

# Retour d'expérience sur l'enseignement de la science des données biologiques en classe inversée

Guyliann Engels 1\*

Philippe Grosjean 2†

## Résumé

R est enseigné dans le cursus en biologie de l'Université de Mons, Belgique. Cinq cours de sciences des données en classe inversée sont dispensés entre le deuxième bachelier et la seconde année de master. Nous utilisons le package `{learnitdown}` afin de gérer nos différents outils d'apprentissage `{bookdown}`, `h5p`, `{leanrr}`, `{shiny}` et les projets individuels ou de groupes sur GitHub. Sur l'année académique 2020-2021, les 133 étudiants ont réalisé 8800 commit et modifié plus de 300000 lignes dans les 33 projets proposés. On dénombre également plus de 300000 entrées dans la base de données associée aux activités H5P, learnR et shiny. Nous pouvons donc suivre finement la progression de chaque étudiant.

**Mots-clefs :** Science des données biologiques – suivi de l'apprentissage – classe inversée – `learnitdown`

## Développement

R est enseigné dans le cursus des biologistes de l'Université de Mons en Belgique. Depuis 2018, les cours classiques de biostatistique ont été remplacés par des cours de sciences des données biologiques donnés en classe inversée. Ces nouveaux cours ont permis d'intégrer des nouvelles notions liées à la reproductibilité, à l'Open Data ou encore à l'Open Science. Ce changement d'approche pédagogique a été de pair avec l'utilisation de nouveaux outils d'apprentissages comme `{bookdown}`, `{shiny}`, `{learnR}`,...

Les activités d'apprentissages sont découpées en 3 niveaux. D'abord, les étudiants lisent les notions théoriques illustrées dans un contexte biologique grâce à `{bookdown}` et répondent à des questions intégrées générées avec H5P ou `{Shiny}` (<https://wp.sciviews.org/>). Ensuite, les étudiants réalisent des tutoriaux `{learnR}` afin de se familiariser avec le code R associé à ces notions théoriques. Pour finir, ils réalisent des travaux individuels ou de groupe afin d'appliquer ces notions théoriques sur un cas pratique en biologie. Les travaux sont hébergés sur GitHub dans une organisation (<https://github.com/BioDataScience-Course>) et générés via GitHub Classroom.

Nous utilisons le package R `{learnitdown}` (<https://github.com/SciViews/learnitdown>) afin de gérer efficacement l'ensemble des outils employés comme une forme de plateforme d'apprentissage (*LMS Learning Management Systems*) dédié à R.

Les résultats des activités H5P, shiny et learnr sont directement collectés dans une base de données. Les projets sont analysés grâce au gestionnaire de versions git. Sur l'année 2020-2021 académique, les 133 étudiants répartis sur 3 années ont réalisé plus de 8800 commit et ont modifié plus de 300000 lignes dans les 33 projets proposés. On dénombre également plus de 300000 entrées dans la base de données associée aux activités H5P, learnR et shiny. L'analyse de ces nombreuses données nous permet de suivre finement le travail réalisé par chaque étudiant. Les étudiants ont accès à leur progression grâce à un rapport de progression individuelle. Ils peuvent également se comparer à la progression moyenne des étudiants. Ce suivi de l'apprentissage en continu combiné à une correction manuelle des projets a permis de supprimer les examens de fin d'années.

Outre l'analyse fine de la progression d'un étudiant dans un cours donné, l'analyse transversale des données sur plusieurs années vont également être très utiles afin d'adapter les activités d'apprentissages. Nous avons actuellement 3 années successives d'étudiants qui ont suivi le même cours. Les premiers outils implémentés

---

\*Service d'écologie numérique, Institut Complexys & Infortech, [guyliann.engels@umons.ac.be](mailto:guyliann.engels@umons.ac.be)

†Service d'écologie numérique, Institut Complexys & Infortech, [philippe.grosjean@umons.ac.be](mailto:philippe.grosjean@umons.ac.be)

sont les travaux sur Github et les tutoriaux learnr. Les 163 étudiants ont travaillé sur 30 projets (8700 commit, 413943 lignes modifiées) et réalisé entre 10 et 15 tutoriaux learnr (227887 entrées). La participation à ce cours est très élevée et le pourcentage de réussite à ce cours est de 80% et plus chaque année.

---

La soumission doit contenir :

1. le titre de la présentation,
2. les noms des auteurs,
3. les contacts des auteurs, incluant l’institut et le courriel,
4. le résumé de l’intervention,
5. les mots-clefs,
6. le développement,
7. d’éventuelles références (Chiquet et al. (2018)).

La soumission de deux pages maximum devra être déposée au format PDF sur le site web de la conférence.

## Références

Chiquet, Julien, Mahendra Mariadassou, Stéphane Robin, and others. 2018. “Variational Inference for Probabilistic Poisson Pca.” *The Annals of Applied Statistics* 12 (4): 2674–98.