

MULTIVARIATE STATISTICS

Camila Pacheco Katrín Björnsdóttir

TODAY

- Explore how **ordination methods can generate hypotheses** about ecological gradients and species interactions
- Use **R** to perform Unconstrained ordinations
- Learn the following methods:
 - Clustering analysis
 - Detrended correspondence analysis (DCA)
 - Principal Component Analysis (PCA)
 - Non-metric Multidimensional Scaling (NMDS)
- Break 😊
- Practice

```
header(); ?>
```

```
<div class="container">  
  <h1 class="text-primary text-prima  
  info"></i> <  
  <hr>  
  <div class="row">  
    <?php
```



REQUIRED MATERIAL

<https://github.com/lacapary/BIO503/>

REQUIRED MATERIAL

You are required to have downloaded and installed

```
1 install.packages(c("vegan",  
2                      "ape",  
3                      "factoextra",  
4                      "dendextend"))
```

REQUIRED MATERIAL

Do not hesitate to ask questions!

THE DATA FOR THIS SESSION?

The data originates from the research of Batterink & Wijffels (1983), published as a report in Dutch.

Table 0.1. Dune Meadow Data. Unordered table that contains 20 relevées (columns) and 30 species (rows). The right-hand column gives the abbreviation of the species names listed in the left-hand column; these abbreviations will be used throughout the book in other tables and figures. The species scores are according to the scale of van der Maarel (1979b).

	00000000011111111112		
	12345678901234567890		
1	Achillea millefolium	13..222..4.....2...	Ach mil
2	Agrostis stolonifera	.48..43..45447...5	Agr sto
3	Aira praecox2.3.	Air pra
4	Alopecurus geniculatus	.272..53..85..4....	Alo gen
5	Anthoxanthum odoratum432..4.....4.4.	Ant odo
6	Bellis perennis	.3222....2.....2..	Bel per
7	Bromus hordaceus	.4..32.2..4.....	Bro hor
8	Chenopodium album1.....	Che alb
9	Cirsium arvense	..2.....	Cir arv
10	Eleocharis palustris4.....458..4	Ele pal
11	Elymus repens	44444..6.....	Ely rep
12	Empetrum nigrum2.	Emp nig
13	Hypochaeris radicata2....2.5.	Hyp rad
14	Juncus articulatus44..33..4	Jun art
15	Juncus bufonius2.4..43.....	Jun buf
16	Leontodon autumnalis	.5223332352222.2562	Leo aut
17	Lolium perenne	75652664267....2..	Lol per
18	Plantago lanceolata555..33.....23..	Pla Lan
19	Poa pratensis	44542344444.2...13..	Poa pra
20	Poa trivialis	2765645454.49..2....	Poa tri
21	Potentilla palustris22.....	Pot pal
22	Ranunculus flammula2....2222...4	Ran fla
23	Rumex acetosa563.2..2.....	Rum ace
24	Sagina procumbens	..5..22.242.....3.	Sag pro
25	Salix repens335	Sal rep
26	Trifolium pratense252.....	Tri pra
27	Trifolium repens	.52125223633261..22.	Tri rep
28	Vicia lathyroides12.....1..	Vic lat
29	Brachythecium rutabulum	..2226222244..44.634	Bra rut
30	Calliergonella cuspidata4.3..3	Cal cus

THE DATA FOR THIS SESSION?

```
1 library(tidyverse)
2 library(vegan)
3
4 # Load the community dataset which we'll use in the exam
5
6 dune2_spe <- read_csv("Data/dune2_spe.csv")
7 dune2_env <- read_csv("Data/dune2_env.csv")
8
9 # Open the dataset and look if you can find any patterns
10 head(dune2_spe)
```

```
# A tibble: 6 × 28
  Achimill Agrostol Airaprae Alopogeni Anthodor Bellpere Bromhord Chenalbu
  <dbl>     <dbl>    <dbl>     <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1     1         0         0         0         0         0         0         0
2     3         0         0         2         0         3         4         0
3     0         4         0         7         0         2         0         0
4     0         8         0         2         0         2         3         0
5     2         0         0         0         4         2         2         0
6     2         0         0         0         3         0         0         0
# i 20 more variables: Cirsarve <dbl>, Comapalu <dbl>, Eleopalu <dbl>,
# Elymrepe <dbl>, Empenigr <dbl>, Hyporadi <dbl>, Juncarti <dbl>,
# Juncbufo <dbl>, Lolipere <dbl>, Planlanc <dbl>, Poaprat <dbl>,
# Poatriv <dbl>, Ranuflam <dbl>, Rumeacet <dbl>, Sagiproc <dbl>,
# Salirepe <dbl>, Scorautu <dbl>, Trifprat <dbl>, Trifrepe <dbl>,
# Vicilath <dbl>
```

1 head(dune2_env)

```
# A tibble: 6 × 5
  A1 Moisture Management Use      Manure
  <dbl>    <dbl> <chr>     <chr>    <dbl>
1 2.8      1 SF       Haypastu   4
2 3.5      1 BF       Haypastu   2
3 4.3      2 SF       Haypastu   4
4 4.2      2 SF       Haypastu   4
5 6.3      1 HF       Hayfield   2
6 4.3      1 HF       Haypastu   2
```

RECAP: LINEAR MODELS

- We learned some models to study at ecological data.
- These models allowed us to ask questions such as:
 - What are the effects of precipitation and temperature on species richness? or
 - How does the abundance of species change between habitats?

WHY MULTIVARIATE? 🌱

Let's compare **univariate** vs **multivariate** approaches.

UNIVARIATE EXAMPLE

You study how **soil pH** affects the **abundance of one species** (*Species A*):

```
1 # Univariate linear regression  
2 model <- lm(species_A_abundance ~ soil_pH, data = your_d  
3 summary(model)
```

This tells you:

👉 *How soil pH affects the abundance of one species.*

WHY THAT'S LIMITING IN ECOLOGY

In real ecosystems:

- You usually have **many species** per site.
- Environmental variables may affect **multiple species simultaneously**.
- Species may interact (e.g., competition, facilitation).

MOVING TO MULTIVARIATE

“What if I want to know how **soil pH, moisture, and light** influence the **whole plant community**? ”

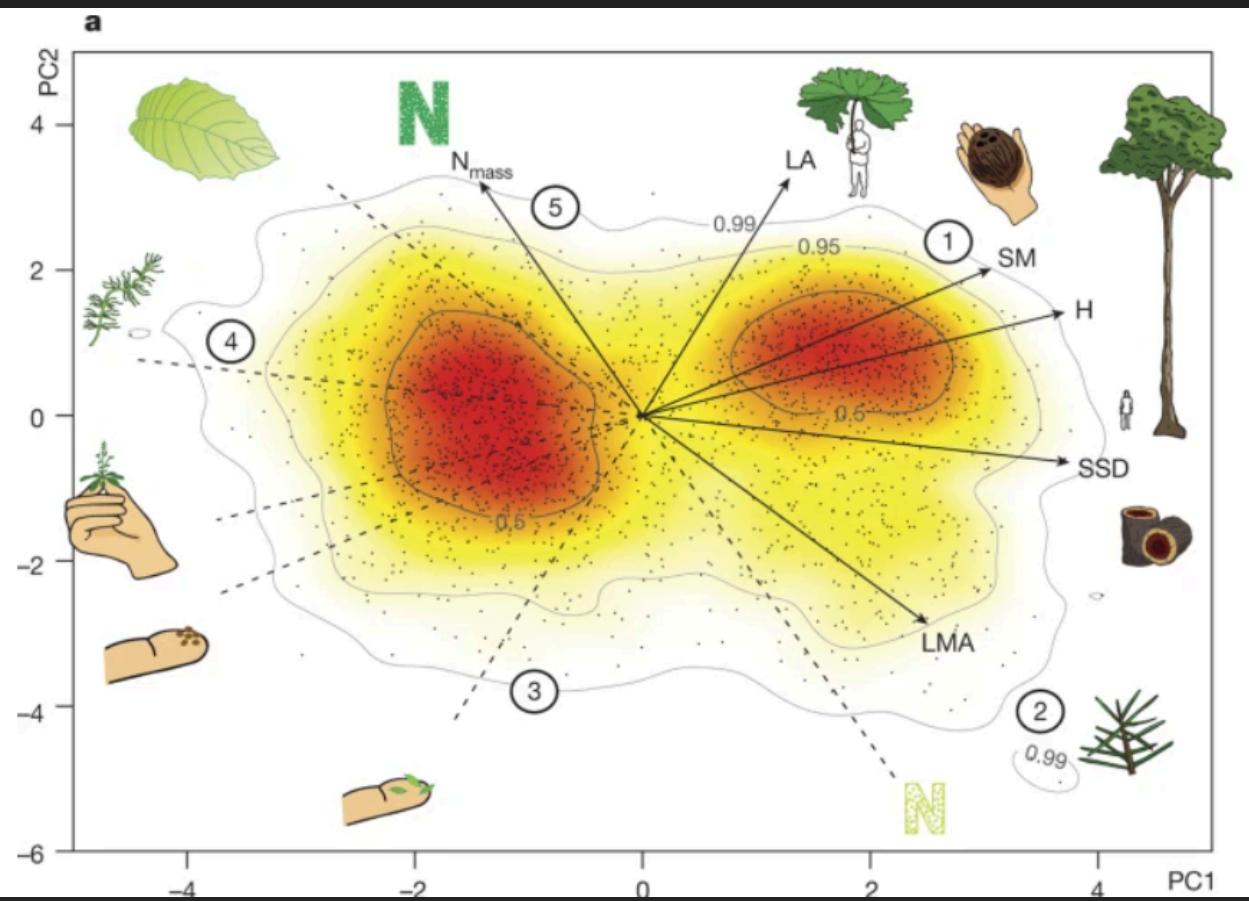
That’s where **ordination methods (e.g., PCA, NMDS)** and **clustering** shine. They summarize patterns across multiple species simultaneously.

EXAMPLES OF QUESTIONS IN MULTIVARIATE STATISTICS

Our research question might be:

- What are the global patterns of how vascular plants grow and function?

Remember this:



EXAMPLES OF QUESTIONS IN MULTIVARIATE STATISTICS

Our research question might be:

- How does the plants composition change along an elevation gradient?
- What is the composition dissimilarity of plants communities?
- How closely-related are local vegetation communities in terms of their composition ?

HOW DOES THE DATA LOOK?

Matrix Species

Site	Species 1	Species 2	...	Species n
1	abundance 1	abundance 2	...	abundance n
2	abundance 1	abundance 2	...	abundance n
m

Matrix locations

Site	Temperature	Precipitation	...	Driver n
1	Temperature 1	Precipitation 2	...	Driver n
2	Temperature 1	Precipitation 2	...	Driver n
m

MULTIVARIATE METHODS

- *Ordination* and *clustering* are the two main classes of multivariate methods that community ecologists employ.
- To some degree, these two approaches are *complementary*.
- Hierarchical data clustering allows you to explore your data and look for discontinuities (e.g. gaps in your data), gradients and meaningful ecological units (e.g. groups or subgroups of species).
- Given the continuous nature of communities, ordination can be considered a more natural approach.

CLUSTERING

- Hierarchical clustering offers insight into how your biodiversity data are organized and can help you to disentangle different patterns and the scales at which they can be observed.
- Its results can be represented as dendograms (tree-like diagrams), which describe how closely observations are.



source: Prasad Pai

CLUSTERING



source: Prasad Pai



source: Prasad Pai

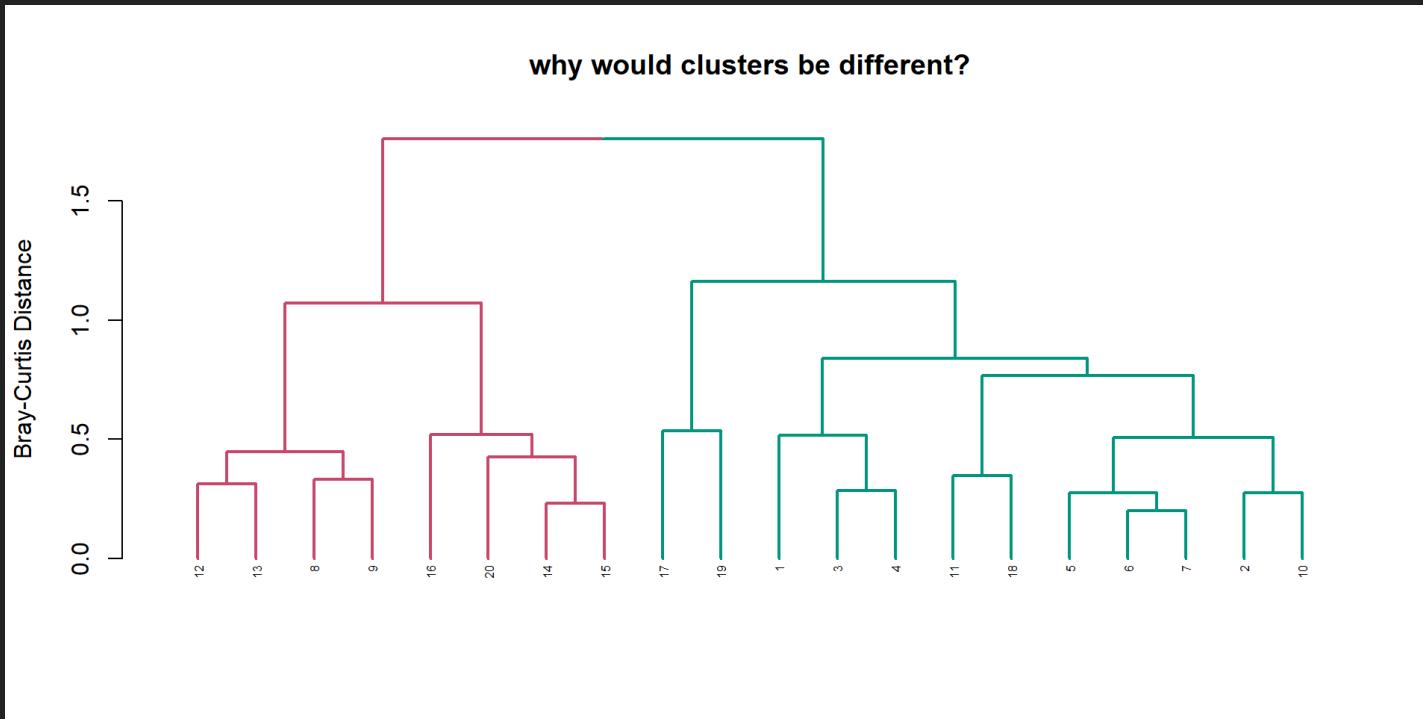
CLUSTERING

```
1 library(dendextend)
2
3
4 dis_data<-dune2_spe %>%
5   vegdist(method = "bray", upper=FALSE)
6
7 dend <- dis_data %>%
8   hclust(method="ward.D2") %>%
9   as.dendrogram()
10
11 dend
```

