

# A (brief) introduction to ordination and the vegan package

Eduard Szöcs

Institute for Environmental Sciences - University of Koblenz-Landau



SEFS9, July 5th 2015

Datasets  
oooooo

Unconstrained Ordination  
oooooooooooooooooooo

Constrained Ordination  
oo

Permutation Tests  
o

Datasets  
oooooo

Unconstrained Ordination  
oooooooooooooooooooo

Constrained Ordination  
oo

Permutation Tests  
o

# Topics addressed

	Raw data	Transformed data	Unimodal	Distance-based
Unconstrained	PCA	tb-PCA	CA, DCA	PCoA, NMDS
Constrained	RDA	tb-RDA	CCA	db-RDA

Datasets  
oooooo

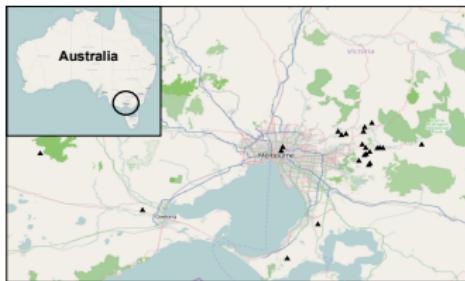
Unconstrained Ordination  
oooooooooooooooooooo

Constrained Ordination  
oo

Permutation Tests  
o

# Datasets

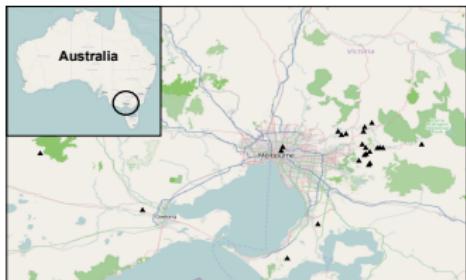
## Exercise: Salinization and Pesticides



- ▶ Macroinvertebrates
  - ▶ 24 sites
  - ▶ covering a salinity and toxicity gradient

The dataset is published in: Szöcs, E., Kefford, B.J., Schäfer, R.B., 2012. Is there an interaction of the effects of salinity and pesticides on the community structure of macroinvertebrates? *Science of the Total Environment* 437, 121–126.

## Exercise: Salinization and Pesticides



- ▶ Macroinvertebrates
  - ▶ 24 sites
  - ▶ covering a salinity and toxicity gradient

## Questions:

- ▶ Interaction between salinization and pesticides?
  - ▶ Which species are affected?
  - ▶ Other influences?

The dataset is published in: Szöcs, E., Kefford, B.J., Schäfer, R.B., 2012. Is there an interaction of the effects of salinity and pesticides on the community structure of macroinvertebrates? *Science of the Total Environment* 437, 121–126.

## Exercise: Salinization and Pesticides

```
setwd('yourpath/3-IntroVeganPackage/data/')
abu <- read.table('melbourneAbu.csv', sep = ';', header = TRUE)
env <- read.table('melbourneEnv.csv', sep = ';', header = TRUE)
```

