

# Science des données II : module 5



Analyse des données

Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique  
Laboratoire d'Écologie numérique



<https://wp.sciviews.org>  
[sdd@sciviews.org](mailto:sdd@sciviews.org)

## Reprise...

C'est la reprise... au programme, l'AA consacrée à l'**analyse des données**.



- Informations sur les mises-à-jour logicielles et matérielles
- Réinstallation de l'application **BioDataScience** indispensable !
- Annotations dans le cours
- Déroulement de la semaine

L'analyse des données concerne l'exploration de gros jeux de données multivariés. L'information est trop volumineuse pour être digérée telle quelle. L'analyse des données vise à **synthétiser** cette information et en **extraire les caractéristiques intéressantes**.

- Les techniques de **regroupement** ou de **classification** remplacent un grand nombre d'individus par des groupes ou classes,
- Les techniques d'**ordination** et de **réduction des dimensions** synthétisent les variables en un plus petit nombre et représentent les données sur des graphiques (cartes) plus lisibles.
- Ces techniques sont **exploratoires**. Elles sont utilisées très tôt dans l'analyse, juste après, voir pendant la phase de description des données.
- Elles sont généralement suivies par de l'inférence (SDD I, partie 2) ou de la modélisation (SDD II, partie 1)

Cette semaine : module 5, classification non supervisée avec la CAH