

BING-F4002 Acquisition et analyse de données

Fiche TP 3 : Ordinations

Dufrêne M. - Gilbert M.

(avec la collaboration initiale de - Barbier N. - Deblauwe V.)

Version d'octobre 2015

Solutions proposées (mais il y en d'autres ...)

TP : Q1

Réalisez une ACP sur le fichier Xylobios_ecologie.txt en ne prenant pas en compte les variables X, Y, Sampl_structure, Alt et Region ?

Comment interpréter-vous les résultats et que décidez-vous de faire ?

A priori, on vérifie d'abord les distributions des variables pour éventuellement décider de les transformer. Ici, on décide de faire une première ACP sur les données brutes.

Solution potentielle pour la Q1

Initiation et remise à zéro des données précédentes

```
rm(list=ls())
```

on élimine les données stockées

Lecture de données

```
data=read.table('Xylobios_ecologie.txt', h=T, sep="\t", row.names="Sitecode")
```

Supprimer 5 variables

```
data$X<- NULL
```

```
data $Y<- NULL
```

```
data $Sampl_structure <- NULL
```

```
data $Alt <- NULL
```

```
data $Region <- NULL
```

```
data$Cov_flr <- NULL
```

Appel de l'ACP

```
library(vegan)
```

```
acp <- rda(data, scale=TRUE) # scale=true => matrice de corrélation;
```

```
summary(acp)
```

Partitioning of correlations:

	Inertia	Proportion
Total	35	1
Unconstrained	35	1

Eigenvalues, and their contribution to the correlations

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	8.5829	6.2260	4.1927	2.91649	2.78974	1.94293
Proportion Explained	0.2452	0.1779	0.1198	0.08333	0.07971	0.05551
Cumulative Proportion	0.2452	0.4231	0.5429	0.62623	0.70594	0.76145

⇒ Variance des deux premiers axes très faibles ...

Graphique amélioré

plot(acp)

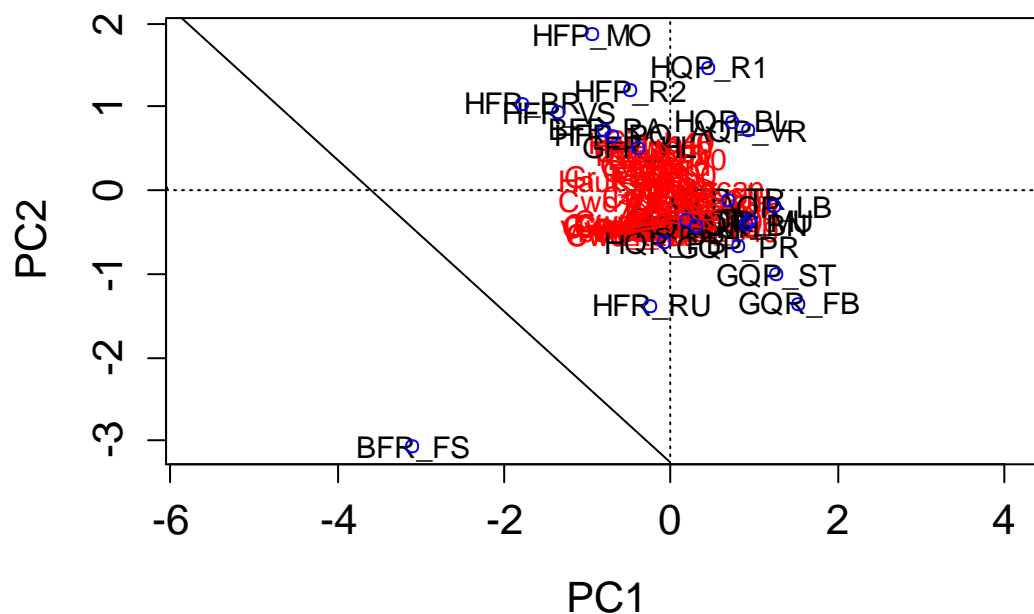
coordo <- scores(acp)

coordovar = data.frame(coordo\$species)

arrows(0,0, coordovar\$PC1, coordovar\$PC2, col='red')

coordoobj = data.frame(coordo\$sites)

points(coordoobj\$PC1, coordoobj\$PC2, col='blue')



Dessin très confus et avec visiblement une station « outlier » (BFR_FS). On va donc vérifier les contributions des variables aux axes.