**dim(env)**

[1] 24 23

**dim(abu)**

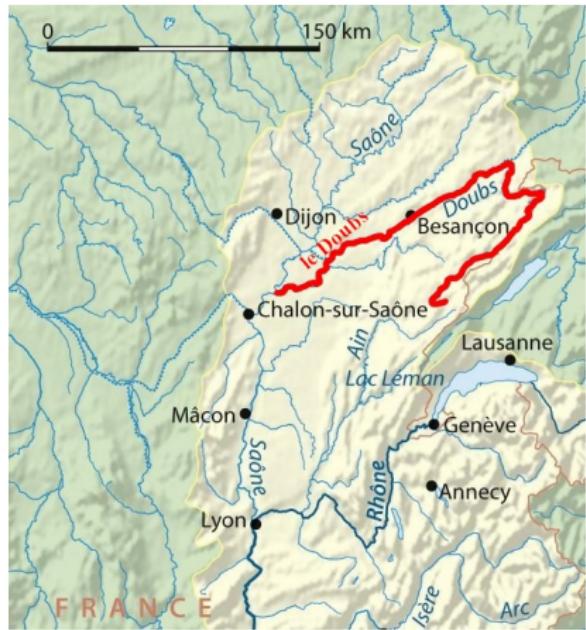
[1] 24 76

24 sites, 22 environmental variables, 75 taxa

```
head(env[, 1:10])
```

ID	T	pH	oxygen	Depth	maxwidth	minwidth	rifperc	poolperc	Bedrock	
1	1-11	16.8	7.67	80.1	0.9	15	12.0	0	100	0
2	2-11	16.5	7.29	83.0	0.9	30	15.0	0	100	0
3	3-11	17.3	7.20	77.9	0.4	4	2.5	0	100	0
4	4-11	15.6	7.84	72.0	0.7	8	2.5	0	100	0
5	5-11	17.2	6.97	69.9	0.9	7	4.0	0	100	0
6	6-11	15.5	7.26	80.0	0.2	3	2.0	5	95	0

# Demonstration: Doubs river fish communities



- ▶ Fish
- ▶ 30 sites along the Doubs River

---

Verneaux, J. (1973) Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. These d'état, Besançon. 1–257.

Datasets



Unconstrained Ordination



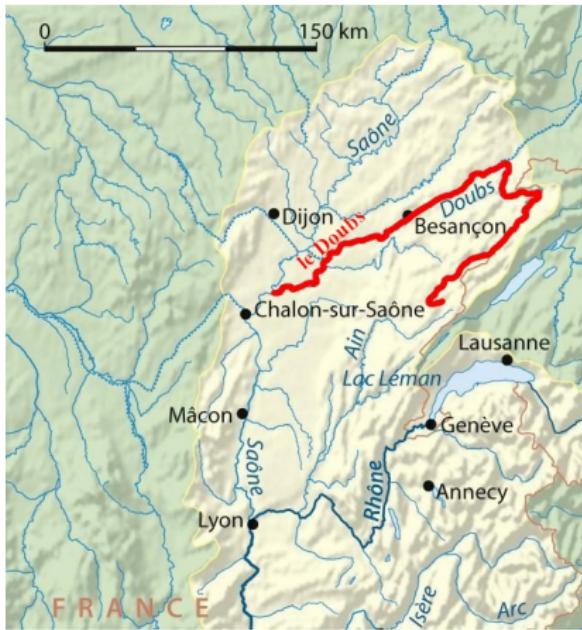
Constrained Ordination



Permutation Tests



# Demonstration: Doubs river fish communities



- ▶ Fish
- ▶ 30 sites along the Doubs River

## Questions

- ▶ How does fish composition change downstream?
- ▶ Environmental drivers?

Verneaux, J. (1973) Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. These d'état, Besançon. 1–257.

Datasets



Unconstrained Ordination



Constrained Ordination



Permutation Tests



## Demonstration: Doubs river fish communities — Species

```
setwd('your/workingdirectory')
Dabu <- read.table('doubtsAbu.csv', sep = ',', header = TRUE)
Denv <- read.table('doubtsEnv.csv', sep = ',', header = TRUE)
Dspa <- read.table('doubtsSpa.csv', sep = ',', header = TRUE)
```

**dim(Dabu)**

[1] 30 27

30 sites, 27 taxa

```
head(Dabu[, 1:18])
```

	CHA	TRU	VAI	LOC	OMB	BLA	HOT	TOX	VAN	CHE	BAR	SPI	GOU	BRO	PER	BOU	PSO	ROT
1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	5	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	5	5	5	0	0	0	0	0	0	0	0	0	1	0	0	0	
4	0	4	5	5	0	0	0	0	0	1	0	0	1	2	2	0	0	
5	0	2	3	2	0	0	0	0	5	2	0	0	2	4	4	0	0	
6	0	3	4	5	0	0	0	0	1	2	0	0	1	1	1	0	0	

Demonstration: Doubs river fish communities — Environment

$\dim(\text{Denv})$

[1] 30 11

30 sites, 11 variables

**head(Deny)**

	das	alt	pen	deb	pH	dur	pho	nit	amm	oxy	dbo
1	0.3	934	48.0	0.84	7.9	45	0.01	0.20	0.00	12.2	2.7
2	2.2	932	3.0	1.00	8.0	40	0.02	0.20	0.10	10.3	1.9
3	10.2	914	3.7	1.80	8.3	52	0.05	0.22	0.05	10.5	3.5
4	18.5	854	3.2	2.53	8.0	72	0.10	0.21	0.00	11.0	1.3
5	21.5	849	2.3	2.64	8.1	84	0.38	0.52	0.20	8.0	6.2
6	32.4	846	3.2	2.86	7.9	60	0.20	0.15	0.00	10.2	5.3

