

Apprendre R et les statistiques... grâce à R



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>

Rencontres R 2024, Vannes



UMONS

Enseigner R et les statistiques à des étudiants réfractaires

Cursus universitaire en biologie, voir <https://wp.sciviews.org>

UMONS - Université de Mons

Science des données biologiques

Moodle Discord Github E-mail RStudio

Login with BioDataScience

Accueil


Contact

Contenu des cours

- Science des données biologiques 1 (Bab2)
- Science des données biologiques 2 (Bab3)
- Science des données biologiques 3-5 (Ma1&2)

Anciennes versions

Bien débiter...



Avant d'utiliser le matériel didactique lié à votre cours de Science des

Subsection 1

Sondage du public avec Wooclap

Question 1 (retient-on en écoutant ?)



1. Le cours de science des données biologique de l'UMONS est accessible à l'adresse:

56 bonnes réponses
sur 99 répondants

<https://sdb.biodatascience.org>

29%

29 votes



<https://wp.sciviews.org>

57%

56 votes

<https://www.biodatascience.org>

9%

9 votes

<https://sdb.sciviews.org>

5%

5 votes

Question 2 (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

Question 3 (question ouverte)

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

54 répondants



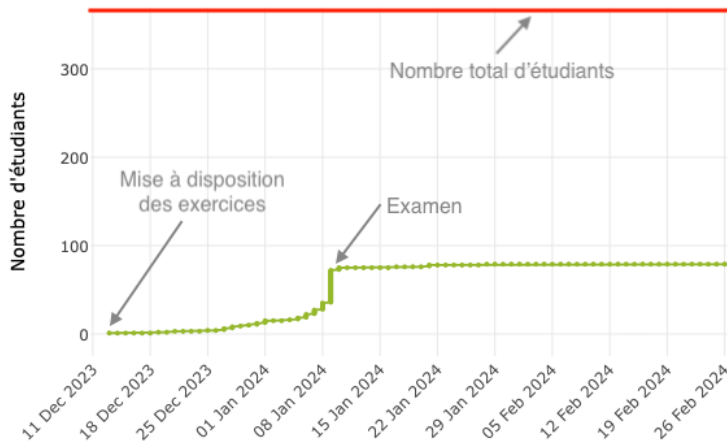
Subsection 2

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...

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`.

<- [(x)]

!

x2

x

duplicated

✓ Vérifier

Participation - des solutions...

