KICK-OFF MEETING

www.biolaweb.com

Acronym:	BIOLAWEB Boosting Institute of Chemistry, Technology and Metallurgy in Water Biomonitoring
Grant No:	101079234
Type of action:	HORIZON Coordination and Support Actions (HORIZON - CSA)
Starting Date:	01/10/2022
Duration:	36 months

Workshop, Belgrade, October 2023

BIOLAWEB
presentation

Funded by
the European Union

Ecological analysis of metabarcoding data

Beta diversity

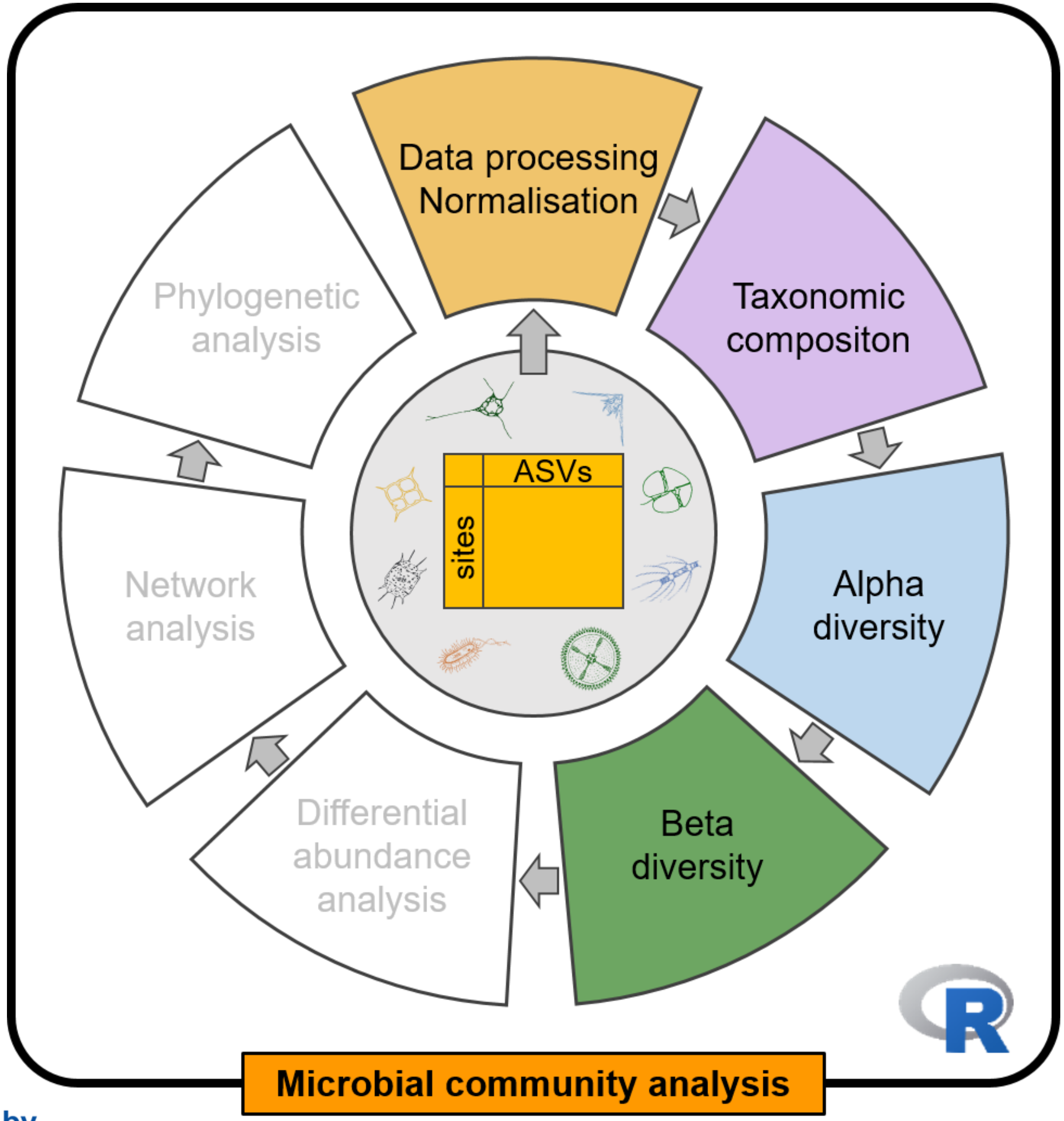
Benjamin Alric

INRAE

ECLA PÔLE ÉCOSYSTÈMES LACUSTRES

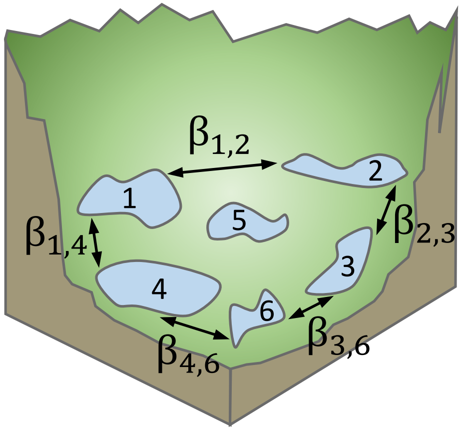
OFB OFFICE FRANÇAIS DE LA BIODIVERSITÉ

Statistical analysis workflow from metabarcoding data



Funded by
the European Union

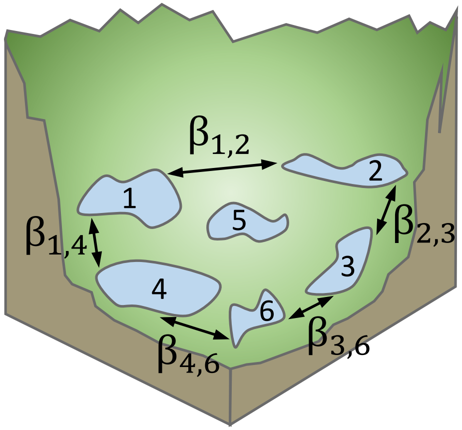
Beta diversity: Analytic workflow



Beta diversity:
Variation in species composition among sites within a geographical area of interest (Whittaker 1960, 1972)

Funded by the European Union

Beta diversity: Analytic workflow

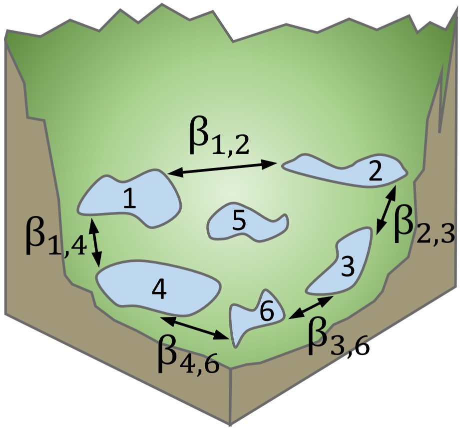


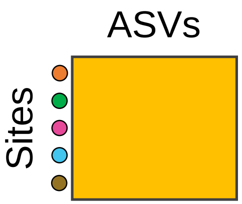
Beta diversity:
Variation in species composition among sites within a geographical area of interest (Whittaker 1960, 1972)

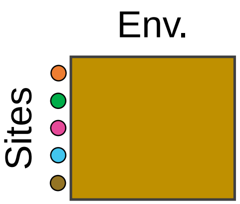
Hypotheses about the processes that generate and maintain biodiversity

Funded by the European Union

Beta diversity: Analytic workflow





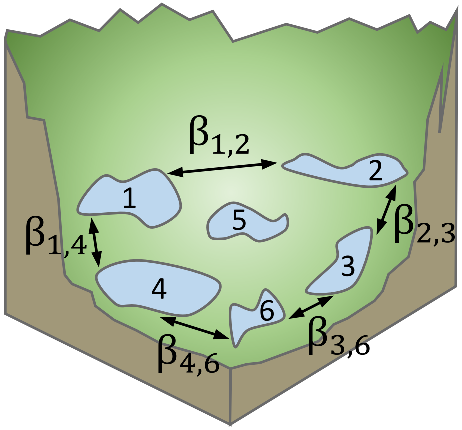


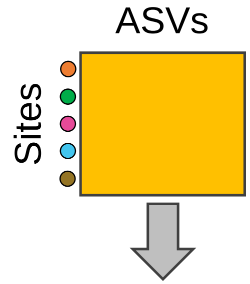
Measuring the variation among commuties

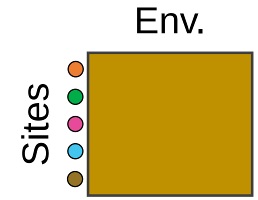
Beta diversity:
Variation in species composition among sites within a geographical area of interest (Whittaker 1960, 1972)

