

Fiche de Biostatistique

Tableaux homogènes

D. Chessel, A.B. Dufour & J. Thioulouse

Résumé

La fiche passe en revue quelques usages de l'analyse en composantes principales sur les tableaux homogènes. On rencontre le non centrage, le décentrage, le double centrage autour des tableaux de pourcentages, de notes, de rangs ou de notes d'abondance.

Plan

1	INTRODUCTION.....	2
2	POURCENTAGES.....	2
3	NOTES.....	8
4	RANGS.....	10
5	UNE SEULE UNITE DE MESURE.....	11
6	NOTES D'ABONDANCE.....	17

1 Introduction

Nous avons défini la famille des analyses en composantes principales. Dans cette famille, le cas le plus utilisé est celui de l'ACP normée ou ACP sur matrice de corrélation. Cette pratique est incontournable quand le tableau contient des variables de nature diverses, pour être simple quand le tableau réunit des kg (ou g ou t), des m (ou mm ou années-lumières), des vitesses, des pommes et des scoubidous. La variance dépendant des unités, elle n'a pratiquement que la fonction de permettre la normalisation, c'est-à-dire sa propre disparition.

Les tableaux homogènes, au contraire comporte dans chaque cellule un nombre comparable au contenu des autres cellules, qu'il s'agisse d'une notation unique d'abondance, une présence-absence, un rang, un pourcentage, etc... L'usage de l'ACP normée peut alors être sans inconvénient ou au contraire obscurcir définitivement l'information. A l'aide d'exemples, la fiche regroupe des cas typiques qui permettra de faire des choix pertinents.

2 Pourcentages

Un tableau X est un tableau de fréquences si la somme des valeurs par lignes ou (exclusif) la somme des valeurs par colonnes vaut l'unité. On notera alors $x_{ij} = f_{j/i}$ dans le premier cas et $x_{ij} = f_{i/j}$ dans le second. Si on hésite, c'est qu'on est sur un problème d'analyse des correspondances (AFC).

```
> data(granulo)

> granulo
$tab
  V1 V2 V3 V4 V5 V6 V7 V8 V9
1  0  2  1  15 260 810 1815 1990  0
2  0  0 12 340 1220 2095 2950 2050 1160
...
11 0 39 130 325 680 460 485 1235 530
...
32  2 13  93 244 299 263 1029 2920 293
33 0 0  0  16 182 1288 1364 3685 382
34 0 0  0  0 433 1209 3295 2464  0
...
48 0 0  0  7 695 1700 4440 2730  0
49 0 0  0 22 950 3005 5365 2230  0

$born
 [1] 0.3 0.5 1.0 2.0 4.0 8.0 16.0 32.0 64.0 128.0
```

Ce tableau comporte 49 lignes - échantillons et 9 colonnes - classe de diamètres (Gaschignard-Fossati 1986). L'échantillon 32 a fourni 2 grammes de grains ayant un diamètre compris entre 0.3 et 0.5 mm (sable fin), ..., 293 grammes de grains compris entre 67 et 128 mm (gros galets). Le poids total de sédiments récolté par la drague n'est que le résultat du hasard et permet de calculer un profil par lignes :

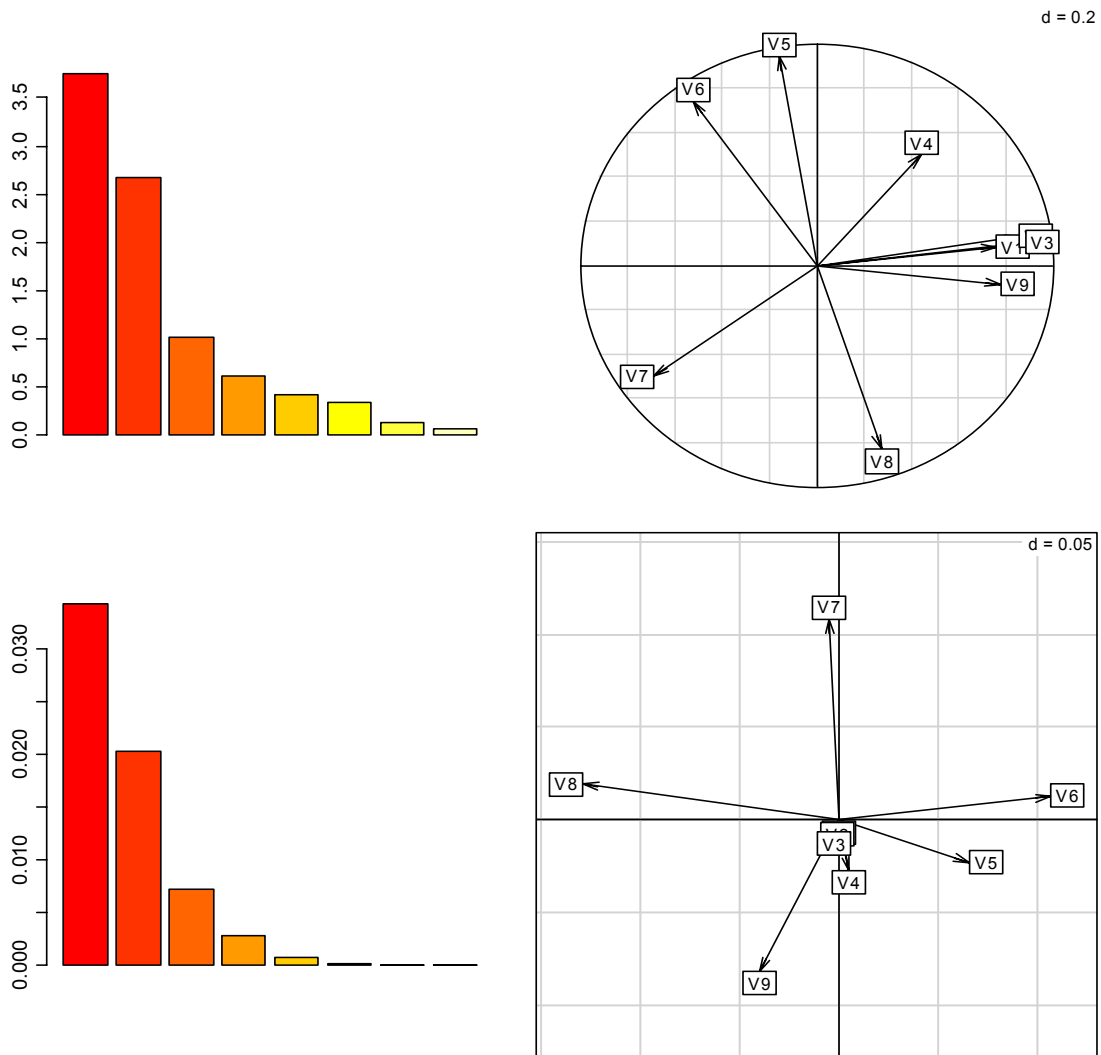
```
> grapc <- t(apply(granulo$tab,1,function(x) x/sum(x)))
```

```
> round(grapc,dig=2)[1:3,]
  V1 V2 V3  V4  V5  V6  V7  V8  V9
1  0  0  0 0.00 0.05 0.17 0.37 0.41 0.00
2  0  0  0 0.03 0.12 0.21 0.30 0.21 0.12
3  0  0  0 0.00 0.16 0.21 0.47 0.13 0.03

> grapc <- data.frame(grapc)
```

Chaque cellule contient une fréquence et le tableau est homogène.

```
> s.corcircle(dudi.pca(grapc,scan=F)$co)
> par(mfrow=c(2,2))
> s.corcircle(dudi.pca(grapc)$co)
Select the number of axes: 2
> s.arrow(dudi.pca(grapc,scal=F)$co)
Select the number of axes: 2
```



Les variances par colonnes sont très différentes :

```
> apply(grapc,2,var)
      V1      V2      V3      V4      V5      V6      V7      V8
5.047e-07 6.827e-06 9.625e-05 2.187e-03 6.208e-03 1.333e-02 1.382e-02 1.953e-02
      V9
1.144e-02
```

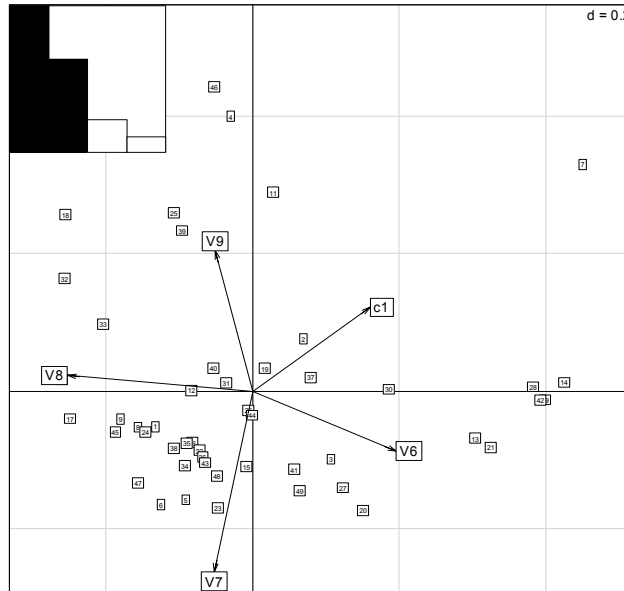
mais ce n'est pas une raison pour s'en débarrasser. La première classe joue dans l'ACP normée un rôle disproportionné à son importance expérimentale. On peut regrouper les cinq premières classes.

```
> granou <- cbind.data.frame(c1,grapc[,6:9])
```

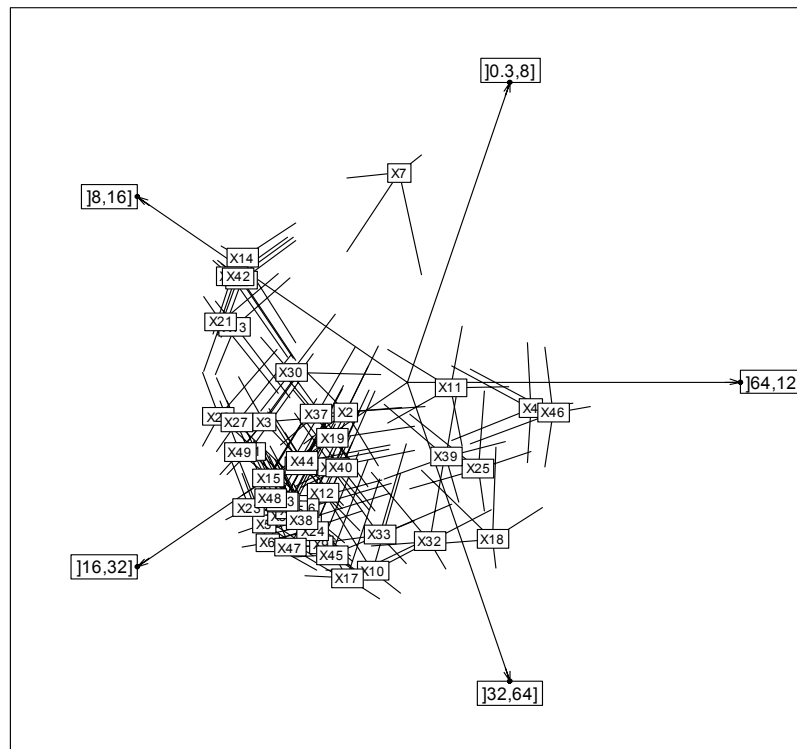
```
> round(granou, dig=2)
  c1  V6  V7  V8  V9
1  0.06 0.17 0.37 0.41 0.00
2  0.16 0.21 0.30 0.21 0.12
3  0.16 0.21 0.47 0.13 0.03
4  0.18 0.06 0.13 0.22 0.41
...
```

