Community ecology - Computer lab II - AB332

Anders K. Krabberød (UiO) and Ramiro Logares (ICM)

October 2023

Computer lab II

First load the necessary packages.

Now load the the data from the previous lab (change path to the file if necessary):

```
load("AB332_lab_I.RData")
```

## Ordination and clustering

PCA

Pca from the rarefied table

```
otu.tab.trans.ss.nozero.pca <- PCAtools::pca(t(otu.tab.trans.ss.nozero), scale = FALSE) # Runs de PCA
biplot(otu.tab.trans.ss.nozero.pca, showLoadings = T, lab = rownames(otu.tab.trans.ss.nozero)) # Plots
```

```
## Warning: Removed 4 rows containing missing values (`geom_segment()`).
```
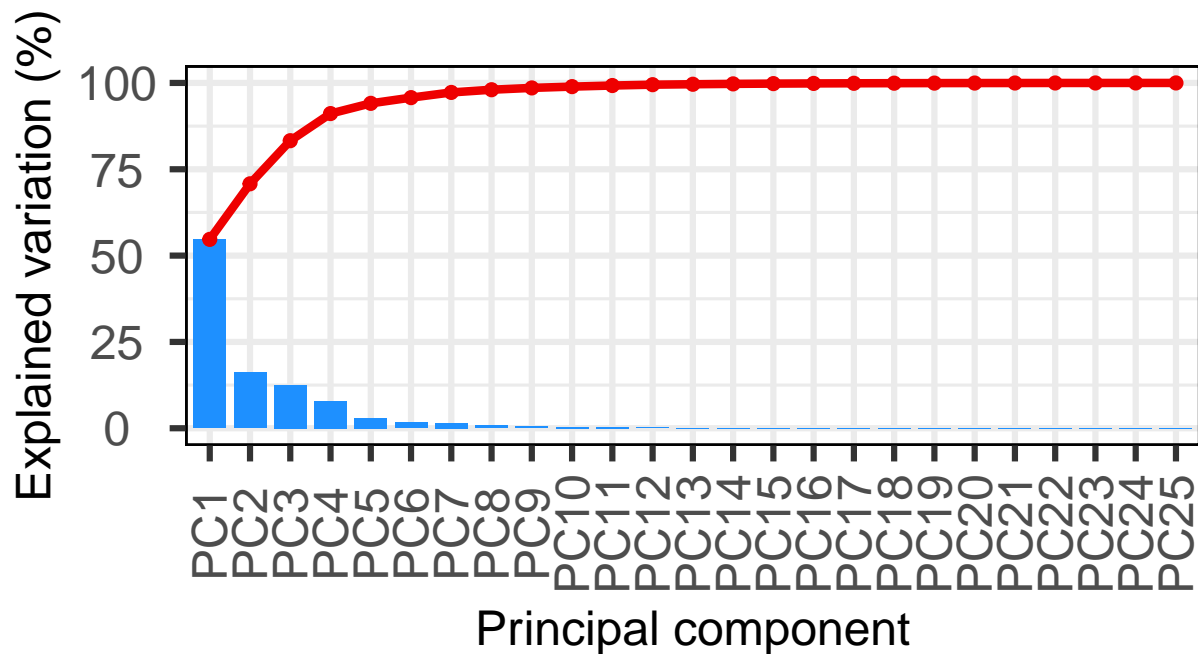
```
## Warning: Removed 4 rows containing missing values (`geom_label_repel()`).
```

```
## Warning: ggrepel: 14 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```
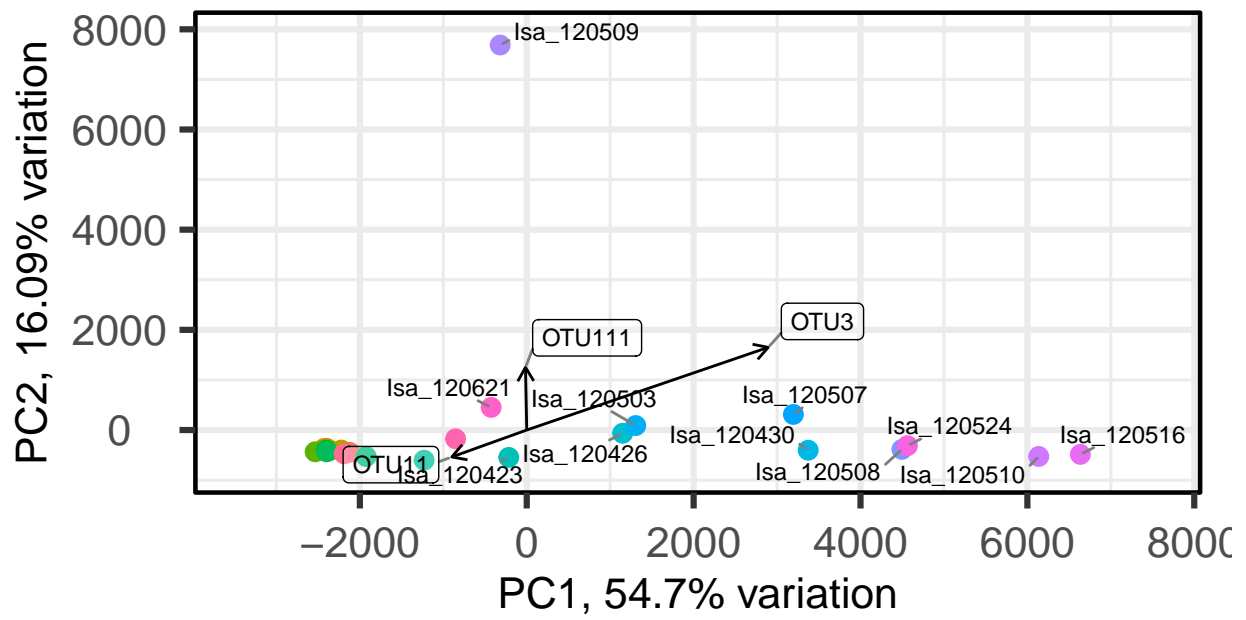
```
screeplot(otu.tab.trans.ss.nozero.pca, axisLabSize = 18, titleLabSize = 22) # We plot the percentage of
```

SCREE plot

NMDS

We will define the function NMDS.scree() that automatically performs a NMDS for 1-7 dimensions and plots the number of dimensions vs. stress