'dendrogram' with 2 branches and 20 members total, at height 1.757535

CLUSTERING

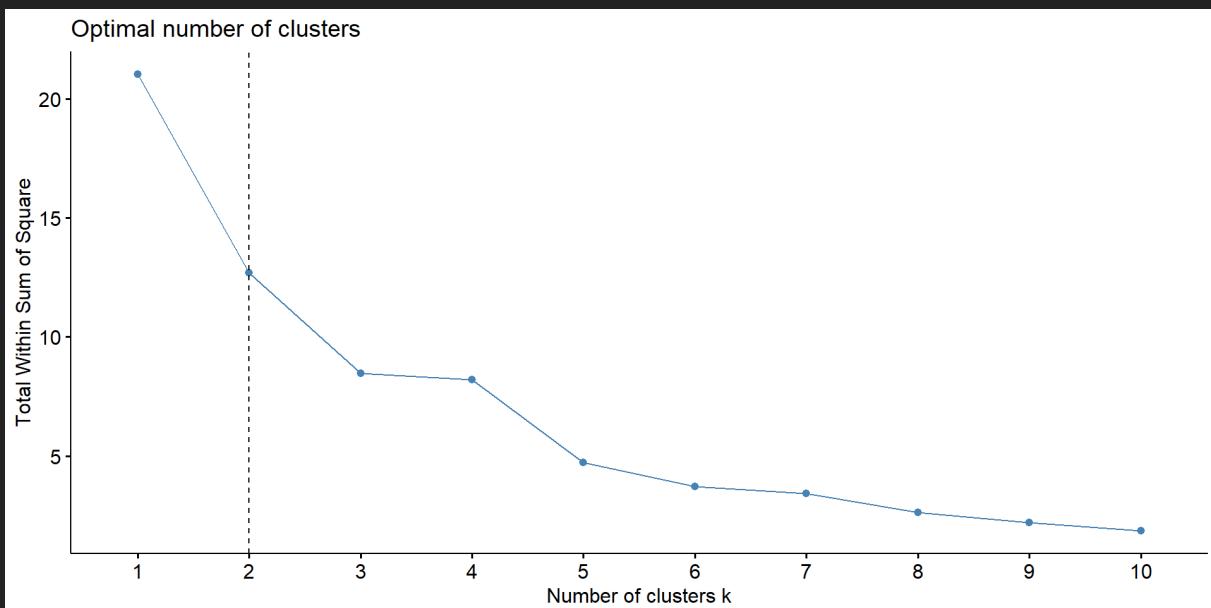
```
1 dend.plot <- dend %>%
2   set("branches_lwd", 2) %>% # Branches line width
3   set("branches_k_color", k = 2) %>% # Color branches b
4   set("labels_cex", 0.5) # Change label size
5
6 plot(dend.plot, ylab = "Bray-Curtis Distance", main = "w")
```



CLUSTERING

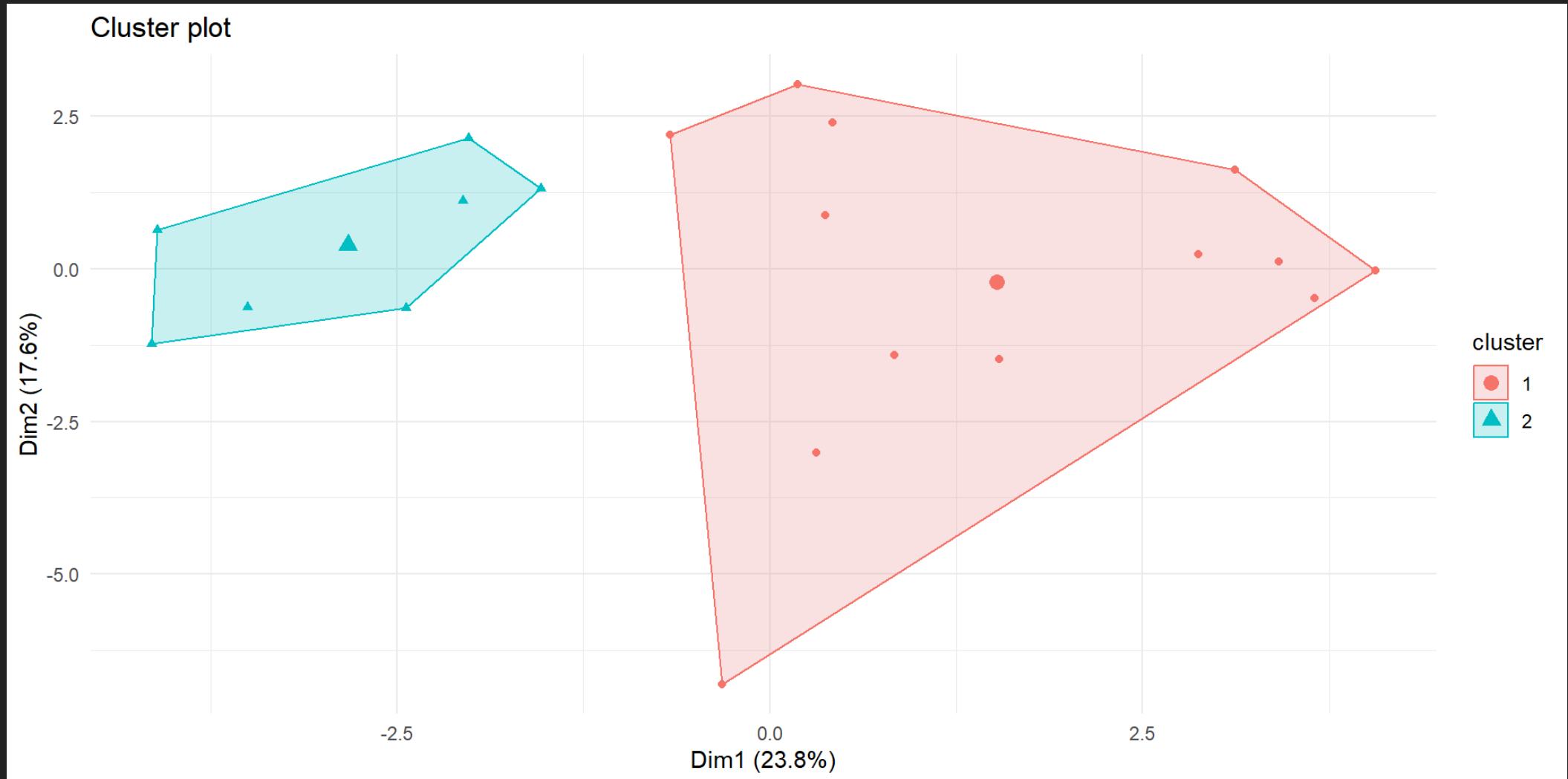
What is the ideal number of clusters?

```
1 library(factoextra)
2
3 varespec_m<-dis_data |> as.matrix()
4
5 fviz_nbclust(varespec_m, kmeans, method = "wss") +
6   geom_vline(xintercept = 2, linetype = 2)
```



CLUSTERING

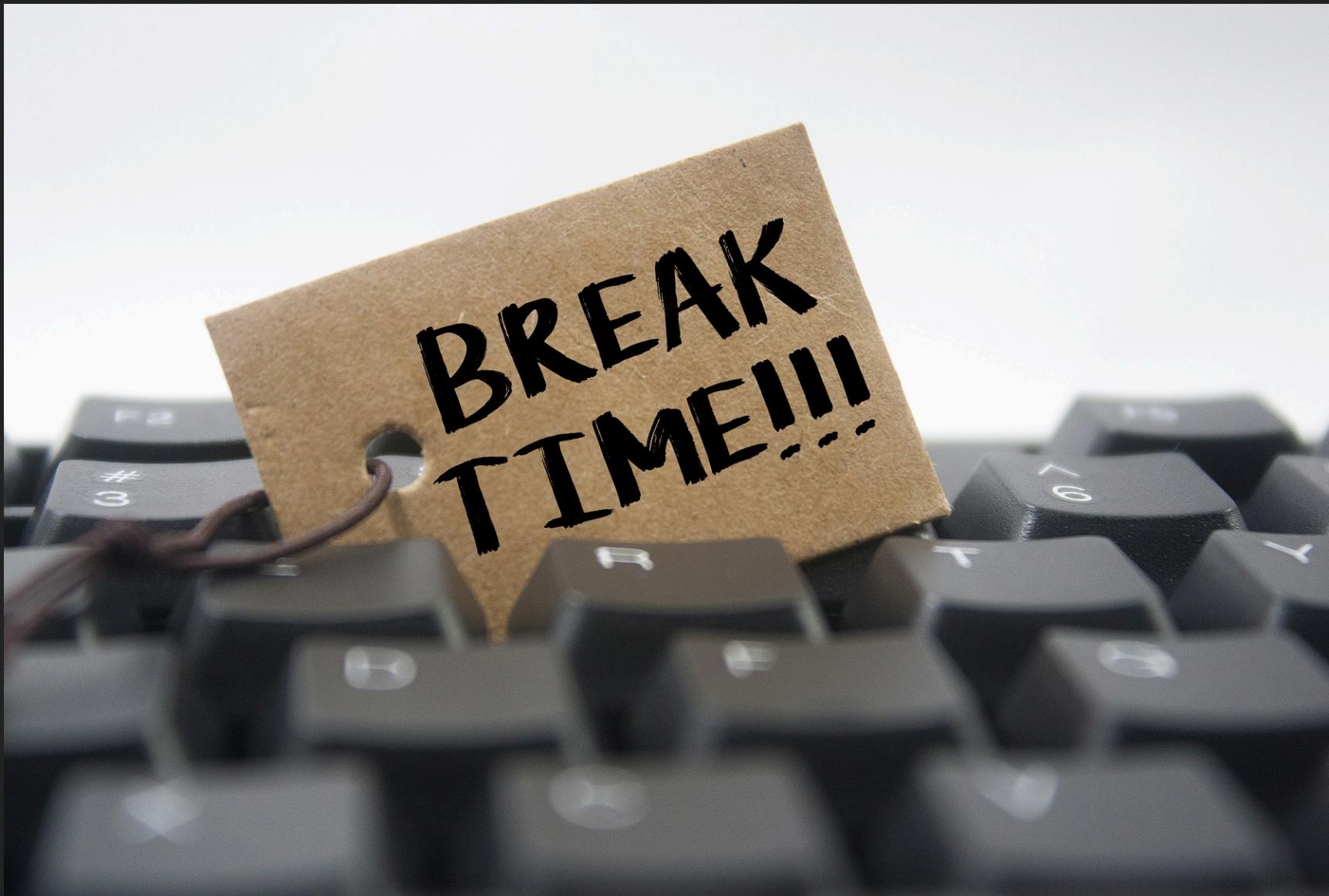
```
1 fviz_cluster(kmeans(varespec_m, centers = 2), geom = "po
```



FROM CLUSTERS TO GRADIENTS

- Clustering tells us **who is similar**
- Ordinations help explain **why they are similar**
- Gradients (like pH, moisture, disturbance) may underlie these groupings

