

# Training in ade4 in R - Module I: Basic methods

## Introduction

Stéphane Dray

2021-04-19



# Material

The content of the course is available at

<https://github.com/sdray/LausanneBasic/>

In R, you can download the course by

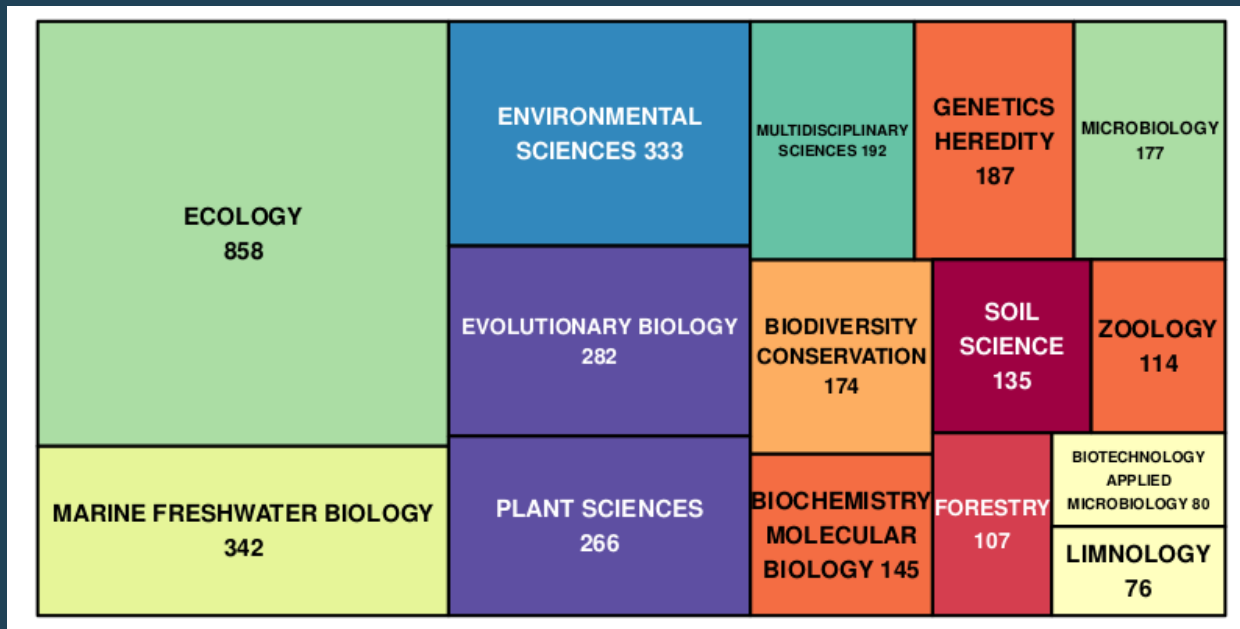
```
usethis::use_course("sdray/LausanneBasic", destdir = "~/Bureau/")
```

# Required packages

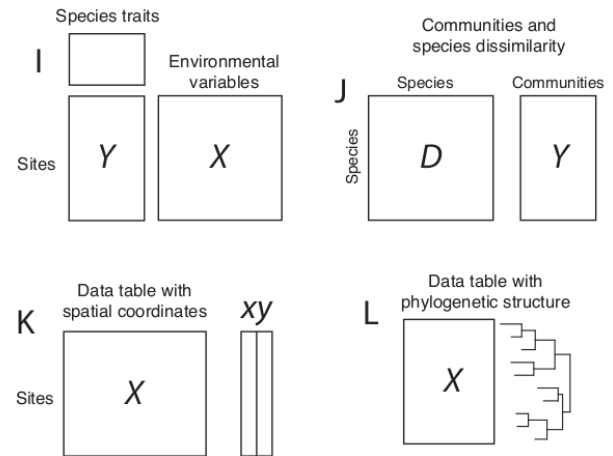
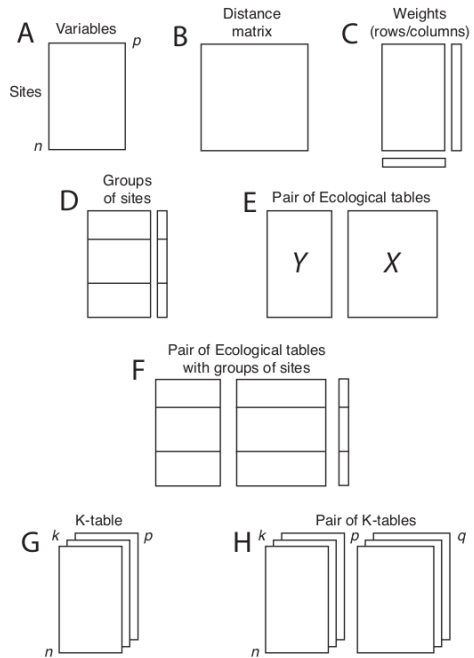
- `ade4` to run the analyses
- `adegraphics` to represent results
- `rgl` to understand multivariate methods in interactive 3D

