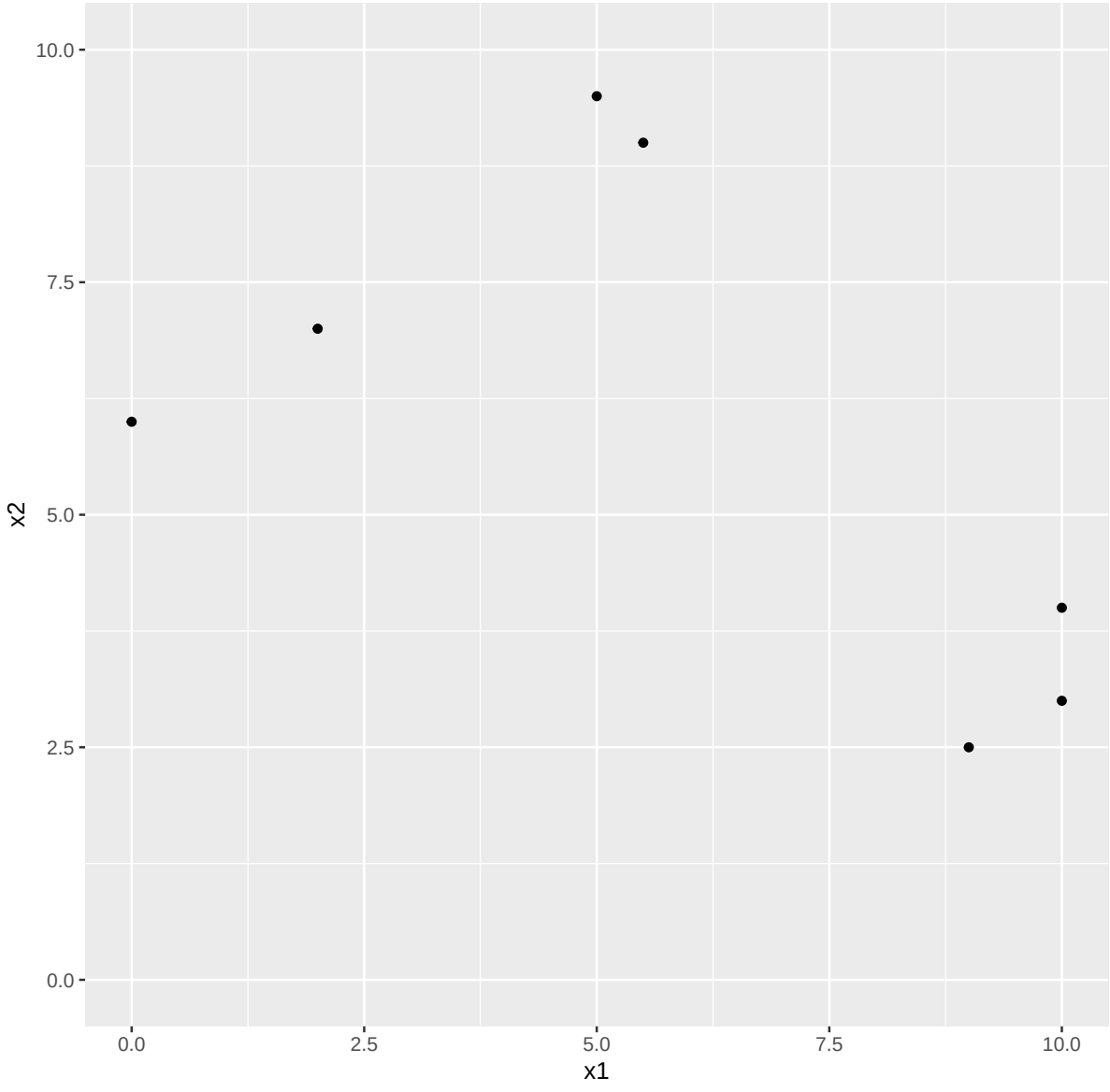
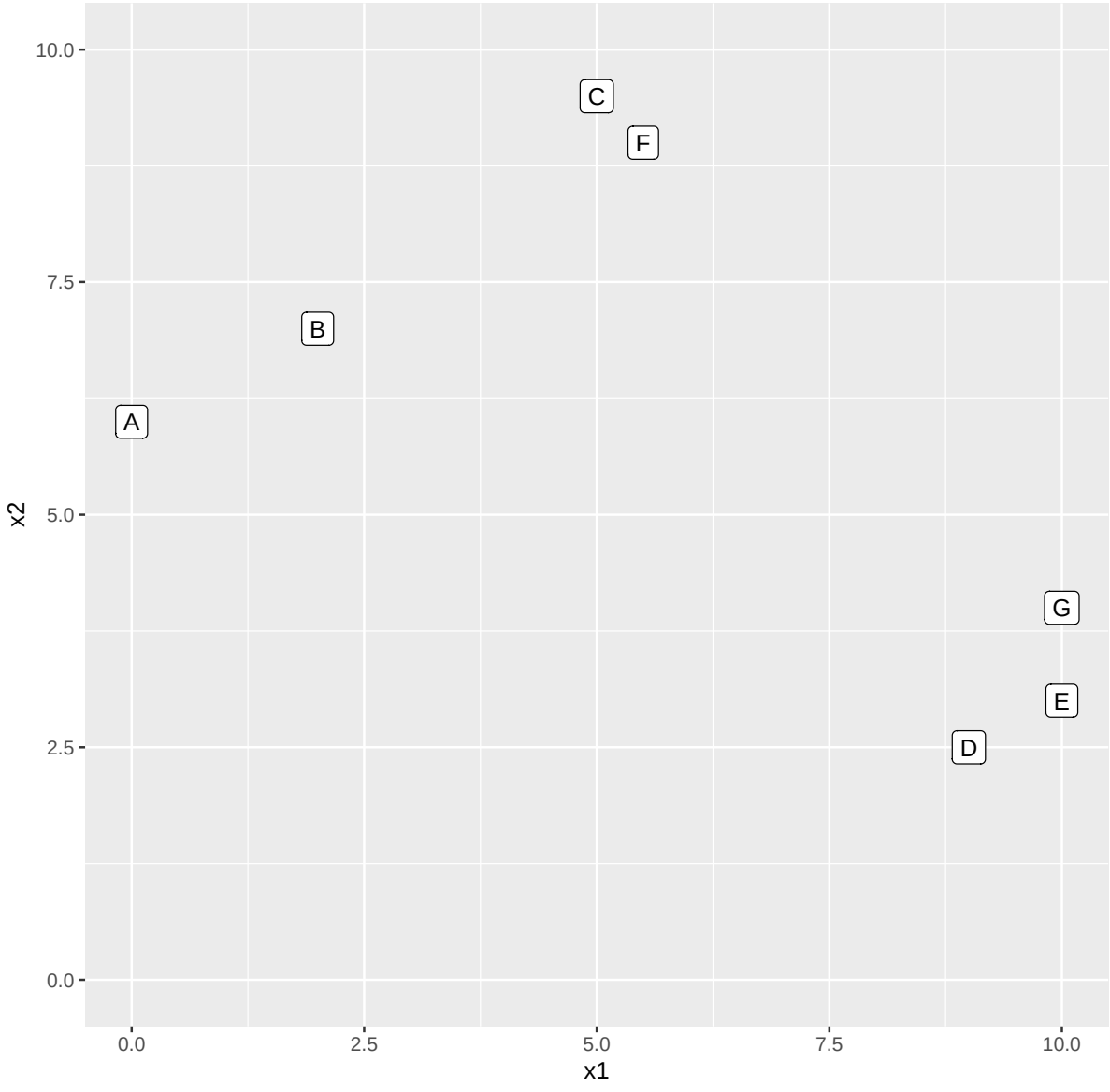
Hierarchical Clustering

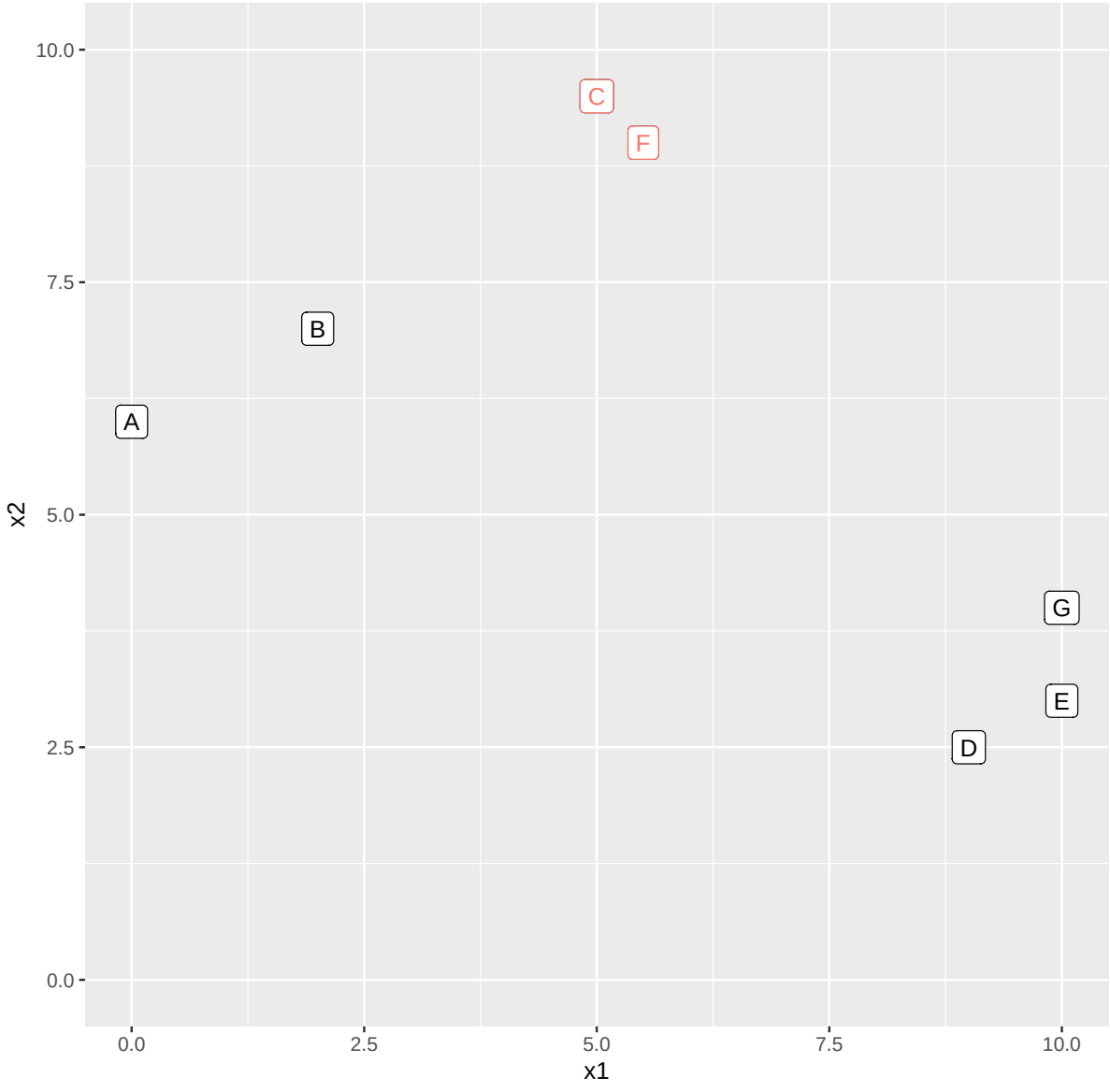
Hierarchical clustering - Centroid Linkage