Varier 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
```

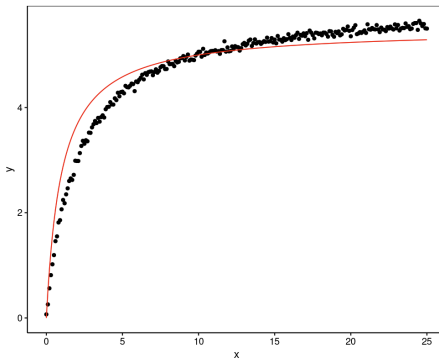
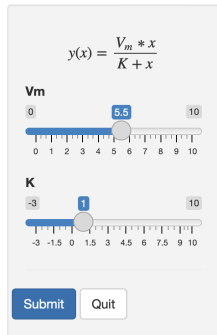
Next Topic

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...

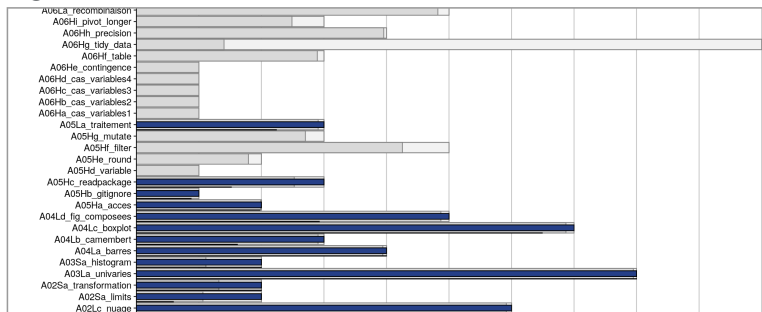
- Exercices directement dans le cours en ligne
- Liste des exercices à la fin de chaque module
- **Points attribués à la réalisation des exercices** (exemple, 5% de la note finale)



Participation - des solutions...

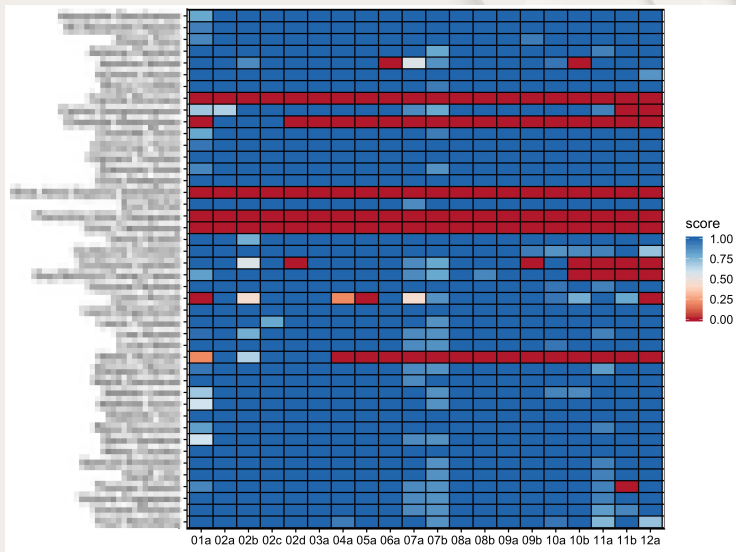
- Exercices directement dans le cours en ligne
- Liste des exercices à la fin de chaque module
- Points attribués à la réalisation des exercices (exemple, 5% de la note finale)
- **Rapport de progression** en temps réel

Progression



Participation - résultat

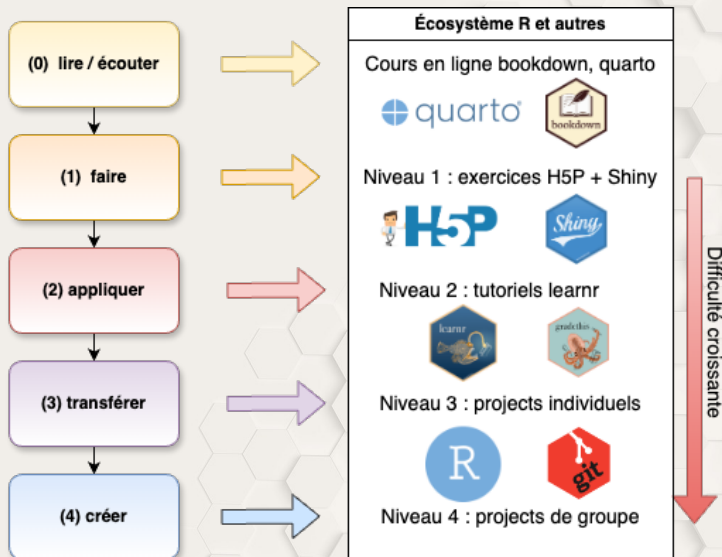
Plus de 90% de participation observée aux exercices en ligne de nos cours



Subsection 3

Progressivité de l'apprentissage

Progressivité : 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 a directory structure with files like '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

Master branch: Go to file, Add file, Code, Use this template

Commits: 9 commits by phgrosjean on Mar 13

File	Description	Time
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](#)

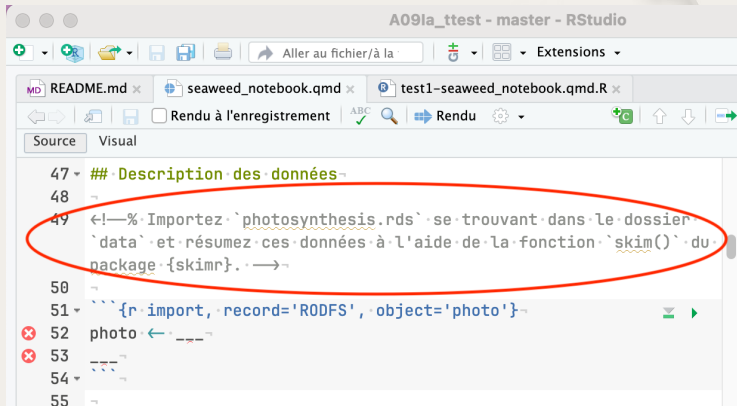
Packages: No packages published. [Publish your first package](#)

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 code editor displays R code with line numbers 47 to 55. A red circle highlights the comment on line 49, which instructs the user to import a file and use the 'skim()' function. The code is as follows:

```
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**

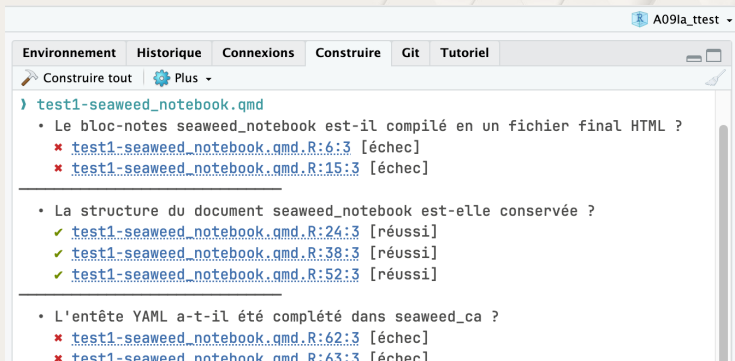
```

56 ▾ ```{r·desccomment,·output='asis'}~
57 select_answer(r"-{~
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·
reprënd·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 'A091a_ttest' session. The 'Construire' (Build) tab is active, displaying the output of a test run for 'test1-seaweed_notebook.qmd'. The output lists several test cases with their results and line numbers.

```

) 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

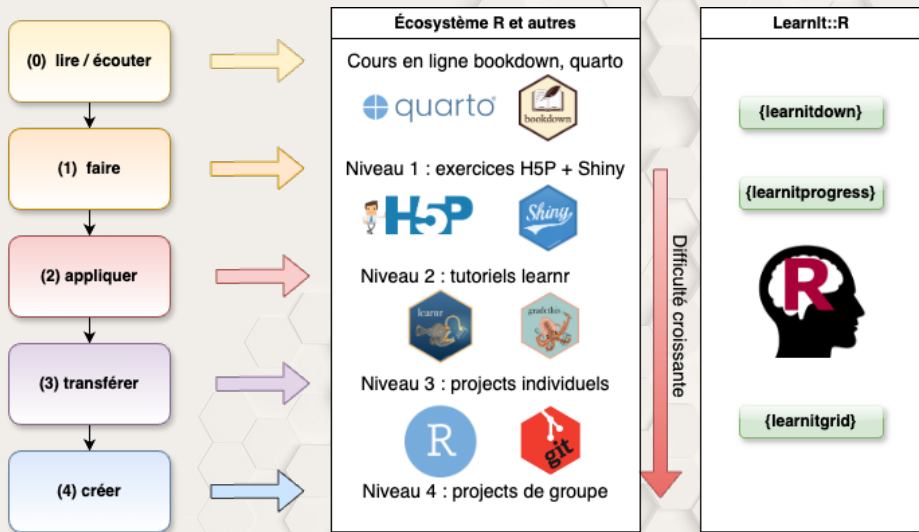
@histo_fact = histogrammes multiples de la variable feret

CSV Excel

Filter: 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>		template repo docs html (get Rmd)	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>		template repo docs html (get Rmd)	eval01	id295

Plateforme pédagogique LearnIt::R



Avez-vous des questions ?



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/>
- Cette présentation : <https://go.sciviews.org/rr2024>