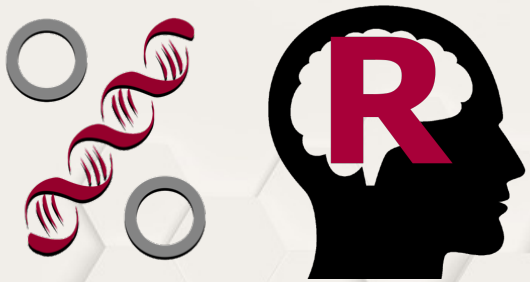
# Apprentissage de la science des données, six ans d'innovation pédagogique, et ensuite ?



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

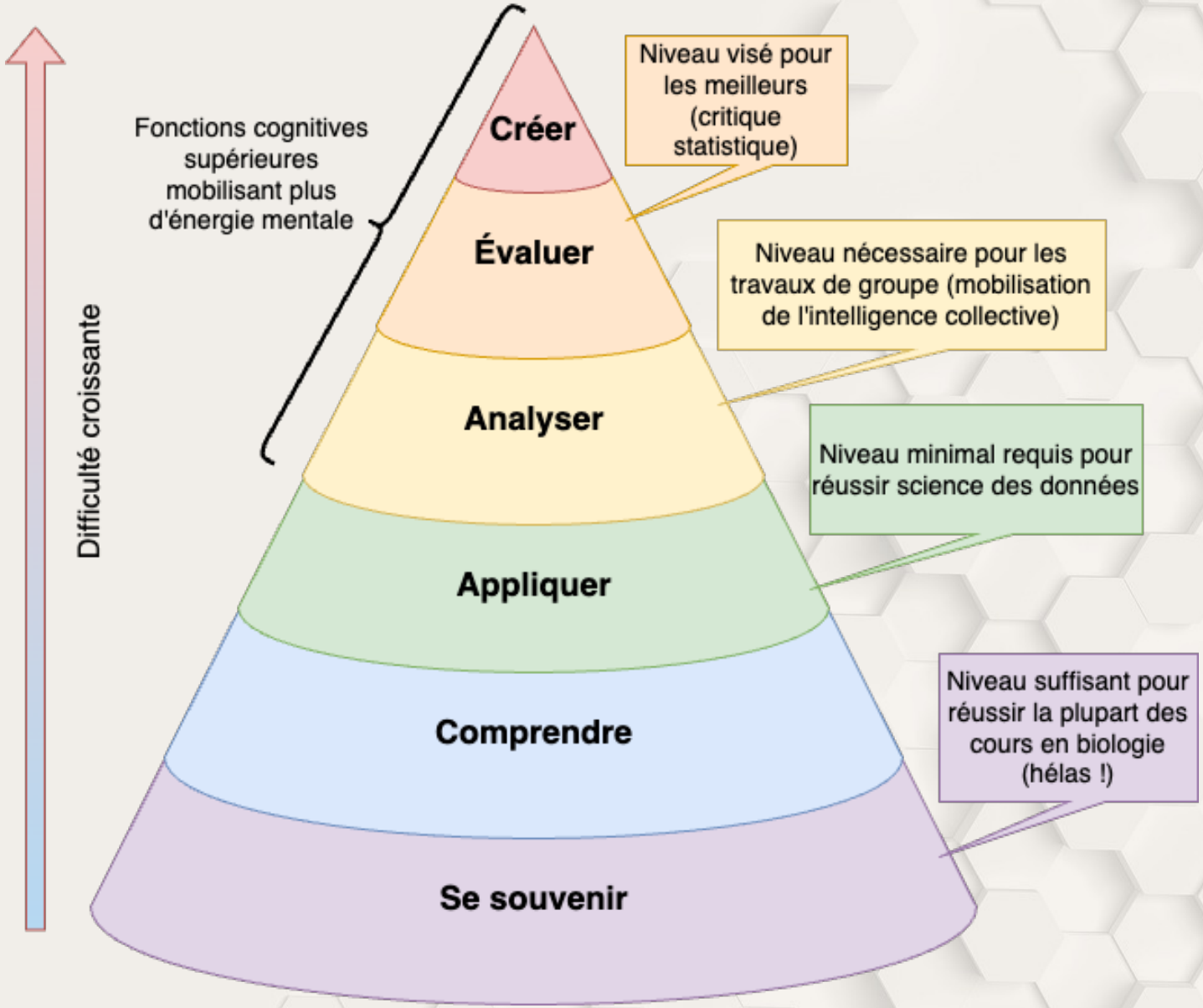
# Objectifs pédagogiques principaux des cours de SDD

- **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).*

UMONS

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




UMONS

Subsection 1

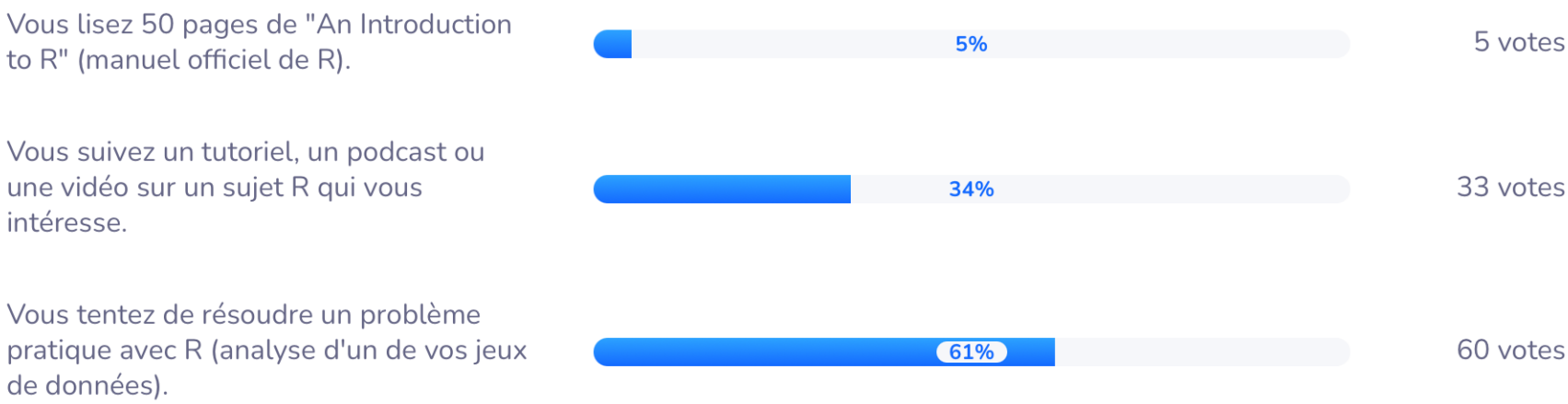
# Sondage de data scientists avec Wooclap

