

# Apprentissage de la science des données, six ans d'innovation pédagogique, et ensuite ?



Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique  
Service d'Ecologie numérique  
<[philippe.grosjean@umons.ac.be](mailto:philippe.grosjean@umons.ac.be), [phgrosjean@sciviews.org](mailto:phgrosjean@sciviews.org)>  
<[guyliann.engels@umons.ac.be](mailto:guyliann.engels@umons.ac.be)>

Bilan avec le SAP, juillet 2024

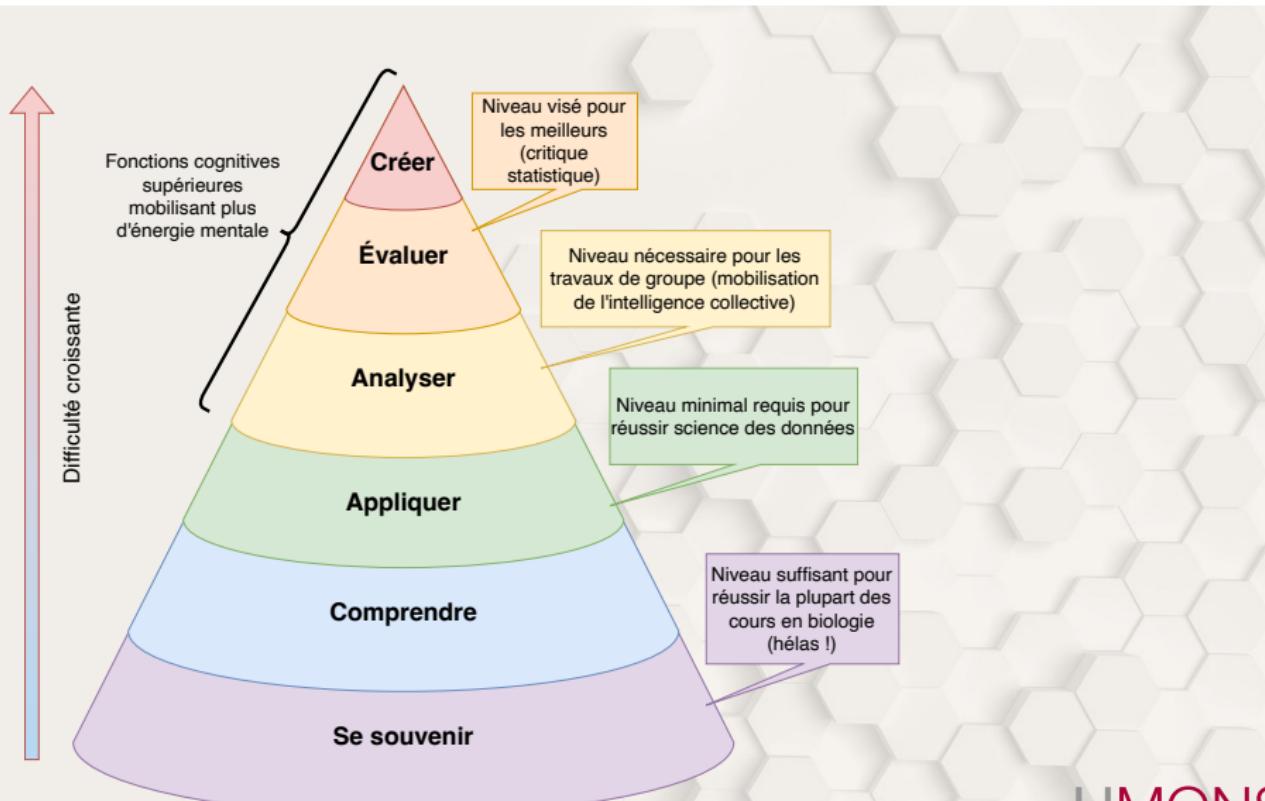


# Objectifs pédagogiques principaux des cours de SDD

- **Connaître** des outils informatiques et statistiques utiles pour manipuler des données biologiques
- **Comprendre** la logique des analyses statistiques
- **Appliquer** ces analyses à de nouvelles données biologiques
- **Analyser et interpréter** les résultats de ces analyses (en groupes)
- En bonus, pour les meilleurs : **évaluer** de manière critique les conclusions d'une analyse statistique

*Ceci correspond aux niveaux 1 à 4 (ou à 5) de la taxonomie de Bloom révisée par Anderson et Krathwohl (2001). Seule le dernier niveau (créer) n'est pas explicitement visé dans les trois premiers cours (mais abordé dans le quatrième cours à option).*

# Objectifs pédagogiques (par rapport à Anderson & Krathwohl 2001)



## Subsection 1

Sondage de data scientists avec Wooclap

UMONS

# Question : mode d'apprentissage préféré

## 2. Vous avez une demi-journée de libre que vous décidez de consacrer à R, ...

98 répondants

Vous lisez 50 pages de "An Introduction to R" (manuel officiel de R).