Contributions relatives des objets aux axes

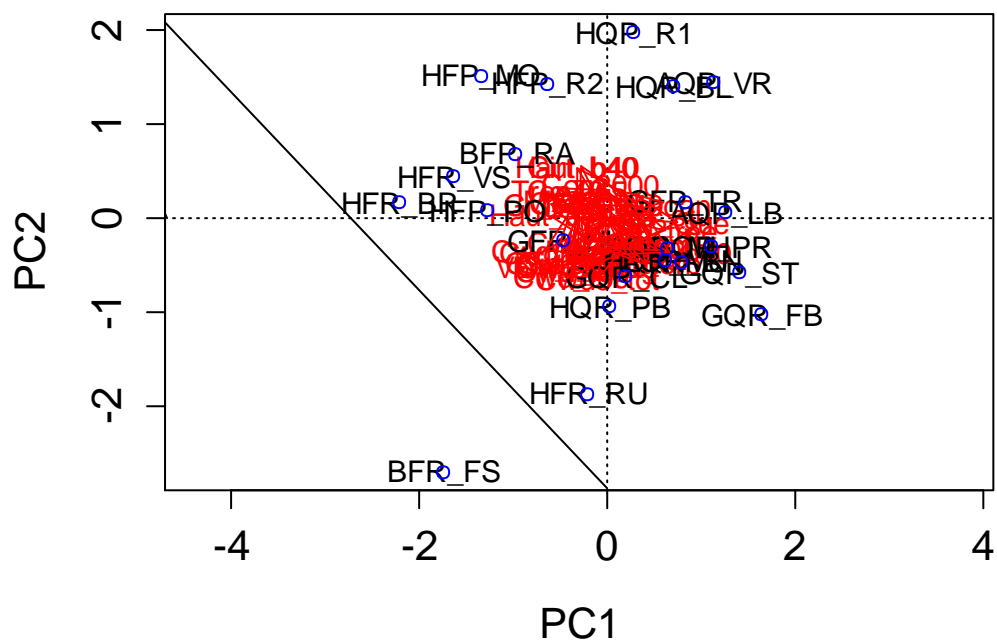
```

Cos2= data.frame(goodness(acp, display = "sites"))
Inertie = data.frame(inertcomp(acp, display = "sites"))
Prd.PC1 = Inertie$CA*Cos2$PC1
CTRObj1 = round(Prd.PC1/sum(Prd.PC1)*100, digits = 2)
Prd.PC2 = Inertie$CA*Cos2$PC2
CTRObj2 = round(Prd.PC2/sum(Prd.PC2)*100, digits = 2)
CTRObj = data.frame(CTRObj1, CTRObj2, row.names=row.names(Cos2))

```

	CTRObj1	CTRObj2
HQP_BL	1.96	2.19
AQR_BN	3.04	2.01
HFR_BR	11.74	8.43
GQR_CL	0.11	0.25
GQR_FB	8.54	7.80
BFR_FS	35.29	35.07
GFR_HL	0.57	0.74
AQP_LB	5.71	3.33
AQR_MH	0.32	0.47
HFP_MO	3.33	7.37
HQR_MU	3.31	2.10
HQR_PB	0.02	0.62
HFP_PO	1.88	1.76
GQP_PR	2.34	2.06
HQP_R1	0.72	3.82
HFP_R2	0.90	2.79
BFP_RA	2.41	2.21
HFR_RU	0.26	3.11
GQP_ST	5.85	4.94
GFP_TR	1.77	1.04
AQP_VR	3.18	2.63
HFR_VS	6.75	5.26

La station BFR_FS a des valeurs exceptionnelles => On transforme en log toutes les variables et on regarde ce que cela donne.