UMONS

# Question : mode d'apprentissage préféré

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

98 répondants




UMONS

# Question ouverte

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

54 répondants


UMONS

# Conséquences : axer un maximum sur la pratique !

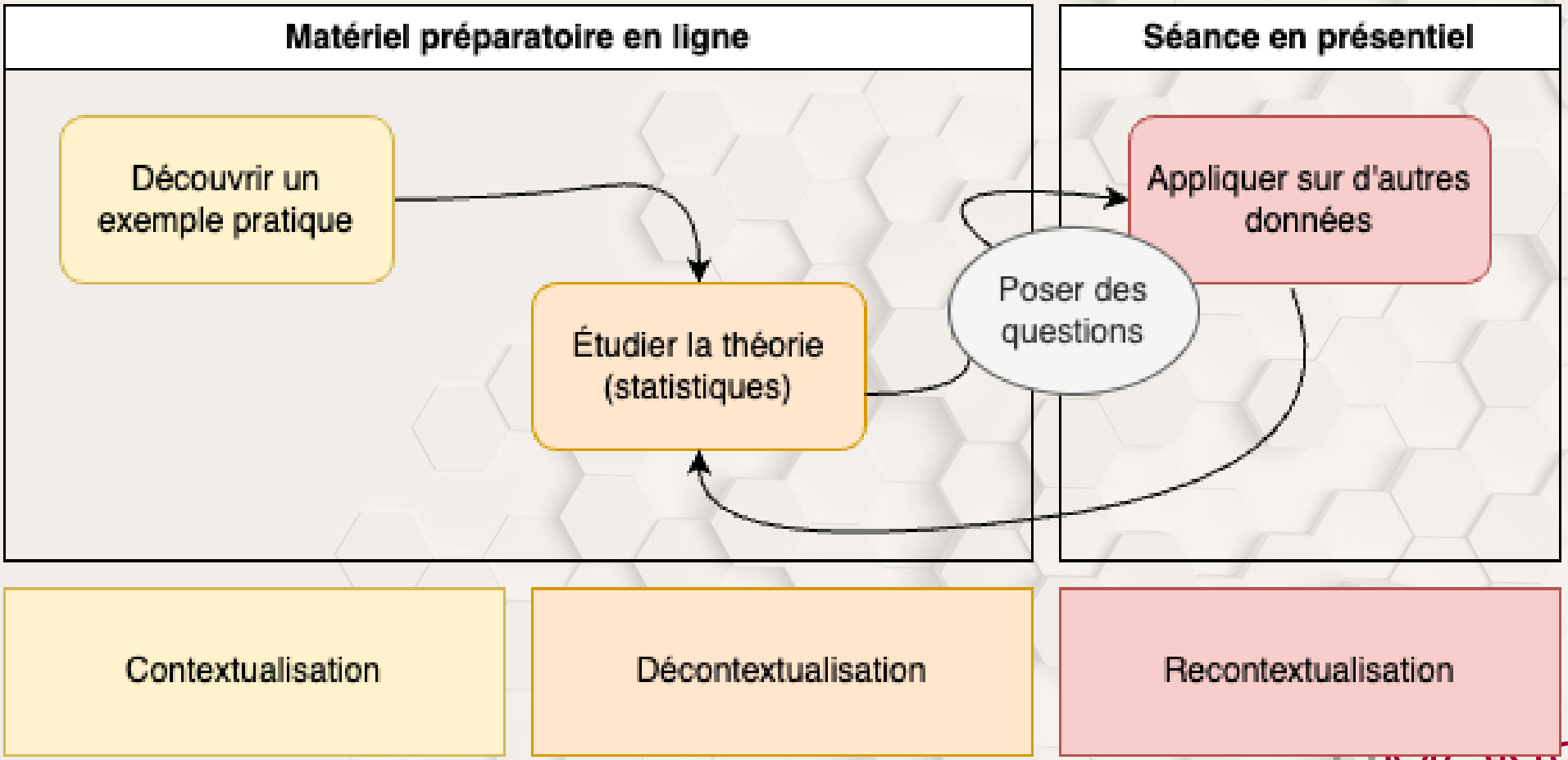
- Apprentissage en classe inversée (*un cours purement théorique avant des exercices pratiques ne sert à rien ;* ***les étudiants n'enregistrent rien***)
- **Exercices en ligne** pour l'autoévaluation
- **Projets** avec analyse de données biologiques réelles comme **activité principale en présentiel**

*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.*

UMONS

# Contextualiser - décontextualiser - recontextualiser

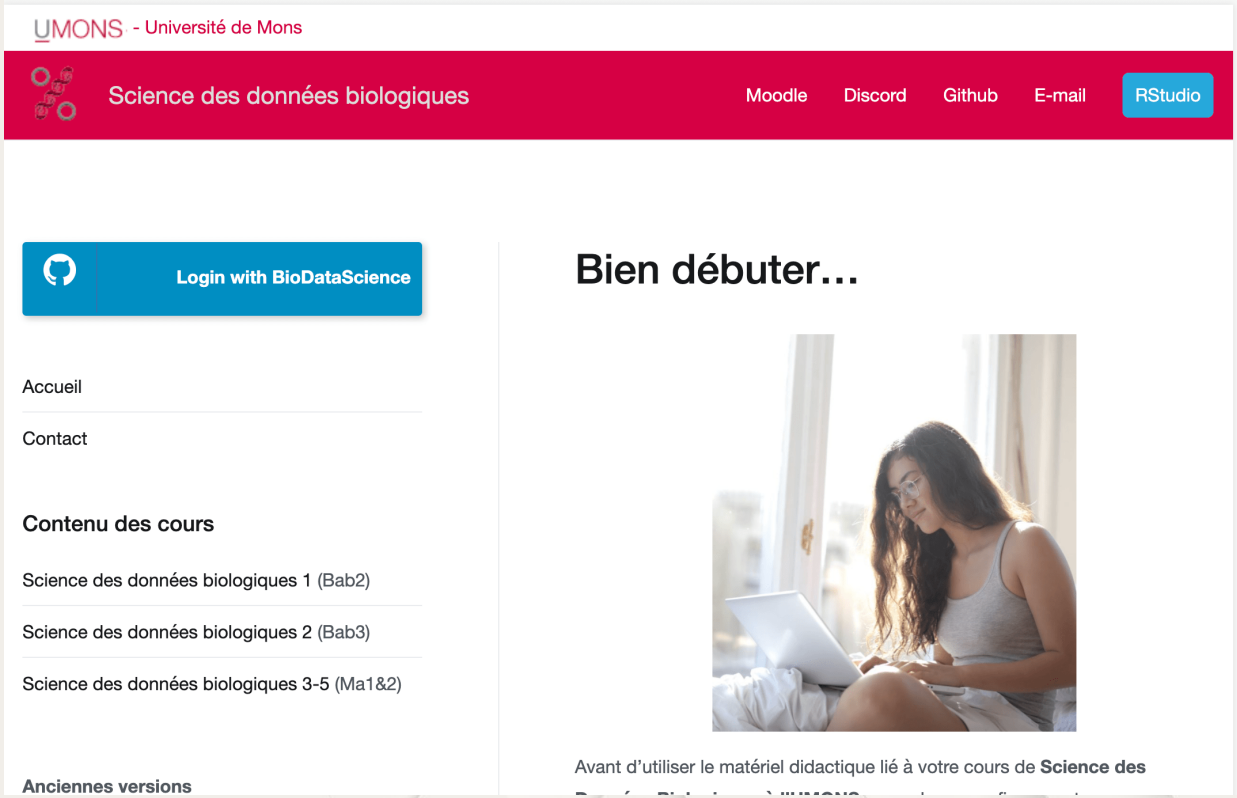
Exemple pratique qui introduit un concept quasi-systématiquement dans le cours en ligne. Recontextualisation en séance.