BREAK

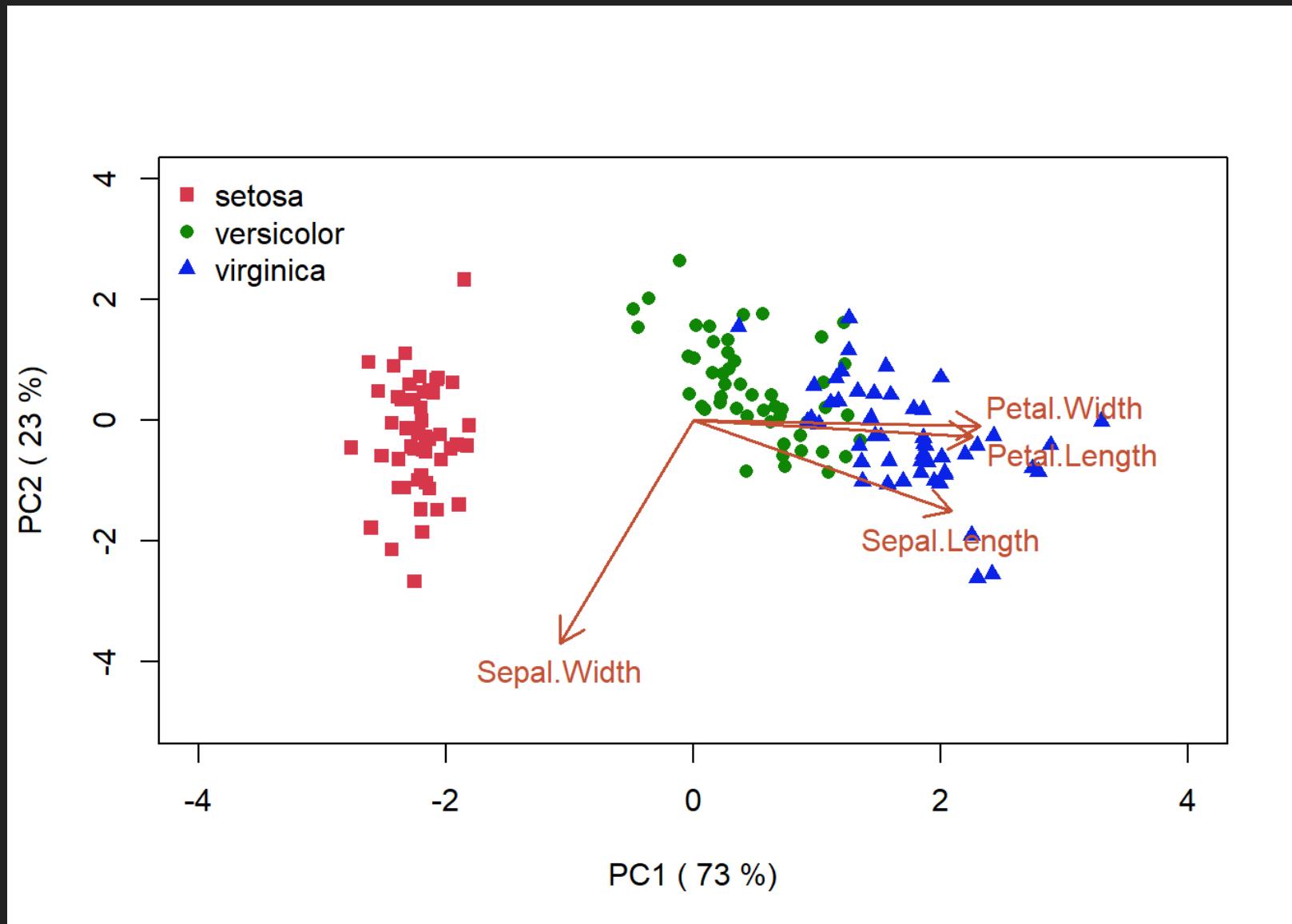


ORDINATIONS

Ordination is a collective term for multivariate techniques which summarize a multidimensional dataset in such a way that when it is projected onto a low dimensional space, any intrinsic pattern the data may possess becomes apparent upon visual inspection (Pielou, 1984).

- In ecological terms, ordination helps us understand community data.
- Find key components of variation among samples, sites, species
- Reduce the number of dimensions in multivariate data while limiting substantial loss of information.
- Create new variables for use in subsequent analyses.

WHAT IS ORDINATION?



DATA

Matrix algebra

3	6	4	2	0	2	8	1	2	5	0	1	7	8	5	2
4	3	2	1	4	7	6	5	3	5	2	3	6	7	1	4
7	0	1	8	6	7	3	4	0	1	6	3	4	6	7	3
3	5	2	8	4	4	0	0	7	3	1	1	6	0	1	3
0	1	7	4	6	3	6	0	5	3	6	2	0	3	2	7
8	2	1	5	1	4	3	1	6	5	1	3	0	4	1	0
2	7	1	1	3	4	8	0	8	0	0	8	1	0	3	6
0	1	2	7	0	4	1	1	0	6	0	8	5	5	0	6
0	6	0	1	0	4	1	8	3	8	3	7	0	0	5	1
2	5	4	7	3	4	3	6	4	8	0	6	2	2	8	4
1	2	5	4	2	2	4	6	2	7	3	8	0	5	3	1
6	7	1	3	0	7	4	2	1	7	4	4	3	0	4	3
0	2	4	0	8	1	6	7	6	0	8	1	5	4	7	0
5	0	5	2	7	3	2	6	4	4	7	2	5	1	0	5
7	6	7	8	6	2	8	2	5	1	1	6	8	0	2	3
5	7	8	4	5	1	3	8	1	1	7	5	8	5	5	0
5	1	1	6	4	1	4	2	5	2	3	2	0	6	1	8
3	7	7	5	8	5	4	5	8	7	3	2	4	3	1	2
3	3	8	1	4	8	8	5	4	1	1	0	6	3	4	2
5	2	3	7	8	7	0	6	2	4	5	0	7	6	7	1
0	7	5	6	0	6	6	3	8	6	7	2	0	8	5	2
1	3	7	6	4	3	7	5	8	7	5	2	0	6	6	2
6	4	2	2	1	7	1	7	8	3	4	8	8	5	4	1
1	0	3	2	6	0	1	1	5	7	6	7	2	6	0	8
2	5	1	6	8	0	2	0	2	2	7	0	2	8	1	2
7	5	1	7	4	5	7	1	5	7	6	4	8	8	0	6

MULTIVARIATE STATISTICS

Matrix algebra

$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ m \end{matrix} & \left[\begin{matrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{matrix} \right] \end{matrix}$$

TYPE OF ORDINATIONS?

- Unconstrained Ordination: we're basically letting the data *speak for itself*. We don't impose any specific relationships or constraints between the variables.
- Constrained Ordination: we impose some restrictions or *constraints* on the analysis based on what we already know or suspect about the data.
- In simple terms, unconstrained ordination lets the data tell its story without interference, while constrained ordination guides the analysis based on what we already know or suspect.
- We are going to focus in Unconstrained Ordinations tion aims at arranging samples or species continuously along gradients.

DOING AN ORDINATION

This ordination goes in two steps:

- First, we will perform an ordination on a species abundance matrix.
- Then we will use environmental data (samples by environmental variables) to interpret the gradients that were uncovered by the ordination.

DIFFERENT ORDINATION TECHNIQUES

- *P*rincipal *C*omponent *A*nalysis (PCA)
- *D*etrended *C*orrespondence *A*nalysis (DCA)
- *N*on-metric *M*ulti*d*imensional *S*caling (NMDS)
- And *MORE*.....
- Why so many ordination methods?
- Different ordination methods = different assumptions + different strengths.

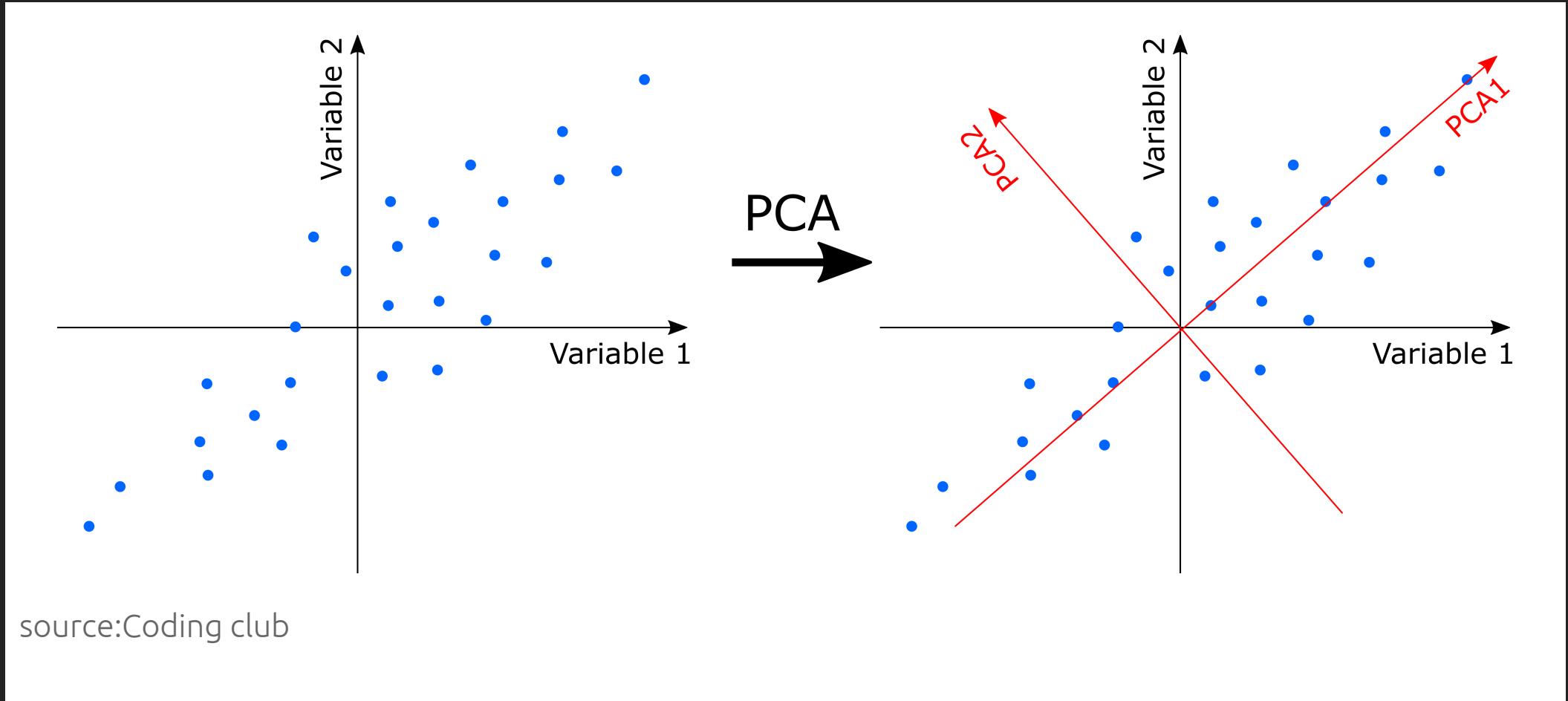
COMPARING ORDINATION METHODS

Method	Type	Best For	Assumes
PCA	Linear	Species respond linearly to gradients	Normality, Euclidean distances
DCA	Unimodal	Species have peak abundances along gradients	Long gradients, curved structure
NMDS	Non-metric	Preserving ranked distances	No assumptions, works with Bray-Curtis
CCA/RDA	Constrained	Link communities to environmental variables	Explaining structure via predictors

PRINCIPAL COMPONENT ANALYSIS (PCA)

- PCA is a linear method that helps us make sense of complex data by reducing its dimensions.
- In a nutshell, the PCA linearly transforms the feature from the original space to a new feature space, containing principal components that explain most of the variance in the dataset.

PRINCIPAL COMPONENT ANALYSIS (PCA)



PRINCIPAL COMPONENT ANALYSIS (PCA)

Euclidean distances among samples

- The axes (also called principal components or PC) are orthogonal to each other (and thus independent).
- Each PC is associated with an eigenvalue.
- The sum of the eigenvalues will equal the sum of the variance of all variables in the data set.
- The eigenvalues represent the variance extracted by each PC, and are often expressed as a percentage of the sum of all eigenvalues (i.e. total variance).

PRINCIPAL COMPONENT ANALYSIS (PCA)

- The relative eigenvalues thus tell how much variation that a PC is able to ‘explain’.
- Axes are ranked by their eigenvalues:
 - the first axis has the highest eigenvalue and thus explains the most variance
 - the second axis has the second highest eigenvalue, etc.

PRINCIPAL COMPONENT ANALYSIS (PCA)

```
1 PCA <- rda(dune2_spe, scale = FALSE) # Use scale = TRUE i
2 # Here, all species are measured on the same scale
3 # So use scale = FALSE
4 PCA
```

Call: rda(X = dune2_spe, scale = FALSE)

-- Model Summary --

	Inertia	Rank
Total	78.97	
Unconstrained	78.97	19

Inertia is variance

-- Eigenvalues --

Eigenvalues for unconstrained axes:

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
24.181	17.678	7.557	6.760	4.274	4.009	2.835	2.584

