

Aula 12 - Ficha de Trabalho 10

Tiago A. Marques

December 3, 2018

Preliminary steps

We load up library `vegan` and we source (i.e. read in) a file that has some useful functions for exploring PCA outputs (this file is in the folder for Aula T23).

```
library(vegan)
source("C:/Users/tam2/Dropbox/Trabalho/DBA/EcologiaNumerica/2018/Aulas/Teoricas/Aula23/brocardfunctions
```

Exercício 1

Recolheram-se dados de abundância de várias espécies de anfíbios em vários distritos de Portugal (DataTP10anfíbios.csv). Efectue uma análise de componentes principais aos dados e retire as principais conclusões.

As usual, we begin by reading the data

```
anfíbios<-read.csv2("DataTP10anfíbios.csv")
```

and we check it all looks OK

```
str(anfíbios)

## 'data.frame': 15 obs. of 7 variables:
## $ Distritos : Factor w/ 15 levels "AVEI","BEJA",...: 7 2 8 13 11 10 5 14 9 1 ...
## $ Rana : num 4.44 2.18 0.47 1.22 4.66 2.68 1.45 0 1 2.35 ...
## $ Triturus : num 0 0 0 2.45 0 0 1.88 0 0 0 ...
## $ Salamandra : num 35.96 41.1 35.38 20.92 5.12 ...
## $ Hyla : num 3.2 0 0.71 0 1.82 0 0 1.5 1.71 6.47 ...
## $ Bufo : num 4.8 16.4 14.8 44.1 23.6 ...
## $ Pleurodeles: num 16 0.48 9.3 11.46 20.76 ...
```

We have 15 observations for 7 variables, which correspond respectively to locations and taxa (genera of amphibians). The first column does not represent a genera, but it contains the labels of the locations

```
anfíbios[,1]

## [1] EVOR BEJA FARO SETU LISB LEIR CBRA VISE GUAR AVEI COIM PORT BRAG BRCA
## [15] VREA
## 15 Levels: AVEI BEJA BRAG BRCA CBRA COIM EVOR FARO GUAR LEIR LISB ... VREA
```

The names of the genera are

```
names(anfíbios)[2:7]

## [1] "Rana" "Triturus" "Salamandra" "Hyla" "Bufo"
## [6] "Pleurodeles"
```

We can also see and if some locations have much more amphibians than others

```
nbyloc=data.frame(loc=anfíbios[,1],n=rowSums(anfíbios[,2:7]))
nbyloc
```

```
##      loc      n
## 1  EVOR 64.40
## 2  BEJA 60.12
## 3  FARO 60.68
## 4  SETU 80.13
## 5  LISB 56.00
## 6  LEIR 64.44
## 7  CBRA 55.34
## 8  VISE 59.52
## 9  GUAR 44.71
## 10 AVEI 61.30
## 11 COIM 67.15
## 12 PORT 72.92
## 13 BRAG 63.86
## 14 BRCA 58.43
## 15 VREA  0.00
```

with Vila Real representing a really weird location, with no data (might that be a mistake?).

We can also evaluate what are the most abundant taxa,

```
colSums(anfibios[,2:7])
```

```
##      Rana      Triturus Salamandra      Hyla      Bufo Pleurodeles
##      21.99      5.44      146.68      36.63      471.83      186.43
```

By now, on lecture TP12, we should be fast at doing all the above.

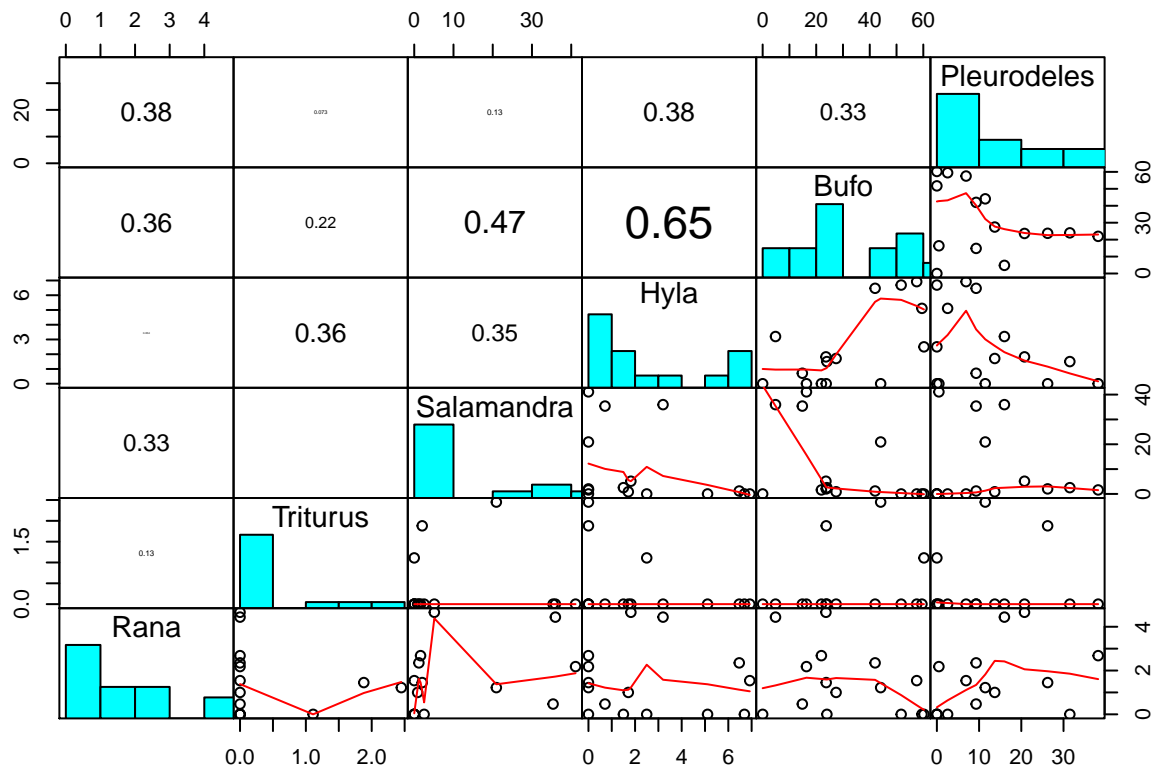
Given that a PCA is a dimension reduction technique, we can see if the genera are highly correlated or not. The analysis will be more efficient if there are high correlations to explore (e.g. two variables with perfect correlation can be summarized in a single component!)

```
panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}
panel.corP <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * r)
}
panel.corS <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y, method="spearman"))
```

```

txt <- format(c(r, 0.123456789), digits = digits)[1]
txt <- paste0(prefix, txt)
if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
text(0.5, 0.5, txt, cex = cex.cor * r)
}
pairs(anfibios[,2:7], lower.panel = panel.smooth, upper.panel = panel.corP,
      gap=0, rowlattice=FALSE,diag.panel = panel.hist)

```



As it turns out, there is actually no correlation between any two variables larger than 0.65, and this is hinting to the fact that the PCA might not be very efficient to reduce the dimensionality of this data set.

For the PCA plots it might be useful to define `row.names` in the object `anfibios`, as these will be used by default in the plotting of biplots, hence making the plots easier to interpret.

```
row.names(anfibios)=anfibios[,1]
```

Now, we implement the PCA itself

```
acp1<-princomp(anfibios[, -1])
acp1
```

```

## Call:
## princomp(x = anfibios[, -1])
##
## Standard deviations:
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6
## 21.3241377 13.8202048  8.8379189  2.0249230  1.0881376  0.5173327
##

```

```
## 6 variables and 15 observations.
```

and we visualize the summary

```
summary(acp1)
```

```
## Importance of components:
##                Comp.1    Comp.2    Comp.3    Comp.4
## Standard deviation  21.324138 13.8202048 8.8379189 2.02492300
## Proportion of Variance 0.623434 0.2618644 0.1070897 0.00562166
## Cumulative Proportion 0.623434 0.8852984 0.9923880 0.99800971
##                Comp.5    Comp.6
## Standard deviation  1.088137650 0.5173326744
## Proportion of Variance 0.001623361 0.0003669335
## Cumulative Proportion 0.999633067 1.0000000000
```

We can also explore some of the components of the PCA output, namely the loadings (i.e. the correlations between the PC's and the original variables)

```
acp1$loadings
```

```
##
## Loadings:
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## Rana                0.375 0.920
## Triturus            -0.239 0.195 -0.951
## Salamandra  0.471 -0.710 0.521
## Hyla                0.893 -0.330 -0.292
## Bufo            -0.861 -0.255 0.434
## Pleurodeles  0.171 0.655 0.732
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## SS loadings      1.000 1.000 1.000 1.000 1.000 1.000
## Proportion Var  0.167 0.167 0.167 0.167 0.167 0.167
## Cumulative Var  0.167 0.333 0.500 0.667 0.833 1.000
```

Note there is an argument to the function that prints the loadings that can be used to force all loadings to appear, since by default only those above 0.1 are displayed

```
print(acp1$loadings,digits = 2, cutoff = 0.05)
```

```
##
## Loadings:
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## Rana                0.06 0.38 0.92 0.10
## Triturus            -0.24 0.20 -0.95
## Salamandra  0.47 -0.71 0.52
## Hyla            -0.08                0.89 -0.33 -0.29
## Bufo            -0.86 -0.26 0.43 -0.06
## Pleurodeles  0.17 0.66 0.73                -0.07
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## SS loadings      1.00 1.00 1.00 1.00 1.00 1.00
## Proportion Var  0.17 0.17 0.17 0.17 0.17 0.17
## Cumulative Var  0.17 0.33 0.50 0.67 0.83 1.00
```

the scores (the coordinates of the locations in the principal components space) - note one could use these to manually plot the results of a PCA

```
acp1$scores
```

```
##          Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## EVOR  35.9372068 -9.447106  4.8185073  3.725788576  0.39758994
## BEJA  25.9322705 -26.132662  1.0883920 -1.005602272  0.47170637
## FARO  25.9645802 -15.938199  3.7883005 -0.737356805 -1.71864114
## SETU  -5.6126292 -11.697835 10.6619966 -3.377098293  0.85090691
## LISB   6.1061455  10.813623  0.4620701  1.348352884  2.53012718
## LEIR   8.9848858  25.213559 10.5964813 -0.466393130  0.23644253
## CBRA   5.5366438  16.597692  2.8113376 -1.813620589  0.31169056
## VISE   6.2936806  19.570766  6.8724397 -0.427544499 -2.25862036
## GUAR  -0.4735304   8.222119 -5.4756022 -0.578501670 -0.04296380
## AVEI -13.9550085   1.242643 -2.2288006  3.256310122  0.24966376
## COIM -30.6908939 -6.783789 -0.2957481 -0.008096814 -0.54770479
## PORT -28.3591494 -3.467991  2.1077881  2.418836310 -0.07153958
## BRAG -31.5838299 -8.565286 -1.7535031 -2.719565086  0.72190123
## BRCA -24.5776694 -6.527845 -5.5590466  1.756994063 -1.07362658
## VREA  20.4972974  6.900310 -27.8946126 -1.372502796 -0.05693222
##          Comp.6
## EVOR -0.534750516
## BEJA  0.673293918
## FARO  0.200646515
## SETU -0.826266873
## LISB  0.511082088
## LEIR  0.709903936
## CBRA -1.076690623
## VISE  0.118997893
## GUAR  0.355027113
## AVEI -0.365408342
## COIM  0.462117161
## PORT -0.008871033
## BRAG  0.208693187
## BRCA -0.265844705
## VREA -0.161929720
```

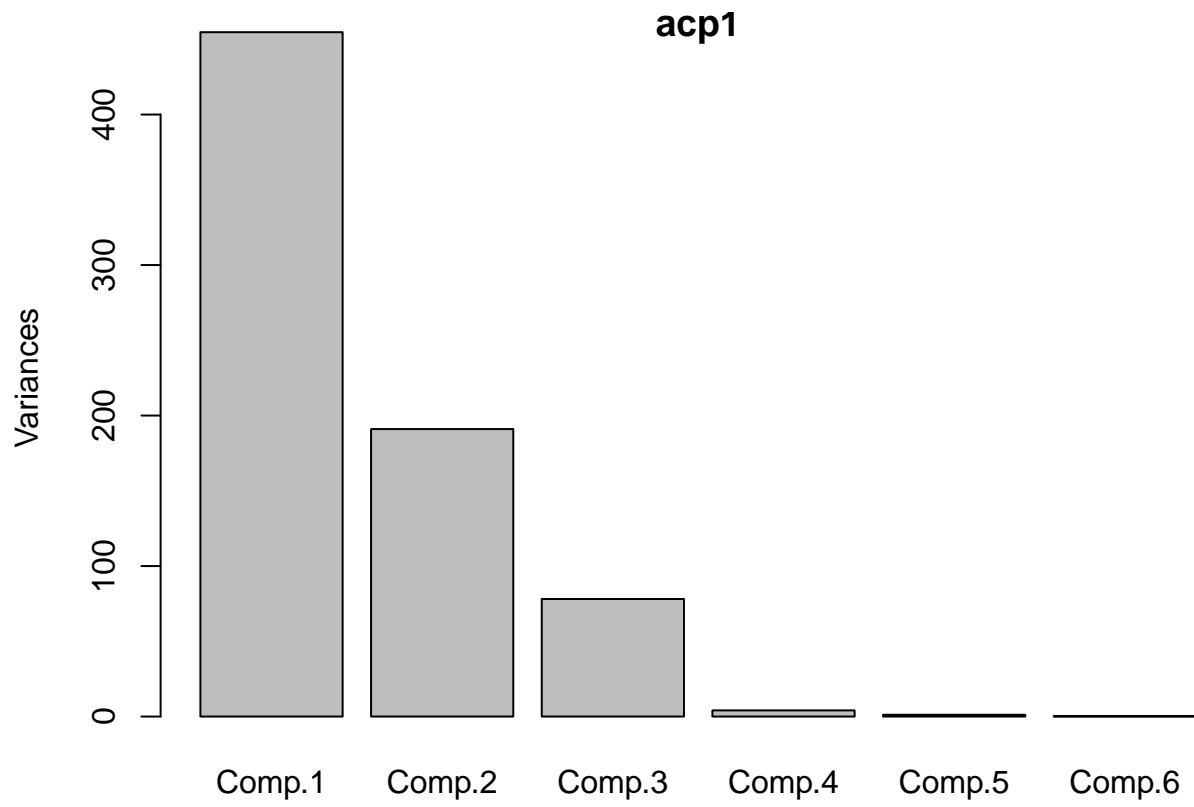
and the standard deviation associated with each PC

```
acp1$sdev
```

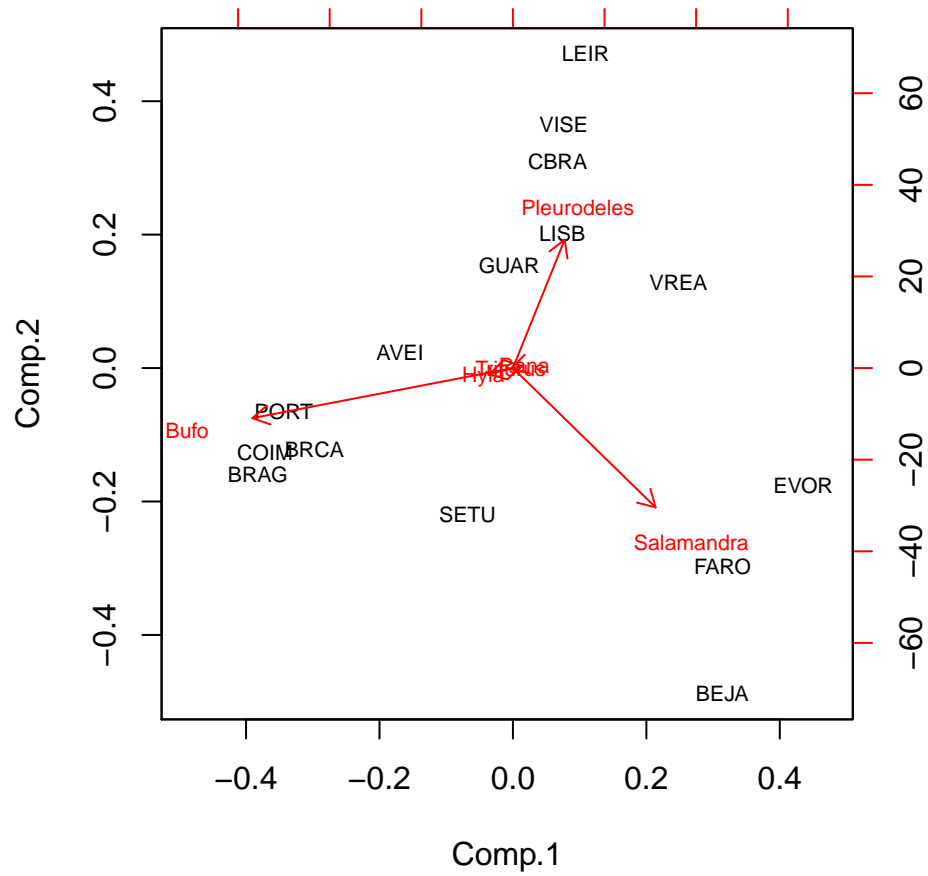
```
##          Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6
## 21.3241377 13.8202048  8.8379189  2.0249230  1.0881376  0.5173327
```

We can plot the objects produced by a PCA in two different ways

```
par(mfrow=c(1,1),mar=c(4,4,0.5,0.5))
plot(acp1)
```



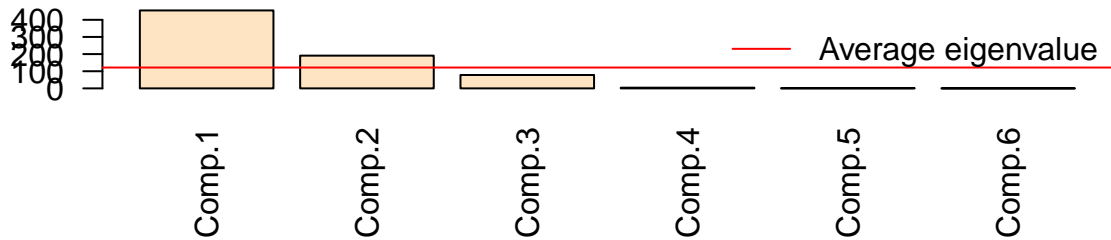
```
biplot(acp1, cex=0.7)
```



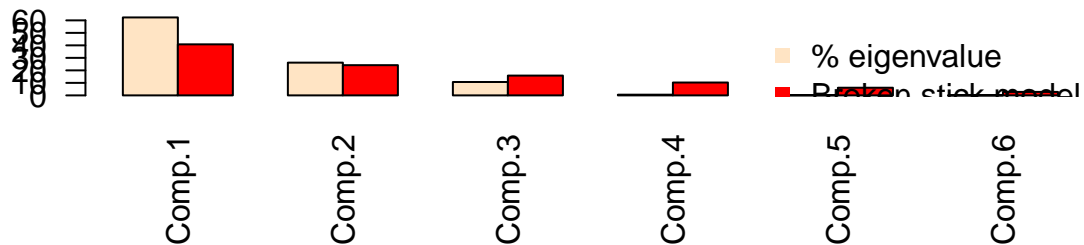
Finally, we can use one of the bespoke functions in file `brocardfunctions.R` to plot the variances associated with each PC and two criteria to choose the number of PCs to interpret

```
evplot(acp1$sdev^2)
```

Eigenvalues



% variation



Both criteria seem to agree that the more sensible would be to interpret just the first two PCs.

