

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



Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique  
Service d'Écologie numérique  
<philippe.grosjean@umons.ac.be>, <phgrosjean@sciviews.org>  
<guyliann.engels@umons.ac.be>

Bilan avec le SAP, juillet 2024



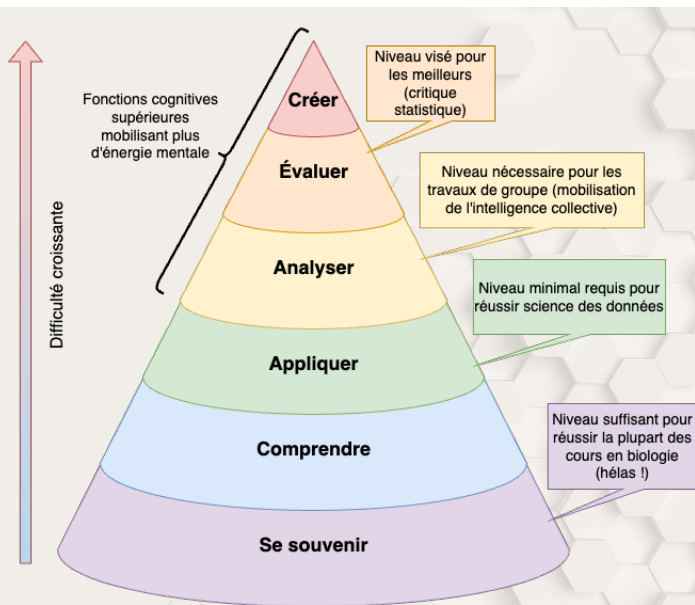
**UMONS**

# 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

## 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%

5 votes

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



34%

33 votes

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



61%

60 votes

54 répondants



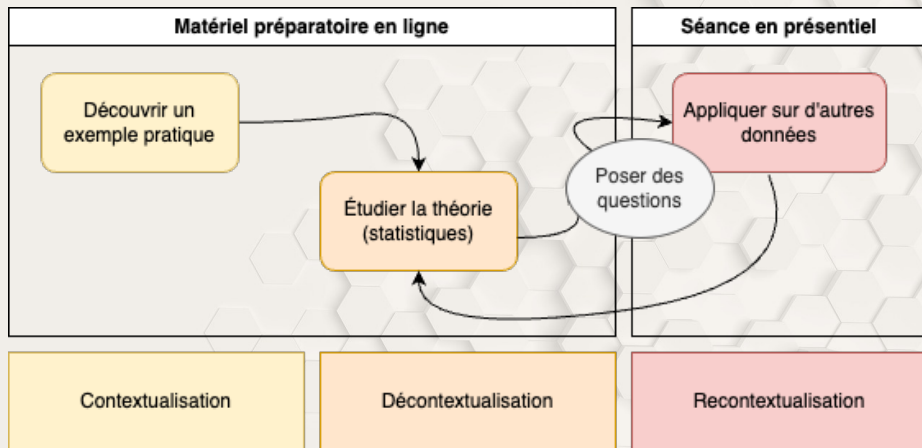
## Conséquences : axer un maximum sur la pratique !

- 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 UMONS website interface. At the top, the header includes the UMONS logo and name, followed by the course title 'Science des données biologiques'. Navigation links for Moodle, Discord, Github, E-mail, and RStudio are present. A blue button for 'Login with BioDataScience' is visible. The main content area is titled 'Bien débuter...' and features a photograph of a woman working on a laptop. Below the photo, there is a note about using didactic material. A sidebar on the left contains links for 'Accueil', 'Contact', 'Contenu des cours' (with sub-links for 'Science des données biologiques 1 (Bab2)', 'Science des données biologiques 2 (Bab3)', and 'Science des données biologiques 3-5 (Ma1 & 2)'), and 'Anciennes versions'.

