

# Two-table methods

## Discrimination and co-ordination

Stéphane Dray

2024-01-08



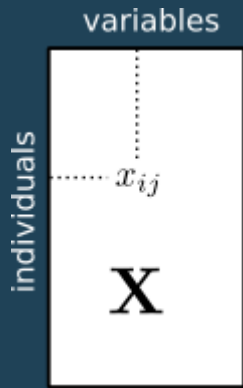
# Introduction

```
library(ade4)
library(adegraphics)
adegpar(paxes.draw = TRUE, pbackground.col = "lightgrey",
        pgrid.col = "white")
data(doubs)
str(doubs, max.level = 1)
```

```
## List of 4
## $ env      : 'data.frame':   30 obs. of  11 variables:
## $ fish     : 'data.frame':   30 obs. of  27 variables:
## $ xy       : 'data.frame':   30 obs. of  2 variables:
## $ species: 'data.frame':   27 obs. of  4 variables:
```

- Species table: abundance of 27 fish species recorded for 30 sites on the Doubs river
- Environmental table: 11 environmental variables for the same sites
- Spatial coordinates

# Introduction



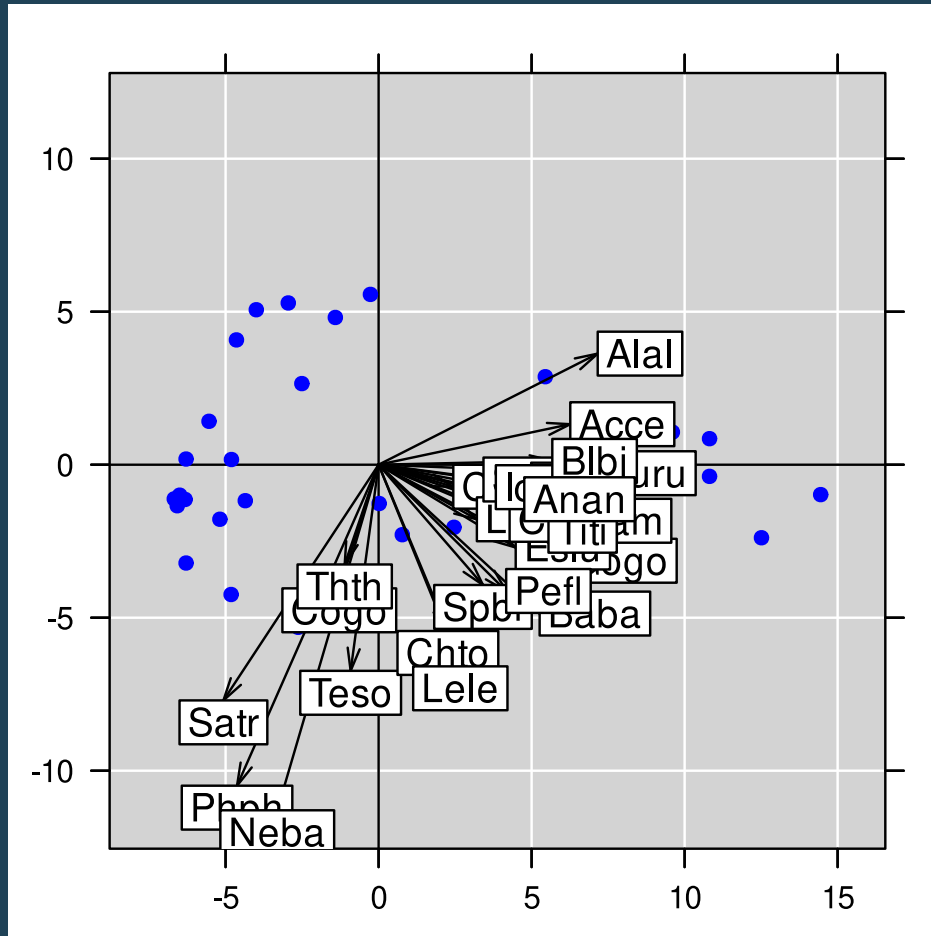
One table with  $p$  variables measured on  $n$  individuals

- quantitative (`dudi.pca`)
- categorical (`dudi.coa` or `dudi.acm`)
- both (`dudi.mix` or `dudi.hillsmith`)