Amélioration et contribution plus équilibrée des différentes stations.

	CTRobj1	CTRobj2
HQP_BL	1.83	4.14
AQR_BN	2.30	1.66
HFR_BR	18.15	10.56
GQR_CL	0.12	0.65
GQR_FB	9.73	7.27
BFR_FS	11.39	17.83
GFR_HL	0.81	0.55
AQP_LB	5.71	3.32
AQR_MH	1.48	1.00
HFP_MO	6.67	7.46
HQR_MU	1.43	1.14
HQR_PB	0.00	1.35
HFP_PO	5.95	3.46
GQP_PR	4.46	2.73
HQP_R1	0.28	6.27
HFP_R2	1.54	4.11
BFP_RA	3.51	2.77
HFR_RU	0.17	5.44
GQP_ST	7.25	4.71
GFP_TR	2.55	1.53
AQP_VR	4.60	5.90
HFR_VS	10.06	6.15

On sort alors un cercle des corrélations plus propre.

Cercle des corrélations

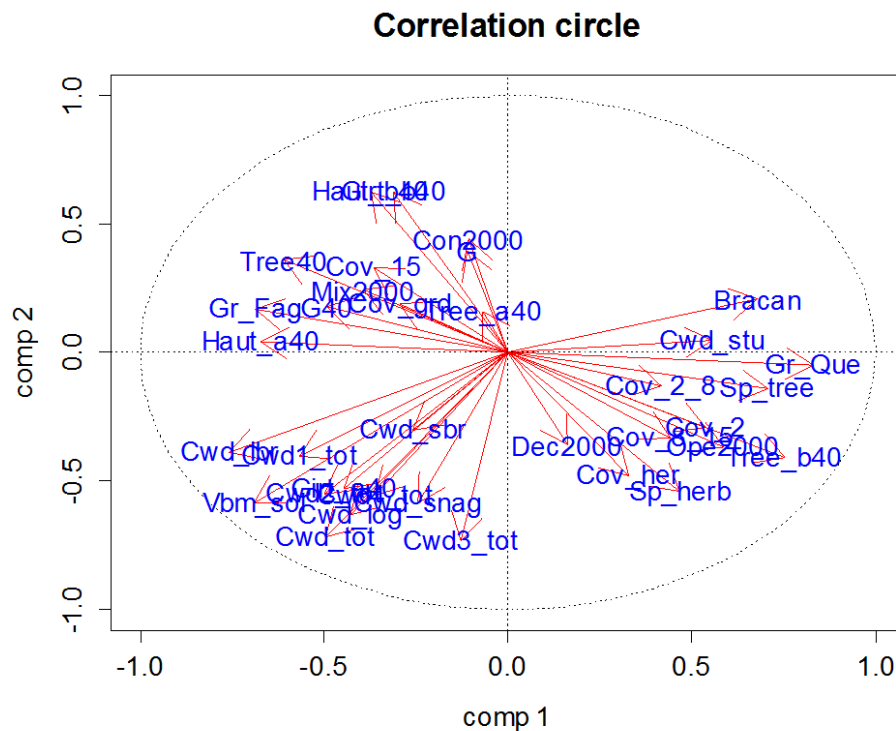
coordo <-scores(acp)

correlation = data.frame(cor(data, coordo\$sites))

```

a <- seq(0,2*pi,length=100)
plot( cos(a), sin(a),
      type = 'l', lty = 3,
      xlab = 'comp 1', ylab = 'comp 2',
      main = "Correlation circle")
arrows(0,0, correlation$PC1, correlation$PC2, col='red')
text(correlation$PC1, correlation$PC2,rownames(correlation), col='blue')
abline(h=0, lty=3) # axe horizontal (lty = type de la ligne)
abline(v=0, lty=3) # axe vertical (lty = type de la ligne)