Pour une première approche, rapide et approximative :

```
> s(dudi.pca(granou, scal=F, scan=F))
```



On a perdu une valeur propre (démontrer que la dernière est toujours nulle). L'analyse indique que la position des variables est celle de la représentation triangulaire étendue à 5 dimensions.



ci-dessus :

```

> x <- cos((1:5)*2*pi/5)
> y <- sin((1:5)*2*pi/5)
> xy <- cbind.data.frame(x,y)
> row.names(xy) <- c("[0.3,8]", "[8,16]", "[16,32]", "[32,64]", "[64,128]")
> xy
      x          y
[0.3,8]  0.309  9.511e-01
[8,16]   -0.809  5.878e-01
[16,32]  -0.809 -5.878e-01
[32,64]   0.309 -9.511e-01
[64,128]  1.000 -2.449e-16
> s.corcircle(xy,grid=F)
> s.distri(xy,data.frame(t(granou)),add.p=T,axesell=T,
  csta=0.2,clab=0.75,cell=0)

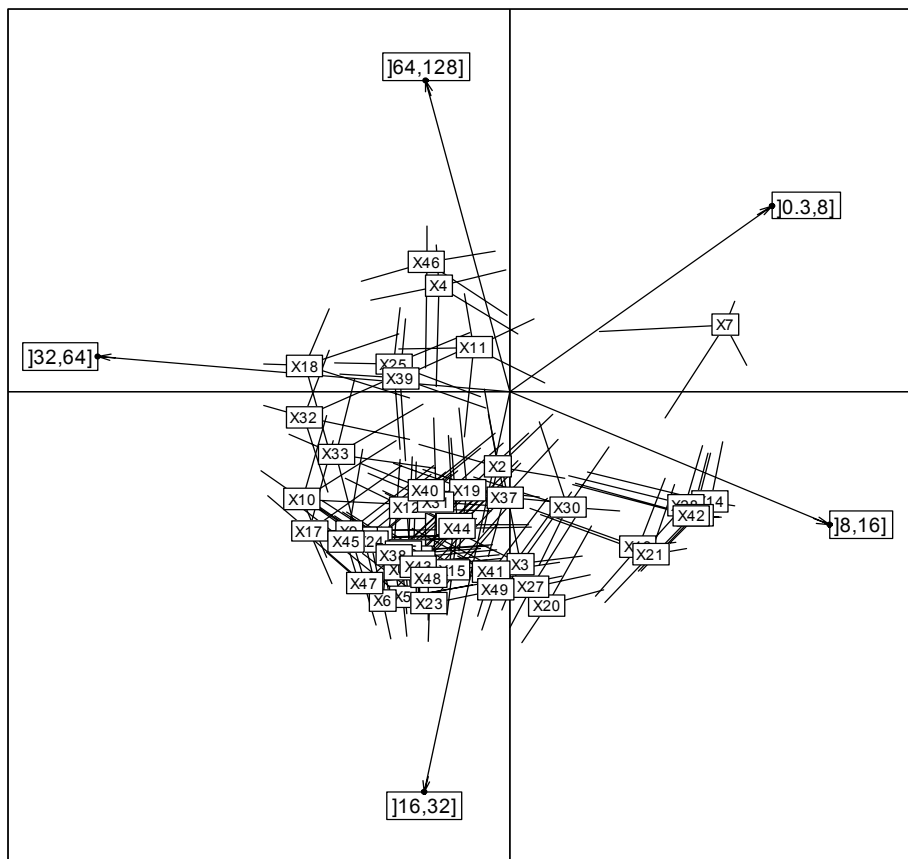
```

ci-dessous :

```

> names(granou) <- row.names(xy)
> xy.pca <- dudi.pca(granou,scal=F,scan=F)$c1
> s.arrow(xy.pca,grid=F)
> s.arrow(xy.pca,grid=F,lab=names(granou))
> s.distri(xy.pca,data.frame(t(granou)),add.p=T,axesell=T,
  csta=0.2,clab=0.75,cell=0)

```



L'image *naïve* (pas tant que ça ?) contient presque la même information que l'image optimale. Les structures des tableaux de pourcentages sont très particulières et supporte mal une normalisation indésirable. Il est logique de privilégier l'expression d'un point au centre de gravité de sa distribution plutôt que de laisser faire le centrage. Mais ce centrage est nécessaire pour éviter l'expression absurde de l'évidence "les données sont positives". La représentation triangulaire s'impose pour trois variables.

```

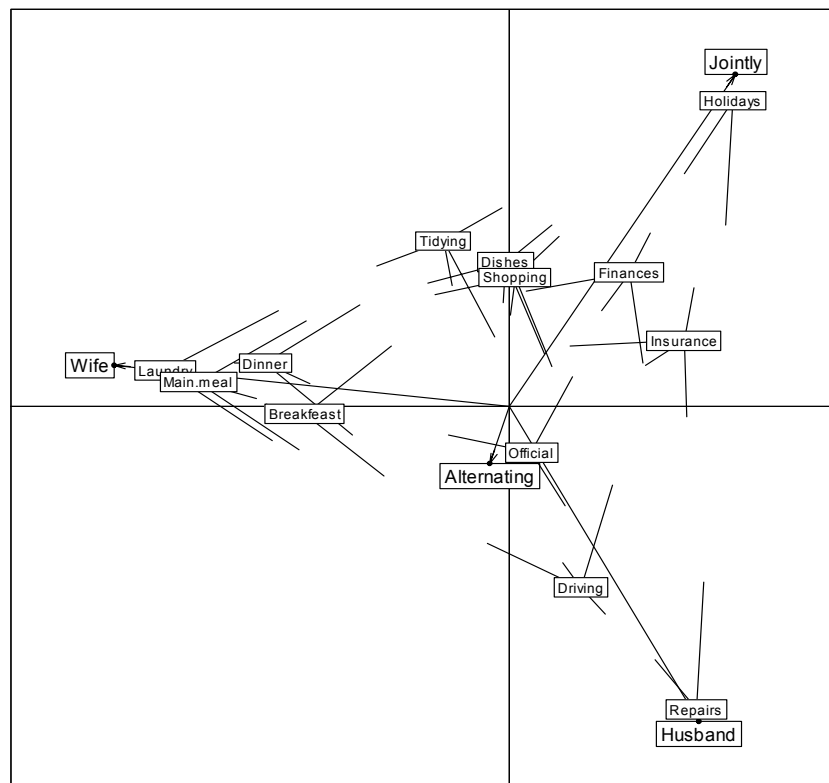
> data(housetasks)
> hc <- t(apply(housetasks,1,function(x) x/sum(x)))
> round(hc,dig=2)

```

Wife Alternating Husband Jointly

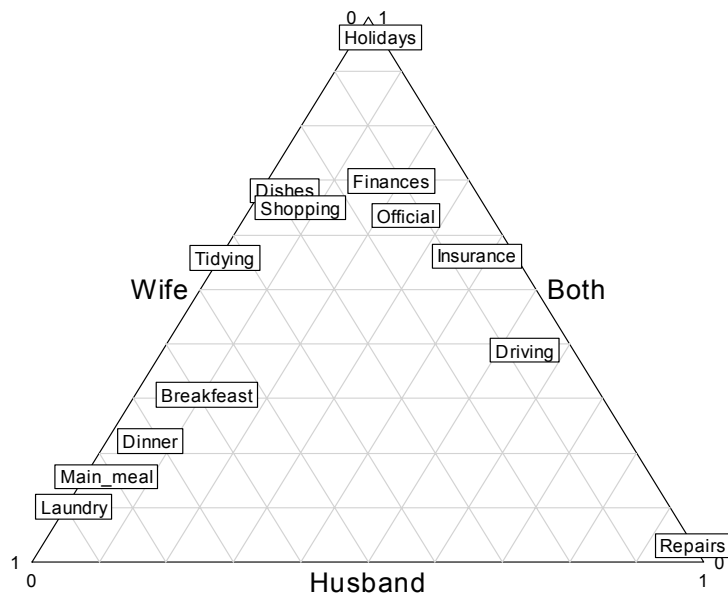
Laundry	0.89	0.08	0.01	0.02
Main_meal	0.81	0.13	0.03	0.03
Dinner	0.71	0.10	0.06	0.12
Breakfast	0.59	0.26	0.11	0.05
Tidying	0.43	0.09	0.01	0.47
Dishes	0.28	0.21	0.04	0.47
Shopping	0.28	0.19	0.08	0.46
Official	0.12	0.48	0.24	0.16
Driving	0.07	0.37	0.54	0.02
Finances	0.12	0.12	0.19	0.58
Insurance	0.06	0.01	0.38	0.55
Repairs	0.00	0.02	0.97	0.01
Holidays	0.00	0.01	0.04	0.96

```
> hc <- data.frame(hc)
> hc.pca <- dudi.pca(hc, scal=F, scan=F)$c1
> s.arrow(hc.pca, grid=F, lab=names(hc))
> s.distri(hc.pca, data.frame(t(hc)),
  add.p=T, axesell=T, csta=0.2, clab=0.75, cell=0)
```



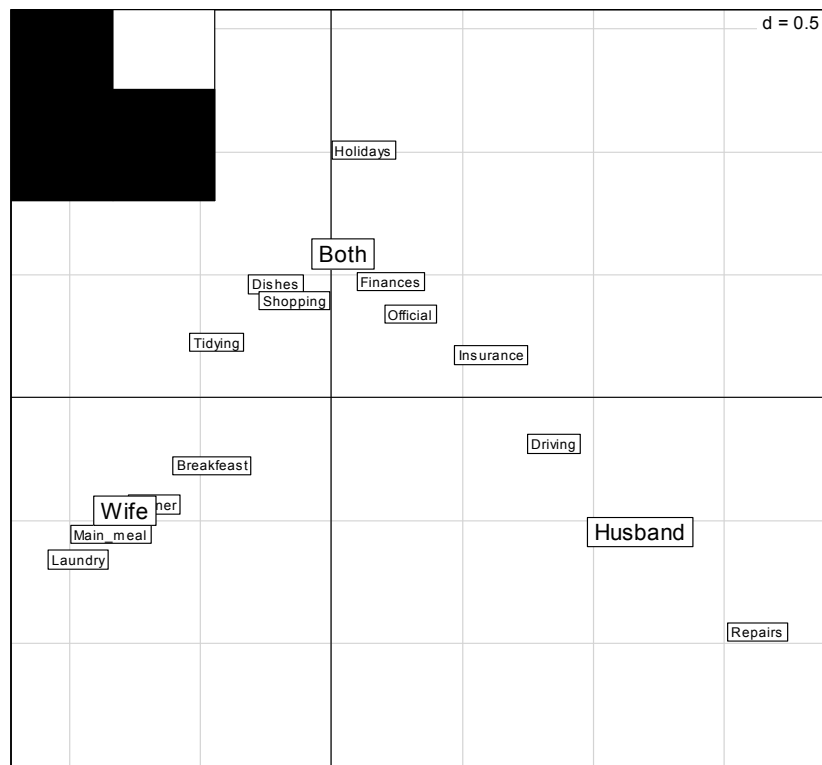
A Madame, la cuisine, à Monsieur la voiture, au couple le reste. Groupons "ensemble" et "chacun à son tour", la représentation triangulaire suffira :

```
> tri <- cbind.data.frame(hc$Wife, hc$Husband, hc$Jointly+hc$Alternating)
> names(tri) <- c("Wife", "Husband", "Both")
> triangle.plot(tri, show=F, clab=1, label=row.names(hc))
```