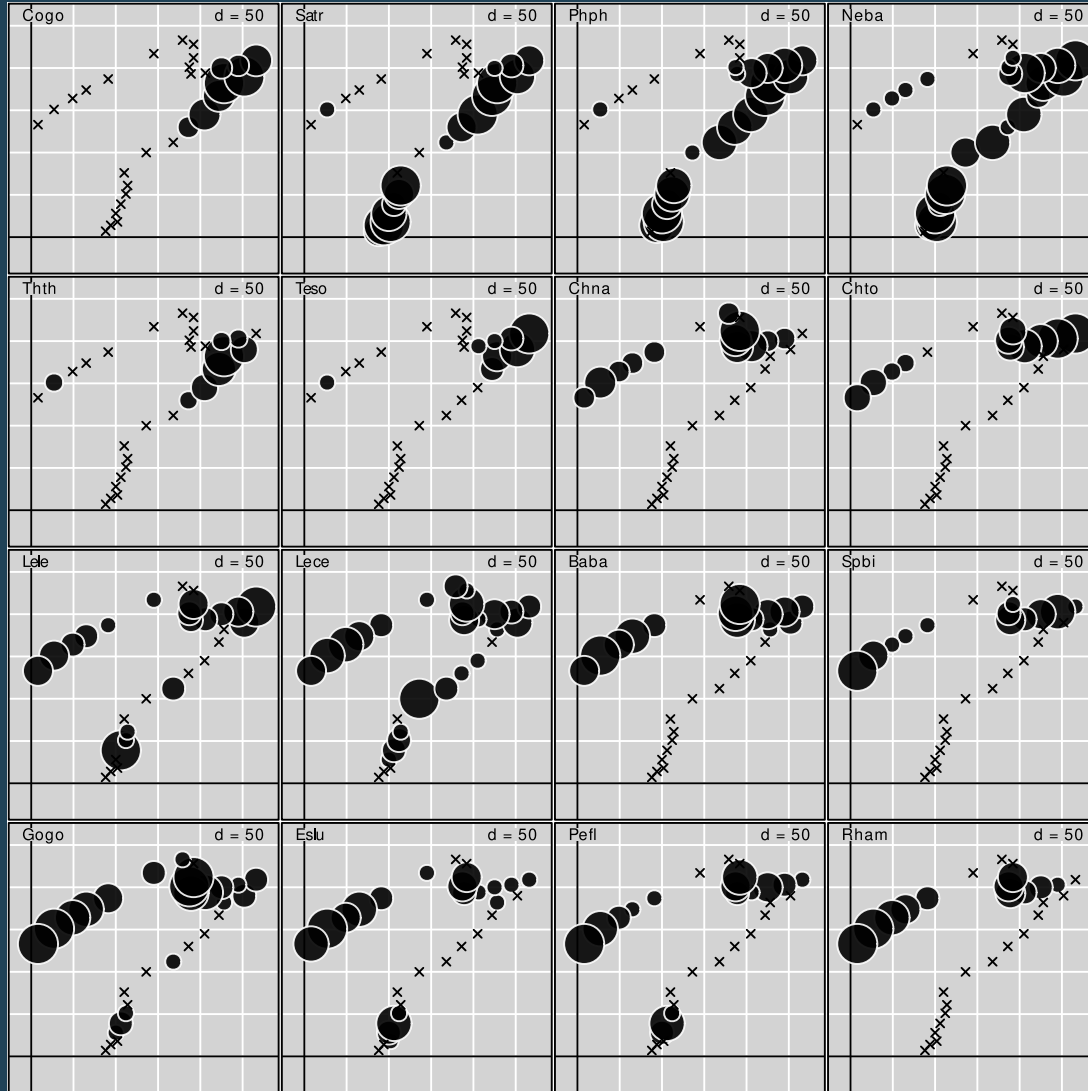
Describe the information contained in the table:

- Identify which variables are the most linked
- Identify the differences/similarities between individuals