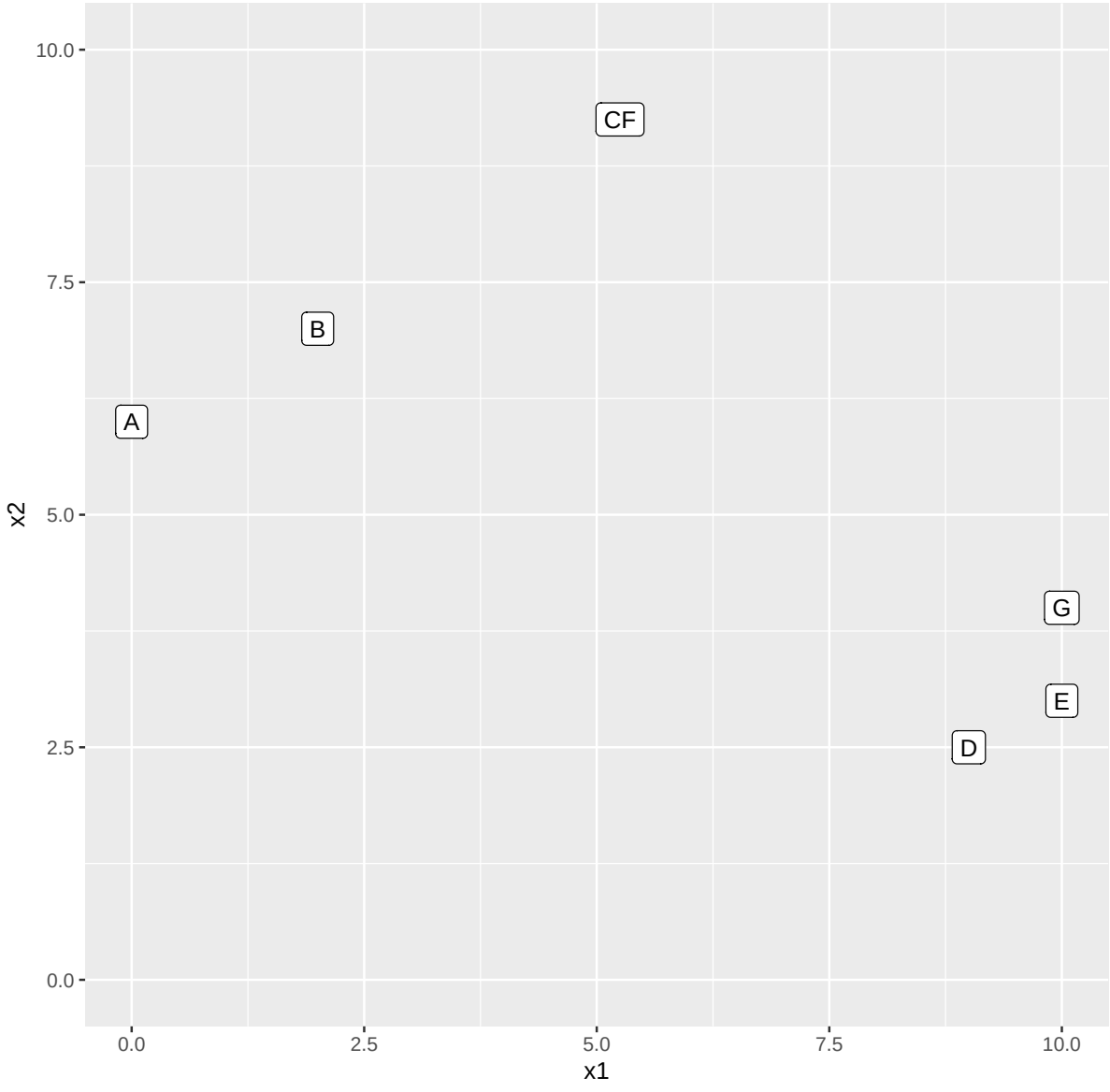
Label the Points



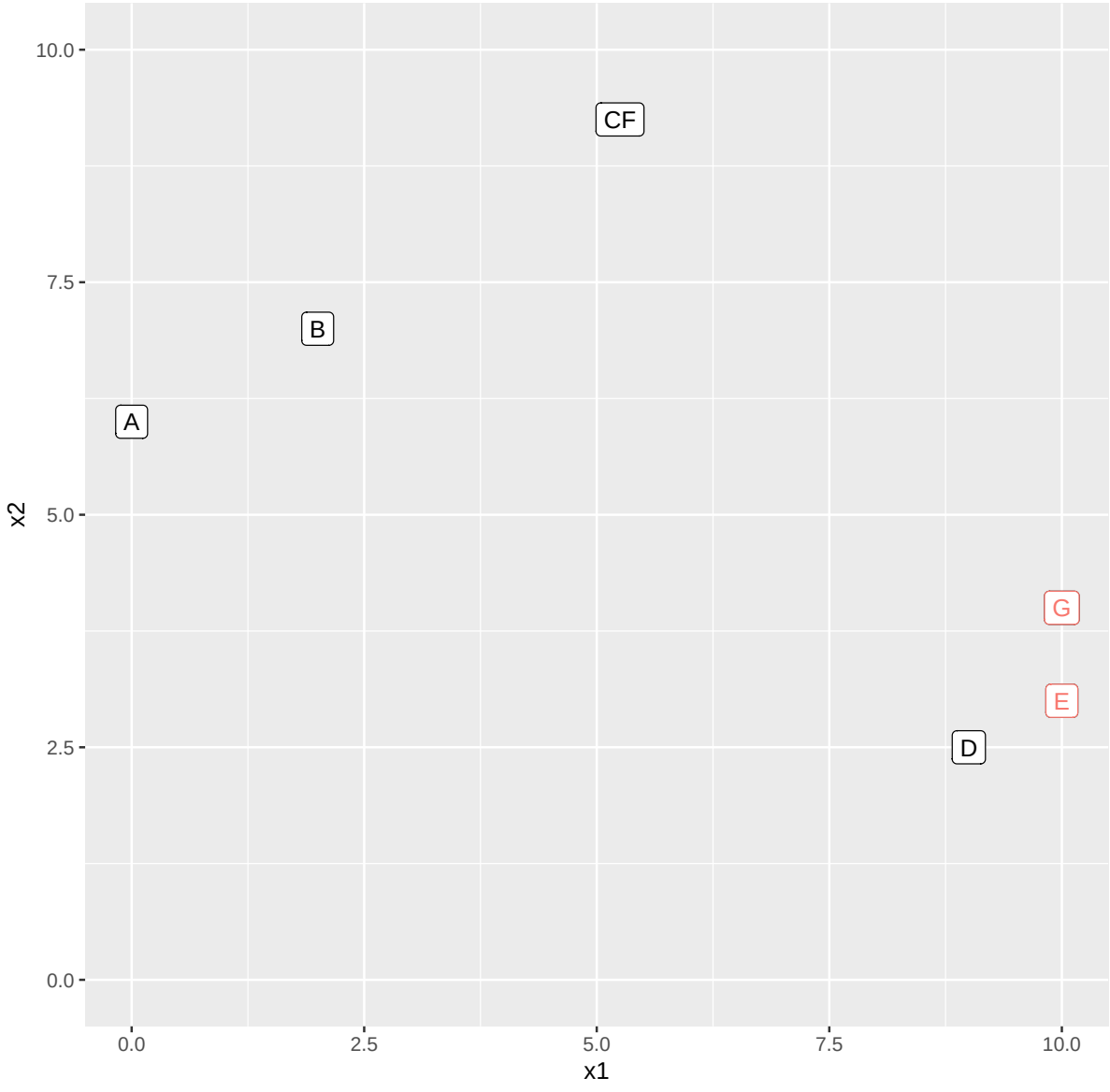
Identify the pair with the smallest distance.



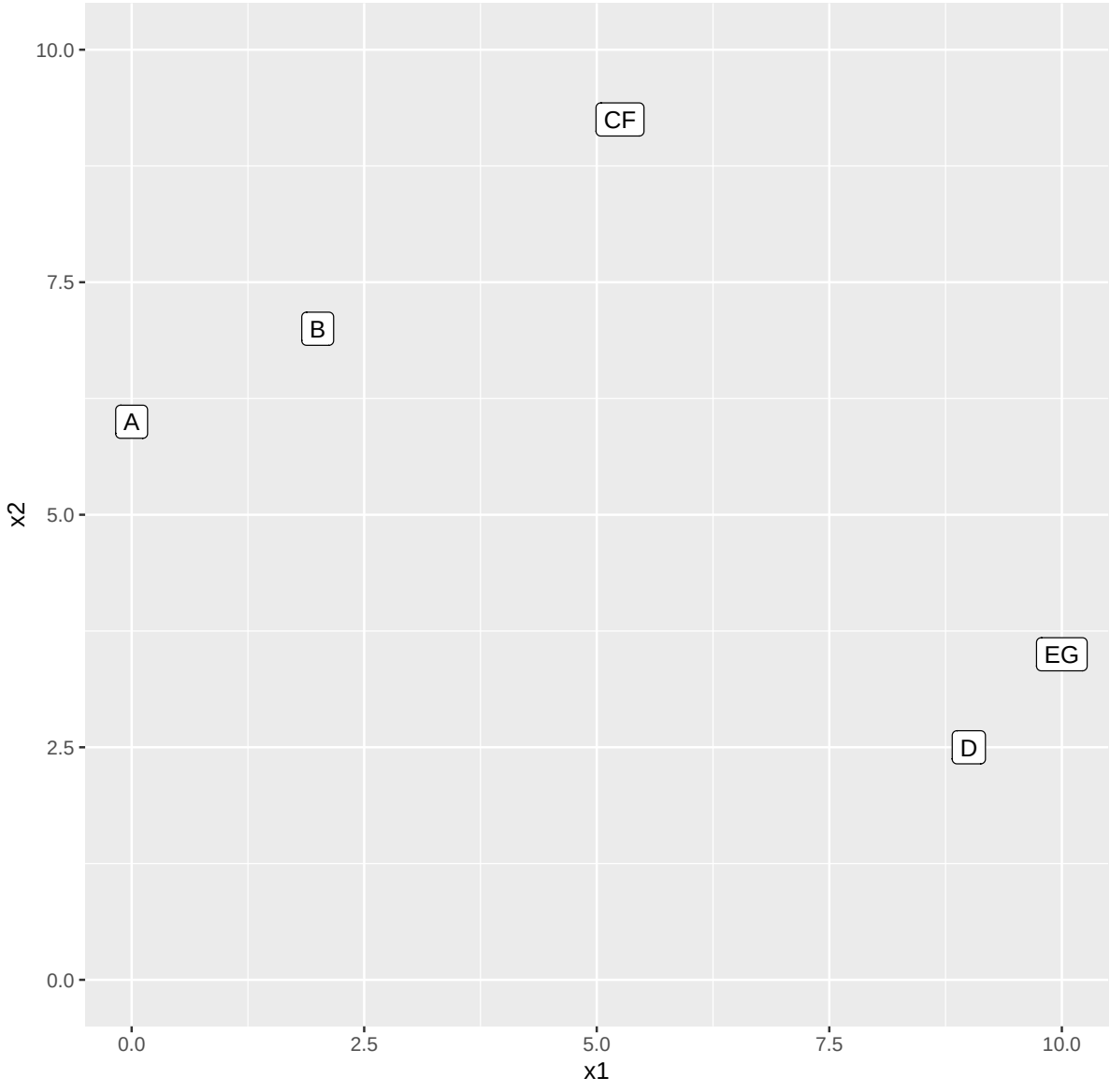
Group the pair



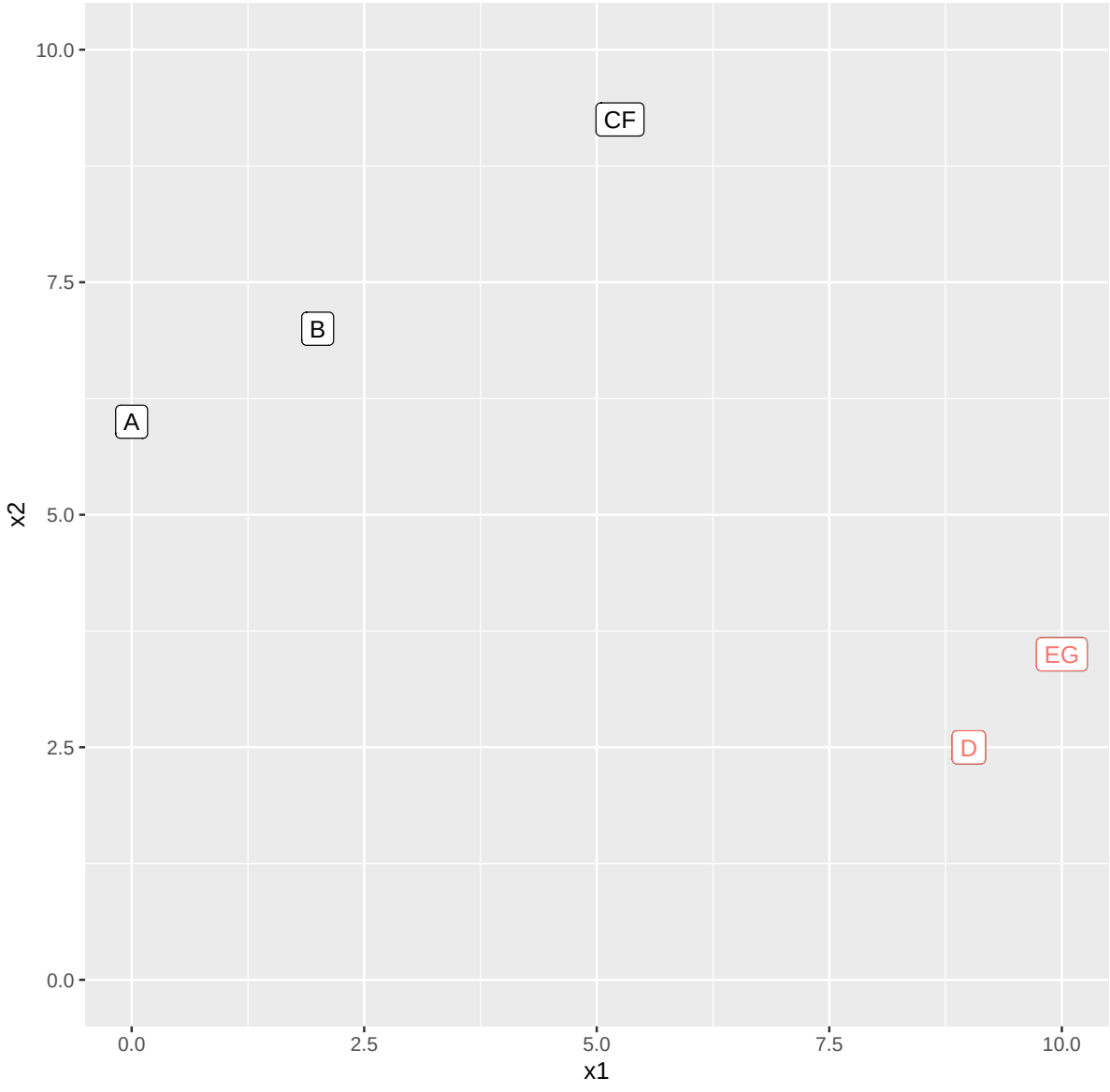
Identify the pair with the smallest distance.