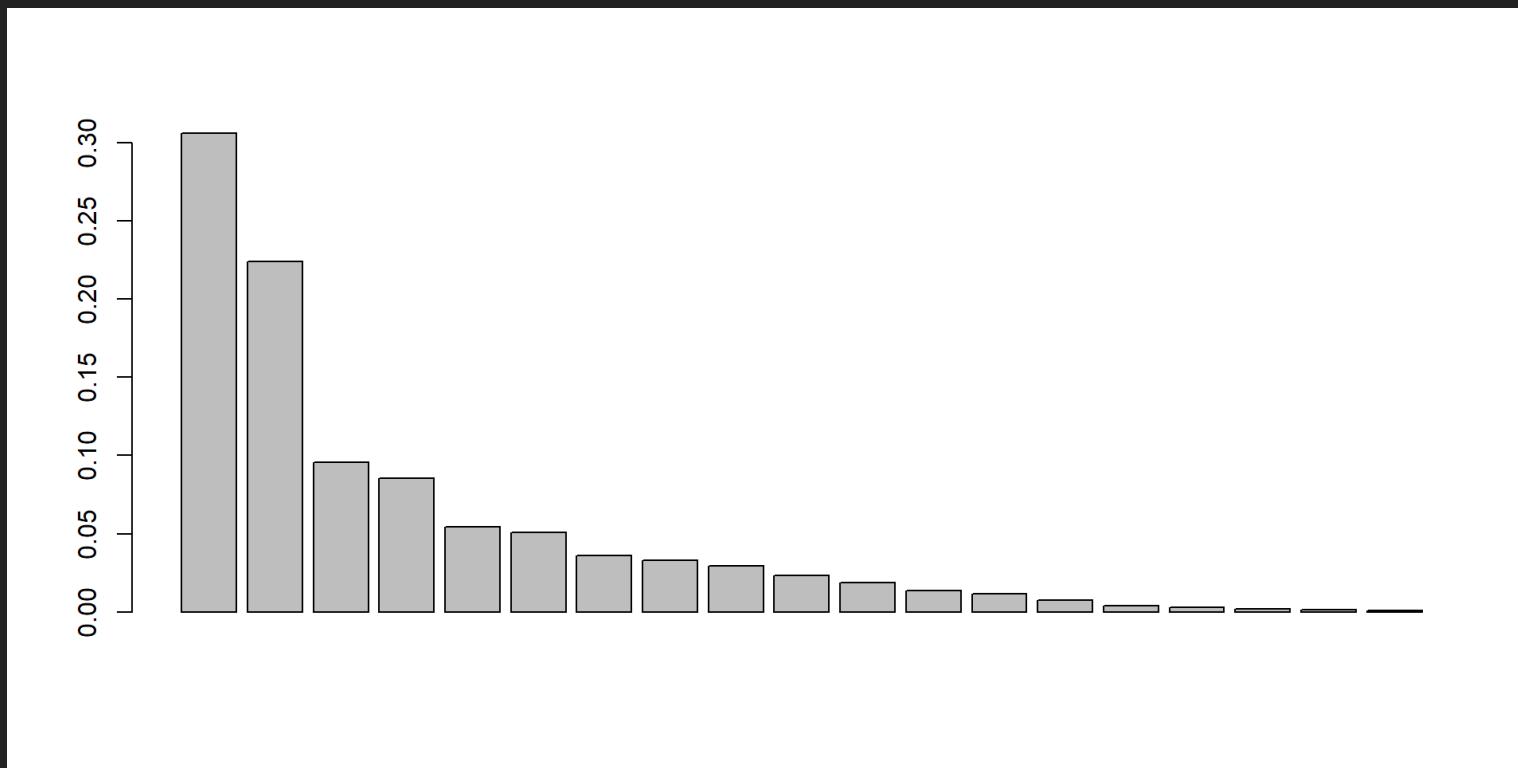
(Showing 8 of 19 unconstrained eigenvalues)

PRINCIPAL COMPONENT ANALYSIS (PCA)

```
1 # Now plot a bar plot of relative eigenvalues. This is t  
2 barplot(as.vector(PCA$CA$eig) / sum(PCA$CA$eig))
```

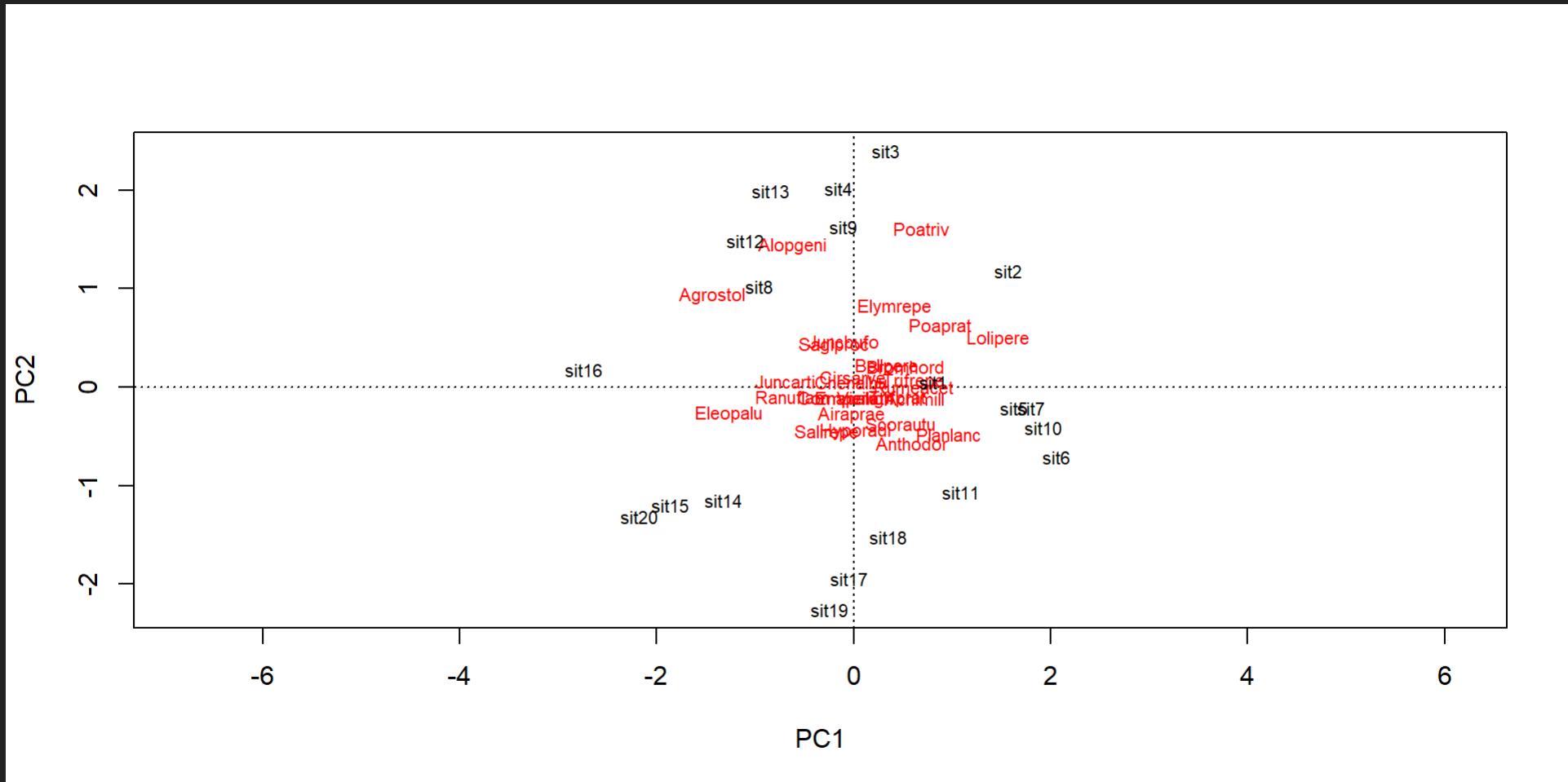
```
1 # Calculate the percent of variance explained by first t  
2 sum((as.vector(PCA$CA$eig) / sum(PCA$CA$eig))[1:2]) # 53%,
```

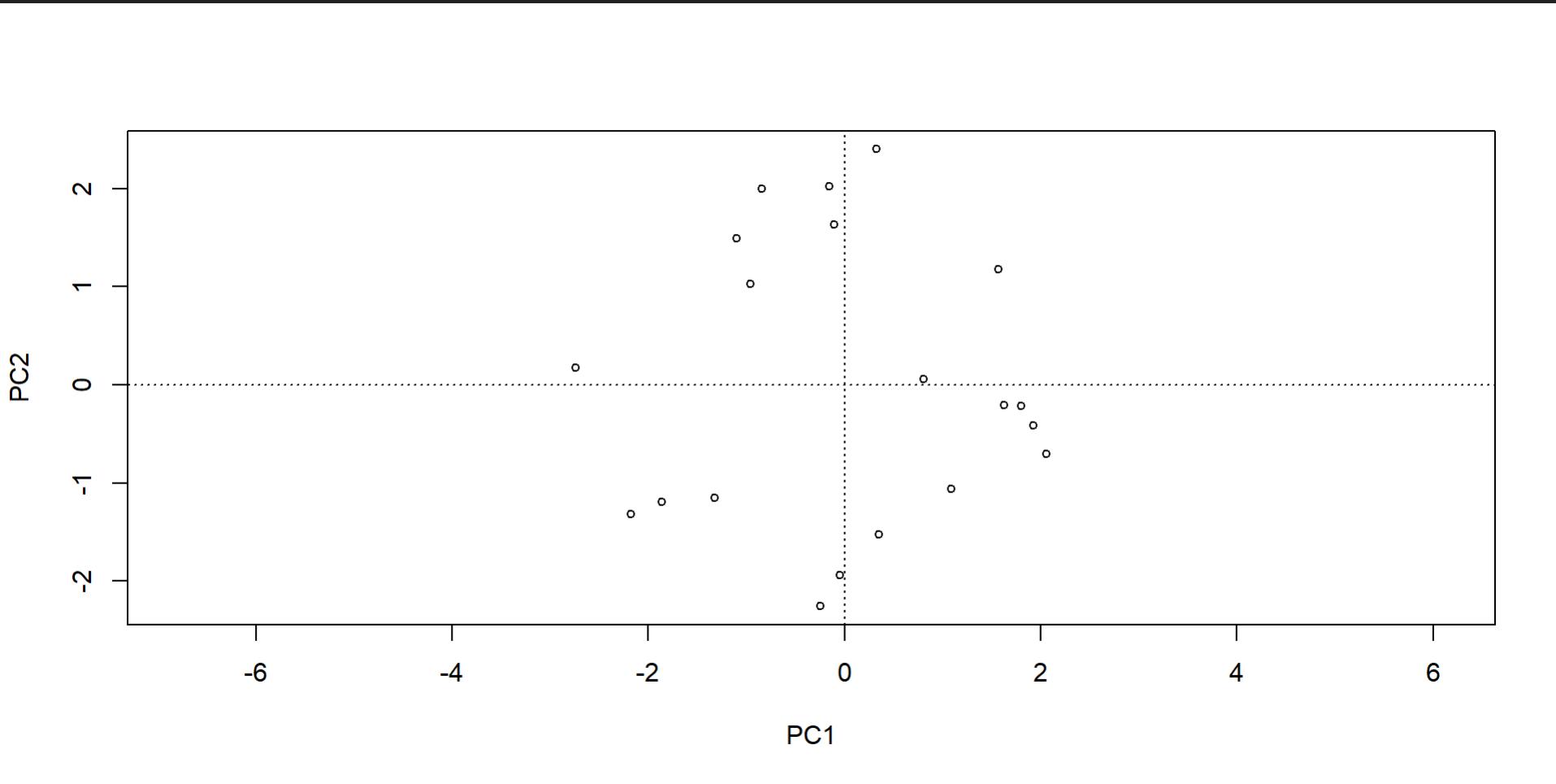
```
[1] 0.5300765
```



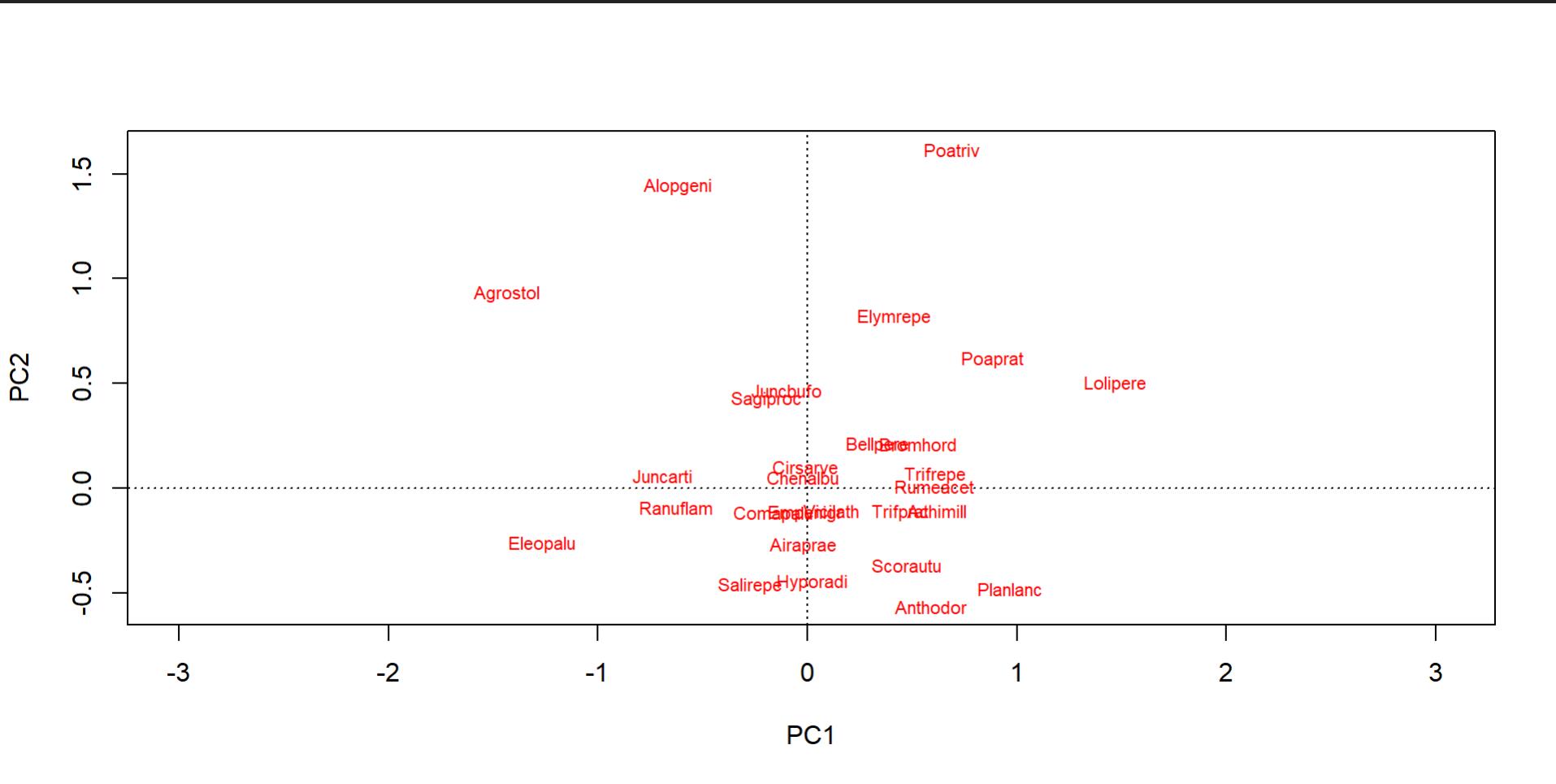
PRINCIPAL COMPONENT ANALYSIS (PCA)

```
1 plot(PCA)
```



