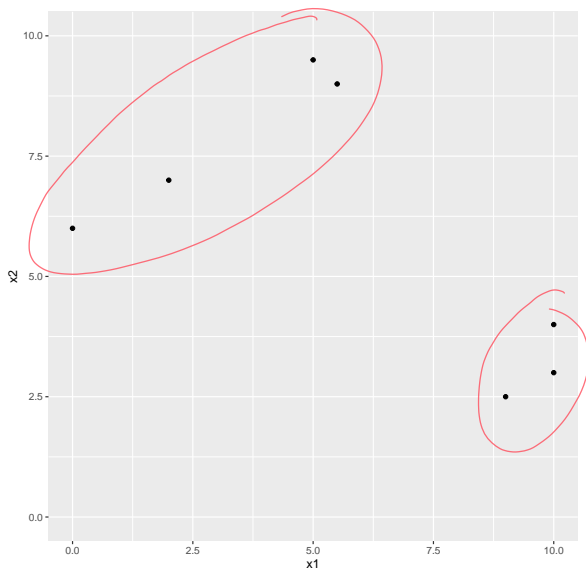
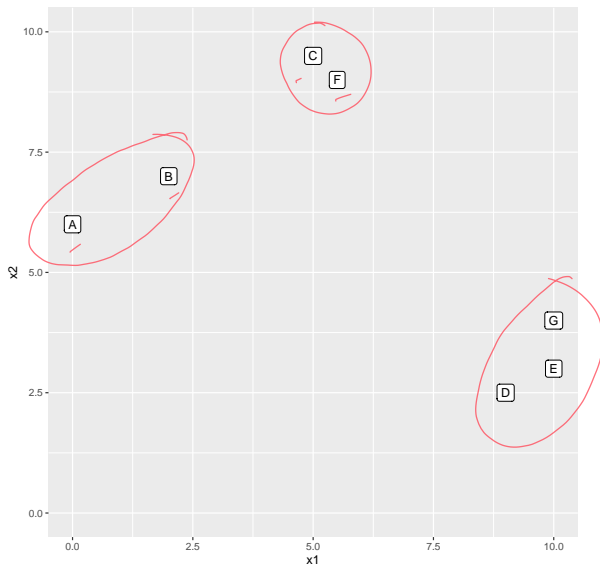