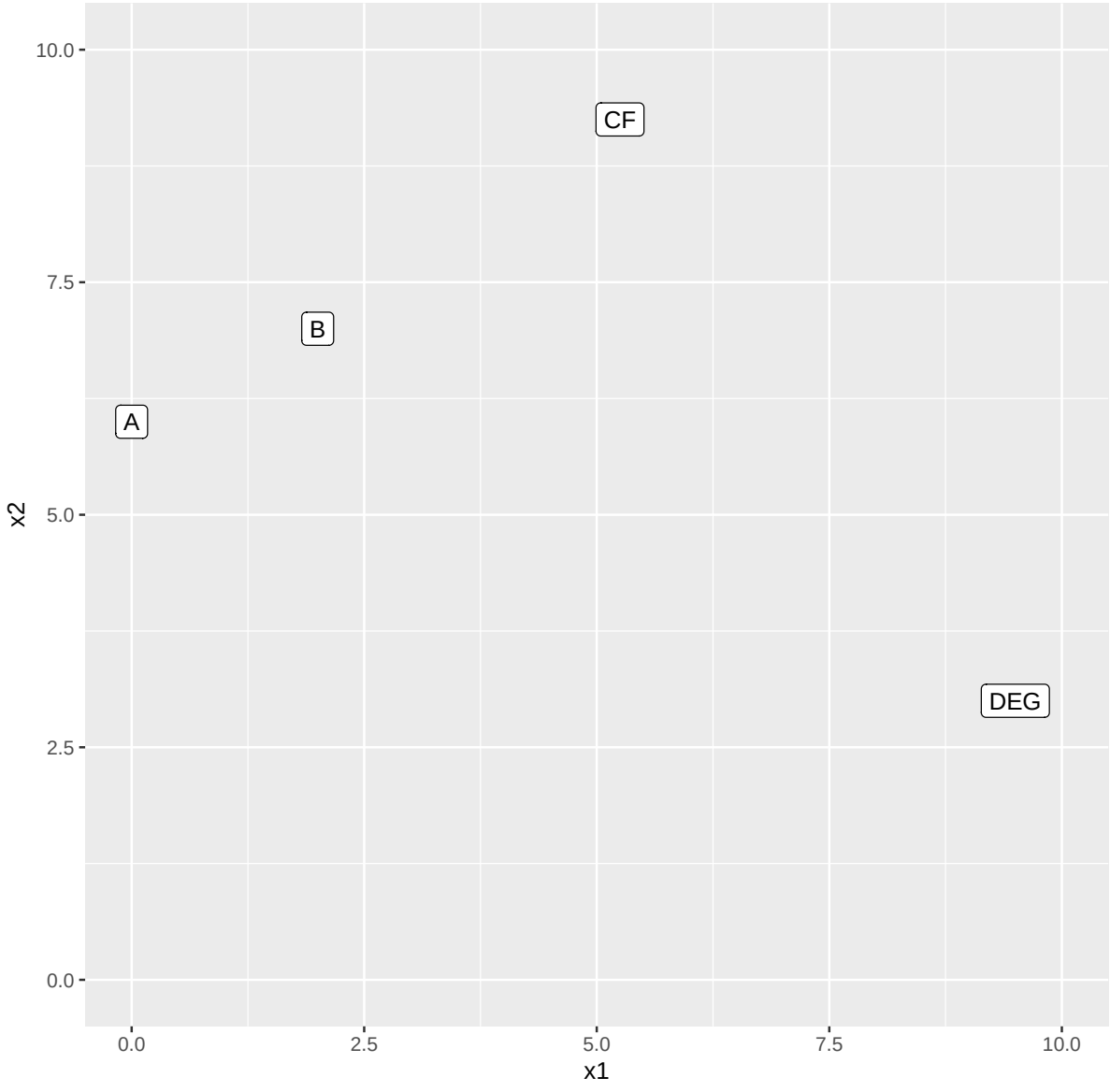
Group the pair



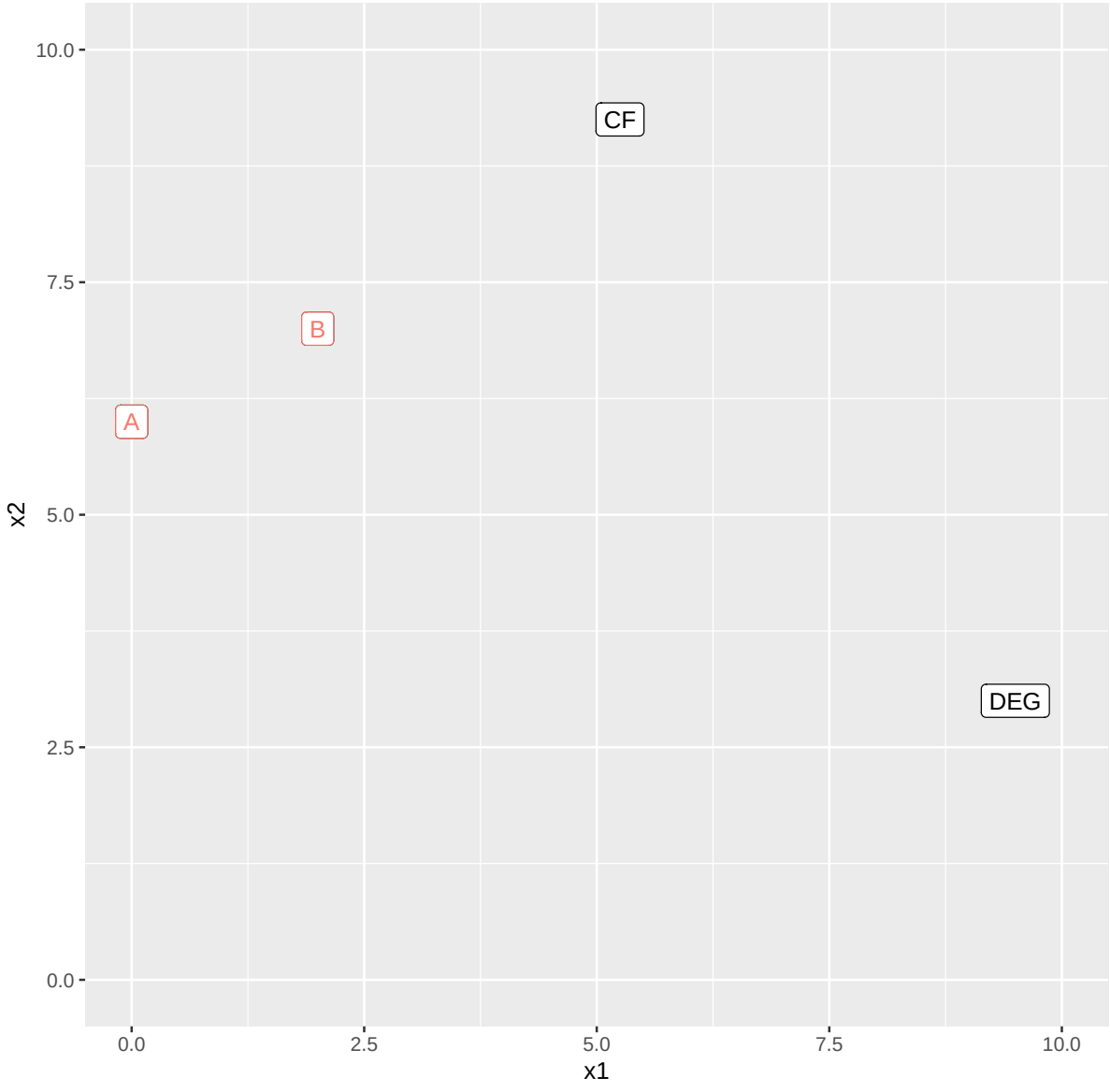
Identify the pair with the smallest distance.



