

Science des données I : module 1



Introduction et approche pédagogique

Philippe Grosjean, Guyliann Engels, Maxime Bodson & Anne-Charlotte Vranckx

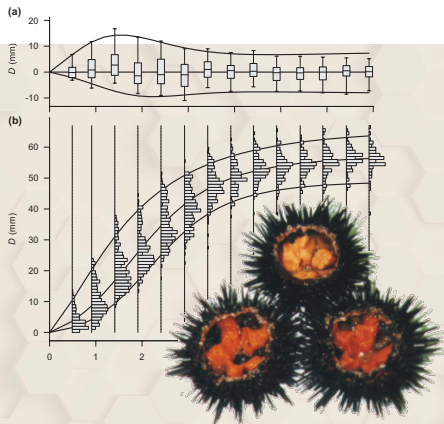
Université de Mons, Belgique
Laboratoire d'Écologie numérique



<https://wp.sciviews.org>
sdd@sciviews.org

Qui sommes-nous ?

Prof. Philippe Grosjean



- **Bioingénieur** + thèse de doctorat en biologie marine (croissance d'oursins)
- Capacités supplémentaires développées en **science des données** durant des post-docs et via de la consultance pendant 4 ans partout en Europe
- **Laboratoire EcoNum** créé en 2004 à l'Université de Mons
- Intéressé par des travaux **interdisciplinaires** : biologie, chimie, modélisation, statistiques, informatique

Guyliann Engels

- **Master** en Biologie des Organismes et Écologie à l'UMONS.
- **Mémoire** effectué dans le laboratoire d'Écologie numérique des Milieux aquatiques sur l'écophysiologie et l'écotoxicologie de la posidonie (*Posidonia oceanica*, une plante marine) en Méditerranée.
- **Thèse de doctorat** en cours sur le plancton dans le même laboratoire.
- **Assistant** en biologie à l'UMONS depuis septembre 2017.



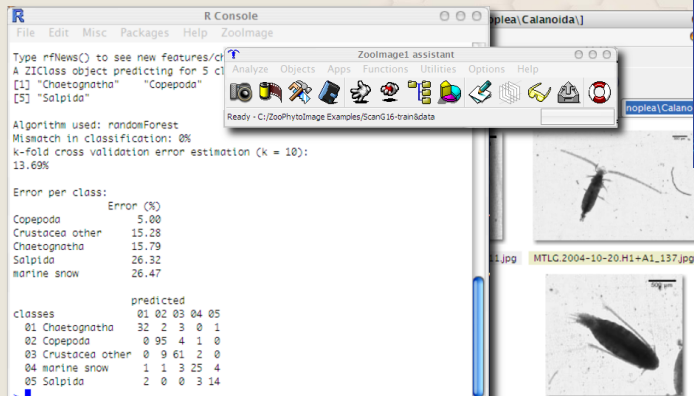
Les “nouveaux”

- Maxime Bodson : assistant pédagogique
- Anne-Charlotte Vranckx : élève assistante actuellement en Bab2 biologie

Recherche : identification automatisée du plancton

Le plancton (constitué des organismes aquatiques qui dérivent en pleine eau) forme des communautés très diversifiées. Un litre d'eau de mer contient typiquement des milliers d'espèces de plancton.

Au laboratoire EcoNum, nous développons des outils pour énumérer automatiquement plancton via l'analyse d'image combinée à la classification supervisée (une technique statistique que nous étudierons en Master 1).