```

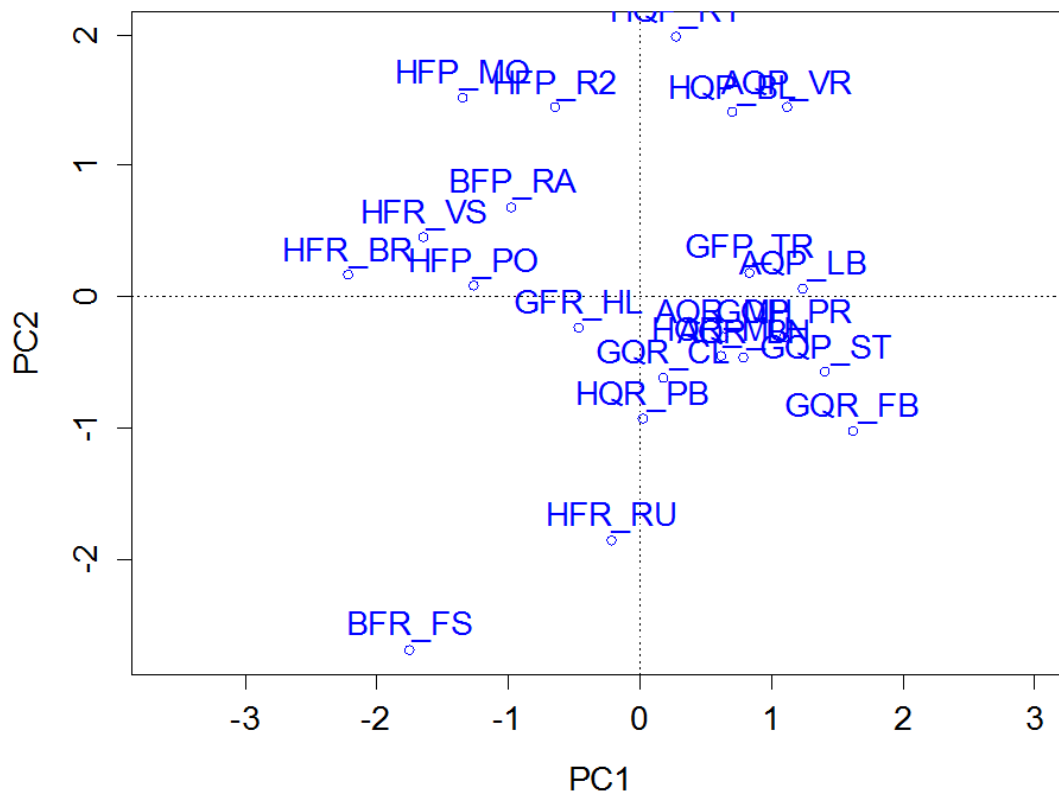


Ainsi qu'un graphique propre avec les noms des stations.

```

# Graphique amélioré
plot(acp, type = 'n')
coordo <- scores(acp)
coordovar = data.frame(coordo$species)
coordobj = data.frame(coordo$sites)
points(coordobj$PC1,
        coordobj$PC2, col='blue')
text(coordobj$PC1,
      coordobj$PC2, col='blue', pos = 3, labels = rownames(data) )

```



TP : Q2

Réalisez une AFC sur le fichier **Demazy_2013_carabides.txt**.

Comment interpréter-vous les résultats et que décidez-vous de faire ?

Pouvez-vous ajouter le résultat d'un groupement sur ce graphique ?