# Hierarchical Clustering

# Hierarchical clustering - Centroid Linkage

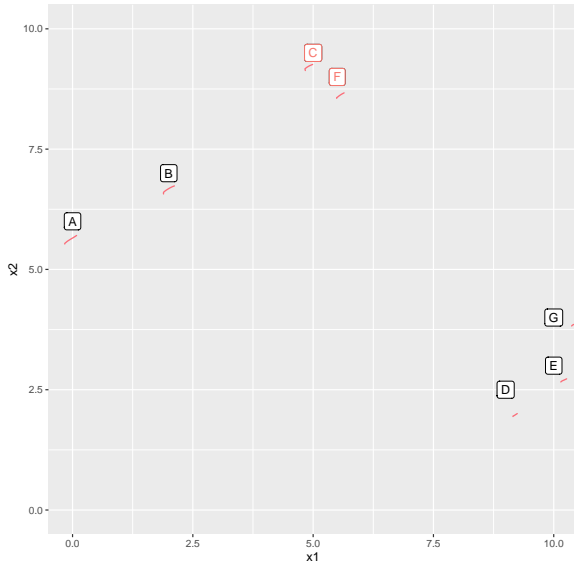


# Label the Points



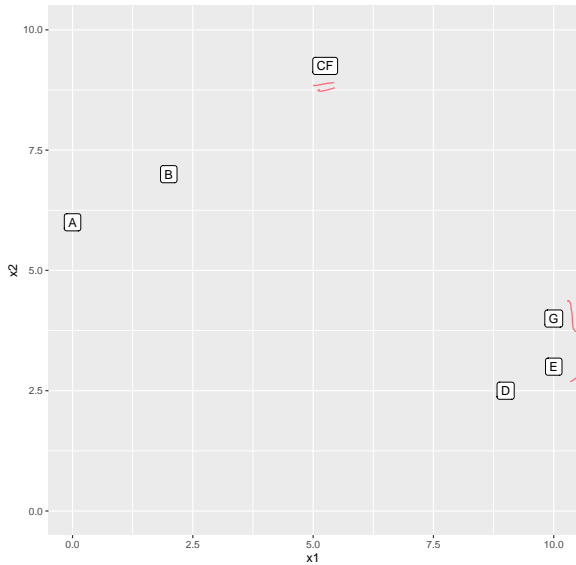
Identify the pair with the smallest distance.

7 clusters

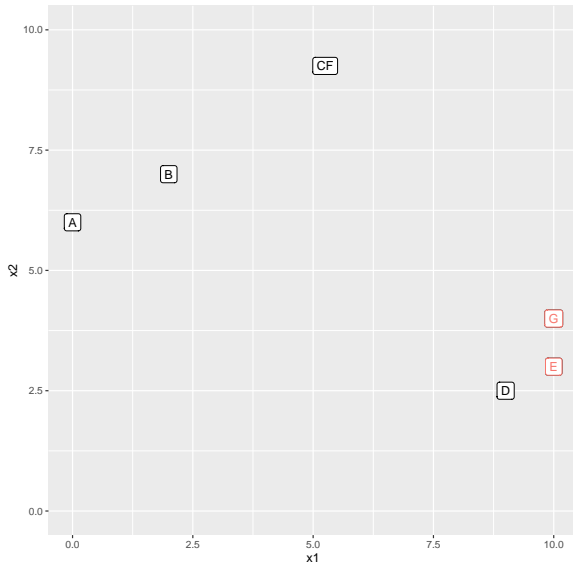


# Group the pair

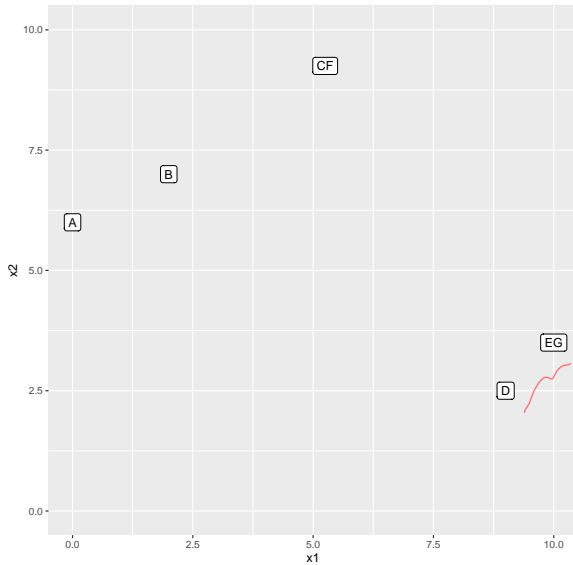
6 clusters



Identify the pair with the smallest distance.