The screenshot displays the R Console window with the following output:

```

Type rfNews() to see new features/changes
A ZIClass object predicting for 5 classes
[1] "Chaetognatha"    "Copepoda"
[5] "Salpida"

Algorithm used: randomForest
Mismatch in classification: 0%
k-fold cross validation error estimation (k = 10):
13.69%

Error per class:
              Error (%)
Copepoda      5.00
Crustacea other 15.28
Chaetognatha  15.79
Salpida       26.32
marine snow   26.47

predicted
classes 01 02 03 04 05
01 Chaetognatha 32  2  3  0  1
02 Copepoda     0 95  4  1  0
03 Crustacea other 0 96  2  0
04 marine snow  1  1  3 25  4
05 Salpida      2  0  0  3 14
  
```

Overlaid on the R Console is the ZoomImage assistant window, which includes a toolbar with icons for Analyze, Objects, Apps, Functions, Utilities, Options, and Help. Below the toolbar, it shows a file path: Ready - C:\ZooPhytoImage Examples\ScanG 16-brain\data. To the right of the R Console, there are two small images of plankton: a copepod and a chaetognath.



Qu'est-ce que la science des données ?

Science des données : une approche pragmatique

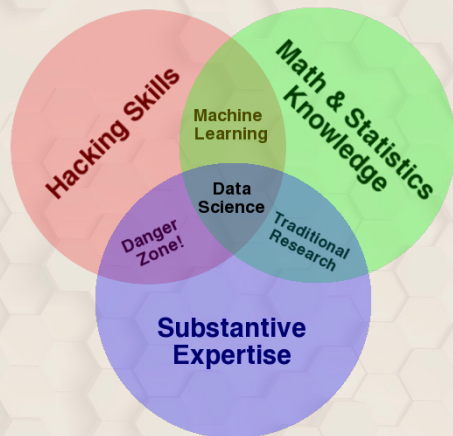
A data scientist is a
statistician who is useful.
— *Hadley Wickham*

JS

Science des données : à l'interface entre plusieurs disciplines

- La Science des Données, c'est la discipline qui s'intéresse à l'analyse de données *sous toutes ses formes*
- Très large et **interdisciplinaire** :
 - (Bio)statistiques et visualisation
 - Utilisation d'outils informatiques
 - Expertise dans le domaine (biologie)
- Il faut maîtriser simultanément les 3 domaines pour être un scientifique des données.

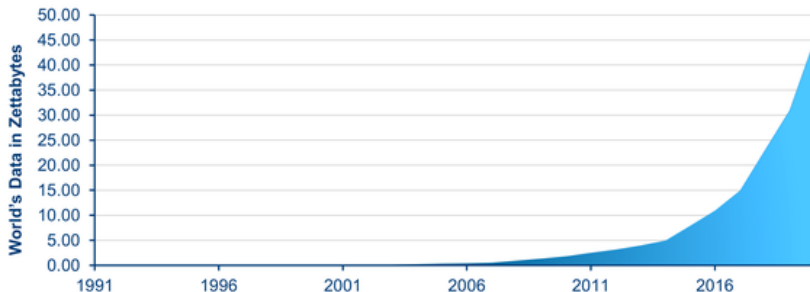
C'est notre objectif durant votre formation "science des données biologiques" qui s'étalera sur 4 année.



Pourquoi la science des données ?

- Discipline à la fois ancienne et **récente**
 - Évolution des statistiques, avec ses prémices dans les années 1960 (John Tukey).
 - Émerge comme science à part : 2001 William S. Cleveland, *"Data Science : An Action Plan for Expanding the Technical Area of field of Statistics"*.
 - Le terme **Data Scientist** n'est d'usage courant que depuis 2008.
- Besoin issu de la **quantité de données** disponibles (1 zettabyte = 1 milliard de terabytes = 1 000 000 000 000 000 000 000 octets).

Data growth



La science de données biologiques

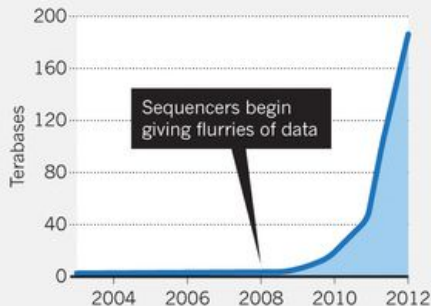
La biologie n'échappe pas au besoin d'analyser des (gros) jeux de données :

- **Génétique**, bases immenses
- **Biodiversité** animale et végétale
- **Etudes écologiques** avec images satellites, capteurs haute vitesse
- **Littérature** scientifique
- etc.