5 votes

Vous suivez un tutoriel, un podcast ou une vidéo sur un sujet R qui vous intéresse.



33 votes

Vous tentez de résoudre un problème pratique avec R (analyse d'un de vos jeux de données).



60 votes

# Question ouverte

3. Comment auriez-vous aimé aborder l'apprentissage de R  
(30 caractères max) ?

54 répondants

En pratiquant sur mes données  
 Sur des cas pratiques Avec une problématique concret Un tutoriel type OpenClassroom  
 avec des projets concrets En jeu en travaillant sur mes données Via un jeu vidéo R moderne Petit projet de zéro  
 Flammekueche Les mains dans le cambouis Sur une problème pratique Sans les stats  
 avec tidyverse Exercices pratiques Mins en faisant du julia  
 tutoriel Cas pratique Sur l'enclume Exercices  
 R pour la statistique avec des cas pratiques Tidyverse MarkdownTuto Learn TRUE == FALSE Tutoriels  
 Exercice pratique Avec un expert sur des données et en pratique Cas pratiques  
 Par des mini challenges pratique Je lis le blog de Bruno Projet déjà fait exercice dataviz  
 Avec des cas réels Travaux pratiques Une suite d'exercice progressif Projet tidyverse  
 Vrais jeux de donnees Des formations avec pratiques Au travers d'un projet  
 En suivant le cours à Mons 😊

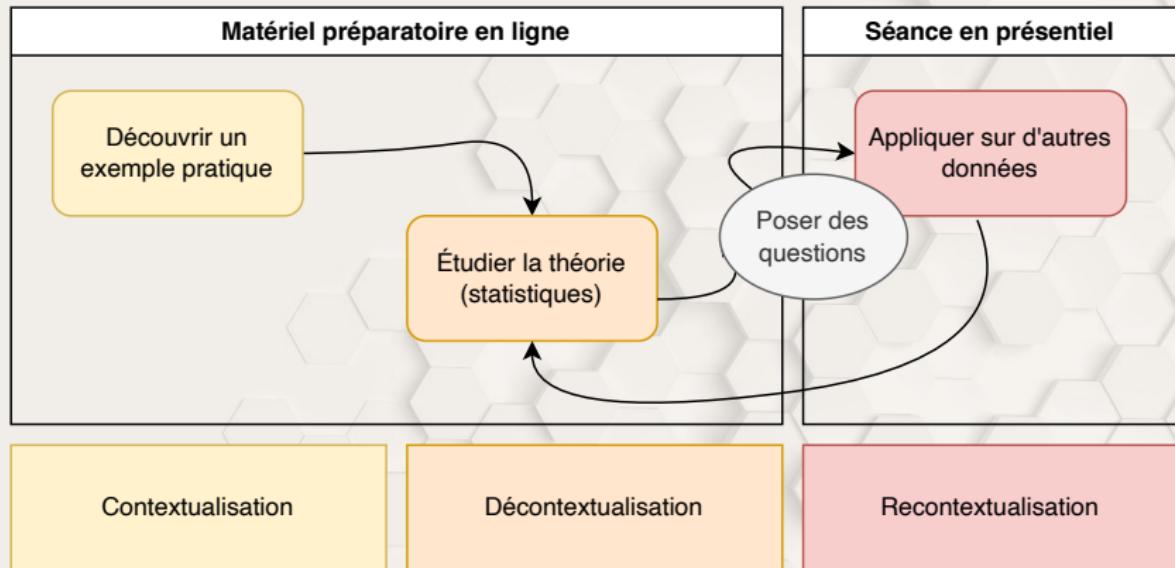
# Conséquences : axer un maximum sur la pratique !

- Apprentissage en classe inversée (*un cours purement théorique avant des exercices pratiques ne sert à rien ; les étudiants n'enregistrent rien*)
- **Exercices en ligne** pour l'autoévaluation
- Projets avec analyse de données biologiques réelles comme **activité principale en présentiel**

*Ceci a nécessité la réécriture complète du matériel pédagogique de nos cours de science des données : PowerPoints utilisés durant les cours théoriques remplacés par du matériel d'auto-apprentissage en classe inversée.*

# Contextualiser - décontextualiser - recontextualiser

Exemple pratique qui introduit un concept quasi-systématiquement dans le cours en ligne. Recontextualisation en séance.



# Matériel pédagogique en ligne

Cours et exercices à l'adresse : <https://wp.sciviews.org>

Plus de **900 pages** de cours, plus de **600 exercices** en ligne, **40 projets** individuels cadrés autocorrigés, **évaluation sur plus de 360 critères** pour chaque étudiant, LRS enregistrant plus de **800.000 évènements** annuellement (et 5 ans de travail !)

