

adegraphics : un package pour la représentation et l'analyse de données multivariées

A. Siberchicot^a, A. Julien-Laferrière^a, J. Thioulouse^a and S. Dray^a

^aLaboratoire de biométrie et biologie évolutive (UMR CNRS 5558)

CNRS - Université Lyon 1

43 bd du 11 novembre 1918, 69622 Villeurbanne, France

{aurelie.siberchicot, alice.julien-laferriere, jean.thioulouse, stephane.dray}@univ-lyon1.fr

Mots clefs : Analyse multivariée, Graphique, Visualisation.

Le package `ade4` [1] propose de nombreuses fonctions utilisées aussi bien pour l'analyse exploratoire de données multivariées que pour leur représentation graphique. Si la partie procédurale a été enrichie grâce au développement de nouvelles méthodes d'analyse, les améliorations concernant les fonctionnalités graphiques ont été minimales. Au fil du temps, l'utilisation du package a donc révélé une faiblesse dans la flexibilité et l'adaptabilité des représentations graphiques et une difficulté à manier des graphes de plus en plus complexes.

Pour optimiser la prise en charge des graphiques liés à `ade4`, nous venons de développer le package `adegraphics`. Ce package réimplémente et améliore l'ensemble des fonctionnalités présentes dans `ade4` en s'appuyant sur l'environnement graphique fourni par le package `lattice` [2] et en adoptant une programmation orientée objet basée sur l'utilisation du format de classe S4. Il en résulte une amélioration de la représentation, de la gestion et de la manipulation des graphiques aussi bien pour les données brutes que pour les sorties d'analyse. Pour le développeur, l'utilisation d'une structure en classes facilite la maintenance du code et offre un environnement favorable à l'implémentation de nouvelles méthodes.

La structure adoptée dans `adegraphics` permet de stocker les graphiques sous la forme de deux grandes classes d'objets : graphique simple et graphique multiple. Cette dernière classe permet de gérer dans un seul type d'objet la juxtaposition, l'insertion et la superposition de graphiques. Un grand nombre de paramètres graphiques permet de modifier *a priori* et *a posteriori* les représentations associées à ces objets (Figure 1). Les structures complexes associées aux données multivariées nous ont conduit à proposer de nouvelles fonctionnalités permettant notamment de décliner facilement un même graphique en une collection via la prise en compte d'une partition des individus en groupe ou de multiples variables ou axes issus d'une analyse multivariée. L'identification de structures dans les données multivariées et leur restitution est ainsi simplifiée.

En adoptant une présentation basée sur l'analyse de jeux de données en écologie, nous présenterons les principaux atouts du package `adegraphics`. Nous explorerons notamment la représentation des données brutes, leur traitement, la restitution graphique des résultats d'analyse et son optimisation.

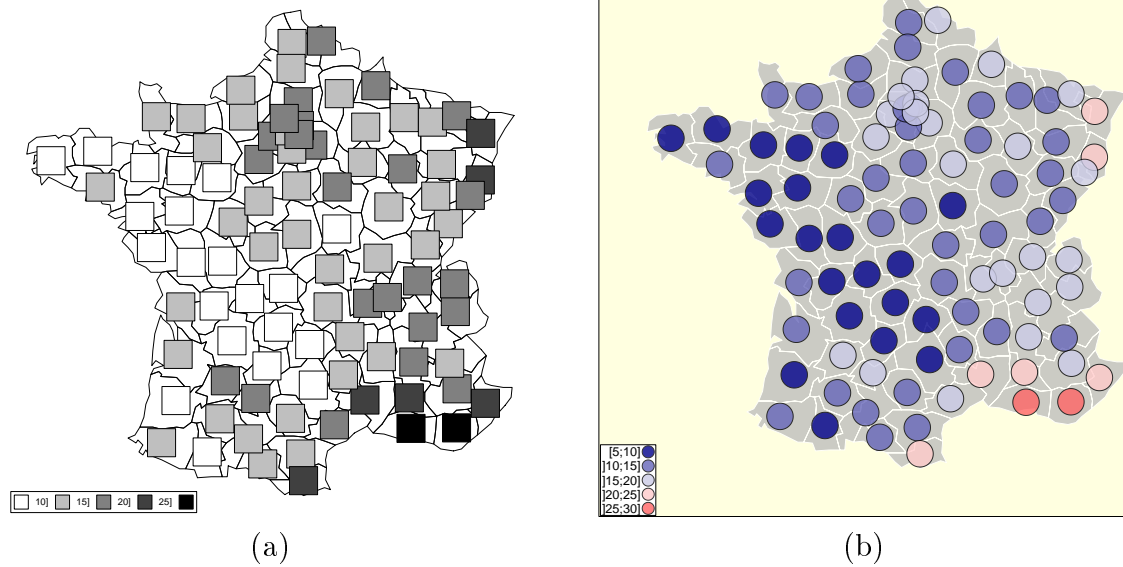


Figure 1 - Exemple de l'évolution graphique avec le package `adegraphics` : cartographie des scores du premier axe d'une analyse en composantes principales. (a) Avec `ade4`, la représentation graphique est figée (le fond de carte, la forme et la couleur des symboles et la position de la légende sont fixés). (b) Avec `adegraphics`, les différents éléments du graphique (carte, légende, symboles) sont modifiables *a priori* et *a posteriori*.

Références

- [1] Dray, S., Dufour, A.-B. (2007). The `ade4` package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**(4), 1-20
- [2] Sarkar, D. (2008). *Lattice: multivariate data visualization with R*. Springer Verlag