Note that we could have implemented the same analysis using function `rda` in library `vegan`, which produces slightly different output, but the same results

```
acpbyrda1<-rda(anfibios[,-1])
acpbyrda1
```

```
## Call: rda(X = anfibios[, -1])
##
##              Inertia Rank
## Total          781.5
## Unconstrained  781.5    6
## Inertia is variance
##
## Eigenvalues for unconstrained axes:
##  PC1  PC2  PC3  PC4  PC5  PC6
## 487.2 204.6 83.7  4.4  1.3  0.3
```

The summary of this object gives us direct access to something that was hidden above, namely the coordinates of the covariates in the PC space

```
summary(acpbyrda1)
```

```
##
## Call:
## rda(X = anfibios[, -1])
##
## Partitioning of variance:
```



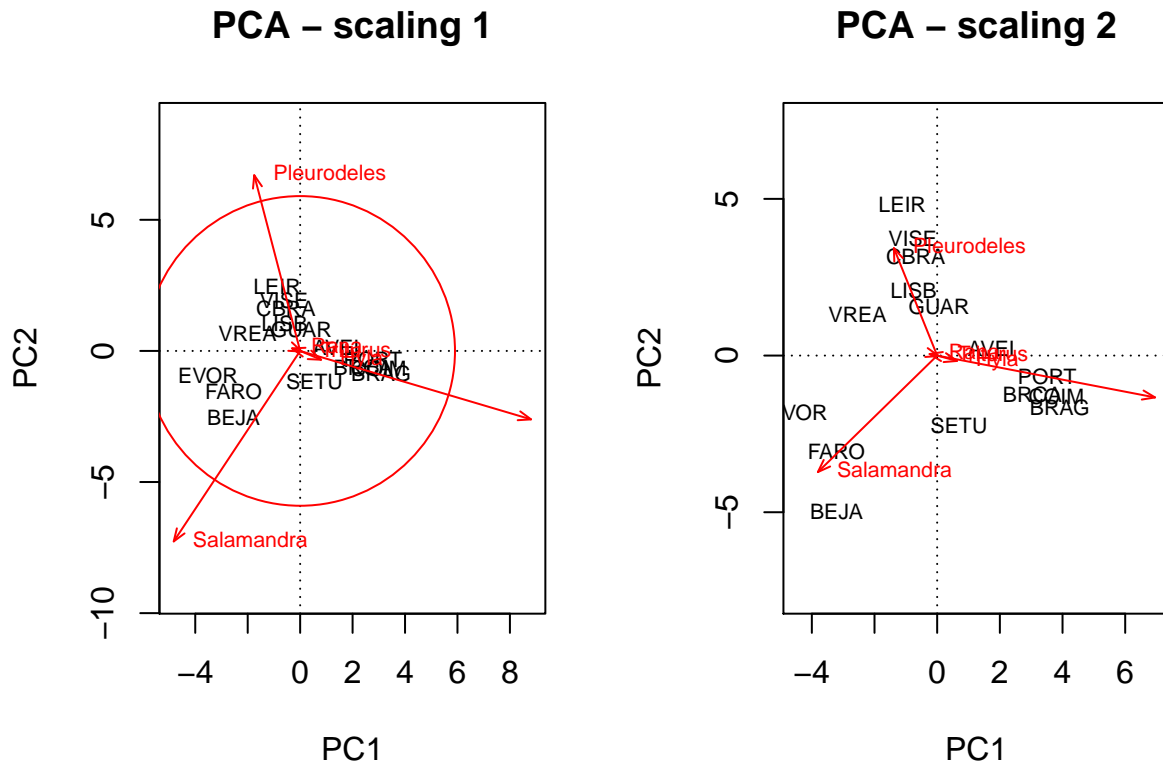
```

##              Inertia Proportion
## Total          781.5          1
## Unconstrained  781.5          1
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5
## Eigenvalue    487.1988 204.6408 83.6880 4.393193 1.268618
## Proportion Explained 0.6234 0.2619 0.1071 0.005622 0.001623
## Cumulative Proportion 0.6234 0.8853 0.9924 0.998010 0.999633
##              PC6
## Eigenvalue    0.2867497
## Proportion Explained 0.0003669
## Cumulative Proportion 1.0000000
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
## * General scaling constant of scores: 10.2273
##
##
## Species scores
##
##              PC1      PC2      PC3      PC4      PC5      PC6
## Rana          -0.24279 0.04961 -0.18772 0.287830 0.378957 -0.0188070
## Triturus      0.04546 -0.01272 -0.08416 -0.183302 0.080555 0.1862655
## Salamandra   -3.80663 -3.71720 -1.74307 0.008117 -0.018248 -0.0002021
## Hyla          0.63170 -0.17200 0.06577 0.685012 -0.135978 0.0572872
## Bufo          6.95334 -1.33604 -1.45330 -0.042191 0.009485 -0.0070138
## Pleurodeles   -1.38213 3.42864 -2.45037 0.021880 -0.028093 0.0008849
##
##
## Site scores (weighted sums of species scores)
##
##              PC1      PC2      PC3      PC4      PC5      PC6
## EVOR -4.45029 -1.8051 -1.43972 4.85876 0.9649 2.72958
## BEJA -3.21133 -4.9933 -0.32520 -1.31139 1.1447 -3.43677
## FARO -3.21533 -3.0454 -1.13190 -0.96158 -4.1708 -1.02418
## SETU 0.69504 -2.2351 -3.18569 -4.40403 2.0650 4.21760
## LISB -0.75616 2.0662 -0.13806 1.75837 6.1401 -2.60877
## LEIR -1.11264 4.8176 -3.16612 -0.60822 0.5738 -3.62364
## CBRA -0.68563 3.1714 -0.84000 -2.36512 0.7564 5.49587
## VISE -0.77938 3.7395 -2.05341 -0.55756 -5.4812 -0.60741
## GUAR 0.05864 1.5710 1.63605 -0.75442 -0.1043 -1.81220
## AVEI 1.72812 0.2374 0.66594 4.24651 0.6059 1.86519
## COIM 3.80061 -1.2962 0.08837 -0.01056 -1.3292 -2.35883
## PORT 3.51186 -0.6626 -0.62978 3.15437 -0.1736 0.04528
## BRAG 3.91119 -1.6366 0.52393 -3.54655 1.7519 -1.06526
## BRCA 3.04358 -1.2473 1.66098 2.29127 -2.6055 1.35698
## VREA -2.53829 1.3185 8.33462 -1.78986 -0.1382 0.82656

```

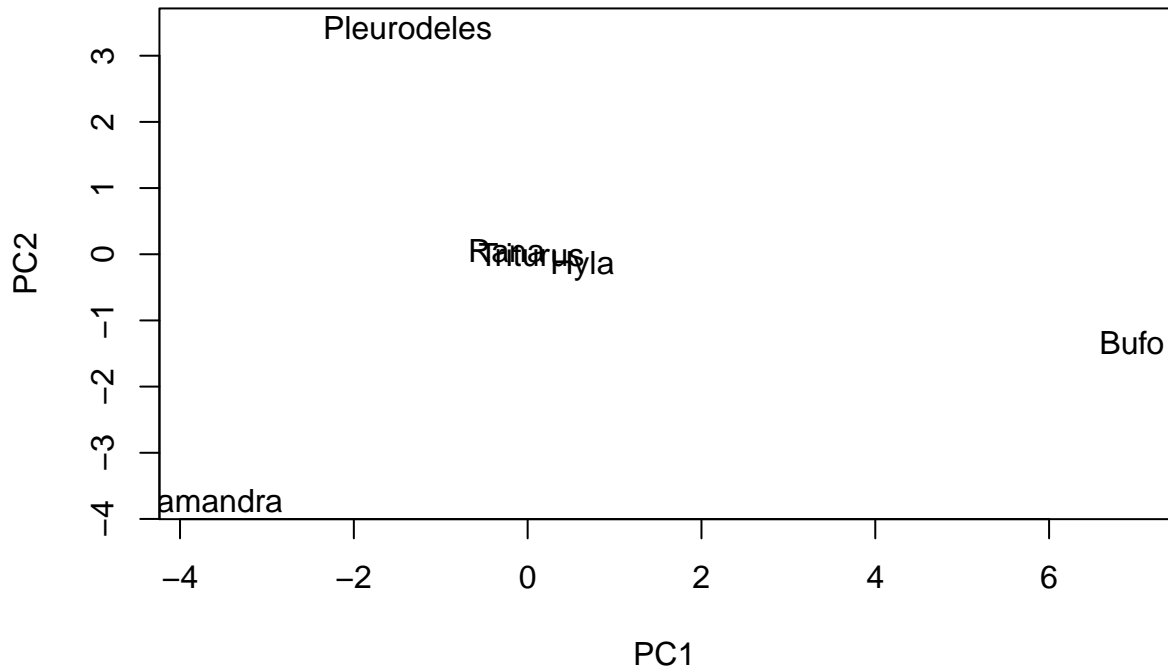
We can use a function in `brocardfunctions.R` to make a clean plot of this object

```
cleanplot.pca(acpbyrda1)
```



and this representation is slightly more useful as it shows us that the variables most important (those which the arrows go outside the “control” circle) are *Pleurodeles*, *Salamandra* and *Bufo* (question: what do these 3 taxa share? They were the most abundant! This is not good.). This last genera name is not printed in the plot, for reasons which are unclear (a possible bug in the function). Nonetheless, the `rda` function creates an object that can help us, because it provides the coordinates of the variables in the PC space. We could therefore plot these in PC space if we so wanted and check that the variable that was not printed above was *Bufo*

```
plot(summary(acpbyrda1)$species[,1:2],type="n")  
text(summary(acpbyrda1)$species[,1:2],names(anfibios[,-1]))
```



Note that the results of the `princomp` and of `rda` might look different, for example, the sites scores. The sites scores in the first component are

```
#princomp
acp1$scores[,1]
```

```
##      EVOR      BEJA      FARO      SETU      LISB      LEIR
## 35.9372068 25.9322705 25.9645802 -5.6126292  6.1061455  8.9848858
##      CBRA      VISE      GUAR      AVEI      COIM      PORT
##  5.5366438  6.2936806 -0.4735304 -13.9550085 -30.6908939 -28.3591494
##      BRAG      BRCA      VREA
## -31.5838299 -24.5776694  20.4972974
```

```
#rda
summary(acpbyrda1)$sites[,1]
```

```
##      EVOR      BEJA      FARO      SETU      LISB      LEIR
## -4.45028832 -3.21132583 -3.21532690  0.69504061 -0.75615526 -1.11264441
##      CBRA      VISE      GUAR      AVEI      COIM      PORT
## -0.68563095 -0.77937869  0.05863969  1.72812015  3.80061052  3.51185867
##      BRAG      BRCA      VREA
##  3.91118736  3.04357864 -2.53828528
```

but what we should keep in mind is that signs in scores are not possible to interpret, and in fact magnitudes are not easily interpreted either. It is only relative to each other that they can be interpreted. And relative to each other, the correlations between the sites scores across methods is perfect - as we can see below all 1 or -1 - so the results just look different at first, but are exactly the same!

```
cor(acp1$scores[,1],summary(acpbyrda1)$sites[,1])
```

```
## [1] -1
```

```
cor(acp1$scores[,2],summary(acpbyrda1)$sites[,2])
```

```
## [1] 1
```

```
cor(acp1$scores[,3],summary(acpbyrda1)$sites[,3])
```

```
## [1] -1
```

```
cor(acp1$scores[,4],summary(acpbyrda1)$sites[,4])
```

```
## [1] 1
```

```
cor(acp1$scores[,5],summary(acpbyrda1)$sites[,5])
```

```
## [1] 1
```

```
cor(acp1$scores[,6],summary(acpbyrda1)$sites[,6])
```

```
## [1] -1
```

Finally, we should note that it might be more sensible, given the data were abundances, to scale the data prior to the analysis. If we do that, say with the `rda` function, this is what we get

```
acpbyrda2<-rda(scale(anfibios[,-1]))
acpbyrda2
```

```
## Call: rda(X = scale(anfibios[, -1]))
```

```
##
```

```
##           Inertia Rank
```

```
## Total           6
```

```
## Unconstrained   6   6
```

```
## Inertia is variance
```

```
##
```

```
## Eigenvalues for unconstrained axes:
```

```
##   PC1   PC2   PC3   PC4   PC5   PC6
```

```
## 2.3287 1.3351 1.1423 0.8014 0.2688 0.1238
```

The summary of this object gives us direct access to something that was hidden above, namely the coordinates of the covariates in the PC space

```
summary(acpbyrda2)
```

```
##
```

```
## Call:
```

```
## rda(X = scale(anfibios[, -1]))
```

```
##
```

```
## Partitioning of variance:
```

```
##           Inertia Proportion
```

```
## Total           6           1
```

```
## Unconstrained   6           1
```

```
##
```

```
## Eigenvalues, and their contribution to the variance
```

```
##
```

```
## Importance of components:
```

```
##           PC1   PC2   PC3   PC4   PC5   PC6
```

```
## Eigenvalue   2.3287 1.3351 1.1423 0.8014 0.26875 0.12379
```

```

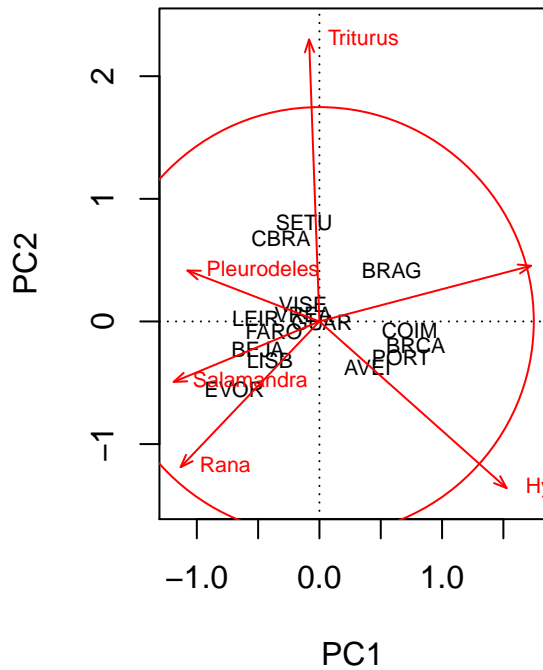
## Proportion Explained  0.3881 0.2225 0.1904 0.1336 0.04479 0.02063
## Cumulative Proportion 0.3881 0.6106 0.8010 0.9346 0.97937 1.00000
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
## * General scaling constant of scores:  3.0274
##
##
## Species scores
##
##           PC1      PC2      PC3      PC4      PC5      PC6
## Rana      -0.70518 -0.5605  0.35220 -0.71704  0.26273  0.09418
## Triturus  -0.05266  1.0845 -0.09913 -0.54777  0.10755 -0.16477
## Salamandra -0.74030 -0.2325 -0.81806 -0.34136 -0.37349 -0.01298
## Hyla      0.95051 -0.6409  0.15236 -0.31626 -0.09727 -0.28401
## Bufo      1.07507  0.2143  0.12125 -0.43920 -0.22054  0.26382
## Pleurodeles -0.67071  0.1960  0.95069  0.02149 -0.36379 -0.05137
##
##
## Site scores (weighted sums of species scores)
##
##           PC1      PC2      PC3      PC4      PC5      PC6
## EVOR -1.11472 -1.16822 -0.2980 -0.9783 -0.05659 -1.1615
## BEJA -0.80364 -0.48583 -1.5122 -0.1381 -0.18387  1.0847
## FARO -0.59632 -0.16407 -1.1444  0.5814 -1.21466 -0.1498
## SETU -0.19929  1.71236 -0.6032 -1.2792 -0.12597 -0.1142
## LISB -0.64957 -0.67438  0.9097 -0.6987  1.13161  1.0599
## LEIR -0.83183  0.04880  1.4780  0.3464 -0.54002  0.9706
## CBRA -0.50210  1.43906  0.6390 -0.3091  0.37601 -1.0968
## VISE -0.21402  0.29417  0.8576  1.1450 -1.39824 -0.4538
## GUAR  0.03154 -0.01817  0.2578  0.6898  0.30029  0.3949
## AVEI  0.63058 -0.78602  0.4346 -0.5638  0.33398 -0.7674
## COIM  1.18052 -0.15962 -0.1215  0.1496 -0.42285  0.6488
## PORT  1.07042 -0.62408  0.3282 -0.5670 -0.21849 -0.1281
## BRAG  0.94564  0.88104 -0.3997 -0.2379  0.34510  1.0314
## BRCA  1.26232 -0.41002 -0.2108  0.1312 -0.16822 -0.7446
## VREA -0.20951  0.11499 -0.6152  1.7287  1.84190 -0.5739

```

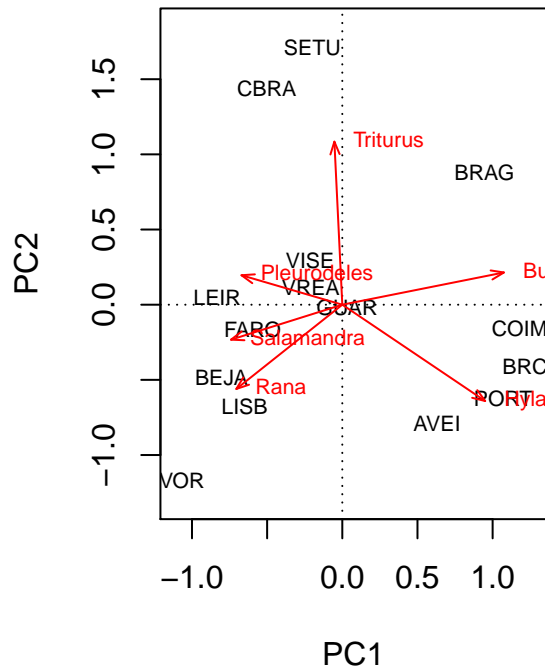
Make a clean plot of the output

```
cleanplot.pca(acpbyrda2)
```

PCA – scaling 1



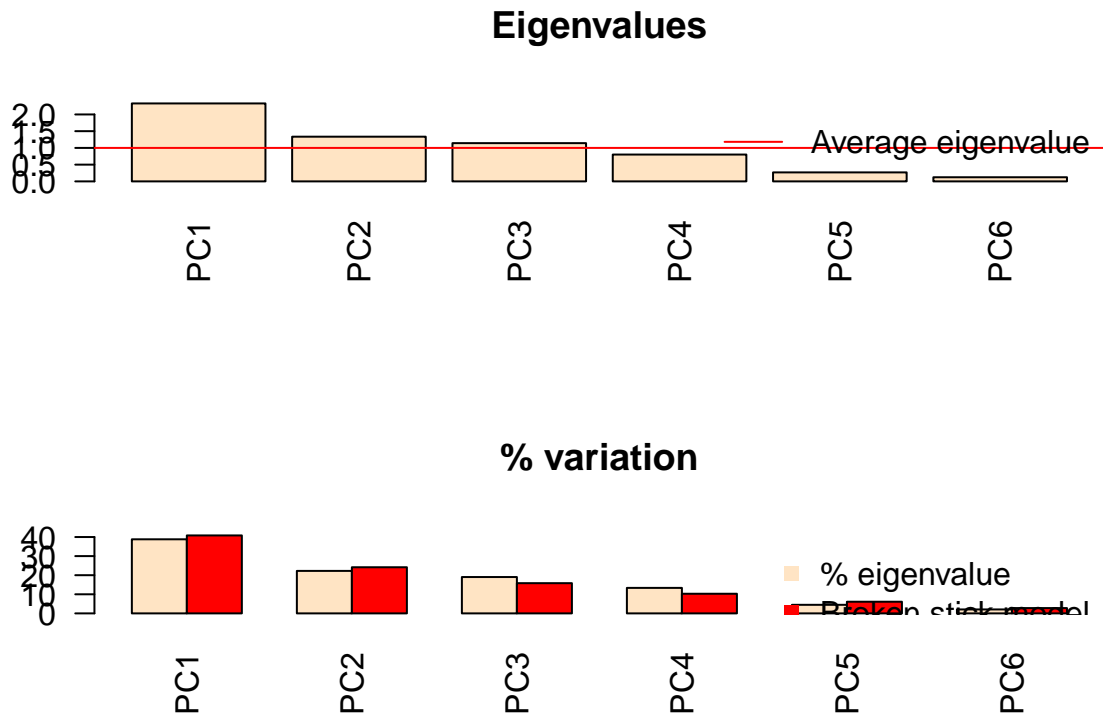
PCA – scaling 2



The results are now considerably different, and this means that prior results were (inadvertently) dominated by the most abundant species. Typically, that is not desired, and therefore this might be a better analysis to interpret.

In this case we might need more than 2 PCs to interpret. Can you explain why this is a better analysis and yet it does not manage to represent as well the original information in just 2 PC's?

```
evplot(acpbyrda2$CA$eig)
```



Exercício 2

Efectue uma análise de componentes principais aos dados das características de diferentes serras portuguesas (DataTP10serras.csv), com base na matriz de variância/covariância e na matriz de correlação. Comente os principais resultados obtidos em cada uma das análises e refira em que circunstâncias cada uma destas abordagens seria mais adequada.

We read in the data

```
serras<-read.csv("DataTP10serras.csv", sep=";")
```

as before, check it is all OK

```
str(serras)
```

```
## 'data.frame': 15 obs. of 6 variables:
## $ Serra : Factor w/ 15 levels "Acor","Alvao",...: 5 8 13 1 15 11 12 3 2 7 ...
## $ Altitude : int 1993 1548 1486 1418 1416 1415 1382 1362 1283 1227 ...
## $ Perc.Floresta: int 20 21 18 17 16 15 25 32 45 54 ...
## $ Dispon.agua : int 5 6 7 8 9 5 4 10 12 14 ...
## $ Matos : int 75 48 76 98 56 47 56 35 34 54 ...
## $ Incendios : int 3 4 6 3 6 5 4 12 15 14 ...
```

We have 15 observations for 6 variables, which correspond respectively to serras and their characteristics. The first column does not represent a characteristic, but it contains the labels of the serras.

As above, we begin with exploring the data

```
serras[,1]
```

```
## [1] Estrela Geres Montesinho Acor Soajo Marao  
## [7] Montemuro Amarela Alvao Gardunha Lousa Bornes  
## [13] Padrela Falperra Gralheira  
## 15 Levels: Acor Alvao Amarela Bornes Estrela Falperra Gardunha ... Soajo
```

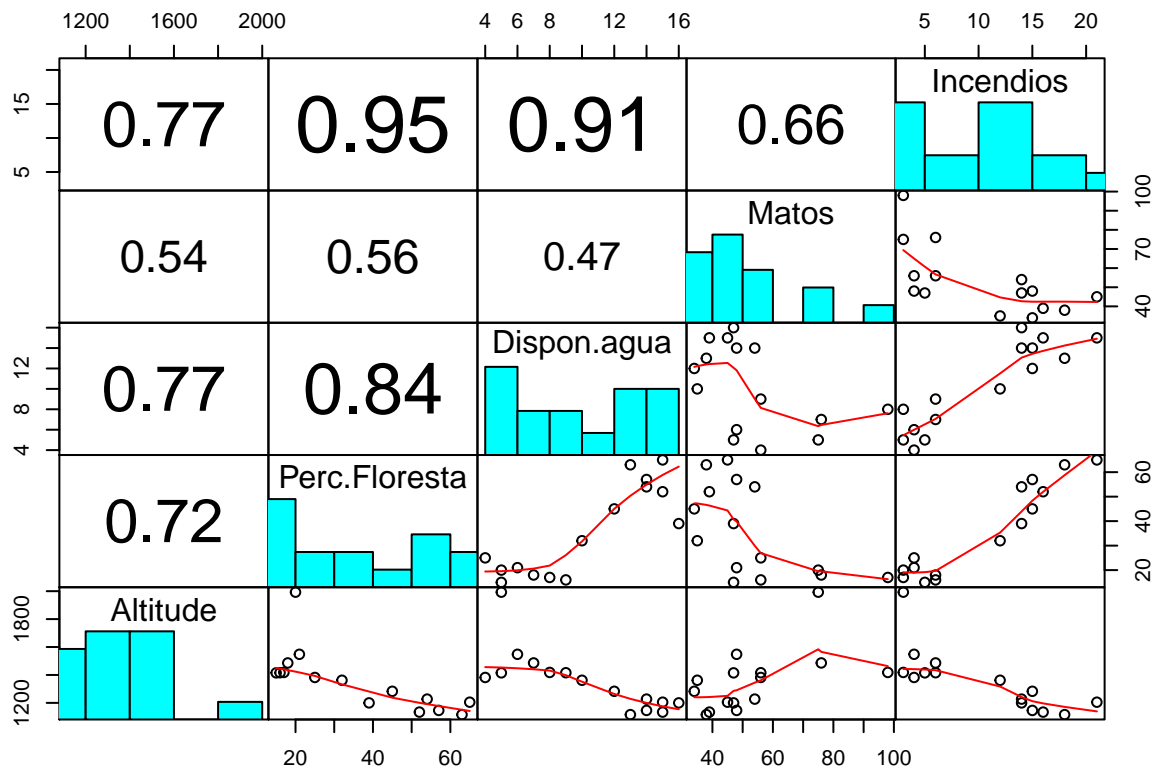
The names of the serras are

```
names(serras)[2:6]
```

```
## [1] "Altitude" "Perc.Floresta" "Dispon.agua" "Matos"  
## [5] "Incendios"
```

What are the correlations between descriptors?

```
pairs(serras[,2:6], lower.panel = panel.smooth, upper.panel = panel.corP,gap=0, rowlattice=FALSE,diag.panel = panel.hist)
```



Some of the correlations are quite high, as high as 0.95, so the PCA technique might be efficient.