The screenshot shows the homepage of the "Science des données biologiques" website. At the top, there is a navigation bar with the University of Mons logo, the title "Science des données biologiques", and links for Moodle, Discord, Github, E-mail, and RStudio. On the left, a sidebar lists "Accueil", "Contact", "Contenu des cours" (with links to Bab2, Bab3, and Ma1&2), and "Anciennes versions". The main content area features a large image of a woman sitting at a desk with a laptop, looking down at it. The text "Bien débuter..." is displayed above the image. Below the image, a caption reads: "Avant d'utiliser le matériel didactique lié à votre cours de Science des".

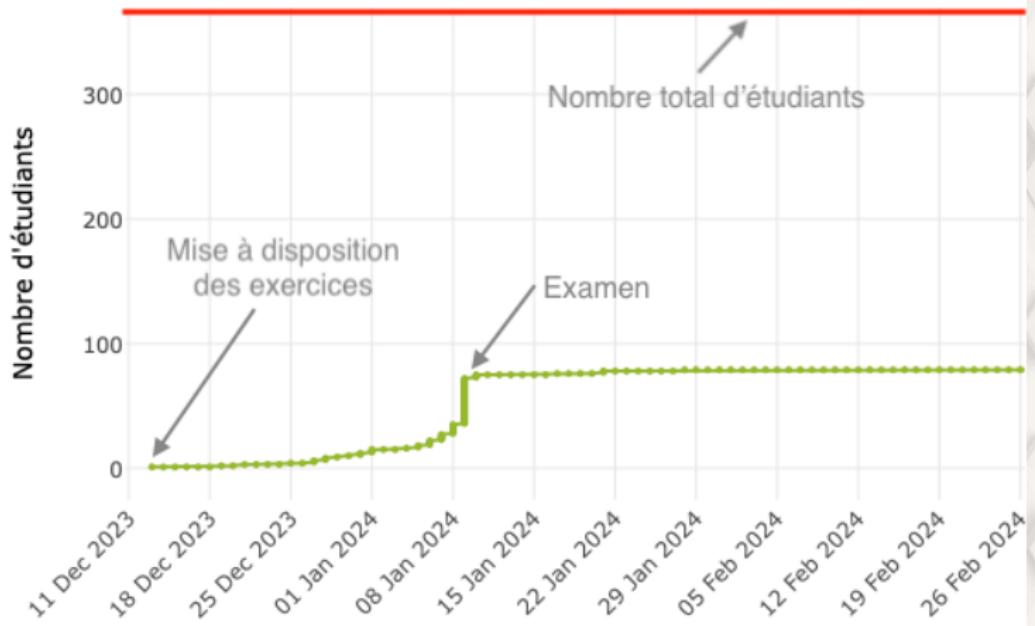
## Subsection 2

### Motiver et capter l'attention

# Exercices facultatifs - peu de participation

Cours de math, exercices interactifs en ligne proposés sans précautions particulières

## Test de révision sur les Nombres Complexes



UMONS

## Participation - des solutions...

Varier le type d'exercices : H5P

- Plus de 50 types d'exercices différents, voir <https://h5p.org>

Construisez une instruction R qui reprend seulement la première occurrence de chaque valeur du vecteur  $x <- c("chat", "chat", "chien", "chat", "cheval", "chien")$  et place le résultat dans  $x2$ .

&lt;-

[

(x)]

!

x2

x

duplicated

VérifierThe logo of the Université catholique de Louvain (UCLouvain) is displayed, consisting of the letters 'UCL' in a stylized font followed by 'LOUVAIN' in a smaller, bold, serif font.

# Participation - des solutions...

## Varier le type d'exercices : H5P + learnr

- Tutoriels avec écriture de code R, voir <https://rstudio.github.io/learnr/>

### duplicated()

Philippe Grosjean & Guyliann Engels

2021-10-07

Utilisation simple de duplicated()

Utilisation de duplicated() sur un vecteur

Utilisation de duplicated() sur un tableau

Utilisation créative de duplicated()

### Utilisation simple de `duplicated()`

Voici un vecteur `v` :

```
print(v)
```

```
## [1] "chien"   "chat"    "chien"   "chat"    "chat"    "cheval"  "chien"
```

Indiquez à quelle position se situe la première occurrence de chaque mot dans `v` en retournant un vecteur `v1` de même taille contenant des valeurs logiques (`TRUE` ou `FALSE`). Utilisez `duplicated()` pour y arriver.

