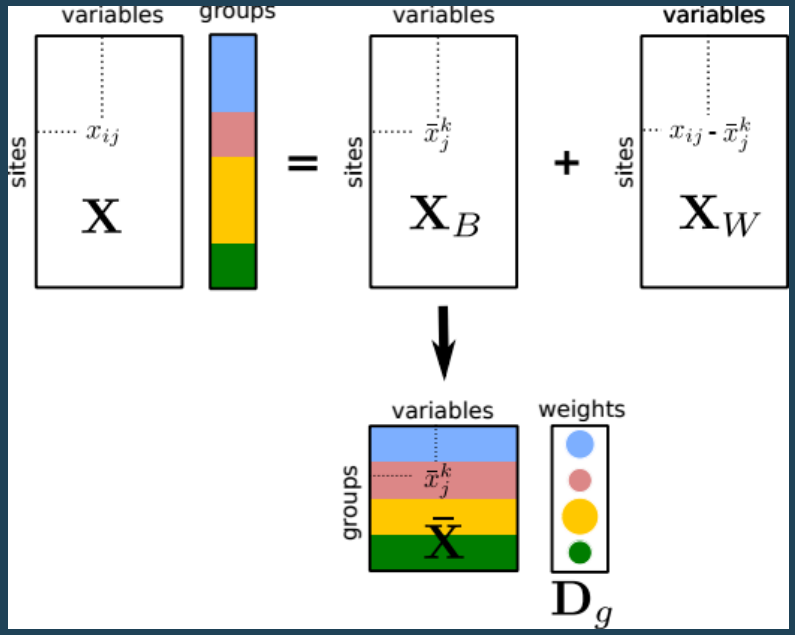
Methods partitioning individuals

in practice

Stéphane Dray

2025-10-19

Strategies of decomposition



- Within-class analysis focuses on $\mathbf{X}_W$
- Between-class analysis focuses on $\mathbf{X}_B$ maximizing B
- Discriminant analysis focuses on $\mathbf{X}_B$ maximizing B/T

Within-Class Analysis

Perform the analysis

```
library(ade4)
library(adegraphics)
data(meau)
pca_env <- dudi.pca(meau$env, scannf = FALSE)
wca.season <- wca(pca_env, meau$design$season, scannf = FALSE)
```

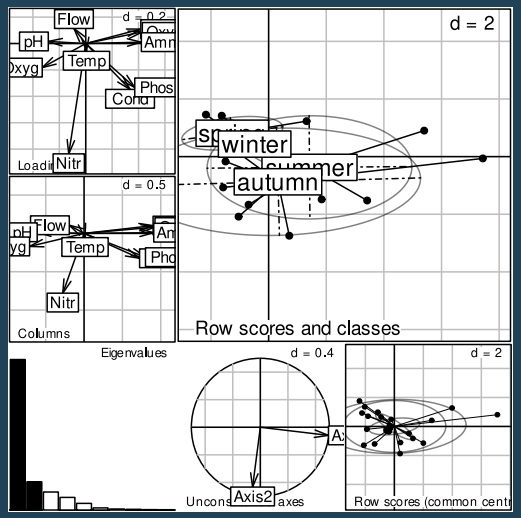
Have a look to the summary

```
summary(wca.season)
```

```
## Within-class analysis
##
## Class: within dudi
## Call: wca.dudi(x = pca_env, fac = meau$design$season, scannf = FALSE)
##
## Total inertia: 6.814
##
## Eigenvalues:
##     Ax1     Ax2     Ax3     Ax4     Ax5
##  4.6505  0.8701  0.5565  0.3900  0.2055
##
## Projected inertia (%):
##     Ax1     Ax2     Ax3     Ax4     Ax5
##  68.248  12.769   8.167   5.724   3.015
##
...
```