As before, we add the right labels to the row.names

```
row.names(serras)=serras[,1]
```

Now, we implement the PCA itself, first on the covariance matrix

```
acp.serras<-princomp(serras[,-1])  
#this is the same as  
#acp.serras<-princomp(serras[,-1],cor=FALSE)  
acp.serras
```



```
## Call:
## princomp(x = serras[, -1])
##
## Standard deviations:
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## 215.332725 15.664152 11.295886 2.235848 1.010098
##
## 5 variables and 15 observations.
```

We can look at the summary of the resulting object

```
summary(acp.serras)
```

```
## Importance of components:
##
##           Comp.1      Comp.2      Comp.3      Comp.4
## Standard deviation 215.3327250 15.664151562 11.295885593 2.2358479945
## Proportion of Variance 0.9918929 0.005248781 0.002729514 0.0001069373
## Cumulative Proportion 0.9918929 0.997141723 0.999871237 0.9999781741
##
##           Comp.5
## Standard deviation 1.010098e+00
## Proportion of Variance 2.182589e-05
## Cumulative Proportion 1.000000e+00
```

We can also look at some of the components of the object created

```
acp.serras$loadings
```

```
##
## Loadings:
##           Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## Altitude      0.997
## Perc.Floresta 0.555 0.779 -0.263 0.116
## Dispon.agua   0.130 0.792 0.593
## Matos         -0.805 0.589
## Incendios      0.186 0.171 0.551 -0.795
##
##           Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings 1.0 1.0 1.0 1.0 1.0
## Proportion Var 0.2 0.2 0.2 0.2 0.2
## Cumulative Var 0.2 0.4 0.6 0.8 1.0
```

Apparently, the main driver of the variability would be **Altitude**. But in fact, the truth is that the altitude is measured with a higher order of magnitude (1000's) than the other variables (10's or 100's). And that means this is not really desirable.

We can look at the scores of the analysis

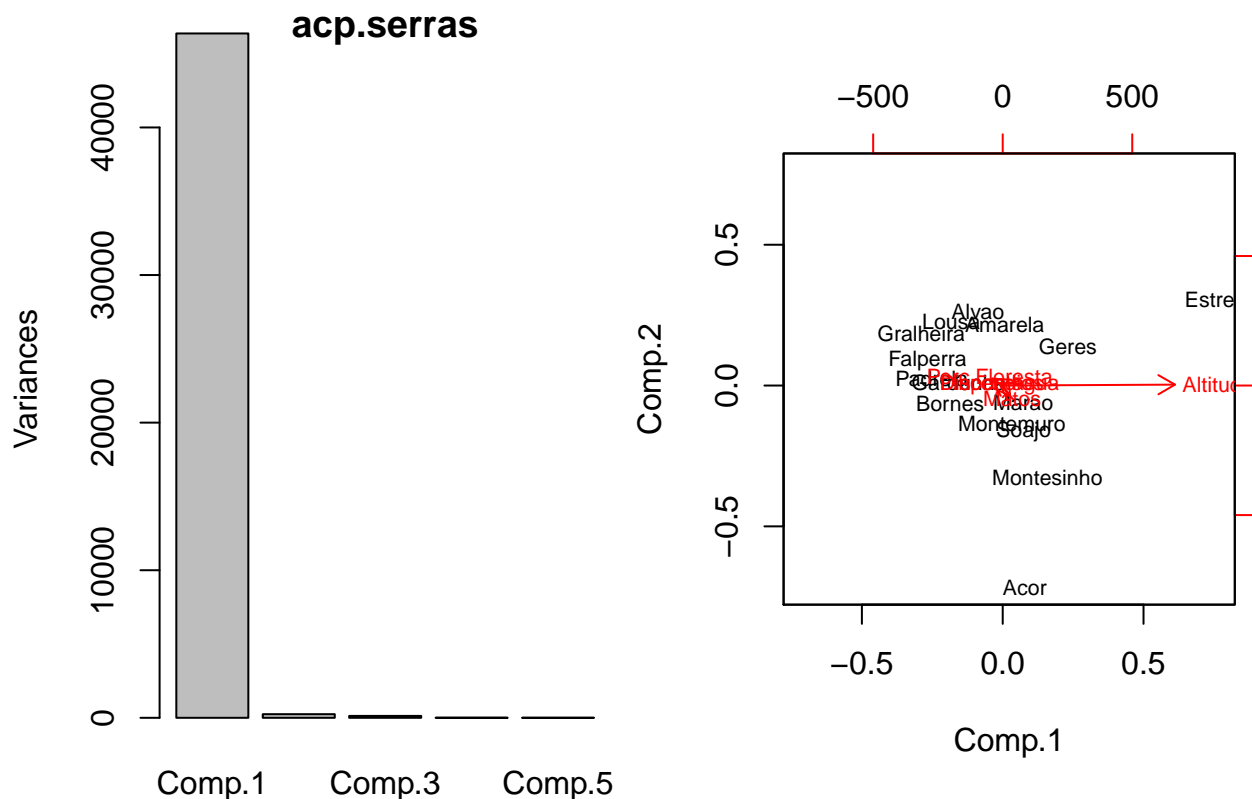
```
acp.serras$scores
```

```
##           Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
## Estrela 637.772883 18.577037 15.8304839 0.6159163 0.401905823
## Geres 192.883507 8.458511 -11.0379411 -1.6814589 1.333573829
## Montesinho 132.409285 -19.853092 1.9160680 0.9369192 -1.609093482
## Acor 65.680113 -43.598861 11.8774196 0.1551196 -0.022792148
## Soajo 61.852382 -9.868268 -13.0582419 2.3130933 0.366487802
## Marao 60.605771 -3.704939 -19.8566483 -1.2639212 -0.835996979
## Montemuro 27.529712 -8.078386 -7.9617387 -5.3448661 0.003224666
## Amarela 6.004445 13.130049 -13.2810904 1.5715964 -0.852477944
```

```
## Alvao      -73.671336  16.039050  -5.1146234  0.8413669 -0.563993596
## Gardunha  -129.183895  0.768920  12.2499784 -0.6179373  1.312161511
## Lousa     -152.331717  13.870548  16.2403720  0.8775077 -1.907480327
## Bornes    -155.527864  -3.763524  -4.0235285  4.6340288  1.106989884
## Padrela   -210.395347  1.503908  9.0297213  -1.4836588  1.111213174
## Falperra  -222.481752  5.347614  -0.1898093  0.9765820  0.807563740
## Gralheira -241.146186  11.171433  7.3795786  -2.5302879 -0.651285955
```

and the plot it

```
par(mfrow=c(1,2),mar=c(4,4,0.5,0.5))
plot(acp.serras)
biplot(acp.serras,cex=0.7)
```



and then on the correlation matrix

```
acp.serras2<-princomp(serras[, -1], cor=TRUE)
```

which we note in passing is equivalent with an analysis on the covariance matrix of the scaled data (scaled data have, by definition, variances of 1, and hence the covariances and correlations are the same)

```
#which is the same as
acp.serras2<-princomp(scale(serras[, -1]))
```

Then we can look at the object resulting

```
acp.serras2
```

```
## Call:
## princomp(x = serras[, -1], cor = TRUE)
```

```
##
## Standard deviations:
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## 1.9774983 0.7738141 0.5699272 0.3709750 0.1681455
##
## 5 variables and 15 observations.
```

```
summary(acp.serras2)
```

```
## Importance of components:
##               Comp.1   Comp.2   Comp.3   Comp.4
## Standard deviation    1.9774983 0.7738141 0.5699272 0.37097499
## Proportion of Variance 0.7820999 0.1197576 0.0649634 0.02752449
## Cumulative Proportion 0.7820999 0.9018575 0.9668209 0.99434542
##               Comp.5
## Standard deviation    0.168145492
## Proportion of Variance 0.005654581
## Cumulative Proportion 1.000000000
```

```
acp.serras2$loadings
```

```
##
## Loadings:
##               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## Altitude      0.436          0.867 0.218
## Perc.Floresta -0.472 -0.171 0.401 -0.617 0.454
## Dispon.agua   -0.463 -0.367          0.739 0.313
## Matos         0.359 -0.907          -0.157 -0.153
## Incendios     -0.494          0.281          -0.820
##
##               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings    1.0    1.0    1.0    1.0    1.0
## Proportion Var 0.2    0.2    0.2    0.2    0.2
## Cumulative Var 0.2    0.4    0.6    0.8    1.0
```

the summary of the object

```
summary(acp.serras2)
```

```
## Importance of components:
##               Comp.1   Comp.2   Comp.3   Comp.4
## Standard deviation    1.9774983 0.7738141 0.5699272 0.37097499
## Proportion of Variance 0.7820999 0.1197576 0.0649634 0.02752449
## Cumulative Proportion 0.7820999 0.9018575 0.9668209 0.99434542
##               Comp.5
## Standard deviation    0.168145492
## Proportion of Variance 0.005654581
## Cumulative Proportion 1.000000000
```

and some of its components, as before, the loadings

```
acp.serras2$loadings
```

```
##
## Loadings:
##               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## Altitude      0.436          0.867 0.218
## Perc.Floresta -0.472 -0.171 0.401 -0.617 0.454
```

```
## Dispon.agua  -0.463 -0.367          0.739  0.313
## Matos        0.359 -0.907          -0.157 -0.153
## Incendios    -0.494           0.281          -0.820
##
##              Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings   1.0    1.0    1.0    1.0    1.0
## Proportion Var 0.2    0.2    0.2    0.2    0.2
## Cumulative Var 0.2    0.4    0.6    0.8    1.0
```

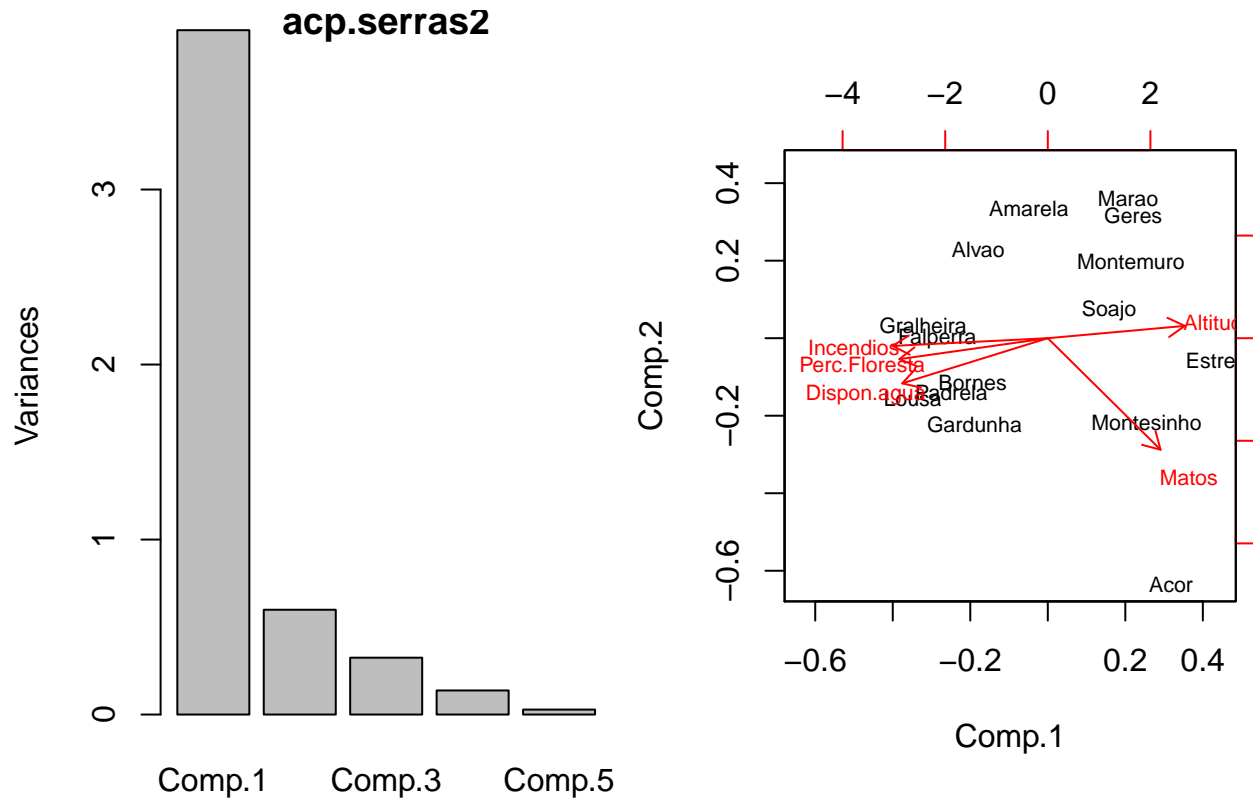
and the scores

```
acp.serras2$scores
```

```
##              Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
## Estrela      3.3843786 -0.1701555286  1.72232398  0.07514264  0.07994403
## Geres        1.6882639  0.9496873031  0.06625469  0.01739621  0.24404943
## Montesinho   1.9511841 -0.6533639671 -0.19031455 -0.02727515 -0.28748056
## Acor         2.4374279 -1.9059101727 -0.65104457 -0.07036841 -0.02629920
## Soajo        1.2117037  0.2167791652 -0.43885015  0.52022847 -0.01106022
## Marao        1.5864603  1.0781288710 -0.57641648 -0.08816721 -0.12645700
## Montemuro    1.6429874  0.5896419318 -0.57088183 -0.72723636  0.10318056
## Amarela      -0.3735097  1.0037137166  0.04488904  0.26602348 -0.17258861
## Alvao        -1.3746378  0.6836330476  0.20072489  0.09967202 -0.10143626
## Gardunha     -1.4495942 -0.6636659398  0.13075576 -0.08309614  0.23338604
## Lousa        -2.6682495 -0.4638399285  0.65613646 -0.24793196 -0.29699170
## Bornes       -1.4855502 -0.3421301766 -0.25978905  0.83528254  0.06845791
## Padreira     -1.9024131 -0.4209728434 -0.07063063 -0.21719952  0.21754209
## Falperra     -2.1819737 -0.0009751006 -0.14613111  0.20411953  0.11004382
## Gralheira    -2.4664777  0.0994296218  0.08297355 -0.55659014 -0.03429033
```

Now, all variables have a reasonable influence on the first component. We can plot the resulting object

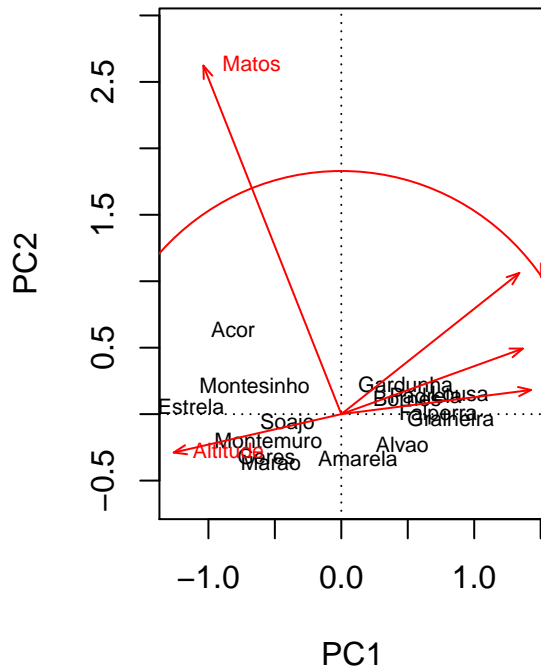
```
par(mfrow=c(1,2),mar=c(4,4,0.5,0.5))
plot(acp.serras2)
biplot(acp.serras2,cex=0.7)
```



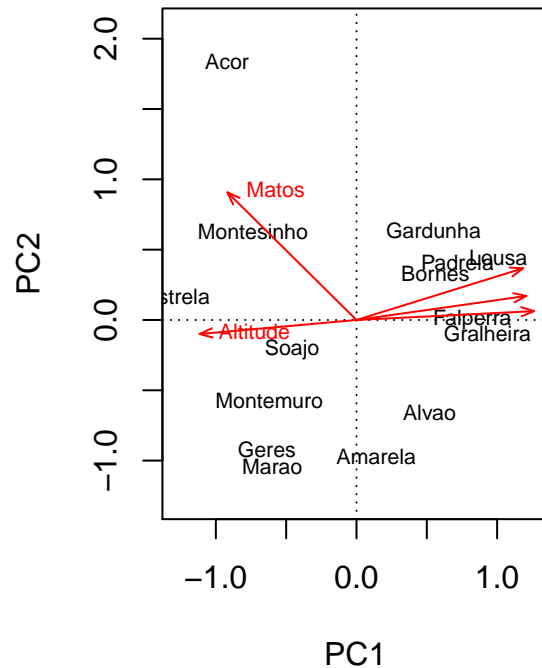
In this case, given all the variables are measured in different units, it would only make sense to interpret the analysis over the correlation matrix, since the analysis based on the covariances suffers a strong impact from the measuring units of each variable (in other words, the analysis would return different results if say altitude was measured in km and in meters!). Leveraging on the nice clean biplot function (`cleanplot.pca`), for the scaled analysis we get the following biplots

```
rda.serras2s<-rda(scale(serras[, -1]))
cleanplot.pca(rda.serras2s)
```

PCA – scaling 1



PCA – scaling 2



showing that matos seems to be the main driver of the variability in Serras. Notice that the conclusions would be extremely different if interpreting the analysis based on the covariance and the correlation matrix.

Just to finish, note how the different principal components are indeed independent (i.e. uncorrelated, and while not shown here for both, this obviously happens for both covariance and correlation based analysis)

```
round(cor(acp.serras$scores), 4)
```

##	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
## Comp. 1	1	0	0	0	0
## Comp. 2	0	1	0	0	0
## Comp. 3	0	0	1	0	0
## Comp. 4	0	0	0	1	0
## Comp. 5	0	0	0	0	1

This happens by construction, i.e. a PCA is the mathematical procedure that obtains the linear combinations of the original covariates which simultaneously respect two conditions:

- the variance is largest for the 1st component, then the 2nd, then the 3rd, and so on
- and all the components are orthogonal (i.e. all independent of each other).

Exercício 3

Recolheu dados relativos a densidades de várias espécies de anfíbios e répteis em três tipos florestais (pinhal, eucaliptal e sobre) (DataTP10herpetofauna.csv). Realize uma análise de correspondências e comente os resultados obtidos. Efectue uma análise de componentes principais aos mesmos dados compare os resultados com os obtidos na análise de correspondências.

We read the data in

```
herpeto <- read.csv2("DataTP10herpetofauna.csv")
```

and we look at the data

```
str(herpeto)
```

```
## 'data.frame':  22 obs. of  11 variables:
## $ Habitat      : Factor w/ 22 levels "Eucalip1","Eucalip2",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Chalcides    : num  0.05 0.02 0.01 0.01 0.02 0.08 0.03 0.02 0 0.01 ...
## $ Triturus     : num  0.22 0.18 0.18 0.15 0.12 0.03 0 0.02 0.03 0 ...
## $ Bufo         : num  0 0.04 0.05 0.01 0.04 0 0 0 0.01 0.01 ...
## $ Rana         : num  0.2 0.13 0.11 0.1 0.02 0.02 0.02 0.02 0.01 0 ...
## $ Salamandra  : num  0.03 0 0 0.01 0 0.02 0 0 0.01 0.01 ...
## $ Lacerta     : num  0.06 0.15 0.13 0.17 0.37 0.21 0.19 0.19 0.2 0.23 ...
## $ Podarcis    : num  0.16 0 0.09 0.1 0.12 0.12 0.34 0.2 0.26 0.11 ...
## $ Coluber     : num  0.03 0.05 0.08 0.04 0.08 0.16 0.11 0.31 0.25 0.21 ...
## $ Malpolon    : num  0.06 0.04 0.06 0.05 0.02 0.05 0.1 0.03 0.08 0.1 ...
## $ Psamodromus: num  0.16 0.36 0.32 0.33 0.16 0.32 0.17 0.2 0.14 0.27 ...
```

PCA

Given the way the material was presented in the lectures, we first implement the PCA.

The labels of the locations are

```
herpeto[,1]
```

```
## [1] Eucalip1 Eucalip2 Eucalip3 Eucalip4 Eucalip5 Eucalip6 Eucalip7
## [8] Eucalip8 Pinhal1 Pinhal2 Pinhal3 Pinhal4 Pinhal5 Pinhal6
## [15] Pinhal7 Pinhal8 Sobro1 Sobro2 Sobro3 Sobro4 Sobro5
## [22] Sobro6
## 22 Levels: Eucalip1 Eucalip2 Eucalip3 Eucalip4 Eucalip5 ... Sobro6
```

and this seems to imply there are 3 types of habitats, Eucalipal, Pinhal e Sobro. We can create a separate variable to hold the type of habitat - this could be useful for plotting say points with different colors per habitat

```
herpeto.hab=substr(herpeto[,1],1,4)
herpeto.hab
```

```
## [1] "Euca" "Euca" "Euca" "Euca" "Euca" "Euca" "Euca" "Euca" "Pinh" "Pinh"
## [11] "Pinh" "Pinh" "Pinh" "Pinh" "Pinh" "Pinh" "Sobr" "Sobr" "Sobr" "Sobr"
## [21] "Sobr" "Sobr"
```

The names of the genera are

```
names(herpeto)[-1]
```

```
## [1] "Chalcides" "Triturus" "Bufo" "Rana" "Salamandra"
## [6] "Lacerta" "Podarcis" "Coluber" "Malpolon" "Psamodromus"
```

We can also see and if some locations have much more animals than others

```
nbyloc=data.frame(loc=herpeto[,1],n=rowSums(herpeto[,-1]))
nbyloc
```

```
##      loc      n
```

```
## 1 Eucalip1 0.97
## 2 Eucalip2 0.97
## 3 Eucalip3 1.03
## 4 Eucalip4 0.97
## 5 Eucalip5 0.95
## 6 Eucalip6 1.01
## 7 Eucalip7 0.96
## 8 Eucalip8 0.99
## 9 Pinhal1 0.99
## 10 Pinhal2 0.95
## 11 Pinhal3 0.98
## 12 Pinhal4 0.96
## 13 Pinhal5 0.96
## 14 Pinhal6 0.92
## 15 Pinhal7 0.92
## 16 Pinhal8 0.95
## 17 Sobro1 0.93
## 18 Sobro2 0.92
## 19 Sobro3 0.90
## 20 Sobro4 0.91
## 21 Sobro5 0.91
## 22 Sobro6 1.00
```

and in fact these numbers are extremely suspicious to me, as they are way too close to 1 to being true... Not sure what happened here, but I suspect some scaling must have happened at some level.