Hypotheses about the processes that generate and maintain biodiversity

Beta diversity: Analytic workflow







Measuring the variation among commuties

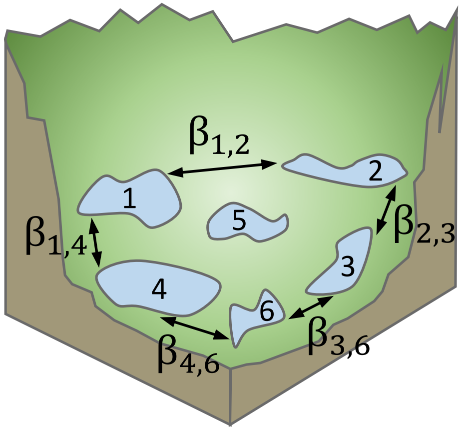
Bringing out a structure

Beta diversity:
Variation in species composition among sites within a geographical area of interest (Whittaker 1960, 1972)

Hypotheses about the processes that generate and maintain biodiversity

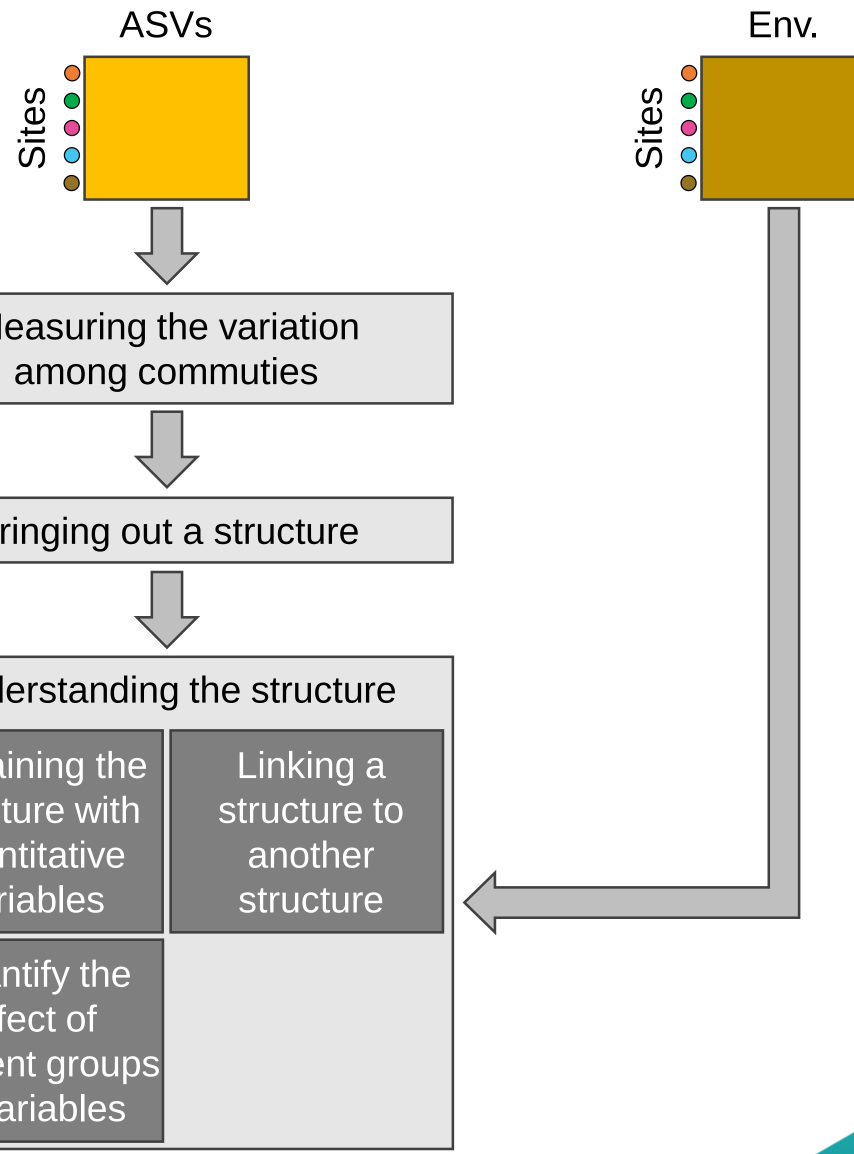
Funded by the European Union

Beta diversity: Analytic workflow



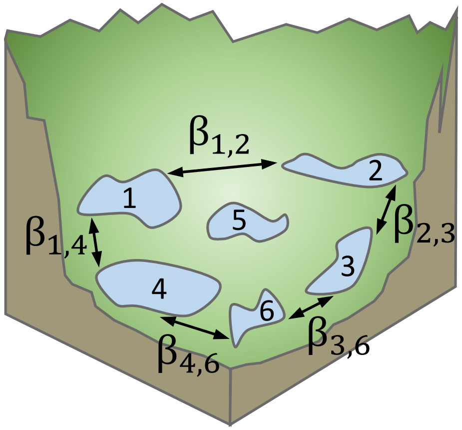
Beta diversity:
Variation in species composition among sites within a geographical area of interest (Whittaker 1960, 1972)

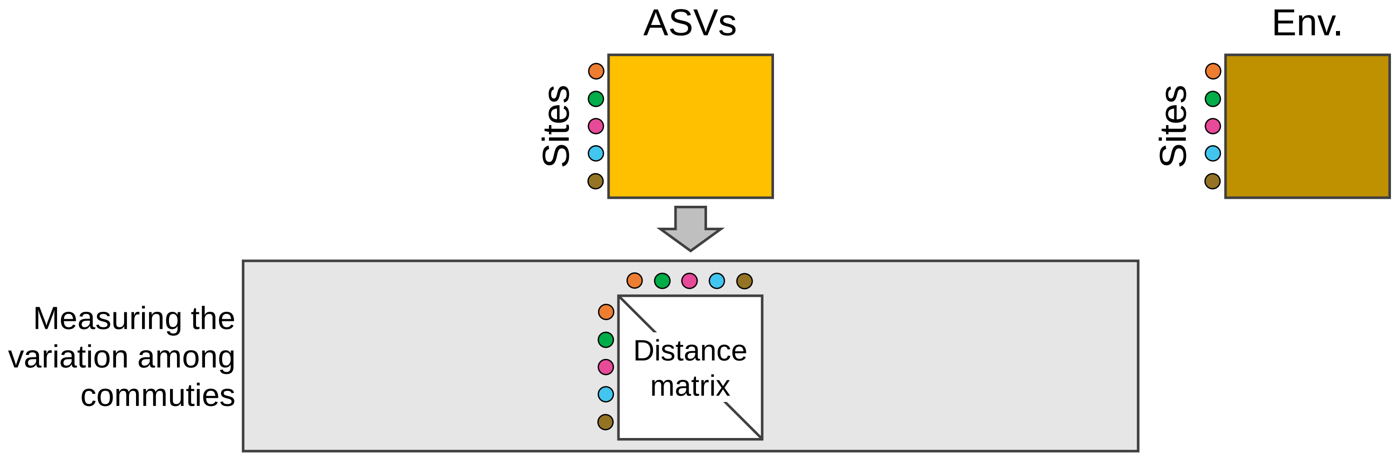
Hypotheses about the processes that generate and maintain biodiversity