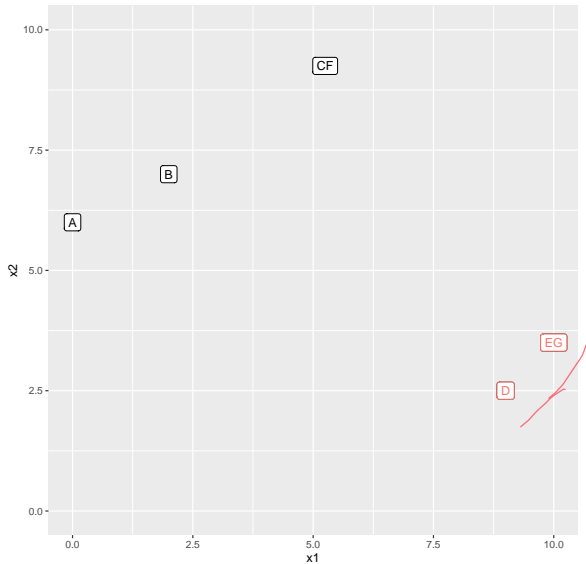


# Group the pair



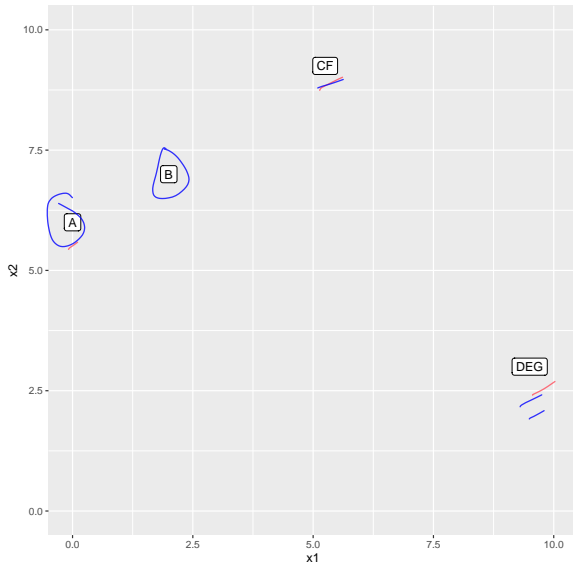
3 clusters

Identify the pair with the smallest distance.



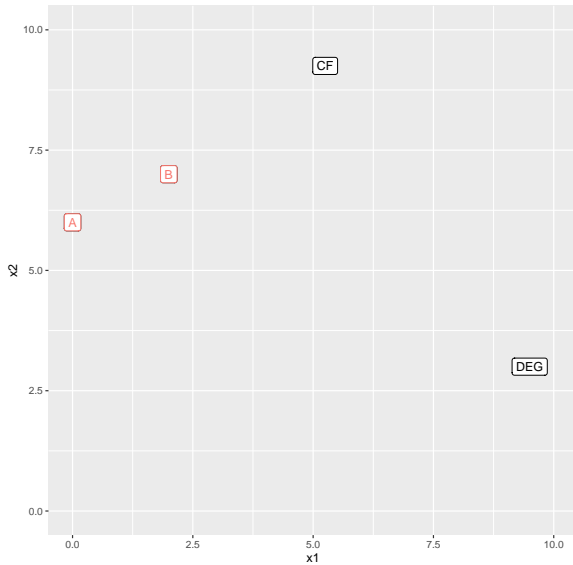


## Group the pair



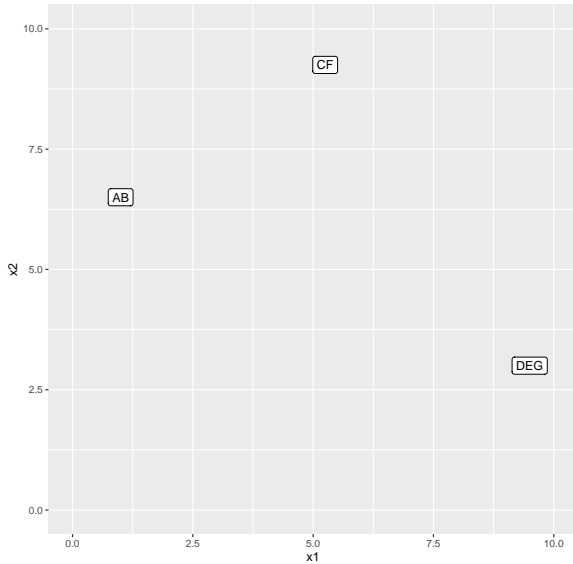
4 clusters

Identify the pair with the smallest distance.