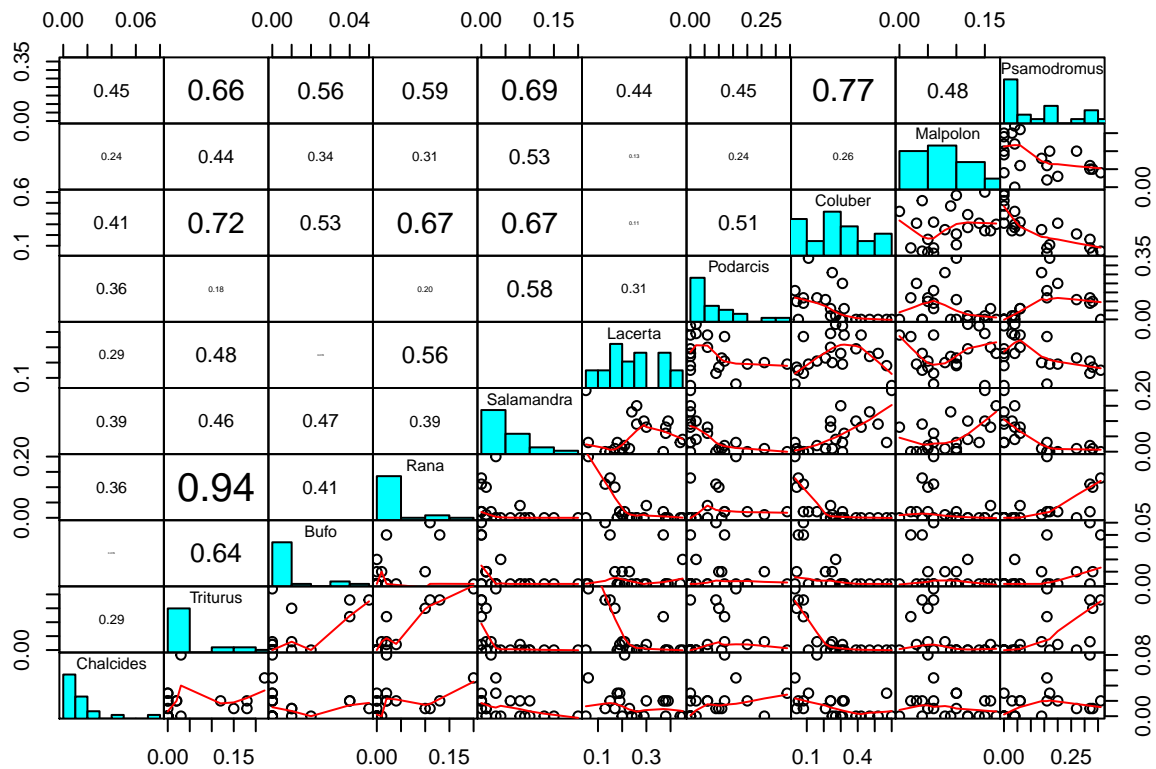
We can also evaluate what are the most abundant taxa, and show them in increasing order

```
sort(colSums(herpeto[,-1]))
```

```
##      Bufo   Chalcides      Rana   Triturus Salamandra   Podarcis
##      0.18     0.34     0.67     0.95     1.17     1.66
## Malpolon Psamodromus   Lacerta   Coluber
##      1.89     2.70     5.52     5.97
```

Given that a PCA is a dimension reduction technique, we can see if the genera are highly correlated or not. The analysis will be more efficient if there are high correlations to explore (e.g. two variables with perfect correlation can be summarized in a single component!)

```
pairs(herpeto[,-1], lower.panel = panel.smooth, upper.panel = panel.corP,
      gap=0, rowlattice=FALSE,diag.panel = panel.hist)
```

As it turns out, there are a couple of large correlations, but many low correlations too, so it's unclear how the PCA might perform.

For the PCA plots it might be useful to define `row.names` in the object `herpeto`, as these will be used by default in the plotting of biplots, hence making the plots easier to interpret.

```
row.names(herpeto)=herpeto[,1]
```

Now, we implement the PCA itself

```
acph1<-princomp(herpeto[, -1])
acph1
```

```
## Call:
## princomp(x = herpeto[, -1])
##
## Standard deviations:
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6
## 0.222253341 0.116857752 0.084798534 0.065174317 0.049421709 0.027323770
##   Comp.7   Comp.8   Comp.9   Comp.10
## 0.019466369 0.012803054 0.007711547 0.005515373
##
## 10 variables and 22 observations.
```

and we visualize the summary

```
sumacp1=summary(acph1)
sumacp1
```

```
## Importance of components:
##                Comp.1   Comp.2   Comp.3   Comp.4
## Standard deviation 0.2222533 0.1168578 0.08479853 0.06517432
## Proportion of Variance 0.6307611 0.1743747 0.09182164 0.05424020
## Cumulative Proportion 0.6307611 0.8051358 0.89695745 0.95119765
##                Comp.5   Comp.6   Comp.7   Comp.8
## Standard deviation 0.04942171 0.027323770 0.019466369 0.012803054
## Proportion of Variance 0.03118917 0.009533439 0.004838807 0.002093127
## Cumulative Proportion 0.98238683 0.991920266 0.996759072 0.998852199
##                Comp.9   Comp.10
## Standard deviation 0.0077115470 0.0055153726
## Proportion of Variance 0.0007593664 0.0003884347
## Cumulative Proportion 0.9996115653 1.0000000000
```

We can also explore some of the components of the PCA output, namely the loadings (i.e. the correlations between the PC's and the original variables)

```
print(acph1$loadings,digits = 2, cutoff = 0.2)
```

```
##
## Loadings:
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## Chalcides                0.75  0.58  0.31
## Triturus  0.25        -0.38        -0.45        -0.34  0.33  0.35
## Bufo                -0.43  0.33
## Rana                -0.25 -0.25 -0.28        0.31 -0.64  0.36
## Salamandra                -0.27  0.33 -0.81                0.29
## Lacerta                -0.87                0.36
## Podarcis  0.25                0.87                0.31
## Coluber  -0.71  0.41                0.36                0.35
## Malpolon                -0.40  0.64  0.53                0.33
## Psamodromus 0.51                0.68  0.36                0.30
##          Comp.10
## Chalcides
## Triturus  -0.45
## Bufo       0.81
## Rana       0.31
## Salamandra
## Lacerta
## Podarcis
## Coluber
## Malpolon
## Psamodromus
##
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
## SS loadings  1.0   1.0   1.0   1.0   1.0   1.0   1.0   1.0
## Proportion Var  0.1   0.1   0.1   0.1   0.1   0.1   0.1   0.1
## Cumulative Var  0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8
##          Comp.9 Comp.10
## SS loadings  1.0   1.0
## Proportion Var  0.1   0.1
## Cumulative Var  0.9   1.0
```

the scores (the coordinates of the locations in the principal components space) - note one could use these to manually plot the results of a PCA

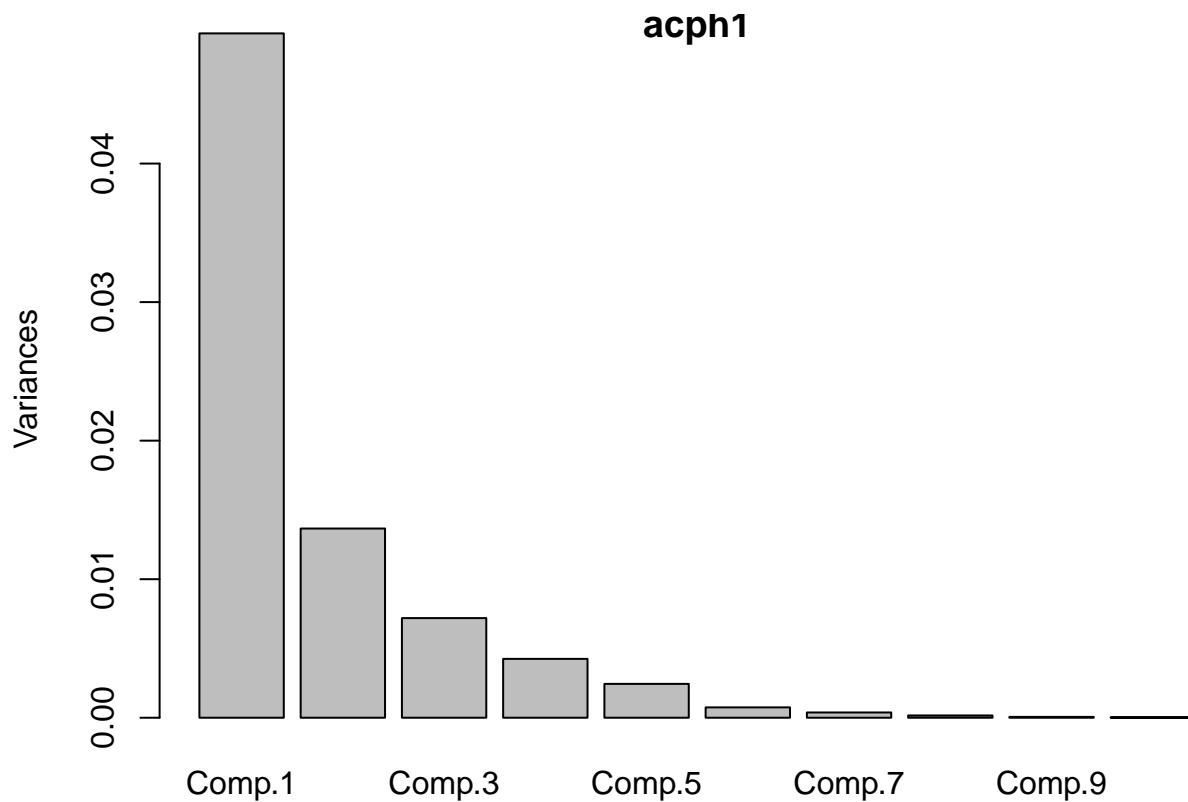
acph1\$scores

##	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
## Eucalip1	0.33036012	0.1190746337	-0.0536501941	-0.173598501	-0.090526132
## Eucalip2	0.34584060	0.0517842147	-0.1721385189	0.058846301	0.009082210
## Eucalip3	0.32476098	0.0779560507	-0.0841678556	0.019993831	-0.005314638
## Eucalip4	0.34143688	0.0258898542	-0.0655258483	0.025832367	0.013281281
## Eucalip5	0.17800655	-0.1704727058	-0.0023542392	0.002300882	-0.068136960
## Eucalip6	0.20429760	0.0187811424	0.0293535323	0.106670305	0.061945854
## Eucalip7	0.21143613	-0.0009541545	0.2399409213	-0.066183602	0.019471640
## Eucalip8	0.06037370	0.0814274760	0.1267152444	0.075954562	-0.032648361
## Pinhal1	0.07974812	0.0415702696	0.1736002759	-0.017164678	-0.022469305
## Pinhal2	0.12084691	0.0036613757	0.0441539446	0.089594180	0.082344779
## Pinhal3	-0.03808968	-0.0671599482	-0.0092753287	-0.076925610	0.044006065
## Pinhal4	-0.10939491	-0.2021268920	-0.0189188778	-0.001490877	0.002516037
## Pinhal5	-0.11011557	-0.1101631123	0.0179693053	0.027027647	-0.028576384
## Pinhal6	-0.16187489	-0.1856887962	-0.0266235105	-0.018859311	-0.003899037
## Pinhal7	-0.20549424	-0.0722419492	-0.0272707540	0.078608606	-0.073942208
## Pinhal8	-0.09507087	-0.1578061781	-0.0281760532	-0.060250364	0.050353478
## Sobro1	-0.25176387	0.0667264114	-0.0264528531	-0.007841541	-0.006905388
## Sobro2	-0.12468269	-0.0152867492	-0.0482293582	-0.075773264	0.091258354
## Sobro3	-0.20139685	-0.0139815965	-0.0293047807	-0.030779206	0.009969235
## Sobro4	-0.29326921	0.1501800341	0.0001913402	0.040198823	-0.043555560
## Sobro5	-0.27961522	0.0673598930	-0.0156762094	0.045156030	-0.062687574
## Sobro6	-0.32633958	0.2914707263	-0.0241601824	-0.041316579	0.054432614
##	Comp.6	Comp.7	Comp.8	Comp.9	
## Eucalip1	0.0003849084	0.0340361565	-0.0057888571	-0.0002534738	
## Eucalip2	0.0119767752	-0.0030303004	-0.0061535217	-0.0058286363	
## Eucalip3	0.0168104838	-0.0340484229	0.0052779494	0.0111880085	
## Eucalip4	-0.0099672205	-0.0091393928	-0.0152754544	-0.0041853267	
## Eucalip5	-0.0261431923	-0.0267621951	0.0395666453	-0.0044229935	
## Eucalip6	-0.0198959109	0.0530281565	0.0213244051	0.0070892394	
## Eucalip7	-0.0027819539	0.0040587949	-0.0018174578	-0.0099902259	
## Eucalip8	0.0012641350	0.0018305978	-0.0121284771	0.0067218530	
## Pinhal1	0.0064195679	-0.0319496192	-0.0026199560	0.0049651031	
## Pinhal2	0.0190603464	-0.0059887697	-0.0101173348	-0.0098615361	
## Pinhal3	0.0159393830	0.0123083101	-0.0081942702	0.0105938377	
## Pinhal4	0.0241696557	-0.0097383806	-0.0060933523	0.0095122430	
## Pinhal5	-0.0201931467	0.0136854862	-0.0031822026	0.0080259948	
## Pinhal6	0.0516448822	-0.0013358857	-0.0074970018	0.0016667964	
## Pinhal7	-0.0479037860	0.0075133298	-0.0171264087	-0.0014230612	
## Pinhal8	-0.0100291625	0.0110885427	0.0065037999	0.0035186253	
## Sobro1	-0.0258301614	-0.0041949108	-0.0013418082	-0.0033168842	
## Sobro2	-0.0248019521	-0.0072512355	0.0027701275	-0.0114785828	
## Sobro3	0.0016150264	-0.0030127952	0.0002010651	-0.0116901771	
## Sobro4	0.0773930484	0.0168840405	0.0195055872	-0.0065897458	
## Sobro5	-0.0142070233	0.0005152552	-0.0058044624	-0.0063703662	
## Sobro6	-0.0249247029	-0.0184967623	0.0079909856	0.0121293084	
##	Comp.10				
## Eucalip1	-0.0008319545				
## Eucalip2	0.0092156639				
## Eucalip3	0.0061611483				
## Eucalip4	-0.0159141142				
## Eucalip5	-0.0025343131				

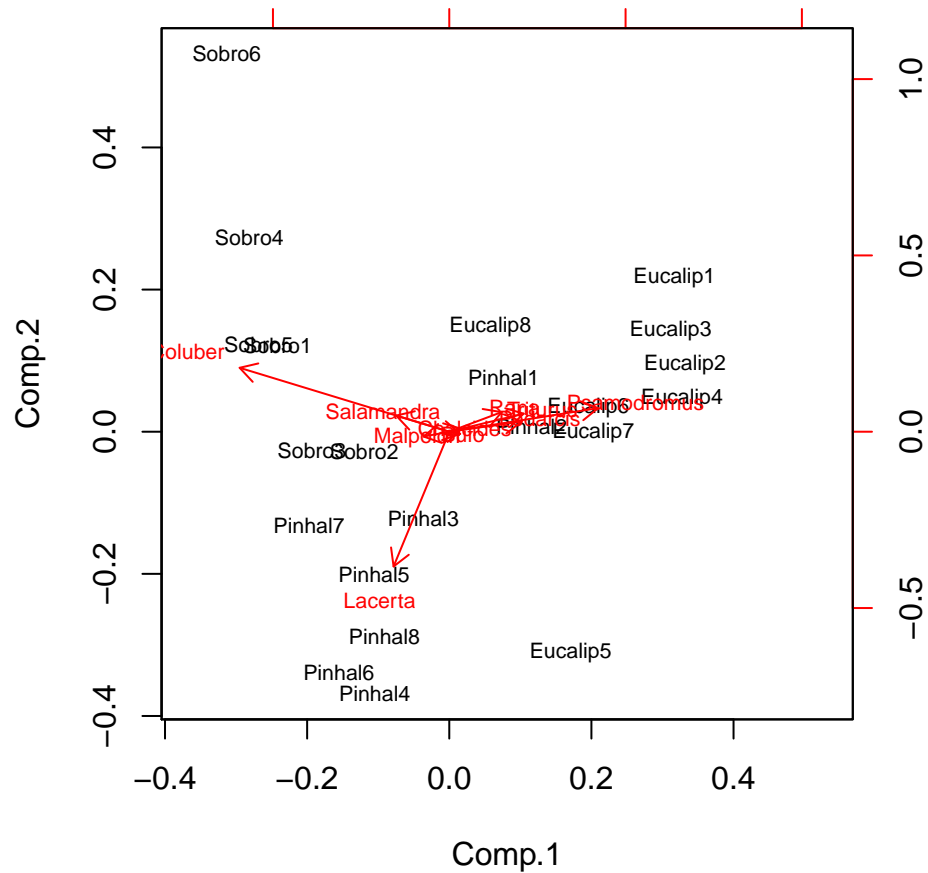
```
## Eucalip6 -0.0015222878
## Eucalip7  0.0090531127
## Eucalip8 -0.0030570981
## Pinhal1  -0.0044147046
## Pinhal2  0.0005058102
## Pinhal3  0.0023686708
## Pinhal4  0.0083093312
## Pinhal5 -0.0002895905
## Pinhal6 -0.0085563999
## Pinhal7  0.0028535561
## Pinhal8 -0.0019751886
## Sobro1   0.0015824262
## Sobro2  -0.0007066841
## Sobro3  -0.0002147896
## Sobro4  -0.0010813382
## Sobro5   0.0025031725
## Sobro6  -0.0014544285
```

We can plot the objects produced by the PCA

```
par(mfrow=c(1,1),mar=c(4,4,0.5,0.5))
plot(acph1)
```



```
biplot(acph1,cex=0.7)
```

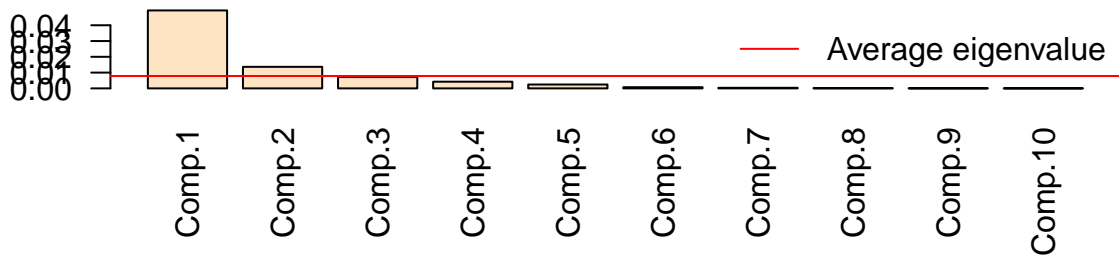


as for previous exercises, and not surprisingly given the unscaled analysis (i.e. using the covariance matrix), **Lacerta** and **Coluber**, the most abundant species, are the ones driving most of the variability on the first PC.

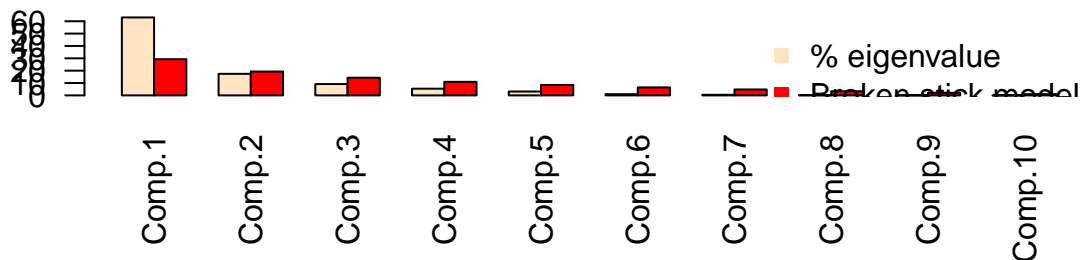
Finally, we can use one of the bespoke functions in file `brocardfunctions.R` to plot the variances associated with each PC and two criteria to choose the number of PCs to interpret

```
evplot(acph1$sdev^2)
```

Eigenvalues



% variation



Both criteria seem to agree that the more sensible would be to interpret just the first two PCs.

