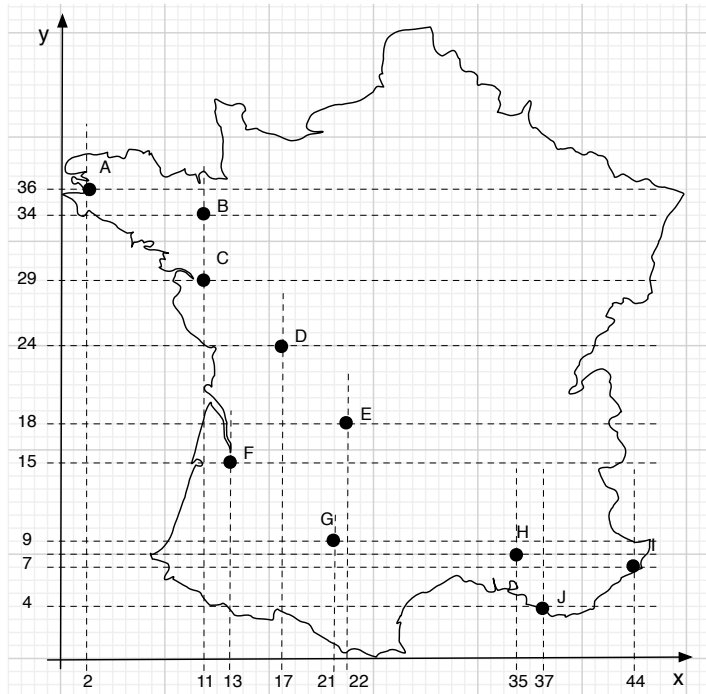


TD1 : Classification hiérarchique ascendante

Données : la carte ci-dessous, les points de A à I représentent des villes. Les axes x et y fournissent les coordonnées des villes sur la carte.



Questions :

- 1- Donner les attributs de cet individu
- 2- Choisir une distance entre individus
- 3- Calculer le centre de gravité g_l et l'inertie I du nuage de points :

$$I = \sum_{i=1}^{i=N} d^2(x_i, g)$$

- 4- Avec Excel faire les étapes de l'algorithme ci-dessous jusqu'à obtenir 3 classes.
On utilisera le saut minimum pour la mise à jour des distances.

Algorithme :

- chaque individu correspond à une classe $g_l = \frac{1}{N_l} \sum_{i \in C_l} x_i$ tous les points (10x10)
- calculer une matrice des distances entre
- chercher la plus petite distance $d(i, j)$ avec $i \neq j$ et $i < j$
- tant que ($d(i, j) < \text{seuil}$) et (le nombre de classes est > 1) faire
 - o fusionner la classe i et la classe j dans la classe i en renommant la colonne i et la ligne i de la matrice des distances en ij
 - o mettre à jour les distances de la ligne i : $d(i, k) = \min(d(i, k), d(j, k))$
 - o mettre à jour les distances de la colonne i : $d(k, i) = \min(d(k, i), d(k, j))$
 - o supprimer la colonne j et la ligne j
 - o chercher la plus petite distance $d(i, j)$ avec $i \neq j$ et $i < j$

Exemple sur un autre jeu de données (http://ouestgenopuces.univ-rennes1.fr/formations/analyse_partie2c.pdf)

Matrices des distances



BA : Bari
FL : Florence
MI : Milan
NA : Naples
RM : Rome
TO : Turin

	BA	FL	MI	NA	RM	TO
BA	-					
FL	662	-				
MI	877	295	-			
NA	255	468	754	-		
RM	412	268	564	219	-	
TO	996	400	138	869	669	-

$$L(MI, TO) = 138$$



	BA	FL	MI/TO	NA	RM
BA	-				
FL	662	-			
MI/TO	877	295	-		
NA	255	468	754	-	
RM	412	268	564	219	-

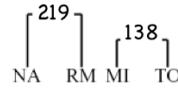


HCL single linkage

	BA	FL	MI/TO	NA	RM
BA	-				
FL	662	-			
MI/TO	877	295	-		
NA	255	468	754	-	
RM	412	268	564	219	-



$$L(NA, RM) = 219$$



$$L(BA, NA/RM) = 255$$



	BA	FL	MI/TO	NA/RM
BA	-			
FL	662	-		
MI/TO	877	295	-	
NA/RM	255	268	564	-

HCL single linkage



$$L(BA/NA/RM, FL) = 268$$



	BA/NA/RM	FL	MI/TO
BA/NA/RM	-		
FL	268	-	
MI/TO	564	295	-

	BA/NA/RM/FL	MI/TO
BA/NA/RM/FL	-	
MI/TO	295	-



$$L(BA/NA/RM/FL, MI/TO) = 295$$