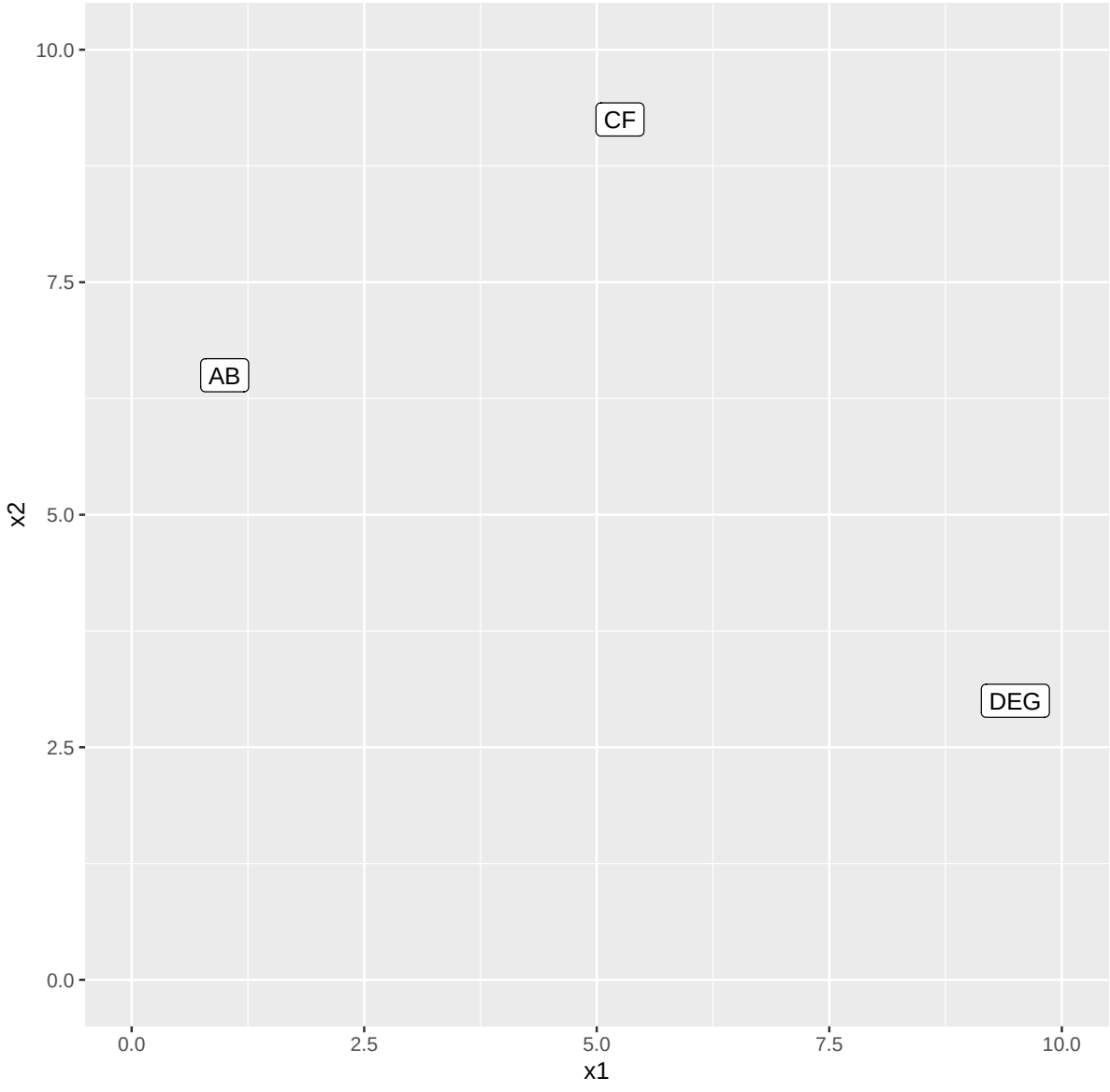
Group the pair



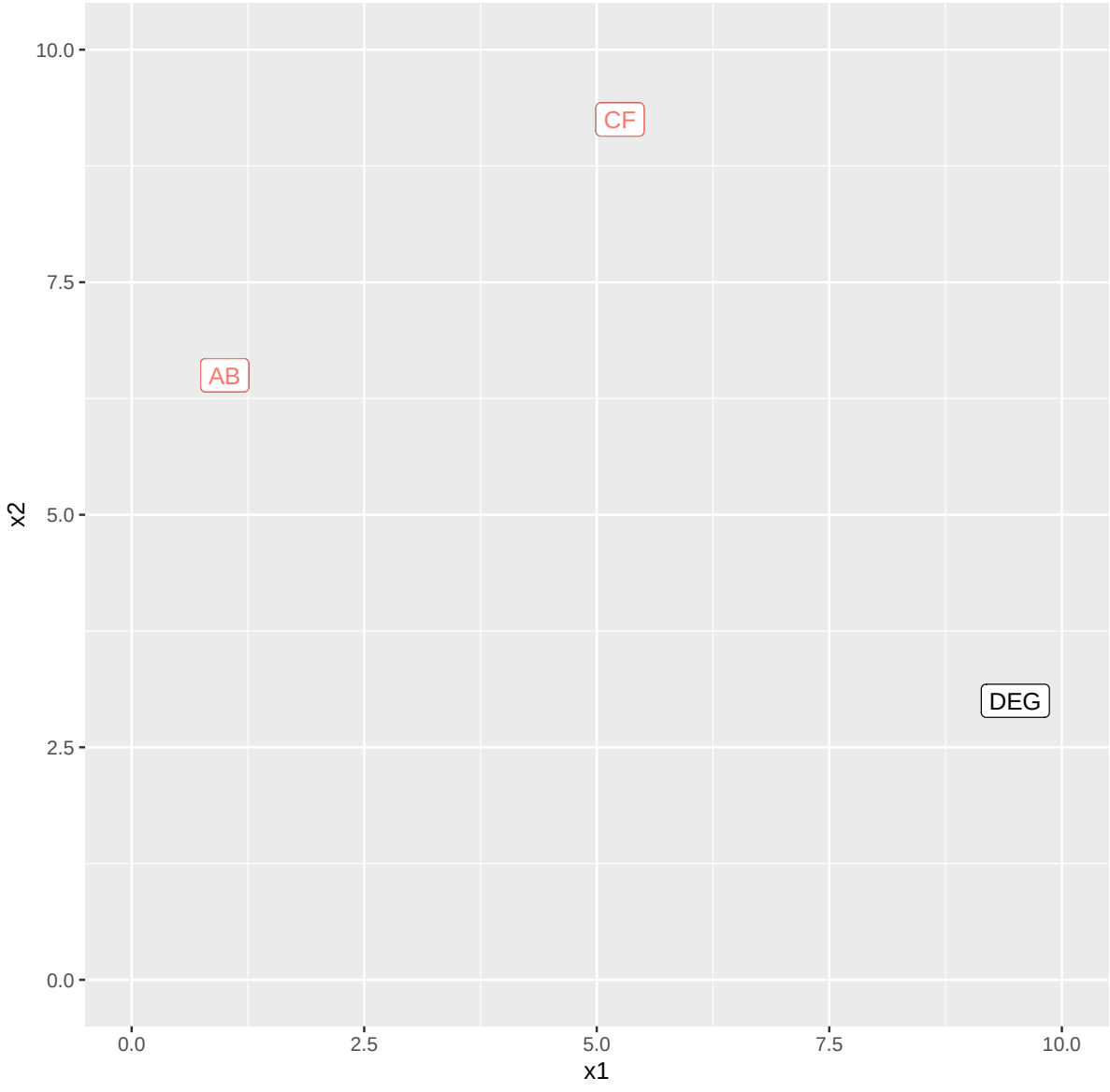
Identify the pair with the smallest distance.



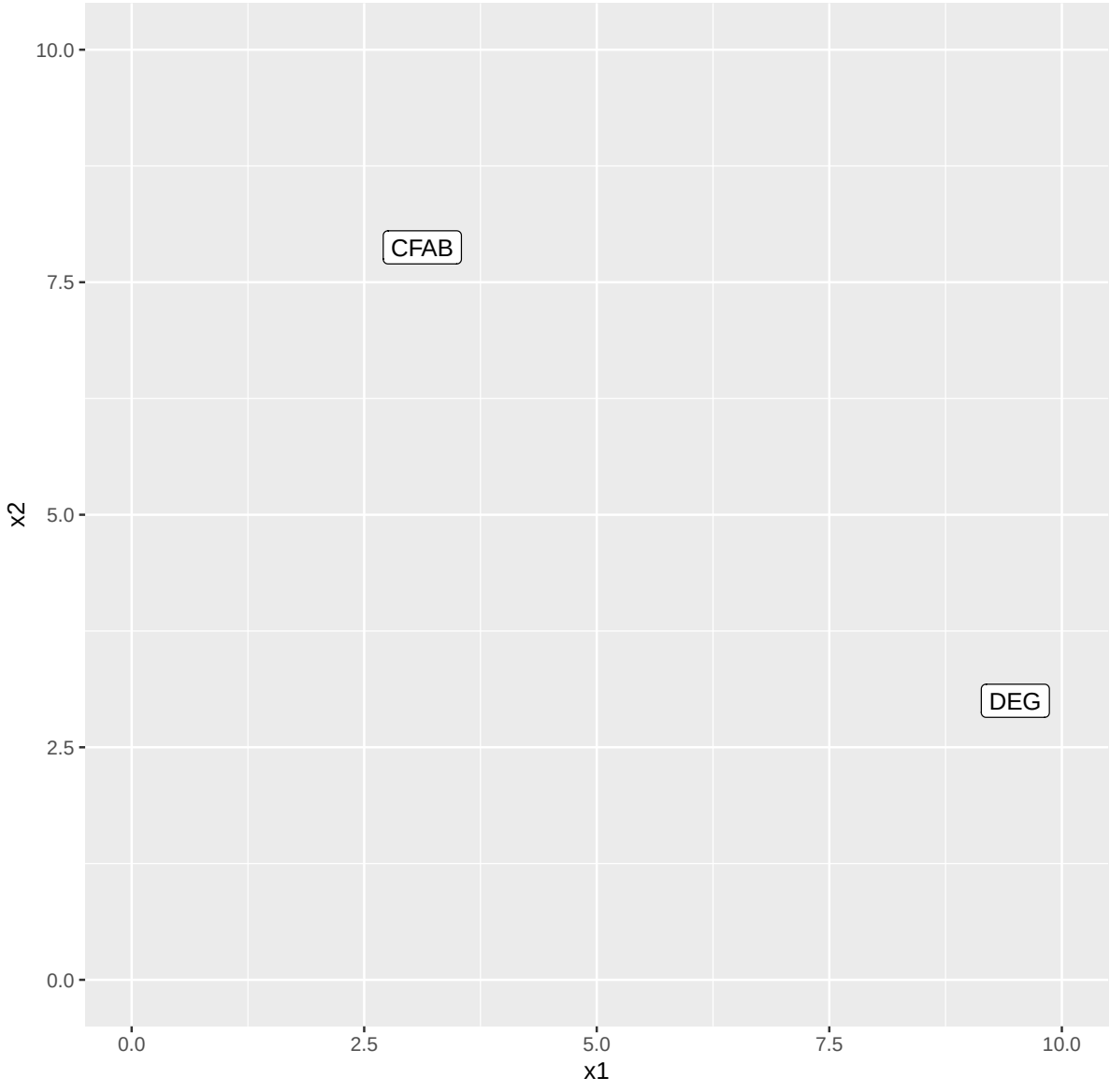
Group the pair



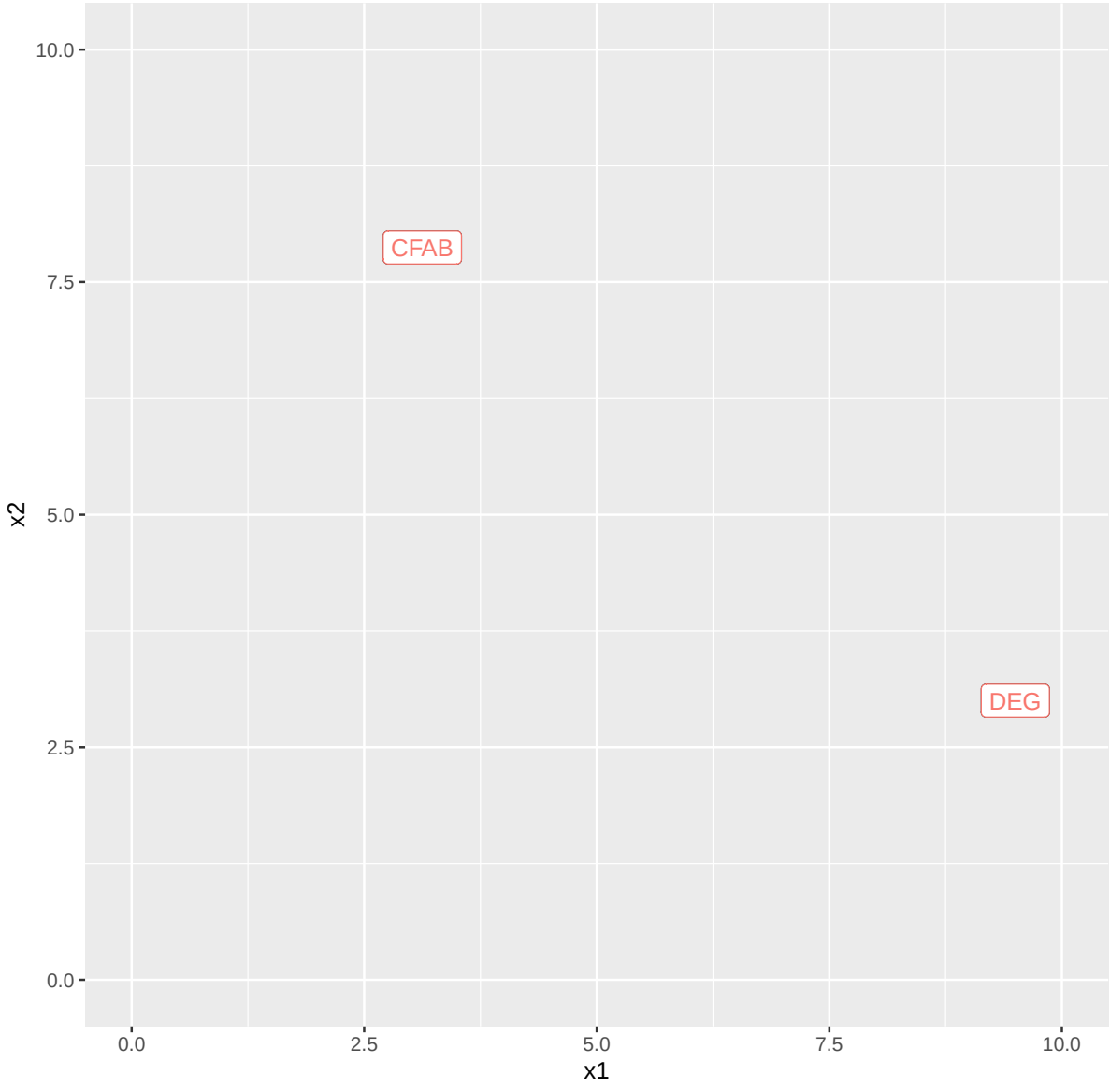
Identify the pair with the smallest distance.