Note that we could have implemented the same analysis using function `rda` in library `vegan`, which produces slightly different output, but the same results

```
acphbyrda1<-rda(herpeto[, -1])
acphbyrda1
```

```
## Call: rda(X = herpeto[, -1])
##
##              Inertia Rank
## Total          0.08204
## Unconstrained 0.08204   10
## Inertia is variance
##
## Eigenvalues for unconstrained axes:
##  PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8    PC9
## 0.05175 0.01431 0.00753 0.00445 0.00256 0.00078 0.00040 0.00017 0.00006
##   PC10
## 0.00003
```

The summary of this object gives us direct access to something that was hidden above, namely the coordinates of the covariates in the PC space

```
summary(acphbyrda1)
```

```
##
## Call:
## rda(X = herpeto[, -1])
```

```

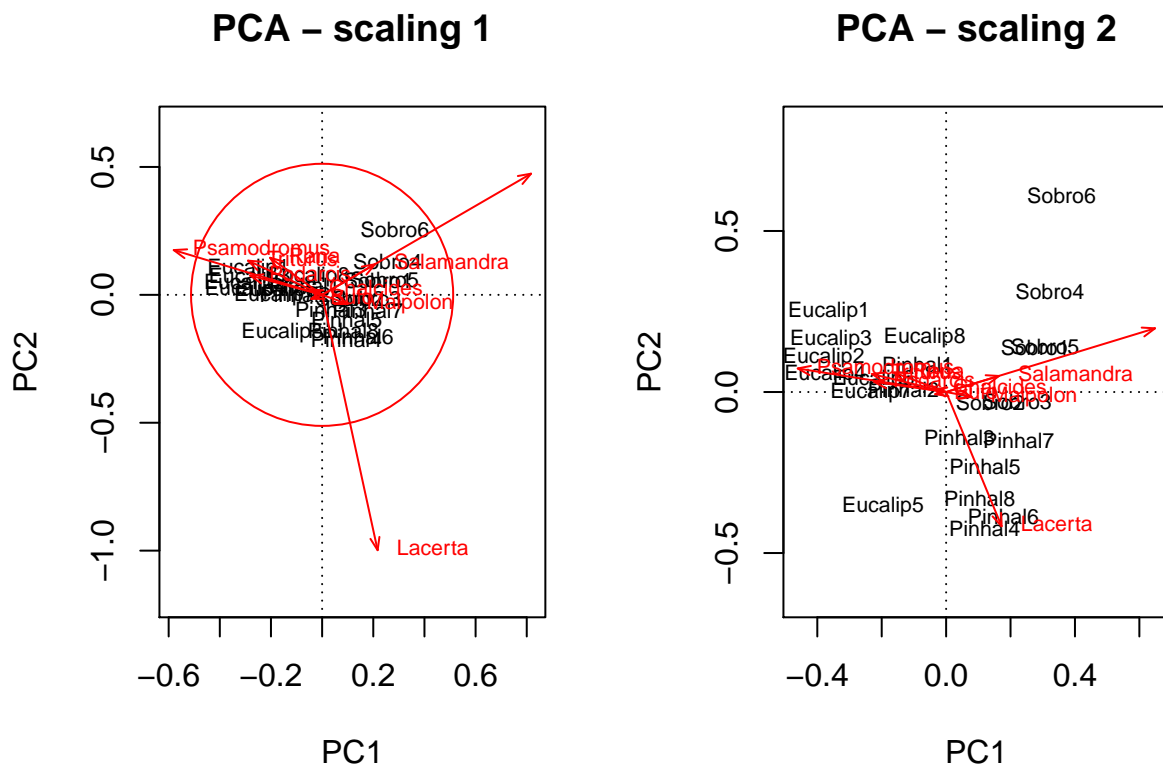
##
## Partitioning of variance:
##           Inertia Proportion
## Total           0.08204           1
## Unconstrained 0.08204           1
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##           PC1      PC2      PC3      PC4      PC5      PC6
## Eigenvalue      0.05175 0.01431 0.007533 0.00445 0.002559 0.0007821
## Proportion Explained 0.63076 0.17437 0.091822 0.05424 0.031189 0.0095334
## Cumulative Proportion 0.63076 0.80514 0.896957 0.95120 0.982387 0.9919203
##           PC7      PC8      PC9      PC10
## Eigenvalue      0.000397 0.0001717 0.0000623 3.187e-05
## Proportion Explained 0.004839 0.0020931 0.0007594 3.884e-04
## Cumulative Proportion 0.996759 0.9988522 0.9996116 1.000e+00
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
## * General scaling constant of scores: 1.145681
##
##
## Species scores
##
##           PC1      PC2      PC3      PC4      PC5      PC6
## Chalcides      -0.03776 0.008854 0.0098076 0.002762 -0.0003448 0.002596
## Triturus       -0.23020 0.055851 -0.1308192 -0.049064 0.0904079 0.003909
## Bufo           -0.03385 -0.006658 -0.0220420 0.012559 0.0069510 0.006214
## Rana           -0.16181 0.060539 -0.0870802 -0.066131 0.0566041 0.015235
## Salamandra     0.16429 0.049853 -0.0560908 -0.072646 -0.0673632 -0.090889
## Lacerta        0.17234 -0.416903 -0.0031983 0.053345 0.0222889 -0.005150
## Podarcis       -0.22365 0.031512 0.3006997 -0.047010 0.0342179 -0.012577
## Coluber         0.64874 0.197801 0.0315186 0.097320 0.0332375 0.014769
## Malpolon       0.07940 -0.014905 0.0002864 -0.106045 -0.1296873 0.059417
## Psamodromus   -0.46076 0.073078 -0.0259752 0.182206 -0.0733715 -0.005168
##
##
## Site scores (weighted sums of species scores)
##
##           PC1      PC2      PC3      PC4      PC5      PC6
## Eucalip1      -0.36307 0.248894 -0.1545380 -0.650611 0.44741 0.003441
## Eucalip2      -0.38008 0.108241 -0.4958405 0.220544 -0.04489 0.107066
## Eucalip3      -0.35692 0.162946 -0.2424433 0.074933 0.02627 0.150277
## Eucalip4      -0.37524 0.054116 -0.1887455 0.096814 -0.06564 -0.089102
## Eucalip5      -0.19563 -0.356328 -0.0067813 0.008623 0.33676 -0.233706
## Eucalip6      -0.22453 0.039257 0.0845521 0.399778 -0.30616 -0.177859
## Eucalip7      -0.23237 -0.001994 0.6911435 -0.248043 -0.09624 -0.024869
## Eucalip8      -0.06635 0.170202 0.3649999 0.284662 0.16136 0.011301
## Pinhal1       -0.08764 0.086892 0.5000510 -0.064330 0.11105 0.057388
## Pinhal2       -0.13281 0.007653 0.1271843 0.335781 -0.40698 0.170389
## Pinhal3        0.04186 -0.140380 -0.0267173 -0.288301 -0.21749 0.142490
## Pinhal4        0.12023 -0.422492 -0.0544953 -0.005588 -0.01244 0.216064

```

```
## Pinhal5  0.12102 -0.230267  0.0517601  0.101294  0.14123 -0.180516
## Pinhal6  0.17790 -0.388133 -0.0766883 -0.070681  0.01927  0.461678
## Pinhal7  0.22584 -0.151003 -0.0785527  0.294609  0.36545 -0.428234
## Pinhal8  0.10448 -0.329852 -0.0811604 -0.225806 -0.24887 -0.089655
## Sobro1   0.27669  0.139474 -0.0761967 -0.029388  0.03413 -0.230908
## Sobro2   0.13703 -0.031953 -0.1389234 -0.283983 -0.45103 -0.221716
## Sobro3   0.22134 -0.029225 -0.0844116 -0.115354 -0.04927  0.014437
## Sobro4   0.32231  0.313911  0.0005512  0.150657  0.21527  0.691853
## Sobro5   0.30730  0.140798 -0.0451549  0.169236  0.30982 -0.127003
## Sobro6   0.35865  0.609242 -0.0695928 -0.154846 -0.26903 -0.222814
```

We can make a clean plot of this object

```
cleanplot.pca(acphbyrda1)
```



The corresponding analysis based on the correlation matrix is perhaps more sensible

```
sacphbyrda1<-rda(scale(herpeto[,-1]))
sacphbyrda1
```

```
## Call: rda(X = scale(herpeto[, -1]))
##
##              Inertia Rank
## Total              10
## Unconstrained    10   10
## Inertia is variance
##
## Eigenvalues for unconstrained axes:
```



```
## PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 PC10
## 4.937 1.387 1.315 0.800 0.692 0.421 0.224 0.198 0.021 0.006
```

The summary of this object gives us direct access to something that was hidden above, namely the coordinates of the covariates in the PC space

```
summary(sacphbyrda1)
```

```
##
## Call:
## rda(X = scale(herpeto[, -1]))
##
## Partitioning of variance:
##           Inertia Proportion
## Total           10           1
## Unconstrained    10           1
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##           PC1    PC2    PC3    PC4    PC5    PC6    PC7
## Eigenvalue      4.9365 1.3874 1.3152 0.79974 0.69153 0.42064 0.22441
## Proportion Explained 0.4937 0.1387 0.1315 0.07997 0.06915 0.04206 0.02244
## Cumulative Proportion 0.4937 0.6324 0.7639 0.84389 0.91304 0.95510 0.97754
##           PC8    PC9    PC10
## Eigenvalue      0.19814 0.020642 0.0057846
## Proportion Explained 0.01981 0.002064 0.0005785
## Cumulative Proportion 0.99736 0.999422 1.0000000
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
## * General scaling constant of scores: 3.806754
##
##
## Species scores
##
##           PC1    PC2    PC3    PC4    PC5    PC6
## Chalcides  -0.6148 -0.66989 0.28721 -0.020219 0.65342 -0.2655
## Triturus   -1.0417 0.47923 0.24165 0.014366 0.10989 0.2037
## Bufo       -0.7323 0.69684 -0.42256 0.062473 -0.18893 -0.3362
## Rana       -0.9712 0.32579 0.47437 0.091677 0.16752 0.3383
## Salamandra 0.9168 0.30657 0.56875 0.018431 0.03683 -0.0234
## Lacerta    0.5502 0.04452 -0.92847 0.232372 0.42470 0.1977
## Podarcis   -0.6334 -0.80017 -0.08305 0.144766 -0.53992 0.2077
## Coluber    1.0287 0.02834 0.21519 -0.506741 -0.11054 -0.1168
## Malpolon   0.6740 0.02272 0.34552 0.902175 -0.07640 -0.1687
## Psamodromus -1.0743 -0.05322 -0.05785 0.002388 -0.08965 -0.3810
##
##
## Site scores (weighted sums of species scores)
##
##           PC1    PC2    PC3    PC4    PC5    PC6
## Eucalip1 -1.4296 -0.17228 1.859965 0.18866 0.996881 2.079766
## Eucalip2 -1.4284 1.46077 0.081130 -0.17814 0.333498 -0.786716
```

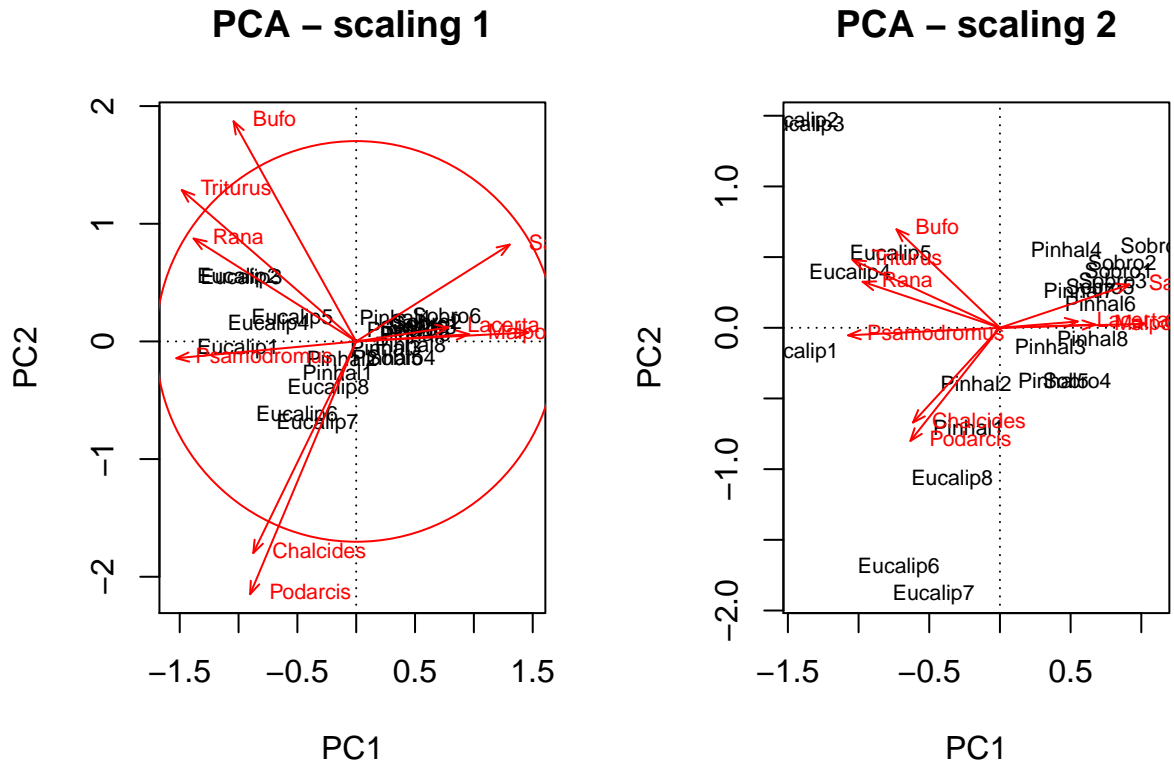
```

## Eucalip3 -1.3878  1.43203 -0.083985  0.16053 -0.787980 -0.814436
## Eucalip4 -1.0597  0.38443  0.208352  0.02952 -0.238384  0.581143
## Eucalip5 -0.7728  0.51739 -1.530892 -0.20115  0.198427 -0.063827
## Eucalip6 -0.7151 -1.69389  0.267106 -0.41076  1.747837 -1.880656
## Eucalip7 -0.4685 -1.88450 -0.100348  0.79593 -1.224439  0.344216
## Eucalip8 -0.3374 -1.07102 -0.212962 -0.99714 -0.774187  0.180882
## Pinhal1 -0.2211 -0.70609 -0.396428  0.08884 -1.867745  0.523299
## Pinhal2 -0.1692 -0.39642 -0.461122  0.33845 -0.749594 -1.098069
## Pinhal3  0.3565 -0.13499  0.364301  1.30791  0.482541  0.177416
## Pinhal4  0.4695  0.55664 -1.255823  0.82222  0.237081  0.002148
## Pinhal5  0.3795 -0.37481 -0.620047 -0.41217  0.718626  0.394531
## Pinhal6  0.7116  0.17490 -0.857397  0.92070  0.523985  0.668720
## Pinhal7  0.5600  0.26689 -0.791365 -1.67646  0.452221  0.934844
## Pinhal8  0.6562 -0.07702 -0.144873  1.19855  0.940616 -0.018770
## Sobro1  0.8416  0.40817  0.448639 -0.60596 -0.180739  0.134376
## Sobro2  0.8648  0.46235  0.643822  1.14046 -0.136818 -0.281960
## Sobro3  0.7979  0.33708  0.134159  0.15440 -0.009382  0.217083
## Sobro4  0.5456 -0.37114  0.508297 -0.86208  0.443939 -0.800235
## Sobro5  0.7097  0.29638 -0.001117 -1.50433 -0.086574  0.445470
## Sobro6  1.0966  0.58513  1.940588 -0.29798 -1.019809 -0.939224

```

We can make a clean plot of this object

```
cleanplot.pca(sacphbyrda1)
```



and we can see clearly that the separation on the first PC is between sites from Eucaliptal versus Pinhal versus Sobro, without any clear separation regarding habitat on the second component.

Making a biplot by hand has the advantage of letting us use bespoke colors as well as fine tune the limits of the plot. Check where are the species and locations scores

```
names(summary(sacphbyrda1))
```

```
## [1] "species"      "sites"        "call"         "tot.chi"      "unconst.chi"
## [6] "cont"         "scaling"      "digits"       "inertia"      "method"
```

and then use those to make the biplot

```
#a couple of useful objects to make the code more readable
```

```
myacp=summary(sacphbyrda1)
```

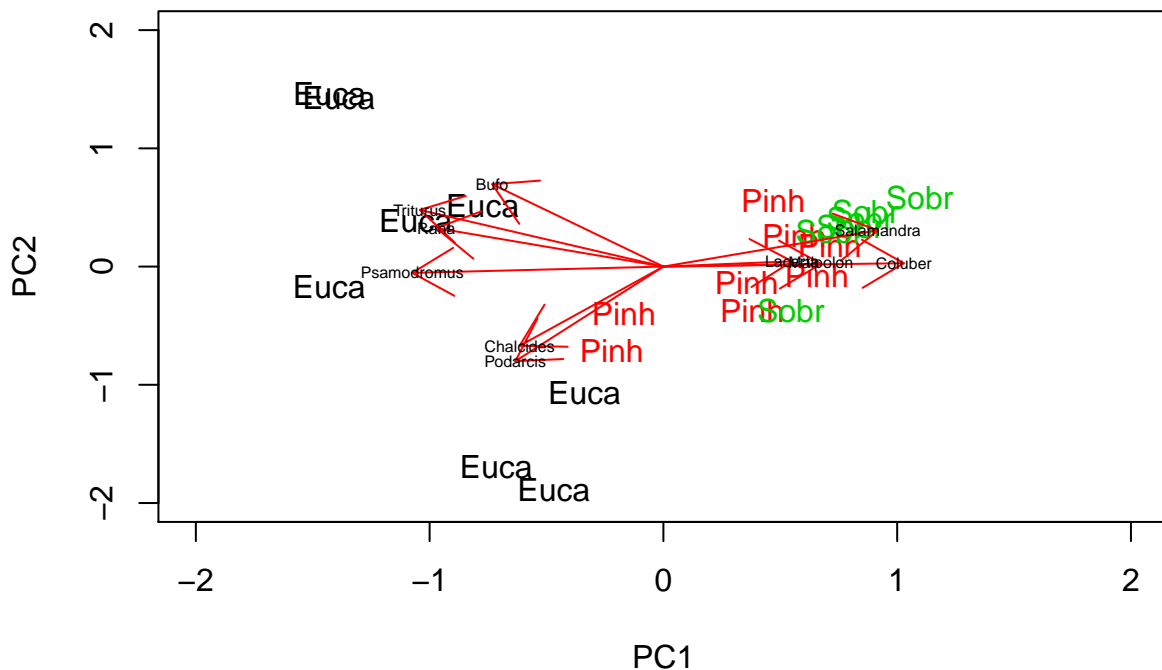
```
plot(myacp$sites[,1:2],type="n",ylim=c(-2,2),xlim=c(-2,2))
```

```
text(myacp$sites[,1:2],labels=substr(herpeto[,1],1,4),col=as.numeric(as.factor(substr(herpeto[,1],1,4))))
```

```
#species
```

```
arrows(x0=0,y0=0,x1=myacp$species[,1],y1=myacp$species[,2],col=2)
```

```
text(myacp$species[,1:2],labels=names(herpeto[-1]),cex=0.5)
```



CA

We now implement the CA

```
ca.herp<-cca(herpeto[, -1])
```

```
summary(ca.herp)
```

```
##
```

```
## Call:
```

```
## cca(X = herpeto[, -1])
```

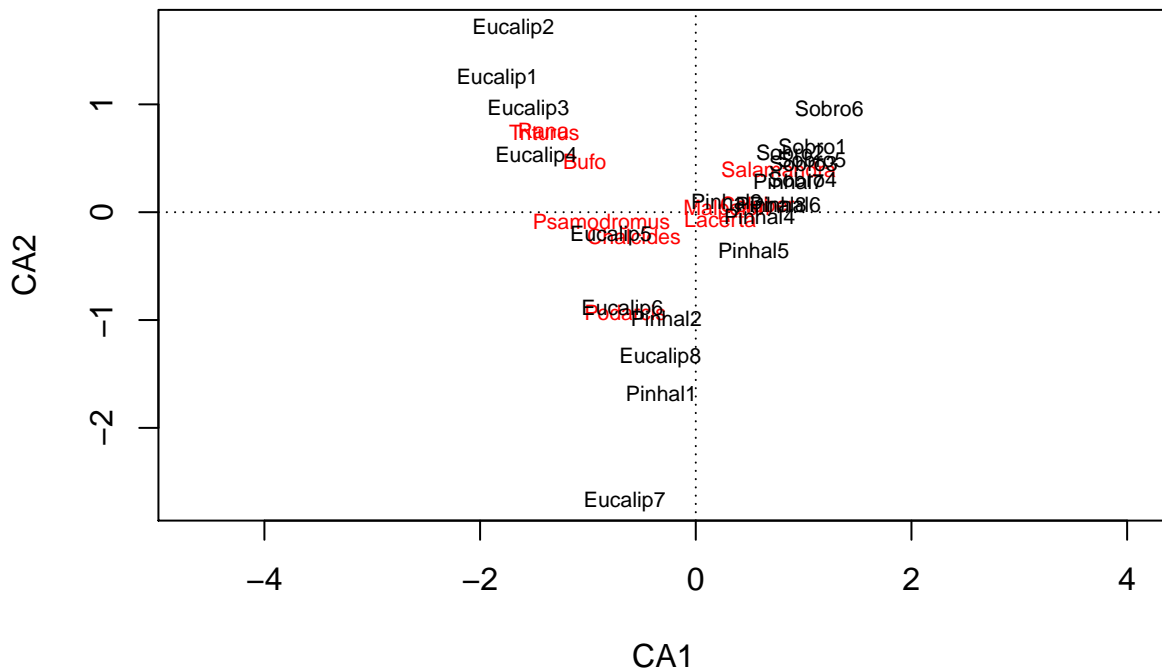
```

##
## Partitioning of scaled Chi-square:
##           Inertia Proportion
## Total           0.7634           1
## Unconstrained  0.7634           1
##
## Eigenvalues, and their contribution to the scaled Chi-square
##
## Importance of components:
##           CA1      CA2      CA3      CA4      CA5      CA6
## Eigenvalue      0.4453 0.1260 0.06964 0.04409 0.03147 0.02391
## Proportion Explained 0.5833 0.1651 0.09122 0.05776 0.04123 0.03132
## Cumulative Proportion 0.5833 0.7484 0.83963 0.89739 0.93862 0.96993
##           CA7      CA8      CA9
## Eigenvalue      0.01227 0.008908 0.001775
## Proportion Explained 0.01607 0.011669 0.002326
## Cumulative Proportion 0.98601 0.997674 1.000000
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
##
##
## Species scores
##
##           CA1      CA2      CA3      CA4      CA5      CA6
## Chalcides      -0.5736 -0.22736 0.20757 0.04381 0.14096 0.908077
## Triturus       -1.4050 0.73688 0.05765 -0.16664 -0.22457 -0.089970
## Bufo           -1.0305 0.47098 -0.79879 0.31354 -0.20087 -0.851458
## Rana            -1.4129 0.76139 0.37817 -0.35195 -0.10910 0.163806
## Salamandra     0.7689 0.40083 0.33082 -0.09962 0.30992 -0.110610
## Lacerta        0.2275 -0.06517 -0.35909 -0.13754 -0.04889 0.050076
## Podarcis       -0.6532 -0.92747 0.34514 -0.19702 -0.10903 -0.126929
## Coluber        0.5715 0.06924 0.16295 0.16178 -0.13829 0.002801
## Malpolon       0.2864 0.02326 0.06220 -0.22884 0.34774 -0.078629
## Psamodromus   -0.8750 -0.09060 -0.11225 0.36752 0.19678 0.005858
##
##
## Site scores (weighted averages of species scores)
##
##           CA1      CA2      CA3      CA4      CA5      CA6
## Eucalip1      -1.8378 1.24107 1.9682714 -2.2844 -0.8864 1.46841
## Eucalip2      -1.6903 1.70073 -0.7683320 1.1289 0.3499 -0.18031
## Eucalip3      -1.5521 0.95266 -0.2880435 0.6300 -0.1375 -1.59192
## Eucalip4      -1.4784 0.51800 -0.1493270 0.3650 0.5072 -0.16045
## Eucalip5      -0.7855 -0.21678 -1.6392392 -0.5009 -1.2758 -0.90399
## Eucalip6      -0.6798 -0.90007 -0.1168264 1.5494 1.3665 2.67892
## Eucalip7      -0.6587 -2.68435 1.0167252 -0.9794 0.2880 -0.42184
## Eucalip8      -0.3277 -1.34982 0.6323698 0.9567 -0.8997 0.14569
## Pinhal1       -0.3174 -1.67710 0.7070600 -0.2640 -0.7725 -1.62401
## Pinhal2       -0.2729 -0.98691 -0.5607742 1.4227 1.2752 -0.38206
## Pinhal3        0.2864 0.10054 0.0322563 -1.2341 1.0312 0.42323
## Pinhal4        0.5922 -0.03768 -1.6523057 -0.7708 -0.1172 -0.43146
## Pinhal5        0.5368 -0.35369 -0.6373104 -0.2149 -0.5060 0.84825