Solution potentielle pour la Q2

Initiation et remise à zéro des données précédentes

```
rm(list=ls())
```

on élimine les données stockées

Lecture de données

```
data=read.table(' Demazy_2013_carabides.txt ', h=T, sep="\t", row.names="Placettes")
```

AFC

```
esp.ca <- cca(esp)
```

```
summary(esp.ca)
```

```
call:
cca(X = esp)
```

Partitioning of mean squared contingency coefficient:

	Inertia	Proportion
Total	1.713	1
Unconstrained	1.713	1

Eigenvalues, and their contribution to the mean squared contingency coefficient

Importance of components:

	CA1	CA2	CA3	CA4	CA5	CA6
Eigenvalue	0.4838	0.3663	0.2348	0.12149	0.11175	0.08964
Proportion Explained	0.2824	0.2138	0.1371	0.07092	0.06523	0.05233
Cumulative Proportion	0.2824	0.4962	0.6332	0.70417	0.76940	0.82173

Les deux premiers axes expliquent presque 50% de la variance du jeu de données.

Graphique amélioré

```
plot(esp.ca, type = "n") # plot vide
# plot(esp.ca, type = "n", xlim=c(-2, 2), ylim=c(-2, 2)) # plot vide avec des axes limités
coordolig <- as.data.frame(scores(esp.ca, scaling=1, display="wa"))
coordocol <- as.data.frame(scores(esp.ca, scaling=1, display="spe"))
```

coordonnées des colonnes

```
points(coordocol$CA1, coordocol$CA2,
       col='darkred',
       pch = 16)
```

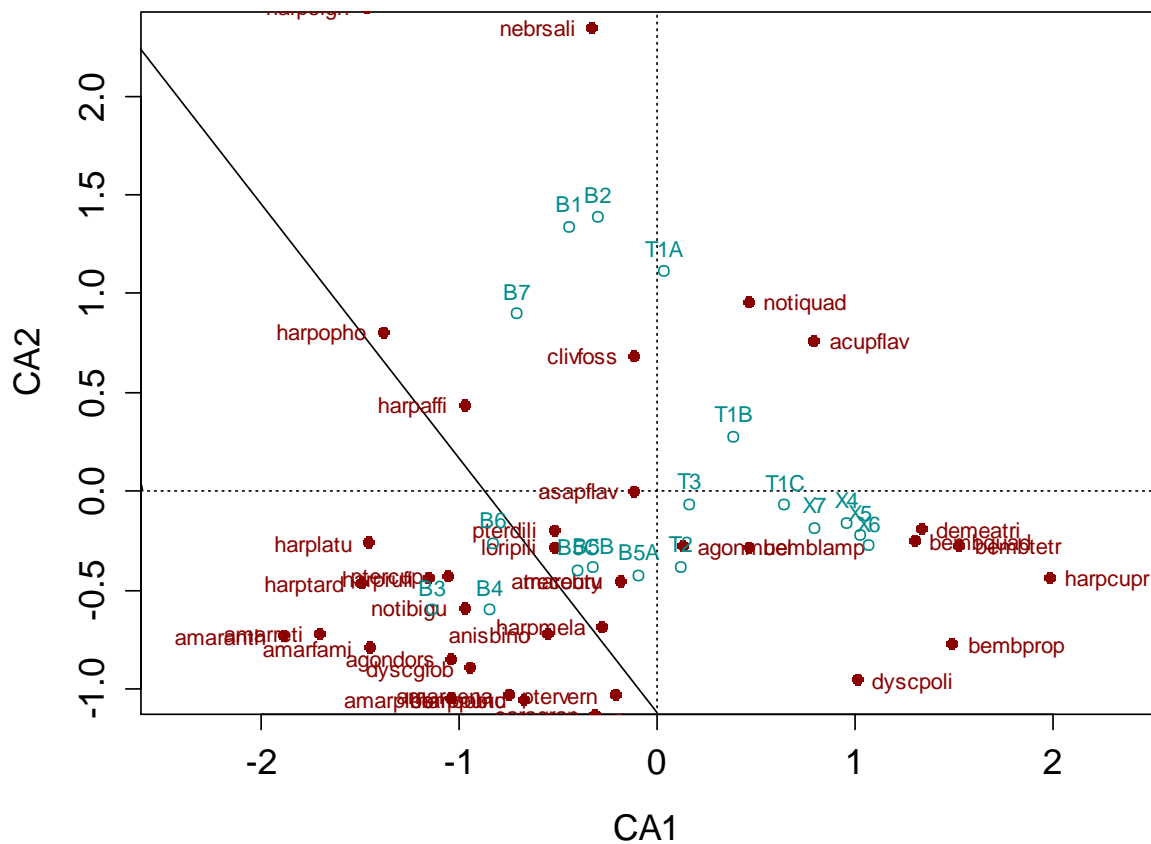
```
coordocol$pos <- ifelse(coordocol$CA1 < 0, 2, 4)
```

```
text(coordocol$CA1, coordocol$CA2,
     labels = rownames(coordocol), # label de stations sur le graphique
     col = 'darkred', # couleur du label (peut être un vecteur)
     pos = coordocol$pos, # position du label
     cex = 0.6) # taille du label
```