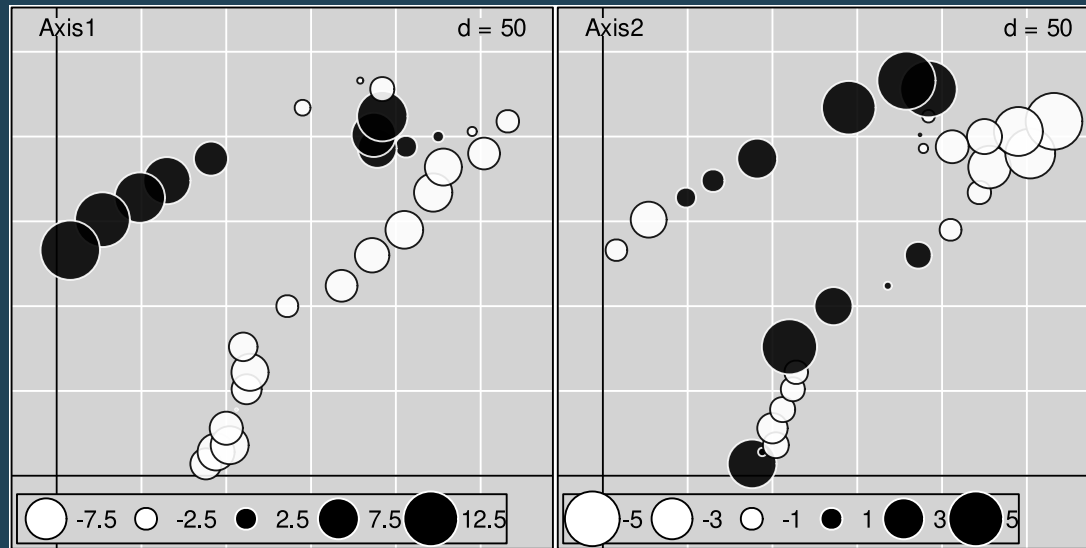
```
pca_fish <- dudi.pca(doubs$fish, scale = FALSE, scannf = FALSE)
biplot(pca_fish, ppoints.col = "blue", row.label.cex = 0,
       posieig = "none")
```



```
s.value(doubs$xy, doubs$fish[, 1:16], centerpar = TRUE,
        symbol = "circle", plegend.drawKey = FALSE, paxes.draw = FALSE)
```

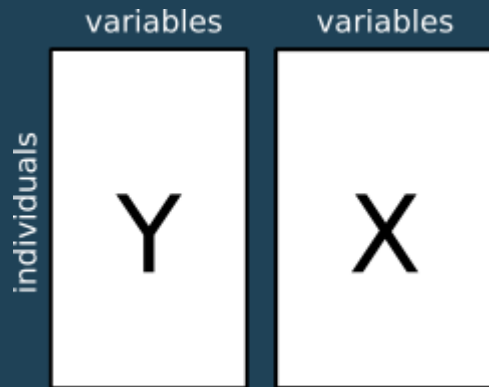


```
s.value(doubs$xy, pca_fish$li, symbol = "circle",  
paxes.draw = FALSE)
```



Which factors can explain these structures?

# Two tables



One table with  $p$  variables measured on  $n$  individuals

One table with  $m$  variables measured on the same individuals

Identify what is the information common to both tables:

- Identify which variables are the most linked
- Identify how this link varies for individuals

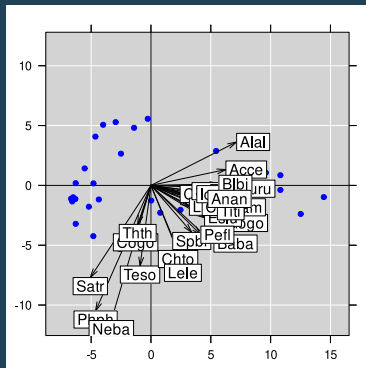
## But how?