Plot the results

```
g1 <- plot(wca.season)
```

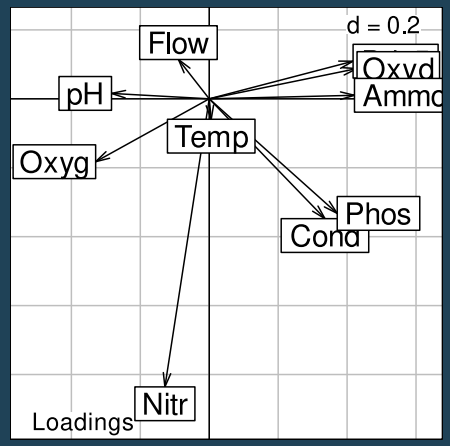



```
names(g1)
```

```
## [1] "loadings" "col"      "eig"      "row"      "Xax"      "ccrow"
```

Loadings for variables

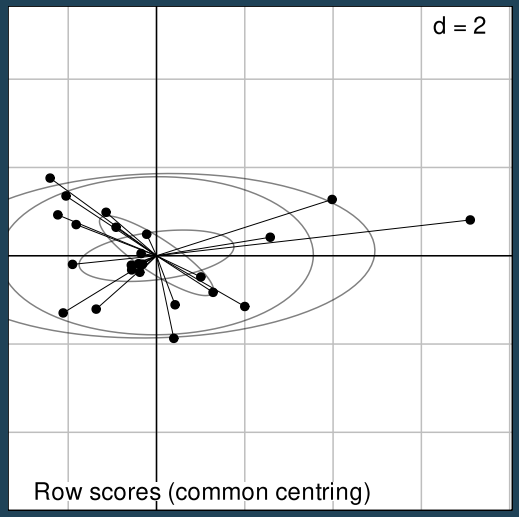
```
g1$loadings
```



$\mathbf{A}$: coefficients (loadings) for the variables of $\mathbf{X}_W$ (`wca.season$c1`)

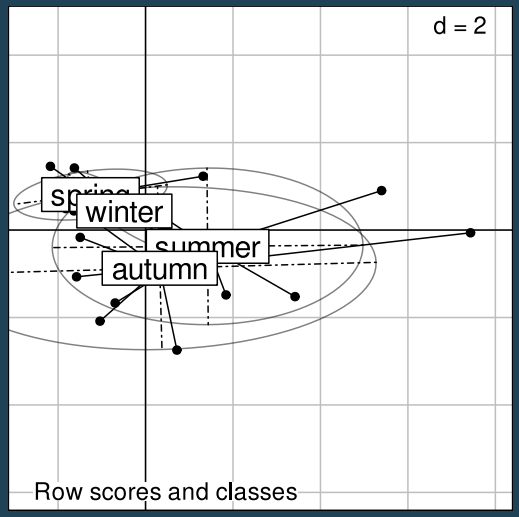
Scores for individuals

```
g1$ccrow
```



$\mathbf{X}_W\mathbf{QA}$: scores of individuals (`wca.season$li`)

```
g1$row
```



$\mathbf{XQA}$: projections of individuals (`wca.season$ls`)

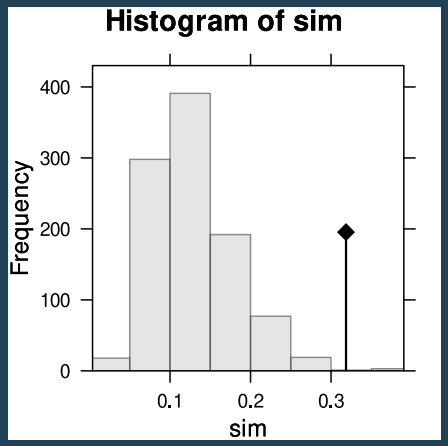
Between-Class Analysis

Perform the analysis

```
bca.season <- bca(pca_env, meau$design$season, scannf = FALSE)
```

Test the significance of the link

```
rt.bca <- randtest(bca.season)
plot(rt.bca)
```




```
rt.bca
```

```
## Monte-Carlo test
## Call: randtest.between(xtest = bca.se
##
## Observation: 0.3185858
##
## Based on 999 replicates
## Simulated p-value: 0.005
## Alternative hypothesis: greater
##
##     Std.Obs Expectation    Variance
## 3.750947355 0.129063528 0.002552928
```