coordonnées des lignes

```
points(coordolig$CA1, coordolig$CA2,
       col='darkcyan',
       pch = 1)
```

```
text(coordolig$CA1, coordolig$CA2,
     labels = rownames(coordolig), # label de stations sur le graphique
     col = 'darkcyan', # couleur du label (peut être un vecteur)
     pos = 3, # position du label
     cex = 0.6) # taille du label
```



Qualité des représentations

```
Cos2= data.frame(goodness(esp.ca, display = "species"))
```

```
Inertie = data.frame(inertcomp(esp.ca, display = "species"))
```

#sum(Inertie.spe) = variance du jeu de données

```
species = rownames(Cos2)
```

$$\text{Prd.CA1} = \text{Inertie\$CA} * \text{Cos2\$CA1}$$

CTR.CA1 = round(Prd.CA1/sum(Prd.CA1)*100, digits = 2)

$$\text{Prd.CA2} = \text{Inertie\$CA} * \text{Cos2\$CA2}$$

CTR.CA2 = round(Prd.CA2/sum(Prd.CA2)*100, digits = 2)

#Prd.CA3 = Inertie\$CA*Cos2\$CA3 # si on veut un 3eme axe

```
#CTR.CA3 = round(Prd.CA3/sum(Prd.CA3)*100, digits = 2)
```

Cos2\$CA1 = round(Cos2\$CA1*100, digits = 2)

```
Cos2$CA2 = round(Cos2$CA2*100, digits = 2)
```

```
#Cos2$CA3 = round(Cos2$CA3*100, digits = 2)
```

```
#Qual = data.frame(species, CTR.CA1, CTR.CA2, CTR.CA3, Cos2$CA1, Cos2$CA2, Cos2$CA3 )
```

```
Qual = data.frame(species, CTR.CA1, CTR.CA2, Cos2$CA1, Cos2$CA2)
```

Qual

	species	CTR.CA1	CTR.CA2	Cos2.CA1	Cos2.CA2
1	acupbrun	0.11	0.11	7.16	11.92
2	acupflav	0.02	0.02	2.17	3.71