Un biologiste analyse des données pratiquement quotidiennement sous une forme ou l'autre !

DATA EXPLOSION

The amount of genetic sequencing data stored at the European Bioinformatics Institute takes less than a year to double in size.



Approche pédagogique

Structuration en module

Un module vise au développement spécifique **d'une ou plusieurs compétences**

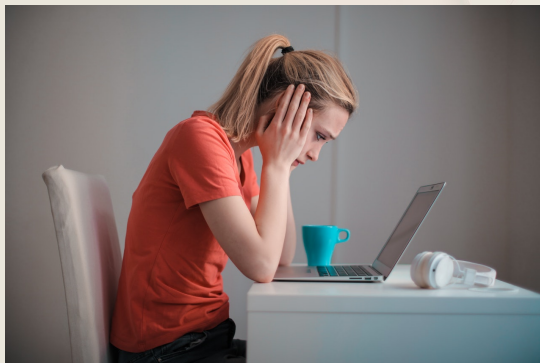
Le travail du module est concentré sur **une semaine**.

- Six modules au Q1 (**AA visualisation**)
- Six modules au Q2 (**AA inférence**)

Exercices et projets cotés en évaluation continue, pas d'examen à la fin.
présence aux séances obligatoire. (voir plan de cours)

De quoi allons-nous discuter ?

... de **pédagogie**. Comment apprendre ?



À vous la parole, quel est le meilleur moyen d'apprendre ?

Et la classe inversée ?

Vous avez une minute ?

Pour comprendre
La classe inversée

0:03 / 1:18



(lien vers la vidéo)

Classe inversée et pédagogie active

Notre approche : **pédagogie active en classe inversée** (vous apprenez *d'abord* à la maison, nous appliquons *ensuite* en présentiel -quand on n'est pas confinés-).

I hear and I forget.

I see and I remember.

I do and I understand.

— Confucius

C'est quoi la pédagogie active ?

Les
pédagogies
actives
pourquoi ne
pas essayer?



0:00 / 3:05



(lien vers la vidéo)

UMONS

Et moi, je fais quoi dans tout cela ?

Lisez ceci... et réagissez (question Wooclap juste après) !



- Vous êtes **acteur de votre apprentissage**, les enseignants sont des **facilitateurs** (plus en retrait par rapport à l'approche classique).
- Plus de séparation entre **cours théorique** et **exercices** ; vos échanges avec le professeur et le ou les assistants sont similaires.
- Les **élèves-assistants** sont coachés tout autant que vous pour vous faciliter l'apprentissage de manière active.
- **Vous posez les questions**, et vos enseignants vous répondent **individuellement**.

ECTS

European Credits Transfer System, créé en 1988 de manière standardisée par la Commission Européenne comme correspondant à une **charge de travail totale** pour l'étudiant de **25 à 30 heures**.



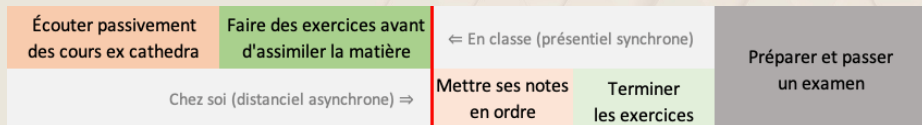
Optimisation du temps de travail

Comment voulez-vous passer vos 25-30h/ECTS ?

Note : 12 modules pour 6 ECTS dans notre cours, donc 1/2 ECTS par module.