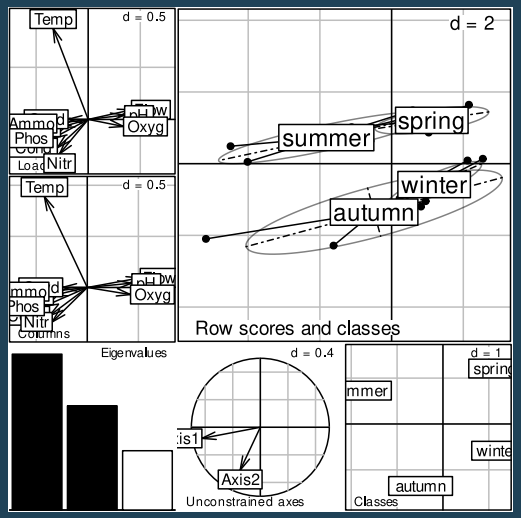
Have a look to the summary

```
summary(bca.season)
```

```
## Between-class analysis
##
## Class: between dudi
## Call: bca.dudi(x = pca_env, fac = meau$design$season, scannf = FALSE)
##
## Total inertia: 3.186
##
## Eigenvalues:
##     Ax1     Ax2     Ax3
##  1.5551  1.0390  0.5918
##
## Projected inertia (%):
##     Ax1     Ax2     Ax3
##   48.81   32.61   18.57
##
...
```

Plot the results

```
g1 <- plot(bca.season)
```

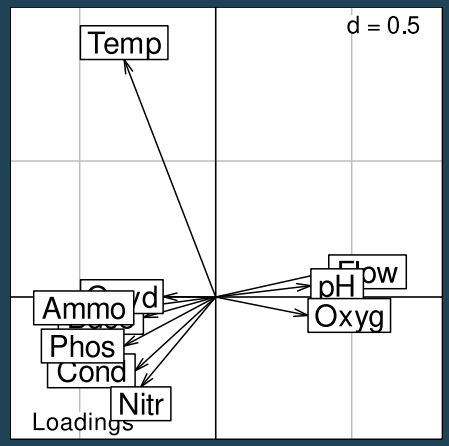



```
names(g1)
```

```
## [1] "loadings" "col"      "eig"      "row"      "Xax"      "class"
```

Loadings for variables

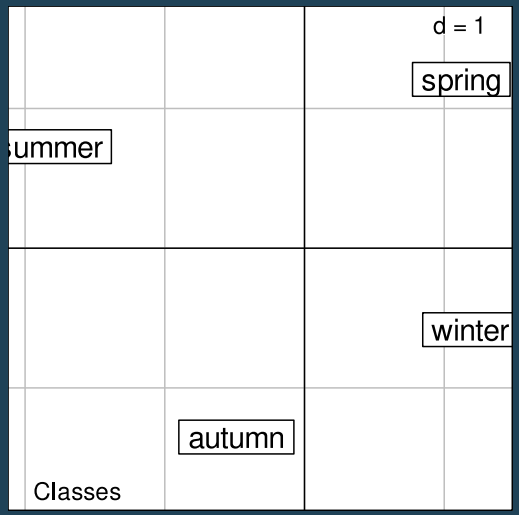
```
g1$loadings
```



$\mathbf{A}$: coefficients (loadings) for the variables of $\mathbf{X}_B$ (`bca.season$c1`)