```
set.seed(666) # Set a seed to make results reproducible
NMDS.scree <- function(x) { # x is the name of the distance matrix
  plot(rep(1, 7), replicate(7, metaMDS(x, autotransform = FALSE, k = 1)$stress), xlim = c(1, 7), ylim =
  for (i in 1:7) {
    points(rep(i + 1, 7), replicate(7, metaMDS(x, autotransform = FALSE, k = i + 1)$stress))
  }
}
```

Using the function to determine the optimal number of dimensions Using the rarefied table

We calculate NMDS for k(dimensions)=2 Rarefied table (we use the dataframe to have access to sample and OTU names)

```
otu.tab.trans.ss.nozero.bray.nmds <- metaMDS(otu.tab.trans.ss.nozero, k = 2, trymax = 100, trace = FALSE
otu.tab.trans.ss.nozero.bray.nmds
```

```
##
## Call:
## metaMDS(comm = otu.tab.trans.ss.nozero, distance = "bray", k = 2,     trymax = 100, autotransform =
```

```
##
## global Multidimensional Scaling using monoMDS
##
## Data:        otu.tab.trans.ss.nozero
## Distance: bray
##
## Dimensions: 2
## Stress:      0.06397485
## Stress type 1, weak ties
## Best solution was not repeated after 100 tries
## The best solution was from try 77 (random start)
## Scaling: centring, PC rotation, halfchange scaling
## Species: expanded scores based on 'otu.tab.trans.ss.nozero'
```
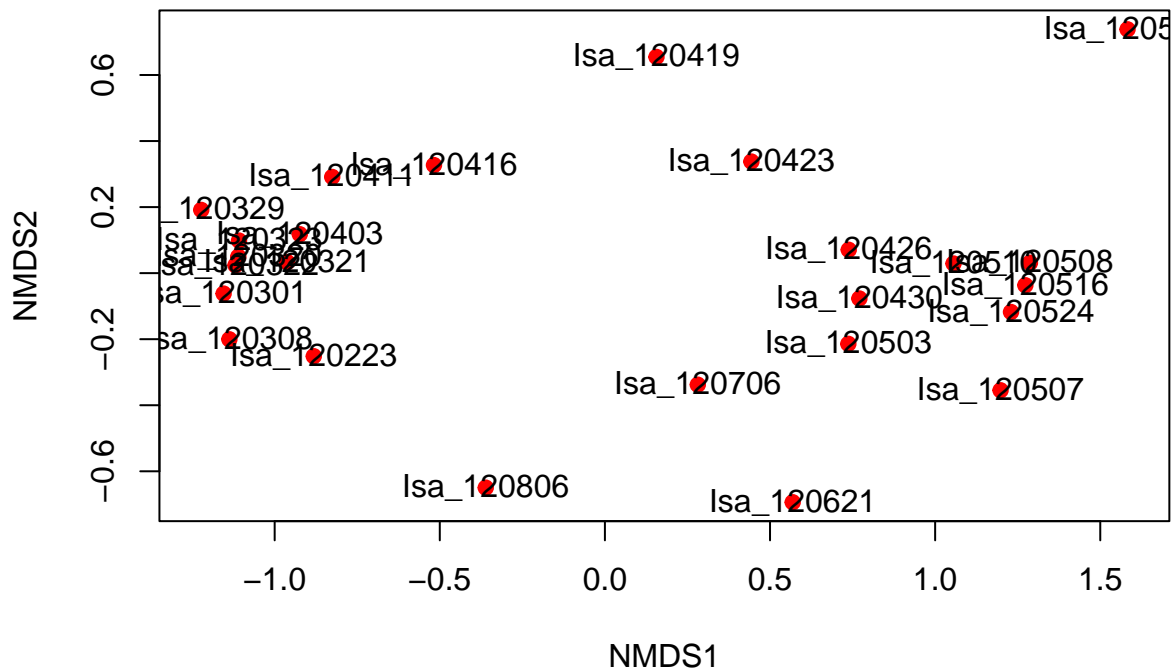
Make stressplot