Funded by
the European Union

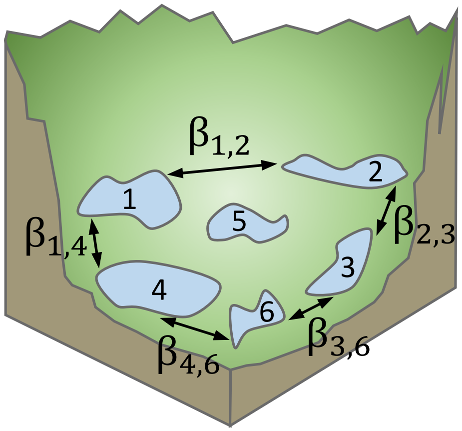
Measuring variation among communities

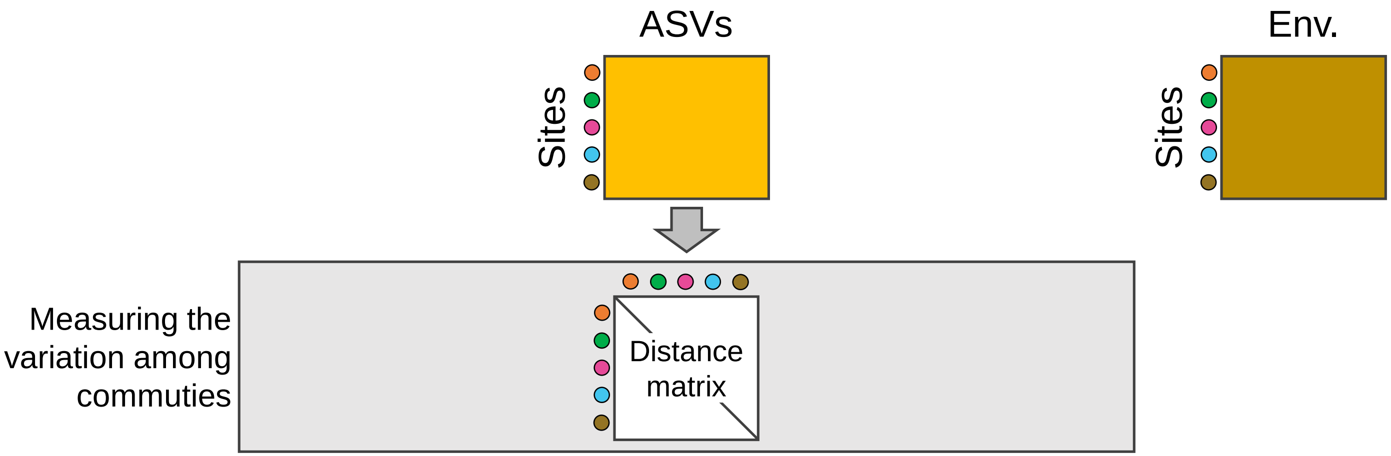




Funded by
the European Union

Measuring variation among communities





> 20 distance metrics available to estimate dissimilarity between communities (Legendre and Legendre 2012)

Bray-Curtis dissimilarity

$$D_{BC}(x_1, x_2) = \frac{\sum_{j=1}^{p} |y_{1j} - y_{2j}|}{\sum_{j=1}^{p} (y_{1j} + y_{2j})}$$

```
library(vegan)
D-BC <— vegdist(tab, method = "bray")
```

Hellinger distance

$$D_H(x_1, x_2) = \sqrt{\sum_{j=1}^{p} \left[\sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}} \right]^2}$$

```
library(vegan)
D-H <— vegdist(tab, method = "hellinger")
```

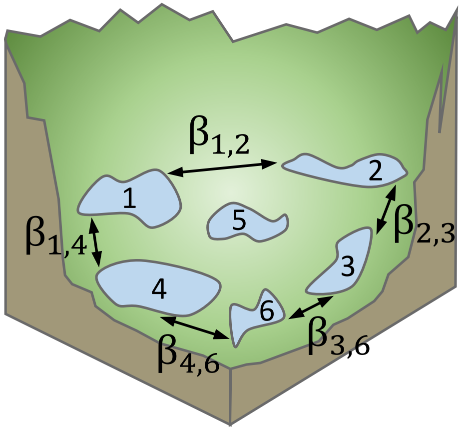
Aitchison distance

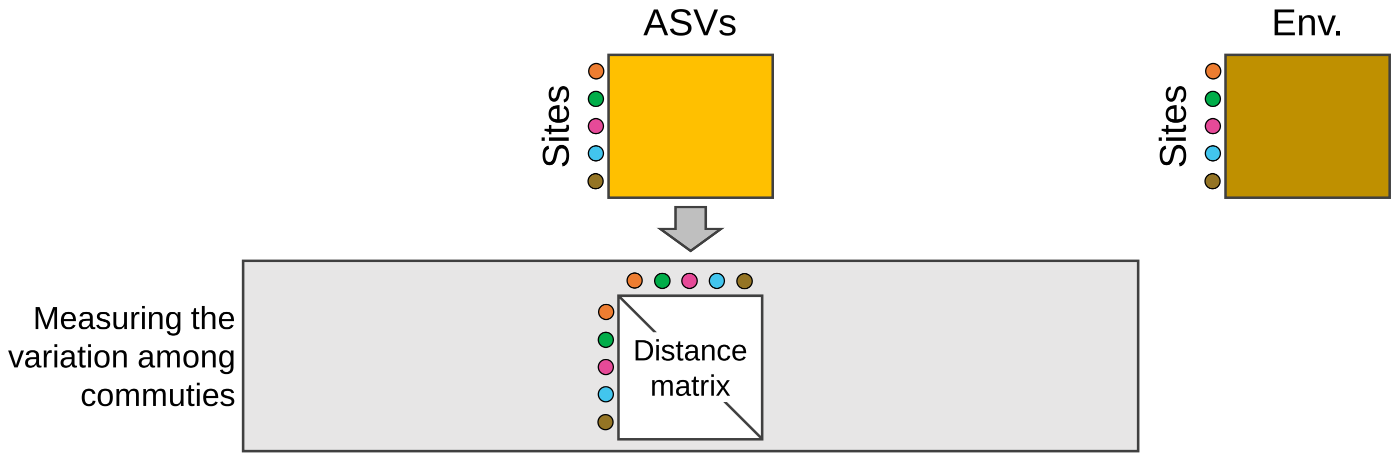
$$D_A(x_1, x_2) = \sqrt{\sum_{j=1}^{D} \left[ln \frac{x_{j1}}{g(x_1)} - ln \frac{x_{j2}}{g(x_2)} \right]^2}$$

```
library(vegan)
D-A <— vegdist(tab, method = "aitchison")
```

Funded by the European Union

Measuring variation among communities





> 20 distance metrics available to estimate dissimilarity between communities (Legendre and Legendre 2012)

Bray-Curtis dissimilarity

$$D_{BC}(x_1, x_2) = \frac{\sum_{j=1}^{p} |y_{1j} - y_{2j}|}{\sum_{j=1}^{p} (y_{1j} + y_{2j})}$$

```
library(vegan)
D-BC <— vegdist(tab, method = "bray")
```

Absolute abundances

Hellinger distance

$$D_H(x_1, x_2) = \sqrt{\sum_{j=1}^{p} \left[\sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}} \right]^2}$$

```
library(vegan)
D-H <— vegdist(tab, method = "hellinger")
```