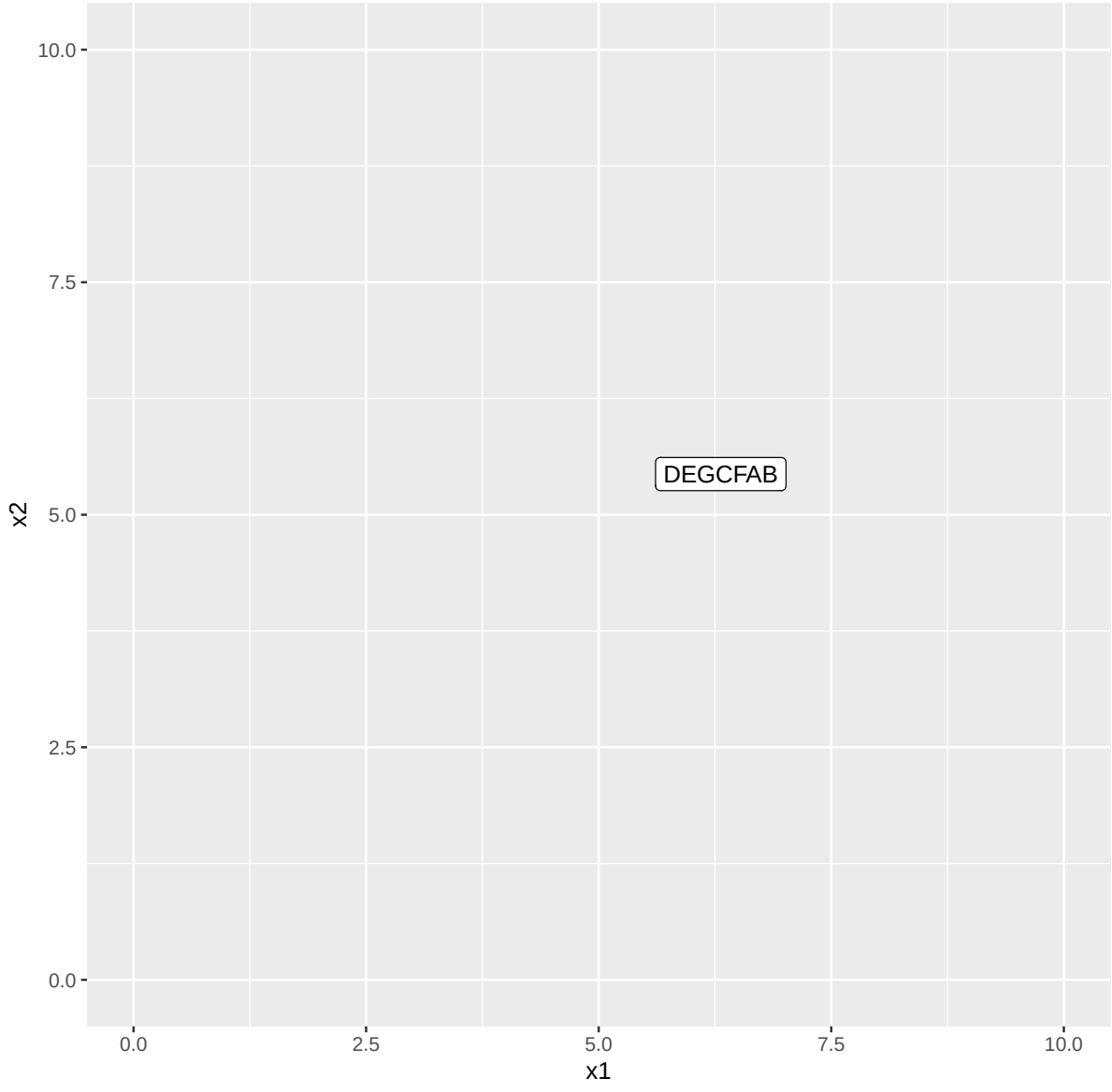
Group the pair



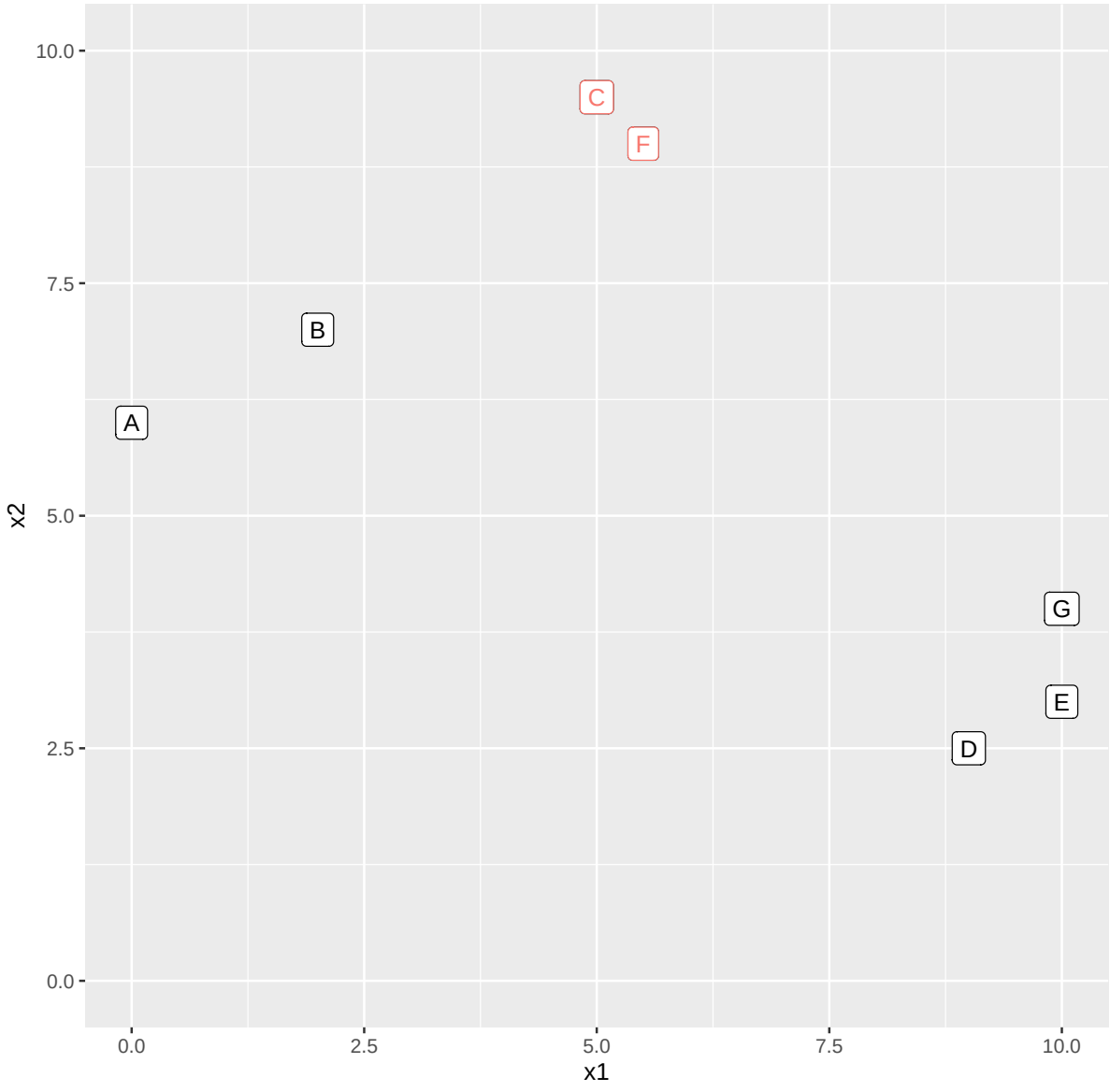
Identify the pair with the smallest distance.