# ade4

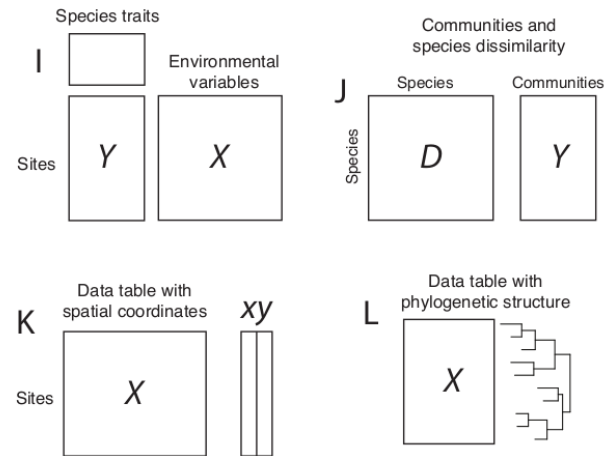
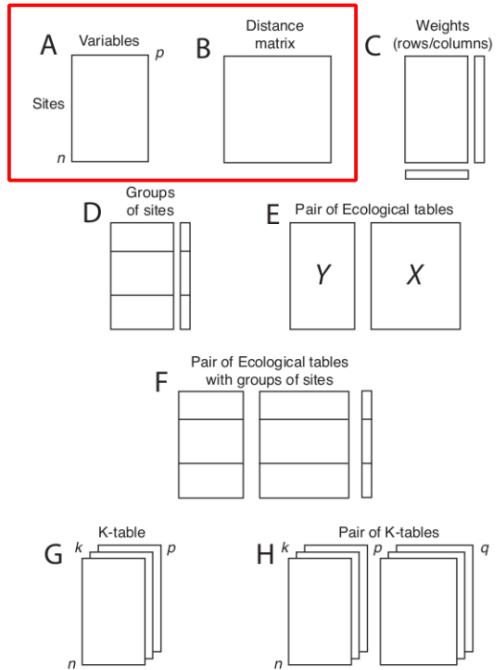
- R package since 2002
- Exploratory analysis of ecological data
  - Multivariate methods
  - Graphical functions



# Data structure



# Data structure



# Module 1: course outline

We will explore the geometric properties, outputs and interpretation of multivariate analysis focusing on one-table methods. Last afternoon for case studies.



- Environmental variables
  - Quantitative variables → Principal Component Analysis (`dudi.pca`)
  - Categorical variables → Multiple Correspondence Analysis (`dudi.acm`)
  - Mix of both → Hill-Smith Analysis (`dudi.hillsmith`)
- Species table
  - Contingency table → Correspondence Analysis (`dudi.coa`)
  - Distance matrix → Principal Coordinates Analysis (`dudi.pco`)



# ade4: the French way

IMS Lecture Notes–Monograph Series

## Multivariate Data Analysis: The French Way

Susan Holmes\*,  
*Stanford University*



*Journal of Statistical Software*

September 2007, Volume 22, Issue 4.