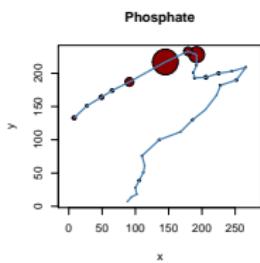
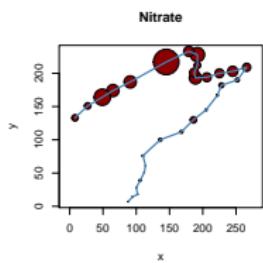
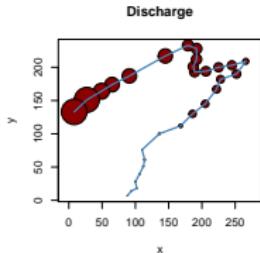
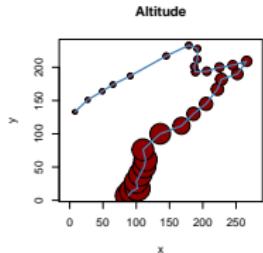
# Unconstrained Ordination

Principal Components Analysis (PCA)

Principal coordinates analysis (PCoA)

Nonmetric Multidimensional Scaling (NMDS)

# Principal Components Analysis (PCA) — Why?



- ▶ 11 variables

## Questions:

- ▶ Which variable are correlated?
- ▶ Which sites have similar conditions?
- ▶ How do conditions change downstream?

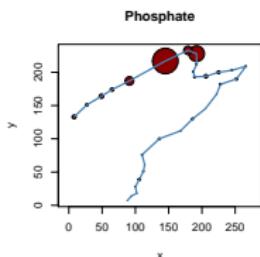
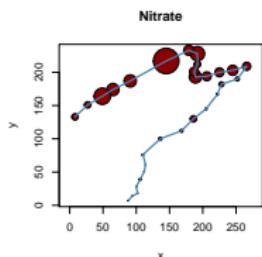
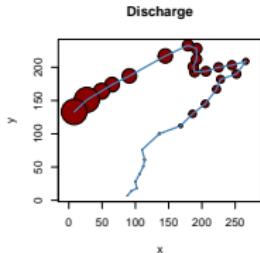
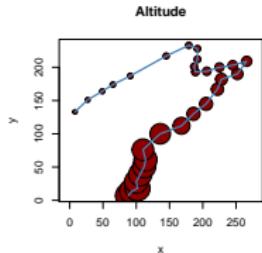
Datasets  
oooooo

Unconstrained Ordination  
●oooooooooooooooooooo

Constrained Ordination  
○○

Permutation Tests  
○○

# Principal Components Analysis (PCA) — Why?



- ▶ 11 variables

## Questions:

- ▶ Which variable are correlated?
- ▶ Which sites have similar conditions?
- ▶ How do conditions change downstream?
- ▶ pairwise comparisons
- ▶ 3D possible
- ▶ more than 3 dimensions?

# Principal Components Analysis (PCA) — What?

- ▶ *"Look from another angle on the data"*
- ▶ PCA is just a rotation of the coordinate system
- ▶ The rotation is done so that the first axis contains as much variation as possible
- ▶ Second axis than most of remaining variation

Datasets  
ooooo

Unconstrained Ordination  
○●oooooooooooooooooooo

Constrained Ordination  
○○

Permutation Tests  
○

# Principal Components Analysis (PCA) — What?

- ▶ *"Look from another angle on the data"*
- ▶ PCA is just a rotation of the coordinate system
- ▶ The rotation is done so that the first axis contains as much variation as possible
- ▶ Second axis than most of remaining variation

Maths:

- ▶ The covariance (or correlation) matrix is decomposed into its Eigenvectors and Eigenvalues.
- ▶ The Eigenvectors give the rotation needed
- ▶ The Eigenvalues stretch the axes