# Indirect ordination

Y

Summarize by a simple ordination method



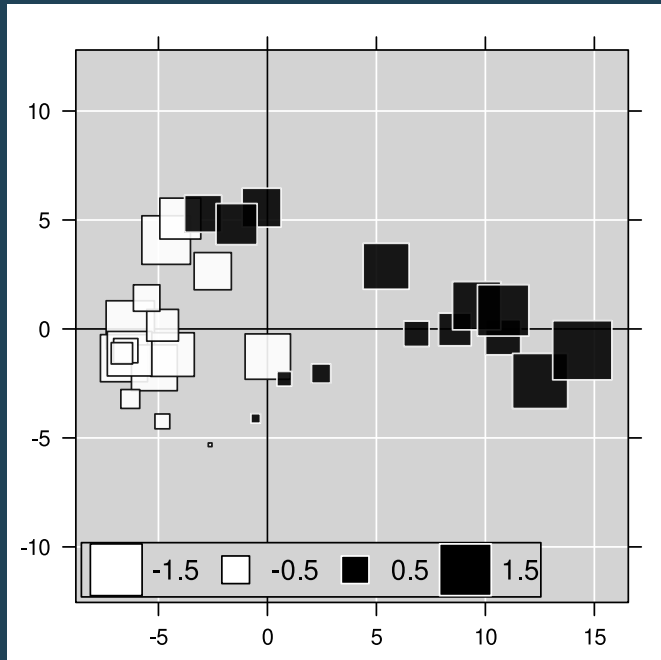
X

and interpret the structures with the second table using simple tools

- Graphical interpretation
- Correlation
- Regression

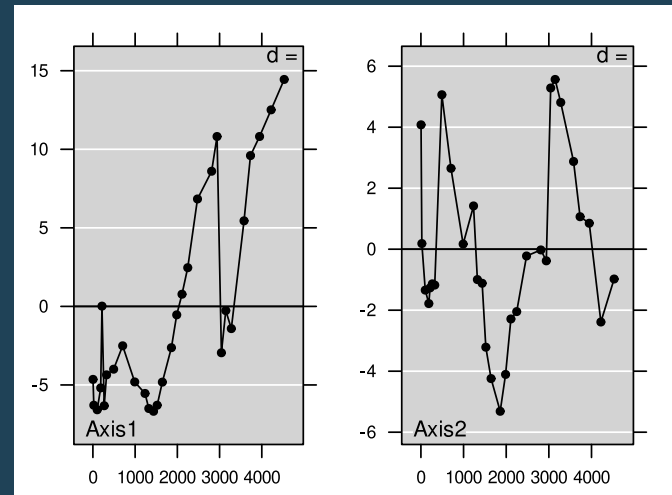
# Graphics

```
s.value(pca_fish$li,  
        scale(doubs$env$dfs))
```



# Correlation

```
s1d.curve(pca_fish$li,  
          p1.horizontal = FALSE,  
          at = doubs$env$dfs)
```

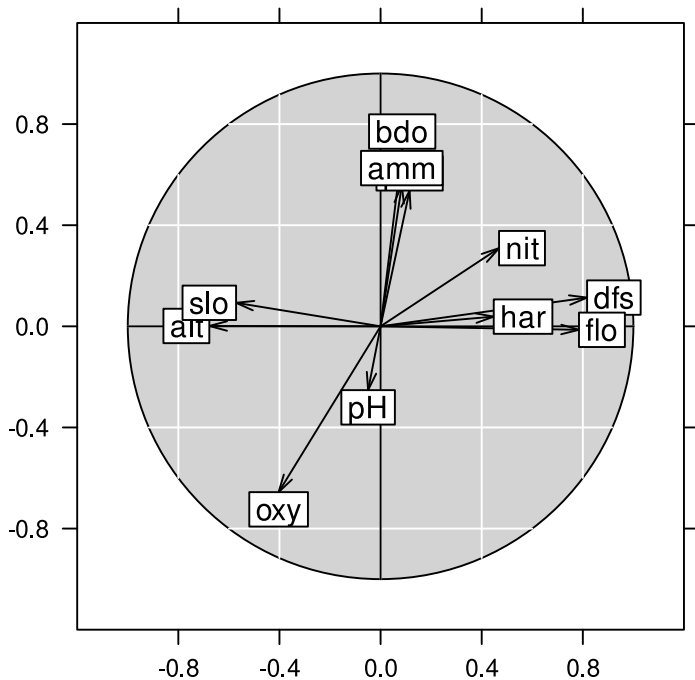


```
cor(pca_fish$li, doubs$env$dfs)
```

```
##           [,1]  
## Axis1 0.8169072  
## Axis2 0.1136704
```