R Code

Start Over

Hints

Run Code

Submit Answer

```
1 v1 <- ____()
2 v1
3
```

[Start Over](#)

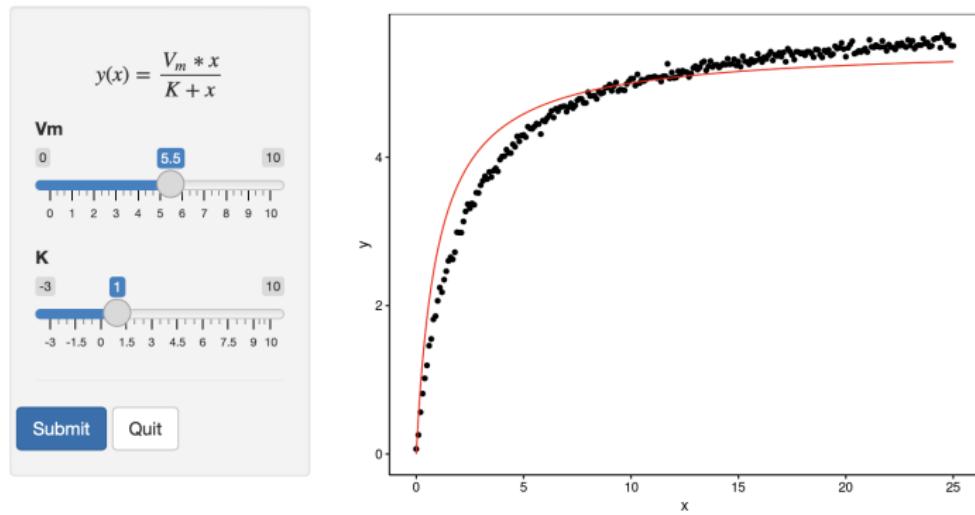
[Next Topic](#)

# Participation - des solutions...

Varier le type d'exercices : H5P + learnr + Shiny

- Démonstration de concepts statistiques avec <https://shiny.rstudio.com/>

Ajustement manuel d'un modèle : Michaelis-Menten



Modèle paramétré :

$$y(x) = \frac{5.50 * x}{1.00 + x}$$

Somme des carrés des résidus  
(valeur à minimiser) :

22.29

# Participation - des solutions...

## ■ Exercices directement dans le cours en ligne

À vous de jouer !

H-P

Qualifiez la situation suivante : le dépistage d'une maladie donne un résultat positif sur un patient, alors qu'en réalité, ce patient n'est pas malade.

Il s'agit d'un vrai positif.

✗ Il s'agit d'un faux négatif

Un faux négatif est un test négatif alors que le patient est malade.

Il s'agit d'un vrai négatif.

Il s'agit d'un faux positif.



0/1



Afficher la solution



Recommencer

UMONS

# Participation - des solutions...

- Exercices directement dans le cours en ligne
- **Liste des exercices** à la fin de chaque module

## 3.6 Récapitulatif des exercices

Ce module 3 vous a permis de réaliser différents graphiques uni- et bivariés afin de visualiser la *distribution* de variables quantitatives seules ou en fonction des niveaux d'une variable qualitative (facteur). Pour évaluer votre compréhension de cette matière, vous aviez les exercices suivants à réaliser :

- **H-P** Les fonctions `chart()` et `geom_histogram()`
- **H-P** Modes et symétries
- Nombre de classes d'un histogramme
- Graphiques univariés
- **H-P** La fonction `chart()` et `geom_density()`
- Graphiques de distribution des données
- **H-P** La fonction `chart()` et `geom_violin()`
- **H-P** Deux versions d'un même fichier
- **H-P** Résolution d'un conflit
- **H-P** Résolution d'un conflit (suite)
- Analyse de données (partie II)

## Participation - des solutions...

- Exercices directement dans le cours en ligne
- Liste des exercices à la fin de chaque module
- **Points attribués à la réalisation des exercices** (exemple, 6% de la note finale)