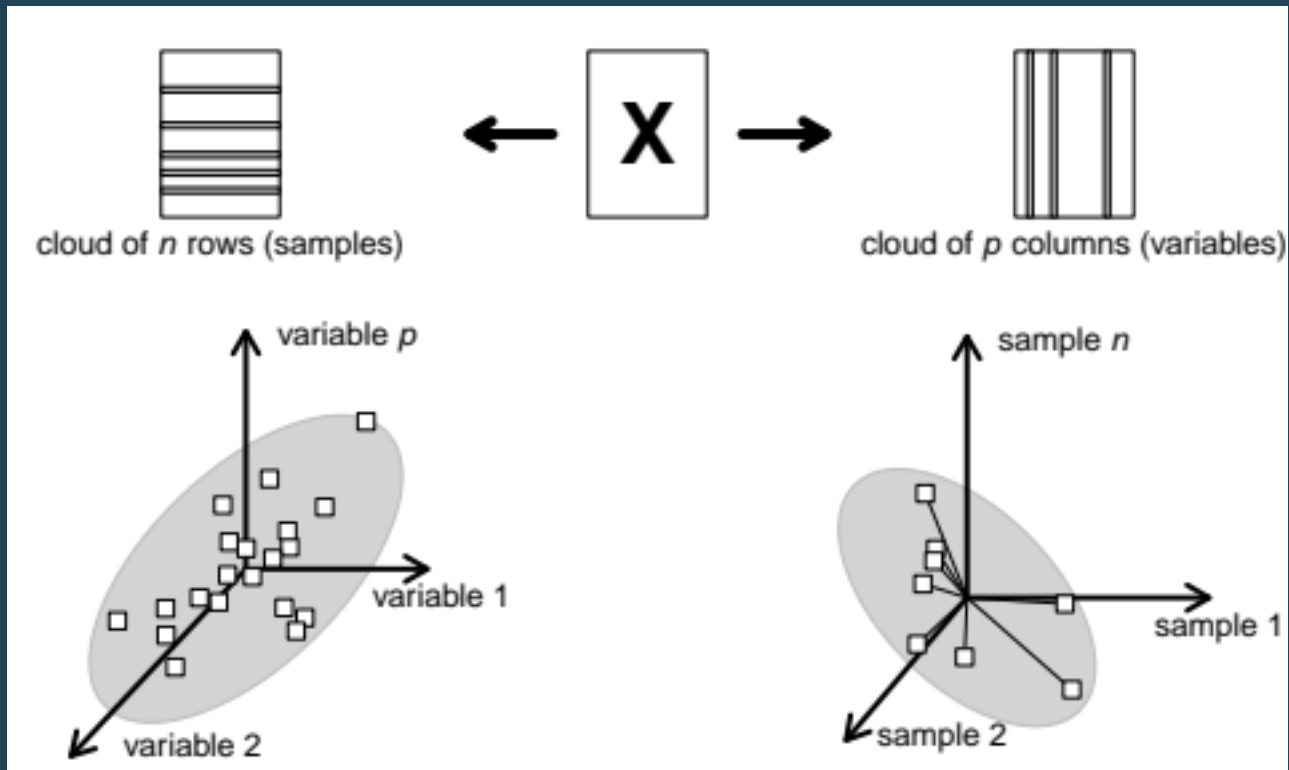
<http://www.jstatsoft.org/>

The ade4 Package: Implementing the Duality  
Diagram for Ecologists

Implementation of functions in **ade4** follows the duality diagram theory

More details are provided in the paper published in Journal of Statistical Software available [here](#)

# Two geometric views



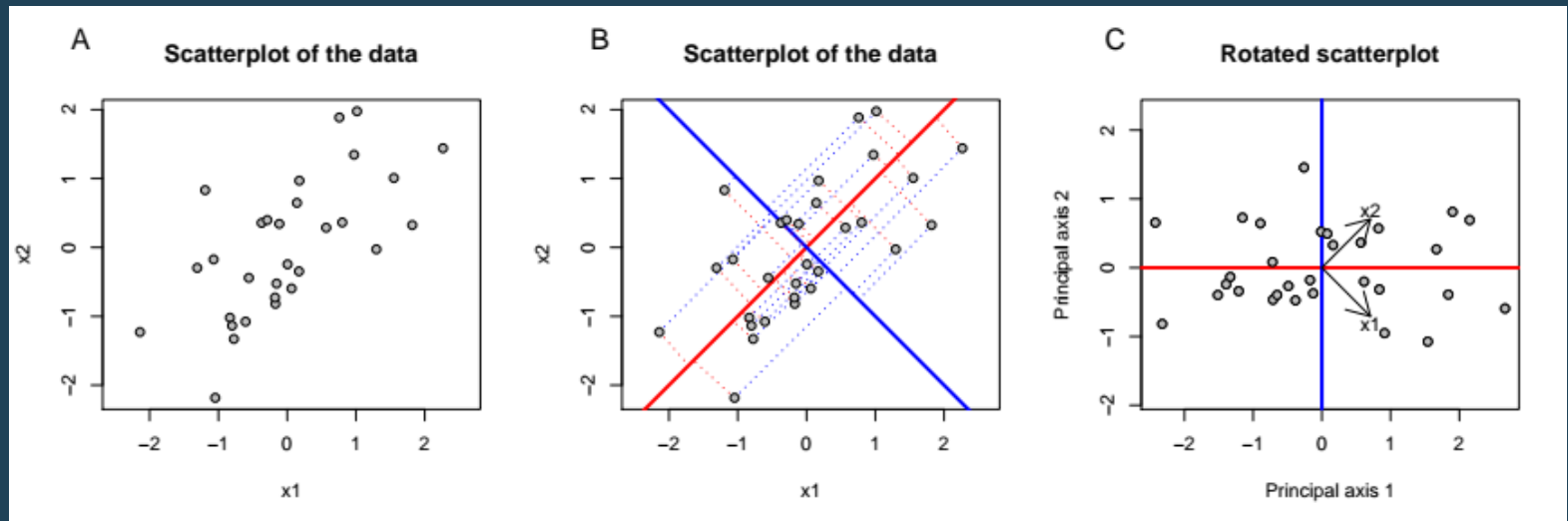
what are the main similarities and differences between the individuals ?