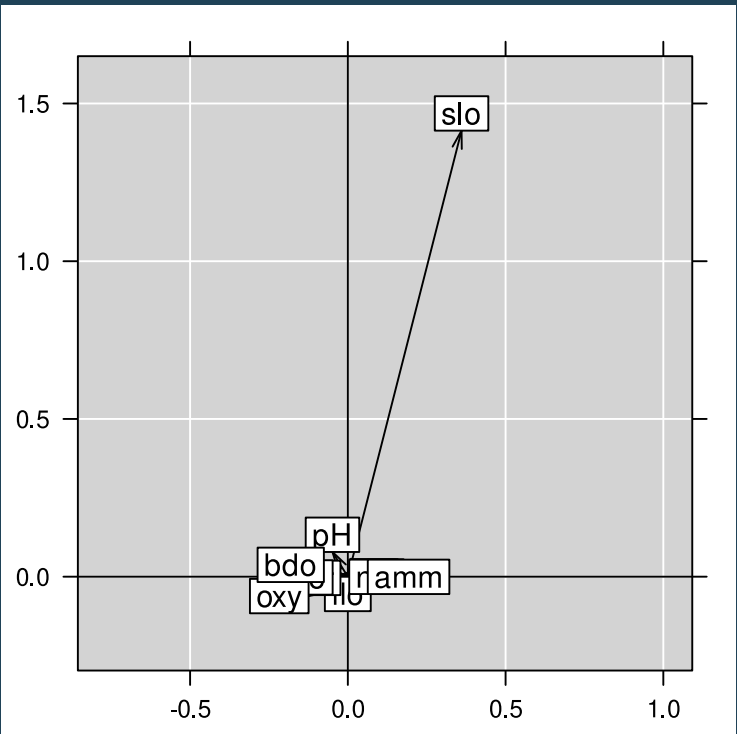
# Correlation

```
s.corcircle(cor(doubs$env,  
pca_fish$li))
```

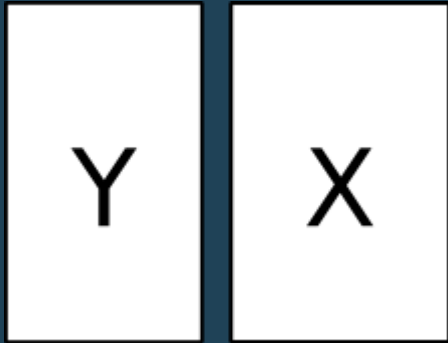


# Regression

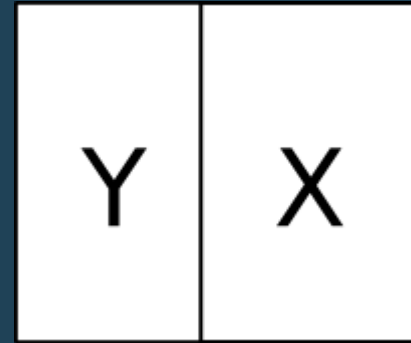
```
lm1 <- lm(as.matrix(pca_fish$li)  
as.matrix(doubs$env))  
s.arrow(coefficients(lm1)[-1,  
], labels = names(doubs$env))
```