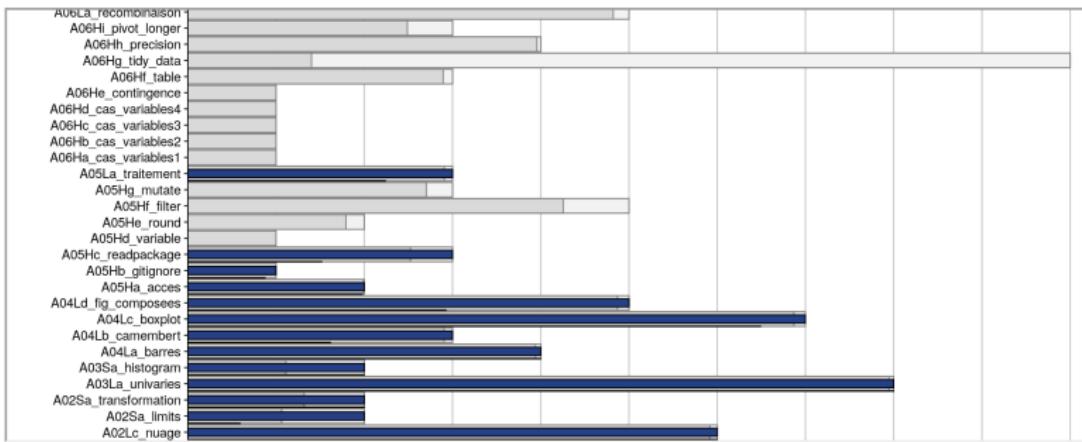


**UMONS**

# Participation - des solutions...

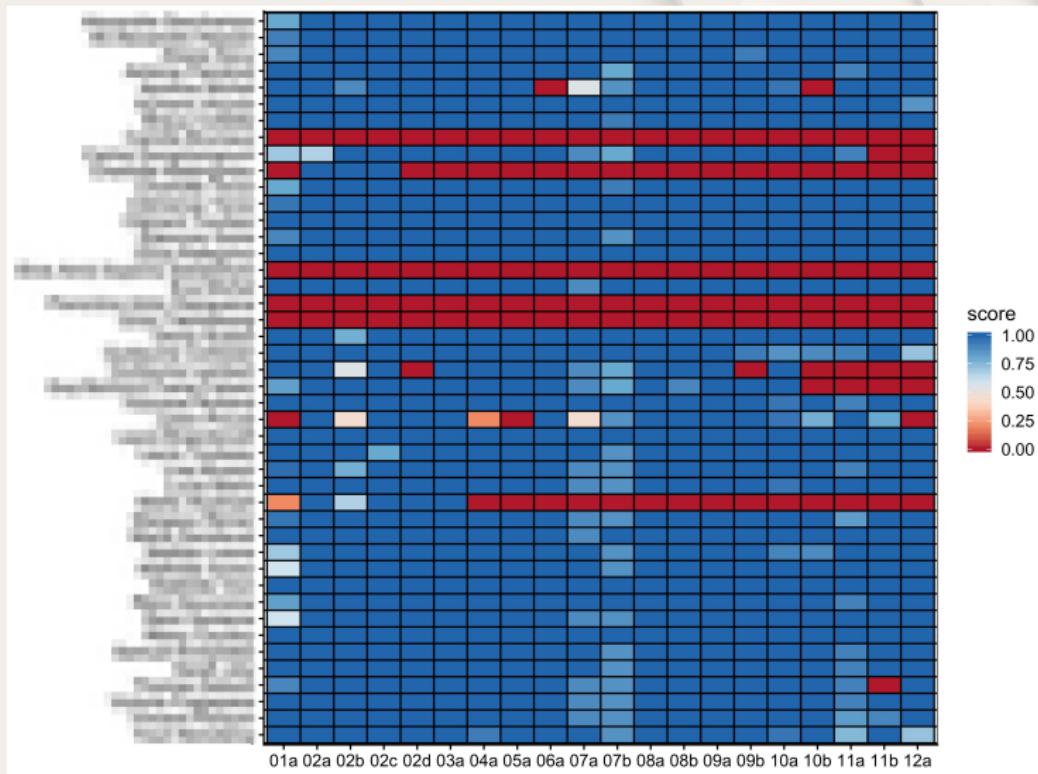
- Exercices directement dans le cours en ligne
- Liste des exercices à la fin de chaque module
- Points attribués à la réalisation des exercices (exemple, 6% de la note finale)
- **Rapport de progression en temps réel**

## Progression



## Participation - résultat

Plus de 90% de participation observée aux exercices en ligne de nos cours

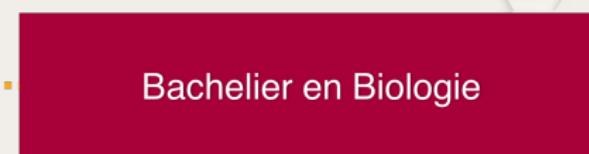


### Subsection 3

## Progressivité de l'apprentissage

# Progressivité : formation sur 4 années (200h présentiel, 500h total)

- **SDD I** en Bab2 : 10 modules, 70h présentiel (avec remédiation), 7 ects
- **SDD II** en Bab3 : 10 modules, 60h présentiel, 6 ects
- **SDD III** en Ma1 : 5 modules, 30h présentiel, 3 ects
- **SDD IV (option)** en Ma2 : 5 modules, 30h présentiel, 3 ects



