

Classification Automatique

TP 4 : CAH sous R

Un compte rendu est à rendre à la fin du TP. Il devra être constitué des programmes R et d'un rapport au format PDF. Les TP sont à réaliser en binôme. **La qualité de la rédaction (programmes et rapport) sera prise en compte.** Les programmes et le rapport devront porter les noms des binômes et être envoyés par mail à gwandji@polytech-lille.fr.

1. Charger les données du fichier `hotel.txt`, identifier les différentes variables du fichier et leur nature.
2. Donner les différentes statistiques descriptives et liaisons entre les variables quantitatives du fichier.
3. En utilisant l'aide de R, décrire les fonctions `hclust` et `rect.hclust`.
4. Faire la classification hiérarchique ascendante des observations en utilisant les distances euclidienne et Manhattan et les liaisons simples, complète et de Ward.
 - Identifier le nombre de classe optimal.
 - Représenter les clusters retenus sur un dendrogramme.
 - Commenter les résultats obtenus.
5. Reprendre la question précédente sur les variables.