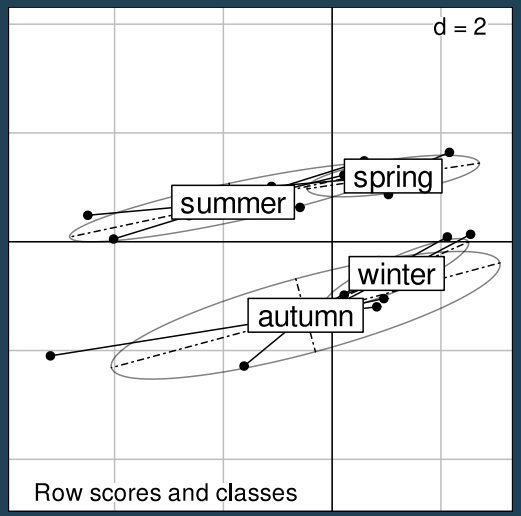
Scores for individuals and classes

```
g1$class
```



$\mathbf{X}_B\mathbf{QA}$: scores of classes
(`bca.season$li`)

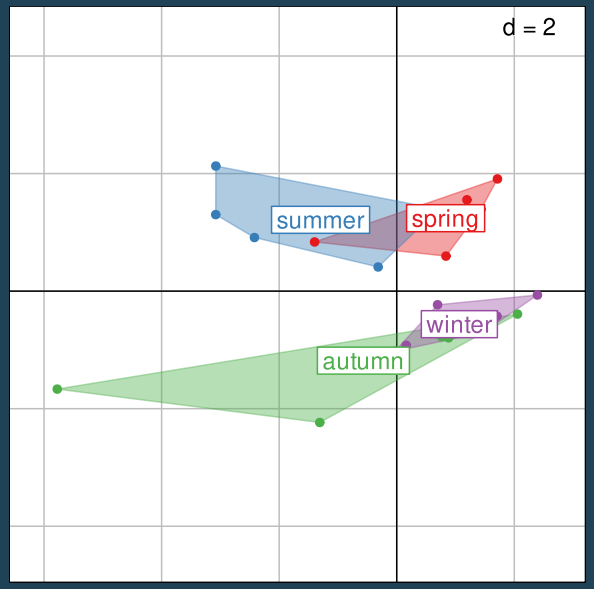
```
g1$row
```



$\mathbf{XQA}$: projections of individuals
(`bca.season$ls`)

Cross-validation

```
xval <- loocv(bca.season)
s.class(xval$XValCoord, meau$design$season, col = TRUE,
    star = 0, ell = 0, chull = 1)
```



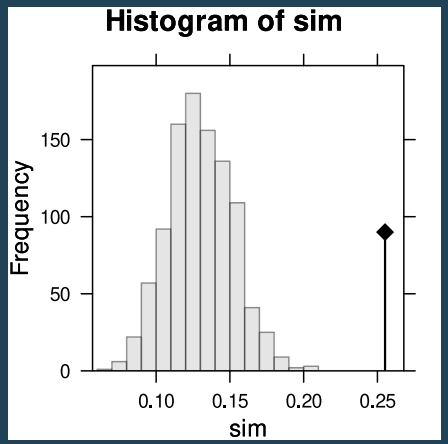
Discriminant Analysis

Perform the analysis

```
dis.season <- discrimin(pca_env, meau$design$season,
    scannf = FALSE)
```

Test the significance of the link

```
rt.dis <- randtest(dis.season)
plot(rt.dis)
```

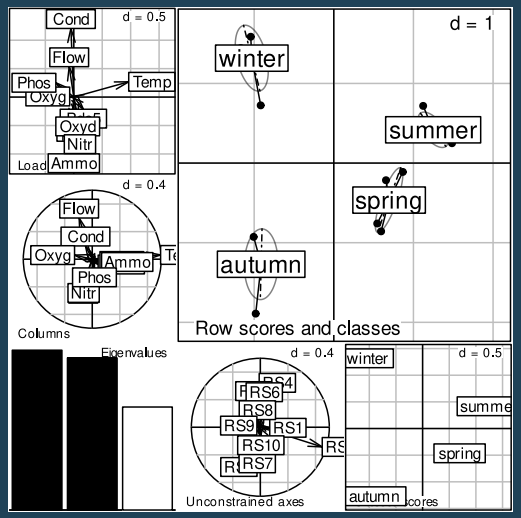



```
rt.dis
```

```
## Monte-Carlo test
## Call: randtest.discrimin(xtest = dis.
##
## Observation: 0.2551751
##
## Based on 999 replicates
## Simulated p-value: 0.001
## Alternative hypothesis: greater
##
##      Std.Obs  Expectation     Varianc
## 5.6807216790 0.1298225308 0.000486921
```

