

Petite bibliographie

Pascal Bessonneau

06/2015

Documentation libre sur internet

Ouvrages payants

Références

Les manuels

Comme évoqué lors de la formation, le site R-project.org abrite les manuels de R. Ils sont une solide base sur R couvrant l'installation jusqu'à la création de paquets. On y trouve aussi la liste des fonctions R de base qui fait quelques milliers de pages.

Ils sont réalisés par le « noyau dur » des développeurs de R. Toutefois, en dehors de la liste des fonctions de R, ils restent assez succints.

Les documents suggérés. . .

Ils se trouvent sur la page contributed documentation.

Les ouvrages pour débiter les plus appréciés sont généralement :

- R pour les débutants, d'Emmanuel Paradis
- Brise Glace-R, traduction d'IcebreakR.
- R pour les sociologues
- ...

Ces documents ont une approche basés sur les exemples essentiellement. Ils permettent de maîtriser les fonctions de base de R mais ne permettent généralement pas d'appréhender tout le potentiel de R.

Les documents suggérés. . .

Un ouvrage offrant un plus de distance est l'ouvrage de Vincent Goulet.

Pour ceux qui ne redoutent pas l'anglais, la lecture des documents de J. Faraway et F. Harrell Jr. sont très intéressants notamment pour ceux intéressés par les méthodes de régression.

Les documents de C. Genolini sont remarquables mais nécessitent une certaine maîtrise de R.

Ouvrages spécialisés

D'autres documents sont plus spécifiques comme ceux sur l'économétrie, l'actuariat, ... Quelques références supplémentaires :

- Pour l'analyse de questionnaires conatifs et cognitifs, le site Personality Project qui abrite un ouvrage de très bonne facture en cours de rédaction.
- Oeuvre du RUG Element-R, un ouvrage sur la cartographie en français
- L'ouvrage de G. Sanchez sur le PLS Path Modeling
- TraMineR, pour l'analyse de trajectoires
- ...

Statistiques

Dans cette section sont indiqués quelques livres intéressants avec un bref commentaire. L'opinion des formateurs n'étant pas parole d'évangiles vous pourrez trouver des informations sur les livres cités ci-dessous sur la page Books de R.

Couvrant une large palette de méthodes statistiques, l'ouvrage de Crawley (2013) est un livre très intéressant qui permet de trouver rapidement comment réaliser ces méthodes avec R.

Le livre de Albert et Rizzo (2012) illustre différents méthodes à travers quelques exemples. Néanmoins sa qualité est un peu en retrait par rapport au livre de Crawley (2013).

L'ouvrage de Dalgaard (2008) est un livre d'initiation à la fois aux statistiques (simples) et à R. L'auteur fait partie des développeurs de R. Il est didactique et intéressant.

Langages et programmation

Le livre de Gentleman (2009) est un livre remarquable qui couvre le niveau débutant à avancé. Toutefois son approche est plus informatique que statistique. Il est possible que sa lecture soit déroutante si on a pas de bonnes connaissances en programmation. Le livre de Chambers (2008) est intéressant et s'adresse aussi à un public intermédiaire et averti. Il couvre notamment la manipulation de données.

L'ouvrage de Genolini (2010) est précieux. Toutefois il s'adresse à un public averti. Il s'agit de la version papier de documents librement téléchargeables sur le web (voir partie précédente). Pour les parisiens, l'auteur est dans le RUG Semin-R.

Langages et programmation

Le livre de Matloff (2011) est un must-have. Remarquable, il couvre les possibilités offertes par R en tant que langage fonctionnel. Il couvre aussi le débogage des fonctions R ainsi que la programmation objet. Mais il s'adresse à un public venant plus de l'informatique que des statistiques.

Les *blue book* de B. Ripley et Venables (2002), illustrant des exemples statistiques avec R et le *yellow book* de (B. D. Ripley & Venables, 2000) sont des classiques. Le *yellow book* ne s'intéresse qu'au langage lui-même. L'intérêt de ces livres est surtout historique.

Le must-have est le livre de (Wickham, 2015)

Graphiques

Le must-have est le livre de (Murrell, 2011). Il couvre toutes les possibilités graphiques de R. Attention à bien acheter la deuxième version infiniment plus intéressante car elle couvre les graphiques traditionnels de R (présentés lors de la formation), le paquet *grid* et les packages *Lattice* et *ggplot2*.

Graphiques

Le package *Lattice* permet de réaliser des graphiques avancés. Il permet notamment d'automatiser la création de graphiques par variables catégorielles, faire des tableaux de graphiques aisément, ... Ce paquet développé par un membre du R Core est en perte de vitesse. Son concurrent est *ggplot2* qui présente les mêmes fonctionnalités avec une qualité graphique supérieure et plus d'options.

Le paquet *Lattice* a son livre (Sarkar D., 2008) comme *ggplot2* (Wickham, 2009).