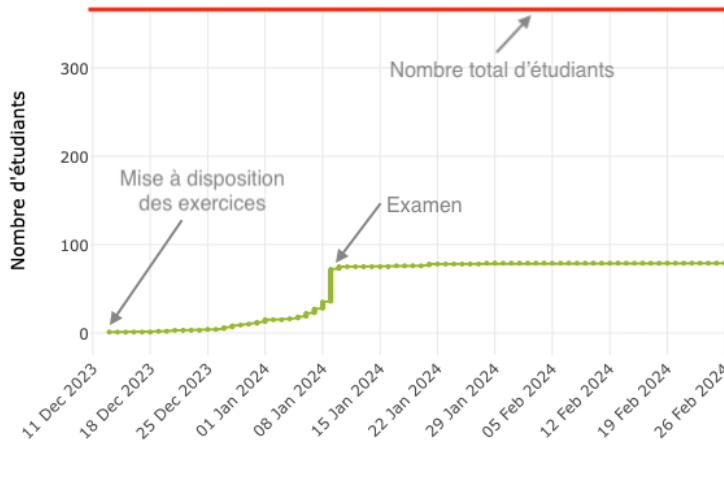
## Subsection 2

# 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



## Participation - des solutions...

**Vari**er 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`.

<-  [  (x)]

!

x2

x

duplicated

✓ Vérifier

# Participation - des solutions...

**Vari**er le type d'exercices : **H5P** + **learnr**

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

## uplicated()

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()