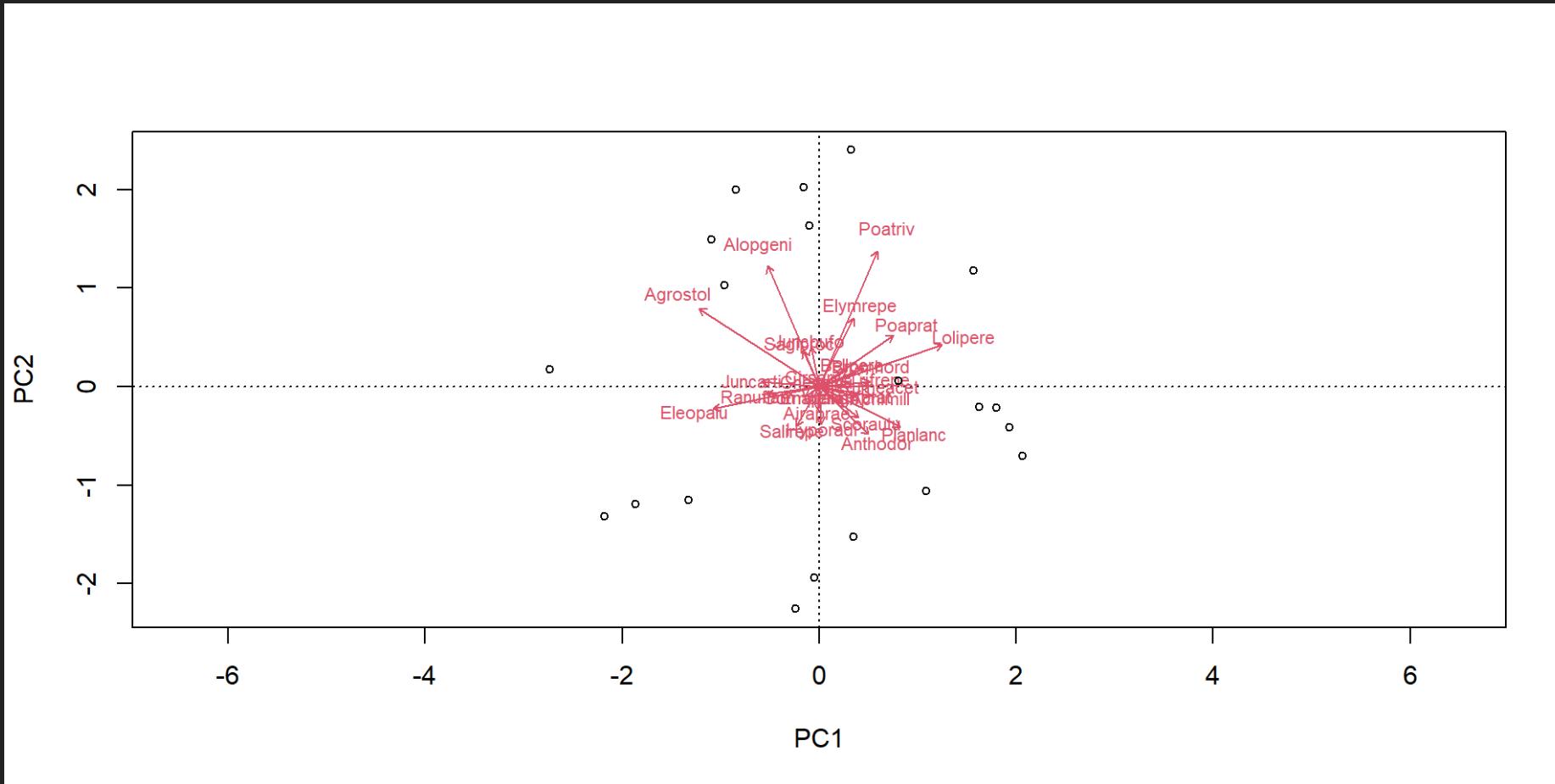


```
1 plot(PCA, display = "species", type = "text")
```



PRINCIPAL COMPONENT ANALYSIS (PCA)

```
1 # In a biplot of a PCA, species' scores are drawn as arr  
2 # that point in the direction of increasing values for t  
3 biplot(PCA, choices = c(1,2), type = c("text", "points"))
```



PRINCIPAL COMPONENT ANALYSIS (PCA)

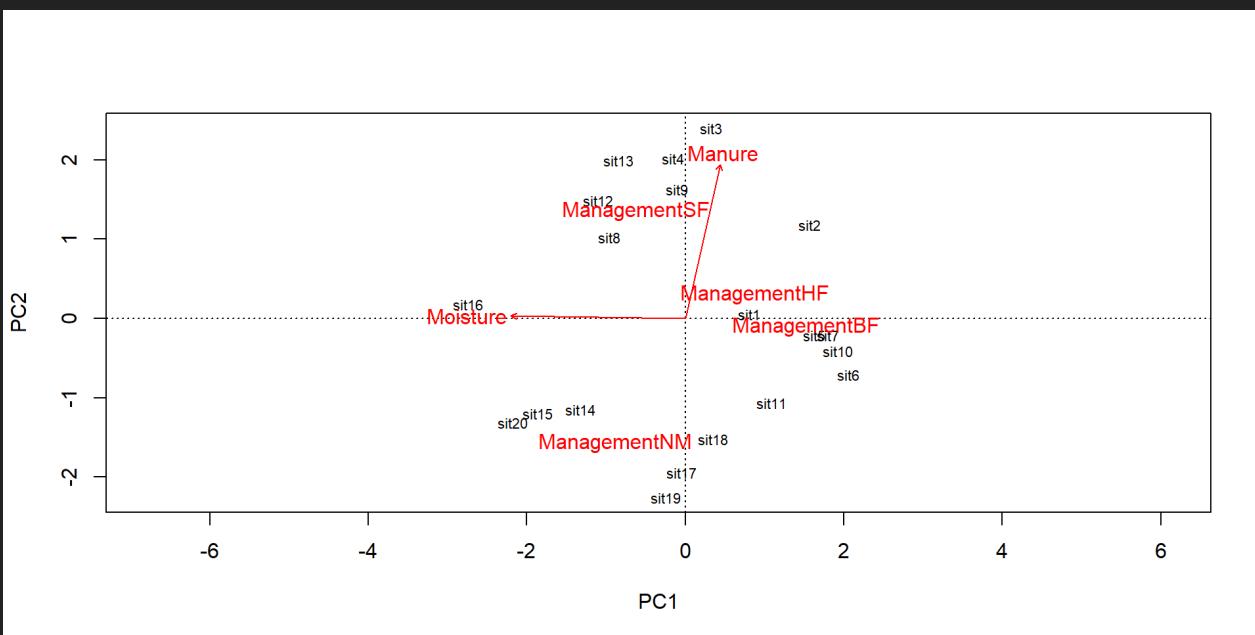
- This implies that the abundance of the species is continuously increasing in the direction of the arrow, and decreasing in the opposite direction.
- Thus PCA is a linear method.
- PCA is extremely useful when we expect species to be linearly (or even monotonically) related to each other.
- Unfortunately, we rarely encounter such a situation in nature.

ENVIRONMETAL VARIABLES AND TRIPLOT

```
1 fit <- envfit(PCA, dune2_env, perm = 999)  
2 scores(fit, "vectors")
```

	PC1	PC2
A1	-0.4981858	-0.11968978
Moisture	-0.8609679	0.01183937
Manure	0.1718683	0.75960897

```
1 plot(PCA, dis="site")  
2 plot(fit, p.max = 0.05, col = "red")
```



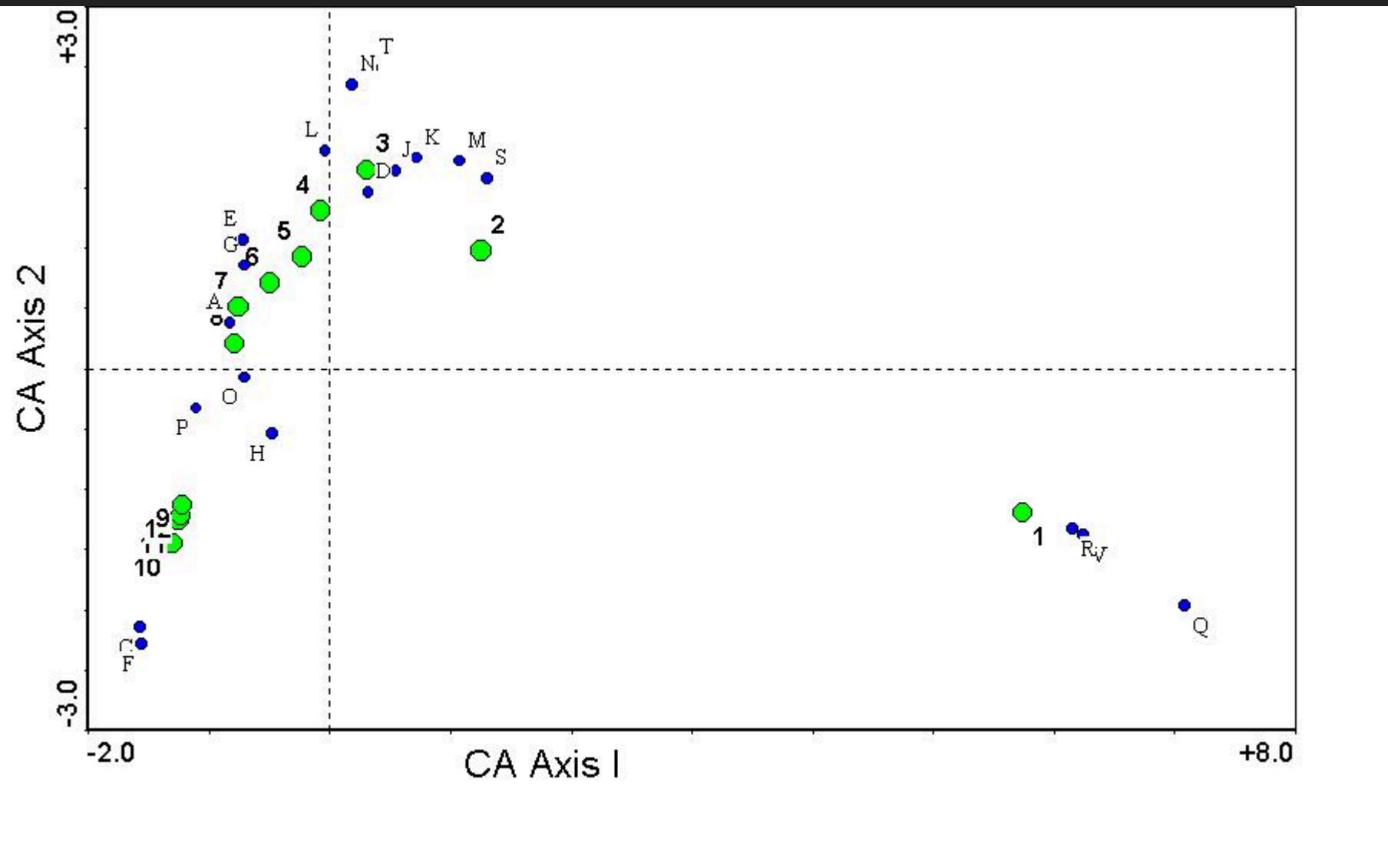
DETRENDED CORRESPONDENCE ANALYSIS (DCA)

chi-square distance metric among samples

DCA is an ordination method built on *Correspondence Analysis* (CA), which arranges samples and species in the same space to explore their relationships.

It is especially useful when species show unimodal responses along environmental gradients (i.e., they peak at some optimal conditions).