Le livre de Wickham (2009) est le manuel du package. Une approche plus didactique est proposée dans le livre de (Chang, 2012). Basé sur des exemples, il est pratique pour débiter mais l'ouvrage de (Wickham, 2009) reste la référence du langage de *ggplot2*.

A noter que le paquet *ggplot2*, a été développé dans l'esprit du livre de Wilkinson (2005). Ce livre est très intéressant pour acquérir les bonnes pratiques en matière de graphiques.

Régression

Les ouvrages de Harrell, Jr. (2001), (Faraway J. J., 2005) et Faraway J. J. (2006) sont assez anciens. Mais ce sont d'excellents livres de référence et qui couvrent les régressions linéaires, logistique, ... Outre l'utilisation de R, ce sont d'excellents livres concernant la régression¹.

Un excellent livre en français sur la régression linéaire avec R de (Cornillon & Matzner-Lober, 2011) est paru chez Springer. Il mêle théorie et pratique avec R.

Plus complexe, l'ouvrage de Sheather (2009) est très intéressant.

1. Ils sont assez techniques

Applications particulières

Dans (Lumley, 2010), le package *survey* de l'auteur est décrit à travers des exemples de sondage simple, stratifié et à plusieurs degrés. Le problème est que l'auteur a énormément travaillé sur son package et beaucoup de fonctionnalités de son package ne sont pas présents dans le livre.

Le livre de référence pour l'analyse de données spatiales et la cartographie est l'ouvrage de Bivand, Pebesma, et Gómez-Rubio (2008).

Le livre de Robert et Casella (2010) est très intéressant et illustre bien les méthodes Monte-Carlo dans R.

Pour les analyses longitudinales, l'ouvrage de Long (2012) offre beaucoup d'informations pour ceux qui sont intéressés par les mesures répétées et les études longitudinales.

Applications particulières

Le livre d'Albert (2009) couvre le problème des statistiques bayésiennes et la façon de réaliser les calculs dans R.

Le livre de Pinheiro et Bates (2000) est un classique... Un must-have pour tous ceux qui veulent travailler avec des modèles mixtes sous R. Le livre est basé sur le paquet *nlme*. Un package plus récent *lme4* existe. Les deux paquets présente beaucoup de similitudes donc... Quelques nouveautés sont présentes *lme4* mais dont le temps d'exécution est beaucoup plus lent que *nlme*.

Un livre (Zuur, Ieno, Walker, Saveliev, & Smith, 2009) plus récent, orienté écologie mais d'une grande qualité, couvre les modèles mixtes, hiérarchiques, les GLMs, ...

Applications particulières

Le livre de Mc Callum et Weston (2012) est très spécifique. Il traite de la parallélisation des calculs sous R. Il permet de découvrir les différentes possibilités pour faire du calcul parallèle sous R. Sa parution est juste antérieure à l'intégration du package *parallel* par défaut dans R. Par conséquent il couvre la version beta de *parallel*. Si la parallélisation des calculs est très aisé sous GNU/Linux, il traite aussi des méthodes plus complexes (disponibles sous Windows et sous GNU/Linux) pour faire du calcul parallèle. Pour l'analyse de questionnaire, un livre est disponible et a été écrit par un grand statisticien français. Il s'agit de l'ouvrage de Falissard (2011). Il s'adresse à un public de niveau débutant à modéré. On y trouve les codes pour réaliser des analyses factorielles (à l'anglaise, sur variables latentes) et la validation de questionnaire. Il est assez orienté vers la médecine, l'auteur étant directeur d'un laboratoire de recherche en psychiatrie.

Applications particulières

Pour le paquet FactoMineR, pour l'analyse de données à la française, deux livres sont disponibles.

Le premier sur les méthodes traditionnelles (ACP, ACM, ...) (Husson, Lê, & Pagès, 2009) et le second sur l'analyse factorielle multiple et l'analyse de données mixtes (Pagès, 2015).

Citations

- Albert, J. (2009). *Bayesian computation with r*. Springer.
- Albert, J., & Rizzo, M. (2012). *R by example*. Springer.
- Bivand, R., Pebesma, E., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with r*. Springer.
- Chambers, J. (2008). *Software for data analysis*. Springer.
- Chang, W. (2012). *R graphics cookbook*. O'Reilly.
- Cornillon, P.-A., & Matzner-Lober, E. (2011). *Régression avec r*. Springer.
- Crawley, M. (2013). *The r book*. Wiley.
- Dalgaard, P. (2008). *Introductory statistics with r*. Springer.
- Falissard, B. (2011). *Analysis of questionnaire data with r*. CRC Press.
- Faraway J. J. (2005). *Linear models with r*. Chapman and Hall.
- Faraway J. J. (2006). *Extending the linear models with r*. Chapman and Hall.
- Genolini, C. (2010). *Petit traité de programmation orientée objet sous r. programmation, construction de packages et bonnes*