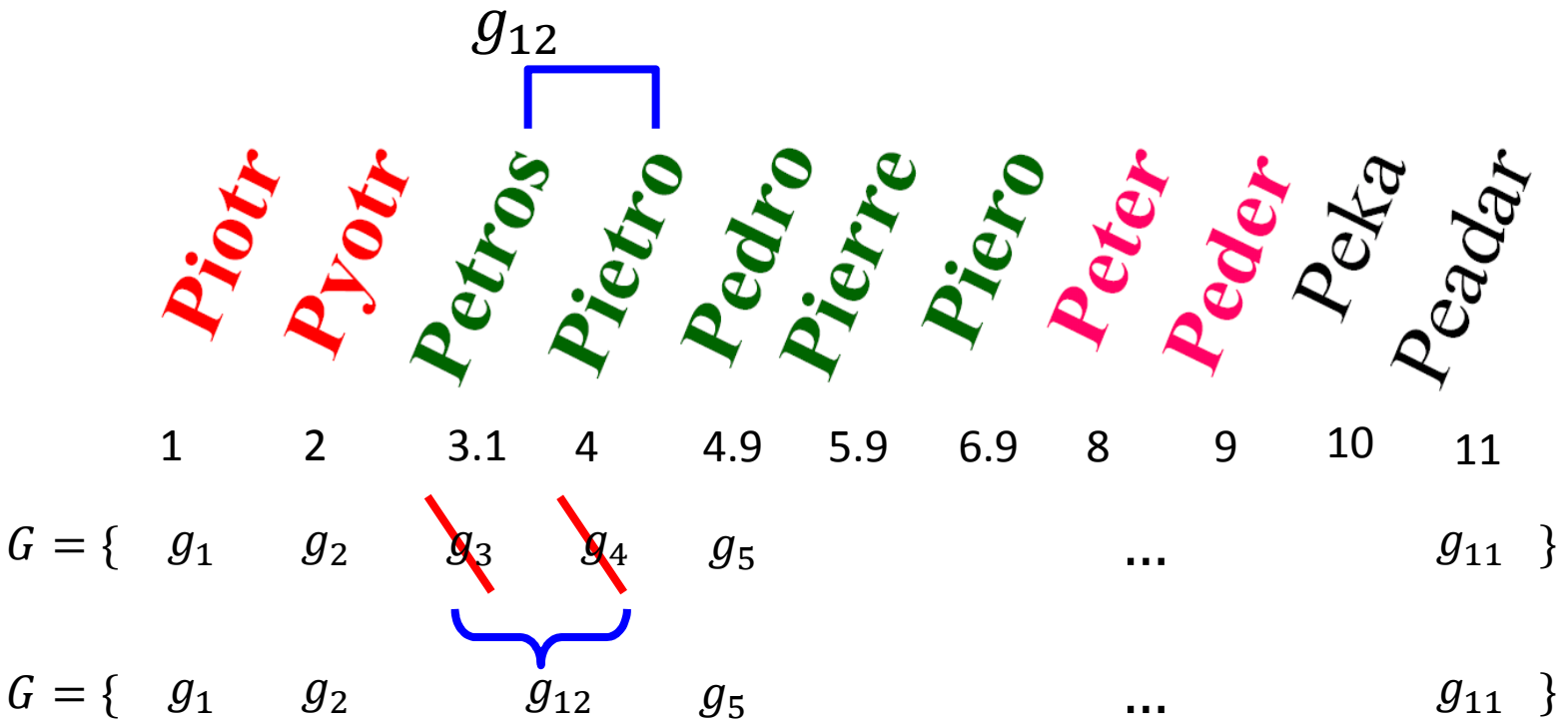
- Do
  - Find two clusters to merge: $i, j = argmin_{1 \le i,j \le |G|} \; D(g_i, g_j)$
  - Merge the two clusters to a new cluster: $g \leftarrow g_i \cup g_j$
  - Remove the merged clusters: $G \leftarrow G \setminus g_i, G \leftarrow G \setminus g_j$
  - Add the new cluster: $G \leftarrow G \cup \{g\}$

keep track of relations

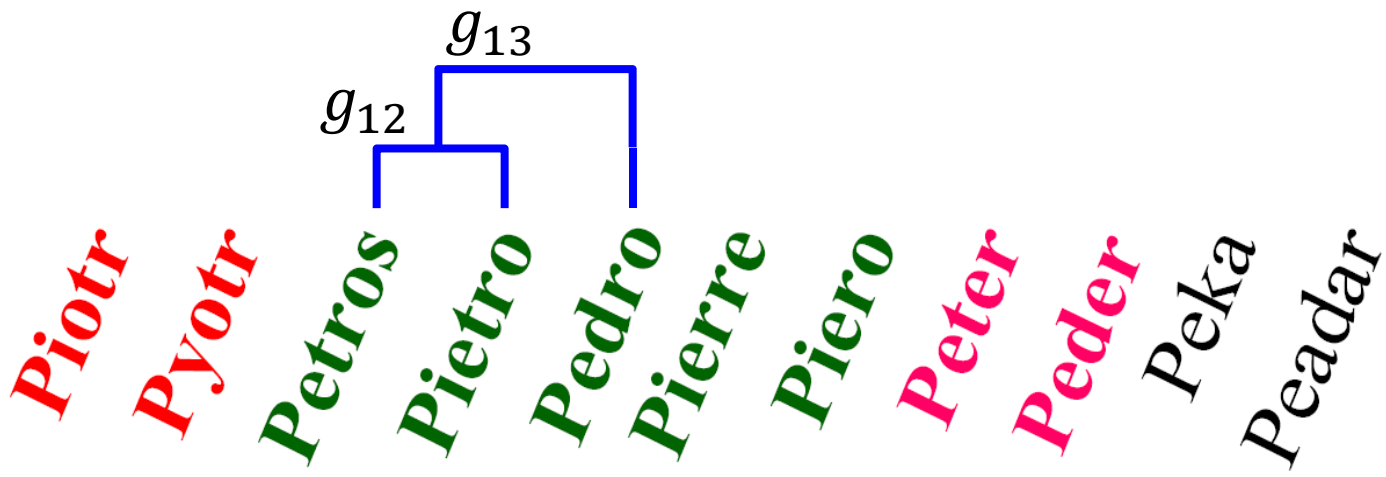



- While $|G| > 1$

# Hierarchical clustering: step-2



# Hierarchical clustering: step-3




$G = \{ \; g_1 \quad g_2 \quad g_3 \quad g_4 \quad g_5 \quad \ldots \quad g_{11} \; \}$

$G = \{ \; g_1 \quad g_2 \quad g_{12} \quad g_5 \quad \ldots \quad g_{11} \; \}$

$G = \{ \; g_1 \quad g_2 \quad g_{13} \quad \ldots \quad g_{11} \; \}$