3	agondors	0.12	0.10	8.61	12.87
4	agonmuel	0.10	0.25	0.94	4.36
5	agonsepx	0.01	0.09	0.13	3.51
6	amaraena	1.03	1.42	5.19	12.56
7	amaranth	1.18	0.74	29.97	33.34
8	amareury	0.00	0.01	0.31	1.65
9	amarfami	0.47	0.32	10.47	12.78
10	amarpleb	2.55	2.55	14.68	25.82
11	amarreti	0.11	0.07	10.13	11.47
12	anisbino	0.73	0.94	5.04	11.40
13	asapflav	0.02	0.01	0.50	0.50
14	bemblamp	4.06	2.97	36.33	46.64
15	bembobtu	0.02	0.03	2.43	6.87
16	bembprop	0.57	0.39	30.86	37.10
17	bembquad	4.67	2.73	35.48	36.46
18	bembtetr	43.31	25.24	84.98	87.02
19	caraaura	3.21	2.37	35.75	46.28
20	caragran	0.01	0.09	0.58	6.20
21	clivfoss	0.02	0.30	0.48	12.88
22	demeatri	0.06	0.04	8.38	8.51
23	dyscglob	2.98	2.83	41.30	68.82
24	dyscpoli	0.07	0.07	7.35	12.28
25	harpaffi	3.63	2.38	23.81	27.47
26	harpcupr	0.14	0.08	8.79	9.11
27	harplat	1.48	0.86	47.92	49.05
28	harpmea	0.10	0.31	1.05	5.79
29	harpopho	0.42	0.30	14.61	18.35
30	harppunc	0.02	0.03	2.43	6.87
31	harprufi	5.05	3.19	64.82	71.84
32	harpsign	0.24	0.42	4.89	15.14
33	harptard	2.55	1.55	32.94	35.31
34	loripili	0.21	0.15	5.15	6.26
35	nebrsali	1.49	33.12	2.53	98.38
36	notibigu	1.51	1.09	45.46	57.76
37	notiquad	0.10	0.25	4.77	20.56
38	ptercupr	8.42	5.37	52.05	58.31
39	pterdili	0.09	0.06	2.04	2.24
40	ptermela	0.30	0.22	34.33	42.98
41	ptervern	0.03	0.33	0.92	16.68
42	ptervers	8.66	6.49	48.51	63.90
43	stompumi	0.11	0.11	7.16	11.92
44	trecobtu	0.00	0.01	0.31	1.65

On a une espèce qui contribue bcp aux axes. Peut-être intérêt à faire aussi une transformation log préalable.

Ajouter un groupement ...

TP : Q3

Réalisez une PCOA sur le fichier **Demazy_2013_carabides.txt**.

Comment interpréter-vous les résultats et que décidez-vous de faire ?

Quel comparaison pouvez-vous faire par rapport à celui de l'AFC ?

Pouvez-vous ajouter le résultat d'un groupement sur ce graphique ?

Solution potentielle pour la Q3

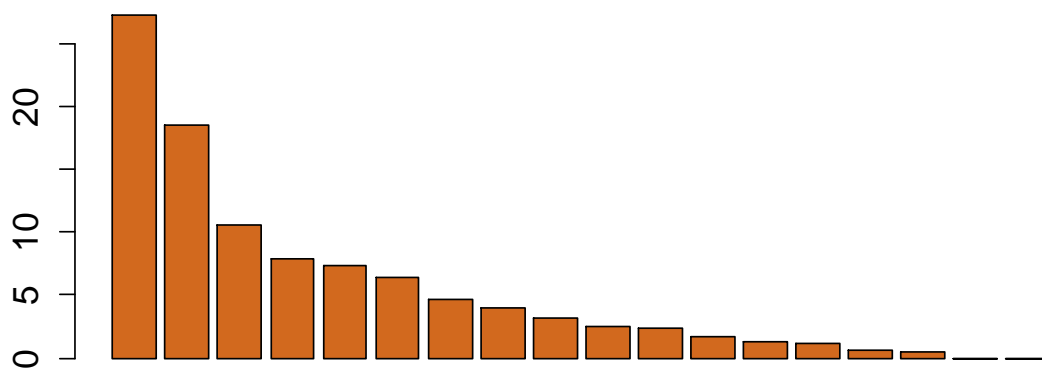
```

# Initiation et remise à zéro des données précédentes
rm(list=ls())                                     # on élimine les données stockées
# Lecture de donnees
data=read.table(' Demazy_2013_carabides.txt ', h=T, sep="\t", row.names="Placettes")

# Matrice de distance
d14=vegdist(data, method="bray")
# PCoA
pcoa <- cmdscale(d14, add= TRUE, eig= TRUE)

# Calcul des valeurs propres
valeurs_propres = round(pcoa$eig/sum(pcoa$eig)*100, digits = 2)
barplot(valeurs_propres, col = "chocolate")