# 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 !)



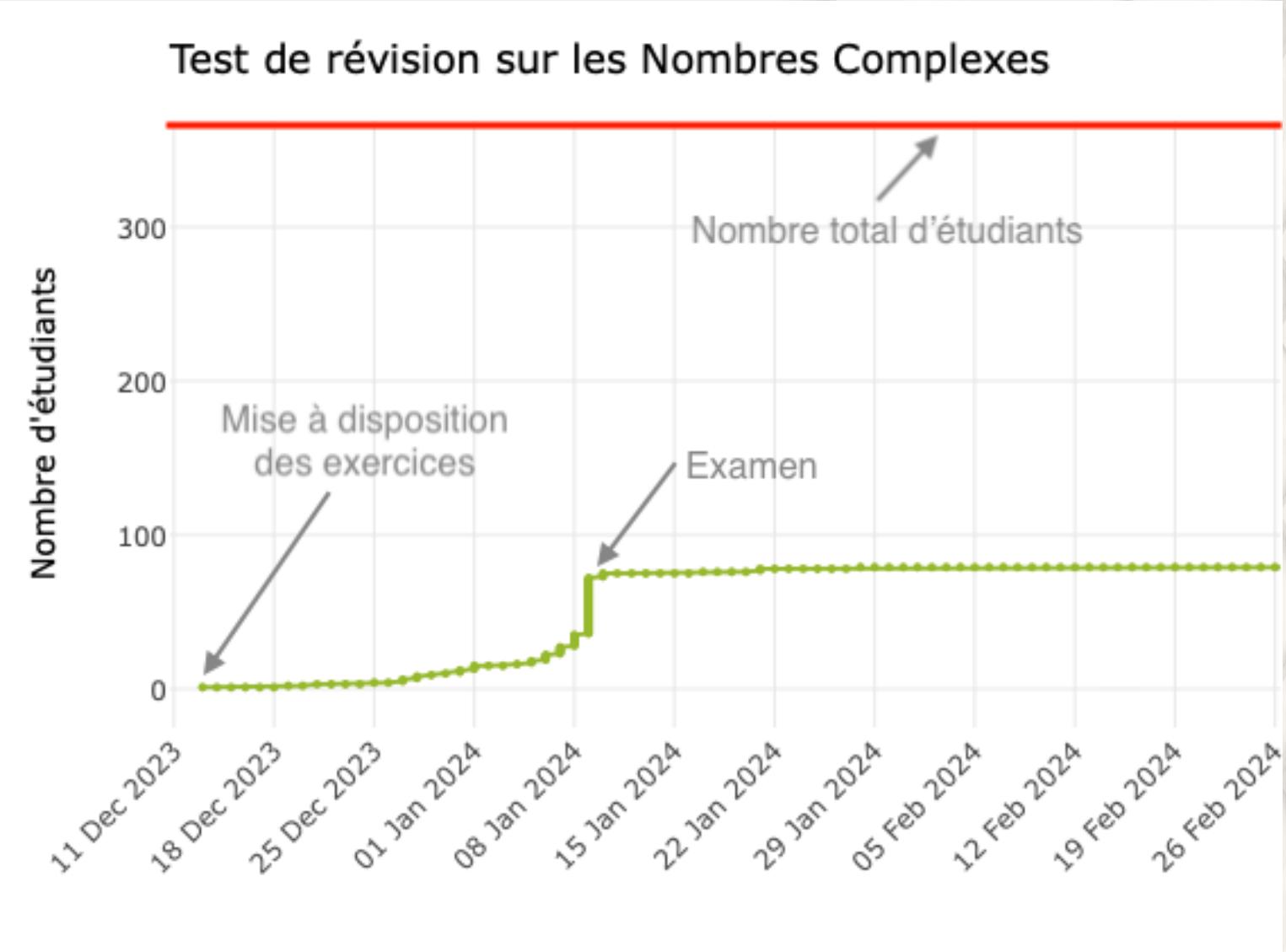
Subsection 2

# Motiver et capter l'attention

UMONS

# Exercices facultatifs - peu de participation

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

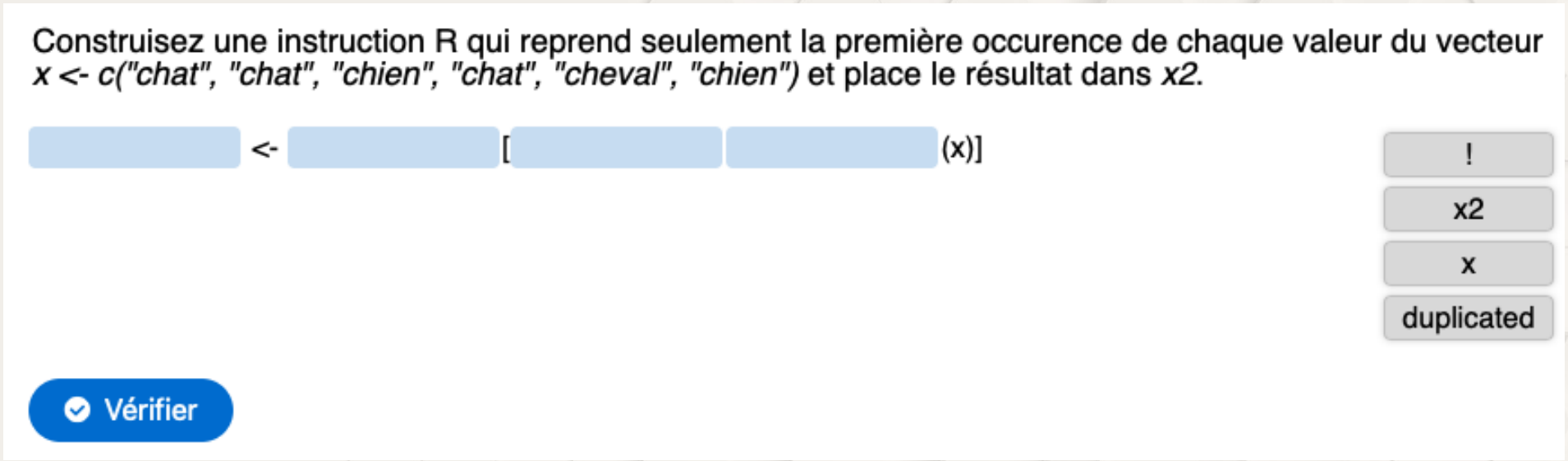



UMONS

# Participation - des solutions…

**Varier** le type d'exercices : H5P

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



# Participation - des solutions…

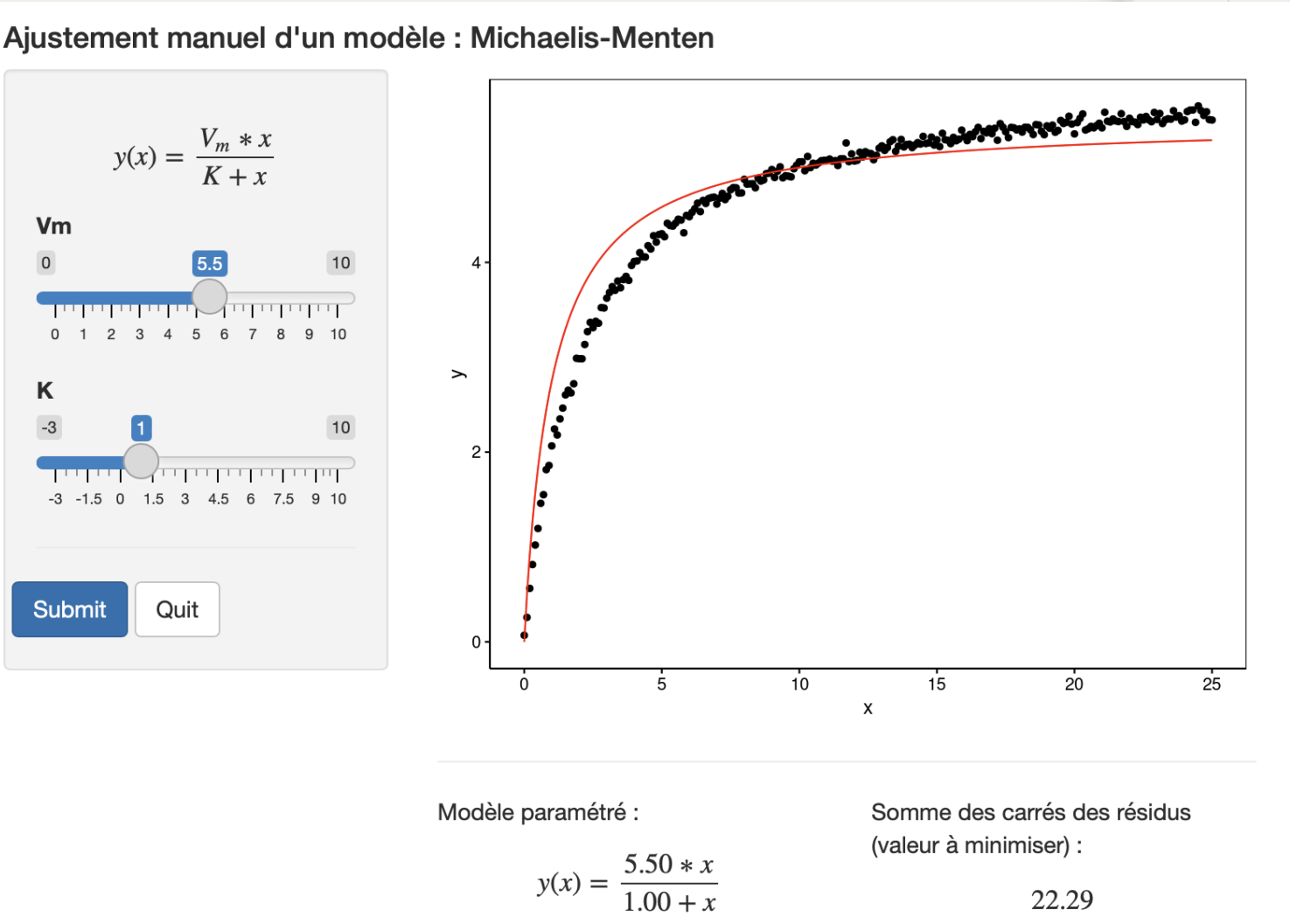
**Varier** le type d'exercices : **H5P** + learnr