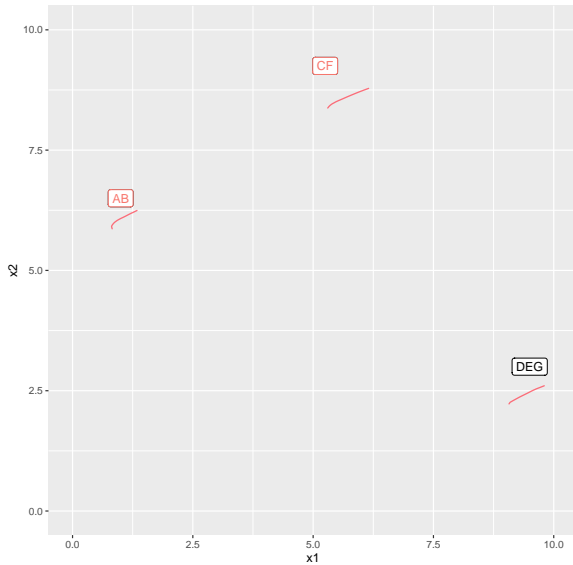


# Group the pair

3 clusters

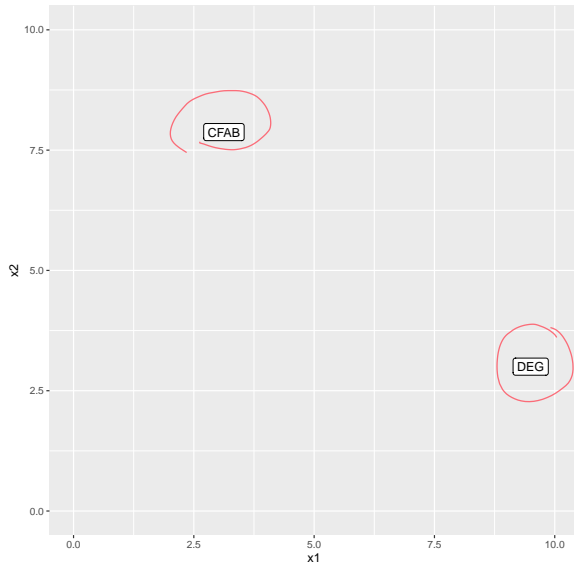


Identify the pair with the smallest distance.

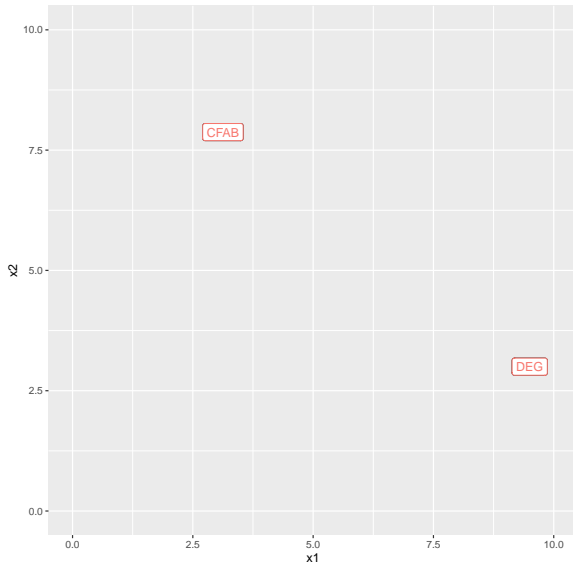


## Group the pair

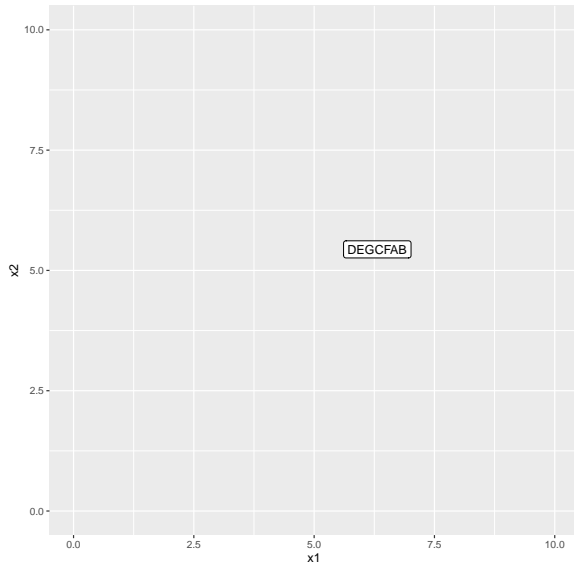
2 clusters



Identify the pair with the smallest distance.