Une dissymétrie certaine. Ci-dessous, l'analyse des correspondances.

```
> row.names(tri) <- row.names(hc)
> scatter(dudi.coa(tri))
```



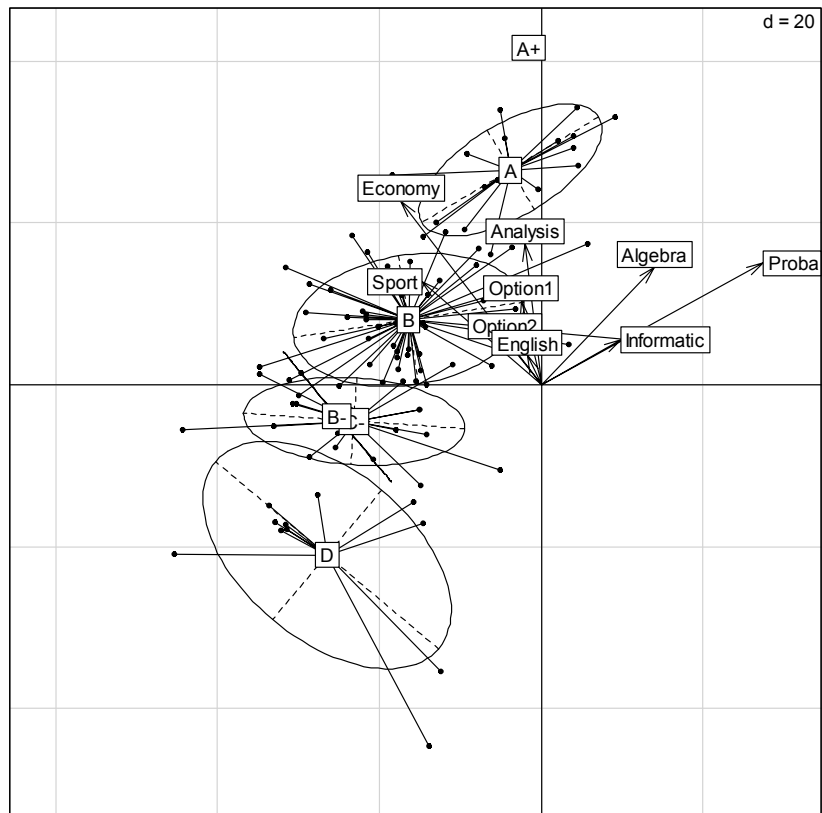
En indiquant comment, pour une tâche donnée, les couples se répartissent, on indique clairement une intention.

3 Notes

deug est une liste à trois composantes :

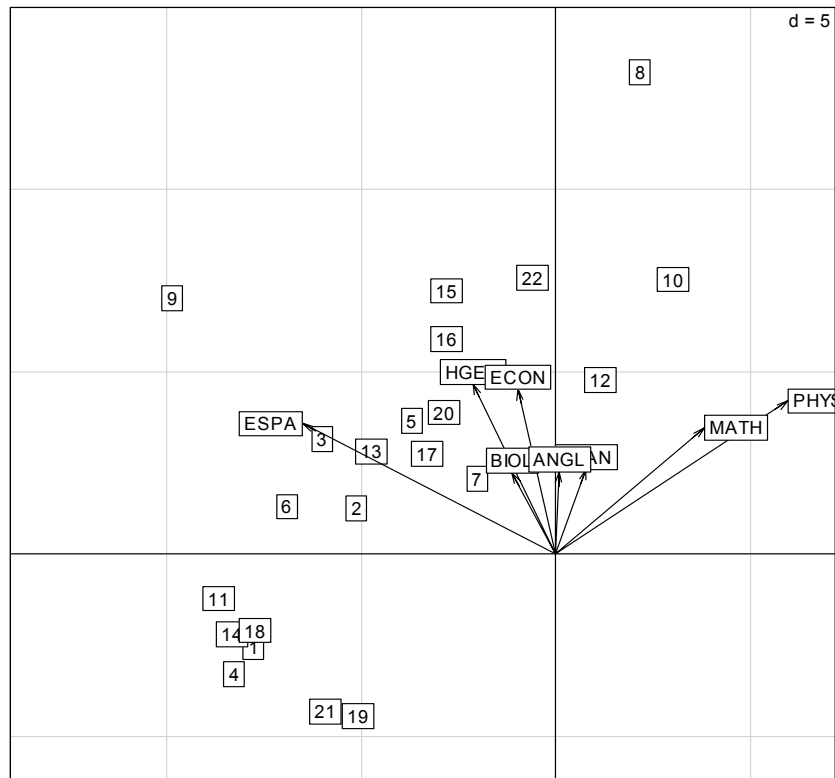
```
$tab data.frame 118 étudiants et 8 matières
$result facteur à 118 composantes donnant le résultat à l'examen (A+, A, B, B-, C-, D)
$cent vecteur nombre de points par matière nécessaire pour avoir exactement 10/20 avec le coefficient de la matière
```

```
> data(deug)
> pca1 <- dudi.pca(deug$tab,scal=FALSE, center=deug$cent, scan=FALSE)
> s.class(pca1$li,deug$result)
> s.arrow(40*pca1$c1,add.plot=T)
```



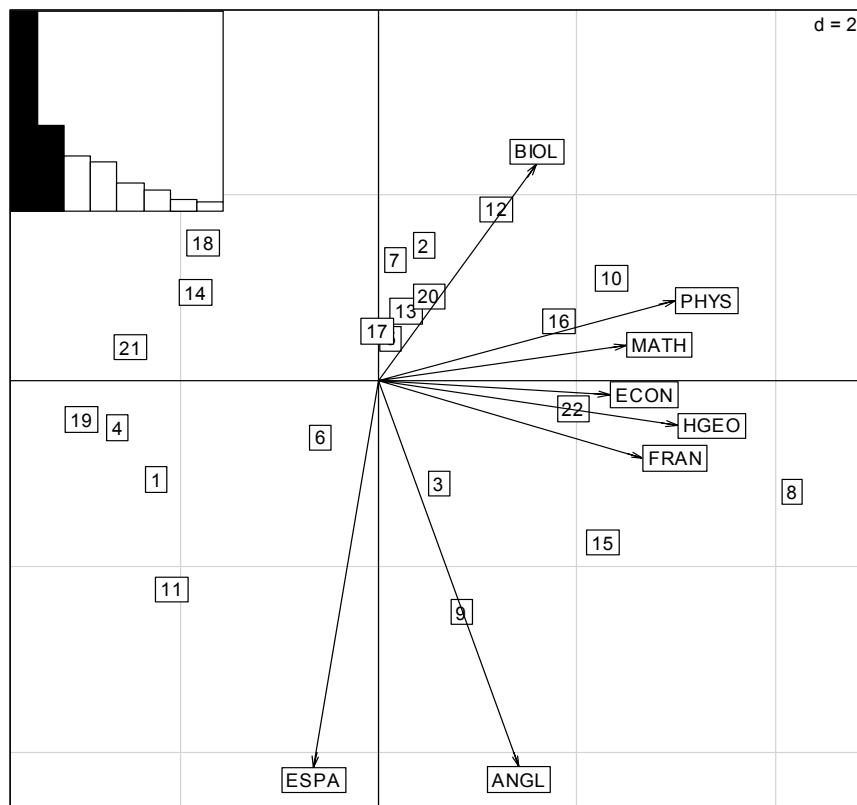
L'origine n'est plus le centre de gravité du nuage. On y perdrait la distinction entre les matières qui font la forme du nuage et celles qui donnent sa position. On retrouve cet aspect dans les données 'seconde'.

```
> data(seconde)
> pca2 <- dudi.pca(seconde, center=rep(10,8), scale=F, scan=FALSE)
> s.label(pca2$li)
> s.arrow(10*pca2$c1,add.plot=T)
```



Cette figure ne modifie pas la notion de compromis (toutes les matières valorisent peu ou prou les bons élèves) mais souligne que Maths et Physique excluent d'abord (les élèves en haut à droite) et que le professeur d'espagnol est d'une bienveillance évidente. Ce qui diffère sensiblement de l'ACP normée qui mettait en évidence l'originalité et la solidarité des professeurs de langue :

```
> scatter(dudi.pca(seconde, scan=F), clab.r=1)
```



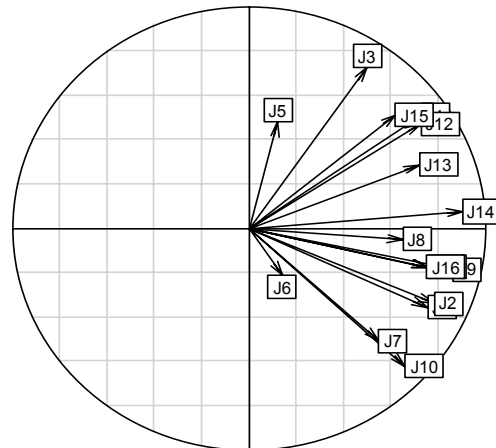
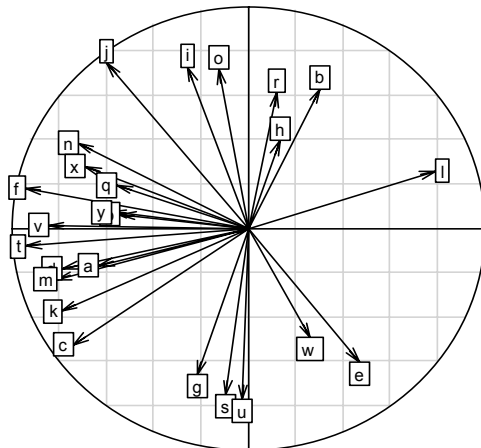
Quand un vecteur de valeurs définit un point signifiant de l'espace devant servir de référence, il convient d'en faire l'origine.