```

```
## Pinhal6  0.8310  0.06962 -1.3868831 -1.1190 -0.2211  0.38989
## Pinhal7  0.8616  0.28254 -0.7186440  0.5525 -1.5195  0.52699
## Pinhal8  0.6973  0.07484 -0.7684806 -1.2209  1.2522  0.57878
## Sobro1   1.0841  0.60666  0.6022968  0.2312 -0.1758 -0.36525
## Sobro2   0.8779  0.55443  0.1752309 -0.6501  2.0473 -0.72125
## Sobro3   0.9984  0.44945 -0.0006649 -0.3582  0.1627 -0.22688
## Sobro4   0.9901  0.30751  0.7987011  1.0692 -1.3731  1.22595
## Sobro5   1.0784  0.48887  0.3728823  0.7538 -1.4952  0.06327
## Sobro6   1.2374  0.96757  2.2301775  0.8152  0.9128 -1.24361
```

```
plot(ca.herp)
```



we could have used package `ca`, but we note these analysis are exactly the same, and as before for different PCA implementations, only the structure of the output objects is slightly different

```
library(ca)
caca.herp<-ca(herpeto[-1])
summary(caca.herp)
```

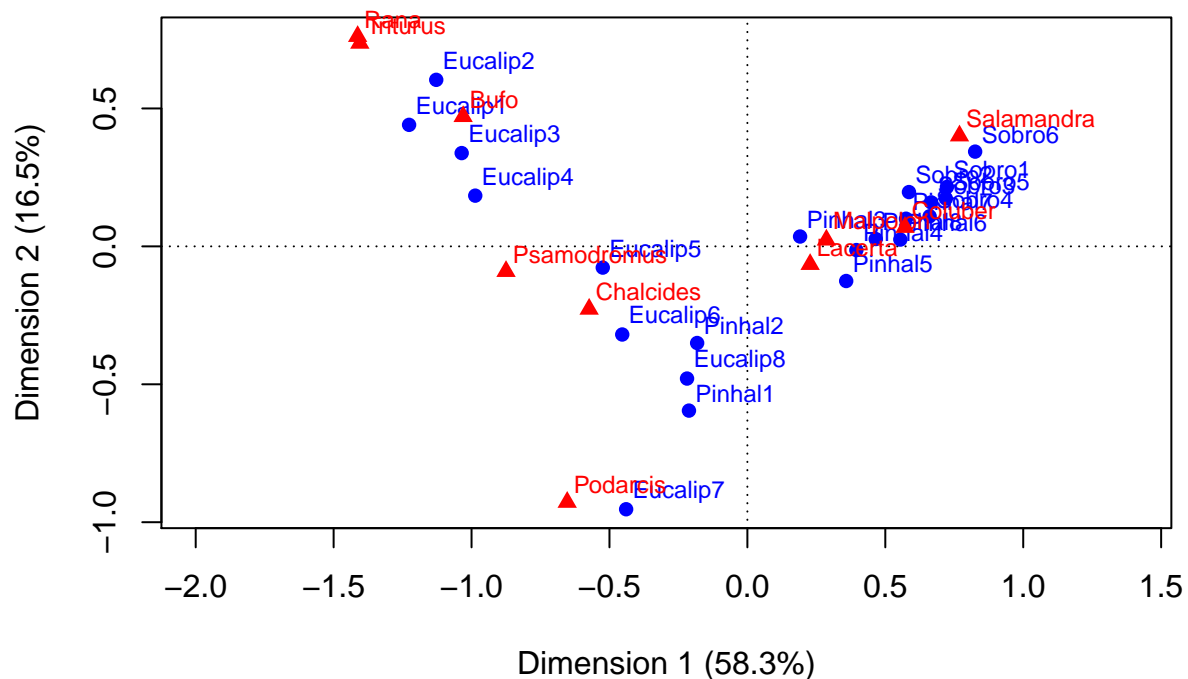
```
##
## Principal inertias (eigenvalues):
##
## dim   value    %  cum%  scree plot
## 1     0.445292 58.3 58.3  *****
## 2     0.126030 16.5 74.8  ****
## 3     0.069639  9.1 84.0  **
## 4     0.044091  5.8 89.7  *
## 5     0.031472  4.1 93.9  *
```

```

## 6      0.023908  3.1  97.0  *
## 7      0.012269  1.6  98.6
## 8      0.008908  1.2  99.8
## 9      0.001775  0.2 100.0
##
## -----
## Total: 0.763383 100.0
##
##
## Rows:
##   name  mass  qlt  inr   k=1 cor ctr   k=2 cor ctr
## 1 | Ecl1 |  46 746 137 | -1226 661 156 | 441 85 71 |
## 2 | Ecl2 |  46 938 105 | -1128 729 132 | 604 209 133 |
## 3 | Ecl3 |  49 923  82 | -1036 834 118 | 338 89 44 |
## 4 | Ecl4 |  46 937  65 | -987 906 101 | 184 31 12 |
## 5 | Ecl5 |  45 443  37 | -524 433  28 | -77  9  2 |
## 6 | Ecl6 |  48 460  42 | -454 307  22 | -320 152 39 |
## 7 | Ecl7 |  46 899  73 | -440 158  20 | -953 741 329 |
## 8 | Ecl8 |  47 732  23 | -219 126  5 | -479 605 86 |
## 9 | Pnh1 |  47 766  32 | -212 86  5 | -595 680 132 |
## 10 | Pnh2 |  45 458  20 | -182 97  3 | -350 360 44 |
## 11 | Pnh3 |  47 252  9 | 191 244  4 | 36  9  0 |
## 12 | Pnh4 |  46 399  23 | 395 398 16 | -13  0  0 |
## 13 | Pnh5 |  46 686  13 | 358 611 13 | -126 75  6 |
## 14 | Pnh6 |  44 561  31 | 555 559 30 | 25  1  0 |
## 15 | Pnh7 |  44 635  31 | 575 617 32 | 100 19  3 |
## 16 | Pnh8 |  45 563  23 | 465 561 22 | 27  2  0 |
## 17 | Sbr1 |  44 933  35 | 723 857 52 | 215 76 16 |
## 18 | Sbr2 |  44 691  32 | 586 621 34 | 197 70 13 |
## 19 | Sbr3 |  43 982  27 | 666 928 43 | 160 53  9 |
## 20 | Sbr4 |  43 595  43 | 661 579 42 | 109 16  4 |
## 21 | Sbr5 |  43 830  37 | 720 784 50 | 174 46 10 |
## 22 | Sbr6 |  48 640  78 | 826 546 73 | 343 94 44 |
##
## Columns:
##   name  mass  qlt  inr   k=1 cor ctr   k=2 cor ctr
## 1 | Chlc |  16 257  31 | -574 222 12 | -227 35  7 |
## 2 | Trtr |  45 956 156 | -1405 750 200 | 737 206 194 |
## 3 | Bufo |  9 399  36 | -1030 330 20 | 471 69 15 |
## 4 | Rana |  32 875 123 | -1413 678 143 | 761 197 146 |
## 5 | Slmn |  56 692  79 | 769 544 74 | 401 148 71 |
## 6 | Lcrt | 262 265  72 | 227 245 30 | -65 20  9 |
## 7 | Pdrc |  79 872 152 | -653 289 76 | -927 583 538 |
## 8 | Clbr | 284 818 151 | 571 806 208 | 69 12 11 |
## 9 | Mlpl |  90 259  38 | 286 257 17 | 23  2  0 |
## 10 | Psmd | 128 802 162 | -875 793 221 | -91  9  8 |

```

```
plot(caca.herp)
```

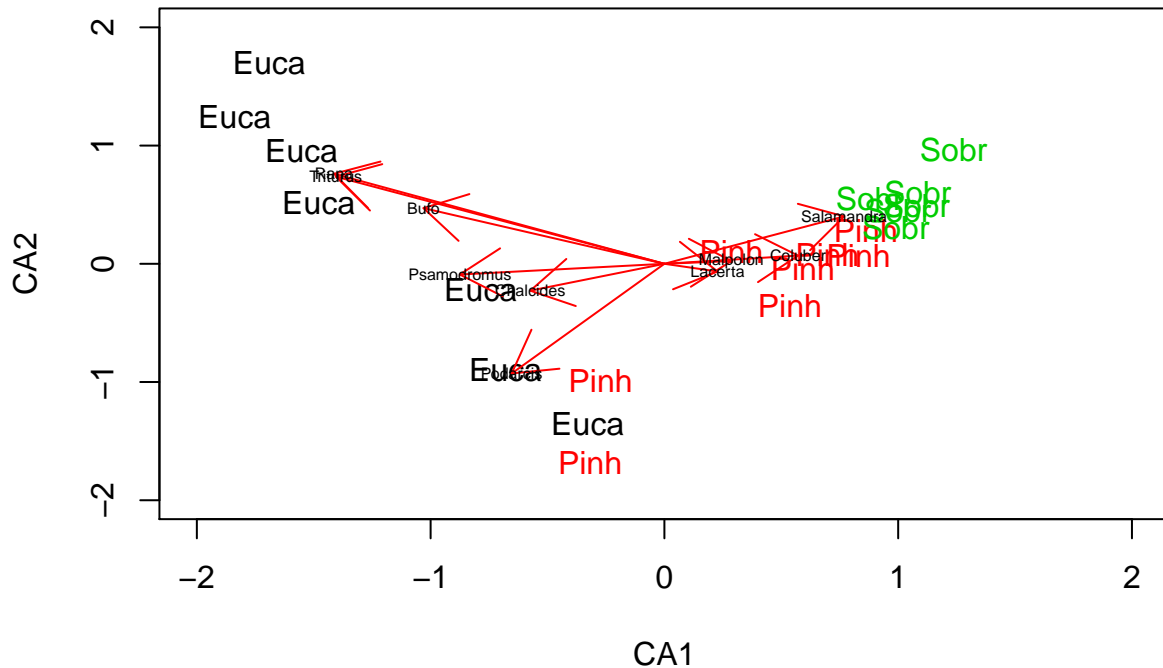


We note that the separation per habitat is perhaps not as clean as for the PCA. Again, we can make our own biplot

```
#a couple of useful objects to make the code more readable
names(summary(ca.herp))

## [1] "species"      "sites"        "call"         "tot.chi"     "unconst.chi"
## [6] "cont"         "scaling"     "digits"      "inertia"     "method"

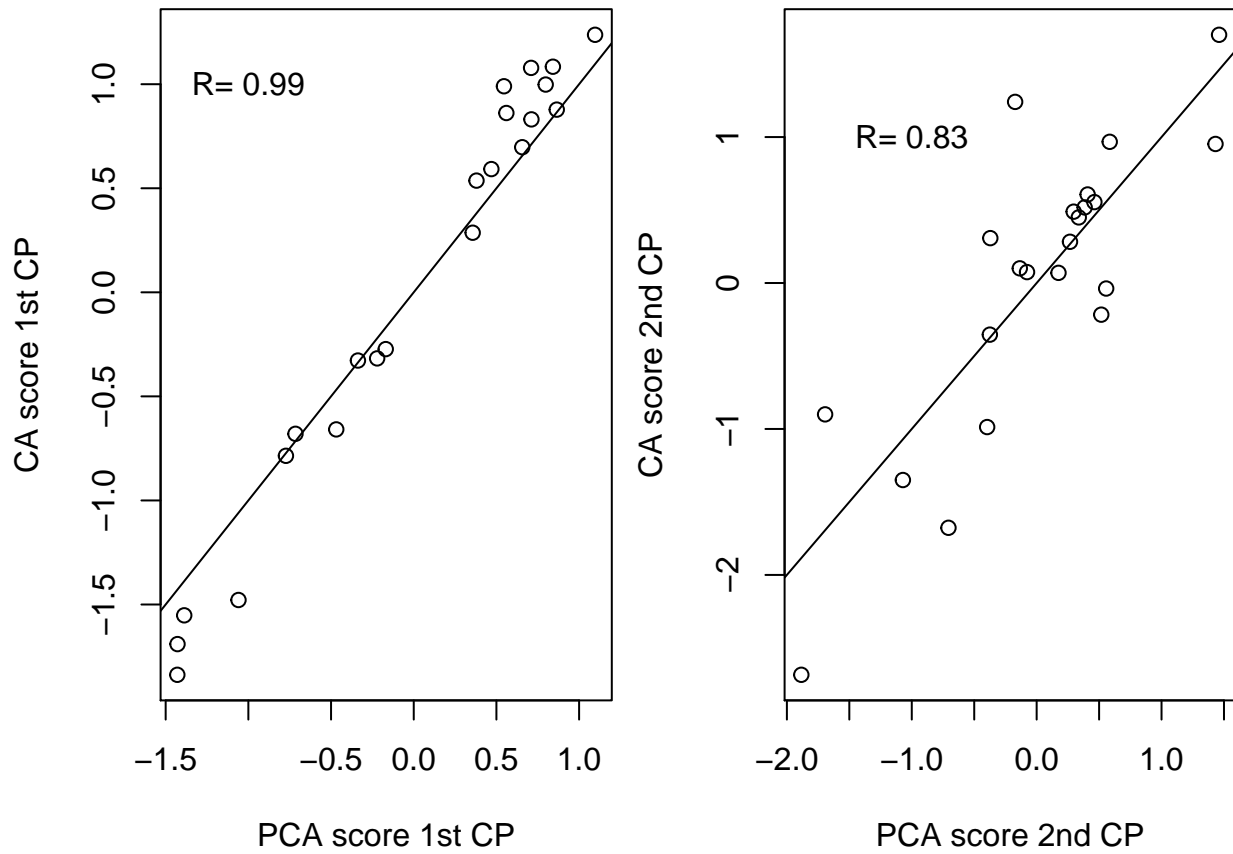
mycas=summary(ca.herp)
#the biplot
plot(mycas$sites[,1:2],type="n",ylim=c(-2,2),xlim=c(-2,2))
text(mycas$sites[,1:2],labels=substr(herpeto[,1],1,4),col=as.numeric(as.factor(substr(herpeto[,1],1,4)))
#species
arrows(x0=0,y0=0,x1=mycas$species[,1],y1=mycas$species[,2],col=2)
text(mycas$species[,1:2],labels=names(herpeto[-1]),cex=0.5)
```



Comparison CA vs. PCA

The results from both analysis are not that different in this case. While this is not a standard procedure (i.e. I have never seen anyone else doing it explicitly), we could compare the relations between the scores in the PCA and the CA, for the first and second axis.

```
par(mfrow=c(1,2),mar=c(4,4,0.5,0.5))
plot(myacp$sites[,1],mycas$sites[,1],xlab="PCA score 1st CP",ylab="CA score 1st CP")
abline(0,1)
text(-1,1,paste0("R= ",round(cor(myacp$sites[,1],mycas$sites[,1]),2)))
plot(myacp$sites[,2],mycas$sites[,2],xlab="PCA score 2nd CP",ylab="CA score 2nd CP")
text(-1,1,paste0("R= ",round(cor(myacp$sites[,2],mycas$sites[,2]),2)))
abline(0,1)
```

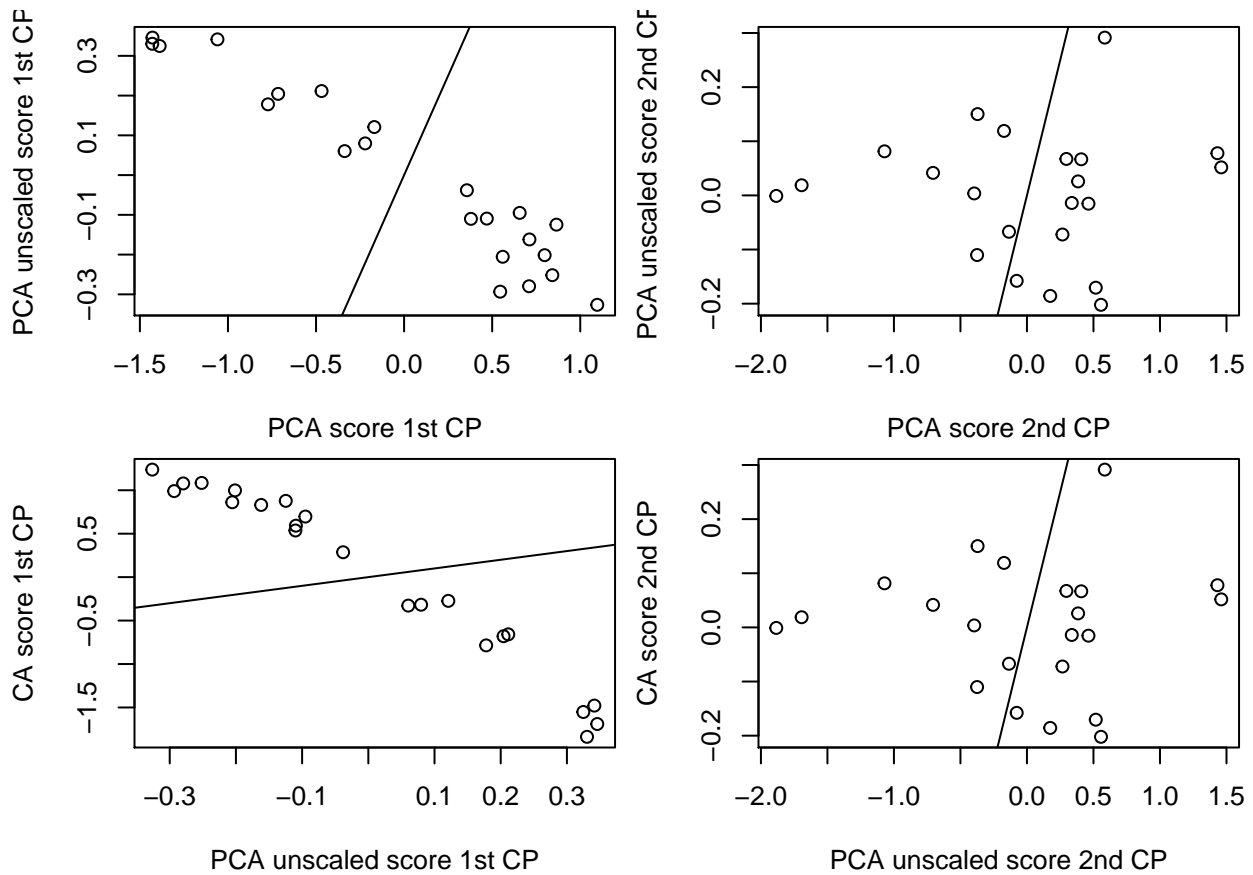



With a correlation of over 99% in the first axis and over 80% in the second, there is nothing structurally different in both analysis.

In general, a CA might be more sensible to use directly over count data.

Actually, in this case, similar plots to the pair above comparing the PCA with and without scaling (i.e., over the correlation and over the covariance matrices) and between the covariance-based PCA and the CA would also be informative.

```
par(mfrow=c(2,2),mar=c(4,4,0.5,0.5))
plot(myacp$sites[,1],sumacp1$scores[,1],xlab="PCA score 1st CP",ylab="PCA unscaled score 1st CP")
abline(0,1)
text(-1,1,paste0("R= ",round(cor(myacp$sites[,1],sumacp1$scores[,1]),2)))
plot(myacp$sites[,2],sumacp1$scores[,2],xlab="PCA score 2nd CP",ylab="PCA unscaled score 2nd CP")
text(-1,1,paste0("R= ",round(cor(myacp$sites[,2],sumacp1$scores[,2]),2)))
abline(0,1)
# PCA unscaled vs CA
plot(sumacp1$scores[,1],mycas$sites[,1],xlab="PCA unscaled score 1st CP",ylab="CA score 1st CP")
abline(0,1)
text(-1,1,paste0("R= ",round(cor(sumacp1$scores[,1],mycas$sites[,1]),2)))
plot(myacp$sites[,2],sumacp1$scores[,2],xlab="PCA unscaled score 2nd CP",ylab="CA score 2nd CP")
text(-1,1,paste0("R= ",round(cor(sumacp1$scores[,2],mycas$sites[,2]),2)))
abline(0,1)
```



These plots show us what we already knew. The results of the unscaled PCA would not be sensible, being very different from those of the CA and the scaled PCA. This was as expected, as both the CA and the unscaled PCA are not dominated by the raw abundances, unlike the unscaled PCA (covariance matrix based).

Exercício 4

Efectue uma análise canônica de correspondências aos dados da alínea anterior, incluindo também as características dos locais (DataTP10herpamb.csv). Discuta os resultados. Em que situações esta análise é preferível em relação às anteriores técnicas?