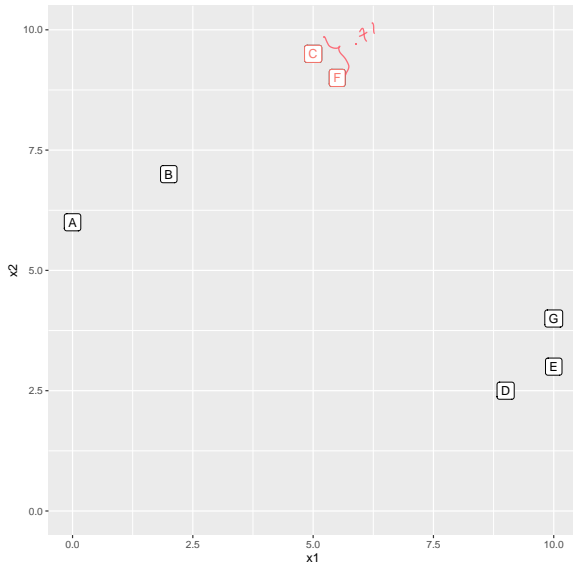
## Group the pair





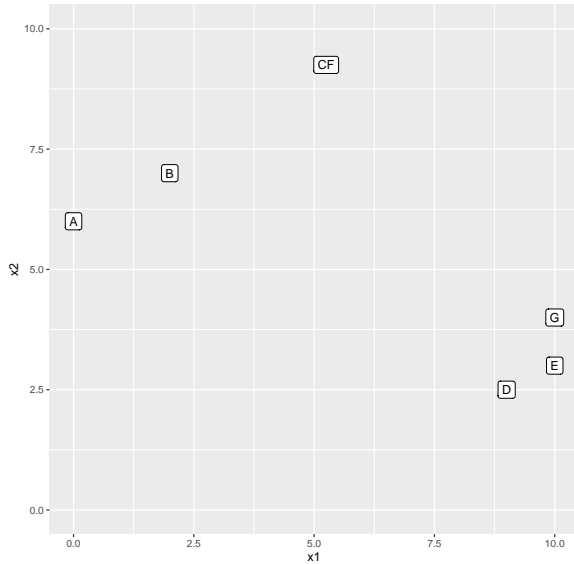


# Dendrogram



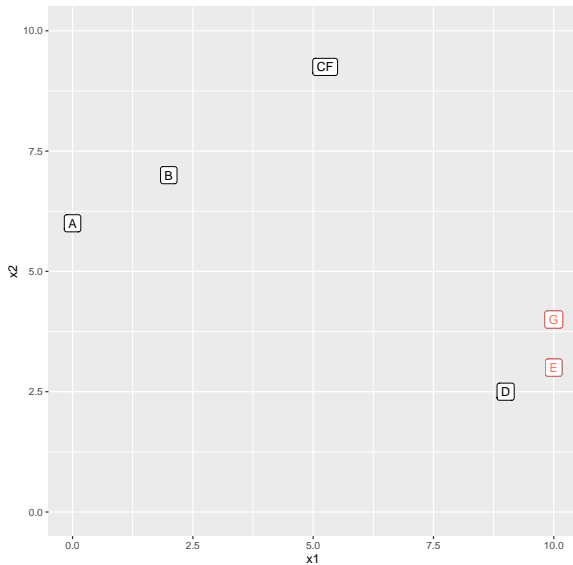
► Distance: 0.71

## Group the pair



► Distance: 0.71

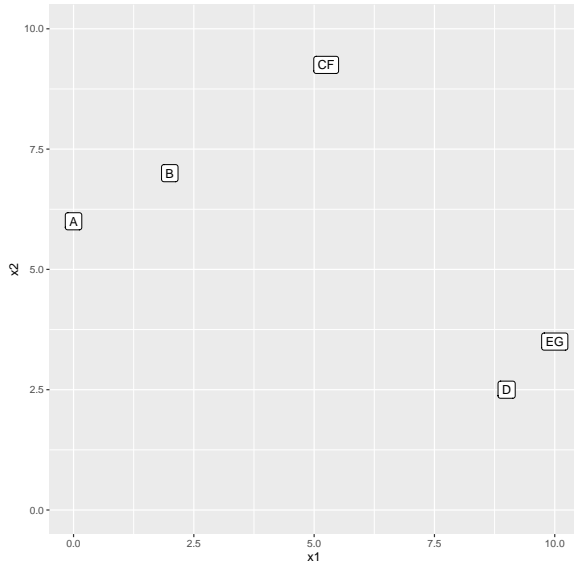
