

---

**Objectif :** L'objectif de ce travail est la mise en pratique concrète d'un certain nombre de techniques d'analyse de données quantitatives, à travers l'étude d'un cas pratique nécessitant l'utilisation de logiciels de traitement statistique de données tel que R.

---

## Sujet

---

**Description des données :** Bases-Partiel contient différents ensembles de données avec une classification connue. Les ensembles de données sont assez simples et de faible dimension. Les algorithmes de classification, tels que la Classification Ascendante Hiérarchique et les K-moyennes mobiles (Kmeans) ne sont pas toujours en mesure de regrouper tous les ensembles de données correctement. Ces bases de données peuvent servir d'évaluation de ces algorithmes et de tests pour de nouveaux algorithmes de classification où l'enjeu est de classer au mieux les données. La description des différentes bases de données est donnée dans la table 1 ci-après :

**Questions :** A partir des fichiers suivants

data set	# objects	# variables	# clusters
Hepta	212	3	7
Lsun	400	2	3
Tetra	400	3	4
Chainlink	1000	3	2
Atom	800	3	2
EngyTime	4096	2	2
Target	770	2	6
TwoDiamonds	800	2	2
WingNut	1070	2	2
GolfBall	4002	3	1

TABLE 1 – Description des données : toutes les bases sont de format ASCII text : \*.lrn la matrice des données et \*.cls le vecteur des labels (les classes).

1. Visualiser le nuage des individus un plan.
2. Appliquer la CAH en utilisant le lien minimum et le critère de Ward.
3. Visualiser l'arbre hiérarchique obtenu par chaque méthode. Que peut-on dire ?
4. Appliquer Kmeans sur le tableau initial.
5. Visualiser les classes sur le premier plan factoriel.
6. Utiliser les boxplots pour interpréter les classes. Que peut-on dire ?
7. Comparer les résultats obtenus par les CAH et Kmeans.

### Consignes importantes

1. Ce projet est à réaliser par binôme, ou seul.
2. Deux projets qui se ressemblent seront lourdement sanctionnés.
3. Ne pas se contenter de lister des graphiques et des tableaux sans commentaires.
4. Un rapport de 15 pages au maximum au format pdf (incluant le code R). Le rapport doit être soigné et clair.
5. Lors de votre envoi de votre rapport, mettre comme objet "Projet L3" dans votre message à l'adresse mohamed.nadif@parisdescartes.fr. Date limite de l'envoi du projet 24 avril à 18 :00.
6. Toute initiative dans ce projet sera appréciée.