Datasets  
ooooo

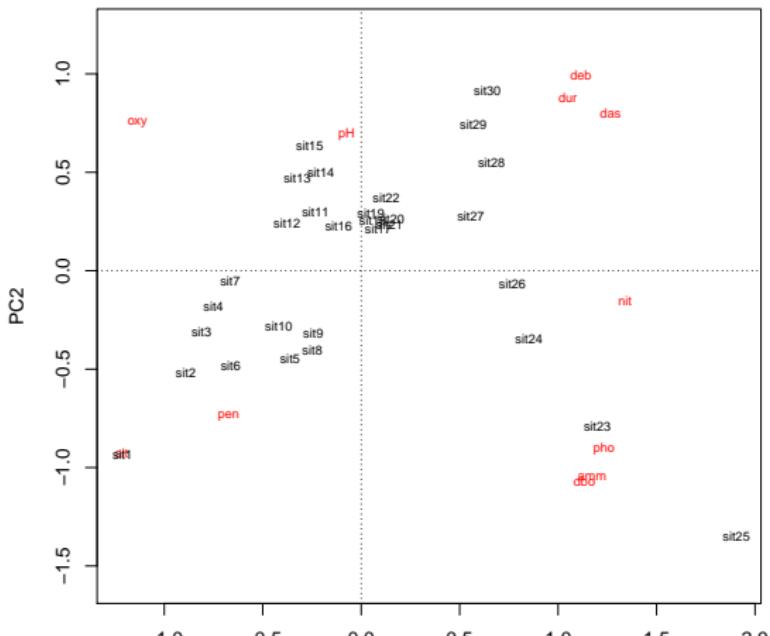
Unconstrained Ordination  
○●oooooooooooooooooooo

Constrained Ordination  
○○

Permutation Tests  
○

# Principal Components Analysis (PCA) — How?

```
require(vegan)  
PCA <- rda(Denv, scale = TRUE)  
plot(PCA, scaling = 3)
```



Datasets  
oooooo

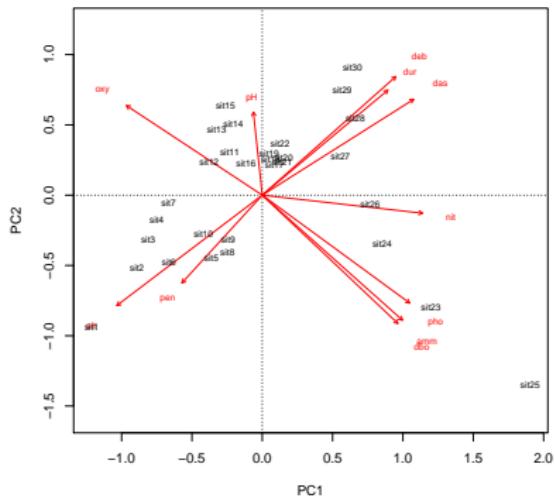
Unconstrained Ordination  
oooooooooooooooooooo

PC1  
Constrained Ordination  
oo

Permutation Tests  
o

## Principal Components Analysis (PCA) — Interpretation? (I)

```
biplot(PCA, cex = 5, scaling = 3)
```



## Datasets

oooooo

## Unconstrained Ordination

## Constrained Ordination oo

## Permutation Tests

# Principal Components Analysis (PCA) — Interpretation? (II)

```
summary(PCA, display = NULL, scaling = 3)
```

Call:

```
rda(X = Denv, scale = TRUE)
```

Partitioning of correlations:

	Inertia	Proportion
Total	11	1
Unconstrained	11	1

Eigenvalues, and their contribution to the correlations

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	5.9687	2.1638	1.06516	0.73873	0.40027	0.33565	0.1727
Proportion Explained	0.5426	0.1967	0.09683	0.06716	0.03639	0.03051	0.0157
Cumulative Proportion	0.5426	0.7393	0.83616	0.90331	0.93970	0.97022	0.9859
	PC8	PC9	PC10	PC11			
Eigenvalue	0.10821	0.02368	0.01707	0.005993			
Proportion Explained	0.00984	0.00215	0.00155	0.000540			
Cumulative Proportion	0.99575	0.99790	0.99946	1.000000			

Scaling 3 for species and site scores

- \* Both sites and species are scaled proportional to eigenvalues on all dimensions
- \* General scaling constant of scores:

# Your turn!

Load the Melbourne dataset (only environmental variables).