Relative abundances

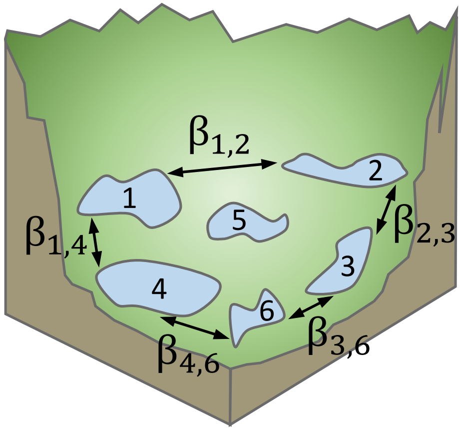
Aitchison distance

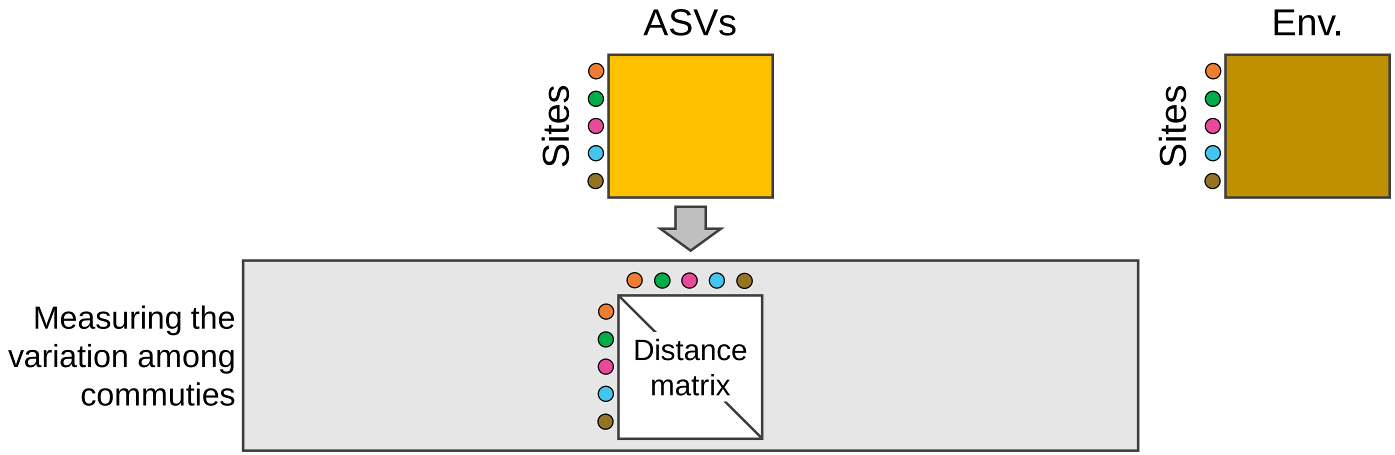
$$D_A(x_1, x_2) = \sqrt{\sum_{j=1}^{D} \left[ln \frac{x_{j1}}{g(x_1)} - ln \frac{x_{j2}}{g(x_2)} \right]^2}$$

```
library(vegan)
D-A <— vegdist(tab, method = "aitchison")
```

Funded by the European Union

Measuring variation among communities





> 20 distance metrics available to estimate dissimilarity between communities (Legendre and Legendre 2012)

Bray-Curtis dissimilarity

$$D_{BC}(x_1, x_2) = \frac{\sum_{j=1}^{p} |y_{1j} - y_{2j}|}{\sum_{j=1}^{p} (y_{1j} + y_{2j})}$$

```
library(vegan)
D-BC <— vegdist(tab, method = "bray")
```

Absolute abundances

Hellinger distance

$$D_H(x_1, x_2) = \sqrt{\sum_{j=1}^{p} \left[\sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}} \right]^2}$$

```
library(vegan)
D-H <— vegdist(tab, method = "hellinger")
```

Relative abundances

Aitchison distance

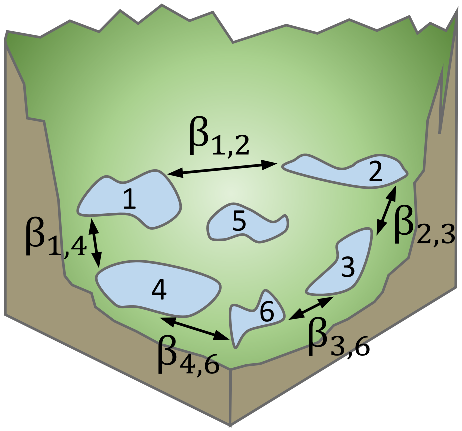
$$D_A(x_1, x_2) = \sqrt{\sum_{j=1}^{D} \left[ln \frac{x_{j1}}{g(x_1)} - ln \frac{x_{j2}}{g(x_2)} \right]^2}$$

```
library(vegan)
D-A <— vegdist(tab, method = "aitchison")
```