4 Rangs

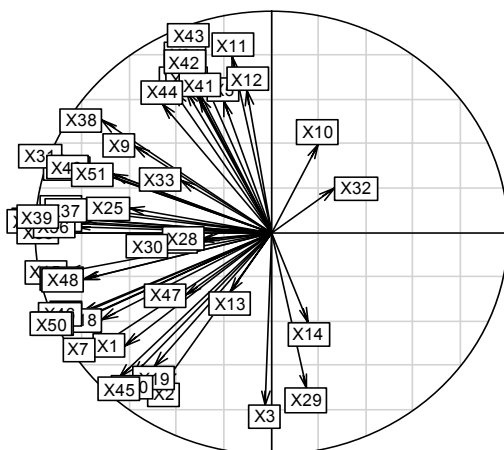
```
> par(mfrow=c(2,2))
> data(macon) # 25 juges classent 8 vins
> s.corcircle(dudi.pca(macon,scan=F)$co)
> data(fruits) # 16 juges classent 28 lots de fruits
> s.corcircle(dudi.pca(fruits$jug,scan=F)$co)
> data(rankrock) # 51 étudiants classent 10 groupes de musique
> s.corcircle(dudi.pca(rankrock,scan=F)$co)
```

d = 0.2

d = 0.2



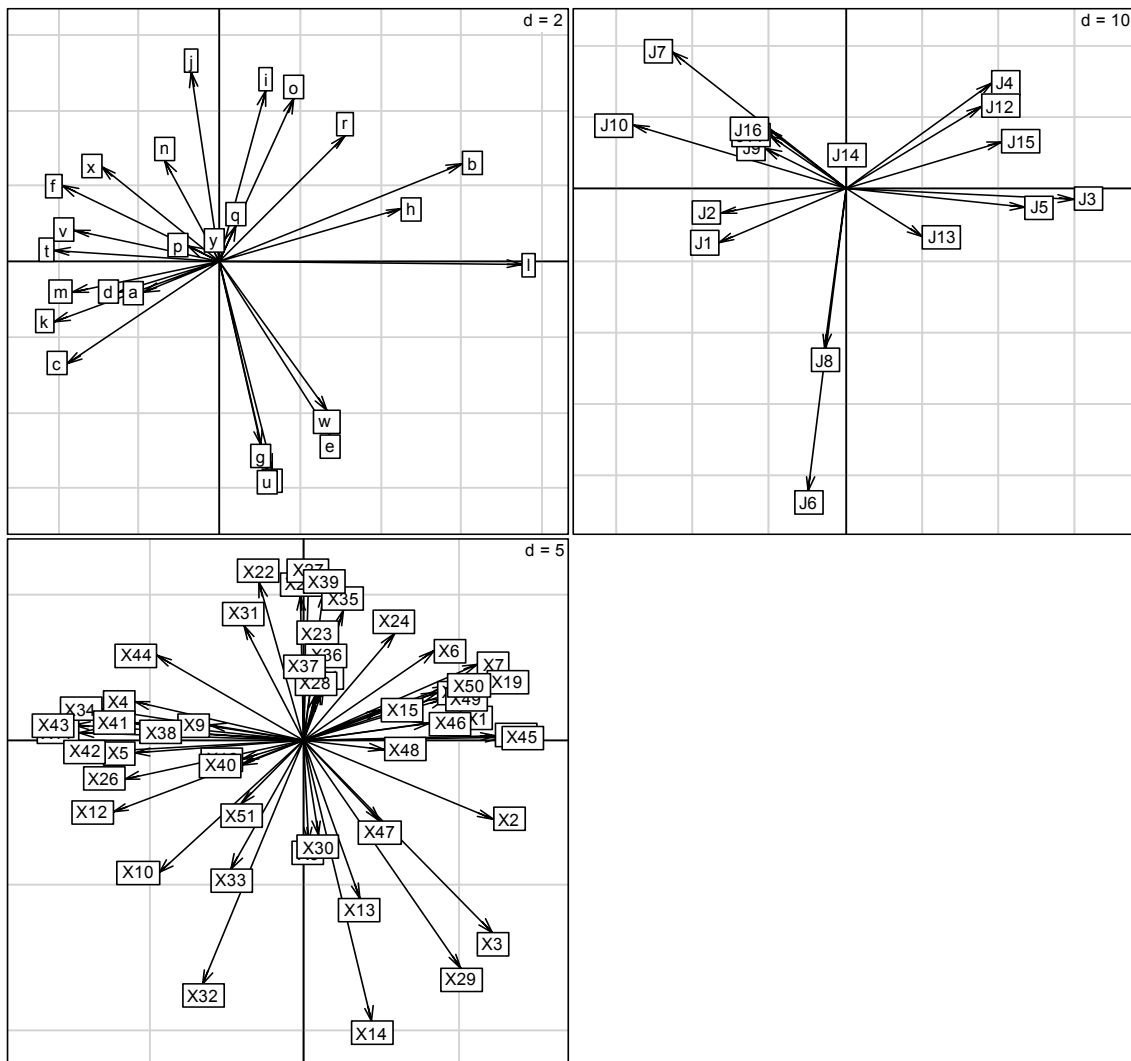
d = 0.2



S'il s'agit de mesurer la cohérence du jury, d'exprimer un compromis entre jugements, un choix collectif (ou plusieurs tendances regroupant des parties du jury), de mettre en évidence la ressemblance entre juges, *les juges sont en colonnes dans une ACP normée*. Les moyennes sont toutes égales, les variances aussi, la normalisation n'est ni nécessaire, ni nuisible. On peut l'utiliser pour tracer les cercles de corrélation.

Si il s'agit de faire une typologie des juges, de mettre en évidence ce qui les oppose, de montrer qu'il existe plusieurs types de jugements, *les juges sont en lignes dans une ACP centrée*. On laissera ainsi dominer dans l'analyse les produits qui ont reçu les appréciations les plus variables. Les deux approches sont antinomiques.

```
> s.arrow(dudi.pca(t(macon),scan=F)$li)
> s.arrow(dudi.pca(t(fruits$jug),scan=F)$li)
> s.arrow(dudi.pca(t(rankrock),scan=F)$li)
```



On remarquera, que derrière le compromis, si il est suffisant pour définir le premier axe, les premières analyses définissent aussi les contradictions entre jugements, mais les secondes sont plus claires.

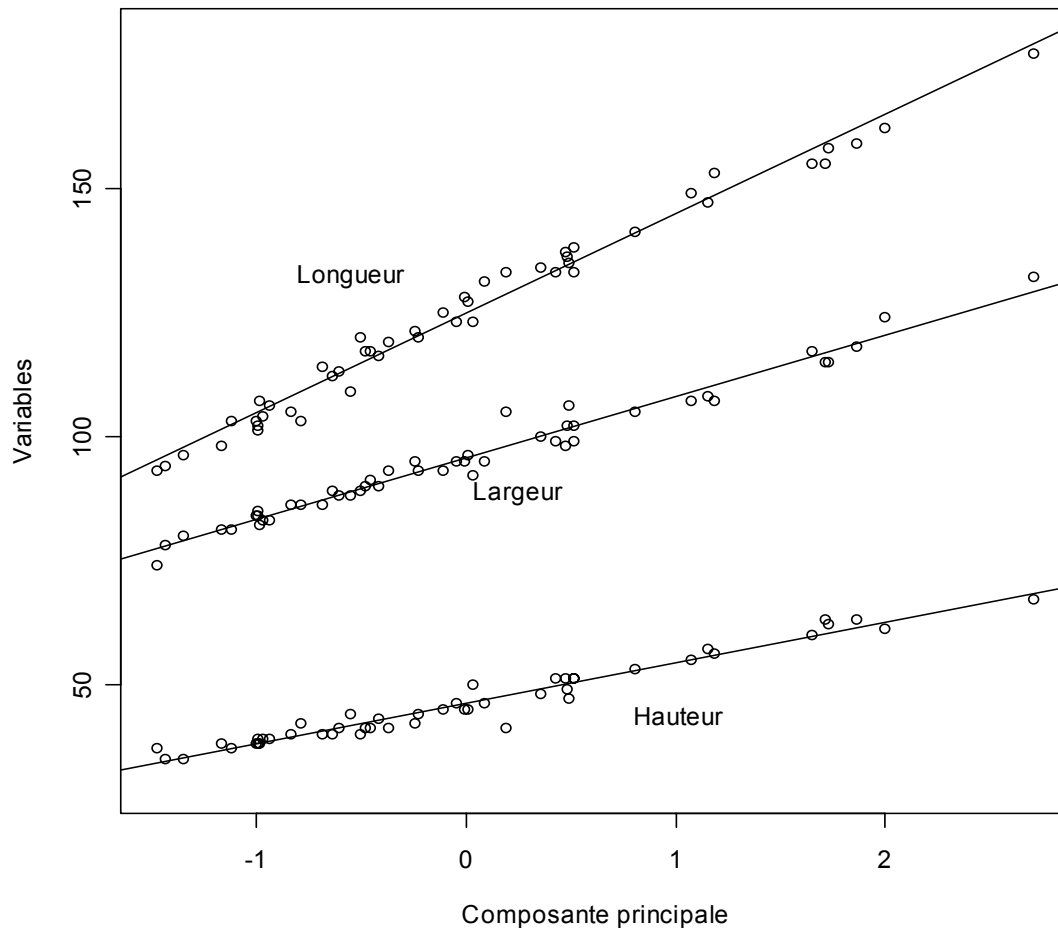
5 Une seule unité de mesure

C'est souvent un problème de morphométrie. Dans 'tortues' les variables sont les trois dimensions de carapaces de tortues mesurées en mm (Jolicœur and Mosimann 1960).

```
> data(tortues)
> ttaille <- tortues[,1:3]
> tsexe <- tortues[,4]
> names(ttaille)
[1] "long" "larg" "haut"

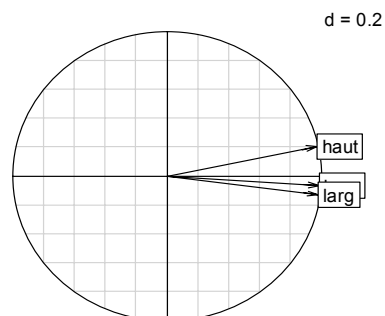
> dudil <- dudi.pca(ttaille,scan=F)
> plot(dudil$li[,1],tortues$long,ylim=c(30,180),
       xlab="Composante principale",ylab="Variables")
> text(locator(1),"Longueur")
> points(dudil$li[,1],tortues$haut)
> text(locator(1),"Hauteur")
```

```
> points(dudi1$l1[,1], tortues$larg)
> text(locator(1), "Largeur")
> abline(lm(tortues$larg~dudi1$l1[,1]))
> abline(lm(tortues$long~dudi1$l1[,1]))
> abline(lm(tortues$haut~dudi1$l1[,1]))
```



L'ACP joue ici son rôle de recherche de *variable latente*. La composante principale prédit les trois variables. C'est une *explicative cachée*. Elle représente la taille théorique de la tortue. C'est la variable cachée qui prédit au mieux les autres, c'est aussi la variable cachée qui est le mieux prédite par toutes les autres.

```
> s.corcircle(dudi1$co)
```

