

OUTILS DE DATAVIZ POUR L'ENSEIGNEMENT

François-Xavier Jollois

*Dpt STID - IUT Paris Descartes — LIPADE - UFR Math-Info Université Paris
Descartes, 143 Avenue de Versailles, 75016 Paris*

Résumé. Dans le cadre des enseignements autour de la dataviz du département STID, nous avons fait le choix de présenter trois façons différentes de travailler. Nous présenterons les outils utilisés et nous dresserons un comparatif de ces approches.

Mots-clés. Enseignement, Dataviz, R, Logiciel Tableau, Javascript

Abstract. As part of the STID department dataviz lessons, we have chosen to present three different ways of working. We will present the tools used and we will compare these approaches.

Keywords. Teaching, Dataviz, R, Tableau Software, Javascript

1 Introduction

Avec la multiplication des données et la venue d'outils toujours plus efficace, la visualisation de données (ou *dataviz*) est de plus en plus présente à l'heure actuelle. Et tirant avantage des navigateurs et possibilités offertes en interface Homme-Machine, les dataviz intègrent de plus en plus d'interactions, permettant à l'utilisateur de personnaliser son exploration visuelle des données. Ainsi, le créateur d'une dataviz, en plus des aspects statistiques, doit maîtriser des aspects informatique et de communication.

Partant de ce constat, nous avons introduit son enseignement dans l'ensemble des formations du département Statistique et Informatique Décisionnelle (STID) de l'IUT Paris Descartes. Nous avons fait le choix de présenter trois outils différents, chacun représentant une voie différente dans la création d'une dataviz. Nous les présentons d'abord, puis nous réalisons une comparaison sur différents aspects.

2 Outils

Le premier outil, Tableau Software [1], est un logiciel complet et payant, destiné à la dataviz, ainsi qu'au reporting. L'idée est de présenter un outil clé en main, permettant de créer des représentations visuelles assez poussées sans connaissances avancées en informatique. La philosophie ici est de simplifier au maximum le travail du créateur de la dataviz, en se concentrant sur les données et la représentation.

Le deuxième outil est l'utilisation du langage R [2] et de bibliothèques spécifiques. Nous utilisons pour cela le package `ggplot2` [3] dédiée à la représentation graphique, ainsi que le package `shiny` pour la création d'applications dynamiques, fonctionnant dans un navigateur. Ici, il est nécessaire d'avoir des connaissances assez avancées en programmation, mais en n'utilisant uniquement le langage R. L'intérêt est d'être très libre dans les choix de représentations, tout en ayant des fonctions permettant la simplification du travail de programmation.

Enfin, le langage JavaScript (ou ECMAScript) [5] est le troisième outil et le plus orienté programmation. Celui-ci permet de créer toutes sortes d'interactions, via un navigateur. En plus du langage de base, nous utilisons la bibliothèque `d3.js` [6], permettant de nombreuses représentations graphiques. Cette voie nécessite par contre de fortes compétences en informatique, et des temps de développement plus longs que pour les deux précédentes possibilités. Au contraire, nous avons ici à l'ensemble des routines graphiques ou textuelles de base, et l'imagination est la seule limite pour la création de `dataviz`.

Ces trois outils peuvent être utilisés dans un même but, mais ce ne sont pas réellement des concurrents directs. En effet, chacun est spécifique et représente une façon de faire différente. Nous dresserons dans notre présentation un tableau des avantages et inconvénients de ces trois outils. Cette liste pourra éventuellement s'élargir à des outils similaires pour chaque cas.

A travers cette présentation et ce comparatif, nous voyons qu'il existe différentes façons pour créer une `dataviz` et la diffuser. Le choix de l'une ou l'autre des solutions dépendra très fortement des contraintes humaines et logicielles, ainsi que de l'environnement dans lequel les `dataviz` produites devront être diffusées.

Bibliographie

[1] Tableau Software, <http://www.tableau.com/>.

[2] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

<https://www.R-project.org/>.

[3] H. Wickham. `ggplot2`: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

<https://CRAN.R-project.org/package=ggplot2>

[4] Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2016). `shiny`: Web Application Framework for R. R package version 0.14.2.

<https://CRAN.R-project.org/package=shiny>

[5] Ecma-262 Edition 5.1, The ECMAScript Language Specification.

<http://www.ecma-international.org>

[6] D3: Data-Driven Documents, a JavaScript library for visualizing data using web standards. <https://d3js.org/>