**Exclude** the variables ID, logCond and logmaxTU.

Perform a PCA.

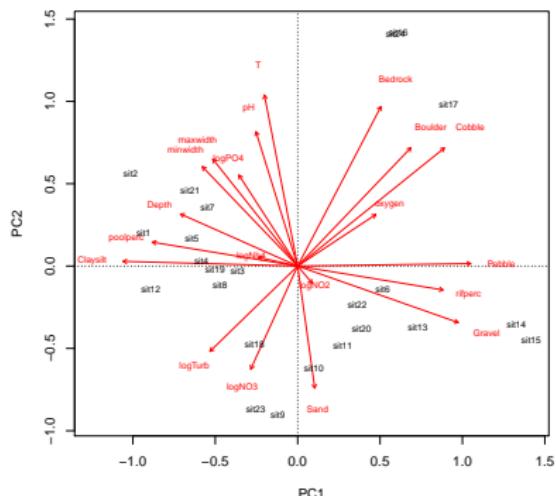
Which variables are correlated?

What other gradients are present in this dataset?

# Exercise

```
take <- env[ , !names(env) %in% c('ID', 'logCond', 'logmaxTU')]  
PCA <- rda(take, scale = TRUE)
```

```
biplot(PCA, scaling = 3)
```



- ▶ multiple variables interrelated
- ▶ 1st axis can be interpreted as *hydrological gradient*
- ▶ 2nd axis can be interpreted as *chemistry gradient*

## Excursus — Principal component regression (PCR)

### Question:

- ▶ How is diversity related to salinity, pesticides and other variables?

Datasets  
○○○○○

Unconstrained Ordination  
○○○○○○○○●○○○○○○○○○○○

Constrained Ordination  
○○

Permutation Tests  
○

## Excursus — Principal component regression (PCR)

### Question:

- ▶ How is diversity related to salinity, pesticides and other variables?

### Problem:

- ▶ Only 24 sites
- ▶ but 22 (partly correlated) explanatory variables
- ▶ strongy hypotheses about salinity and pesticides

### A Solution:

Datasets  
oooooo

Unconstrained Ordination  
oooooooo●oooooooooooo

Constrained Ordination  
○○

Permutation Tests  
○

## Excursus — Principal component regression (PCR)

### Question:

- ▶ How is diversity related to salinity, pesticides and other variables?

### Problem:

- ▶ Only 24 sites
- ▶ but 22 (partly correlated) explanatory variables
- ▶ strongy hypotheses about salinity and pesticides

### A Solution:

- ▶ Reduce number of variables to *Principal Components*
- ▶ regress these

Datasets  
oooooo

Unconstrained Ordination  
oooooooo●oooooooooooo

Constrained Ordination  
○○

Permutation Tests  
○

## Excursus — principal component regression (PCR)

```
div <- diversity(abu[ , -1], index = 'shannon')
pc <- scores(PCA, choices = c(1, 2), scaling = 1, display = 'sites')
model_data <- data.frame(div, pc, logCond = env$logCond, logmaxTU = env$logmaxTU)
model <- lm(div ~ PC1 + PC2 + logCond + logmaxTU, data = model_data)
summary(model)
```