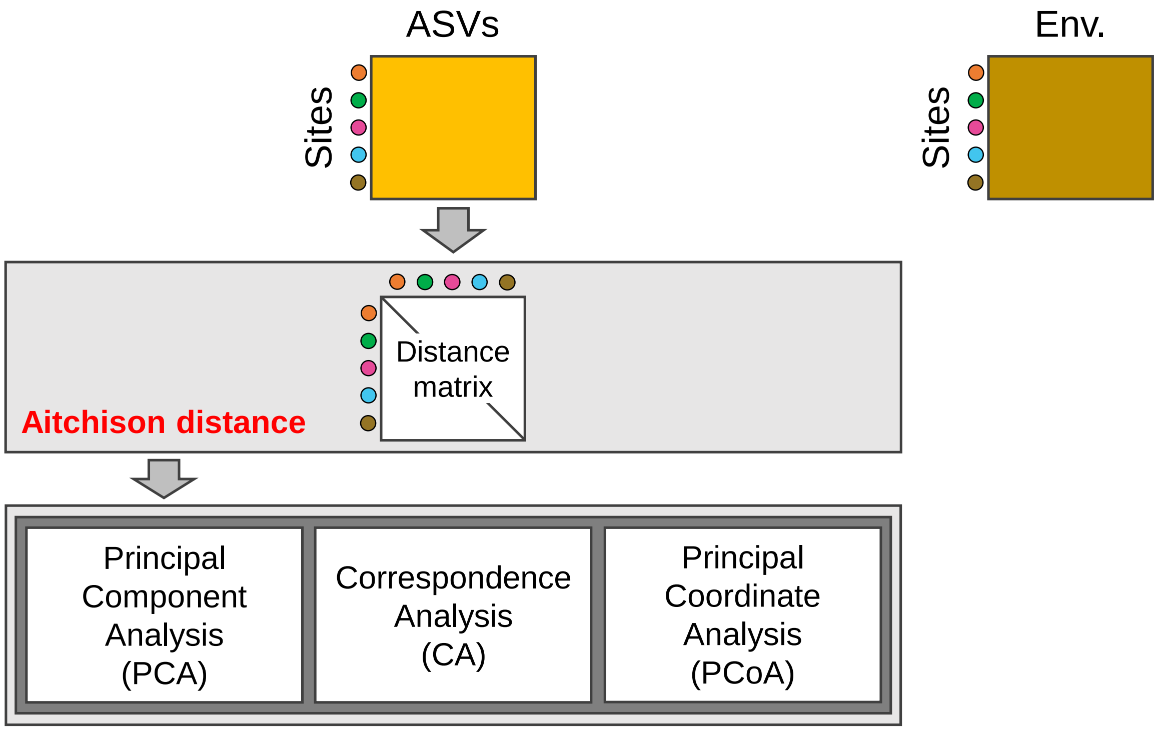
Log-ratio transformation

Funded by the European Union

Bringing out a structure



Measuring the variation among commuties

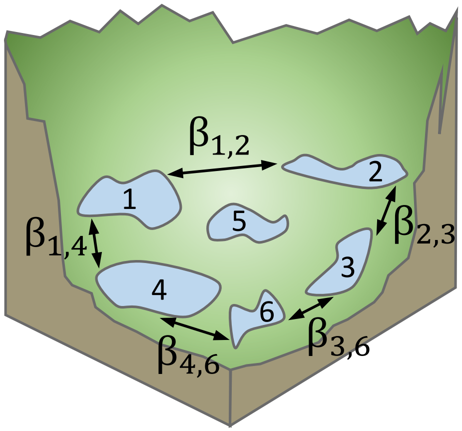


Unconstrained ordinations

Bringing out a structure

Funded by the European Union

Bringing out a structure



Measuring the variation among commuties

Distance matrix

Aitchison distance

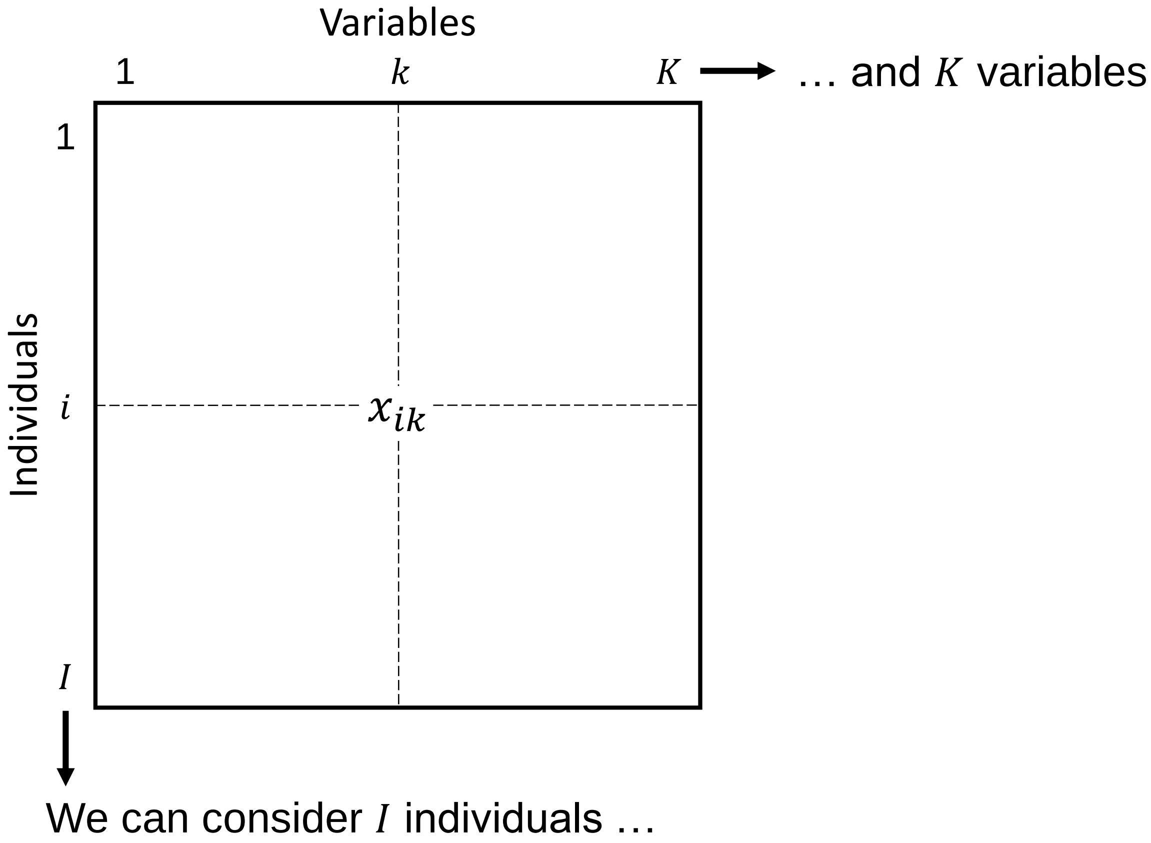
Bringing out a structure

Principal Component Analysis (PCA)	Correspondence Analysis (CA)	Principal Coordinate Analysis (PCoA)

Unconstrained ordinations

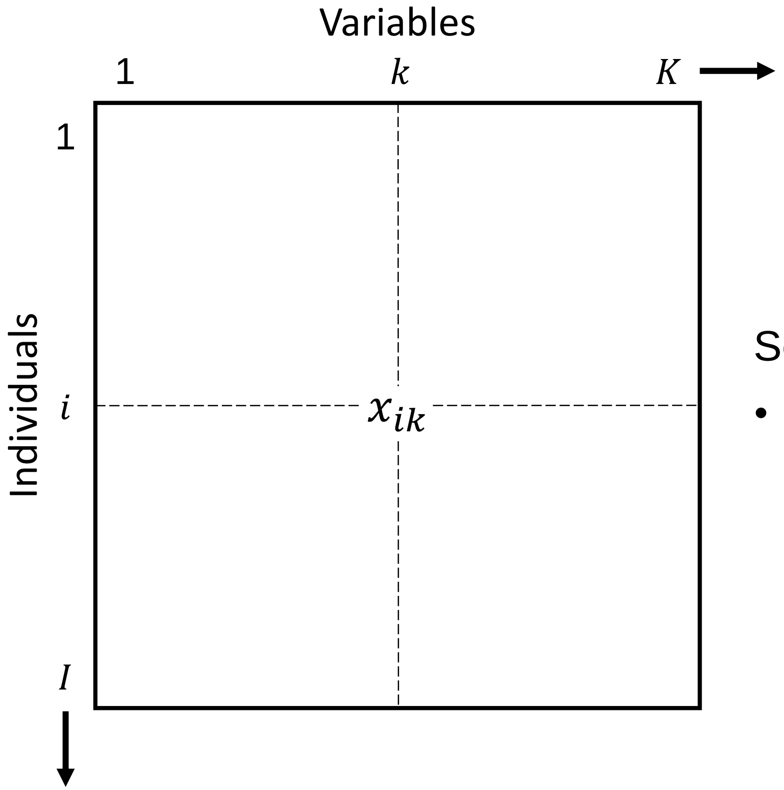
Funded by the European Union

Bringing out a structure – PCA – type of data



Funded by
the European Union

Bringing out a structure – PCA – type of data



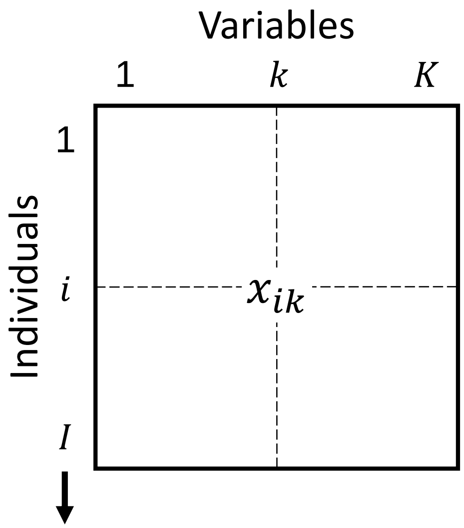
… and K variables

We can consider I individuals …

Several fields:

- Ecology: abundance of species k in lake i

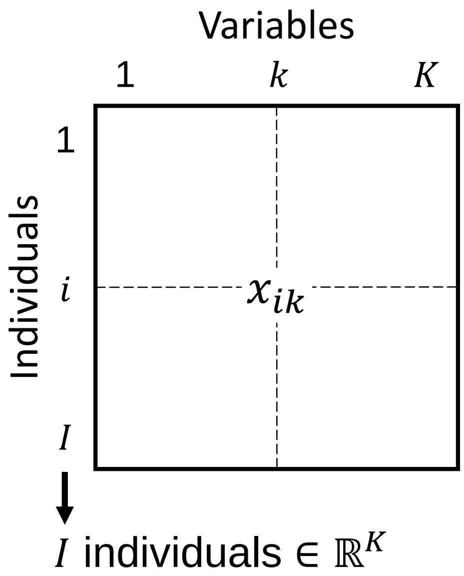
Bringing out a structure – PCA – problems and objectives



I individuals $\in \mathbb{R}^K$

Funded by
the European Union

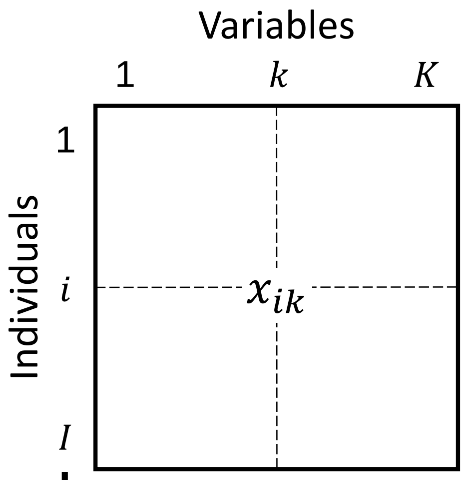
Bringing out a structure – PCA – problems and objectives



	Sp. 1	Sp. 2	Sp. 3	Sp. 4
Lake 1	12	8	14	7
Lake 2	13	6	11	9
Lake 3	10	7	12	8
Lake 4	3	5	4.5	7
Lake 5	1	4.8	3.3	6
Lake 6	2	3	4	9