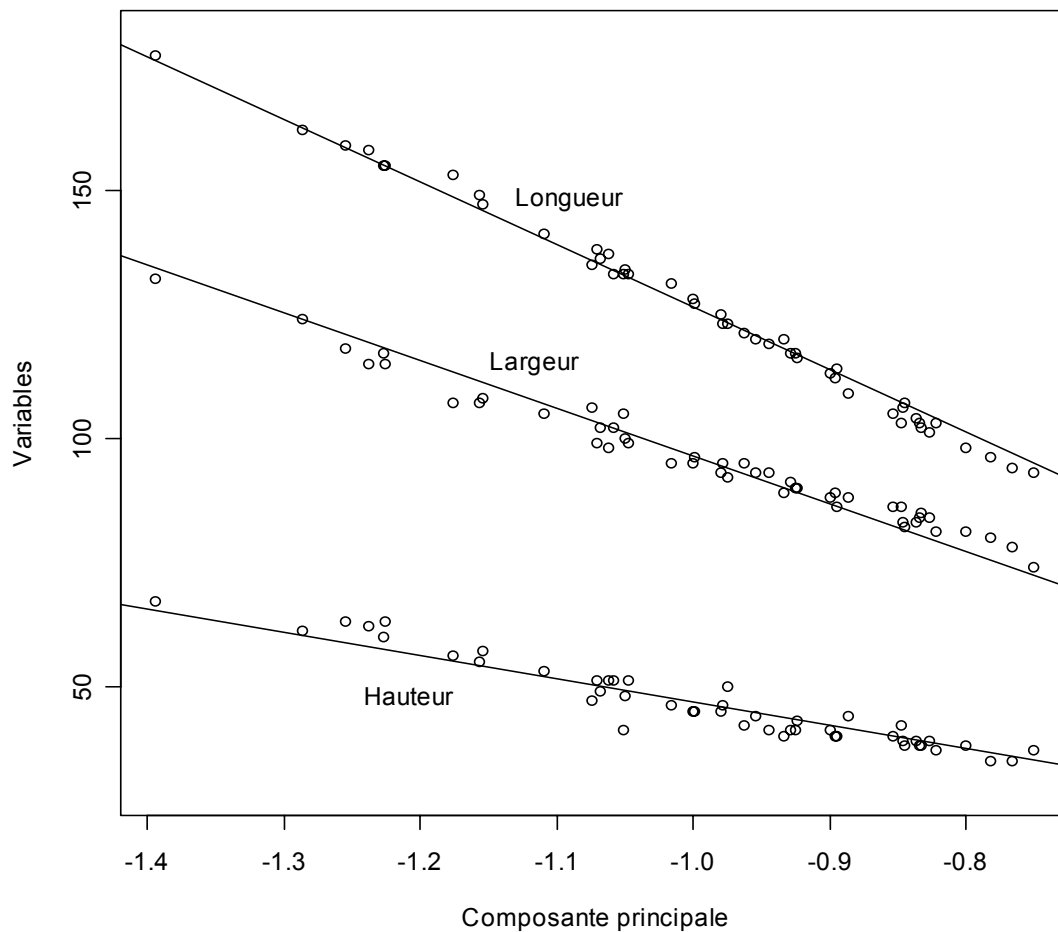


On dit que c'est un effet "taille".

```
> coefficients(lm(tortues$larg~dudi1$l1[,1]))
(Intercept) dudi1$l1[, 1]
  95.50      12.42
> coefficients(lm(tortues$long~dudi1$l1[,1]))
(Intercept) dudi1$l1[, 1]
 124.69      20.09
> coefficients(lm(tortues$haut~dudi1$l1[,1]))
(Intercept) dudi1$l1[, 1]
  46.15      8.12
> apply(ttaille,2,mean)
      long      larg      haut
124.69  95.50  46.15
```

Si on veut une variable latente qui fera des prédictions nulles pour une valeur nulle (régression par l'origine) on fera une ACP non centrée.

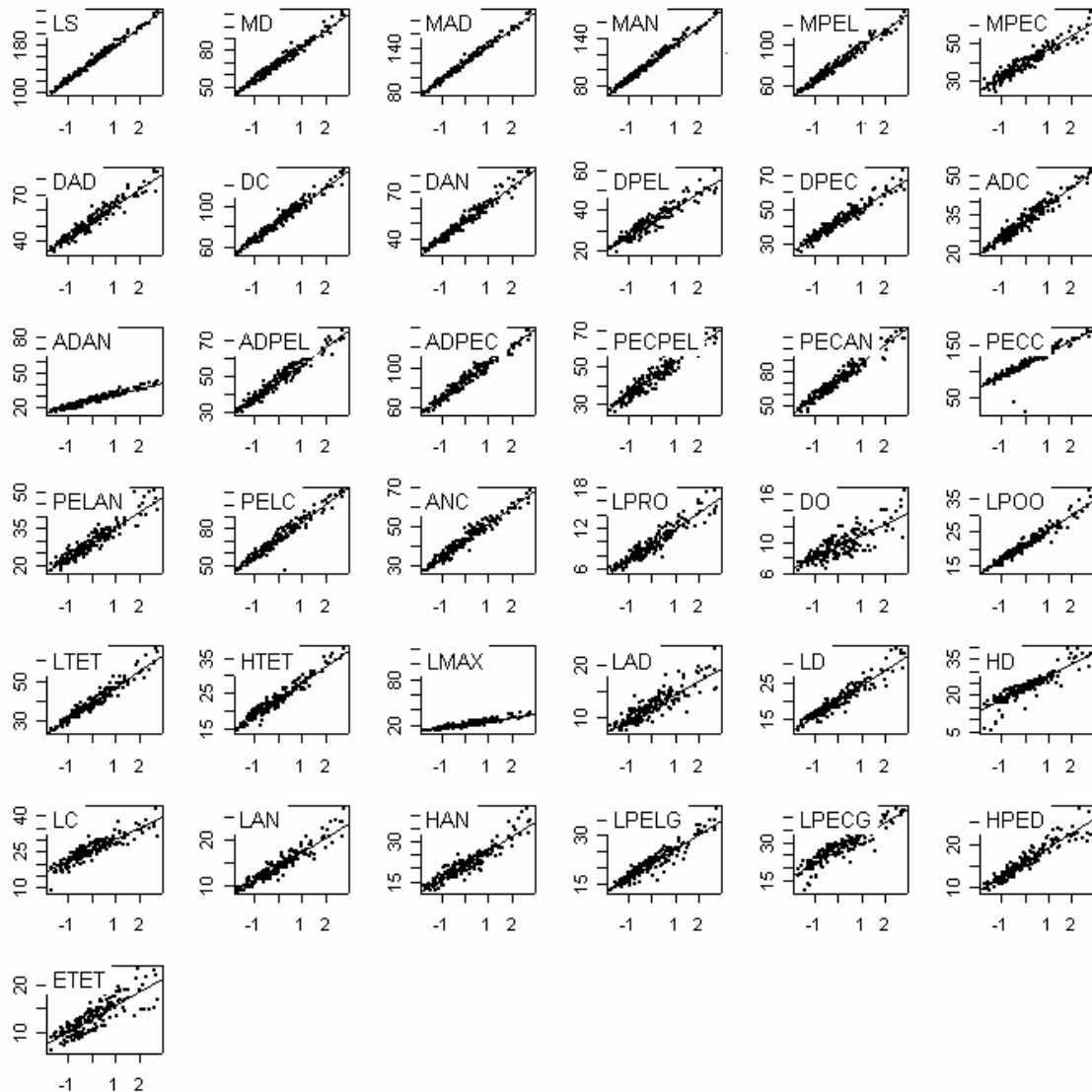
```
> dudi2 <- dudi.pca(ttaille,scan=F,cent=F,scal=F)
> plot(dudi2$l1[,1],tortues$long,ylim=c(30,180),
      xlab="Composante principale",ylab="Variables")
> text(locator(1),"Longueur")
> points(dudi2$l1[,1],tortues$haut)
> text(locator(1),"Hauteur")
> points(dudi2$l1[,1],tortues$larg)
> text(locator(1),"Largeur")
> abline(lm(tortues$larg~-1+dudi2$l1[,1]))
> abline(lm(tortues$long~-1+dudi2$l1[,1]))
> abline(lm(tortues$haut~-1+dudi2$l1[,1]))
```



Cet *auto-modèle* (créé par les données pour modéliser les données) est plus réaliste mais moins bon (les erreurs ont une organisation, liée à la présence des deux sexes).

Dans les données de J.M. Lascaux (1996), l'effet taille est omniprésent :

```
> data(lascaux)
> m <- na.omit(lascaux$morpho)
> m.acp <- dudi.pca(m, scan=F, nf=3)
> score(m.acp)
```



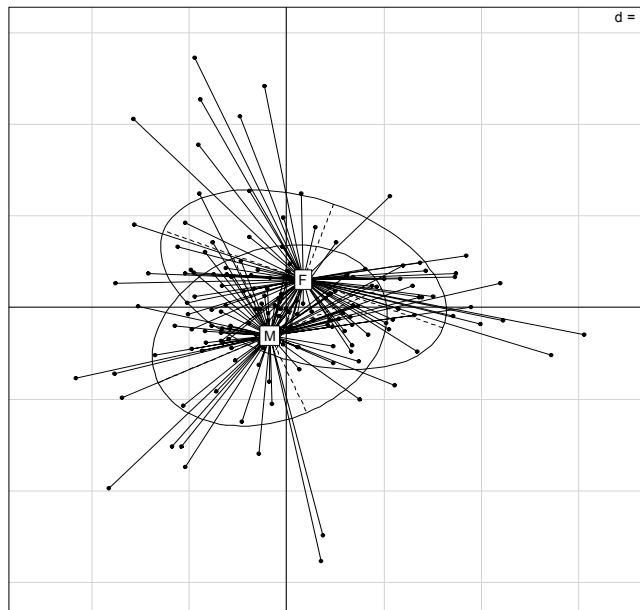
Une truite qui a perdu un morceau de nageoire dans une bataille est facilement repérée ! Ce qui reste quand on enlève l'effet taille concerne, grossièrement parlant, *la forme*. Les variables 13, 18, 20 et 21 présentent des *outliers*. Elles sont exclues :

```
> m <- na.omit(lascaux$morpho)[, -c(13, 18, 20, 27)]
> m.acp <- dudi.pca(m, scan=F, nf=3)
> score(m.acp) # vérifier que le problème est résolu, avec brutalité, certes !

> names(lascaux$sex) <- row.names(lascaux$morpho) # le sexe
> names(lascaux$gen) <- row.names(lascaux$morpho) # le statut génétique
```

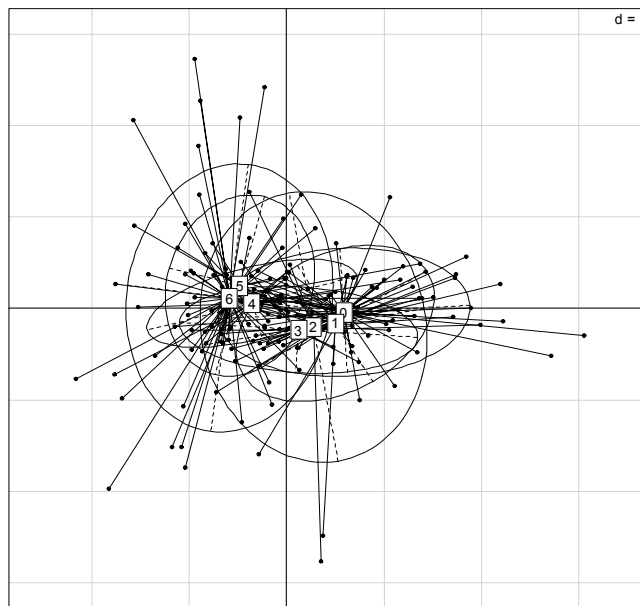
Enlever des données le modèle issu de l'effet taille ou dépouiller l'analyse à partir du facteur 2 donne strictement le même résultat. C'est l'illustration qu'une ACP peut faire deux choses de nature radicalement différente. Le plan 2-3 est directement une analyse de la forme :

```
> s.class(m.acp$li[, 2:3], lascaux$sex[row.names(m.acp$li)])
```