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

# Participation - des solutions…

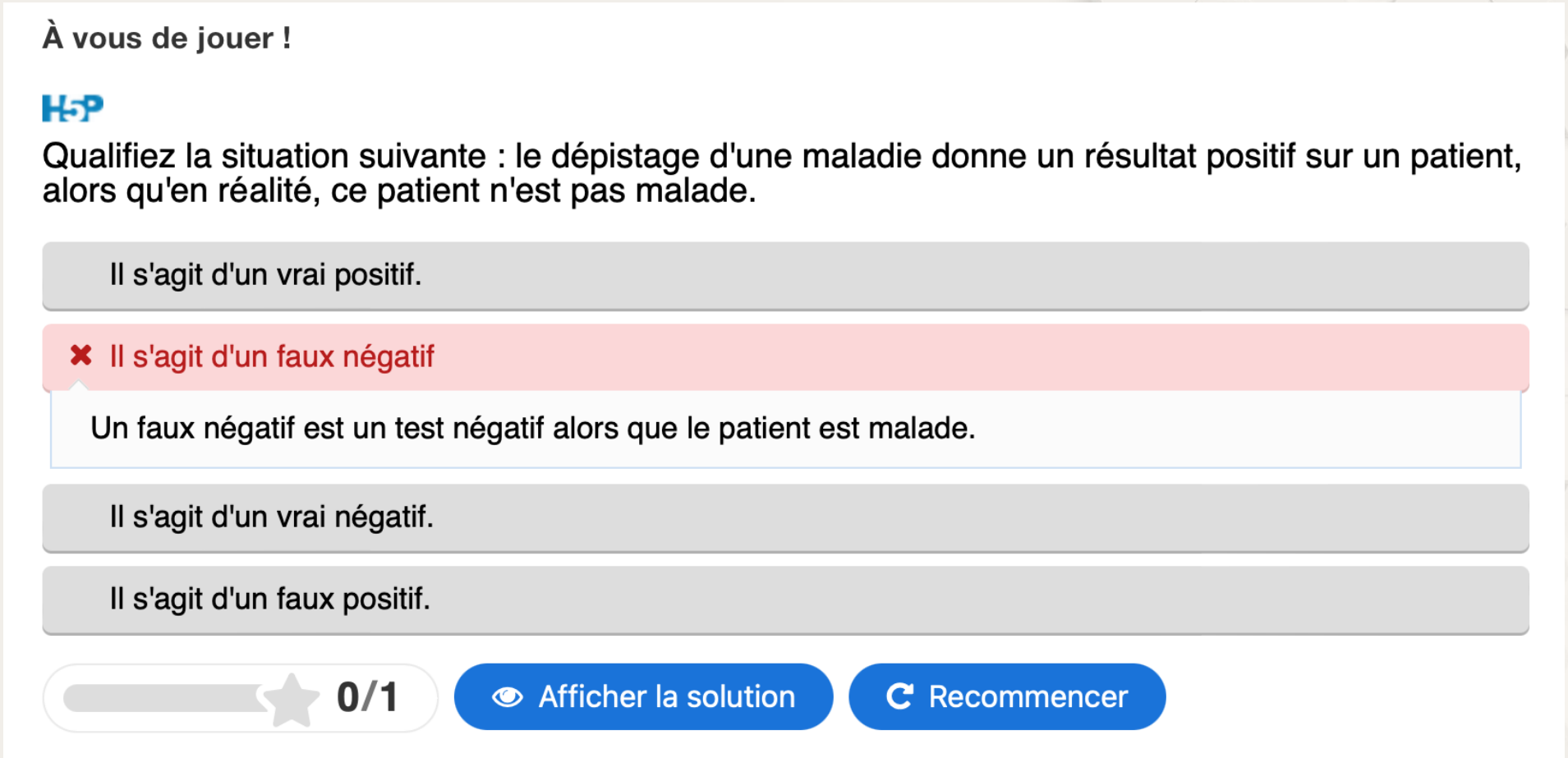
**Varier** le type d'exercices : **H5P** + **learnr** + Shiny

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



# Participation - des solutions…

- Exercices directement **dans le cours en ligne**




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 :

- H5P Les fonctions chart() et geom_histogram()
- H5P Modes et symétries
- Nombre de classes d'un histogramme
- Graphiques univariés
- H5P La fonction chart() et geom_density()
- Graphiques de distribution des données
- H5P La fonction chart() et geom_violin()
- H5P Deux versions d'un même fichier
- H5P Résolution d'un conflit
- H5P Résolution d'un conflit (suite)
- Analyse de données (partie II)