First we read the environmental variables

```
herpamb <- read.csv("DataTP10herpamb.csv", sep=";")
```

and take a quick peak at it to make sure all is good

```
str(herpamb)
```

```
## 'data.frame':  22 obs. of  6 variables:
## $ Habitat      : Factor w/ 22 levels "Eucalip1","Eucalip2",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Perc.floresta : int  75 85 49 65 68 67 59 58 45 45 ...
## $ Agua         : int  2 1 1 1 1 0 0 0 0 0 ...
## $ Altura       : int  12 10 9 8 12 15 14 15 13 14 ...
## $ Idade        : int  9 8 9 8 7 8 9 8 8 9 ...
## $ Activ.agricola: int  1 1 1 2 2 1 1 1 2 2 ...
```

Then, we implement the CCA using function `cca` from package `vegan`

```
cca.herp<-cca(herpeto[,-1]~.,data=herpamb[,-1])
```

We can look at the summary of the results

```
summary(cca.herp)
```

```
##
## Call:
## cca(formula = herpeto[, -1] ~ Perc.floresta + Agua + Altura + Idade + Activ.agricola, data = herpamb[, -1])
##
## Partitioning of scaled Chi-square:
##           Inertia Proportion
## Total           0.7634      1.0000
## Constrained     0.4915      0.6439
## Unconstrained   0.2719      0.3561
##
## Eigenvalues, and their contribution to the scaled Chi-square
##
## Importance of components:
##           CCA1    CCA2    CCA3    CCA4    CCA5    CA1
## Eigenvalue      0.3658 0.09177 0.0200 0.009837 0.00413 0.1080
## Proportion Explained 0.4792 0.12022 0.0262 0.012885 0.00541 0.1415
## Cumulative Proportion 0.4792 0.59938 0.6256 0.638467 0.64388 0.7853
##           CA2    CA3    CA4    CA5    CA6    CA7
## Eigenvalue      0.05750 0.03331 0.02914 0.01788 0.01235 0.008604
## Proportion Explained 0.07532 0.04363 0.03818 0.02343 0.01618 0.011271
## Cumulative Proportion 0.86067 0.90430 0.94247 0.96590 0.98208 0.993351
##           CA8    CA9
## Eigenvalue      0.004061 0.001015
## Proportion Explained 0.005320 0.001329
## Cumulative Proportion 0.998671 1.000000
##
## Accumulated constrained eigenvalues
## Importance of components:
##           CCA1    CCA2    CCA3    CCA4    CCA5
## Eigenvalue      0.3658 0.09177 0.02000 0.009837 0.004130
## Proportion Explained 0.7442 0.18671 0.04069 0.020012 0.008402
## Cumulative Proportion 0.7442 0.93090 0.97159 0.991598 1.000000
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
##
## Species scores
##           CCA1    CCA2    CCA3    CCA4    CCA5    CA1
## Chalcides    -0.6113 -0.29111 -0.572651 0.19521 -0.33274 -0.06196
## Triturus     -1.2902 0.71449 -0.009936 0.02656 0.02877 -0.48284
## Bufo         -0.9295 0.23703 0.280710 -0.36696 0.15084 -0.15629
## Rana         -1.3053 0.71249 -0.114317 0.18142 0.06693 -0.54825
## Salamandra   0.7705 0.43184 0.118288 0.02470 -0.11919 0.08011
## Lacerta      0.1164 -0.10010 0.173742 0.06916 -0.01033 0.34121
## Podarcis     -0.5020 -0.63199 -0.091435 0.11269 0.07508 -0.61205
```

```

## Coluber      0.5613  0.06190 -0.116579 -0.04956  0.03247  0.11979
## Malpolon    0.2333  0.07466 -0.055202  0.01425  0.01962  0.15423
## Psamodromus -0.7509 -0.19039  0.031421 -0.17627 -0.05380 -0.40467
##
##
## Site scores (weighted averages of species scores)
##
##           CCA1      CCA2      CCA3      CCA4      CCA5      CA1
## Eucalip1 -2.0152  1.87412 -2.8930  4.81834  1.10819 -0.81906
## Eucalip2 -1.8793  1.65598  0.6424 -3.91983 -1.31085  0.50931
## Eucalip3 -1.6983  1.00010  0.2777 -3.16262  2.05190 -1.49156
## Eucalip4 -1.6230  0.50958  0.4475 -1.65147 -0.43073 -0.87193
## Eucalip5 -0.9153 -0.38134  2.3268  0.35012  0.95296  0.60554
## Eucalip6 -0.7387 -1.32824 -1.5781 -1.51391 -7.43262  0.02136
## Eucalip7 -0.6709 -2.79755 -1.5910  2.85385  2.85197 -0.95704
## Eucalip8 -0.3082 -1.53374 -1.5519 -0.66192  2.00687 -0.37185
## Pinhal1  -0.2947 -1.69826 -0.7898  0.65662  5.24818 -1.38288
## Pinhal2  -0.2811 -1.37301  0.3499 -3.18278 -0.73277 -0.78594
## Pinhal3   0.2879  0.21061  0.2525  2.17733 -0.94590  0.23520
## Pinhal4   0.5812 -0.19994  2.5969  0.87950  1.02682  1.61298
## Pinhal5   0.5601 -0.48814  0.9088  1.36086 -1.23298  0.69684
## Pinhal6   0.8351 -0.01716  1.9636  1.96754  1.23469  2.41756
## Pinhal7   0.9260  0.17637  1.5098  0.04353 -0.52020  0.46418
## Pinhal8   0.7119  0.06910  1.6312  2.19994 -3.05349  0.94164
## Sobro1    1.2139  0.79587 -0.1443 -0.24031 -0.24692  0.48075
## Sobro2    0.9624  0.73900  1.0770  0.24228 -2.53746 -0.19991
## Sobro3    1.0866  0.57212  0.5625  0.55470  0.02748  0.65598
## Sobro4    1.1086  0.34669 -2.9852 -0.86866  1.34425  0.86836
## Sobro5    1.2042  0.57493 -0.3852 -0.55785  1.13225 -1.35850
## Sobro6    1.4536  1.41329 -2.2943 -1.95167 -0.46756 -0.91294
##
##
## Site constraints (linear combinations of constraining variables)
##
##           CCA1      CCA2      CCA3      CCA4      CCA5      CA1
## Eucalip1 -1.80552  1.5084 -0.832575  1.58192  0.21614 -0.81906
## Eucalip2 -2.05667  0.8323 -0.195676 -1.03400 -1.38874  0.50931
## Eucalip3 -0.77643  0.5318  0.902637 -0.84872  1.90616 -1.49156
## Eucalip4 -1.20256  1.0830  0.685743 -0.89444 -0.16342 -0.87193
## Eucalip5 -0.99294  0.4472 -0.227171  0.80894 -0.71615  0.60554
## Eucalip6 -0.74313 -1.4356 -0.617769 -0.20748 -0.87298  0.02136
## Eucalip7 -0.53871 -1.3844 -0.292870 -0.54116 -0.05337 -0.95704
## Eucalip8 -0.39734 -1.5486 -0.369172 -0.04068 -0.07734 -0.37185
## Pinhal1   0.18782 -1.2099  0.318893 -0.03086  0.52915 -1.38288
## Pinhal2   0.23282 -1.4062 -0.061265  0.33855  0.52477 -0.78594
## Pinhal3   0.27892 -0.2193 -0.638832  3.05769  0.44809  0.23520
## Pinhal4   0.06598 -0.9885  0.643808 -0.43734  0.35672  1.61298
## Pinhal5   0.29016 -0.7996  0.599201  0.22062 -0.44519  0.69684
## Pinhal6  -0.35292  0.6012  1.057962  0.42225  0.97228  2.41756
## Pinhal7   0.77362 -0.7442  1.476561  0.14841  0.65444  0.46418
## Pinhal8   0.33392  0.1281  0.987506  0.12912 -2.17349  0.94164
## Sobro1    0.76593  1.3335 -0.064933  0.28099  2.07343  0.48075
## Sobro2    1.19962  0.6491 -0.228898 -0.72637 -0.21643 -0.19991
## Sobro3    0.82461  0.6593  0.002506 -1.84708  0.46705  0.65598

```

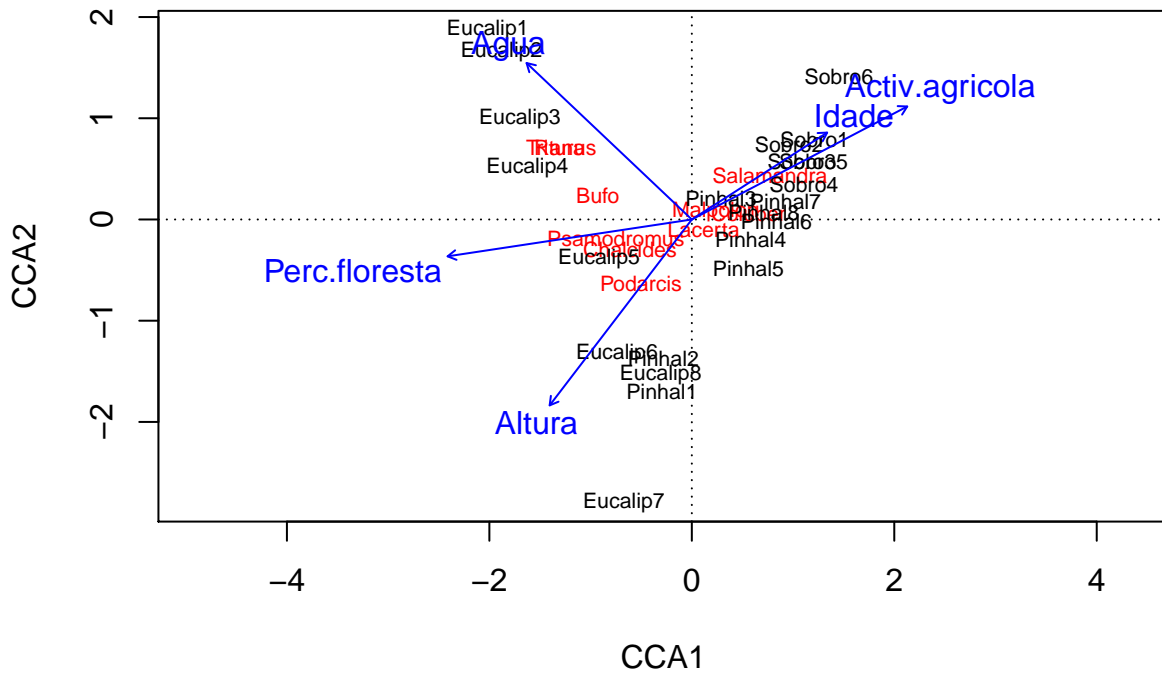
```

## Sobro4    0.87917  0.1031 -3.775111 -0.95601  0.39607  0.86836
## Sobro5    1.99442  1.0410  0.400087  0.92582 -1.50583 -1.35850
## Sobro6    1.41722  1.0218  0.134275 -0.41928 -0.92556 -0.91294
##
##
## Biplot scores for constraining variables
##
##          CCA1    CCA2    CCA3    CCA4    CCA5 CA1
## Perc.floresta -0.9402 -0.1418  0.06674  0.1045 -0.28363  0
## Agua          -0.6362  0.6024 -0.01142  0.3892  0.28406  0
## Altura        -0.5474 -0.7157  0.06631  0.4279  0.02579  0
## Idade          0.5211  0.3345 -0.70051 -0.3370  0.11056  0
## Activ.agricola 0.8294  0.4350 -0.26062 -0.1403 -0.18782  0

```

and finally we plot the triplot that contains the information on the sites and the species information, as the previous analysis did, but now it also contains directly the information about the environmental variables. The ecological interpretation would now have to be done considering the 3 sets of variables, as the environmental variables were used to constrain the axis.

```
plot(cca.herp)
```



Exercício 5

Obteve dados relativos à abundância de diversas espécies de aves em três tipos de habitat. Efectue várias análises de ordenação (análise de componentes principais, análise de correspondências e análise canónica de

correspondências). A análise canônica de correspondências permite integrar informação de uma segunda matriz de dados. Indique em que situações esta análise poderá ser mais informativa e como poderia tentar obter informação análoga nas outras duas análises efectuadas.

We read the data in, both the species matrix and the environmental variables matrix

```
aves <- read.csv("DataTP10aves.csv", sep=";")
avesamb <- read.csv("DataTP10avesamb.csv", sep=";")
row.names(aves)=row.names(avesamb)=aves[,1]
```

We look at both of these, first the species data

```
str(aves)

## 'data.frame': 22 obs. of 11 variables:
## $ Habitat : Factor w/ 22 levels "Bosq 1","Bosq 2",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Erithacus : num 0.05 0.02 0.01 0.01 0.02 0.08 0.03 0.02 0 0.01 ...
## $ Turdus : num 0.22 0.18 0.18 0.15 0.12 0.03 0 0.02 0.03 0 ...
## $ Passer : num 0 0.04 0.05 0.01 0.04 0 0 0 0.01 0.01 ...
## $ Parus : num 0.2 0.13 0.11 0.1 0.02 0.02 0.02 0.02 0.01 0 ...
## $ Athene : num 0.03 0 0 0.01 0 0.02 0 0 0.01 0.01 ...
## $ Anas : num 0.06 0.15 0.13 0.17 0.37 0.21 0.19 0.19 0.2 0.23 ...
## $ Sylvia : num 0.16 0 0.09 0.1 0.12 0.12 0.34 0.2 0.26 0.11 ...
## $ Carduelis : num 0.03 0.05 0.08 0.04 0.08 0.16 0.11 0.31 0.25 0.21 ...
## $ Motacilla : num 0.06 0.04 0.06 0.05 0.02 0.05 0.1 0.03 0.08 0.1 ...
## $ Dendrocopos: num 0.16 0.36 0.32 0.33 0.16 0.32 0.17 0.2 0.14 0.27 ...
```

We have 22 observations for 11 variables, which correspond respectively to locations and taxa (genera of amphibians). The first column does not represent a genera, but it contains the labels of the locations, which are separated by habitats, of which there are the following 3

```
unique(substr(aves[,1],1,4))
```

```
## [1] "Bosq" "Prad" "Ripi"
```

which we assume correspond to “Bosque”, “Pradaria”, and “Ripícola”.

We can see if there are locations with much more abundance than others

```
sort(rowSums(aves[,-1]))

## Ripic 3 Ripic 4 Ripic 5 Prado 6 Prado 7 Ripic 2 Ripic 1 Bosq 5 Prado 8
## 0.90 0.91 0.91 0.92 0.92 0.92 0.93 0.95 0.95
## Prado 2 Prado 5 Bosq 7 Prado 4 Bosq 1 Bosq 2 Bosq 4 Prado 3 Bosq 8
## 0.95 0.96 0.96 0.96 0.97 0.97 0.97 0.98 0.99
## Prado 1 Ripic 6 Bosq 6 Bosq 3
## 0.99 1.00 1.01 1.03
```

(as before, some standardization must have been made before analysis)

or if there are species which dominate the communities.

```
sort(colSums(aves[,-1]))

## Passer Erithacus Parus Turdus Athene Sylvia
## 0.18 0.34 0.67 0.95 1.17 1.66
## Motacilla Dendrocopos Anas Carduelis
## 1.89 2.70 5.52 5.97
```

In fact, *Anas* and *Carduelis* seem to dominate the community in terms of abundance.

Now we look at environmental variables

```
str(avesamb)
```

```
## 'data.frame':  22 obs. of  4 variables:
## $ Habitat : Factor w/ 22 levels "Bosq 1","Bosq 2",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Perc.veg: int  75 85 62 56 59 89 84 86 12 12 ...
## $ Agua    : int  25 23 32 34 35 36 34 28 7 8 ...
## $ Solo    : int  50 52 62 52 54 58 59 63 23 25 ...
```

We have 22 observations for 4 variables, which correspond respectively to locations and environmental variables (genera of amphibians). The first column does not represent a variable, but it contains the labels of the locations.

A number of different analysis could be attempted on this data, as shown below. We use the PCA unscaled, but this is wrong given the raw abundances (for the environmental variables analysis might be equivalent scaled or unscaled, as they seem to be measured in similar scales). Your task is to repeat the analysis with the correlation matrix and compare the results.

Environmental variables

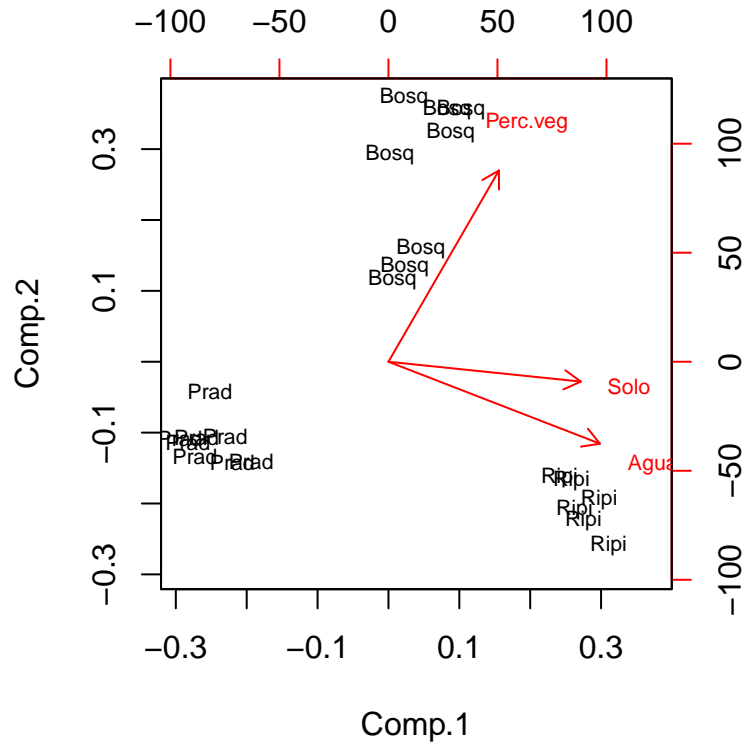
We could see how the sites would group as a function of the environmental variables, both considering a PCA and a CA

PCA

```
pca.avesamb<-princomp(avesamb[,-1])
summary(pca.avesamb)
```

```
## Importance of components:
##
##          Comp.1      Comp.2      Comp.3
## Standard deviation  37.473629 25.5582232 5.3671214
## Proportion of Variance 0.673092  0.3131008 0.0138072
## Cumulative Proportion 0.673092  0.9861928 1.0000000
```

```
biplot(pca.avesamb,cex=0.7,xlabs=substr(avesamb[,1],1,4))
```



We can see a very clear separation of the 3 habitat types based on the sites environmental variables.

CA

```
ca.avesamb<-ca(avesamb[, -1])
ca.avesamb
```

```
##
## Principal inertias (eigenvalues):
##      1      2
## Value  0.093586 0.011924
## Percentage 88.7%  11.3%
##
##
## Rows:
##      Bosq 1   Bosq 2   Bosq 3   Bosq 4   Bosq 5   Bosq 6
## Mass    0.050744 0.054127 0.052774 0.048038 0.050068 0.061908
## ChiDist 0.404234 0.475416 0.200678 0.167570 0.177074 0.365468
## Inertia 0.008292 0.012234 0.002125 0.001349 0.001570 0.008269
## Dim. 1  -1.315845 -1.549911 -0.641840 -0.536668 -0.568257 -1.162157
## Dim. 2   0.338420 0.317972 -0.379574 0.307247 0.308511 0.775381
##      Bosq 7   Bosq 8   Prado 1   Prado 2   Prado 3   Prado 4
## Mass    0.059878 0.059878 0.014208 0.015223 0.016576 0.019621
## ChiDist 0.343336 0.384992 0.316218 0.319674 0.179729 0.197294
## Inertia 0.007058 0.008875 0.001421 0.001556 0.000535 0.000764
```

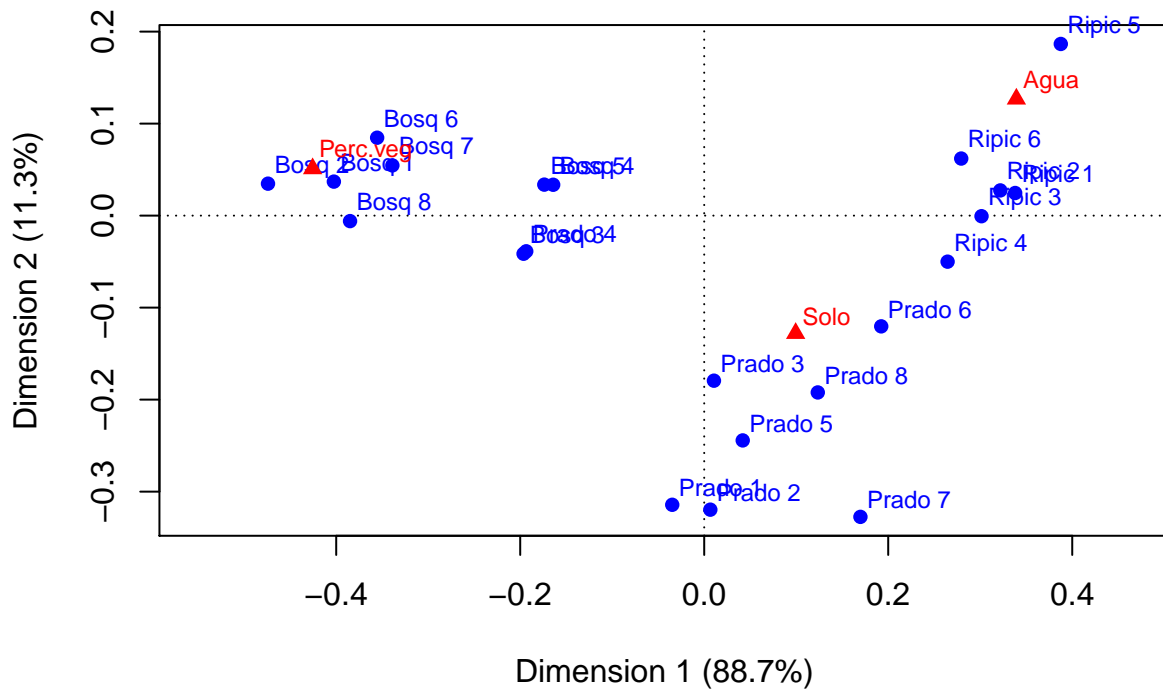


```

## Dim. 1 -1.108038 -1.258329 -0.113654 0.021691 0.034577 -0.632347
## Dim. 2 0.499884 -0.054639 -2.878278 -2.926849 -1.643051 -0.355071
##      Prado 5  Prado 6  Prado 7  Prado 8  Ripic 1  Ripic 2
## Mass 0.020636 0.023681 0.020974 0.015900 0.069012 0.067997
## ChiDist 0.247898 0.226933 0.368830 0.228426 0.338865 0.322988
## Inertia 0.001268 0.001220 0.002853 0.000830 0.007925 0.007094
## Dim. 1 0.136803 0.628879 0.555023 0.403498 1.104777 1.051996
## Dim. 2 -2.237598 -1.102222 -2.998450 -1.760127 0.225210 0.250809
##      Ripic 3  Ripic 4  Ripic 5  Ripic 6
## Mass 0.071719 0.066644 0.072057 0.068336
## ChiDist 0.301397 0.269222 0.430147 0.286161
## Inertia 0.006515 0.004830 0.013332 0.005596
## Dim. 1 0.985218 0.864708 1.266927 0.913179
## Dim. 2 -0.005636 -0.458255 1.708621 0.568026
##
##
## Columns:
##      Perc.veg  Agua  Solo
## Mass 0.316982 0.279432 0.403586
## ChiDist 0.428703 0.362222 0.161991
## Inertia 0.058257 0.036663 0.010590
## Dim. 1 -1.391381 1.109118 0.324890
## Dim. 2 0.467768 1.161271 -1.171424

```

```
plot(ca.avesamb, labels=2)
```



Again, the separation across habitats is very clear.