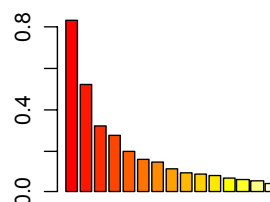
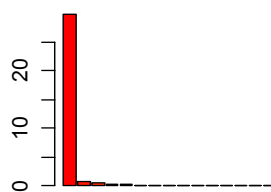
Dans la forme, il y a certainement une composante sexuelle.

```
> s.class(m.acp$li[,2:3], lascaux$gen[row.names(m.acp$li)])
```



Dans la forme, il y a certainement une composante génétique.

```
> par(mfrow=c(1,2)) ; barplot(m.acp$eig[1:15]) ; barplot(m.acp$eig[2:16])
```

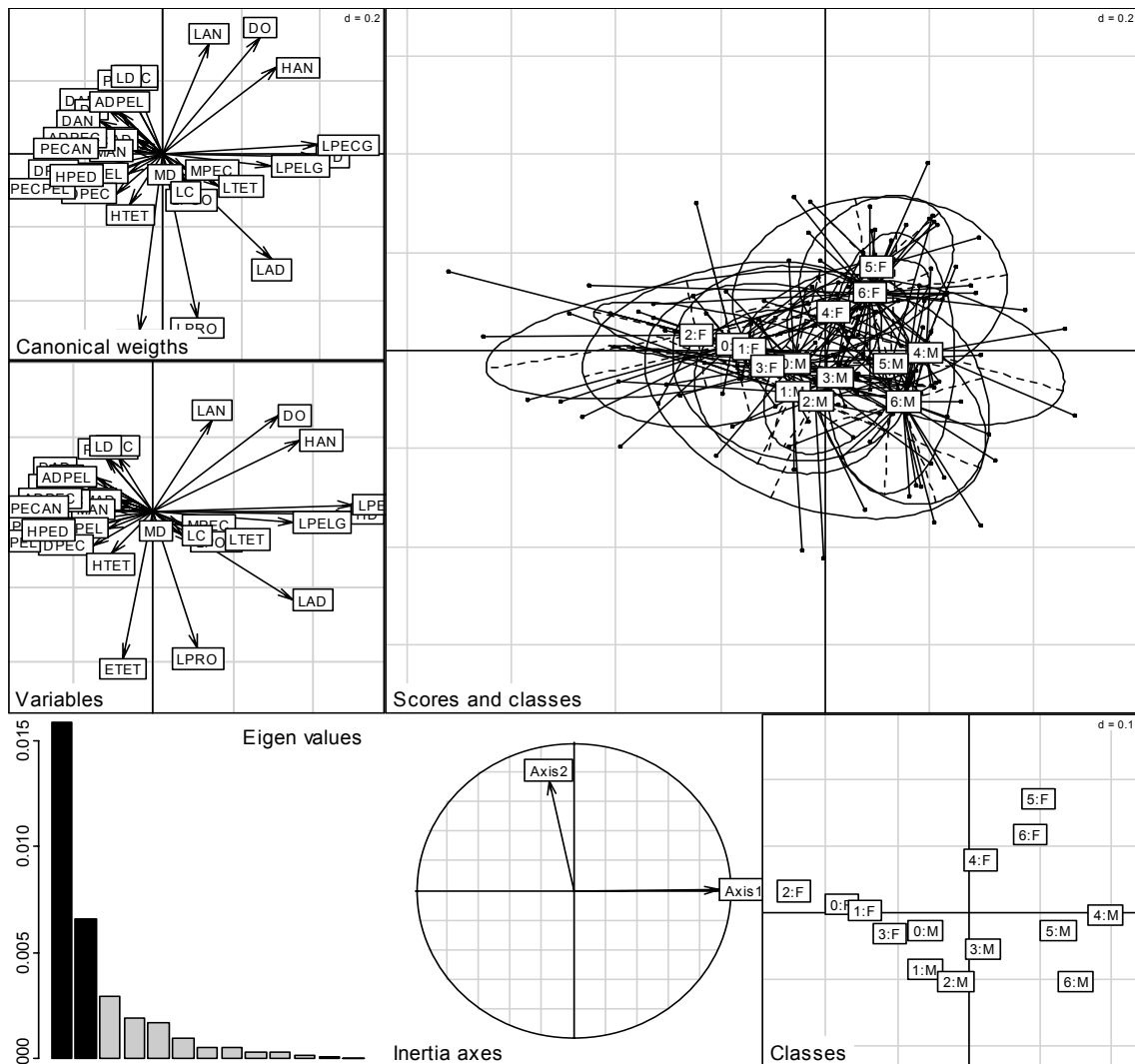


Une ACP peut ainsi en cacher une autre. On peut analyser l'effet taille par une ACP non centrée sur *le tableau des log doublement centré* :

```
> mlog <- log(m)
> mlog0 <- bicenter.wt(mlog)
> mlog0 <- data.frame(mlog0)
> names(mlog0) <- names(m)
> mlog0.acp <- dudi.pca(mlog0, scan=F, scal=F, cent=F)
```

Ceci indique que l'introduction des deux facteurs directement dans l'analyse est pertinente. L'ACP inter-classe est celle des centres de gravité des sous-nuages.

```
> croi <- lascaux$gen[rownames(m.acp$li)]:lascaux$sex[rownames(m.acp$li)]
> plot(between(mlog0.acp, croi, scan=F))
```



Ces truites sont classées en 7 groupes génétiques par le nombre d'allèles méditerranéens. La variable génétique prend les modalités 0 (0 allèle méditerranéen, homozygotes atlantiques, truites dites modernes), 1 à 5 (respectivement 1 à 5 allèles méditerranéens) et 6 (6 allèles méditerranéens, homozygotes méditerranéens, truites dites ancestrales).

Les modernes sont à gauche, les ancestrales à droite, les femelles en haut et les mâles en bas. Le code des variables permettra de poursuivre :

LS Longueur standard
MD Distance bout du museau - insertion de la dorsale

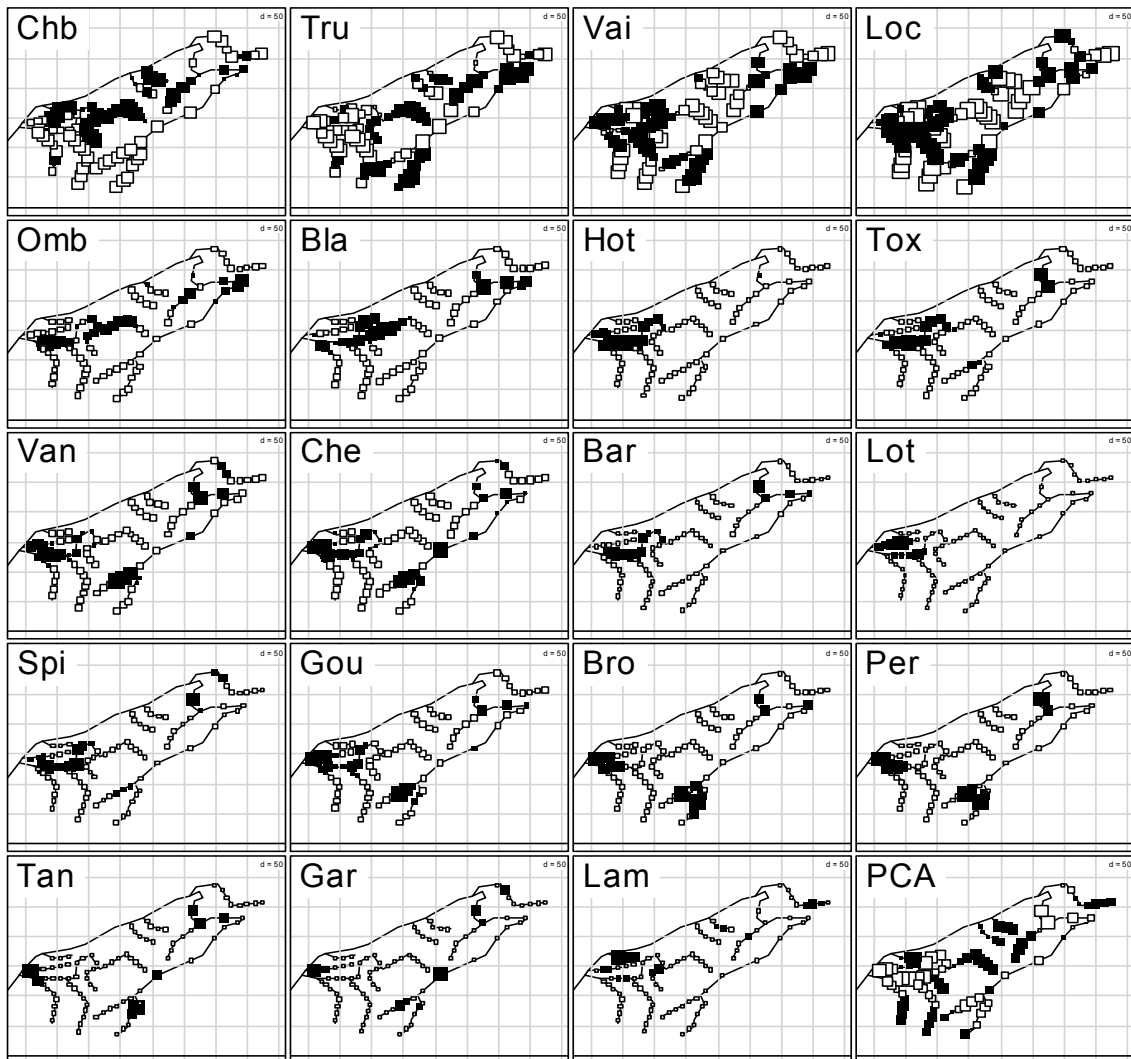
MAD	Distance bout du museau - insertion de l'adipeuse
MAN	Distance bout du museau - insertion de l'anale
MPEL	Distance bout du museau - insertion de la pelvienne
MPEC	Distance bout du museau - insertion de la pectorale
DAD	Distance insertion de la dorsale - insertion de l'adipeuse
DC	Distance insertion de la dorsale - départ de la caudale
DAN	Distance insertion de la dorsale - insertion de l'anale
DPEL	Distance insertion de la dorsale - insertion de la pelvienne
DPEC	Distance insertion de la dorsale - insertion de la pectorale
ADC	Distance insertion de l'adipeuse - départ de la caudale
ADAN	Distance insertion de l'adipeuse - insertion de l'anale
ADPEL	Distance insertion de l'adipeuse - insertion de la pelvienne
ADPEC	Distance insertion de l'adipeuse - insertion de la pectorale
PECPEL	Distance insertion de la pectorale - insertion de la pelvienne
PECAN	Distance insertion de la pectorale - insertion de l'anale
PECC	Distance insertion de la pectorale - départ de la caudale
PELAN	Distance insertion de la pelvienne - insertion de l'anale
PELC	Distance insertion de la pelvienne - départ de la caudale
ANC	Distance insertion de l'anale - départ de la caudale
LPRO	Longueur préorbitale
DO	Diamètre de l'orbite
LPOO	Longueur postorbitale
LTET	Longueur de la tête
HTET	Hauteur de la tête (en passant par au milieu de l'orbite)
LMAX	Longueur de la mâchoire supérieure
LAD	Longueur de l'adipeuse
LD	Longueur de la dorsale
HD	Hauteur de la dorsale
LC	Longueur de la caudale
LAN	Longueur de l'anale
HAN	Hauteur de l'anale
LPELG	Longueur de la pelvienne gauche
LPECG	Longueur de la pectorale gauche
HPED	Hauteur du corps au niveau du pédoncule caudal
ETET	Largeur de la tête (au niveau des orbites)

