Meaningful representation of multivariate analysis output in R : how to solve the trade-off between amount of information and readability?

Timothée Poisot[1,*]

1. Université Montpellier 2, CNRS, Institut des Sciences de l'Évolution, 34095 Montpellier CEDEX 05, France

* Contact author: timothee.poisot@univ-montp2.fr

Keywords: PCA, graphical representation, `match`

"Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away", said the french writer Antoine de Saint-Exupery. How does it apply to graphics? Multivariate analysis have been used for a long time in ecology, because they offer a convenient way to explore the interactions between variables, or the most important factors structuring your data. However, because the purpose of such analyses is to carry the maximum amount of information, their graphical output could be amazingly difficult to read, and the reader is easily overwhelmed by an excess of information. An increasing number of R packages are dedicated to perform such analyses. However, the "out of the box" graphical output is not always easy to customise, and some users may have difficulties to present graphics just the way they want – by including some elements that are not presents by default, or by removing default elements that are useless in their case.

Using datasets from different fields — ecology, physiology, sociology, ... — I present several exemples of visual representation of the same data, and explain how the trade-off between readability and amount of information could be solved, using several ways to approach data representation. The discussion of each exemple is guided by a few questions : How can I do it? What does it tell about my data? Is it informative enough? The use and readability of colors, grey shades, type and sizes of symbols, are discussed. The point of this talk is to adress a fundamental question : How do I convey the maximum quantity of information in an easily readable graphic?

References (of some datasets used)

Poisot & Desdevises (*in revision*). Putative speciation events in *Lamellodiscus* (Monogenea, Diplectanidae) assessed by a morphometric approach. *Parasitology*.

Poisot, Šimková, Hyršl & Morand (*in revision*). Consequences of rapid water temperature increase on immunocompetence, somatic condition, and parasitism in the chub (*Leuciscus cephalus*), a freshwater cyprinid fish. *Journal of Fish Biology*.