Call:

```
lm(formula = div ~ PC1 + PC2 + logCond + logmaxTU, data = model_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.64415	-0.15688	0.02063	0.18219	0.57929

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.83079	0.43429	4.216	0.000468 ***
PC1	0.01971	0.16691	0.118	0.907262
PC2	0.02192	0.19570	0.112	0.911996
logCond	-0.20942	0.13050	-1.605	0.125049
logmaxTU	-0.12572	0.07316	-1.718	0.101994

---

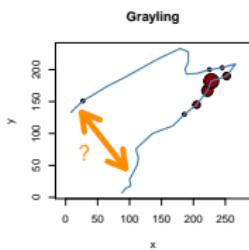
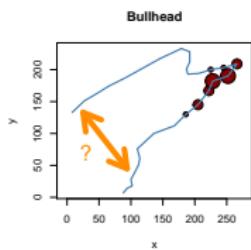
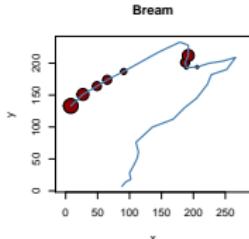
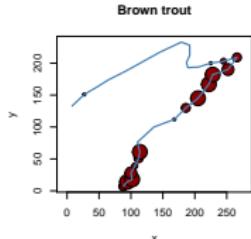
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3645 on 19 degrees of freedom

Multiple R-squared: 0.2682, Adjusted R-squared: 0.1141

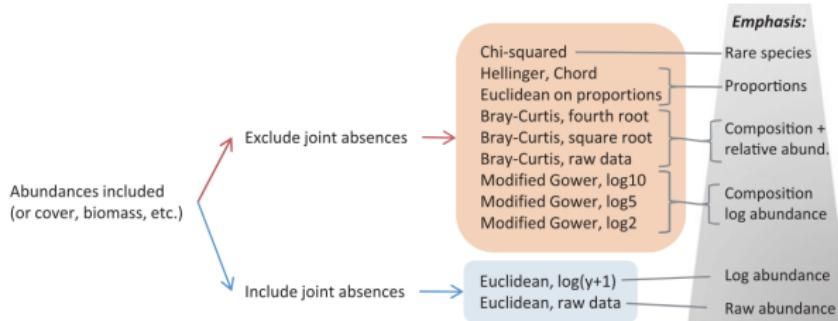
F-statistic: 1.741 on 4 and 19 DF, p-value: 0.1827

# Abundances — The Problem with zeros



- ▶ Species may be absent due to different factors (too high flow, too saline, etc.)
- ▶ *Absence* contains less information than *Presence*
- ▶ PCA preserves the euclidean distance between sites
- ▶ Need another measure of similarity for (raw) abundances

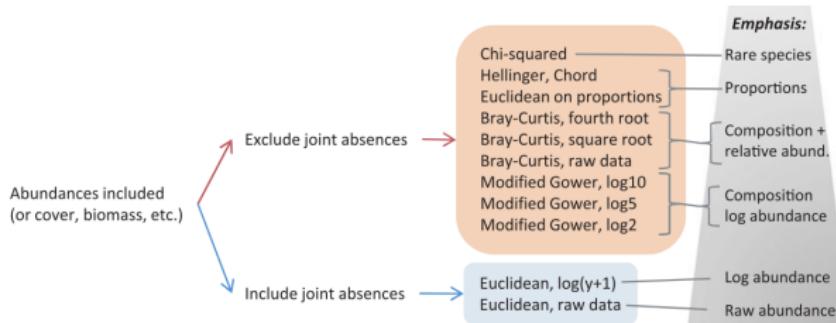
# Dissimilarity measures



---

Anderson, M.J., Crist, T.O., et al. , 2011. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecology Letters* 14, 19–28.

# Dissimilarity measures



	Spe1	Spe2	Spe3
sit1	0	4	8
sit2	0	1	1
sit3	1	0	0

```
vegdist(mat, method = 'euclidean')  
  
          sit1      sit2  
sit2 7.615773  
sit3 9.000000 1.732051
```

```
vegdist(mat, method = 'bray')  
  
          sit1      sit2  
sit2 0.7142857  
sit3 1.0000000 1.0000000
```

Anderson, M.J., Crist, T.O., et al. , 2011. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. Ecology Letters 14, 19–28.

# Principal coordinates analysis (PCoA)

- ▶ Works on distance matrices
- ▶ Species can be added as *weighted averages*
- ▶ Eigenvalue based
- ▶ PCoA with euclidean distance == PCA

Datasets  
○○○○○

Unconstrained Ordination  
○○○○○○○○○○●○○○○○○

Constrained Ordination  
○○

Permutation Tests  
○

# Principal coordinates analysis (PCoA)

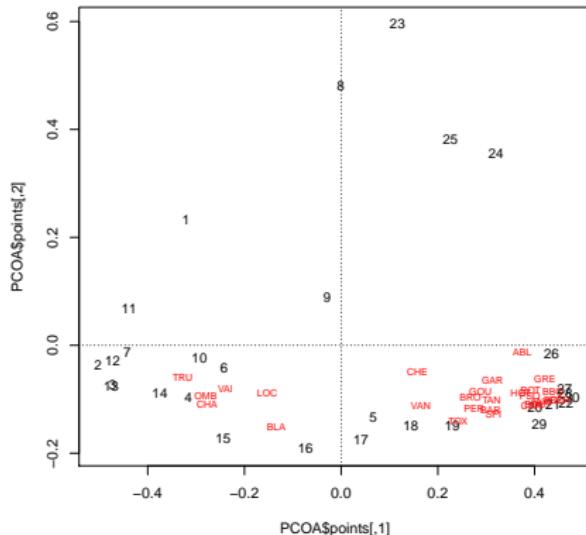
```
# Distance matrix
Dabu_dist <- vegdist(Dabu, method = 'bray')