what are the main relationships between the variables ?

# Explore the space of individuals

[Go to practical 1](#)

# Geometric view for individuals



- Multivariate methods only perform geometric operations (rotations) to obtain the best viewpoint on the data
- When many variables are considered, dimension reduction is also applied to simplify the interpretation

# Statistical triplet

Multivariate methods aim to answer these two questions and seek for small dimension hyperspaces (few axes) where the representations of individuals and variables are as close as possible to the original ones.

To answer the two previous questions, we define

- $\mathbf{Q}$ , a  $p \times p$  positive symmetric matrix, used as an inner product in  $\mathbb{R}^p$  and thus allows to measure distances between the  $n$  individuals
- $\mathbf{D}$ , a  $n \times n$  positive symmetric matrix, used as an inner product in  $\mathbb{R}^n$  and thus allows to measure relationships between the  $p$  variables.

$$(\mathbf{X}, \mathbf{Q}, \mathbf{D})$$

# Duality diagram theory

$$\mathbf{XQX}^\top \mathbf{DK} = \mathbf{B}\mathbf{\Lambda}$$

$$\mathbf{X}^\top \mathbf{DXQ}\mathbf{A} = \mathbf{A}\mathbf{\Lambda}$$

- $\mathbf{B}$  contains the principal components (  $\mathbf{B}^\top \mathbf{DB} = \mathbf{I}_r$  ).
- $\mathbf{A}$  contains the principal axis (  $\mathbf{A}^\top \mathbf{QA} = \mathbf{I}_r$  ).
- $\mathbf{L} = \mathbf{XQ}\mathbf{A}$  contains the row scores (projection of the rows of  $\mathbf{X}$  onto the principal axes)
- $\mathbf{C} = \mathbf{X}^\top \mathbf{DB}$  contains the column scores (projection of the columns of  $\mathbf{X}$  onto the principal components)

Maximization of:

$$Q(\mathbf{a}) = \mathbf{a}^\top \mathbf{Q}^\top \mathbf{X}^\top \mathbf{DXQ}\mathbf{a} = \lambda \text{ and } S(\mathbf{b}) = \mathbf{b}^\top \mathbf{D}^\top \mathbf{XQX}^\top \mathbf{Db} = \lambda$$

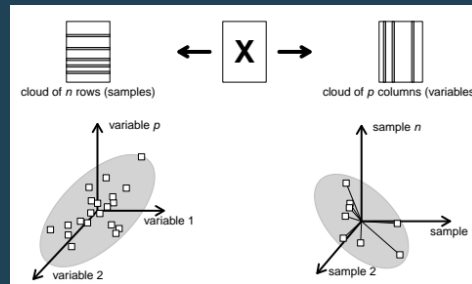
$$\langle \mathbf{XQ}\mathbf{a} | \mathbf{k} \rangle_{\mathbf{D}} = \langle \mathbf{X}^\top \mathbf{Db} | \mathbf{a} \rangle_{\mathbf{Q}} = \sqrt{\lambda}$$

# Implementation in ade4

Computations are performed by the function `as.dudi`. This function takes 3 arguments defining the statistical triplet and returns an object of class `dudi` that contains:

| <b>ade4</b>      | <b>theory</b>               | <b>Definition</b>         |
|------------------|-----------------------------|---------------------------|
| <code>tab</code> | <b>X</b>                    | (transformed) data table  |
| <code>cw</code>  | <b>Q</b>                    | inner product for rows    |
| <code>lw</code>  | <b>D</b>                    | inner product for columns |
| -----            |                             |                           |
| <code>eig</code> | <b><math>\Lambda</math></b> | eigenvalues               |
| <code>l1</code>  | <b>B</b>                    | principal components      |
| <code>c1</code>  | <b>A</b>                    | principal axes            |
| <code>li</code>  | <b>L</b>                    | row scores                |
| <code>co</code>  | <b>C</b>                    | column scores             |