# Analysis of the concanated table

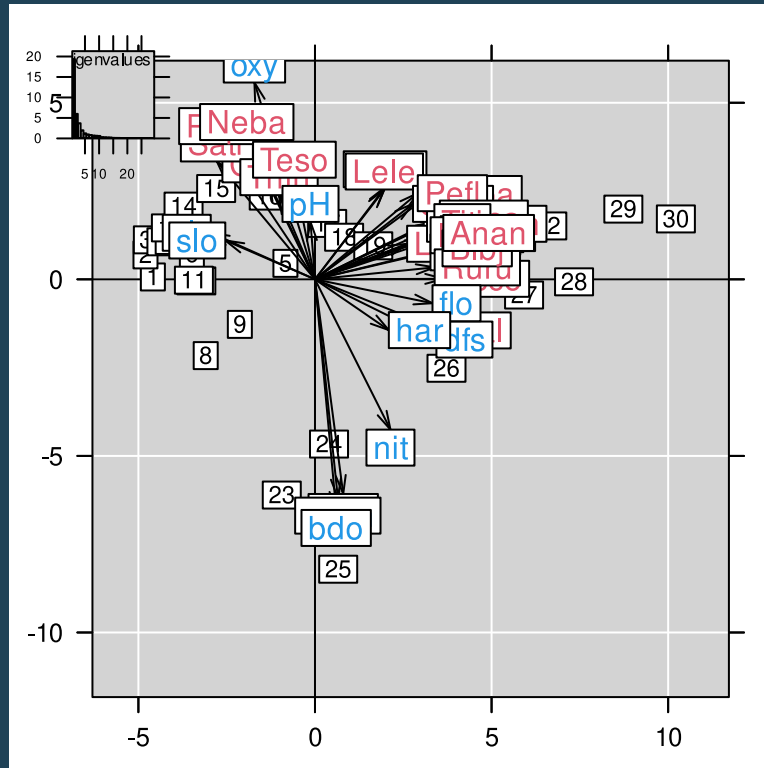


Concatenation

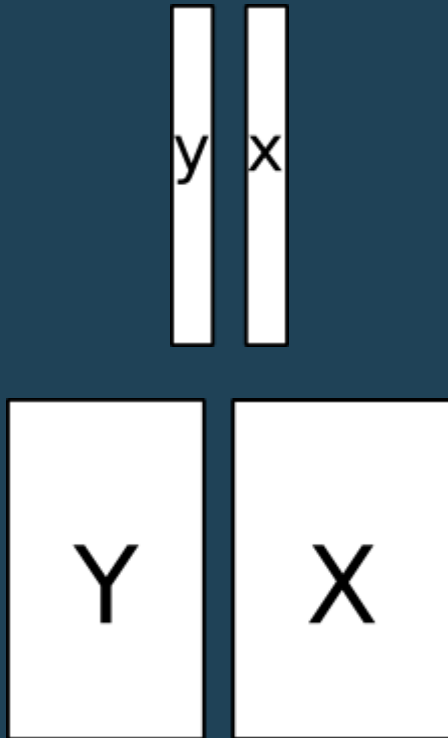


Simple ordination

```
conc <- cbind(doubs$fish, doubs$env)
biplot(dudi.pca(conc, scannf = FALSE), col.label.col = rep(c(2,
  4), c(27, 11)))
```



# From two variables to two tables



- Canonical Correlation Analysis (correlation)
- Co-Inertia Analysis (covariance)
- Analysis on instrumental variables (regression)

# Linear combinations

## PCA

Find a vector of coefficients  $\mathbf{a} = [a_1, \dots, a_p]^\top$  maximizing  $\|\mathbf{XQa}\|_{\mathbf{D}}^2$  with  $\|\mathbf{a}\|_{\mathbf{Q}} = 1$

## Linear regression

Find a vector of coefficients  $\mathbf{a} = [a_1, \dots, a_p]^\top$  maximizing  $\|\mathbf{Xa}\|_{\mathbf{D}}^2$

## Two-table methods

Define two vectors of coefficients  $\mathbf{a}$  and  $\mathbf{b}$  to build two linear combinations  $\mathbf{XQa}$  and  $\mathbf{YMb}$