Funded by
the European Union

Bringing out a structure – PCA – problems and objectives

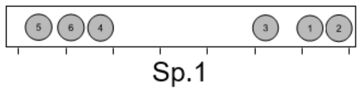


I individuals $\in \mathbb{R}^K$

	Sp. 1	Sp. 2	Sp. 3	Sp. 4
Lake 1	12	8	14	7
Lake 2	13	6	11	9
Lake 3	10	7	12	8
Lake 4	3	5	4.5	7
Lake 5	1	4.8	3.3	6
Lake 6	2	3	4	9

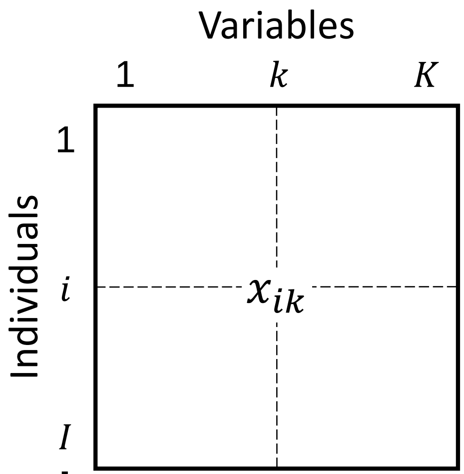
$K = 1$

If K = 1:
axial representation



Funded by
the European Union

Bringing out a structure – PCA – problems and objectives

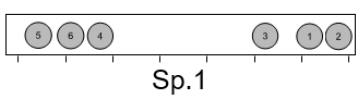


I individuals $\in \mathbb{R}^K$

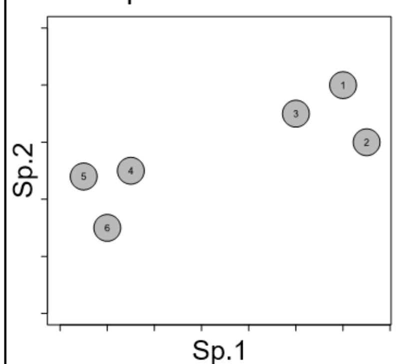
	Sp. 1	Sp. 2	Sp. 3	Sp. 4
Lake 1	12	8	14	7
Lake 2	13	6	11	9
Lake 3	10	7	12	8
Lake 4	3	5	4.5	7
Lake 5	1	4.8	3.3	6
Lake 6	2	3	4	9

K = 2

If K = 1:
axial representation

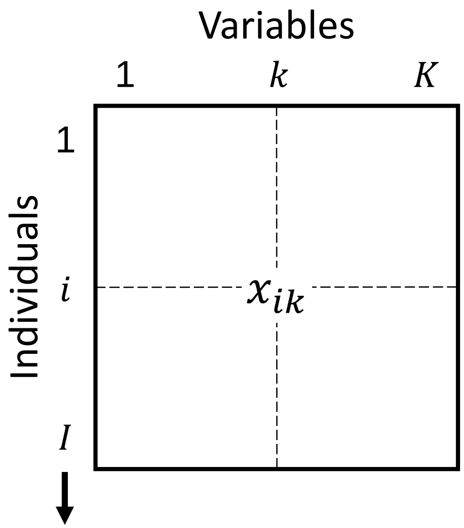


If K = 2:
scatter plot



Funded by
the European Union

Bringing out a structure – PCA – problems and objectives

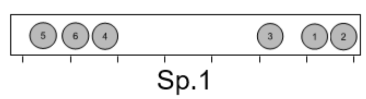


I individuals $\in \mathbb{R}^K$

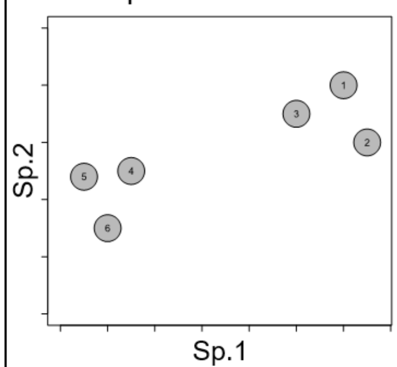
	Sp. 1	Sp. 2	Sp. 3	Sp. 4
Lake 1	12	8	14	7
Lake 2	13	6	11	9
Lake 3	10	7	12	8
Lake 4	3	5	4.5	7
Lake 5	1	4.8	3.3	6
Lake 6	2	3	4	9

$K = 3$

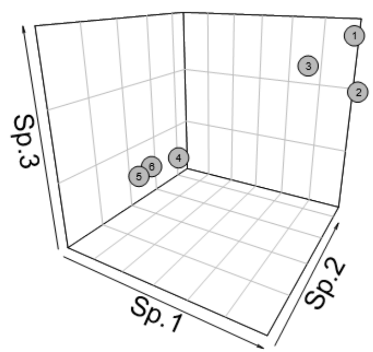
If K = 1:
axial representation



If K = 2:
scatter plot

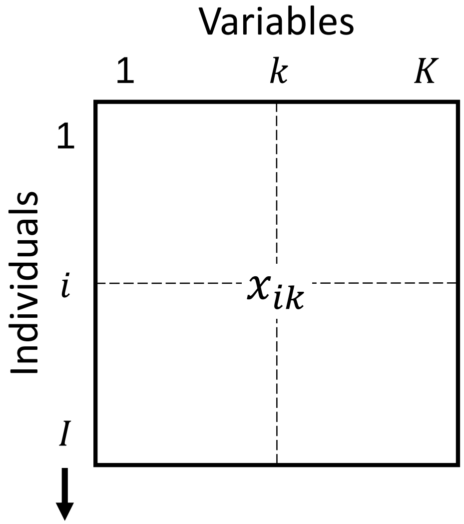


If K = 3:
3-dimensional graph



Funded by
the European Union

Bringing out a structure – PCA – problems and objectives

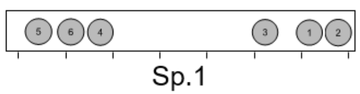


I individuals $\in \mathbb{R}^K$

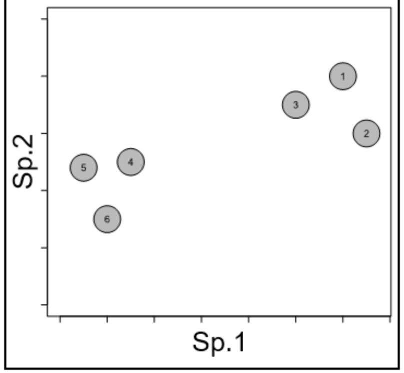
	Sp. 1	Sp. 2	Sp. 3	Sp. 4
Lake 1	12	8	14	7
Lake 2	13	6	11	9
Lake 3	10	7	12	8
Lake 4	3	5	4.5	7
Lake 5	1	4.8	3.3	6
Lake 6	2	3	4	9

$K = 4$

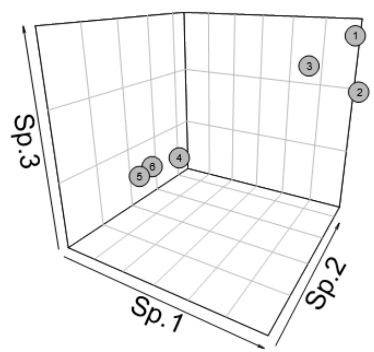
If $K = 1$:
axial representation



If $K = 2$:
scatter plot



If $K = 3$:
3-dimensional graph



If $K \geq 4$:

Impossible to represent or even imagine, BUT simple mathematical concept: K coordinates on K dimensions