Start Over

## 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
```

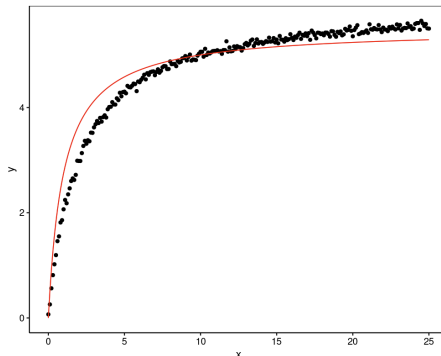
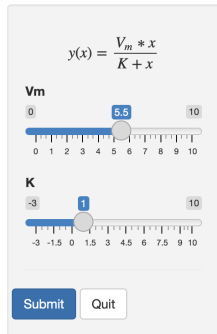
Next Topic

# 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 !



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

# 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 :

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



## Participation - des solutions...

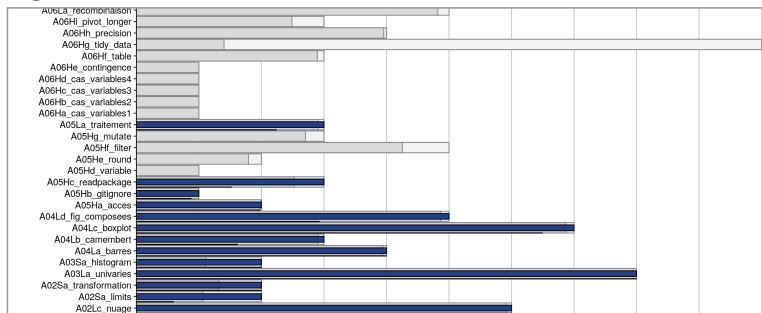
- Exercices directement dans le cours en ligne
- Liste des exercices à la fin de chaque module
- **Points attribués à la réalisation des exercices** (exemple, 6% de la note finale)



# Participation - des solutions...

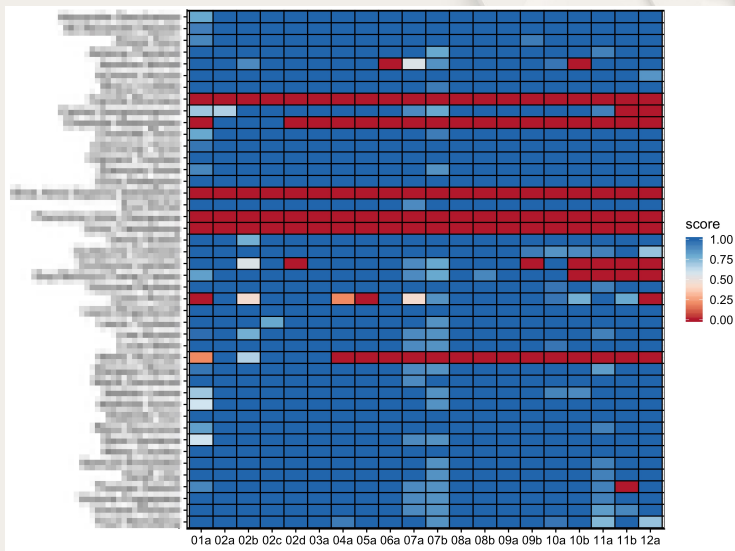
- Exercices directement dans le cours en ligne
- Liste des exercices à la fin de chaque module
- Points attribués à la réalisation des exercices (exemple, 6% de la note finale)
- **Rapport de progression en temps réel**

## Progression



# Participation - résultat

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

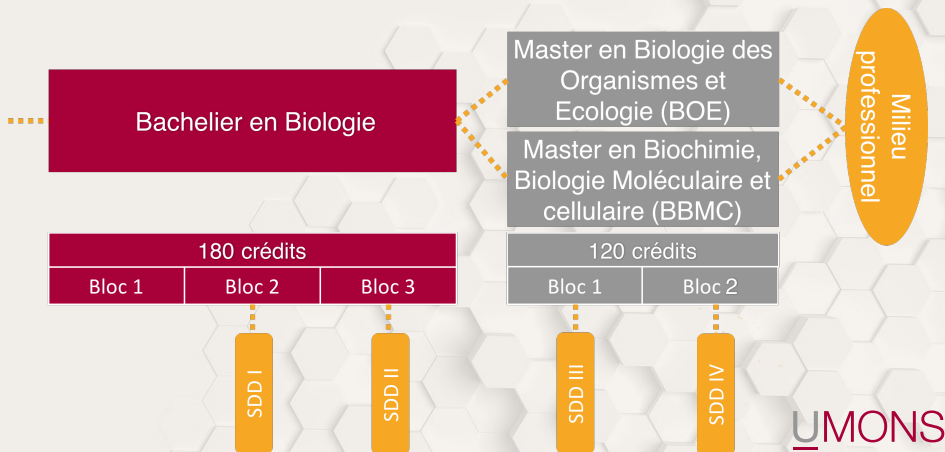


## Subsection 3

# Progressivité de l'apprentissage

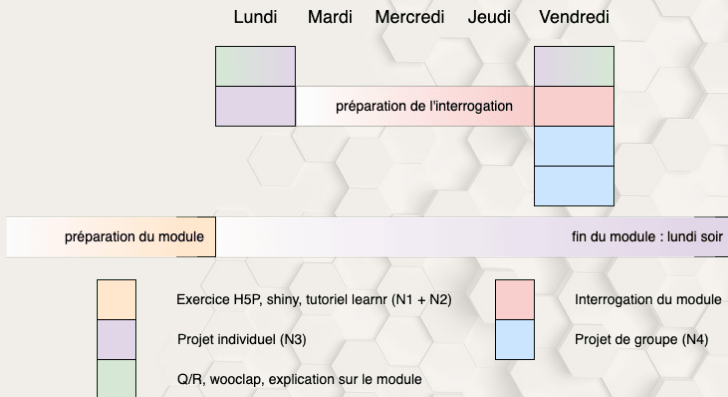
## Progressivité : formation sur 4 années (200h présentiel, 500h total)

- **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



# 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

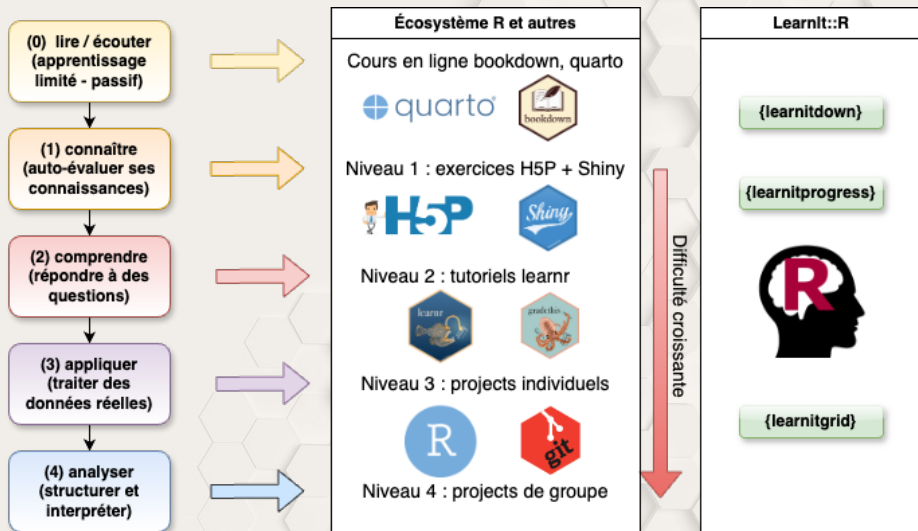


# Evaluation continue

- **Evaluation continue** : une interrogation par module et tous les exercices en ligne et les projets sont également comptabilisés :

Type	Description	Pourcentage
N1	Exercices H5P et shiny	2
N2	Tutoriels avec learnr	4
N3	Projets individuels cadré	6
N4	Projets de groupe	28
/	Evaluation / challenge	5 x 12