```



valeurs_propres

```

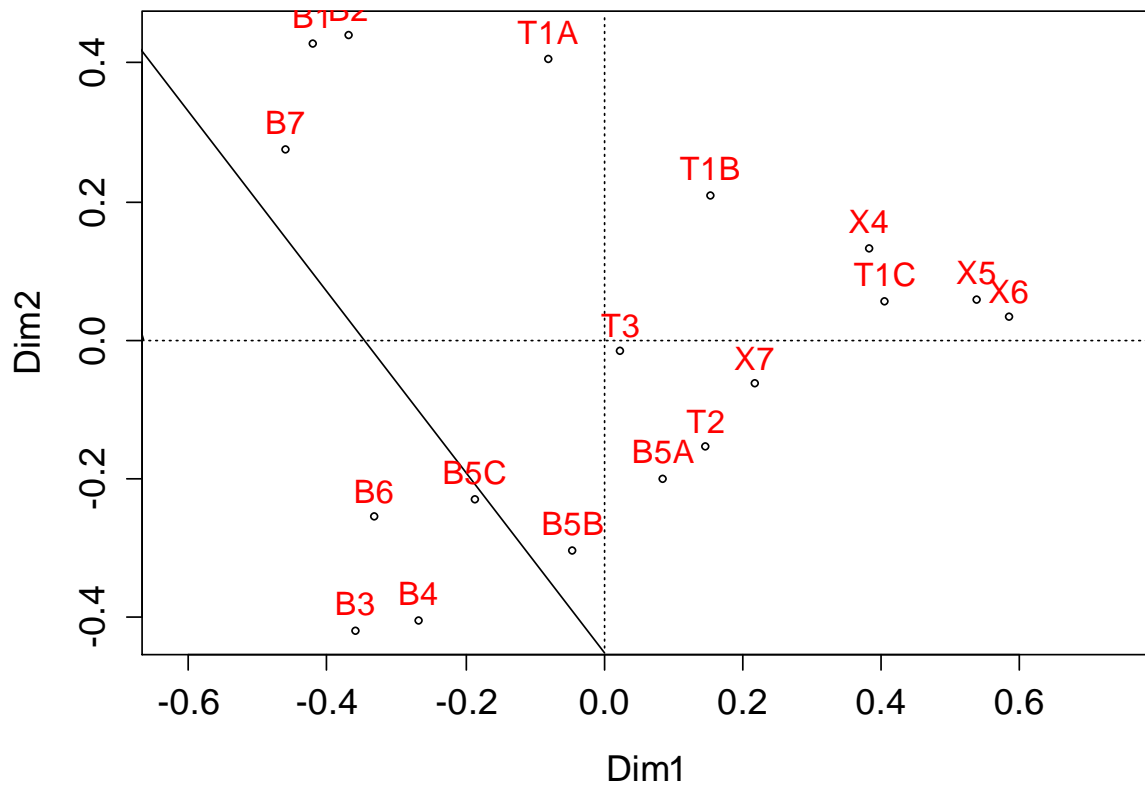
[1] 27.23 18.50 10.57 7.91 7.38 6.46 4.61 3.98 3.14 2.54 2.39
[12] 1.65 1.35 1.18 0.65 0.43 0.00 0.00

```

```

# graphique amélioré
figure <- ordiplot(pcoa, type="points")
abline(h=0, lty=3)                               # axe horizontal (lty = type de la ligne)
abline(v=0, lty=3)                               # axe vertical (lty = type de la ligne)
text(figure, "sites", labels= rownames(esp),col="red", cex=0.9, pos=3)

```



On a une belle opposition des parcelles bios par rapport aux parcelles traditionnelles et en régime de techniques culturales simplifiées.

graphique avec les coordonnées pondérées des espèces

```
ordiplot(pcoa, type="points")
abline(h=0, lty=3); abline(v=0, lty=3)
text(figure, "sites", labels= rownames(esp), col="red", cex=0.9, pos=3)
pcoa.wa <- wascores(coordo, esp)
points(pcoa.wa, col='darkcyan', pch = 1)
text(pcoa.wa, rownames(pcoa.wa), cex=0.7, col="blue")
```


* * *