Optimisation du temps de travail à l'Université

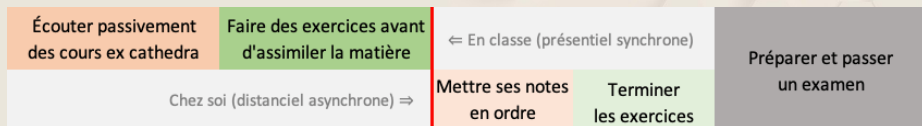
Cours classique *ex cathedra* + séances d'exercices



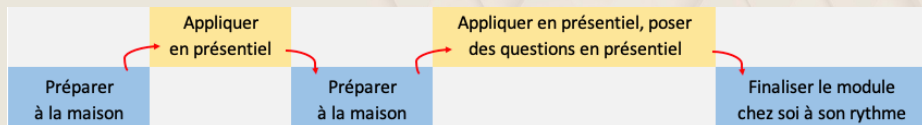
- Le réel apprentissage se déroule **après** les séances de cours et d'exercices
- Un examen est nécessaire pour vérifier vos acquis

Optimisation du temps de travail... comparé à la classe inversée

Cours classique *ex cathedra* + séances d'exercices



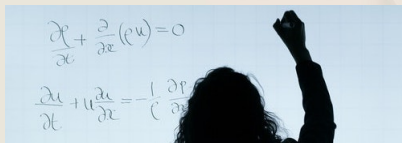
Approche en classe inversée



- Aucune séance en présentiel sans préparation
- Chaque heure de travail pleinement consacrée à l'apprentissage
- Vous êtes actifs **tout le temps** et vous gérez à **votre rythme**
- **Pas besoin d'un examen à la fin** : travail évalué dans sa globalité

Le professeur est un coach et un facilitateur

Mais que font les enseignants alors ? Lisez et réagissez (question Wooclap après).



- Le professeur (et l'assistant) ne mettent **pas** leur savoir en avant. C'est vous qui construisez votre *propre* savoir.
- Ils **ne répondent pas directement** à vos questions : ils vous mettent sur une piste et vous font réfléchir pour trouver la réponse *par vous-même*.
- Ils se mettent en retrait, mais sont **disponibles pour vous aider** (Discord, mail, etc.)

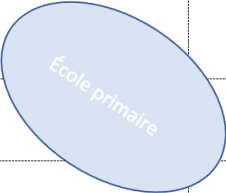
C'est déroutant car le professeur n'apparaît plus comme l'omniscient qui transmet de manière unilatérale son savoir aux étudiants !

Objectifs pour l'acquisition de compétences

Processus cognitif	Types de connaissances			
	A. factuelles	B. conceptuelles	C. procédurales	D. métacognitives
1. Mémoriser				
2. Comprendre				
3. Appliquer				
4. Analyser				
5. Évaluer				
6. Créer				

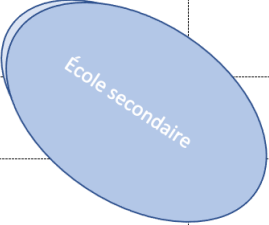
Taxonomie de Bloom modifiée par Krathwohl & Anderson.

Objectifs pour l'acquisition de compétences

Processus cognitif	Types de connaissances			
	A. factuelles	B. conceptuelles	C. procédurales	D. métacognitives
1. Mémoriser				
2. Comprendre				
3. Appliquer				
4. Analyser				
5. Évaluer				
6. Créer				

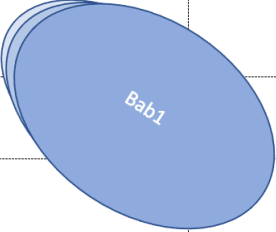
Taxonomie de Bloom modifiée par Krathwohl & Anderson.

Objectifs pour l'acquisition de compétences

Processus cognitif	Types de connaissances			
	A. factuelles	B. conceptuelles	C. procédurales	D. métacognitives
1. Mémoriser	 École secondaire			
2. Comprendre				
3. Appliquer				
4. Analyser				
5. Évaluer				
6. Créer				

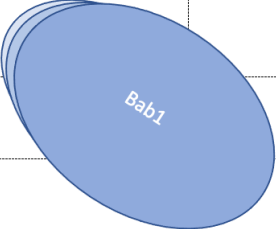

Taxonomie de Bloom modifiée par Krathwohl & Anderson.