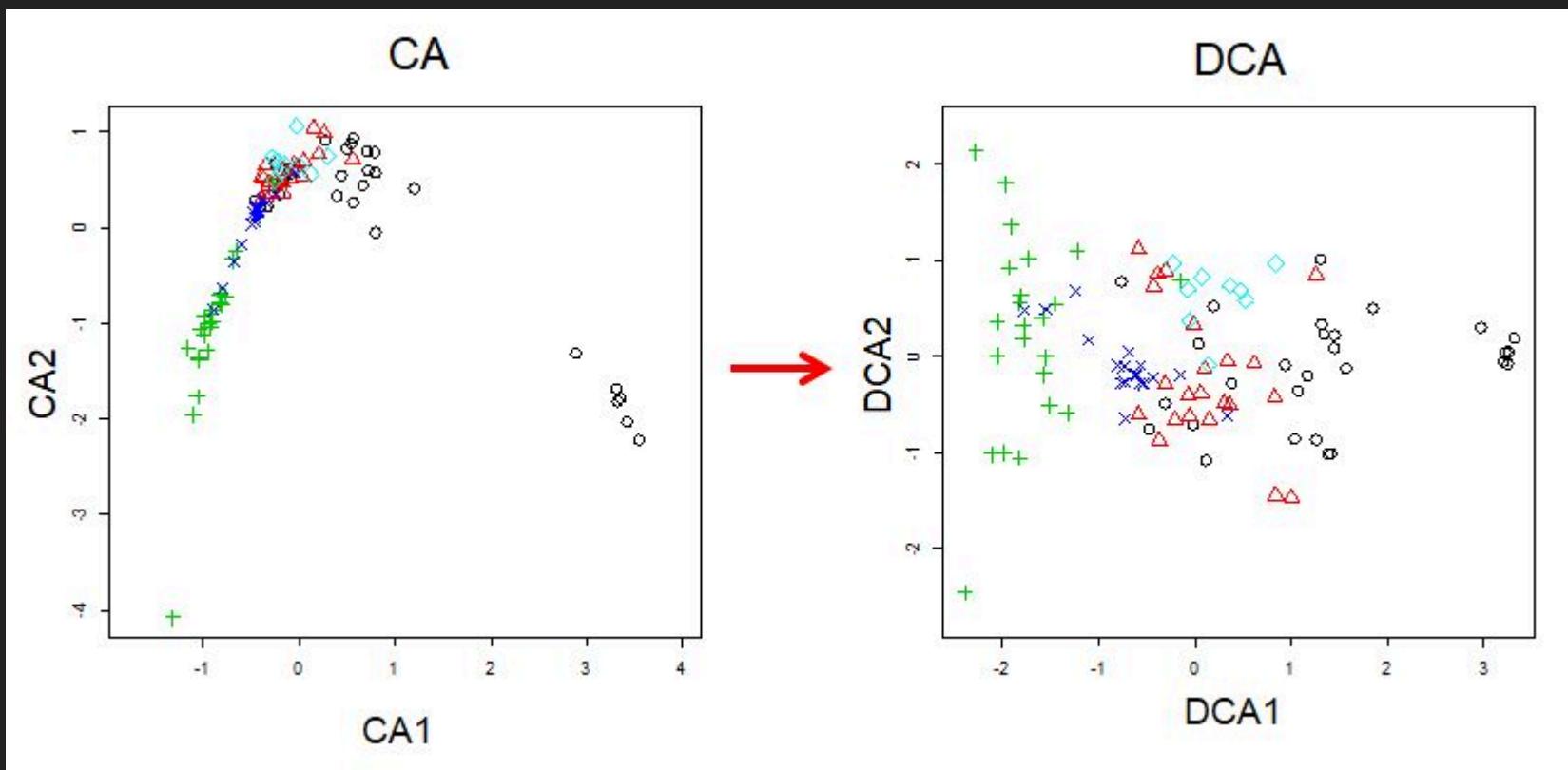
- It has a problem with *CA*: suffers from creating often strong arch artefact in ordination diagrams. Which is caused by a non-linear correlation between first and higher axes



DETRENDED CORRESPONDENCE ANALYSIS (DCA)

Arch can be removed by detrending (*smooths out the data to make it easier to see the main patterns*), which is the base of the detrended correspondence analysis (DCA).

Think of DCA as a corrected version of CA—more reliable for real ecological



gradients.

DETRENDED CORRESPONDENCE ANALYSIS (DCA)

```
1 DCA<-decorana(dune2_spe)
2 DCA
```

Call:

```
decorana(veg = dune2_spe)
```

Detrended correspondence analysis with 26 segments.

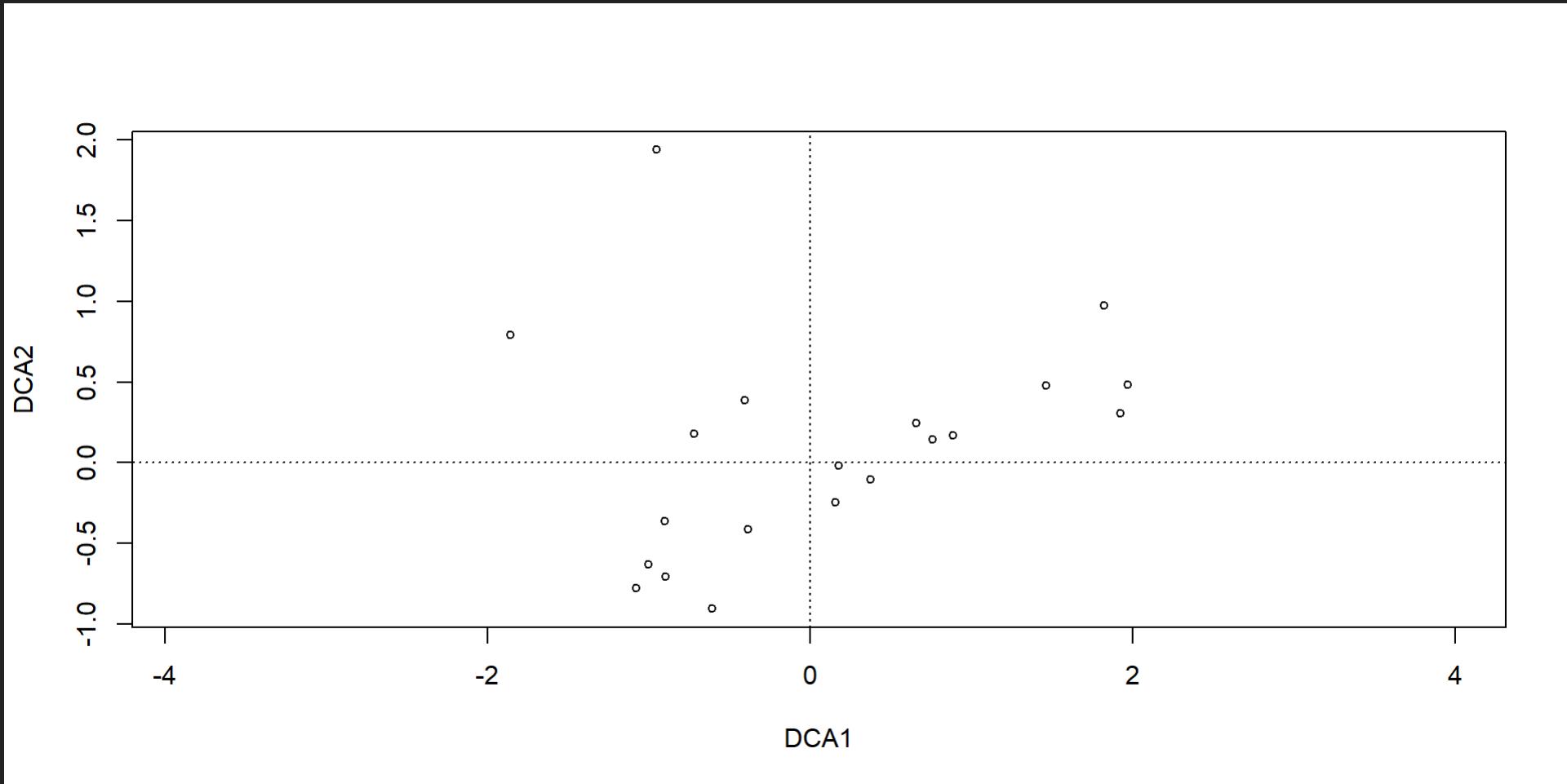
Rescaling of axes with 4 iterations.

Total inertia (scaled Chi-square): 2.1866

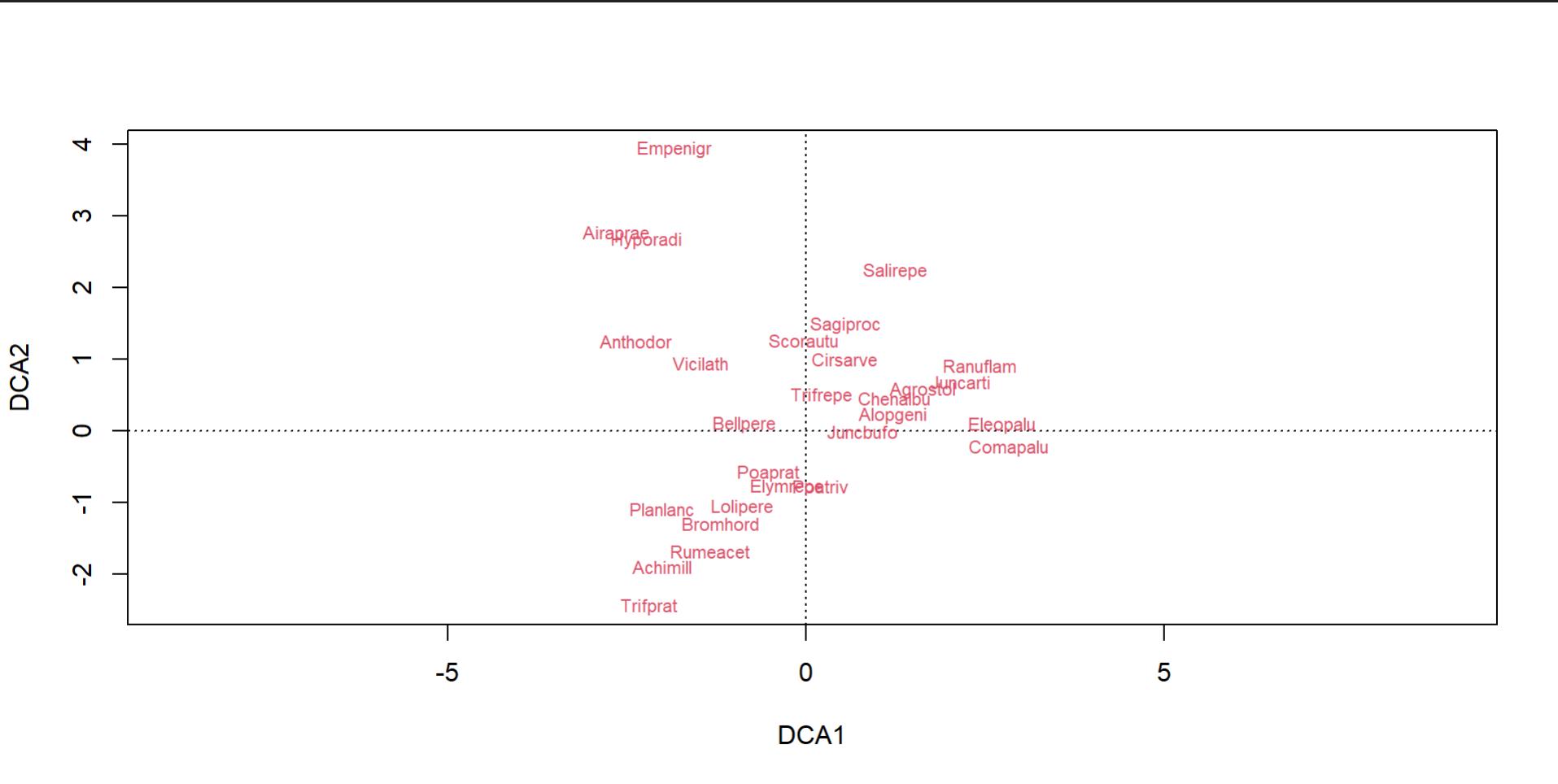
	DCA1	DCA2	DCA3	DCA4
Eigenvalues	0.5392	0.3257	0.16889	0.19567
Additive Eigenvalues	0.5392	0.3175	0.15318	0.18878
Decorana values	0.5636	0.3194	0.07921	0.04138
Axis lengths	3.8264	2.8444	2.03949	2.17577

DETRENDED CORRESPONDENCE ANALYSIS (DCA)

```
1 ordiplot (DCA, display = 'sites', type = 'p')
```



```
1 ordiplot (DCA, display = 'species', type = 't')
```

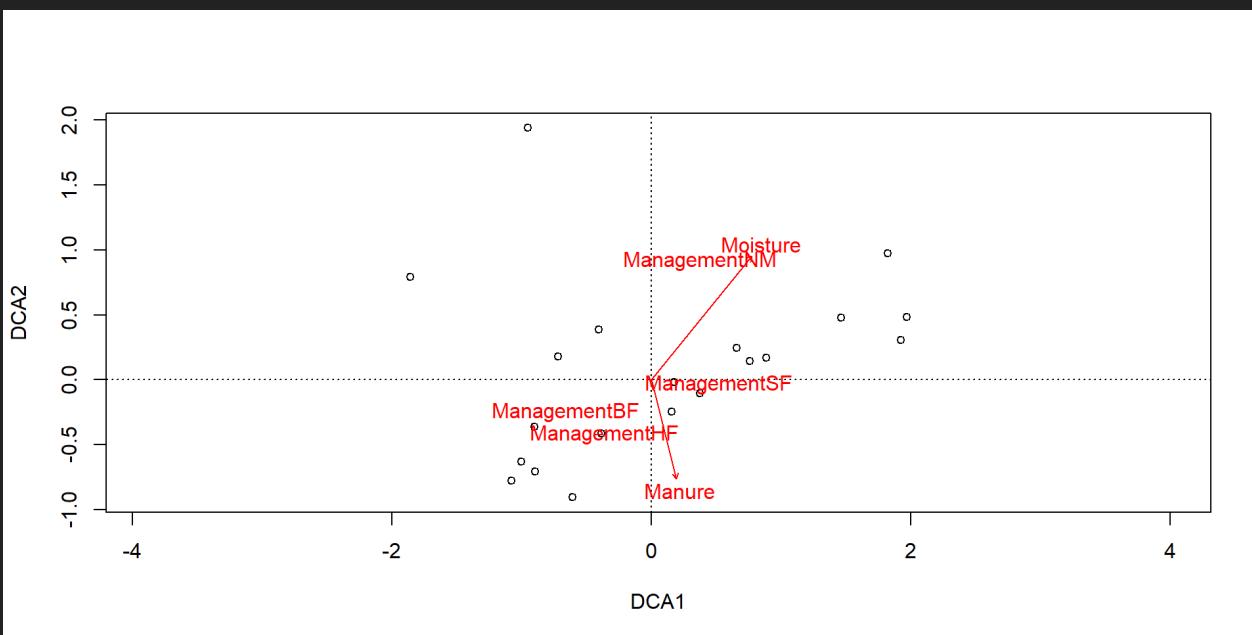


TRI PLOT

```
1 fit <- envfit(DCA, dune2_env, perm = 999)
2 scores(fit, "vectors")
```

	DCA1	DCA2
A1	0.5149060	0.01196127
Moisture	0.5571229	0.68949173
Manure	0.1388042	-0.55260255

```
1 plot(DCA, dis="site")
2 plot(fit, p.max = 0.05, col = "red")
```



NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)

- NMDS is a flexible, non-linear ordination method that shows how similar or different samples are—based on a chosen distance metric (e.g., Bray-Curtis). It tries to preserve the rank order of distances between samples—not the exact values.
- It minimizes “stress”, which measures how well the distances in the reduced space reflect the original data.

NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)

- The lower the stress value (a measure of goodness-of-fit), the better the representation of objects in the ordination-space is.
- `distance` specifies the distance metric to use
- `k` specifies the number of dimensions. You decide how many dimensions `k` to use.

IMPORTANT THINGS TO KNOW

- NMDS does not use eigenanalysis like PCA or DCA.
- So there's no unique solution (you might get slightly different results each time).
- Axes are not ranked by explained variance.
- You must choose the number of dimensions in advance.

NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)

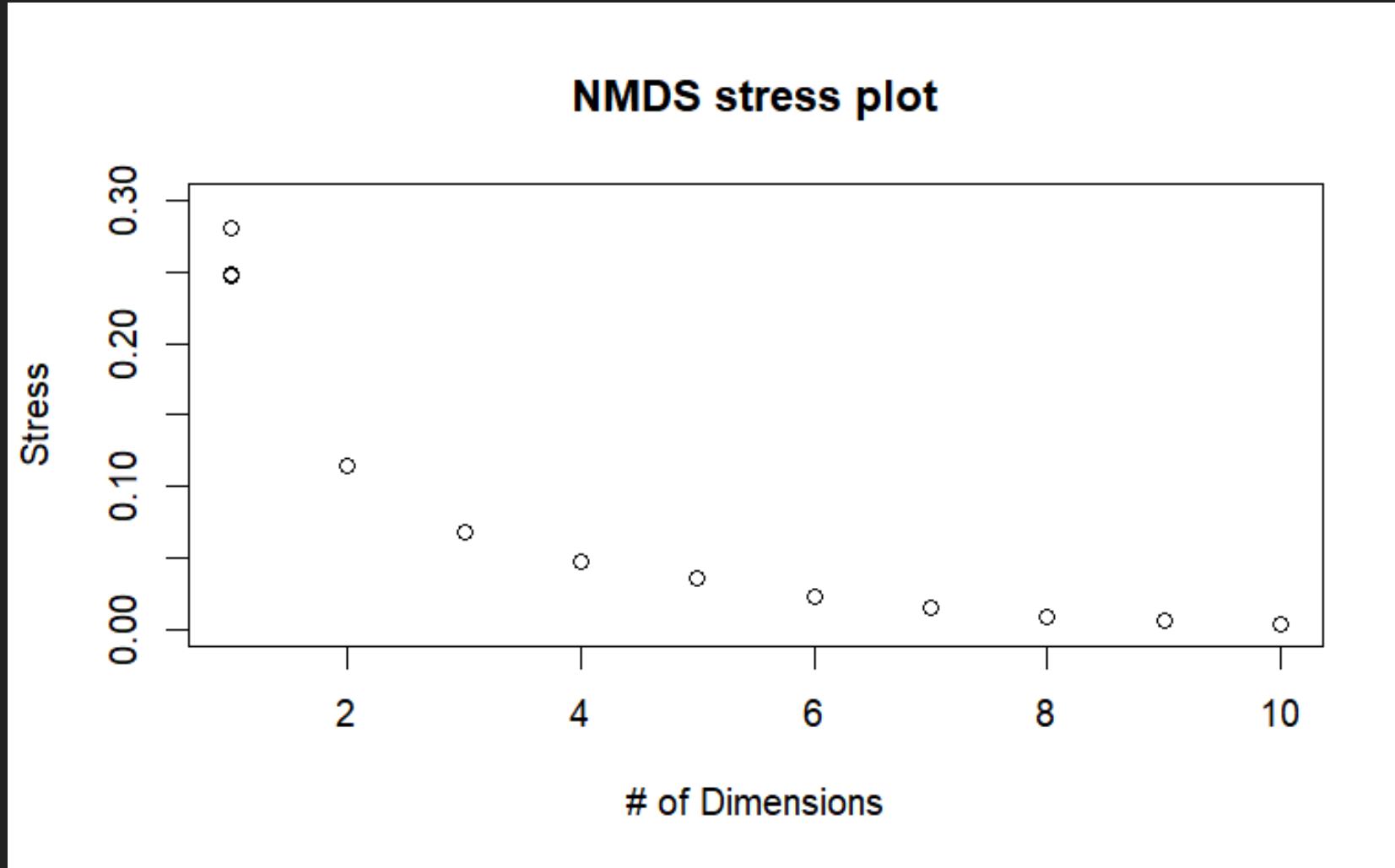
Methodology of NMDS:

Step 1: Perform NMDS with 1 to 10 dimensions
Step 2: Check the stress vs dimension plot
Step 3: Choose the lowest number of dimensions with acceptably low stress (typically < 0.2).
Step 4: Re-run NMDS with that number of dimensions.
Step 5: Check for convergence and interpret the final result.

NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)

```
1 # First step is to calculate a distance matrix. See PCOA
2 # Here we use bray-curtis distance, which is recommended
3 dist <- vegdist(dune2_spe, method = "bray")
4
5 # In this part, we define a function NMDS.scree() that a
6 # performs a NMDS for 1-10 dimensions and plots the nr o
7 NMDS.scree <- function(x) { #where x is the name of the
8   plot(rep(1, 10), replicate(10, metaMDS(x, autotransfor
9   for (i in 1:10) {
10     points(rep(i + 1,10),replicate(10, metaMDS(x, autotr
11   }
12 }
13 NMDS.scree(dist)
```

NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)



NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)

```
1 # Because the final result depends on the initial  
2 # random placement of the points  
3 # we'll set a seed to make the results reproducible  
4 set.seed(2)  
5  
6 # Here, we perform the final analysis and check the resu  
7 NMDS1 <- metaMDS(dist, k = 3, trymax = 100, trace = F)  
8 # Do you know what the trymax = 100 and trace = F means?  
9 # Let's check the results  
10 NMDS1
```

Call:

```
metaMDS(comm = dist, k = 3, trymax = 100, trace = F)
```

global Multidimensional Scaling using monoMDS

Data: dist

Distance: bray

Dimensions: 3

Stress: 0.06826238

Stress type 1, weak ties

Best solution was repeated 5 times in 20 tries

The best solution was from try 9 (random start)
Scaling: centring, PC rotation, halfchange scaling
Species: scores missing

```
1 # If you don't provide a dissimilarity matrix, metaMDS a
2 NMDS2 <- metaMDS(dune2_spe, k = 2, trymax = 100, trace =
3 NMDS2
```

Call:
metaMDS(comm = dune2_spe, k = 2, trymax = 100, trace = F)

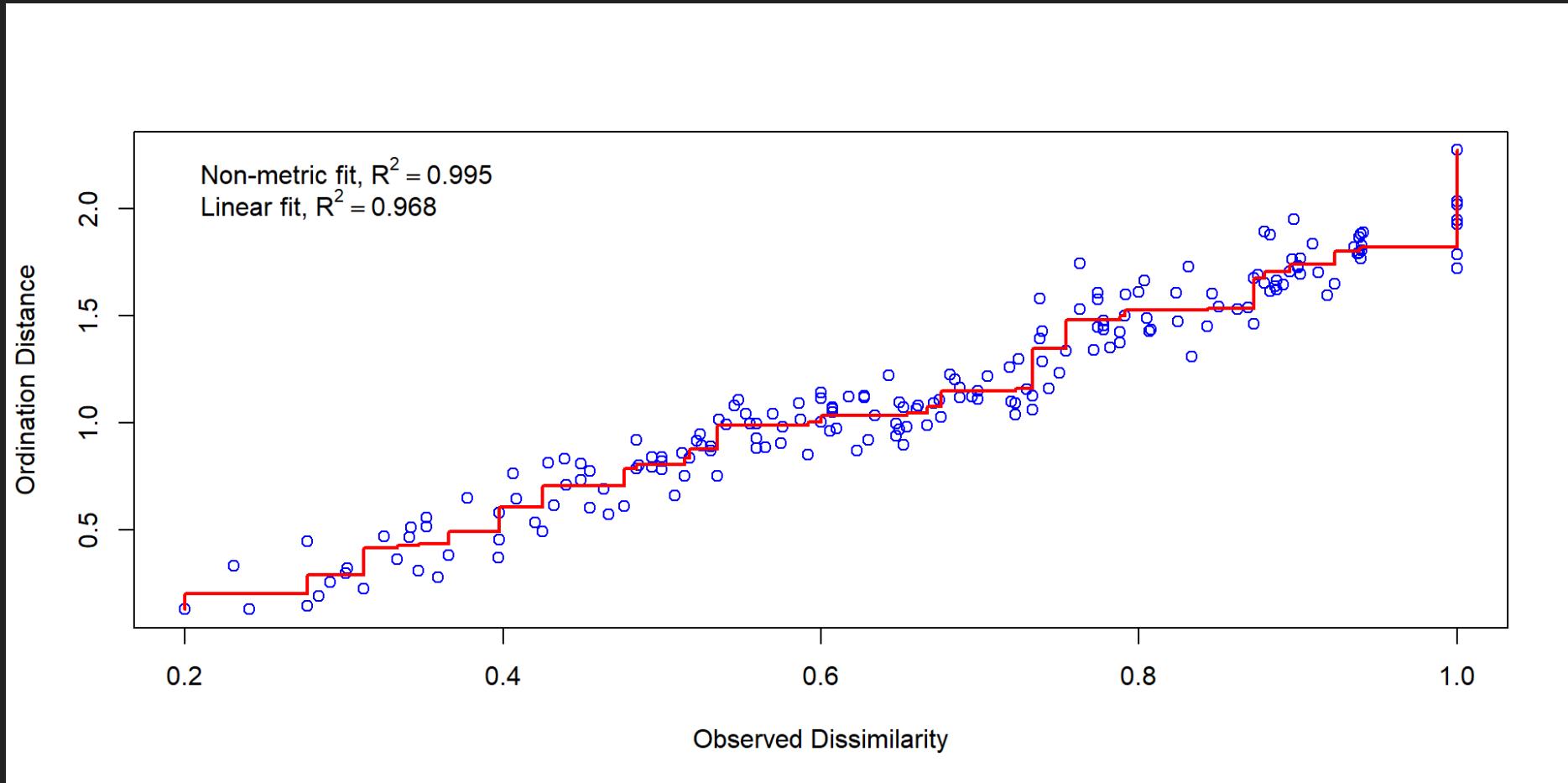
global Multidimensional Scaling using monoMDS

Data: dune2_spe
Distance: bray

Dimensions: 2
Stress: 0.1149964
Stress type 1, weak ties
Best solution was repeated 10 times in 20 tries
The best solution was from try 10 (random start)
Scaling: centring, PC rotation, halfchange scaling
Species: expanded scores based on 'dune2_spe'