```
stressplot(otu.tab.trans.ss.nozero.bray.nmds)
```
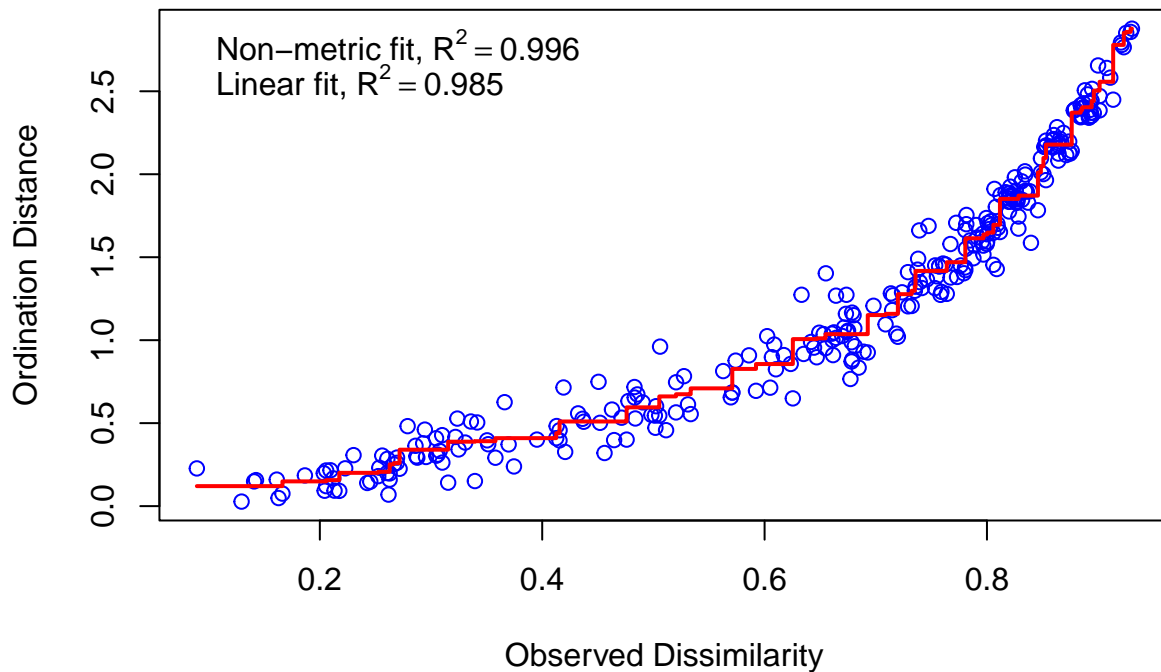

Simple plotting
Rarefied table

```
plot(otu.tab.trans.ss.nozero.bray.nmds, display = "sites", type = "n")
points(otu.tab.trans.ss.nozero.bray.nmds, display = "sites", col = "red", pch = 19)
text(otu.tab.trans.ss.nozero.bray.nmds, display = "sites")
```

Let's make nicer plots

We get the seasons for samples

```
isa.metadata <- read_tsv("https://raw.githubusercontent.com/krabberod/UNIS_AB332_2022/main/computer_lab,
```

```
## Rows: 82 Columns: 26
## -- Column specification ---------------------------------------------------
## Delimiter: "\t"
## chr  (3): Sample_Name, date, seasons
## dbl (23): year, season_nr, month, day, DOY, running_day, template, depth, de...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
isa.metadata <- column_to_rownames(isa.metadata, var = "Sample_Name")
isa.metadata.simp <- isa.metadata[6:30, ]
```

Rarefied table We generate a table of nmds scores and other features

```
otu.tab.trans.ss.nozero.bray.nmds.scores <- as.data.frame(scores(otu.tab.trans.ss.nozero.bray.nmds)$site
otu.tab.trans.ss.nozero.bray.nmds.scores$season <- isa.metadata.simp$seasons
otu.tab.trans.ss.nozero.bray.nmds.scores$month <- as.factor(isa.metadata.simp$month)
otu.tab.trans.ss.nozero.bray.nmds.scores$samples <- rownames(otu.tab.trans.ss.nozero.bray.nmds.scores)
```

Create the plot

```
ggplot(otu.tab.trans.ss.nozero.bray.nmds.scores) +
  geom_point(mapping = aes(x = NMDS1, y = NMDS2, colour = season), size = 3) +
  coord_fixed() + ## need aspect ratio of 1!
  geom_text_repel(
    box.padding = 0.5, aes(x = NMDS1, y = NMDS2, label = samples),
    size = 3
  )
```