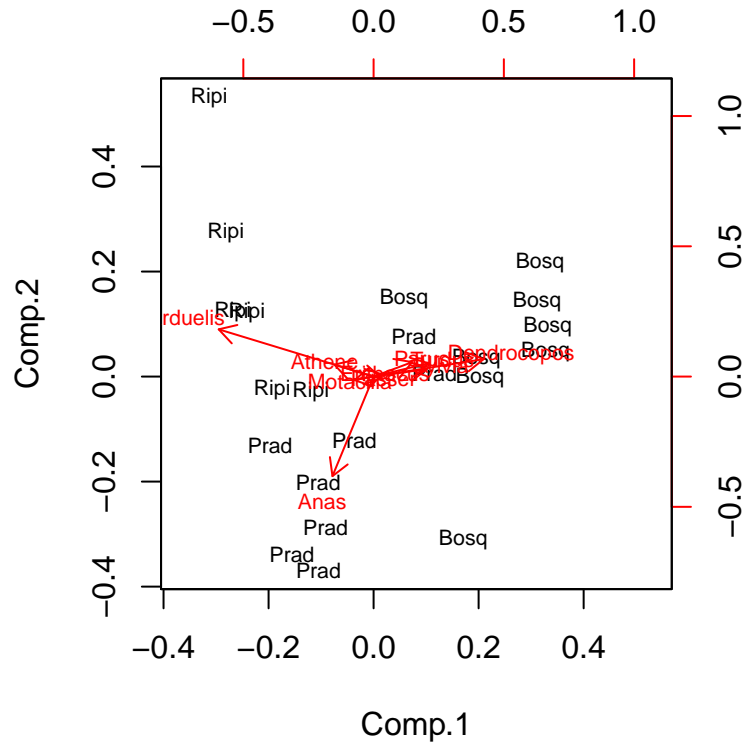
Species data

PCA

```
pca.aves<-princomp(aves[, -1])  
summary(pca.aves)
```

```
## Importance of components:  
##              Comp.1    Comp.2    Comp.3    Comp.4  
## Standard deviation  0.2222533 0.1168578 0.08479853 0.06517432  
## Proportion of Variance 0.6307611 0.1743747 0.09182164 0.05424020  
## Cumulative Proportion 0.6307611 0.8051358 0.89695745 0.95119765  
##              Comp.5    Comp.6    Comp.7    Comp.8  
## Standard deviation  0.04942171 0.027323770 0.019466369 0.012803054  
## Proportion of Variance 0.03118917 0.009533439 0.004838807 0.002093127  
## Cumulative Proportion 0.98238683 0.991920266 0.996759072 0.998852199  
##              Comp.9    Comp.10  
## Standard deviation  0.0077115470 0.0055153726  
## Proportion of Variance 0.0007593664 0.0003884347  
## Cumulative Proportion 0.9996115653 1.0000000000
```

```
biplot(pca.aves, cex=0.7, xlab=substr(avesamb[, 1], 1, 4))
```



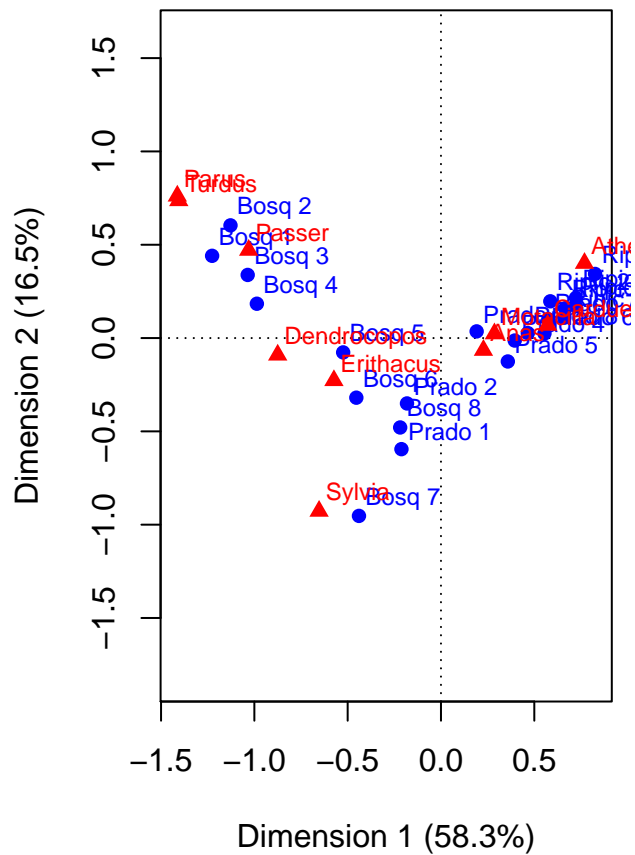
CA

```
ca.aves<-ca(aves[,-1])
ca.aves
```

```
##
## Principal inertias (eigenvalues):
##      1      2      3      4      5      6      7
## Value  0.445292 0.12603 0.069639 0.044091 0.031472 0.023908 0.012269
## Percentage 58.33% 16.51% 9.12% 5.78% 4.12% 3.13% 1.61%
##      8      9
## Value  0.008908 0.001775
## Percentage 1.17% 0.23%
##
##
## Rows:
##      Bosq 1  Bosq 2  Bosq 3  Bosq 4  Bosq 5  Bosq 6
## Mass      0.046081 0.046081 0.048931 0.046081 0.045131 0.047981
## ChiDist   1.508272 1.321314 1.133841 1.036599 0.796308 0.818294
## Inertia   0.104828 0.080451 0.062906 0.049516 0.028618 0.032128
## Dim. 1   -1.837804 -1.690296 -1.552097 -1.478447 -0.785535 -0.679849
## Dim. 2    1.241071 1.700734 0.952661 0.517997 -0.216784 -0.900073
##      Bosq 7  Bosq 8  Prado 1  Prado 2  Prado 3  Prado 4
## Mass      0.045606 0.047031 0.047031 0.045131 0.046556 0.045606
## ChiDist   1.107109 0.615837 0.721845 0.583780 0.386926 0.625996
## Inertia   0.055898 0.017837 0.024506 0.015380 0.006970 0.017872
## Dim. 1   -0.658662 -0.327748 -0.317424 -0.272950 0.286386 0.592153
## Dim. 2   -2.684345 -1.349816 -1.677101 -0.986909 0.100543 -0.037683
##      Prado 5  Prado 6  Prado 7  Prado 8  Ripic 1  Ripic 2  Ripic 3
## Mass      0.045606 0.043705 0.043705 0.045131 0.044181 0.043705 0.042755
## ChiDist   0.458169 0.741372 0.732269 0.620944 0.781261 0.743498 0.691484
## Inertia   0.009573 0.024022 0.023436 0.017401 0.026966 0.024160 0.020443
## Dim. 1    0.536833 0.830987 0.861648 0.697250 1.084080 0.877864 0.998400
## Dim. 2   -0.353691 0.069624 0.282541 0.074841 0.606658 0.554429 0.449455
##      Ripic 4  Ripic 5  Ripic 6
## Mass      0.043230 0.043230 0.047506
## ChiDist   0.868291 0.812487 1.117646
## Inertia   0.032593 0.028538 0.059341
## Dim. 1    0.990091 1.078401 1.237441
## Dim. 2    0.307511 0.488870 0.967568
##
##
## Columns:
##      Erithacus  Turdus  Passer  Parus  Athene  Anas
## Mass      0.016152 0.045131 0.008551 0.031829 0.055582 0.262233
## ChiDist   1.217291 1.622363 1.792900 1.715545 1.042239 0.459361
## Inertia   0.023934 0.118787 0.027487 0.093676 0.060377 0.055334
## Dim. 1   -0.859618 -2.105453 -1.544222 -2.117302 1.152327 0.340921
## Dim. 2   -0.640447 2.075667 1.326694 2.144726 1.129084 -0.183568
##      Sylvia Carduelis Motacilla Dendrocopos
## Mass      0.078860 0.283610 0.089786 0.128266
## ChiDist   1.214767 0.636643 0.565053 0.982436
## Inertia   0.116370 0.114951 0.028667 0.123800
## Dim. 1   -0.978853 0.856413 0.429250 -1.311213
```

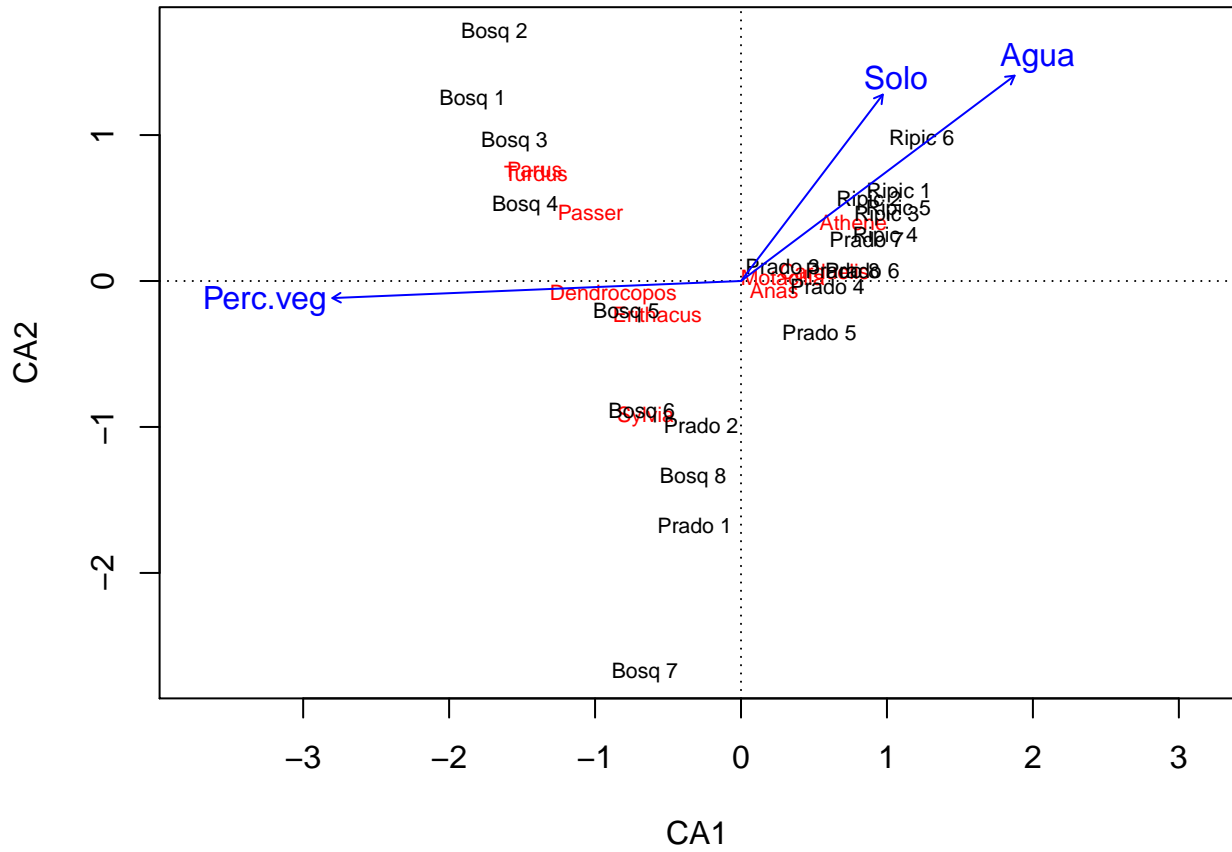
```
## Dim. 2 -2.612553 0.195041 0.065533 -0.255203
```

```
par(mfrow=c(1,2),mar=c(4,4,0.5,0.5))  
plot(ca.aves,labels=2)
```



Note that in a CA we can also overlay on the ordination biplot, if available, a set of environmental covariates. We do this using function `envfit` (to do so we need to fit the CA with function `cca` in package `vegan`)

```
ca.aves2<-cca(aves[, -1])  
ca.aves2env<-envfit(ca.aves2, avesamb[, -1])  
par(mfrow=c(1,1),mar=c(4,4,0.5,0.5))  
plot(ca.aves2)  
plot(ca.aves2env)
```



CCA

```
cca.aves<-cca(aves[, -1] ~ ., data=avesamb[, -1])
summary(cca.aves)
```

```
##
## Call:
## cca(formula = aves[, -1] ~ Perc.veg + Agua + Solo, data = avesamb[, -1])
##
## Partitioning of scaled Chi-square:
##           Inertia Proportion
## Total      0.7634    1.0000
## Constrained 0.3587    0.4699
## Unconstrained 0.4047    0.5301
##
## Eigenvalues, and their contribution to the scaled Chi-square
##
## Importance of components:
##           CCA1    CCA2    CCA3    CA1    CA2    CA3
## Eigenvalue    0.3280 0.02924 0.001425 0.1688 0.08045 0.05299
## Proportion Explained 0.4297 0.03830 0.001866 0.2211 0.10539 0.06941
## Cumulative Proportion 0.4297 0.46802 0.469886 0.6910 0.79640 0.86581
##           CA4    CA5    CA6    CA7    CA8    CA9
## Eigenvalue    0.03458 0.03081 0.01695 0.01137 0.007464 0.001256
## Proportion Explained 0.04530 0.04037 0.02221 0.01489 0.009778 0.001646
```

```

## Cumulative Proportion 0.91111 0.95148 0.97368 0.98858 0.998354 1.000000
##
## Accumulated constrained eigenvalues
## Importance of components:
##           CCA1    CCA2    CCA3
## Eigenvalue      0.3280 0.02924 0.001425
## Proportion Explained 0.9145 0.08152 0.003972
## Cumulative Proportion 0.9145 0.99603 1.000000
##
## Scaling 2 for species and site scores
## * Species are scaled proportional to eigenvalues
## * Sites are unscaled: weighted dispersion equal on all dimensions
##
## Species scores
##
##           CCA1    CCA2    CCA3    CA1    CA2    CA3
## Erithacus  -0.7768 -0.19685 -0.065959 0.18436 -0.077779 0.24382
## Turdus      -0.9648 -0.23347 -0.002096 -1.24513 0.026119 0.10185
## Passer      -0.6850 0.02306 0.113574 -0.95215 -0.142186 -0.92793
## Parus       -1.0307 -0.21659 -0.029620 -1.20241 0.033001 0.55886
## Athene      0.7737 -0.22104 -0.123545 0.06307 -0.227752 0.37841
## Anas        0.1512 0.24331 -0.011338 0.13870 -0.189073 -0.14366
## Sylvia      -0.7323 0.08601 -0.011922 0.35166 0.873994 0.09333
## Carduelis   0.5073 -0.13729 0.031330 0.23570 -0.106902 0.06677
## Motacilla   0.2809 0.06105 0.017772 0.07177 0.002034 0.18827
## Dendrocopos -0.7739 -0.03457 0.011154 -0.32177 0.184739 -0.35045
##
##
## Site scores (weighted averages of species scores)
##
##           CCA1    CCA2    CCA3    CA1    CA2    CA3
## Bosq 1  -1.9920 -3.130118 -8.8184 -1.69906 0.3412784 3.09366
## Bosq 2  -1.7911 -1.888582 2.5962 -1.45636 -1.6684158 -0.61482
## Bosq 3  -1.6737 -1.516559 4.0742 -1.83093 0.4307769 -0.96136
## Bosq 4  -1.6444 -0.864148 -0.9645 -1.65590 0.8319174 -0.71830
## Bosq 5  -0.9269 1.788843 1.0347 -0.36205 -0.0350761 -1.66804
## Bosq 6  -0.9199 -0.002147 -1.9069 1.26653 -0.7150326 -0.64443
## Bosq 7  -0.9902 1.794148 -1.2132 1.99500 1.8709653 0.41297
## Bosq 8  -0.4996 0.098511 4.2427 1.85373 -0.1426594 0.08602
## Prado 1 -0.4851 0.884286 3.5376 0.06650 2.6371451 0.02591
## Prado 2 -0.4075 1.058940 4.9426 -0.23050 1.3348065 -1.39378
## Prado 3 0.3345 0.626284 -3.9496 -0.13035 -0.1440932 1.23049
## Prado 4 0.6681 2.554465 2.2132 0.55476 -1.3515627 -0.33915
## Prado 5 0.5626 1.356450 -1.4583 0.46104 -0.3868010 0.06131
## Prado 6 0.9488 2.468812 2.6743 0.36616 -1.0540694 -0.03454
## Prado 7 0.9990 0.584783 -0.4481 0.35392 -1.1336281 -0.06426
## Prado 8 0.7921 1.646708 -5.7737 0.27631 -0.8460425 0.70205
## Ripic 1 1.3131 -1.080066 -1.8547 -0.03093 0.0005633 0.05120
## Ripic 2 1.0748 -0.077492 -6.5715 -0.22769 0.1093549 0.01896
## Ripic 3 1.1950 0.085138 -1.0070 0.04955 -0.1143608 -0.24653
## Ripic 4 1.1537 -1.536710 9.1861 0.35968 -0.1316139 0.07247
## Ripic 5 1.2864 -0.961352 2.2643 -0.11797 0.1860379 -0.67368
## Ripic 6 1.5511 -3.599707 -2.6757 0.22163 -0.1628344 1.52354

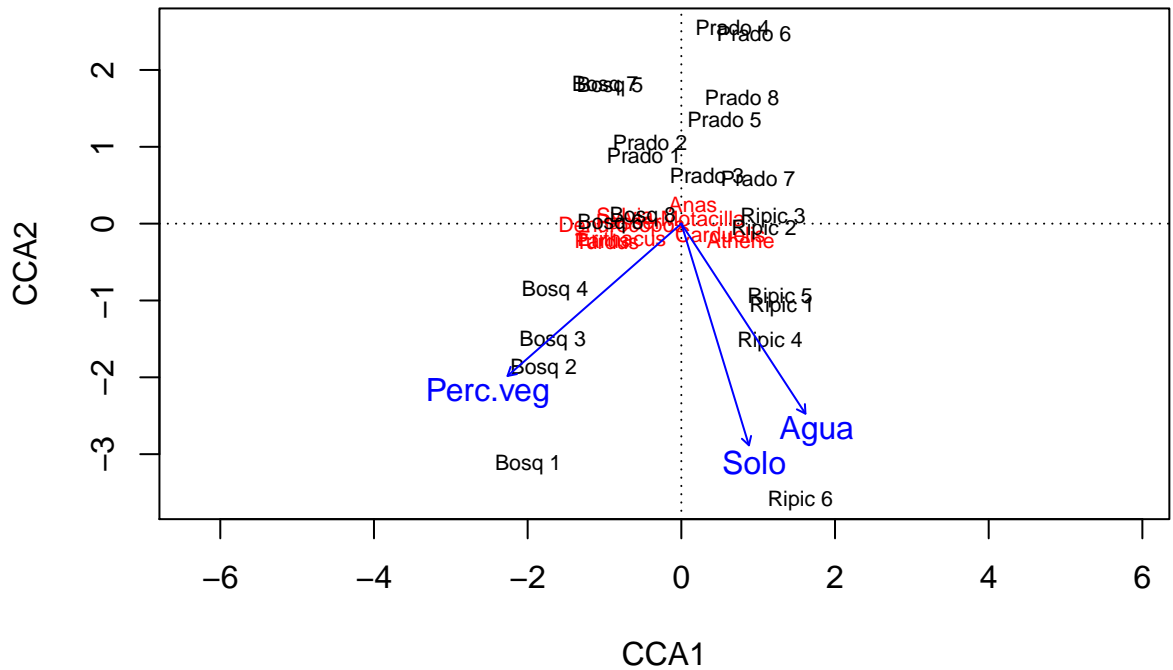
```

```

##
##
## Site constraints (linear combinations of constraining variables)
##
##           CCA1      CCA2      CCA3      CA1      CA2      CA3
## Bosq 1 -1.31648 -0.24662 -0.2960733 -1.69906  0.3412784  3.09366
## Bosq 2 -1.66810 -0.43045 -0.1310699 -1.45636 -1.6684158 -0.61482
## Bosq 3 -0.65583 -0.47859  1.0734735 -1.83093  0.4307769 -0.96136
## Bosq 4 -0.50655 -0.09351 -0.3818036 -1.65590  0.8319174 -0.71830
## Bosq 5 -0.57171 -0.19692 -0.3174287 -0.36205 -0.0350761 -1.66804
## Bosq 6 -1.52015 -0.70390 -0.9236776  1.26653 -0.7150326 -0.64443
## Bosq 7 -1.38010 -0.66702 -0.3701555  1.99500  1.8709653  0.41297
## Bosq 8 -1.51834 -0.80233  0.8309498  1.85373 -0.1426594  0.08602
## Prado 1  0.23409  1.44511  0.1563554  0.06650  2.6371451  0.02591
## Prado 2  0.26899  1.37955  0.3271438 -0.23050  1.3348065 -1.39378
## Prado 3  0.24667  1.37789 -0.2362737 -0.13035 -0.1440932  1.23049
## Prado 4 -0.04446  1.29317 -0.8153716  0.55476 -1.3515627 -0.33915
## Prado 5  0.28423  1.09568  0.6085555  0.46104 -0.3868010  0.06131
## Prado 6  0.47614  1.02432  0.0006532  0.36616 -1.0540694 -0.03454
## Prado 7  0.46977  1.01766  1.2083258  0.35392 -1.1336281 -0.06426
## Prado 8  0.36439  1.41311 -0.2461690  0.27631 -0.8460425  0.70205
## Ripic 1  1.32532 -1.11375  0.3241837 -0.03093  0.0005633  0.05120
## Ripic 2  1.23941 -1.05818  0.2342329 -0.22769  0.1093549  0.01896
## Ripic 3  1.21733 -1.31891  1.0163511  0.04955 -0.1143608 -0.24653
## Ripic 4  1.03051 -1.18296  1.9005736  0.35968 -0.1316139  0.07247
## Ripic 5  1.48182 -0.84024 -3.4193528 -0.11797  0.1860379 -0.67368
## Ripic 6  1.03817 -0.99541 -0.5173174  0.22163 -0.1628344  1.52354
##
##
## Biplot scores for constraining variables
##
##           CCA1      CCA2      CCA3 CA1 CA2 CA3
## Perc.veg -0.7484 -0.6559 -0.09883  0  0  0
## Agua      0.5344 -0.8183 -0.21145  0  0  0
## Solo      0.2911 -0.9544  0.06580  0  0  0

```

```
plot(cca.aves)
```



We conclude that with the direct gradient analysis (i.e. the CCA) the same separation across habitats is present, and we can relate the environmental variables to the abundances. Naturally, the percentage of vegetation is larger on Bosques than on the Pradaria or Ripicola areas, and there is more water near rivers (Ripicola). It is however not easy to see in the CCA output what are the species contributing more to this segregation. That is simpler to do in the CA/PKA plots, where it seems clear that *Turdus* and *Parus* are characteristic of Bosques, as is *Sylvia* (segregating sites on the second component), and *Athene* of Ripicola. Prados are harder to characterize in terms of species in the CA, but the PCA shows that *Anas* seems frequent.