Objectifs pour l'acquisition de compétences

Processus cognitif	Types de connaissances			
	A. factuelles	B. conceptuelles	C. procédurales	D. métacognitives
1. Mémoriser				
2. Comprendre				
3. Appliquer				
4. Analyser				
5. Évaluer				
6. Créer				

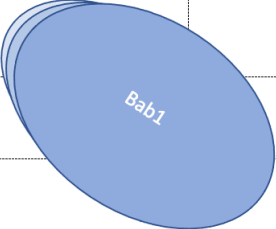
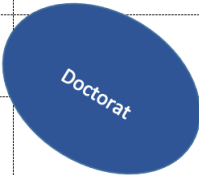
Taxonomie de Bloom modifiée par Krathwohl & Anderson.

Objectifs pour l'acquisition de compétences

Processus cognitif	Types de connaissances			
	A. factuelles	B. conceptuelles	C. procédurales	D. métacognitives
1. Mémoriser				
2. Comprendre				
3. Appliquer				
4. Analyser				
5. Évaluer				
6. Créer				

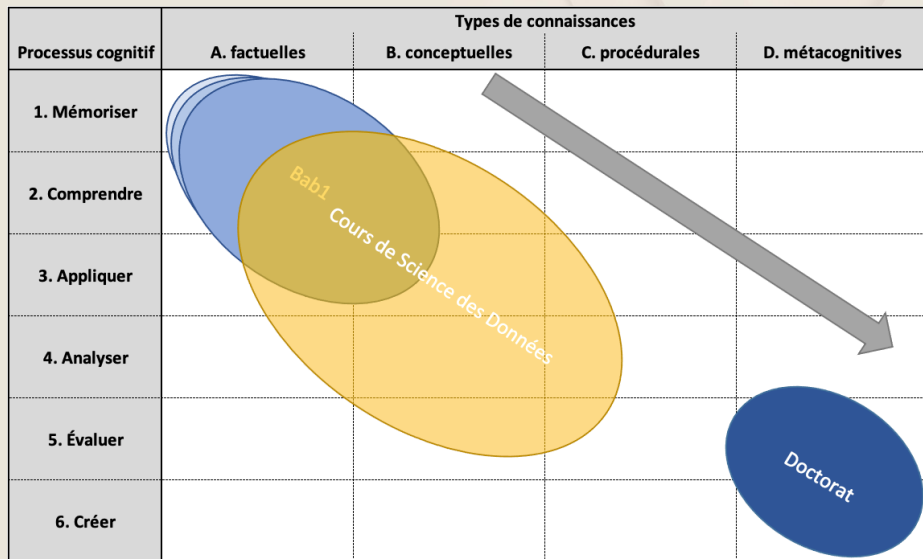
Taxonomie de Bloom modifiée par Krathwohl & Anderson.

Objectifs pour l'acquisition de compétences

Processus cognitif	Types de connaissances			
	A. factuelles	B. conceptuelles	C. procédurales	D. métacognitives
1. Mémoriser				
2. Comprendre				
3. Appliquer				
4. Analyser				
5. Évaluer				
6. Créer				

Taxonomie de Bloom modifiée par Krathwohl & Anderson.

Objectifs pour l'acquisition de compétences



Taxonomie de Bloom modifiée par Krathwohl & Anderson.

Au travail !

Objectif du module 1

Se familiariser avec l'ordinateur et les logiciels que nous utiliserons dans le cadre du cours : R, RStudio, Markdown, git & GitHub !



Préparation

- Préparer son matériel (ordinateur, écouteurs, câble réseau, ...)
- Créer un **compte GitHub**
- **S'enregistrer correctement** dans le site du cours
- Entraidez-vous, si nécessaire !