6 Notes d'abondance

Quand on pense l'ACP comme approche d'un échantillon d'une loi normale multivariée, son usage sur un tableau de notes binaires (0-1) est sans objet. On dit à l'utilisateur "*qu'il n'a pas le droit*". Quand on pense l'ACP comme une opération géométrique ou la maximisation d'une forme quadratique, on dit à l'utilisateur "*qu'il a le droit*". Quand un lecteur formé à une école juge le manuscrit d'un auteur formé à l'autre école, le dialogue est sommaire et tous les coups sont permis. Le même problème se pose pour les notes d'abondance (souvent de 0 à 7 en phytosociologie). L'ACP d'un tableau florofaunistique est valide comme première approche d'objectifs précis. Elle a marquée les pionniers (Goodall 1954) par la facilité avec laquelle elle permet des cartes de synthèse.

```
data(jv73)
par(mfrow=c(5,4))
for(m in 1:19) {
  s.value(jv73$xy, jv73$poi[,m]-mean(jv73$poi[,m]),
    incl=F, contour=jv73$contour, sub=names(jv73$poi)[m],
    possub="topleft", csub=3, cleg=0)
}

s.value(jv73$xy, dudi.pca(jv73$poi, scal=F, scan=F)$li[,1],
  incl=F, contour=jv73$contour, sub="PCA",
  possub="topleft", csub=3, cleg=0)
```

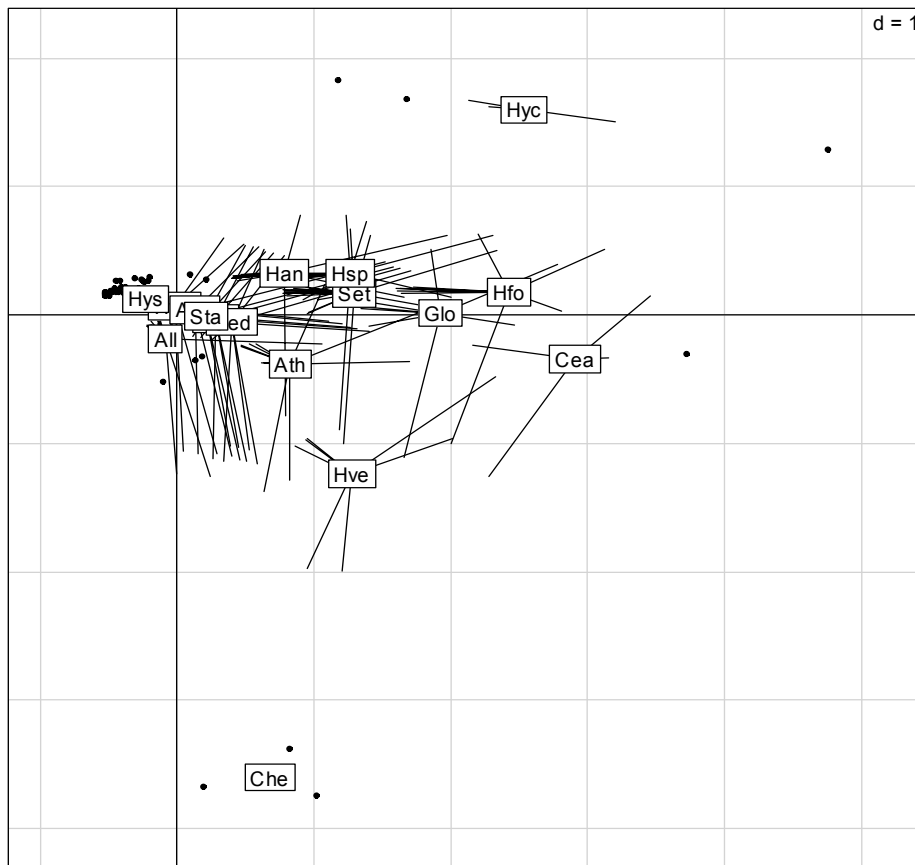


L'ACP est en fait souvent discutée sur ce genre de tableau parce que c'est une méthode qui introduit une grande dissymétrie entre le traitement des lignes et des colonnes. Elle est centrée sur la différence entre les lignes et la redondance entre colonnes. Dans un tableau sites-espèces on s'appuie sur la covariance entre espèces pour discriminer les sites, dans un tableau espèces-sites on fait l'inverse. Seule l'analyse des correspondances fait les deux, c'est à dire, donne un compromis des deux objectifs. L'important est de *ne pas utiliser la procédure à l'aveugle*. La difficulté duale est la grande diversité des structures de ce genre de tableaux et l'objectif est de trouver le particulier par le biais de procédure générale.

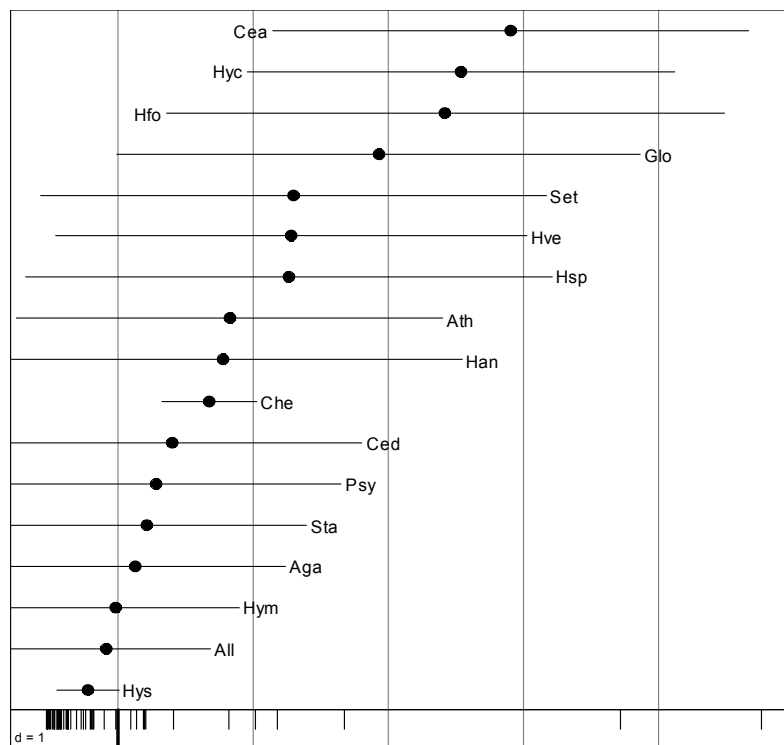
```
> data(trichometeo) # trichometeo$fau comporte 49 pièges et 17 espèces
> t1 <- data.frame(apply(log(trichometeo$fau+1),2,function(x) x/sum(x)))
> s.distri(dudi.pca(t1,scan=F,scale=F)$11,t1,cell=0,csta=0.3,clab=1)
```

Les dénombrements sont généralement transformés par $x \mapsto \log(x+1)$. Le tableau (Usseglio-Polatera and Auda 1987) est passé en pourcentage par colonnes (espèces). Un taxon est une distribution de fréquence entre sites sur une colonne. La moyenne est $1/n$ et la colonne centrée est l'écart entre la distribution de l'espèce et la distribution uniforme. Une composante principale est un score normalisé des sites et la coordonnée de l'espèce est exactement la position moyenne sur ce score. L'analyse maximise la

somme des carrés des écarts des positions moyennes à l'origine qui est la position du taxon indifférent :

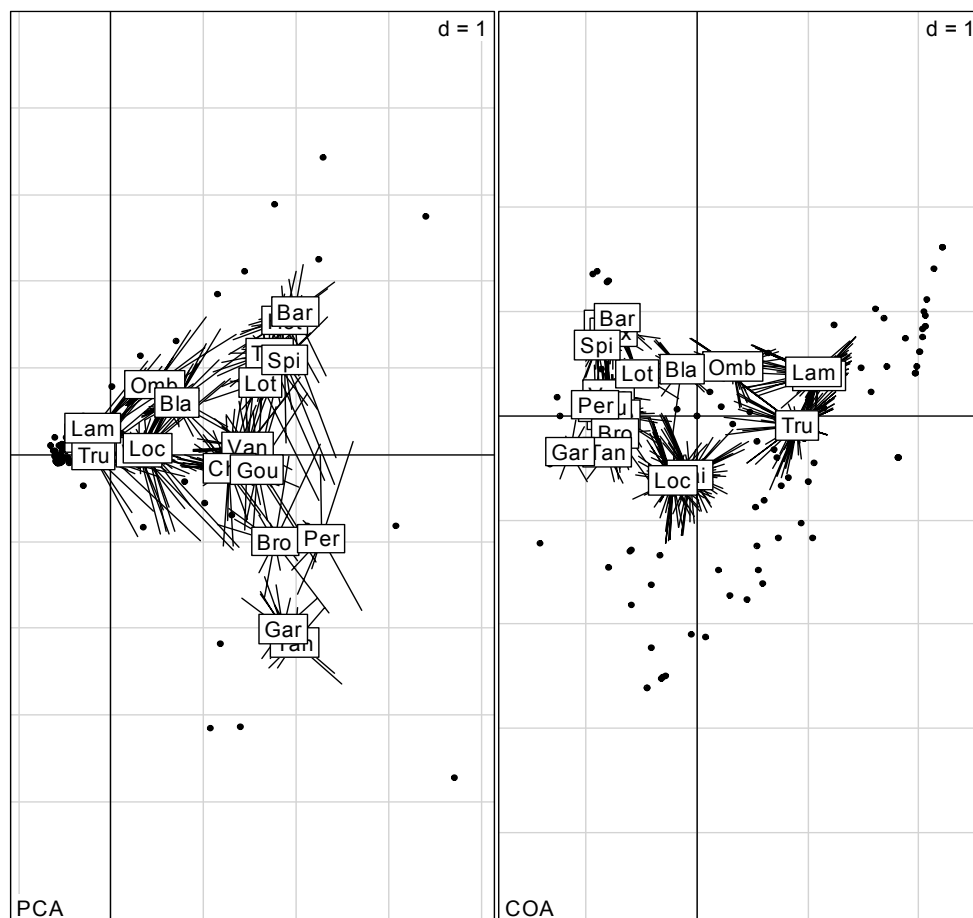


```
> score.distrib(dudi.pca(t1, scan=F, scale=F)$l1[,1], t1)
```



