

# TRAVAUX PRATIQUE SUR LA CLASSIFICATION NON SUPERVISEE: LA CAH et ACM

Ici, nous allons appliquer la **cah** sur un ensemble de données dans l'optique de les classer. Nous utiliserons le langage R et ainsi le logiciel R ou R studio. Pour des personnes ne pouvant installer ces environnements, pas de panique! Plusieurs plateformes en ligne permettent d'exécuter du code R. Dans le cadre de ce Tp, nous ferons usage de la plateforme Datacamp.

## 1. Creation du compte DataCamp et mise en place de l'environnement

- Allez à l'adresse [https://www.datacamp.com/users/sign\\_up](https://www.datacamp.com/users/sign_up) et créez votre compte.

Ceux ayant des compte google pourront également en faire l'usage.

- En suite allez à l'adresse <https://app.datacamp.com/workspace/dashboard>
- Cliquez sur **Create Workspace**, puis donner un nom et choisissez comme langage **R**.
- Validez et l'environnement est configuré.

## 2. Partitionnement avec la CAH

Tout d'abord, nous disposons d'un ensemble d'un échantillon de **2000** personnes récolté en l'année 2003 pensant une étude en France. Il est disponible à l'adresse <https://juba.github.io/questionr/reference/hdv2003.html>. Chacune des personnes est caractérisée par 20 attribut en occurrence le **sexe**, l'**age** et le **niveau d'étude (nivetud)**, ...

L'on désire dans le cadre de ce Tp, savoir comment classer ces différentes personnes. Ce qui pourrait permettre de faire une corrélation entre l'age, le sexe et le niveau d'étude, etc de ces personnes.

Dans la suite nous allons suivre le notebook que nous avons rédigé à l'effet.

<https://app.datacamp.com/workspace/w/2b59fb08-8f8e-47a5-895e-a712efba34bb>