UMONS

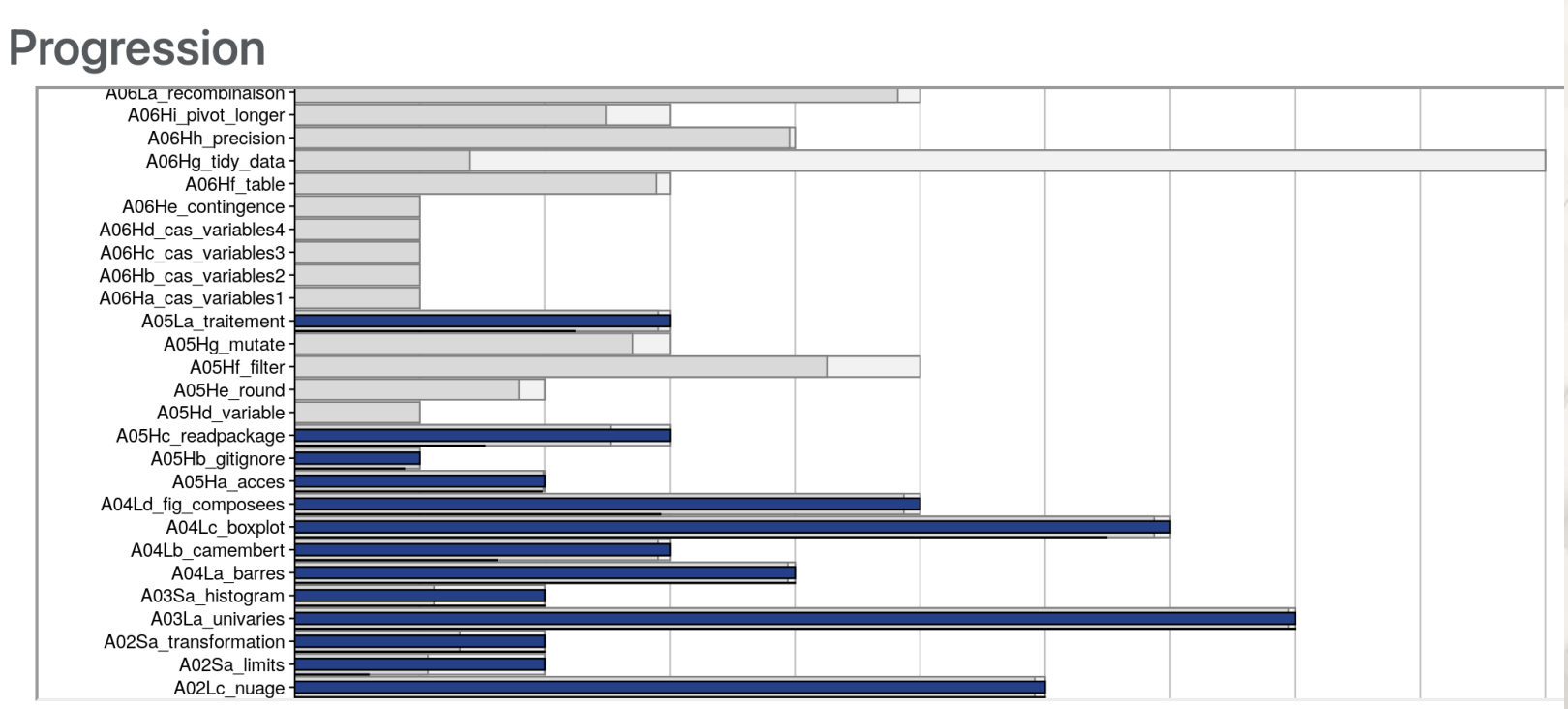
# Participation - des solutions…

- 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)

