**TP 2**

**Analyse factorielle d’un tableau de distances,**

**classification automatique**

**Exercice 1. Analyse factorielle d’un tableau de distances**

Cet exercice a pour but de nous faire comprendre la méthode de l'AFTD en l'utilisant sur des données suivantes:

1. On centre en colonne la matrice X puis on utilise la fonction "dist()" qui calcule le tableau D des distances euclidiennes associé aux données.

On utilise la fonction "as.matrix()" sur le tableau de distances pour le mettre sous forme de matrice.

On multiplie cette matrice D pour avoir la matrice des distance euclidienne D2

2. On calcul W de deux manières différentes, la première, à partir de la formule:

Cette manière suppose d'avoir les données sous forme de matrice comme c'est le cas dans cette exercice. En générale on réalise l'AFTD car on ne dispose qu'un tableau de distances, dans ce cas, on utilise la formule: avec

3. On diagonalise la matrice W avec la fonction "eigen()" , les valeurs propres sont:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| 1.39e+01 | 2.21e-01 | 1.75e-15 | 1.054e-15 | 7.51e-17 | 6.16e-18 | -3.46e-17 | -4.35e-17 |

Les valeurs propres 7 et 8 sont négatives, mais elle sont proches de zéro (10^(-17)) et l'on peut donc les considérer comme nulle. Les valeurs propres deviennent donc:

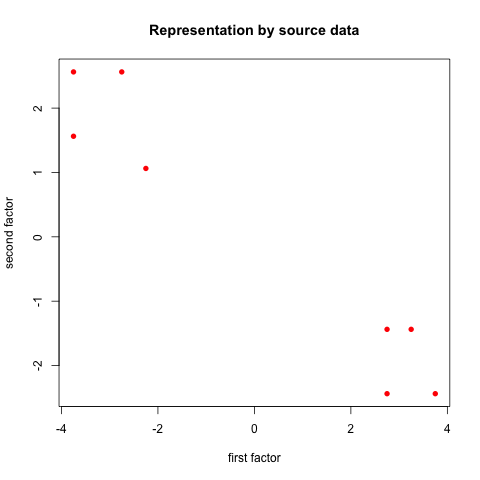
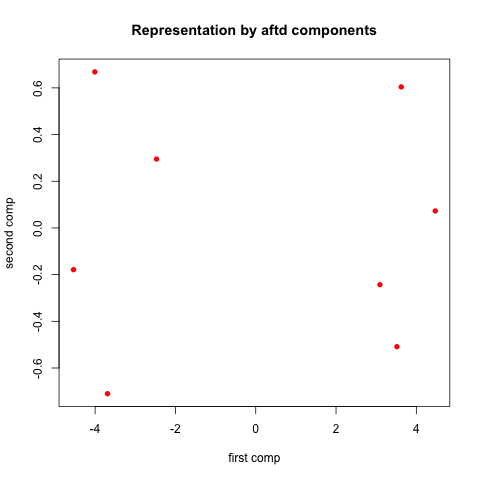
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| 1.39e+01 | 2.21e-01 | 1.75e-15 | 1.054e-15 | 7.51e-17 | 6.16e-18 | 0 | 0 |

On considère donc la matrice comme semi défini positive.

4. La matrice des vecteur propres et la matrice diagonale des valeurs propres sont donnée par la fonction "eigen()" avec les 2 dernières valeurs propres à 0.

5.

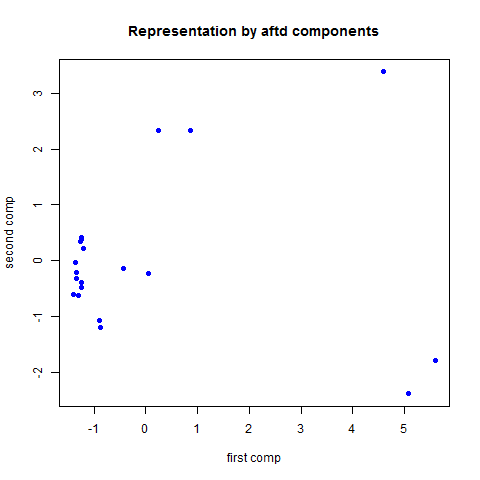
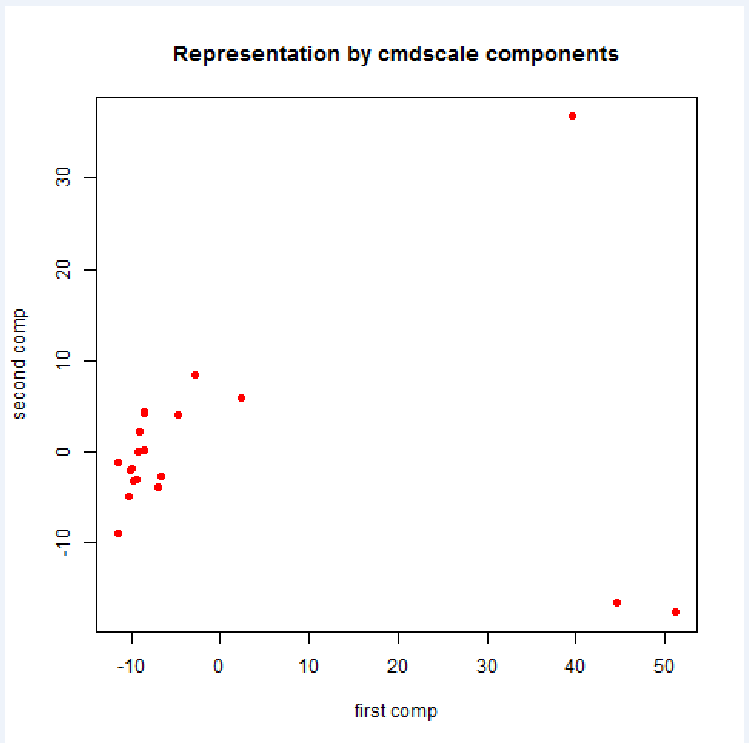
6

. 

On distingue 2 groupe sur les 2 graphiques. Sur La représentation associé au tableau initiale, les points appartenant a un même groupe sont beaucoup plus proche, ce qui donne l'aspect d'avoir deux "paquet" sur le graphique. Il semble que les axes factorielles ne soient pas les même sur les 2 graphiques, c'est pour cela, que les représentation sont différentes.

7.

Données mutations:

1. Les résultats sont quasiment similaire à un facteur d'échelle près. Certains points ne sont pas aux même endroit mais on retrouve les même groupes de points sur les 2 graphiques.

2.