I  
Ias

II  
Ias



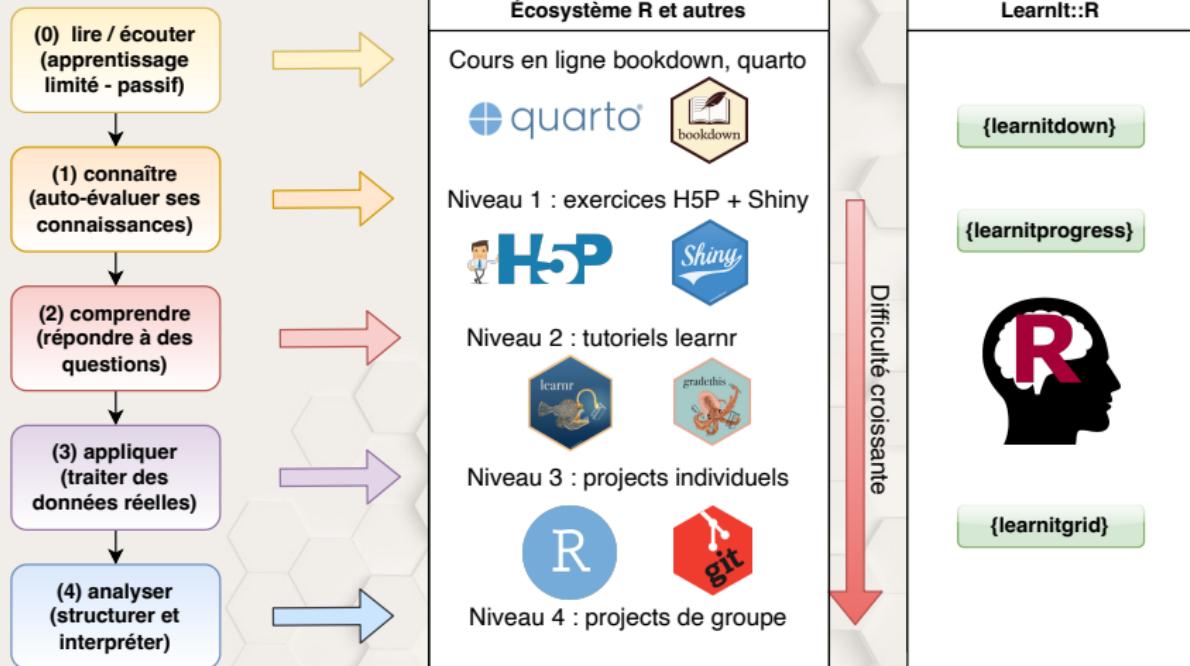
III  
Ias

IV  
Ias

Milieu professionnel

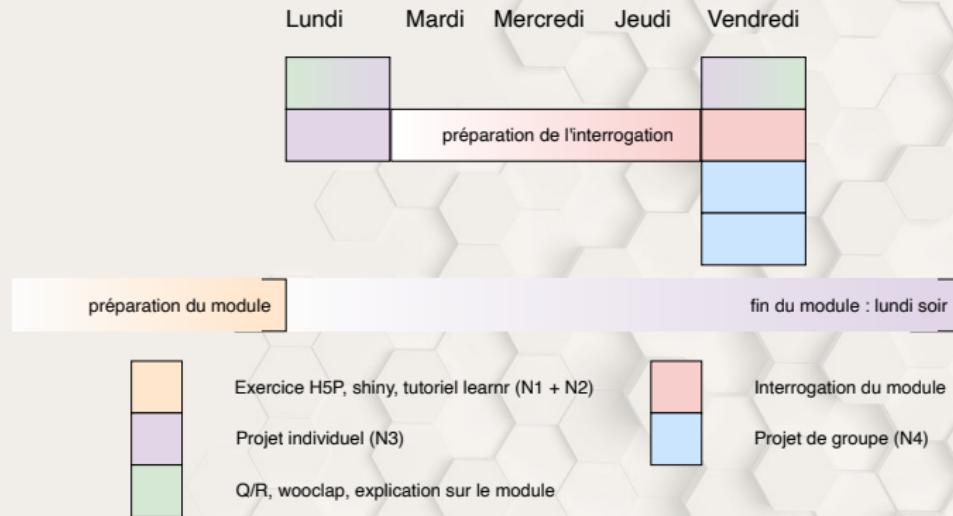
**UMONS**

# Progressivité : 4 niveaux de difficulté croissante



# Progressivité : découpage en 30 modules

- Chaque module est travaillé pendant **une semaine**
- **Deux séances de 2h et de 4h** en présentiel, respectivement et 12 à 15h de travail au total pour l'étudiant (= 0.5 ECTS) par module
- Un module **une semaine sur deux** pour laisser le temps aux étudiants de finaliser le précédent et de préparer le suivant



# Évaluation continue

Évaluation continue prenant en compte **toute l'activité des étudiants** (>360 notes) :

- Exercices N1 & 2 en distanciel
- Projets individuels et en groupe
- Une évaluations sommative par module

Répartition des points **particulièrement soignée** : motiver les étudiants sur les activités les plus importantes, mais sans rien négliger (*chaque exercice donne des points*) !

Niveau	Type	Pourcentage	
Niveau 1	Exercices H5P et shiny	2%	█
Niveau 2	Tutoriels learnr	4%	█
Niveau 3	Projets individuels cadrés	6%	█
Niveau 4	Projets de groupe	28%	██████
Évaluations	Interrogations & challenges	12% x 5	██████████

## Subsection 4

Recontextualisation (N3) = étape clé mais difficile

# Projets GitHub Classroom cadrés (N3)

**BioDataScience-Course / A09la\_ttest** Public template

[Edit Pins](#) [Unwatch](#) 2 [Fork](#) 0 [Star](#) 0

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

[master](#) [Go to file](#) [Add file](#) [Code](#) [Use this template](#)

**About**

No description, website, or topics provided.

