

PLATEFORMES ET NOUVEAUX OUTILS POUR L'APPRENTISSAGE DE LA STATISTIQUE

Vincent Vandewalle ¹

¹ *Université de Lille, vincent.vandewalle@univ-lille.fr*

Résumé. A l'instar de nombreuses disciplines, l'apprentissage en ligne de la statistique connaît un développement croissant ces dernières années. Des supports de plus en plus nombreux et variés sont disponibles : cours, vidéos, mais aussi plus récemment MOOC et plateformes dédiées. Ces nouveaux medias ouvrent la voie à un apprentissage interactif de la *data-science*. En parallèle, les logiciels statistiques, à l'image de R, évoluent. De nouvelles fonctionnalités sont développées pour permettre de compiler facilement texte, formules mathématiques et blocs de code. Les interfaces sont simplifiées et permettent désormais à l'apprenant d'interagir ou de mettre aisément en pratique une notion pédagogique précise. Dans cette présentation, nous passerons en revue ces nouveaux outils et nous étudierons les opportunités qu'ils offrent dans l'apprentissage de la statistique à un large public.

Mots-clés. e-learning, MOOC, apprentissage de la statistique, Rmarkdown

Abstract. As in many disciplines, on-line learning of statistics has been growing steadily for last years. An ever-increasing number of varied training materials is available: courses, videos but also more recently MOOCs as well as dedicated platforms. These new media allow interactive data-science learning. At the same time, software such as R adapt. New features are developed to make it easier to mix text, mathematical formulas and blocks of code. Features are simplified and enable the learner to interact or to easily implement a precise pedagogical notion. In this presentation we will discuss about these new tools and the opportunity to use them for a wide audience in the field of statistics.

Keywords. e-learning, MOOC, learning of statistic, Rmarkdown

1 Introduction

De nos jours, l'apprentissage de la statistique nécessite non seulement l'acquisition de connaissances théoriques en mathématiques et probabilités, mais aussi de compétences en programmation informatique. En effet, avec la multiplication des données disponibles ces dernières années, la maîtrise de logiciels adaptés est désormais indispensable pour procéder aux études statistiques. L'apprentissage de la discipline a ainsi fortement évolué ces vingt dernières années. En parallèle ou en place des traditionnelles formations présentielle ou à distance, il est aujourd'hui possible de se former via de nouvelles plateformes pédagogiques.

D'une part, les MOOC (Massive Open Online Course), souvent réalisés par des établissements d'enseignement supérieur et de recherche, sont de plus en plus nombreux. D'autre part, des plateformes pédagogiques, réalisées par des organismes de formation privés sont en pleine expansion. Ces nouvelles plateformes pédagogiques sont particulièrement adaptées pour l'apprentissage des notions de base en programmation statistique. Leurs interfaces ne se limitent plus à poser des questions sous forme de QCM, mais permettent de tester en ligne différentes solutions et d'avoir un retour en temps réel sur ces dernières. En parallèle, les logiciels statistiques, tels que R, se sont dotés d'interfaces permettant à l'apprenant d'interagir facilement avec les différentes fonctionnalités. On peut maintenant aisément compiler texte, code et formules mathématiques. Avec l'émergence de ces interfaces libres et facilement paramétrables, l'utilisation des logiciels statistiques dès l'enseignement secondaire est simplifiée. Nous dresserons ici un tour d'horizon des principaux MOOC et plateformes de data-science existants. Ensuite, nous discuterons des notebook R sous Rstudio et notebook Python sous Jupyter, outils précieux tant pour l'apprentissage de la statistique que pour la restitution de données. Enfin nous discuterons de la possibilité d'utiliser ces outils dans l'enseignement secondaire.

2 MOOC et plateformes d'apprentissage de la data-science

De nombreuses plateformes générales pour l'apprentissage en ligne existent. La liste suivante n'a pas vocation à être exhaustive mais donne un aperçu des outils existants :

- OpenClassrooms¹ aussi bien en français qu'en anglais qui présente des cours gratuits ainsi que des parcours diplômants
- Udemy² qui présente des cours en Français payants
- Udacity³, en anglais qui présente aussi bien des cours gratuits que des cours payants. Code academy⁴

Ces plateformes sont principalement constituées de vidéos, supports de cours, QCM tests et projets. Les projets donnent en général lieu à une évaluation externe dans le cadre des cours payants.