Identify the pair with the smallest distance.



► Distance: 1

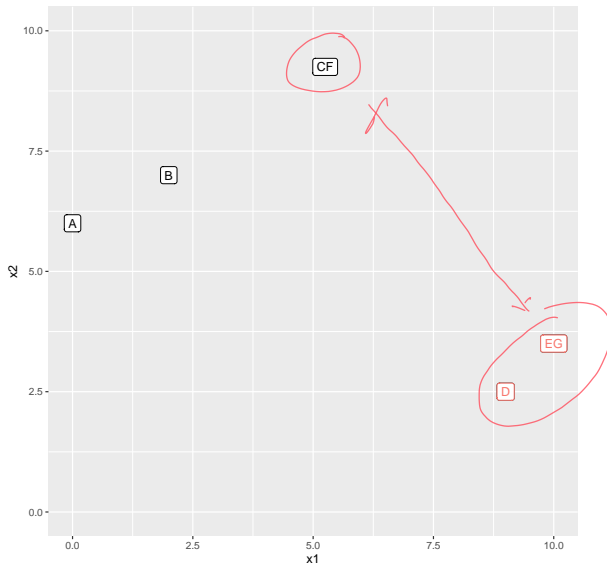


## Group the pair



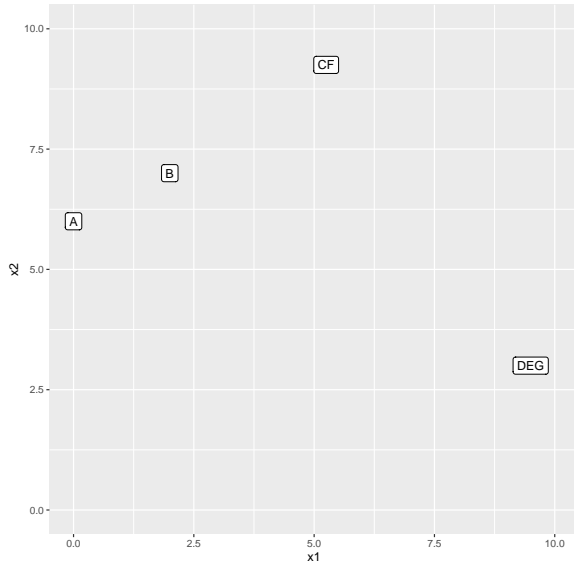
► Distance: 1

Identify the pair with the smallest distance.