- [Readme](#)
- [Activity](#)
- [0 stars](#)
- [2 watching](#)
- [0 forks](#)

[Report repository](#)

---

**Releases**

No releases published [Create a new release](#)

---

**Packages**

No packages published [Publish your first package](#)

---

**Contributors** 2

 [phgrosjean](#) Philippe Grosjean

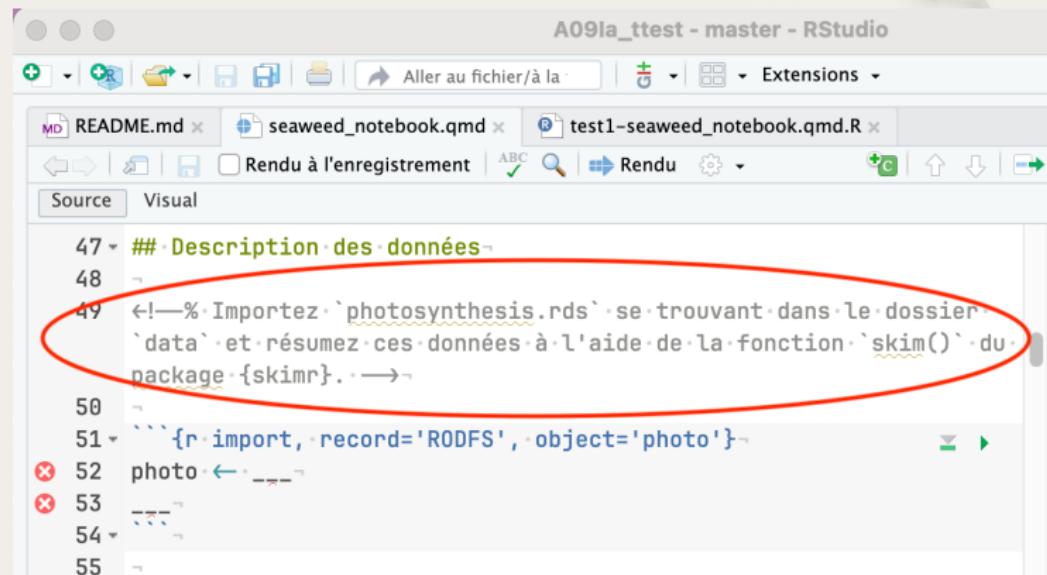
**README.md**

Réponse photosynthétique à un stress thermique chez *Fucus distichus* L., 1767

Connolly et al. (2016) ont étudié l'effet d'un stress thermique chez *Fucus distichus* L. 1767.

# Projets GitHub Classroom cadrés (N3)

## ■ Instructions sous forme de commentaires



A screenshot of the RStudio interface showing a qmd file named "seaweed\_notebook.qmd". The code contains several comments starting with `#` or `##` followed by descriptive text. A large red oval highlights a specific block of code in line 49:

```
47 ## Description des données
48
49 #--% Importez `photosynthesis.rds` se trouvant dans le dossier
  `data` et résumez ces données à l'aide de la fonction `skim()` du
  package {skimr}. -->
50
51 ```{r import, record='RODF5', object='photo'}```
52 photo <- ____
53
54
55
```

UMONS

# Projets GitHub Classroom cadrés (N3)

- Instructions sous forme de commentaires
- Interprétation par **sélection des phrases correctes**

```
56 * ``{r desccomment, output='asis'}  
57 select_answer(r"-{  
58 [] -.- Ce jeu de données ne contient aucune valeur manquante.  
59 [] -.- Une valeur est manquante dans ce jeu de données.  
60 [] -.- Plusieurs valeurs sont manquantes dans ce jeu de données.  
61  
62 [] -.- Ce tableau inclut uniquement des variables numériques.  
63 [] -.- Le tableau comporte uniquement des variables qualitatives.  
64 [] -.- Ce tableau contient deux variables qualitatives et quatre variables quantitatives. Ces variables quantitatives précisent les conditions de l'expérience à l'exception de la dernière qui reprend les résultats obtenus concernant la performance photosynthétique.  
65 [] -.- Ce tableau contient deux variables qualitatives et quatre variables quantitatives. Deux d'entre elles précisent les conditions de l'expérience et les deux autres correspondent aux résultats obtenus concernant la performance photosynthétique.}-")  
66 *  
67
```

# Projets GitHub Classroom cadrés (N3)

- Instructions sous forme de commentaires
- Interprétation par sélection des phrases correctes
- **Évaluation semi-automatique** avec {testthat} + suggestions pour s'améliorer

A screenshot of an RStudio interface showing a GitHub pull request review. The title of the PR is "A09la\_ttest". The review section shows a list of tests for a file named "seaweed\_notebook.qmd".