```
## Warning: ggrepel: 5 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```
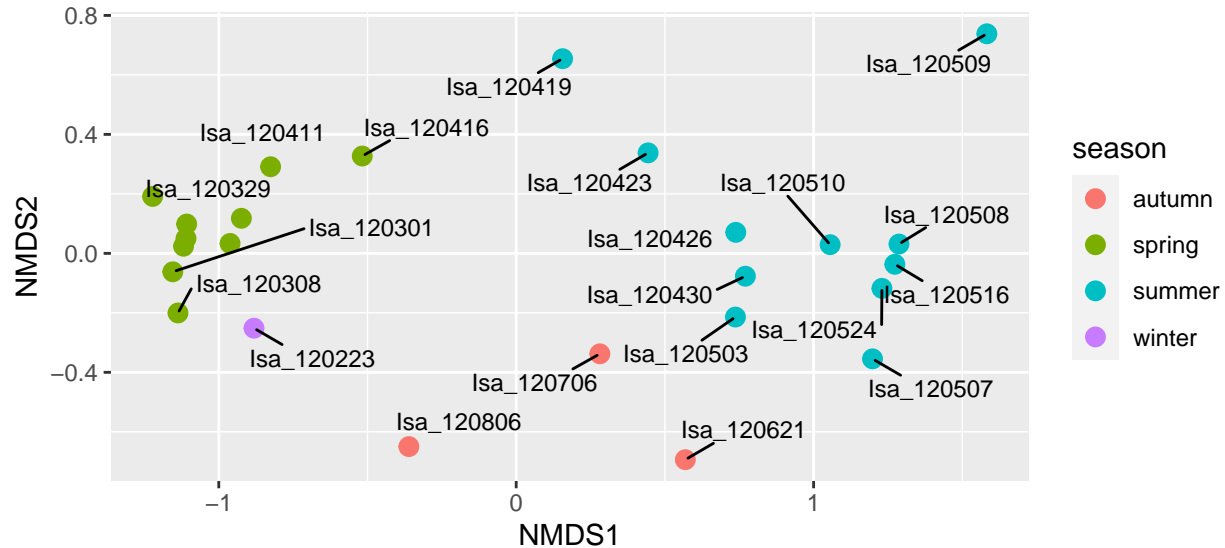

#Clustering of samples

Allows determining the similarity between samples as well as the organization of samples in groups. Hierarchical clustering: samples will be organized in ranks according to their similarity and all samples will be included in a large group Unweighted Pair-Group Method Using Arithmetic Averages (UPGMA): This linkage method will link samples by considering their distance to a subgroup arithmetic average. This is a method widely used in ecology
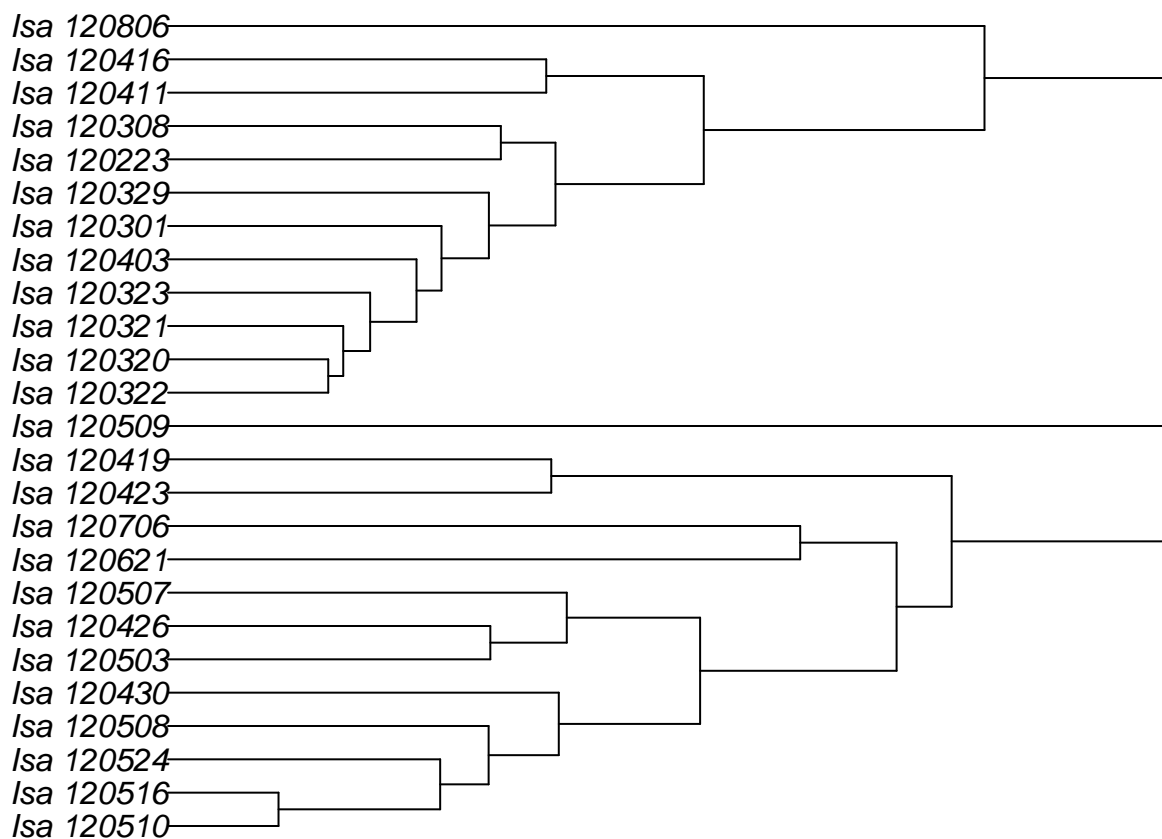
UPGMA