# Progressivité : 4 niveaux de difficulté croissante





# Projets GitHub Classroom cadrés

The screenshot shows a GitHub repository page for 'BioDataScience-Course / A09la\_ttest'. The repository is a public template with 2 unwatchers, 0 forks, and 0 stars. It has 9 commits by phgrosjean on March 13, 3 months ago. The repository contains several files and folders, including 'R', 'bibliography', 'data', 'docs', 'tests', '.gitignore', 'A09la\_ttest.Rproj', 'README.Rmd', and 'README.md'. The 'README.md' file is selected, showing the title 'Réponse photosynthétique à un stress thermique chez *Fucus distichus* L., 1767' and a snippet of text from Smallegange et al. (2016).

**Repository: BioDataScience-Course / A09la\_ttest** (Public template)

**Actions:** Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, Settings

**Files:**

File/Folder	Description	Commit
R	Adaptation of import.R	3 months ago
bibliography	Initialisation du projet	3 months ago
data	Adaptation of import.R	3 months ago
docs	Save objects	3 months ago
tests	Save objects	3 months ago
.gitignore	collection of several objects to automate the correction	3 months ago
A09la_ttest.Rproj	Create A09la_ttest.Rproj	3 months ago
README.Rmd	Some more changes in the README file	3 months ago
README.md	Some more changes in the README file	3 months ago

**README.md**

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

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

**About:** No description, website, or topics provided.

**Releases:** No releases published. [Create a new release](#)

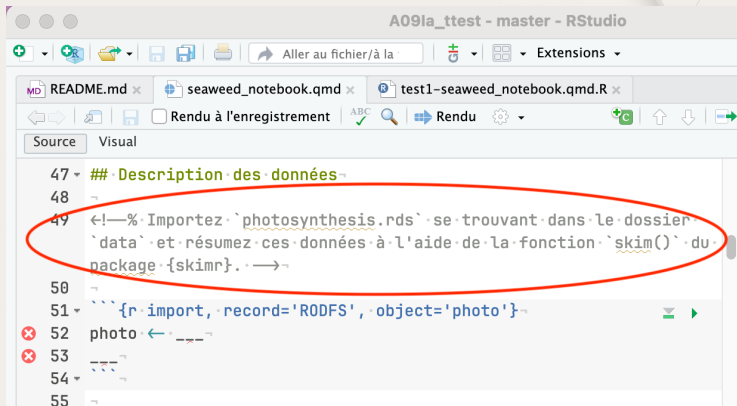
**Packages:** No packages published. [Publish your first package](#)

**Contributors:** 2

**phgrosjean** Philippe Grosjean

# Projets GitHub Classroom cadrés

## ■ Instructions sous forme de commentaires



The screenshot shows the RStudio interface with the file 'test1-seaweed\_notebook.qmd.R' open. The editor displays R code with a comment on line 49 circled in red. The comment instructs the user to import a file named 'photosynthesis.rds' from the 'data' directory and summarize it using the 'skim()' function from the 'skimr' package. The code also includes an R chunk header for 'photo' and a line for creating a data frame.