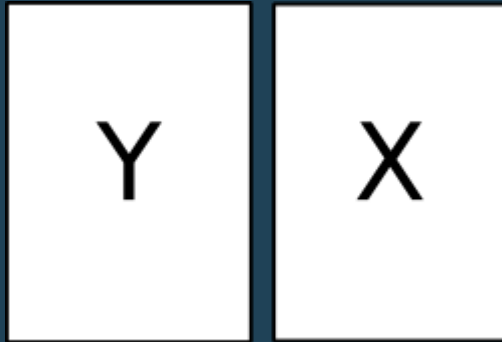
# Canonical correlation analysis



- function `stats:canoncor`
- maximization of  $cor(\mathbf{Xa}, \mathbf{Yb})$



# Co-inertia analysis

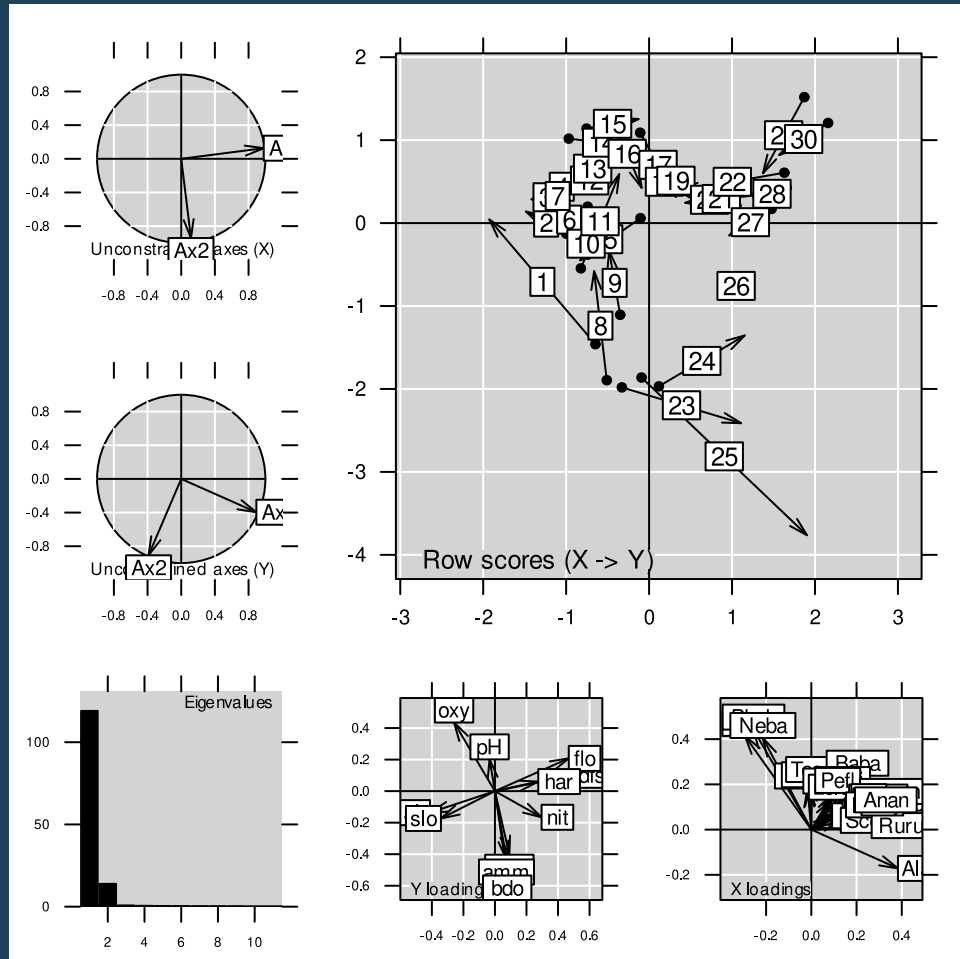


- function `ade4::coinertia`
- maximization of  $cov(\mathbf{XQa}, \mathbf{YMb})$

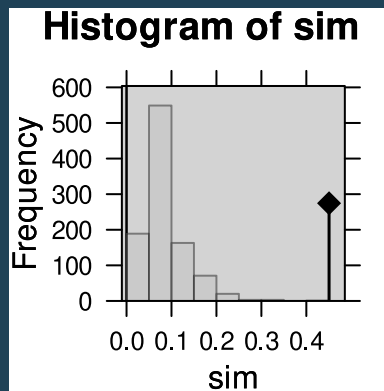
A compromise between canonical correlation analysis, and the two PCAs:

$$cov(\mathbf{XQa}, \mathbf{YMb}) = cor(\mathbf{XQa}, \mathbf{YMb}) \cdot \|\mathbf{XQa}\|_D \cdot \|\mathbf{YMb}\|_D$$

```
pca_env <- dudi.pca(doubs$env, scannf = FALSE)
coi <- coinertia(pca_fish, pca_env, scannf = FALSE)
plot(coi)
```

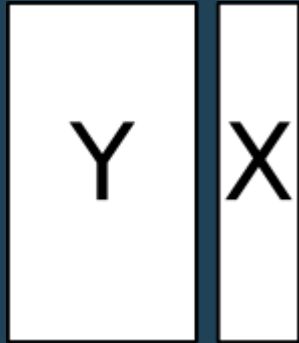


```
rt.coi <- randtest(coi)
plot(rt.coi)
```