Group the pair

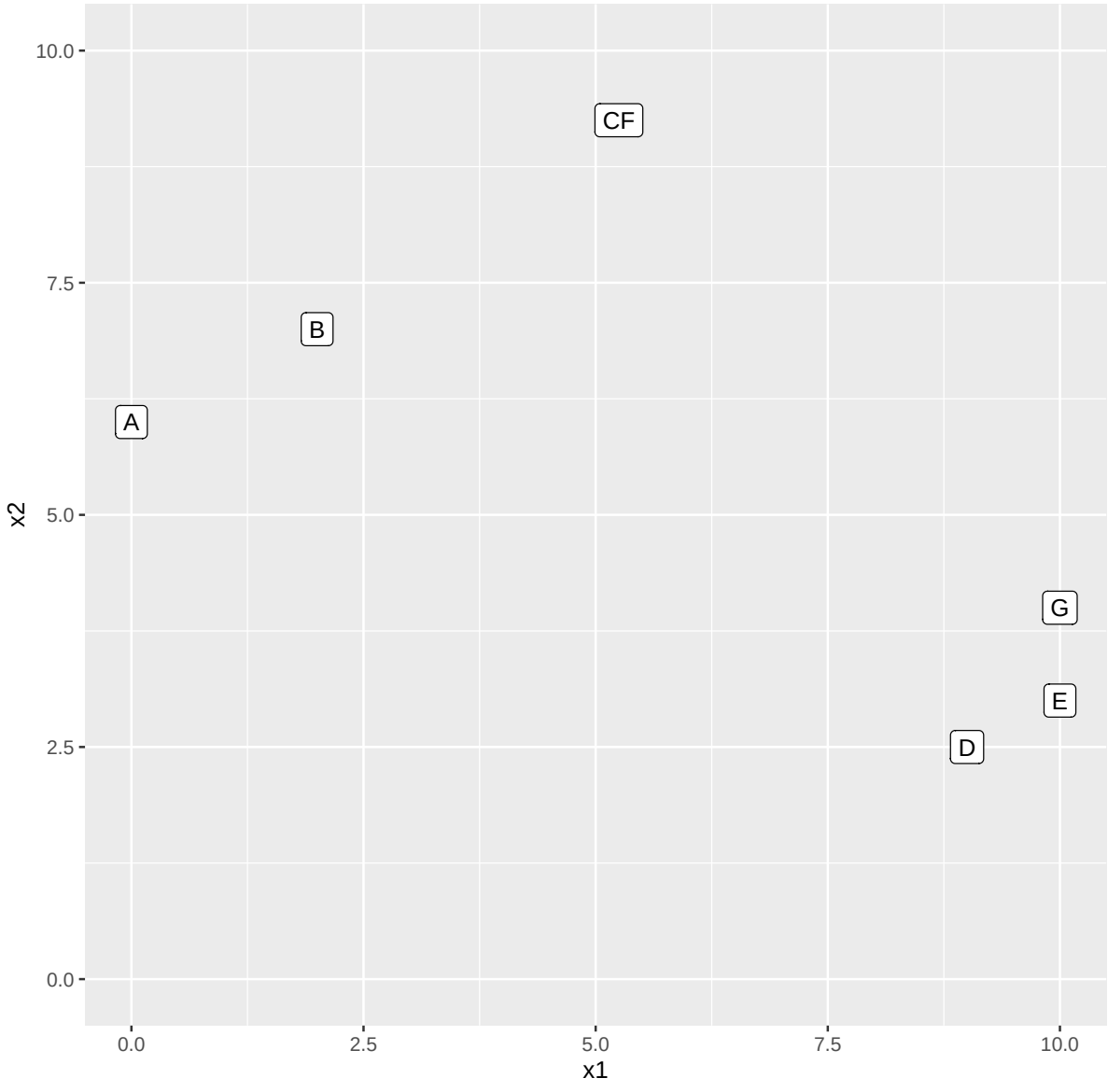


Dendrogram



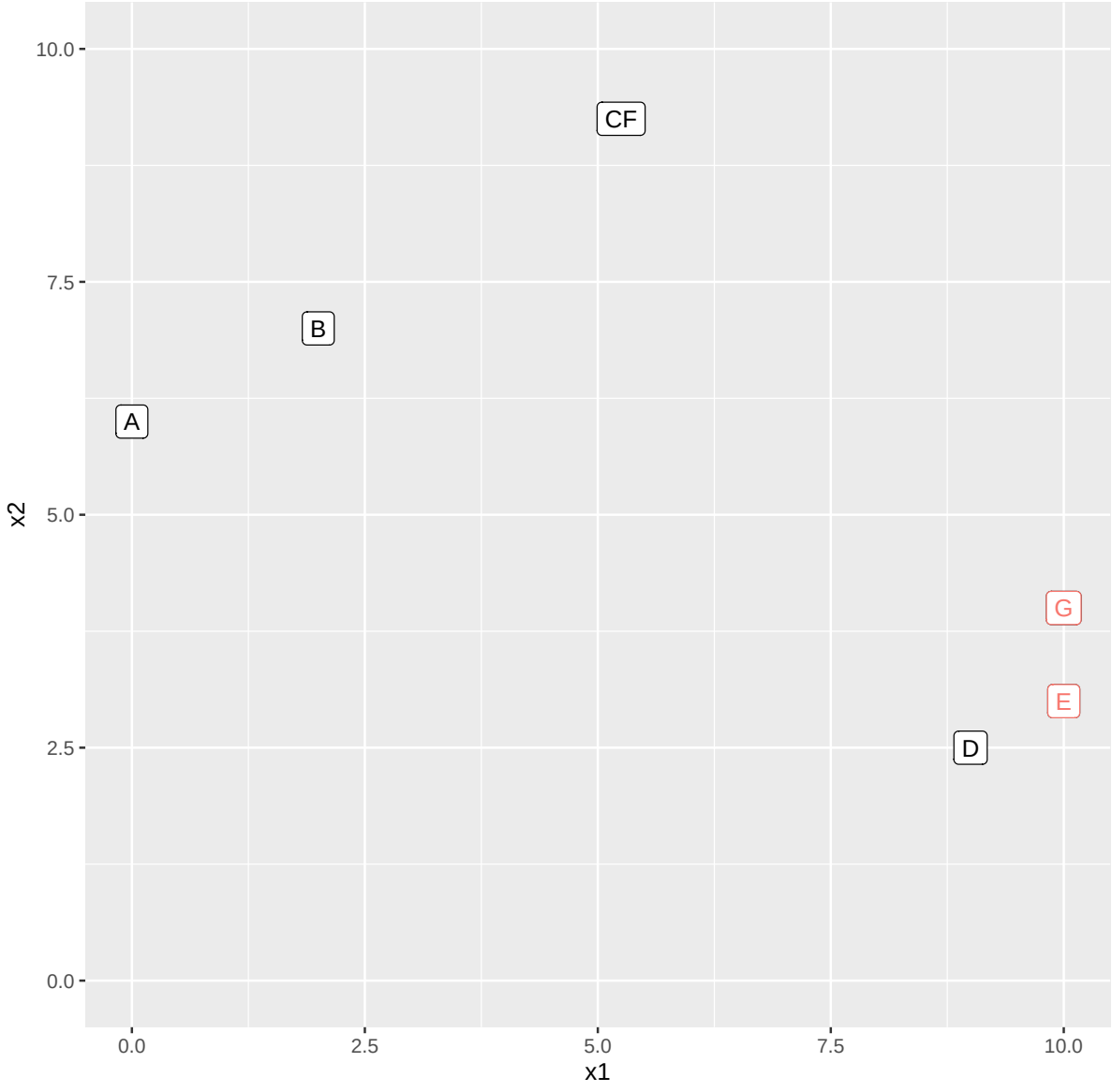
- Distance: 0.71

Group the pair



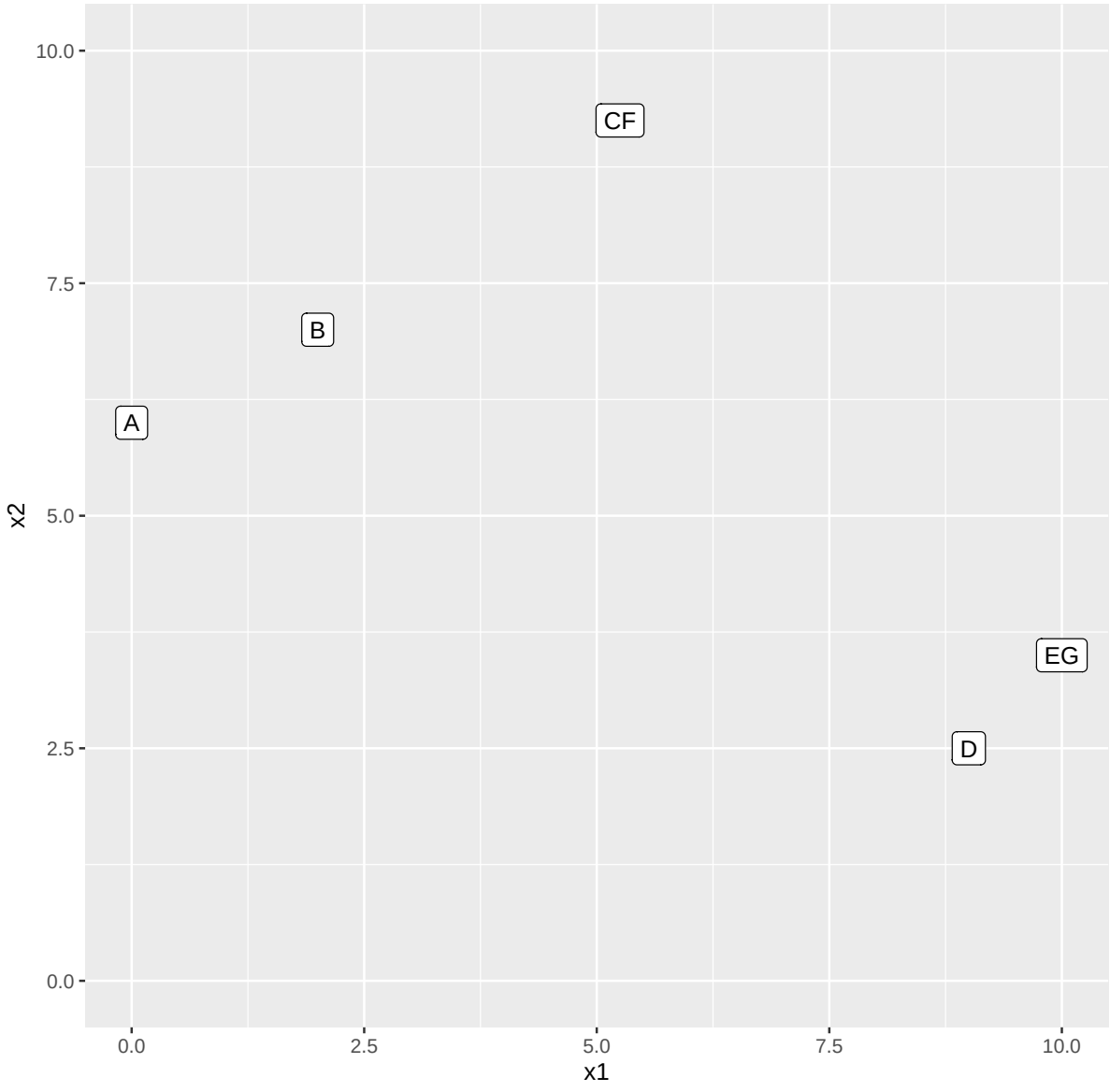
- Distance: 0.71

Identify the pair with the smallest distance.



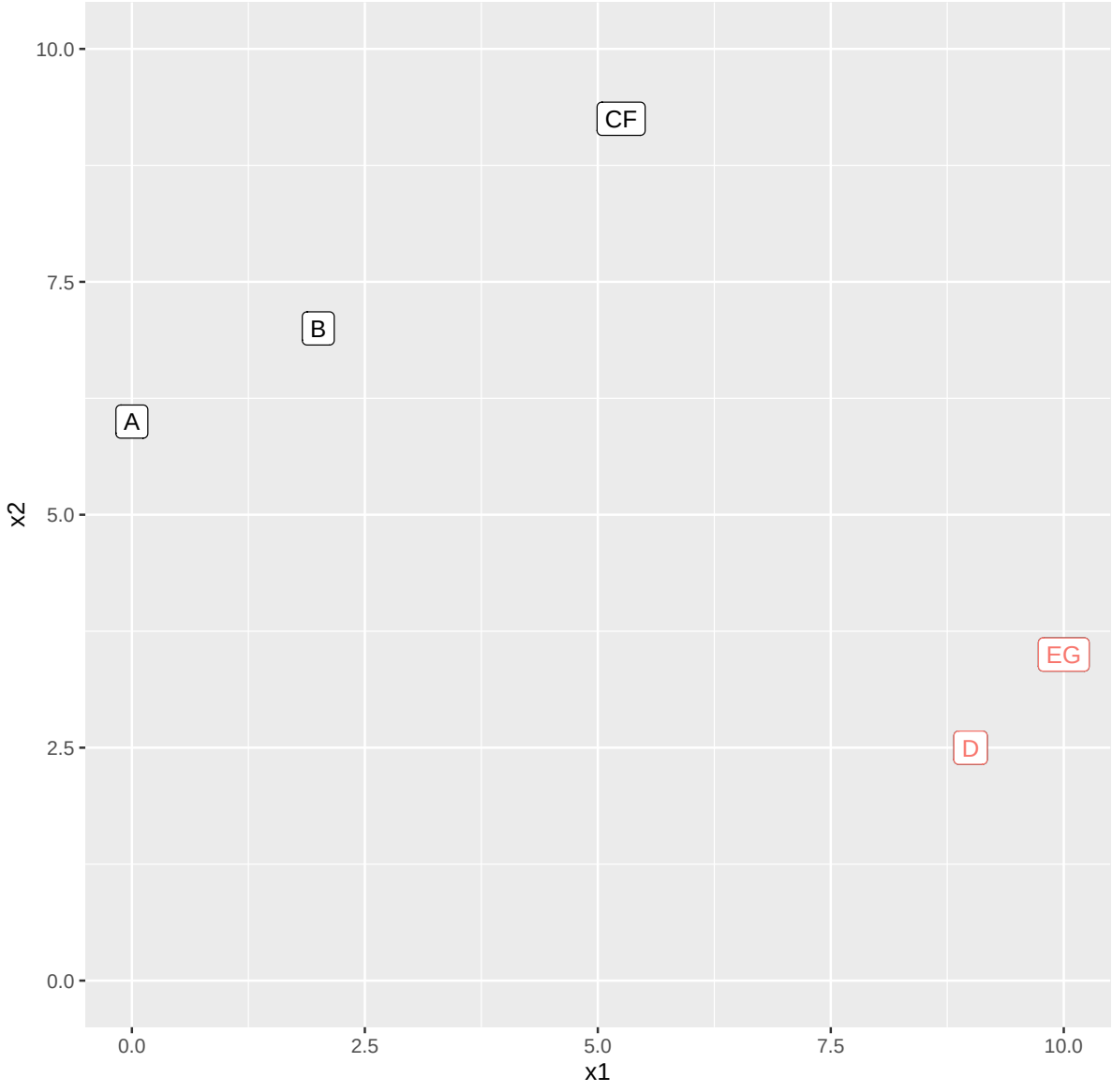
- Distance: 1

Group the pair



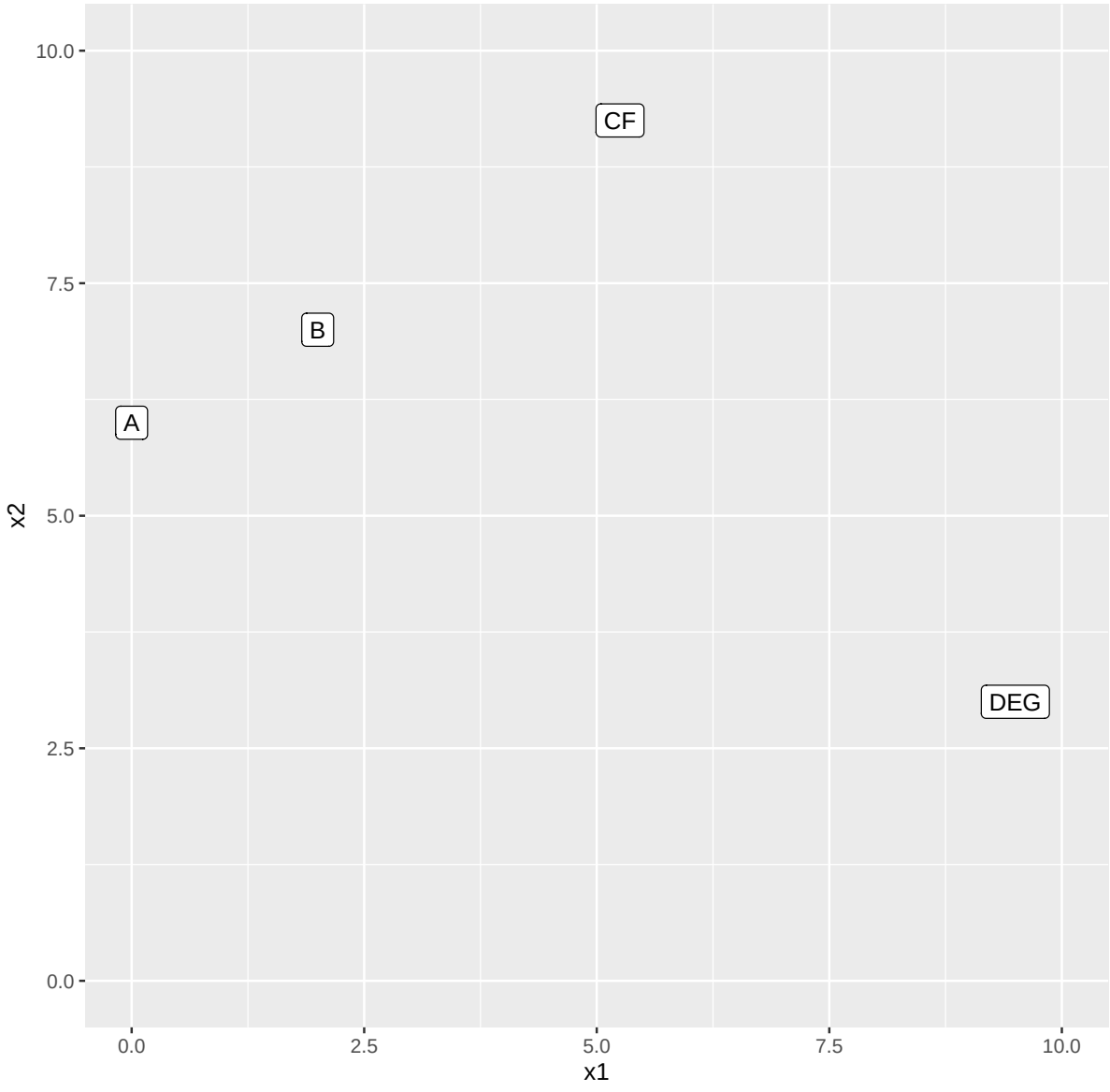
- Distance: 1

Identify the pair with the smallest distance.