```
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='RODFS', object='photo'}
52 photo <- ---
53 ---
54
55
```

# Projets GitHub Classroom cadrés

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

```

56 ▾ ```{r·desccomment,·output='asis'}~
57 select_answer(r"-{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·
représent·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

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

The screenshot shows the RStudio interface with the 'Test' tab selected. The title bar indicates the environment is 'A091a\_ttest'. The top menu bar includes 'Environnement', 'Historique', 'Connexions', 'Construire', 'Git', and 'Tutoriel'. Below the menu, there are buttons for 'Construire tout' and 'Plus'. The main pane displays the test results for 'test1-seaweed\_notebook.qmd'.

```

) test1-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]
  
```

# Correction des projets (grilles critériées)

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

Learnitgrid

☒ Correction par critère

A031A\_22M\_DISTRIBUTIONS  
Set: 2022-12-07

TEMPLATE  
A031a\_distributions

EVALUEUR  
phgrosjean

Entrées pour un critère

@histo\_fact = histogrammes multiples de la variable feret.

CSV Excel

Filtrer : id2

Max	Score	Commentaire	Contenu	Graphique	Liens	Evaluateur	Étudiant/groupe
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_histogram(show.legend = FALSE, bins = 25) +   ylab("Effectifs") + scale_fill_viridis_d()</pre>		<a href="#">template</a> <a href="#">repo</a> <a href="#">docs</a> <a href="#">html (get Rmd)</a>	eval01	id298
2	2 OK.		<pre>chart(data = zoo_sub, ~ feret %fill=% class  class) +   geom_histogram(data = sselect(zoo_sub, ~class), fill = "grey", bins = 25) + geom_histo gram(bins = 25, show.legend = FALSE) ylab("Effectifs")</pre>		<a href="#">template</a> <a href="#">repo</a> <a href="#">docs</a> <a href="#">html (get Rmd)</a>	eval01	id295

## Subsection 4

### Support SAP

# 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

- Reconnaissance du travail réalisé, *ainsi que du rôle pionnier en biologie, voire en Faculté des Sciences à l'UMONS*
- Réticence pour ne pas dire opposition des collègues en biologie
- Réticence pour ne pas dire révolution des étudiants
- Faible niveau en littérature (textes, mais aussi mathématique et numérique) de certains étudiants. *Rien pour leur faire prendre conscience, rien pour les aider ensuite*
- Difficulté d'évaluation des étudiants *(les exercices donnent des points, les étudiants trichent, certains réussissent avec un niveau trop faible)*
- Impossibilité de mise en œuvre de la note absorbante *(révolution chez les étudiants !)*
- Inutilisé constatée du redoublement *(impossible de varier les exercices chaque année)*
- Présence en séance obligatoire => *un ineptie pour nos collègues biologistes, source de tensions*
- Examen en seconde session ou pas ? cf. évaluation continue. *Pousse l'étudiant à travailler pendant l'année, mais il ne comprend pas qu'il n'y ait pas de moyen de rattrapage. Totalement inacceptable pour les collègues.*



## 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>