

Introduction de nouveaux outils (learnr, Github classroom, . . .) dans un cours de Science des Données Biologiques

Guyliann Engels a et Philippe Grosjean a

a Laboratoire d'Écologie numérique des Milieux aquatiques

A-Institut Complexys

A-8 avenue du Champ de Mars, 7000 Mons, Belgique

`guyliann.engels@umons.ac.be`

`philippe.grosjean@umons.ac.be`

Mots clefs : Science des données, biologie, apprentissage, classe inversée.

Les statistiques “classiques” telles qu’enseignées dans un cursus de biologie sont de plus en plus insuffisantes dans le contexte actuel : masse de données en croissance exponentielle, nécessitant des outils spécialisés pour les aborder, crise de la reproductibilité en science [1], Open Science, Open Data, Open Knowledge (Directives européennes). Aujourd’hui, et encore plus demain, nos étudiants devront pouvoir maîtriser les outils simple nécessaires dans ce nouveau contexte. Cela dépasse les statistiques, dans un ensemble plus large nommé science des données [2].

De plus, les outils conventionnels d’apprentissage que sont les livres de référence, les syllabus avec des professeurs omniscients déversant leurs savoirs n’intéressent plus les étudiants. Les étudiants sont toujours plus connectés et chercheront leurs informations sur internet via Google, Youtube, . . .

Ces deux constats nous ont donnés les raisons nécessaires au remplacement des cours de biostatistiques “classiques” par des cours de sciences des données biologiques “Online” à l’UMONS. Ces nouveaux cours doivent à la fois donner des notions, des concepts ou encore des outils simples et performants pour appréhender les métiers de demain et proposer aux étudiants du contenus pédagogiques de qualités avec les outils qu’ils aiment utiliser pour apprendre [3].

Jusqu’à présent, le choix a été fait d’employer une machine virtuelle complètement configurée, appelée SciViews Box (<http://www.sciviews.org/>) afin d’enseigner la biostatistique à l’UMONS. Cette dernière simplifie la configuration de l’ensemble des outils logiciels que les étudiants manipulent, et permet un travail reproductible sans se soucier des problèmes d’installations, de mises à jour, des compatibilités, . . . Cette machine virtuelle est actualisée chaque année afin d’y ajouter les nouveautés proposées par R, RStudio, les nouveaux packages intéressants, . . .

L’apprentissage du langage R n’est pas aisé pour des non-informaticien légèrement déstabilisé lorsque l’on prononce la mot : “code”. La mise à leurs dispositions d’une boîte à outils comprenant une succession de menus (**snippet R Studio**) leurs donnent une impression de simplicité. Prenons par exemple, les snippets portant sur le jeu de données (dataframes) va contenir les instructions de bases permettant de réaliser du remaniement des données (sélection des colonnes, filtre sur les lignes, calcul de nouvelles variables, . . .).

Le nouveau matériel pédagogique se compose en plus d’un ouvrage en ligne de type bookdown (<http://biodatascience-course.sciviews.org/sdd-umons/>) renvoyant vers des tutoriaux vidéo (<http://go.sciviews.org/BioDataScience-videos>) et des documents interactif sous la forme de learnr (<https://github.com/BioDataScience-Course/BioDataScience>). Nous utiliserons R notebook (R Studio) pour fournir un outils de rédaction pour la recherche reproductible. Enfin, nous utiliserons également Github classrom (<https://github.com/BioDataScience-Course>) pour fournir un outils performant vers l’Open science, l’Open Data et l’Open knowledge.

Références

- [1] Baker, M. 2016. “1,500 Scientists Lift the Lid on Reproducibility.” *Nature* 533 (7604): 452–54. doi:10.1038/533452a.
- [2] Cleveland, W.S. 2001. “Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics.” *ISI Review* 69: 21–26. doi:10.1111/j.1751-5823.2001.tb00477.x.
- [3] Leek, Jeffrey T. & Peng, Roger D. 2015. “Opinion: Reproducible research can still be wrong: Adopting a prevention approach”. *Proceedings of the National Academy of Sciences*: 112-6.