Rarefied dataset We generate 100 trees by re-sampling and then, we plot the consensus tree

```
otu.tab.trans.ss.nozero.bray.upgma <- recluster.cons(otu.tab.trans.ss.nozero.bray, tr = 100, p = 0.5, me
plot(otu.tab.trans.ss.nozero.bray.upgma$cons) # plot consensus tree
```
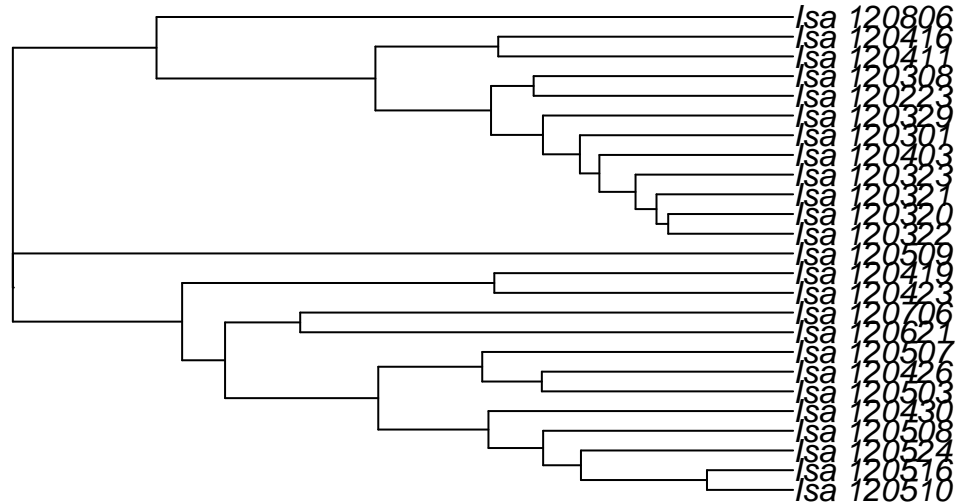

We'll calculate bootstrap support values (0: bad - 100: perfect) This allows us to know how well supported is the branching pattern

```
otu.tab.trans.ss.nozero.bray.upgma.boot <- recluster.boot(otu.tab.trans.ss.nozero.bray.upgma$cons, otu.
recluster.plot(otu.tab.trans.ss.nozero.bray.upgma$cons, otu.tab.trans.ss.nozero.bray.upgma.boot, direct
```

```
save.image("AB332_lab_II.RData")
```