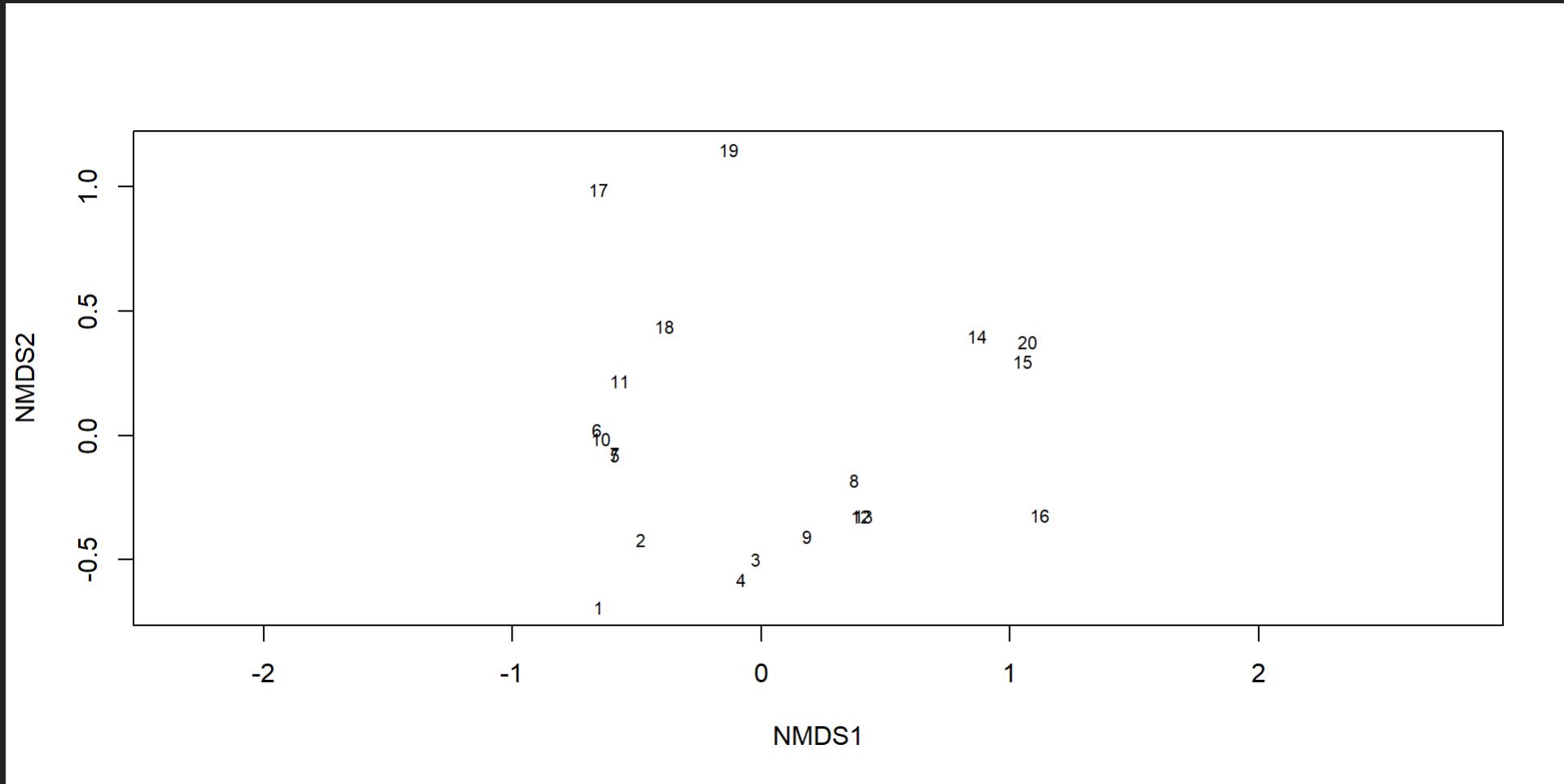
NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)

```
1 stressplot(NMDS1)
```



NON-METRIC MULTIDIMENSIONAL SCALING (NMDS)

```
1 plot(NMDS1, type = "t")
```

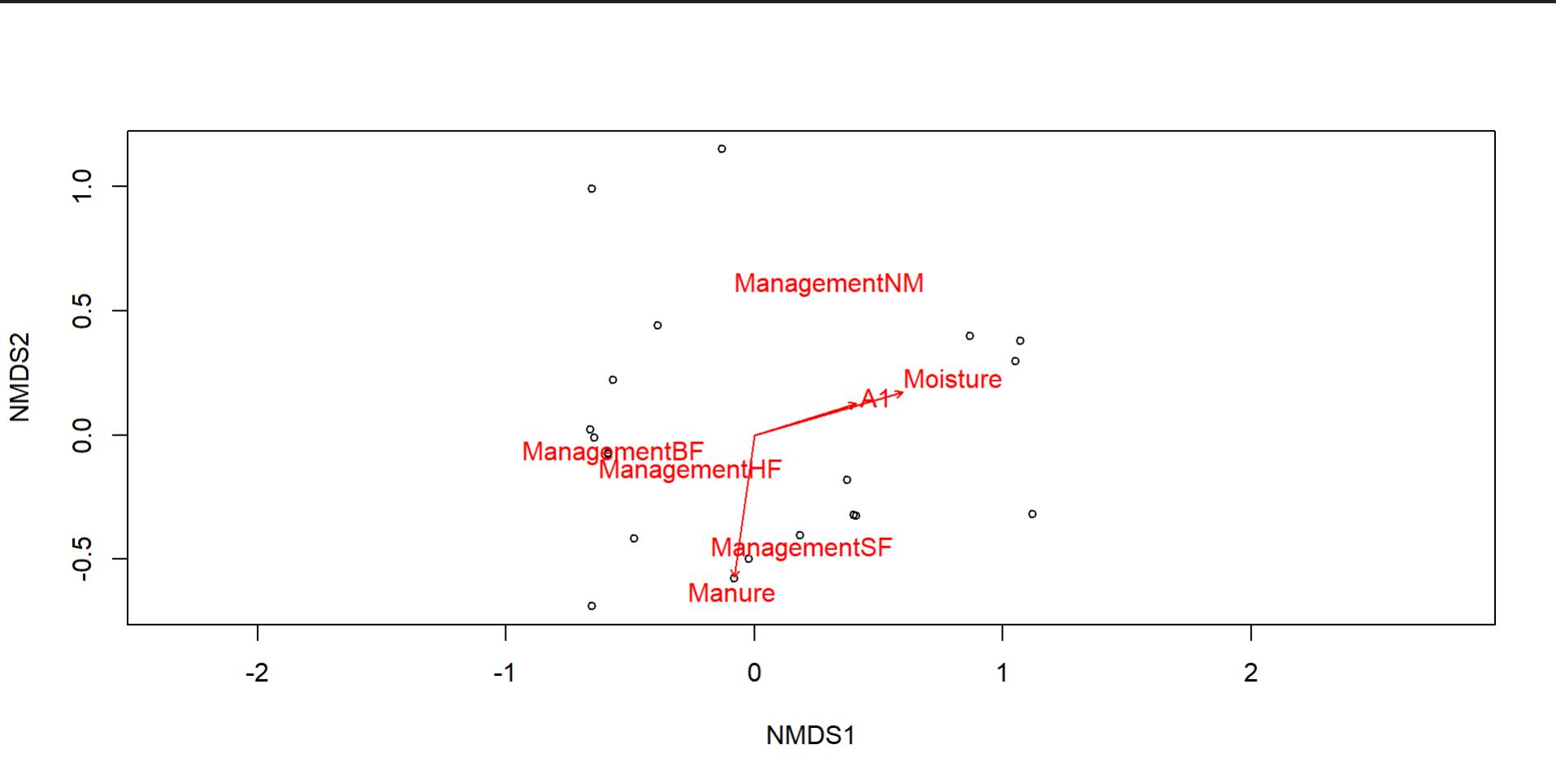


#Triplot

```
1 fit <- envfit(NMDS1, dune2_env, perm = 999)
2 scores(fit, "vectors")
```

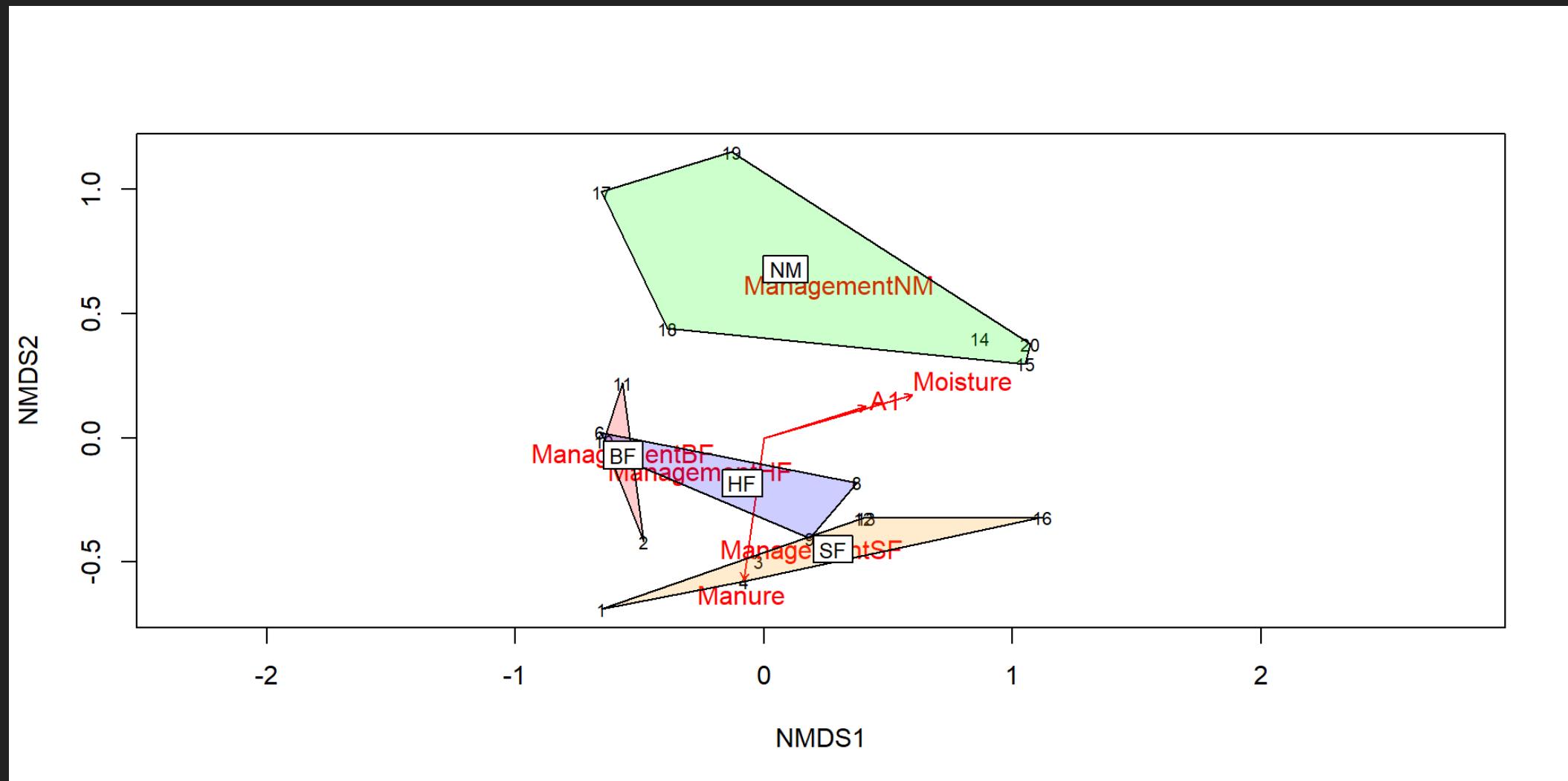
	NMDS1	NMDS2
A1	0.5926574	0.1838978
Moisture	0.8639184	0.2488776
Manure	-0.1197459	-0.8307617

```
1 plot(NMDS1, dis="site")
2 plot(fit, p.max = 0.05, col = "red")
```



ORDIHULL

ORDIHULL



SUMMARY

- In short: Ordination helps ecologists **see the big picture.**
- Use **PCA** when gradients are short and responses are simple.
- Use **DCA** when you suspect **distinct ecological zones** or turnover.
- Use **NMDS** for flexible, non-linear ecological data—common in field studies.
- Use **CCA/RDA** when you want to test specific **hypotheses about drivers.**

|  Think: *What ecological pattern am I trying to uncover or explain?*

MATERIAL

Most of the material comes from :

Introduction to ordinations

Visit Coding club for more examples

BREAK



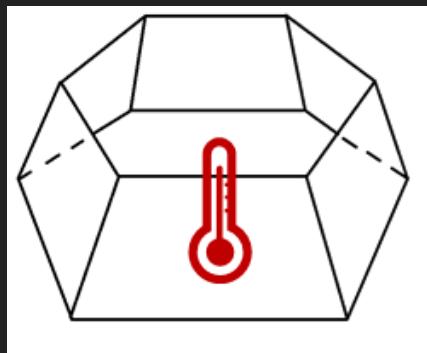
EXERCISE

In this exercise, you will explore community composition using NMDS (Non-metric Multidimensional Scaling) based on a subset of the ITEX dataset.

ITEX

We will be using a subset of the ITEX network dataset we used last time.

- ITEX is a long-term warming experiment that uses standardized protocols to examine impacts of warming on Arctic ecosystems.
- Established in the 1990s - vegetation monitoring over three decades.
- Uses a simple method that is easy to establish in the field - open top chambers



ITEX



LET'S LOOK AT THE DATA

```
1 itex.long <- read_csv("C:/Users/katbj/OneDrive - Univers  
1 head(itex.long)
```

```
# A tibble: 6 × 6  
...#> # ... with 6 variables:  
...#> #   SITE     SUBSITE PLOT  SPECIES_NAME      SumAbund  
...#> #   <dbl>    <chr>   <chr>  <chr>          <dbl>  
1     1 ATQASUK AD      ADC01 Alectoria nigricans      5  
2     2 ATQASUK AD      ADC01 Cassiope tetragona      5  
3     3 ATQASUK AD      ADC01 Cetraria cucullata      6  
4     4 ATQASUK AD      ADC01 Diapensia lapponica      6  
5     5 ATQASUK AD      ADC01 Dicranum acutifolium      1  
6     6 ATQASUK AD      ADC01 Ledum palustre        6
```

```
1 length(unique(itex.long$SITE)) #this code tells you how many sites there are
```

[1] 2

```
1 unique(itex.long$SITE) #you can also do this and then it will tell you what they are
```

[1] "ATQASUK" "BARROW"

EXERCISE

1. Load the dataset `itex.long.csv` and take a moment to familiarize yourself with its structure.
2. Prepare the data for ordination. Consider how to reshape it for NMDS.
3. Calculate a dissimilarity matrix and run the NMDS.
4. Visualize the NMDS results.
5. Interpret the results: What patterns do you see? Do sub-sites cluster together? What might explain the gradients in the ordination space?