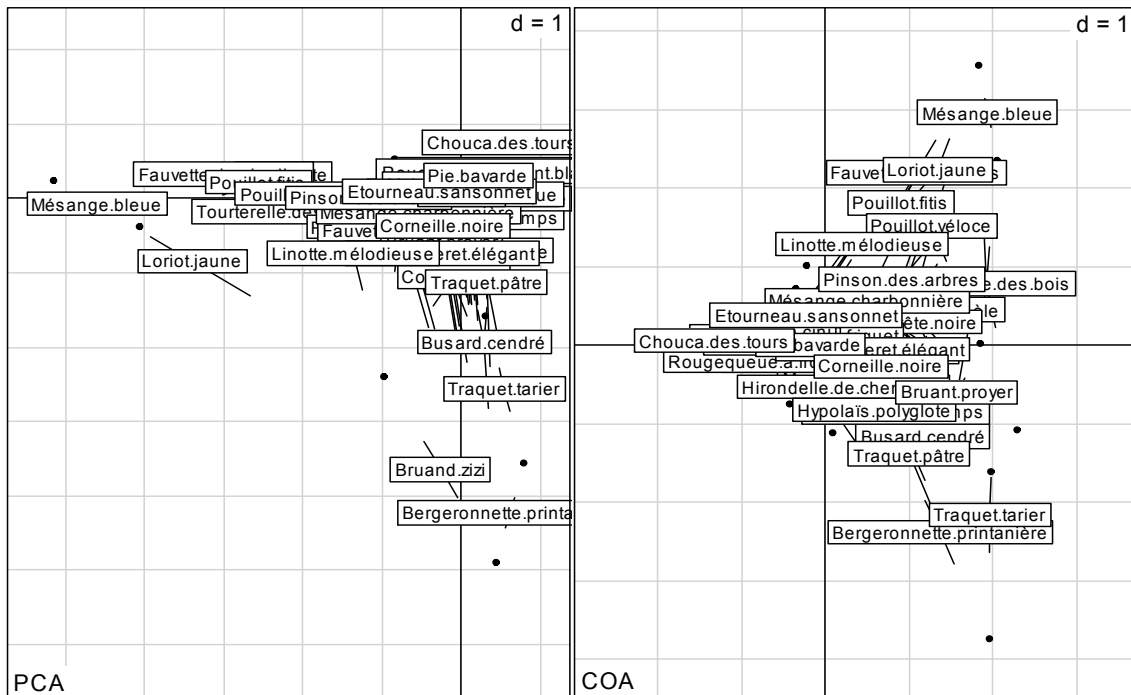
L'essentiel est que certains relevés ont décalé la majorité des espèces dans le même sens (une configuration météorologique particulière provoque l'émergence simultanée des trichoptères). Autre ambiguïté : la covariance de deux variables est une mesure de ressemblance même quand elle est négative (une corrélation de -1 indique qu'on a mesuré exactement la même chose au signe près). Pour deux espèce, c'est au contraire une mesure de différence : une espèce disparaît quand l'autre apparaît. L'ACP a alors le statut de méthode qui fait une double typologie. Le même calcul n'a pas le même sens expérimental.

```
> t1 <- data.frame(apply(jv73$poi,2,function(x) x/sum(x)))
> par(mfrow=c(1,2))
> s.distri(dudi.pca(t1,scan=F,scale=F)$l1,t1,
  cstar=0.3,clab=1,cell=0,axesel=F,sub="PCA")
> s.distri(dudi.coa(jv73$poi,scan=F)$l1,jv73$poi,
  cstar=0.3,clab=1,cell=0,axesel=F,sub="COA")
```



Les deux analyses parlent de la typologie de distributions des espèces. La carte de l'ACP est bien meilleure en ce sens qu'elle indique le gradient amont-aval de richesse croissante (salmonidés-cyprinidés) et dans la seconde partie le gradient de vitesse de courant (lénitique-lentique). Les contraintes de l'AFC déforme sans gain précis cette vision juste du réseau hydrographique.

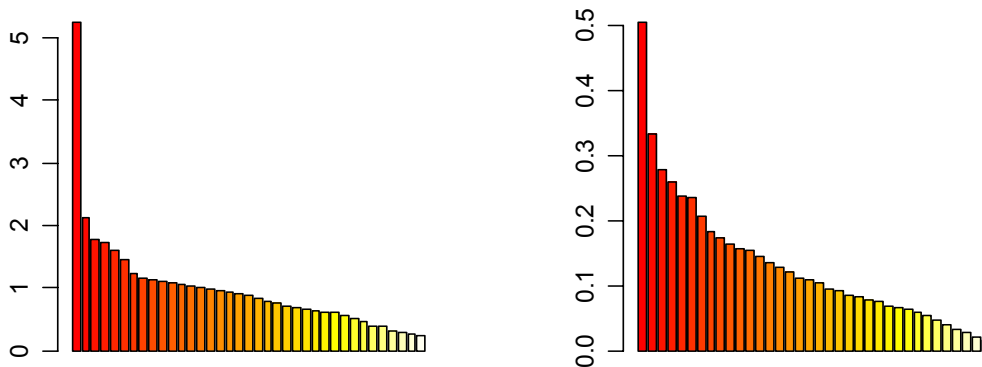
```
> data(aviurba)
> names(aviurba$fau) <- aviurba$species.names.fr
> t1 <- data.frame(apply(aviurba$fau,2,function(x) x/sum(x)))
> par(mfrow=c(1,2))
> s.distri(dudi.pca(t1,scan=F,scale=F)$l1,t1,
  cstar=0.3,clab=0.75,cell=0,axesel=F,sub="PCA")
> s.distri(dudi.coa(aviurba$fau,scan=F)$l1,aviurba$fau,
  cstar=0.3,clab=0.75,cell=0,axesel=F,sub="COA")
```



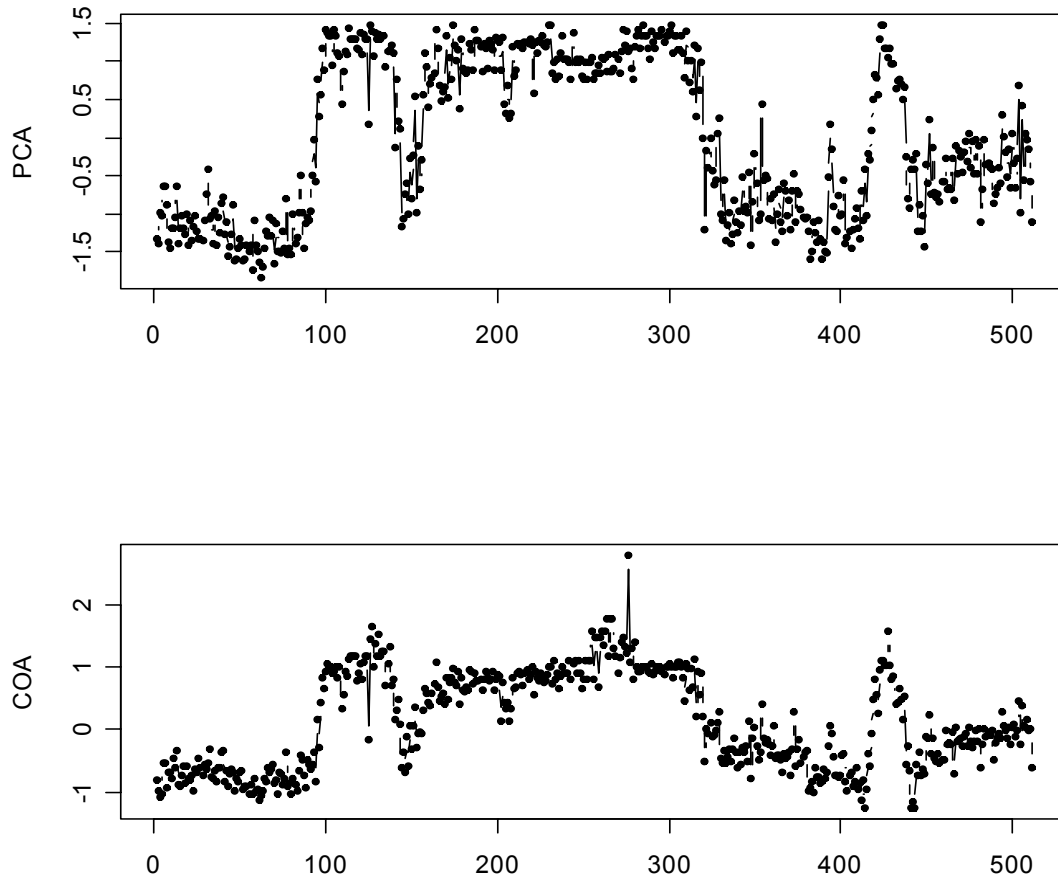
Deux gradients centre-ville -> rural fermé et centre-ville vers rural ouvert pour l'ACP correspondent aux deux gradients urbain-rural et ouvert-fermé pour l'AFC. L'avantage est à la seconde. Il n'y a pas de meilleures méthodes, il n'y a que de meilleures interactions entre une méthode et un objet. Le même raisonnement s'applique aux tableaux espèces-relevés en pourcentage par lignes (profils espèces) centrés par lignes. *Les moyennes se calculent toujours par espèces et la normalisation par espèces est à éviter.* Si on désire d'abord éviter toute réflexion préalable, on peut préférer l'analyse des correspondances (on aura souvent un résultat acceptable mais ce peut être une erreur totale).

> data(steppe)

512 relevés par quadrat sur un transect de 5.12 km ont enregistré la présence absence de 37 espèces (milieu steppique). A gauche l'ACP du tableau non modifié, à droite son AFC :



```
> par(mfrow=c(2,1))
plot(dudi.pca(steppe$tab,scan=F,scale=F)$li[,1],
      pch=20,ylab="PCA",xlab="",type="b")
> plot(dudi.coa(steppe$tab,scan=F)$li[,1],pch=20,ylab="COA",xlab="",type="b")
```



On obtient cette fois un résultat très voisin. L'ACP n'est cependant qu'un point de départ dans la question finalement difficile de l'*ordination* (Estève 1978) des tableaux florofaunistiques. Ces exemples illustrent l'insertion de la procédure dans un comportement adapté à chaque cas.

7 Références

Estève, J. 1978. Les méthodes d'ordination : éléments pour une discussion. Pages 223-250 in J. M. Legay and R. Tomassone, editors. Biométrie et Ecologie. Société Française de Biométrie, Paris.

- Gaschnard-Fossati, O. 1986. Répartition spatiale des macroinvertébrés benthiques d'un bras vif du Rhône. Rôle des crues et dynamique saisonnière. Thèse de doctorat, Université Lyon 1.
- Goodall, D. W. 1954. Objective methods for the classification of vegetation III. An essay in the use of factor analysis. *Australian Journal of Botany* **2**:304-324.
- Jolicœur, P., and J. E. Mosimann. 1960. Size and shape variation in the painted turtle. A principal component analysis. *Growth* **24**:339-354.
- Lascaux, J. M. 1996. Analyse de la variabilité morphologique de la truite commune (*Salmo trutta* L.) dans les cours d'eau du bassin pyrénéen méditerranéen. Thèse de doctorat en sciences agronomiques, INP Toulouse.
- Usseglio-Polatera, P., and Y. Auda. 1987. Influence des facteurs météorologiques sur les résultats de piégeage lumineux. *Annales de Limnologie* **23**:65-79.