- Le bloc-notes seaweed\_notebook est-il compilé en un fichier final HTML ?
  - \* `test1-seaweed_notebook.qmd.R:6:3` [échec]
  - \* `test1-seaweed_notebook.qmd.R:15:3` [échec]
- La structure du document seaweed\_notebook est-elle conservée ?
  - ✓ `test1-seaweed_notebook.qmd.R:24:3` [réussi]
  - ✓ `test1-seaweed_notebook.qmd.R:38:3` [réussi]
  - ✓ `test1-seaweed_notebook.qmd.R:52:3` [réussi]
- L'entête YAML a-t-il été complété dans seaweed\_ca ?
  - \* `test1-seaweed_notebook.qmd.R:62:3` [échec]
  - \* `test1-seaweed_notebook.qmd.R:63:3` [échec]

UMONS

# Correction des projets (grilles critériées)

100 étudiants \* 10 projets \* 30 critères = 30.000 évaluations !

Learnitgrid

Correction par critère

A03IA_22M_DISTRIBUTIONS Set: 2022-12-07	TEMPLATE A03Ia_distributions	EVALUATEUR phgrosjean
Entrées pour un critère		
@histo_fact = histogrammes multiples de la variable feret.		
<a href="#">CSV</a>	<a href="#">Excel</a>	<a href="#">Filtrer : id2</a>
Max	Score	Commentaire
Contenu		Graphique
2	1 Ce n'est pas le graphique demandé. Il manque encore quelques instructions.	<pre>chart(data = zoo_sub, ~ feret %fill=% class   class) +   geom_histogram(data = sselect(zoo_sub, -class), fill = "class", bins = 25) +   geom_histogramshow.legend = FALSE, bins = 25) +   ylab("Effectifs") + scale_fill_viridis_d()</pre>
2	2 OK.	<pre>chart(data = zoo_sub, ~ feret %fill=% class  class) +   geom_histogram(data = sselect(zoo_sub, -class), fill = "grey", bins = 25) + geom_histogram(bins = 25, show.legend = FALSE)   ylab("Effectifs")</pre>

## Subsection 5

Support SAP

UMONS

# Évaluation souhaitée en séance

- **Première séance :** explication de l'approche pédagogique en Bab2, 16/9/2024 13h30 - 15h30, salle Vaughan (*De Vinci*)
- **Module type :** SDD II, module 2 (assez difficile), 21/10/2024 10h30 - 12h30 + 24/10/2024 8h15 à 12h30, salle Vaughan (*De Vinci*)

# Questions et difficultés (1/4 littératie & Bloom)

## ■ Faible niveau en **littératie** de certains étudiants

- en **français** : incapables de comprendre et résumer une trentaine de pages de texte, habitués seulement à annoter des PowerPoints fournis par le prof
- en **mathématiques** : difficultés à comprendre des concepts de base, des équations simples...
- **numérique** : difficulté à utiliser un clavier, faire des actions simples (copier-coller, copie d'écran, recherche de mots dans un texte...)

*Comment leur faire prendre conscience et les aider ensuite ? Pix ?*

## ■ Faible niveau par rapport à la **taxonomie de Bloom** (connaître et un peu comprendre) pour les autres cours

## Questions et difficultés (2/4 évaluation formative vs sommative)

- Tous les exercices donnent des points (reconnaissance du travail), mais les N1-3 sont à visée formative
- Triche possible (et facile en distanciel)
- -> Réussite avec un **niveau trop faible**
- Note absorbante mal perçue
- Points négatifs mal perçus + “fliquage”
- **Inutilité constatée du redoublement** (*impossible de varier les exercices chaque année*)

## Questions et difficultés (3/4 perception)

- **Réticence des étudiants** face à une autre autre pédagogie, incompréhension de l'inexistence d'un ratrappage en seconde session
- **Mauvaise perception de l'évaluation continue par les collègues** : séances obligatoires et pas d'examen en seconde session
- **Examen en seconde session ou pas ?** Cf. évaluation continue. *Si pas pousse les étudiants à travailler pendant l'année, mais très mal perçu.*

## Questions et difficultés (4/4 temps et reconnaissance institutionnelle)

- **Temps énorme à la préparation** (mais l'essentiel est écrit maintenant)
- **Temps énorme à la correction** (learnitgrid nous aide bien + projet d'évaluation par les pairs)
- **Temps énorme au coaching** individuel des étudiants (favoriser l'entre-aide dans la classe)
- **Aide inefficace de la DSI** (serveur LRS, plateforme LearnIt::R, calcul sur le cloud...)
- Souhait de passer à l'**évaluation avec Moodle + Safe Exam Browser** au lieu des interros papier
- **Reconnaissance indispensable du travail réalisé, ainsi que du rôle pionnier en biologie, voire en Faculté des Sciences à l'UMONS**

## Liens utiles



Plateforme pédagogique LearnIt::R : <https://github.com/learnitr>

*en cours d'élaboration sur base du code développé pour nos cours*

- Site web du cours : <https://wp.sciviews.org/>
- Organisation GitHub du cours : <https://github.com/BioDataScience-Course>