UMONS

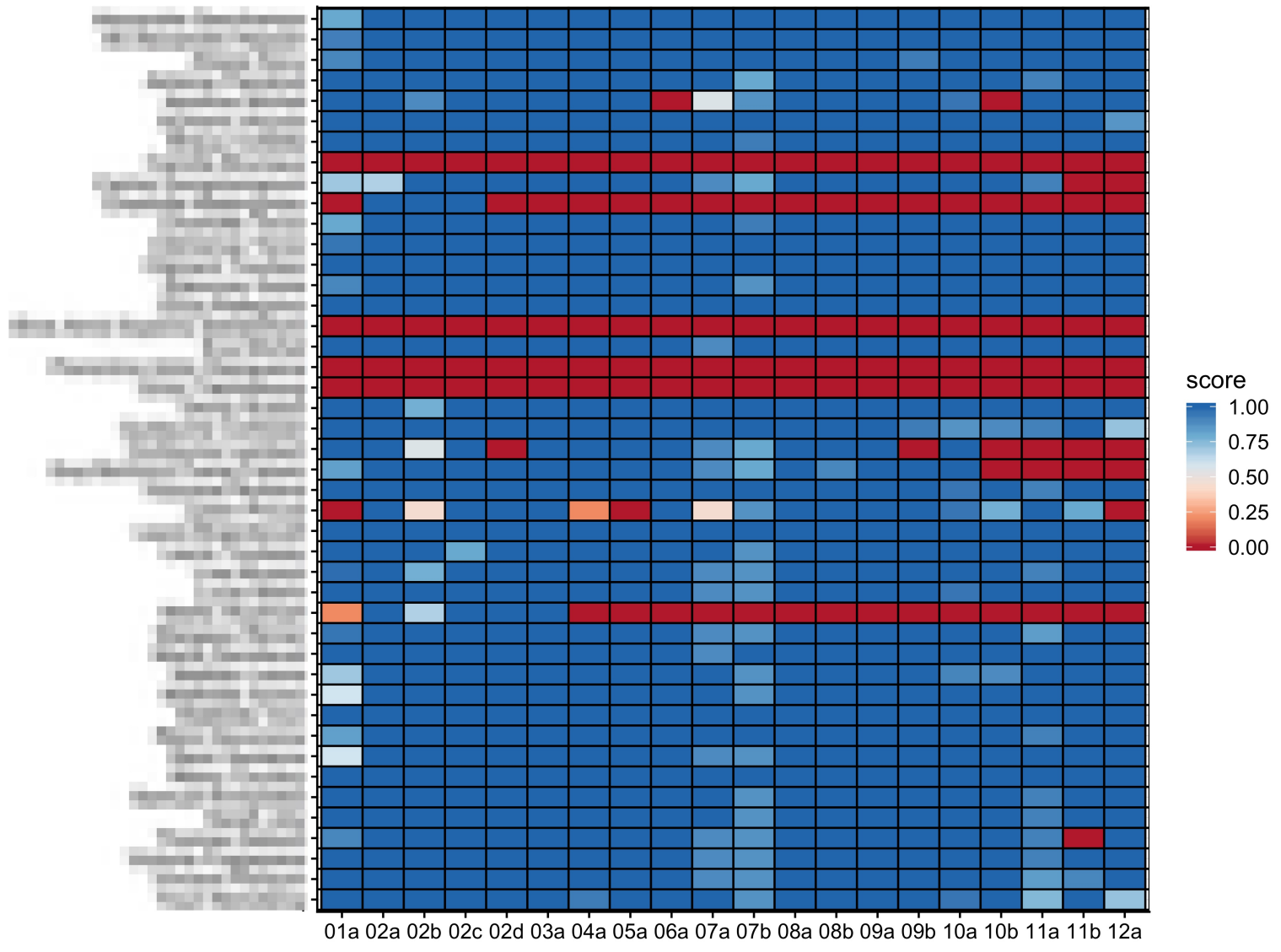
# Participation - des solutions…

- 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



# Participation - résultat

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




UMONS

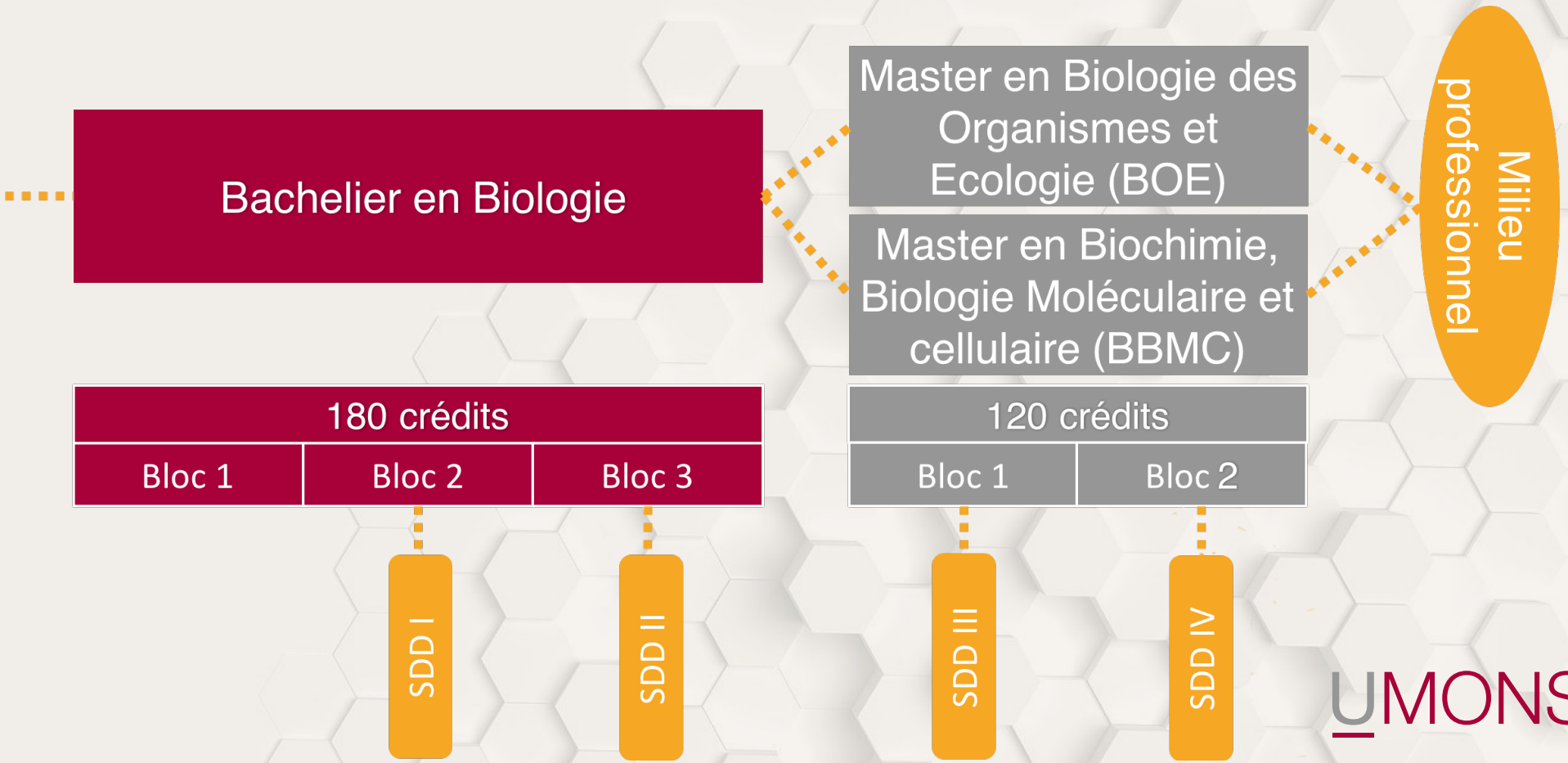
Subsection 3

# Progressivité de l'apprentissage

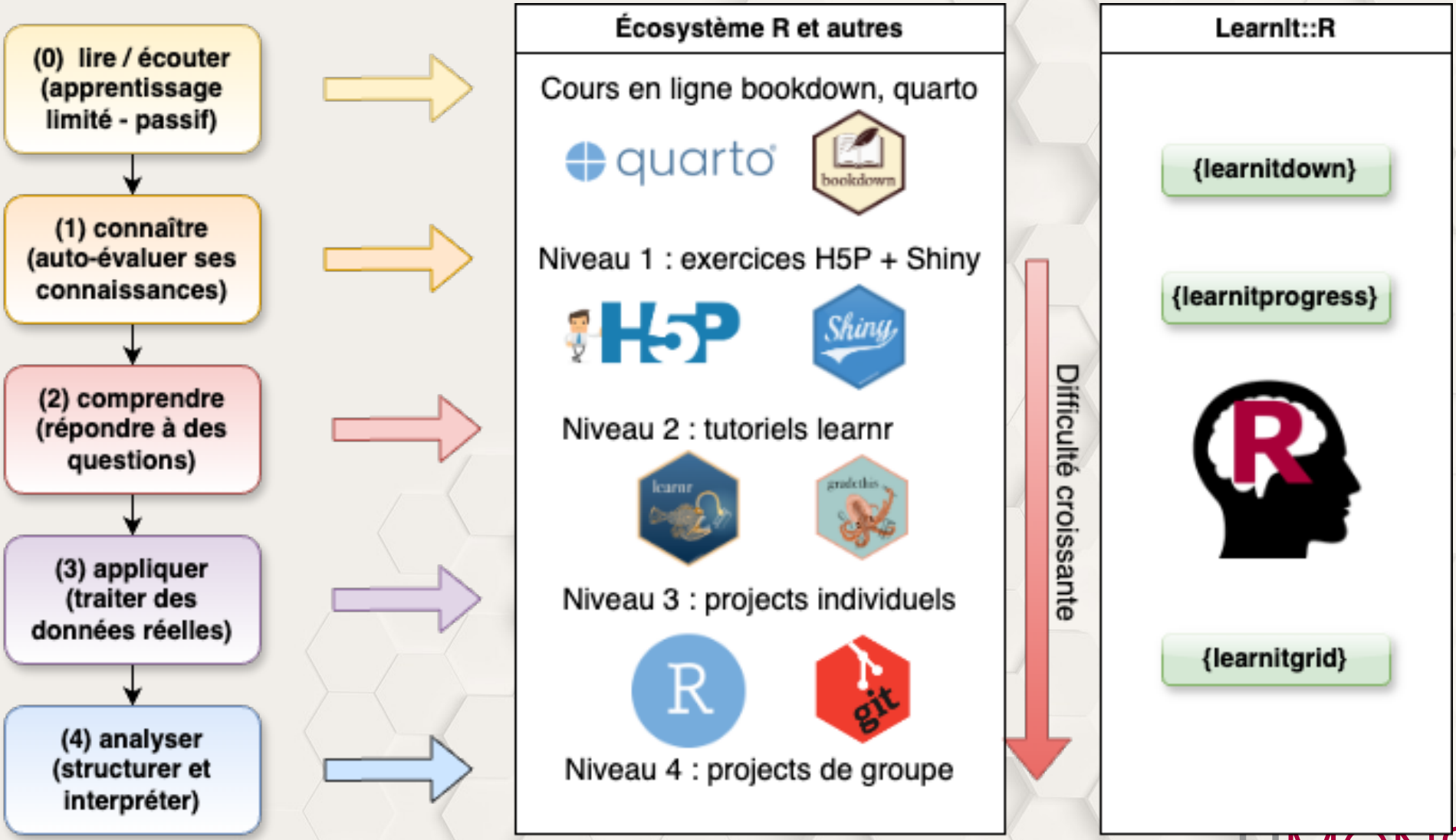
UMONS

# 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

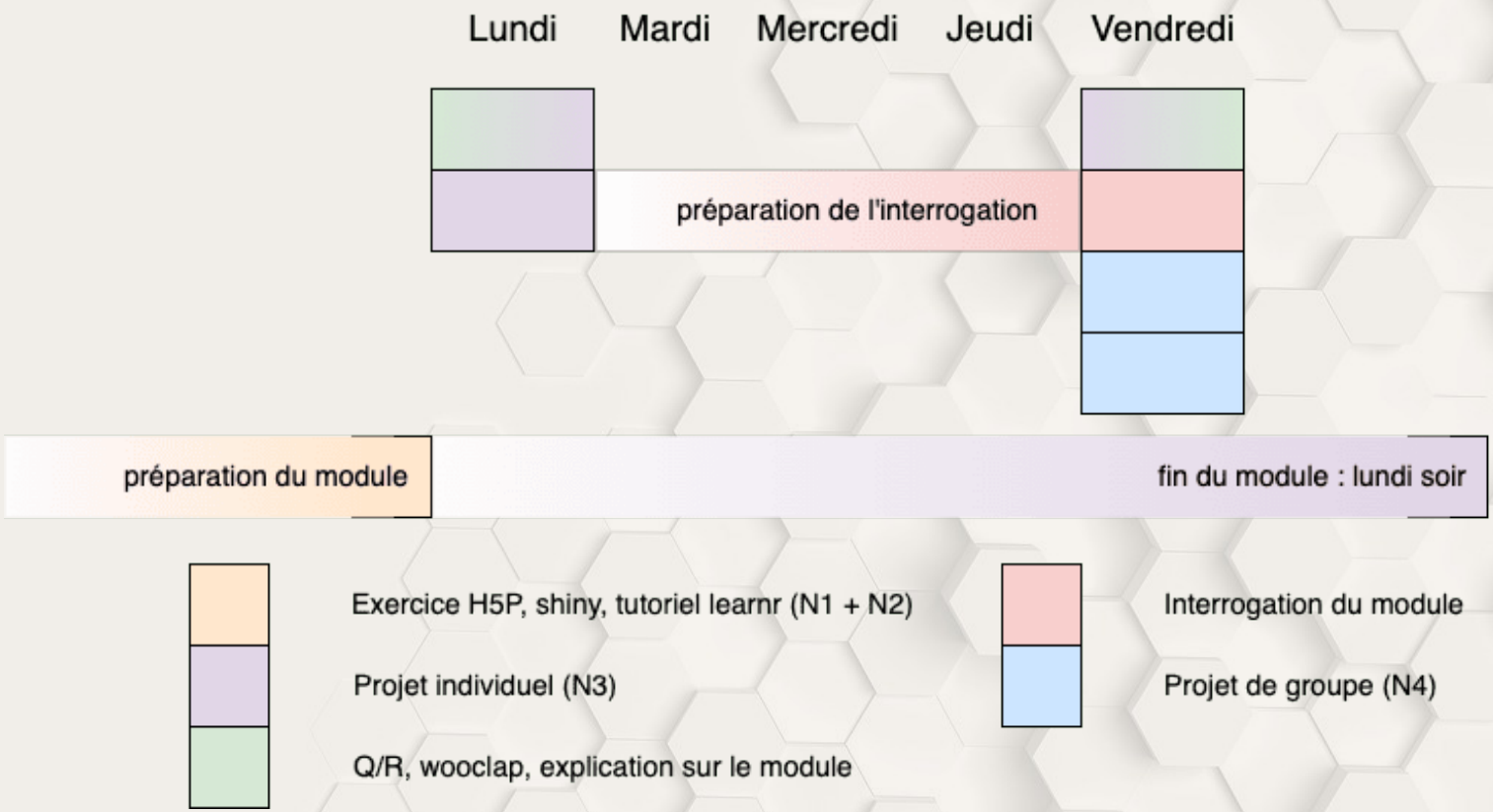


# Progressivité : 4 niveaux de difficulté croissante



# 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




UMONS

# 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 |

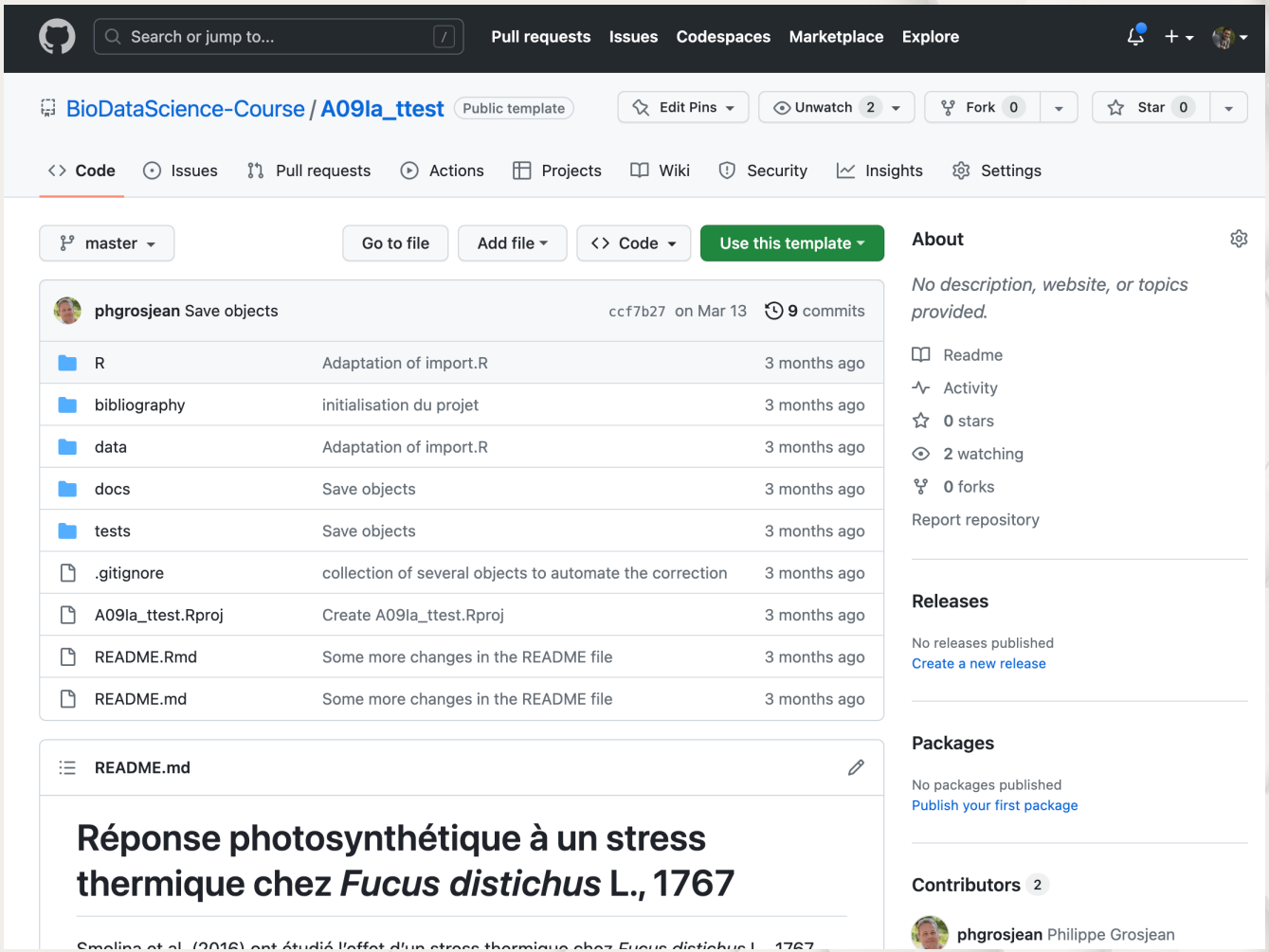
UMONS

Subsection 4

# Recontextualisation (N3) = étape clé mais difficile

UMONS

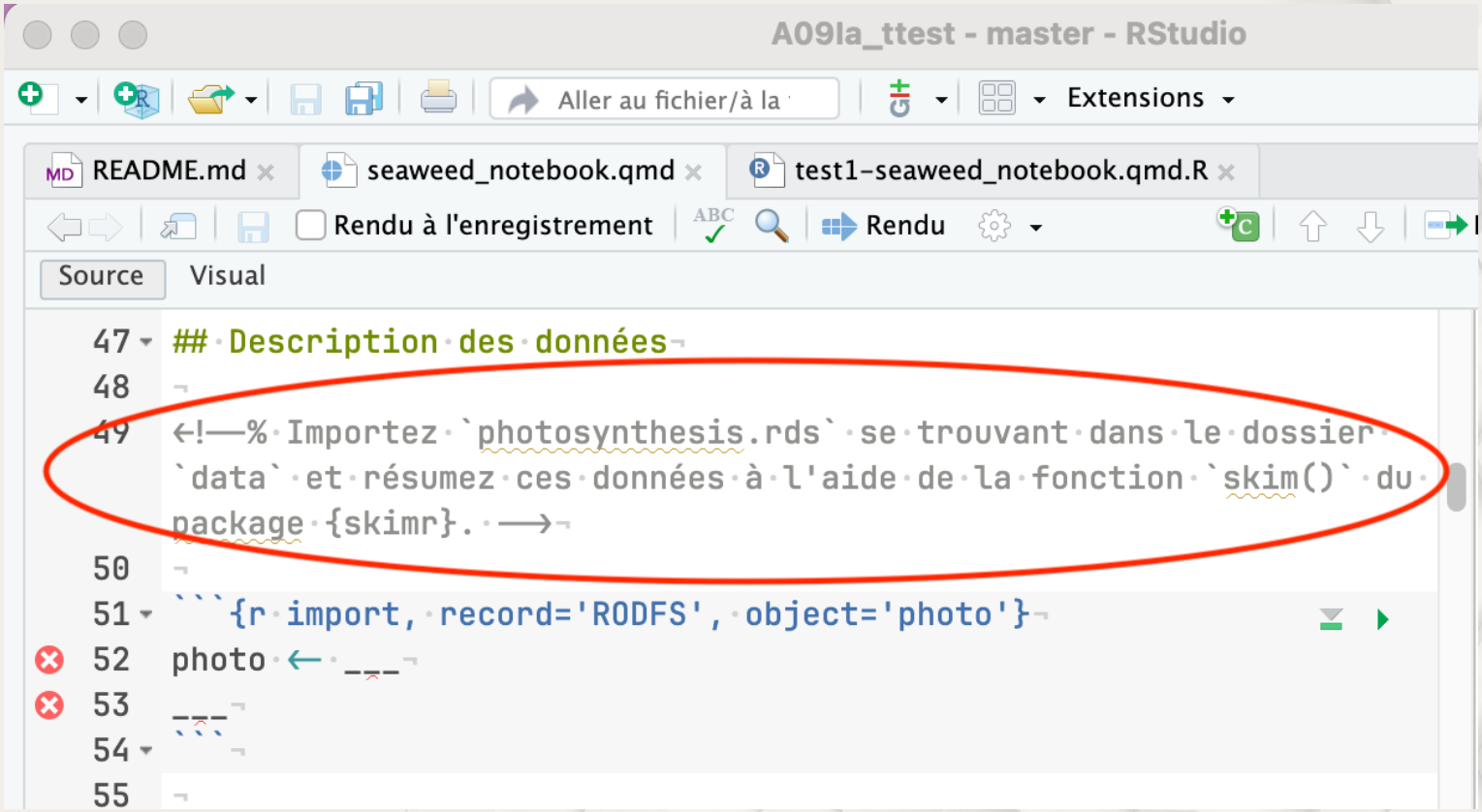