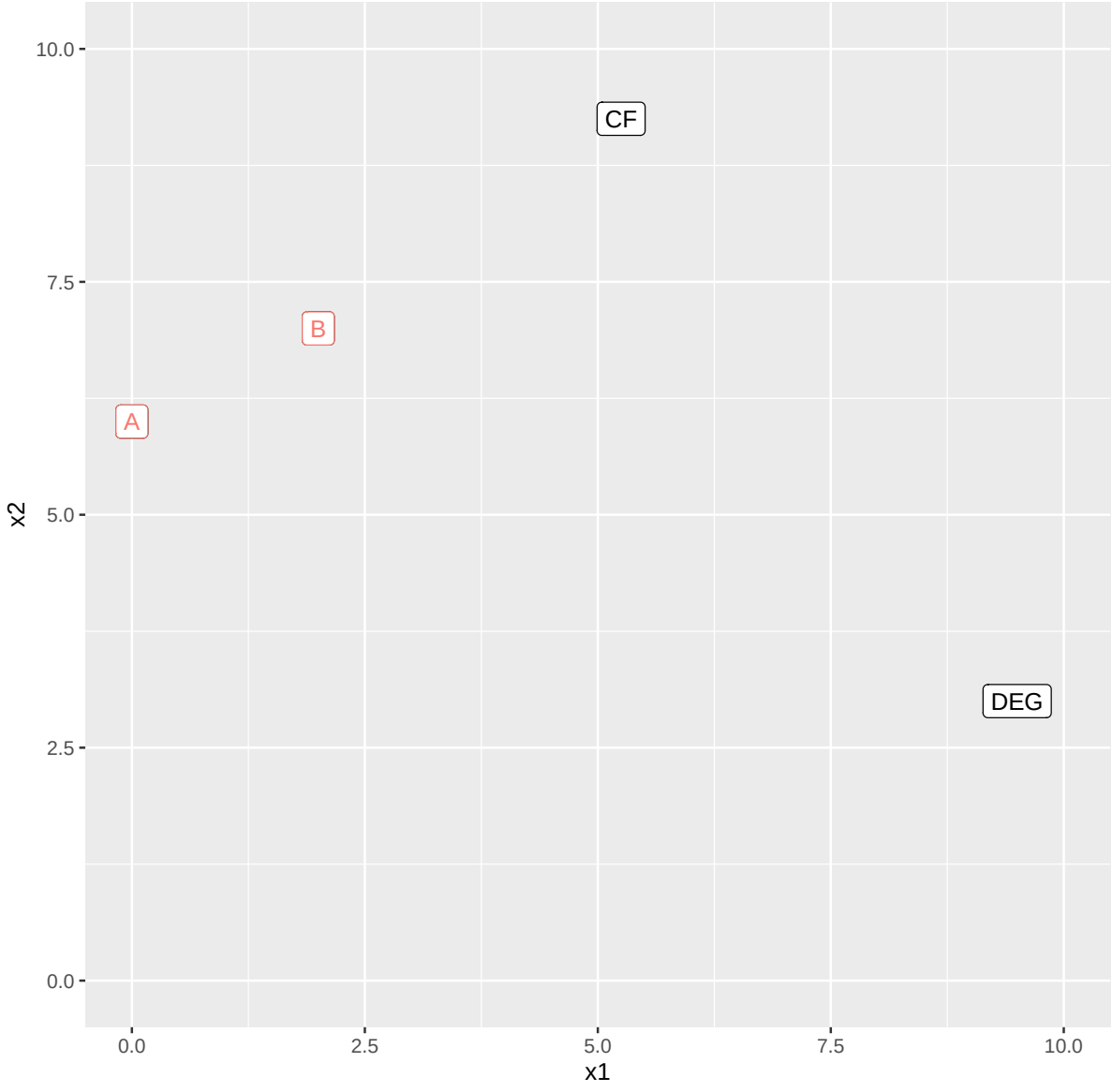
- Distance: 1.41

Group the pair



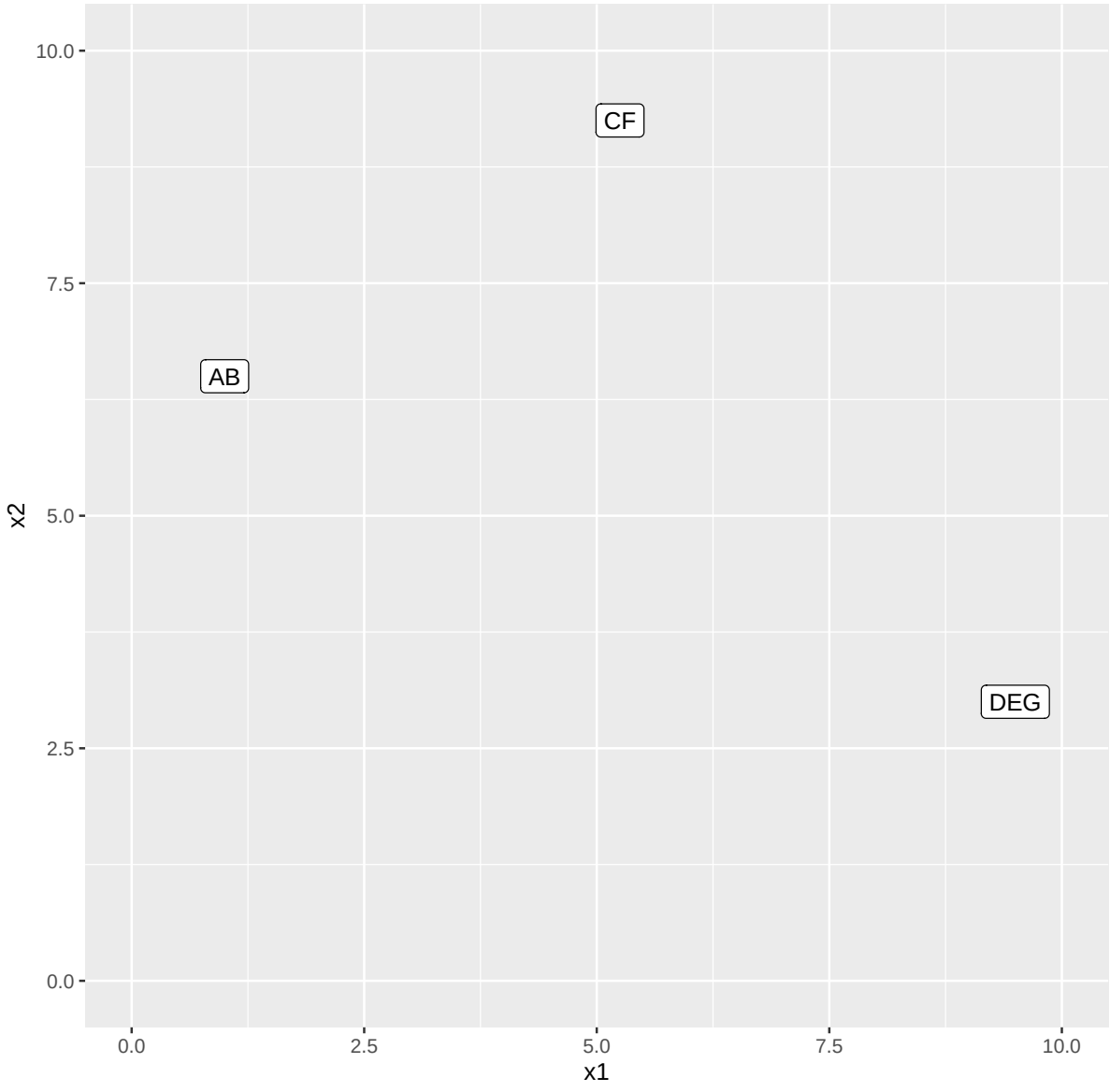
- Distance: 1.41

Identify the pair with the smallest distance.