Try `summary(coi)`

# Analysis on instrumental variables

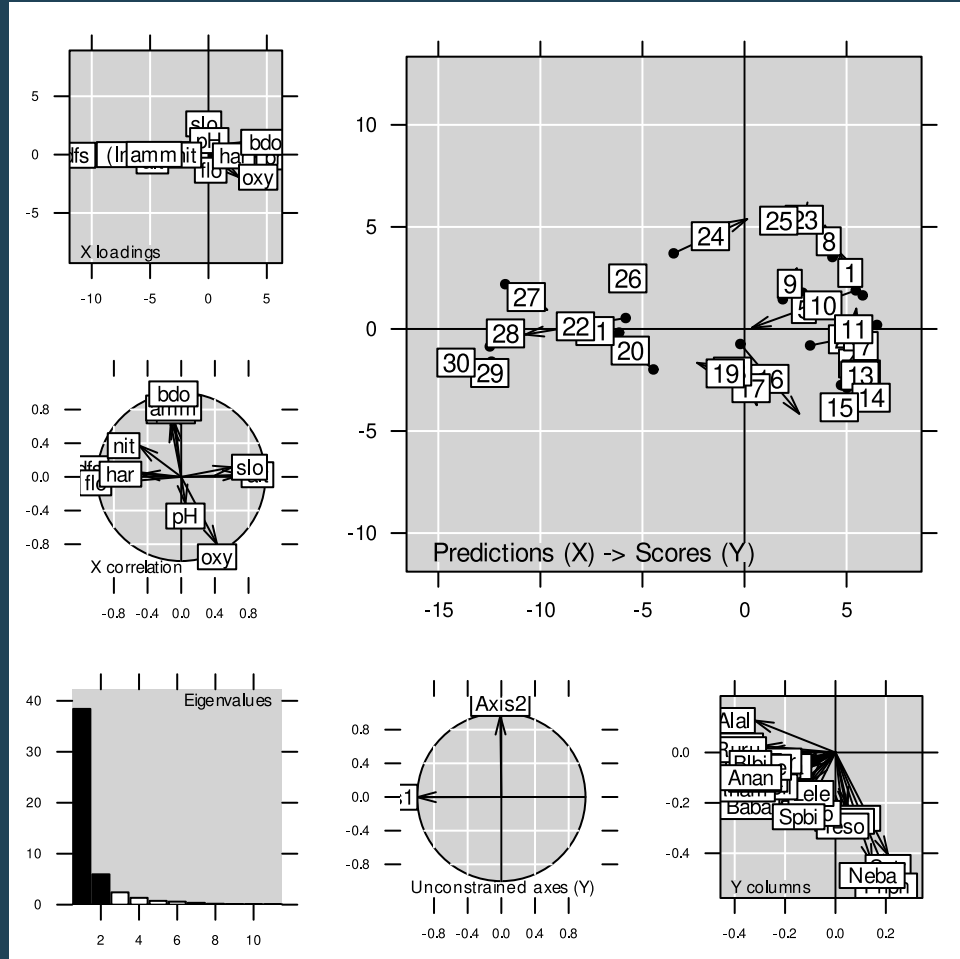


- function `ade4::pcaiv`
- maximization of  $\left\| \widehat{\mathbf{Y}} \mathbf{Q} \mathbf{a} \right\|_{\mathbf{D}}^2$

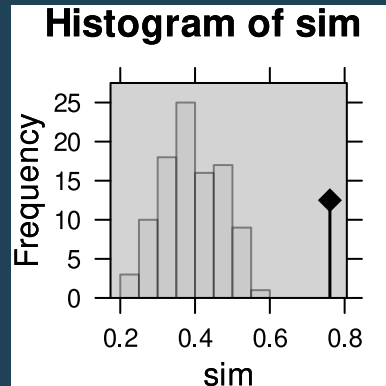
This framework encompasses:

- Redundancy Analysis
- Canonical Correspondence Analysis

```
rda1 <- pcaiv(pca_fish, doubts$env, scannf = FALSE)
plot(rda1)
```



```
rt.rda <- randtest(rda1)
plot(rt.rda)
```

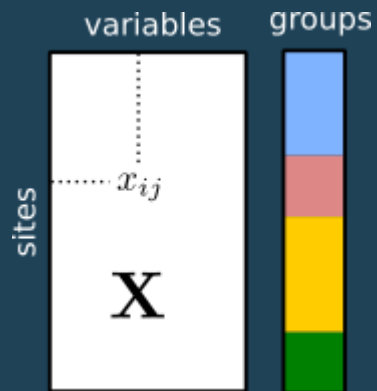


Try `summary(rda1)`

When a PCA is applied on the response table, the method is called Redundancy Analysis

Try to apply a correspondence analysis (`dudi.coa`) instead of a PCA on `doub$fish`. You obtain Canonical Correspondence Analysis.

# One particular case



The second table contains only one categorical variable

The aim is to maximize the separation between the groups

- `cancor`-like method is Discriminant Analysis (`ade4::discrimin`)
- `pcaiv`-like method is Between-Class Analysis (`ade4::bca`)