# From the theory



- The principal axes

$$\mathbf{X}^\top \mathbf{D} \mathbf{X} \mathbf{Q} \mathbf{A} = \mathbf{A} \mathbf{\Lambda}$$

- The row scores

$$\mathbf{L} = \mathbf{X} \mathbf{Q} \mathbf{A}$$

- Maximization of

$$Q(\mathbf{a}) = \mathbf{a}^\top \mathbf{Q}^\top \mathbf{X}^\top \mathbf{D} \mathbf{X} \mathbf{Q} \mathbf{a} = \lambda$$

$$Q(\mathbf{a}) = \|\mathbf{X} \mathbf{Q} \mathbf{a}\|_{\mathbf{D}}^2 = \lambda$$

- The principal components

$$\mathbf{X} \mathbf{Q} \mathbf{X}^\top \mathbf{D} \mathbf{K} = \mathbf{B} \mathbf{\Lambda}$$

- The column scores

$$\mathbf{C} = \mathbf{X}^\top \mathbf{D} \mathbf{B}$$

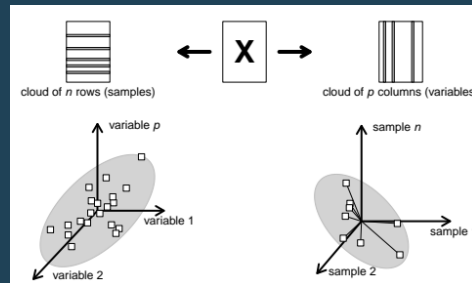
- Maximization of

$$S(\mathbf{b}) = \mathbf{b}^\top \mathbf{D}^\top \mathbf{X} \mathbf{Q} \mathbf{X}^\top \mathbf{D} \mathbf{b} = \lambda$$

$$S(\mathbf{b}) = \|\mathbf{X}^\top \mathbf{D} \mathbf{b}\|_{\mathbf{Q}}^2 = \lambda$$



# To the practice in ade4



- The principal axes

$\$c1$

- The row scores

$\$li$

- Maximization of

$\$eig$

- The principal components

$\$l1$

- The column scores

$\$co$

- Maximization of

$\$eig$

# Compute a PCA by hand

[Go to practical 1](#)

# Methods for `dudi` objects

- `print`
- `summary`
- `screeplot`
- `scatter / biplot`
- `score`
- `is`
- `t`
- `inertia`
- `suprow / supcol`
- `reconst`
- `dist.dudi`

# User-level functions

- The `as.dudi` function is an internal function
- It is called by user-friendly functions corresponding to different analyses
- It can be used by experimented users to define their own analysis

```
library(ade4)  
apropos("dudi.")
```

```
## [1] "dudi.acm"      "dudi.coa"      "dudi.dec"      "dudi.fca"  
## [5] "dudi.fpca"     "dudi.hillsmith" "dudi.mix"      "dudi.nsc"  
## [9] "dudi.pca"      "dudi.pco"      "dudi.type"
```

# Available methods

Different definitions of a statistical triplet correspond to different methods

| Function name  | Analysis name                         |
|----------------|---------------------------------------|
| dudi.pca       | Principal component analysis          |
| dudi.pco       | Principal coordinate analysis         |
| dudi.coa       | Correspondence analysis               |
| dudi.acm       | Multiple correspondence analysis      |
| dudi.dec       | Decentered correspondence analysis    |
| dudi.fca       | Fuzzy correspondence analysis         |
| dudi.fpca      | Fuzzy PCA                             |
| dudi.mix       | Mixed nalysis                         |
| dudi.hillsmith | Hill-Smith analysis                   |
| dudi.nsc       | Non-symmetric correspondence analysis |

# Graphical functions

- Outputs of multivariate methods are usually provided as plots
- `ade4` contains several graphical functions
- they have been re-implemented in a much more flexible way in the package `adegraphics`

A comprehensive overview of the package is available in its vignette available [online](#) or in R by:

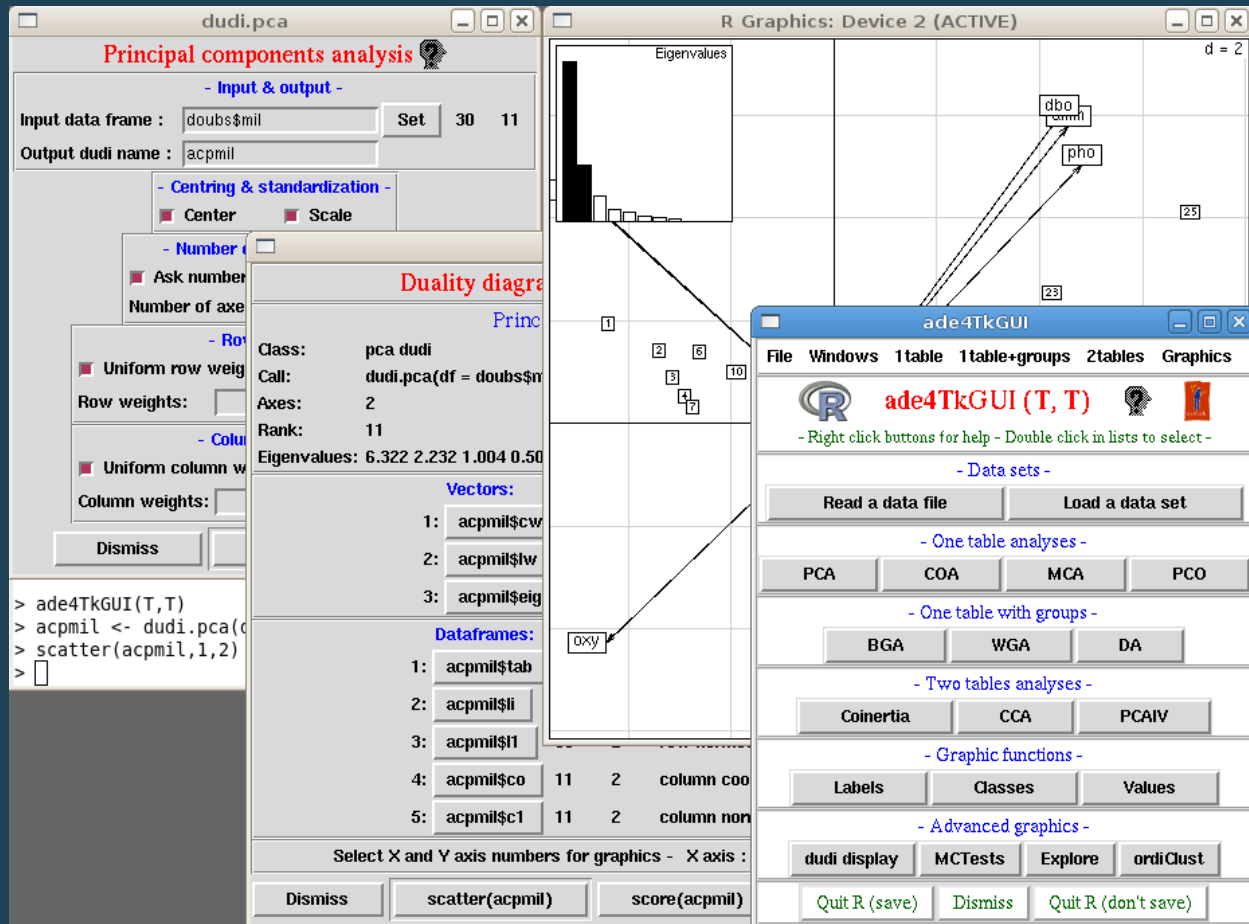
```
vignette("adegraphics")
```

See also the paper published in the R Journal [here](#)

# The ade packages

- `adegraphics`: S4-lattice based multivariate graphics
- `adespatial`: spatial multiscale multivariate analysis
- `adiv`: analysis of diversity
- `adehabitat`: analysis of habitat selection by animals
- `adegenet`: classes and methods for the multivariate analysis of genetic markers
- `adephylo`: exploratory analyses for the phylogenetic comparative method
- `ade4TkGUI`: graphical interface

# ade4TkGUI





# Resources



<https://www.springer.com/fr/book/9781493988488>

- Mailing list:  
<http://listes.univ-lyon1.fr/wws/info/adelist>
- Development:  
<https://github.com/sdray/ade4>
- Courses (in French):  
<http://pbil.univ-lyon1.fr/R/enseignement.html>