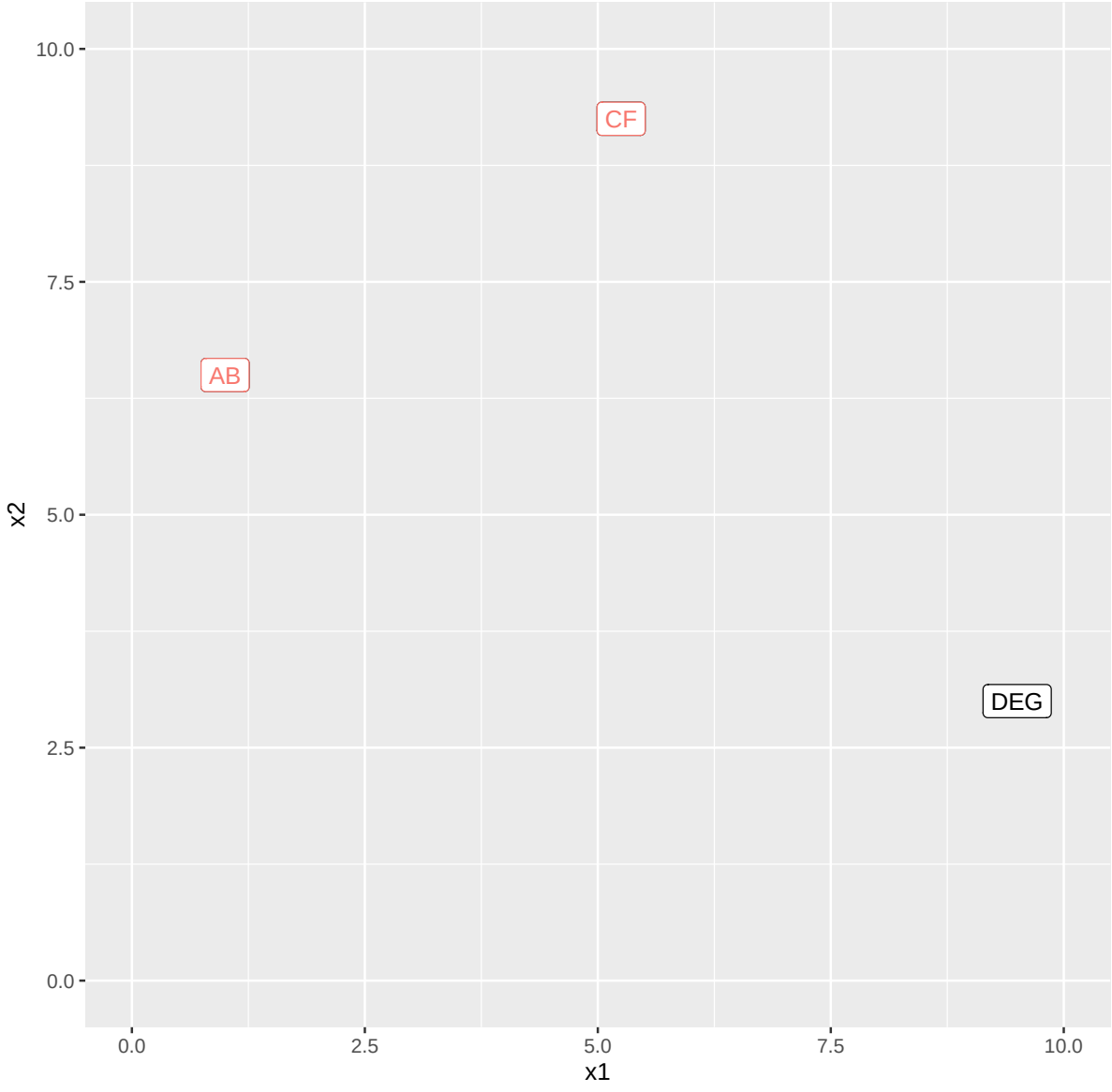
- Distance: 2.24

Group the pair



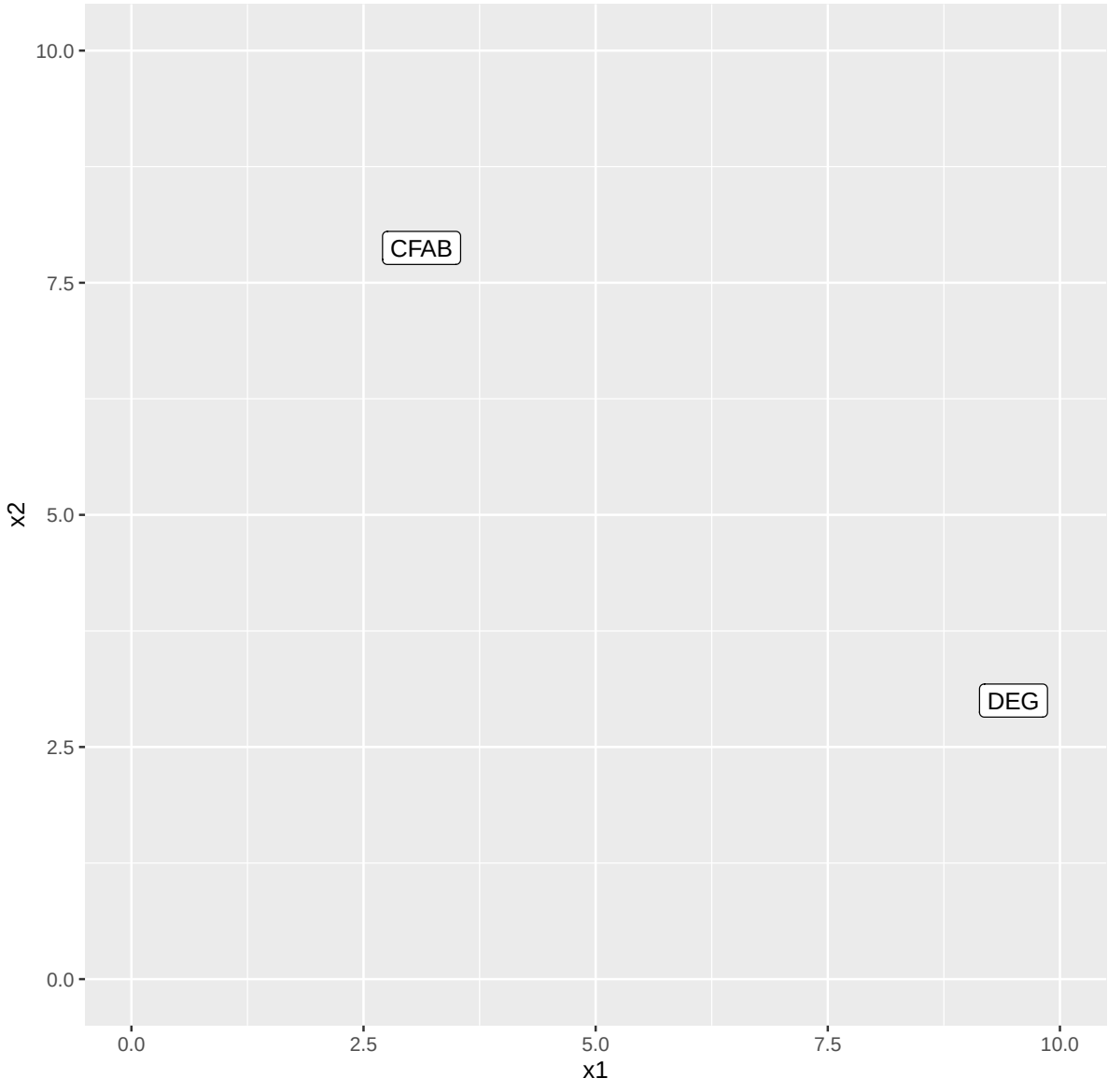
- Distance: 2.24

Identify the pair with the smallest distance.



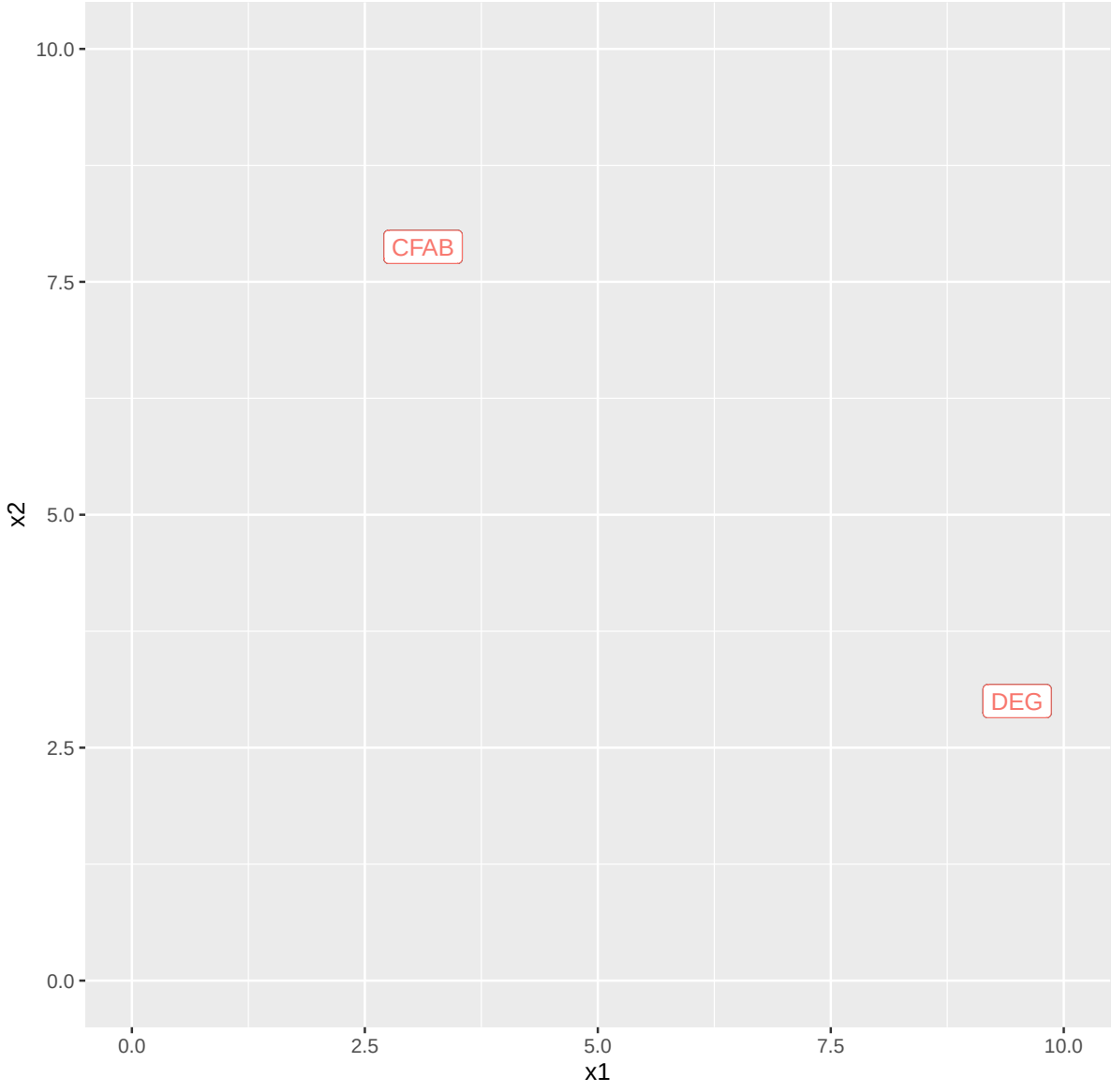
- Distance: 5.06

Group the pair



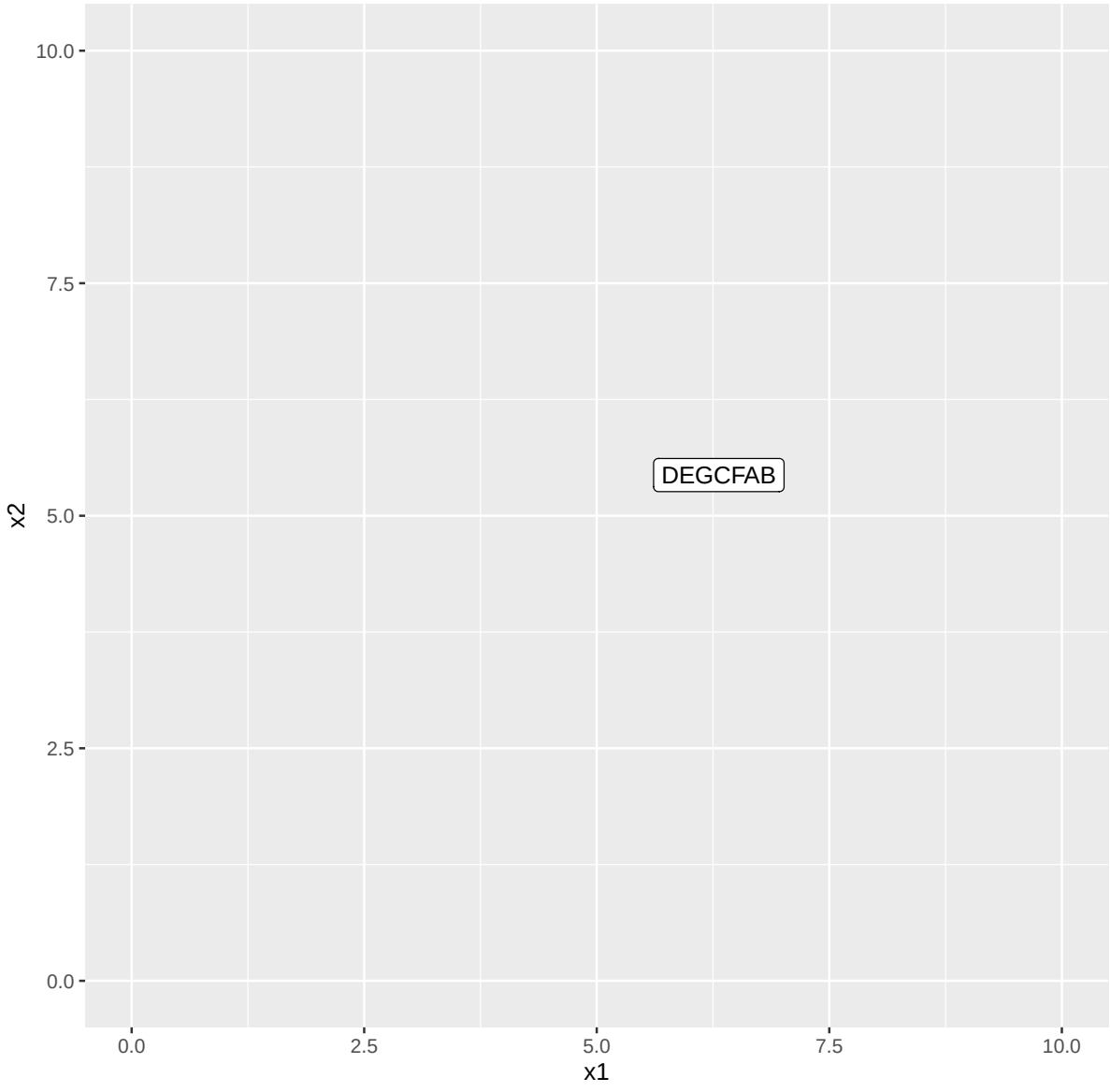
- Distance: 5.06

Identify the pair with the smallest distance.



- Distance: 8.03

Group the pair



- Distance: 8.03

Dendrogram

Linkages

Linkage	*Description*
Complete	Maximal intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the *largest* of these dissimilarities.
Single	Minimal intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the *smallest* of these dissimilarities. Single linkage can result in extended, trailing clusters in which single observations are fused one-at-a-time.
Average	Mean intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the *average* of these dissimilarities.
Centroid	Dissimilarity between the centroid for cluster A (a mean vector of length p) and the centroid for cluster B. Centroid linkage can result in undesirable *inversions*.

Example

You are given the following four pairs of observations: $x_1 = (1,0)$, $x_2 = (1,1)$, $x_3 = (2, 1)$, and $x_4 = (5,10)$.

Calculate the intercluster dissimilarity between x_1, x_2 and x_4 with different linkages and Euclidean distance.