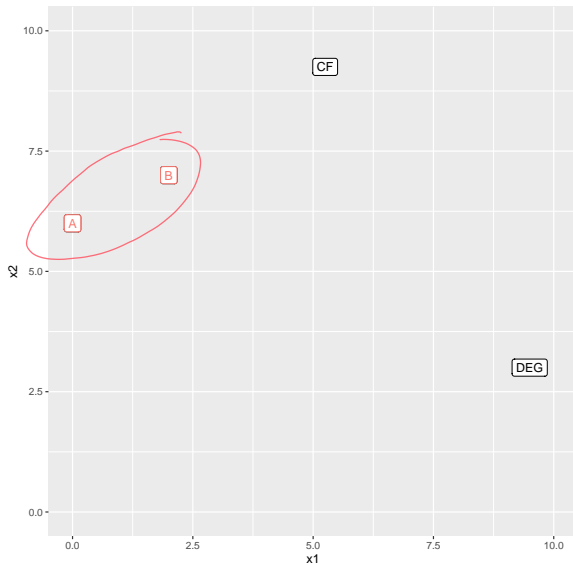
► Distance: 1.41

## Group the pair



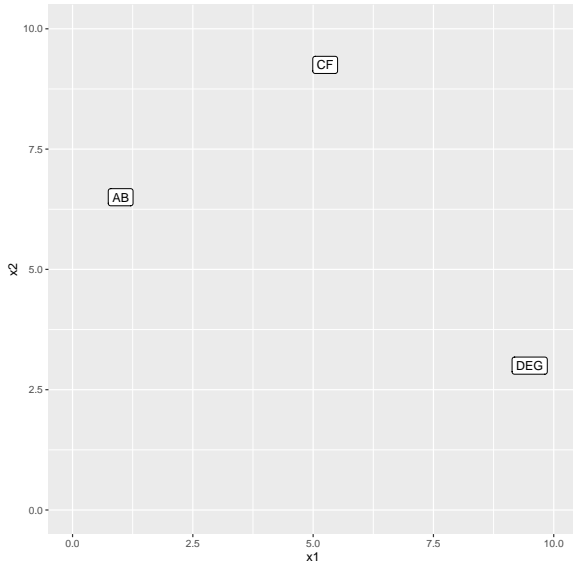
► Distance: 1.41

Identify the pair with the smallest distance.



► Distance: 2.24

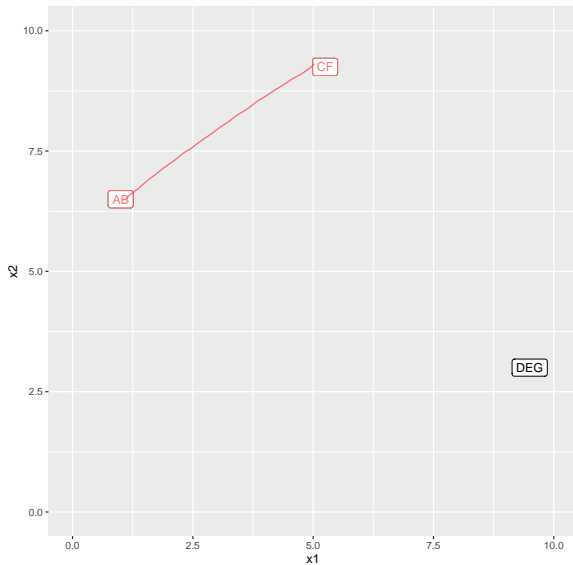
## Group the pair



► Distance: 2.24

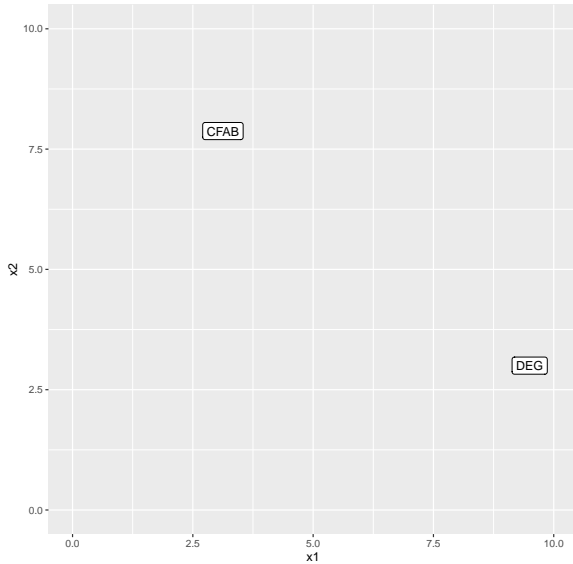


Identify the pair with the smallest distance.