On peut citer la plateforme France université numérique⁵ fondée par le Ministère de l'Enseignement Supérieur et de la Recherche en octobre 2013, et qui vise à fédérer les projets des universités et écoles françaises pour leur donner une visibilité internationale.

¹<https://openclassrooms.com>

²<https://www.udemy.com>

³<https://www.udacity.com>

⁴<https://www.codecademy.com>

⁵<https://www.fun-mooc.fr>

La période d'inscription aux cours est limitée dans le temps, et la participation aux cours peut donner lieu à des évaluations par QCM, mais aussi par des projets donnant lieu à une évaluation, ainsi qu'à des attestations de participation.

Dans le cadre de l'enseignement de la *data-science* on citera notamment la plateforme datacamp⁶ (R, SQL, Git, Shell, Feuilles de calcul) uniquement en anglais et qui dispose en particulier d'une version en ligne de l'IDE R Studio, ainsi que de Jupyter (pour Python et R aussi) et qui permettent de tester le code de manière interactive. Par ailleurs les bonnes réponses aux questions donnent lieu à un gain de points (XP) ce qui motive particulièrement les étudiants, *gamification* oblige. Des études de cas sont aussi disponibles, mais la nécessité de correction automatique limite la profondeur des questions posées, par rapport à des projets plus complets qui donneraient ensuite lieu à une correction humaine. Cependant, la plateforme peut judicieusement venir en support de l'enseignement présentiel, dans ce cas les étudiants peuvent avoir accès gratuitement pendant 6 mois à la version premium.

3 Notebook R et Python

L'émergence des plateformes pédagogiques est concomittente à l'amélioration des interfaces utilisateurs de R et Python. Elles permettent à l'apprenant d'accéder à travers un même document à la fois à des notions de cours, mais aussi à sa mise en pratique. Ce type de document est dénommé notebook. Les plus célèbres sont les notebook Jupyter, en particulier pour le langage informatique Python et plus récemment pour le langage R. L'IDE (Integrated Development Environment) Rstudio a également connu de nombreuses évolutions et permet lui aussi de produire des notebook. Au delà de l'aspect apprentissage, ces notebook permettent à l'apprenant d'avoir à disposition des outils qui lui permettent de produire des rapports reproductibles (Rmarkdown dans R par exemple : Xie Y., Allaire J. J. and Grolemund G., 2018).

L'enseignant peut, quant à lui, développer aisément des outils web pour faciliter l'apprentissage, sans connaissance approfondie en langage web, grâce à des interfaces comme R Shiny. Pour les étudiants, ces interfaces sont utiles pour l'apprentissage de la statistique mais ils pourront également les exploiter en entreprise.

4 Lien avec le secondaire

Bien que le développement des outils précédemment cités nécessite des connaissances en programmation, l'utilisation des interfaces générées peut se faire assez facilement. Des interfaces Shiny peuvent être utilisées dès le secondaire, par exemple pour la simulation

⁶<https://www.datacamp.com>

de loi de probabilités ⁷, ou la réalisation de traitements statistiques simples ⁸. Par ailleurs, les méthodes présentant un aspect ludique et axées sur la gamification, à l'image de Datacamp, sont particulièrement adaptées aux attentes des nouvelles générations en terme d'apprentissage. Le développement de l'IA (Intelligence Artificielle) devrait encore augmenter les possibilités d'apprentissage via ces nouvelles plateformes, en permettant par exemple une correction automatique de projets complexes.

Bibliographie

Xie Y., Allaire J. J., Grolemond G., (2018) *R Markdown: The Definitive Guide*, Editions Chapman & Hall/CRC.

⁷<https://mathsp.tuxfamily.org>

⁸<https://fxjollois.shinyapps.io/explore-data/>