Plot the results

```
g1 <- plot(dis.season)
```

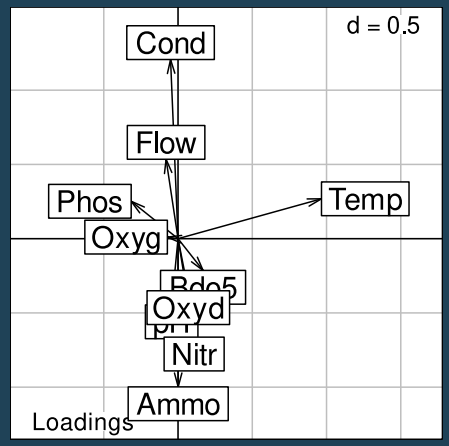



```
names(g1)
```

```
## [1] "loadings" "col"      "eig"      "row"      "Xax"      "class"
```

Loadings for variables

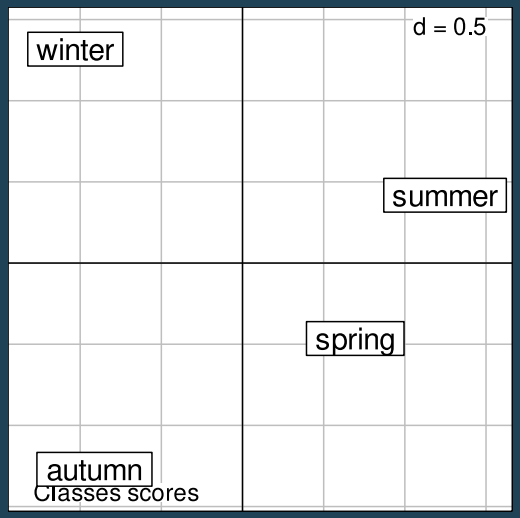
```
g1$loadings
```



$\mathbf{A}^*$: coefficients (loadings) for the variables of $\mathbf{X}_B$ (`dis.season$fa`) with the constraint that $\|\mathbf{X}\mathbf{A}^*\|^2_{\mathbf{D}} = 1$

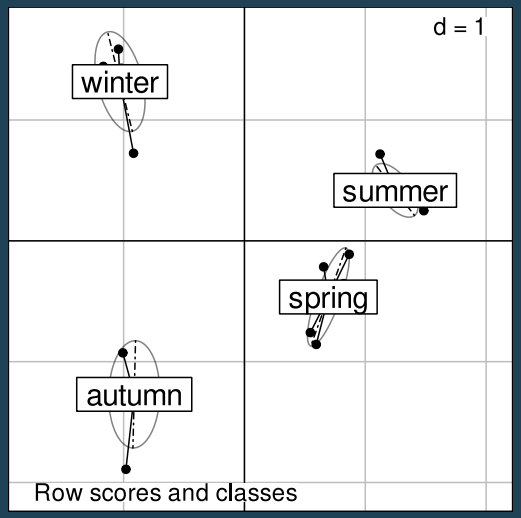
Scores for individuals and classes

```
g1$class
```



$\mathbf{X}_B\mathbf{A}^*$: scores of classes
(`dis.season$gc`)

```
g1$row
```



$\mathbf{X}\mathbf{A}^*$: scores of individuals
(`dis.season$li`)

Your turn

1. Create a Rmd or a R file
2. Create two tables or random numbers (`rnorm`) with 50 individuals and either 10 or 200 variables
3. Create a factor separating the individuals in 5 groups of 10 individuals (`gl`)
4. Perform the between-class analyses of the two tables
5. Look at the outputs and compare the results
6. Display cross-validated maps
7. Interpret