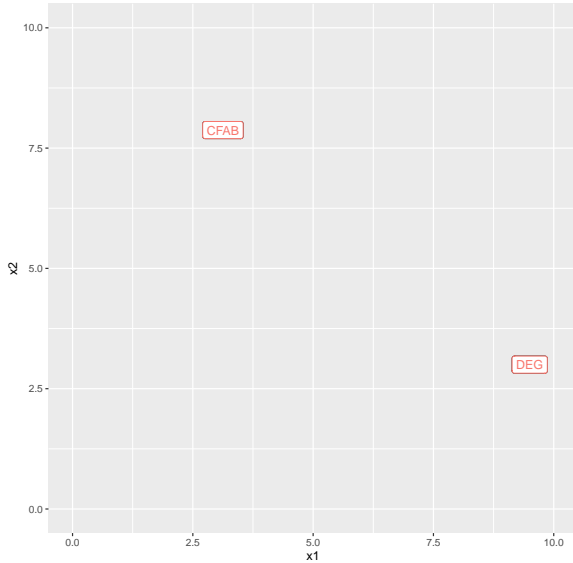
► Distance: 5.06

## Group the pair



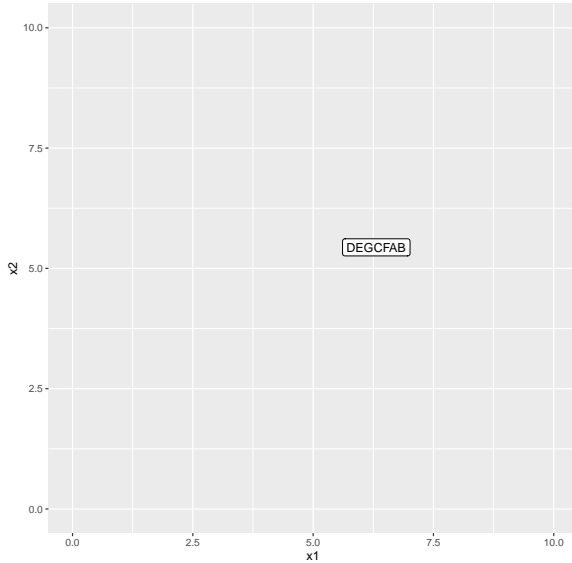
► Distance: 5.06

Identify the pair with the smallest distance.



► Distance: 8.03

## Group the pair



► Distance: 8.03

# Dendrogram

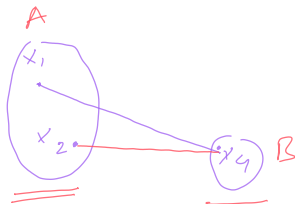
# Linkages

<i>Linkage</i>	<i>Description</i>
Complete	Maximal intercluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the <u>largest</u> 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 <u>smallest</u> of these dissimilarities. Single linkage can result in <u>extended</u> , trailing clusters in which single observations are fused one-at-a-time.
Average	Mean intercluster dissimilarity. Compute all <u>pairwise</u> dissimilarities between the observations in cluster A and the observations in cluster B, and record the <u>average</u> 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 <i>inversions</i> .

## Example

(\*) Complete Linkage

$$\begin{aligned} & \{x_1, x_2\} \quad \{x_4\} \\ x_1 x_4 &= \sqrt{(5-1)^2 + (10-0)^2} = \sqrt{116} \\ x_2 x_4 &= \sqrt{(5-1)^2 + (10-1)^2} = \sqrt{97} \end{aligned}$$



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.

↓  
cluster 1

cluster 2

$$d(\{x_1, x_2\}, \{x_4\}) = \max(\sqrt{116}, \sqrt{97}) = \sqrt{116}$$

(\*) Single linkage:  $d = \min(\sqrt{116}, \sqrt{97}) = \sqrt{97}$

(\*) Average:  $d = \frac{\sqrt{116} + \sqrt{97}}{2}$

Centroid A :  $\left(\frac{1+1}{2}, \frac{1+0}{2}\right) = \left(1, \frac{1}{2}\right)$  M

Centroid B :  $(5, 10)$  N

$$MN = \sqrt{(5-1)^2 + \left(10 - \frac{1}{2}\right)^2} = \dots$$