# PCoA
PCOA <- cmdscale(Dabu_dist, eig = TRUE)

# Create plot
plot(PCOA$points, type = 'n',
      xlab = 'PCOA1', ylab = 'PCOA2')
text(PCOA$points,
     labels = rownames(Dabu), cex = 0.9)
abline(h = 0, lty = 'dotted')
abline(v = 0, lty = 'dotted')

# Add species as weighted averages
wa <- wascores(PCOA$points, Dabu)
text(wa, labels = colnames(Dabu),
     col = 'red', cex = 0.7)
```

```
# explained variance
(PCOA$eig / sum(PCOA$eig))[1:2] * 100
[1] 49.24914 15.95758
```



Datasets  
oooooo

Unconstrained Ordination  
oooooooooooo●ooooo

Constrained Ordination  
○○

Permutation Tests  
○○

# Nonmetric Multidimensional Scaling (NMDS)

- ▶ Similar to PCoA
- ▶ Does not preserve exact distances between objects
- ▶ Possibly better representation in low dimensions
- ▶ **Not** eigenvalue based, iterative algorithm
- ▶ Axes have no meaning, just the relative distances

Datasets  
○○○○○

Unconstrained Ordination  
○○○○○○○○○○○○○○●○○○○

Constrained Ordination  
○○

Permutation Tests  
○

# Nonmetric Multidimensional Scaling (NMDS)

```
# Distance matrix
Dabu_0 <- Dabu[!rowSums(Dabu) == 0, ]
Dabu_dist <- vegdist(Dabu_0, method = 'bray')

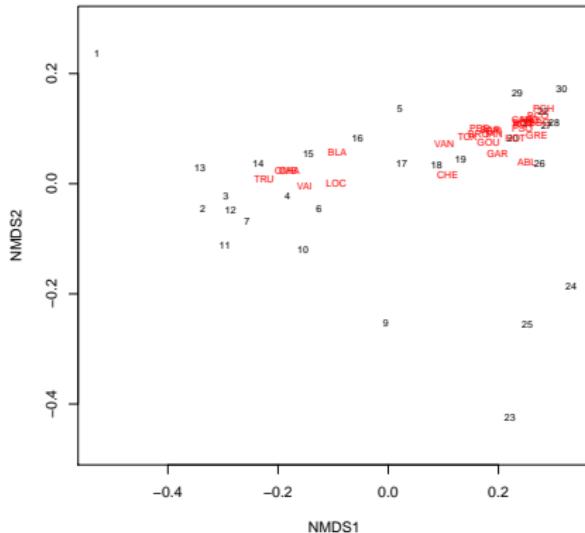
# NMDS
NMDS <- metaMDS(Dabu_dist, k = 2)

# Plot
plot(NMDS, type = 't')

# Add species as weighted averages
wa <- wascores(NMDS$points, Dabu_0)
text(wa, labels = colnames(Dabu),
     col = 'red', cex = 0.7)
```

```
# Stress value
NMDS$stress
```

[1] 0.07376264



# Your turn!

Using the melbourne data.

## Indirect Gradient Analysis

## Datasets

oooooo

## Unconstrained Ordination

## Constrained Ordination

## Permutation Tests

# Your turn!

Using the melbourne data.

# Constrained Ordination

# Constrained Ordination

- ▶ Redundancy analysis (RDA)
- ▶ Transformation-based RDA
- ▶ Distance-based RDA

Datasets  
oooooo

Unconstrained Ordination  
oooooooooooooooooooo

Constrained Ordination  
●○

Permutation Tests  
○

# Your turn!

Using the melbourne data.

# Permutation Tests

# Your turn!

Using the melbourne data.