# Projets GitHub Classroom cadrés (N3)




UMONS

# Projets GitHub Classroom cadrés (N3)

- **Instructions** sous forme de commentaires



# Projets GitHub Classroom cadrés (N3)

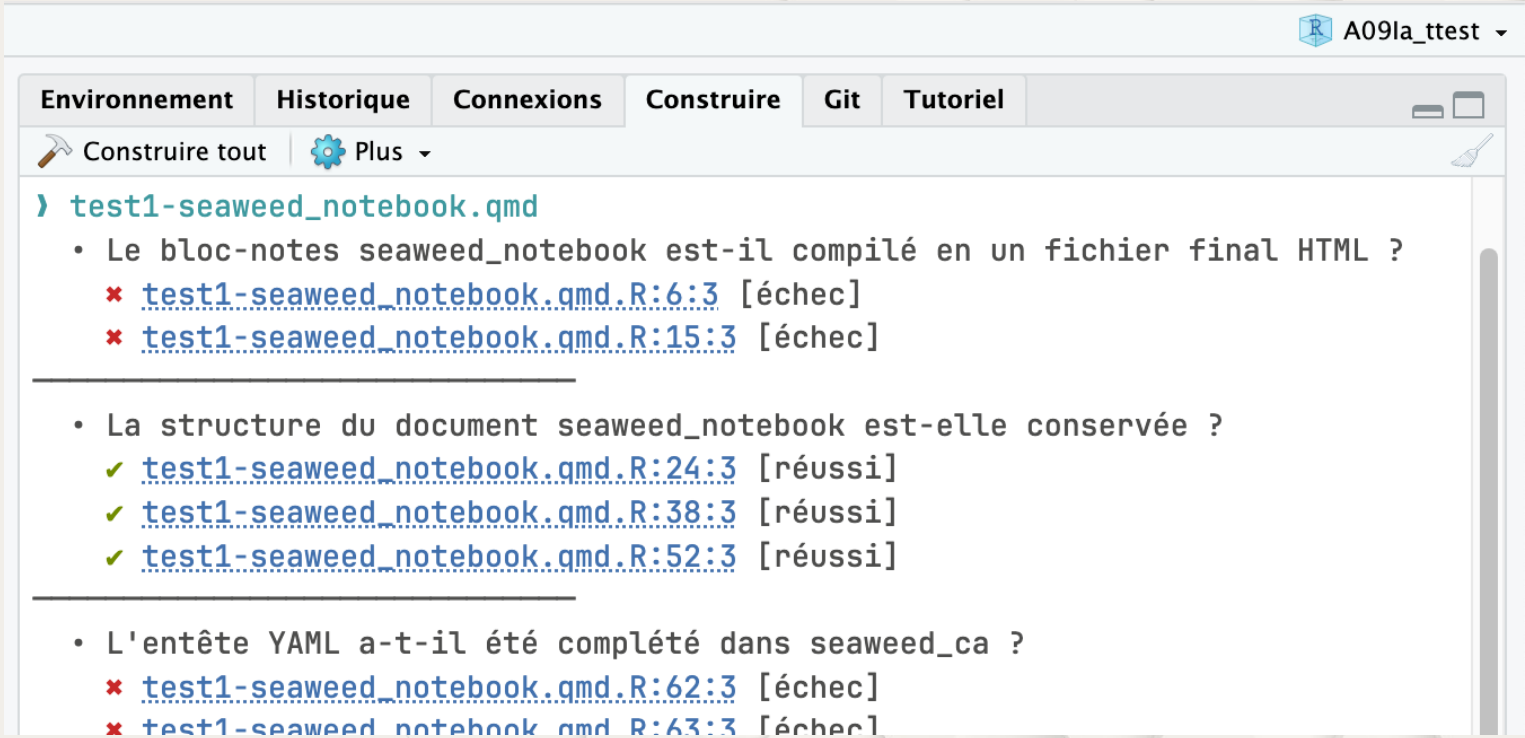
- Instructions sous forme de commentaires
- Interprétation par **sélection des phrases correctes**

```
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
```

UMONS

# 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




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**

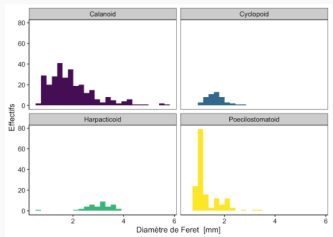
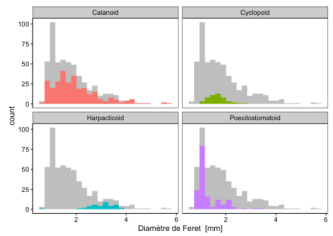
EVALUATEUR
**phgrosjean**

Entrées pour un critère

@histo_fact = histogrammes multiples de la variable feret.

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. | `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()` |  | template repo docs html (get Rmd) | eval01 | id298 |
| 2 | 2 OK. | `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")` |  | template repo docs html (get Rmd) | eval01 | id295 |
| | | |  | | | |

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)*

UMONS

# Questions et difficultés (1/4 litteratie & 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

UMONS

# 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)*

UMONS

# Questions et difficultés (3/4 perception)

- **Réticence des étudiants** face à une autre pédagogie, incompréhension de l'inexistance d'un rattrapage 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.*

UMONS

# 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*

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

UMONS