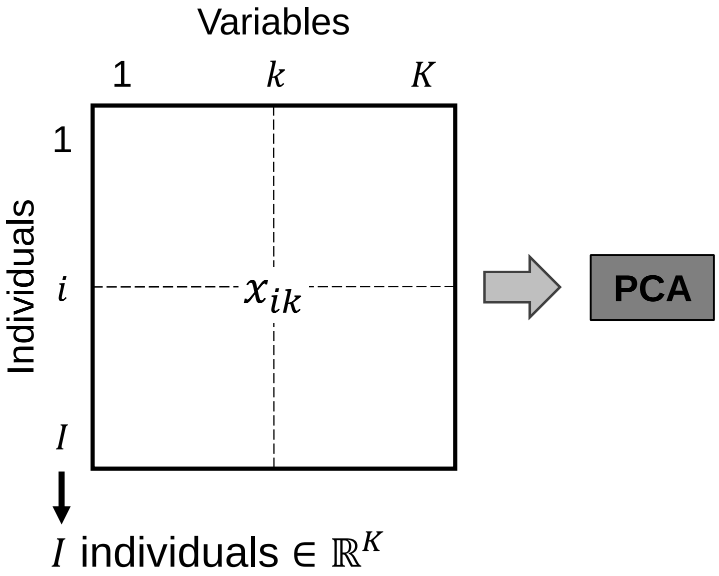
How can we visualize a scatterplot, if it evolves in a very large space ($K \geq 4$)

Funded by the European Union

Bringing out a structure – PCA – problems and objectives

Variables: species

Individuals: lakes

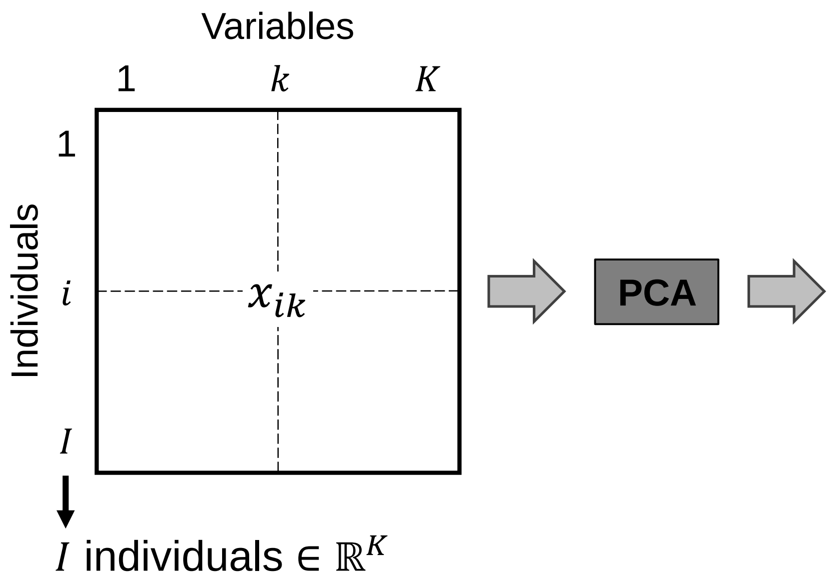


Funded by
the European Union

Bringing out a structure – PCA – problems and objectives

Variables: species

Individuals: lakes



- Exploratory statistics, multidimensional descriptive statistics
- Synthetize, summarize and rank information contained in a table (high dimension, $K \geq 4$)
- Visualize data tables using simple grapics (2D, 3D)

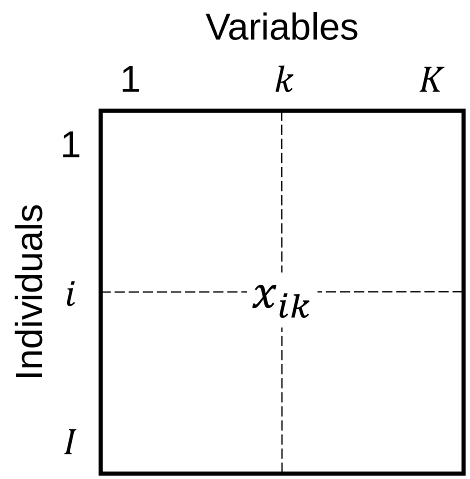
PCA ≡ Analyze a table with K dimensions (or variables, ≥ 4) measured on I individuals

Funded by
the European Union

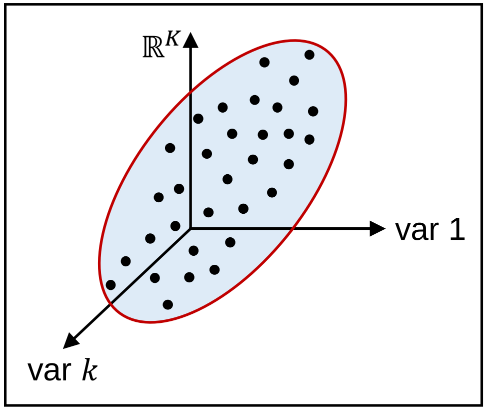
Bringing out a structure – PCA – cloud of individuals

Variables: species

Individuals: lakes



A set of lines …



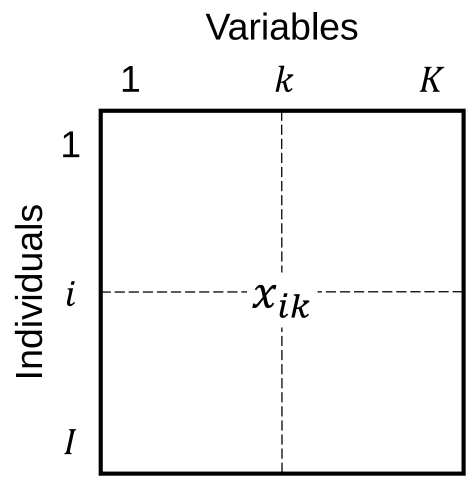
i point-lines $\in \mathbb{R}^K$

Funded by the European Union

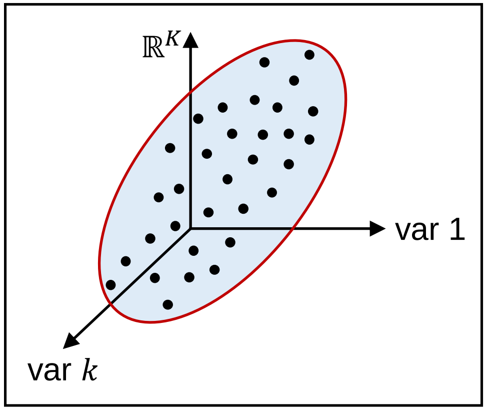
Bringing out a structure – PCA – cloud of individuals

Variables: species

Individuals: lakes



A set of lines …



i point-lines $\in \mathbb{R}^K$

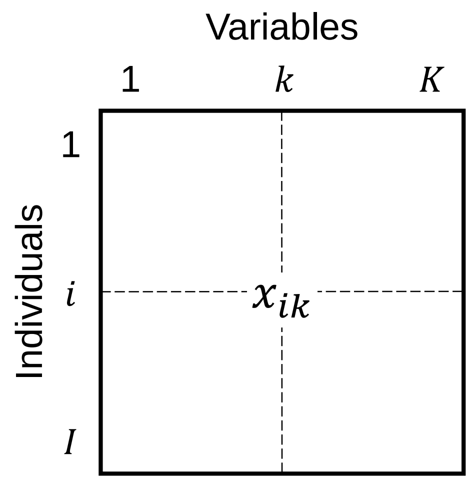
- Similarity between individuals as a function of variables
- Groups of individuals

Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)

Variables: species

Individuals: lakes



i point-lines $\in \mathbb{R}^K$

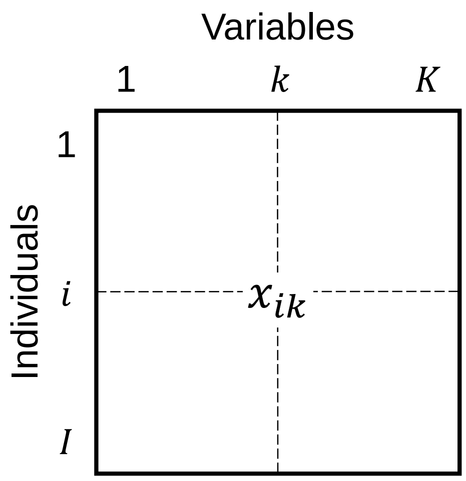
Notion of **similarity** between individuals i and i'.

Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)

Variables: species

Individuals: lakes



i point-lines $\in \mathbb{R}^K$

Notion of **similarity** between individuals i and i'.

$$d^2(i,i') = \sum_{k=1}^{K} (x_{ik} - x_{i'k})^2$$ [Pythagoras]

i: individual 1

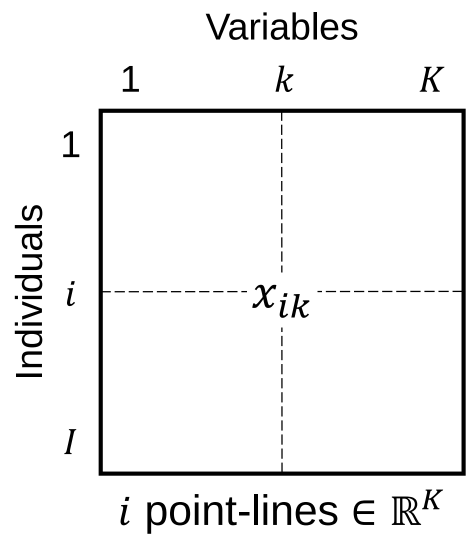
i': individual 2

Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)

Variables: species

Individuals: lakes



i: individual 1

i': individual 2

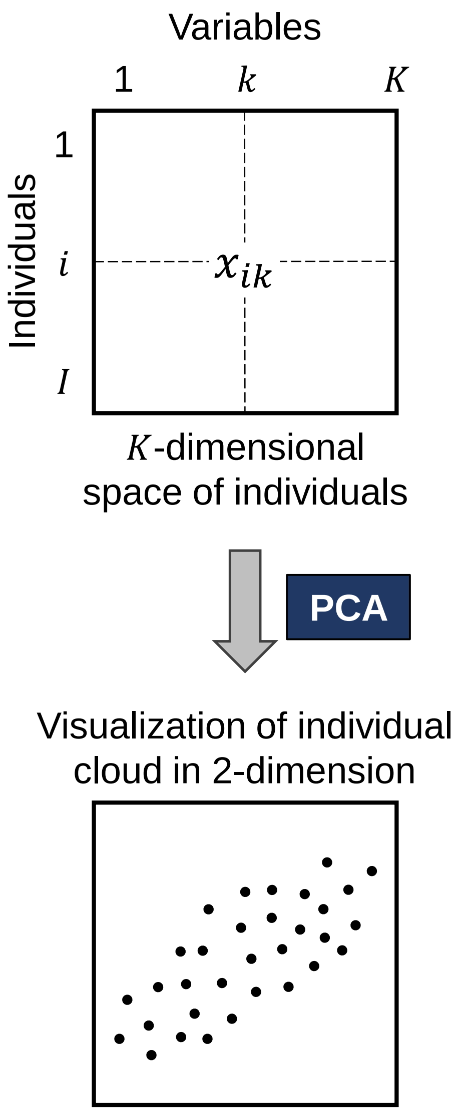
Notion of **similarity** between individuals i and i'.

$$d^2(i,i') = \sum_{k=1}^{K} (x_{ik} - x_{i'k})^2 \quad \text{[Pythagoras]}$$

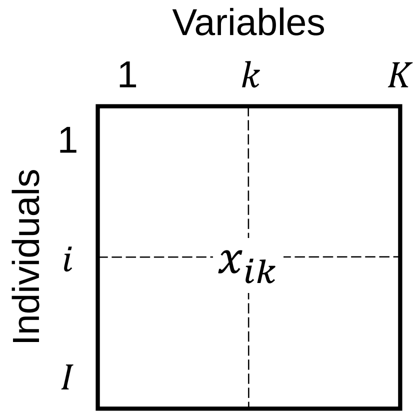
Study of individuals $\equiv$ Study of distance between individuals $\equiv$ Study of the shape of cloud N^I

Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)



Bringing out a structure – PCA – study of individuals (N^I)



K-dimensional space of individuals

PCA

Visualization of individual cloud in 2-dimension

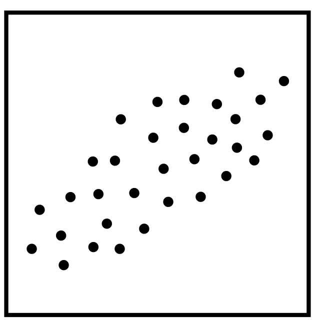
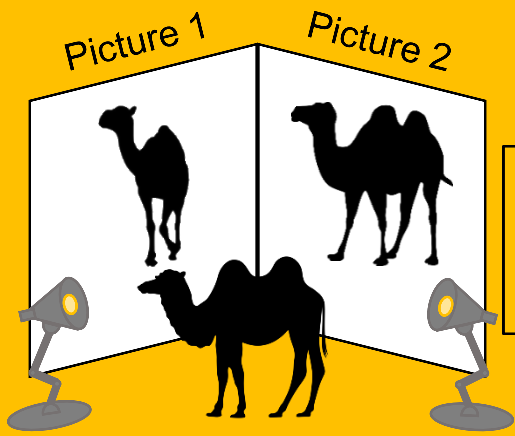


Image quality

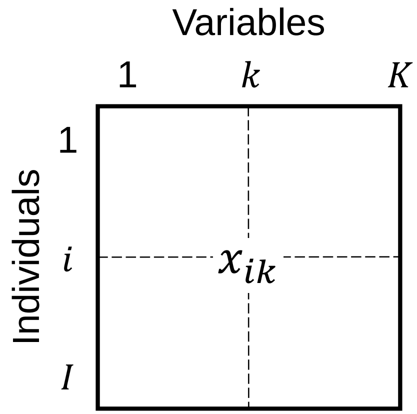


PCA

- Accurately reproduces cloud shape
- Idea of distance between idividuals
- Distances as close as possible

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)



K-dimensional space of individuals

PCA

Visualization of individual cloud in 2-dimension

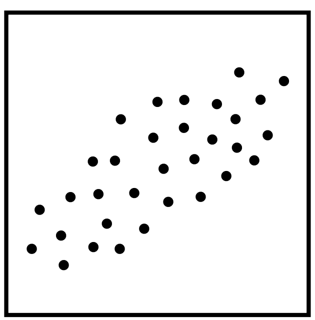
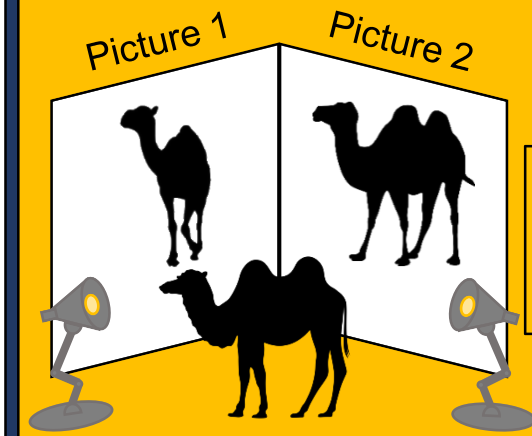
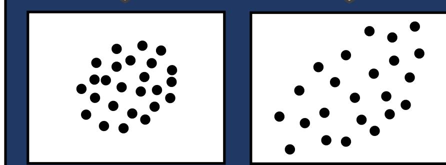


Image quality



PCA

- Accurately reproduces cloud shape
- Idea of distance between idividuals
- Distances as close as possible



Quantify quality

Low dispersion → Poor-fitted picture

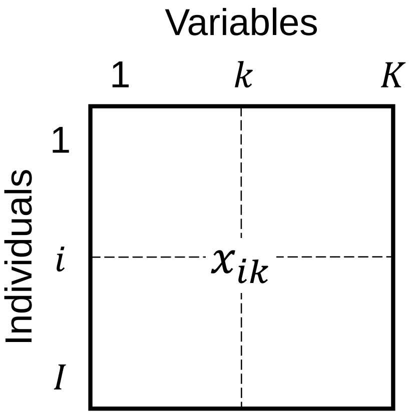
High dispersion → Well-fitted picture

Notion of dipersion ≡ Variability ≡ Inertia (several dimensions)

High dispersion ≡ High variability ≡ Well-fitted picture

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)



K-dimensional space of individuals

PCA

Visualization of individual cloud in 2-dimension

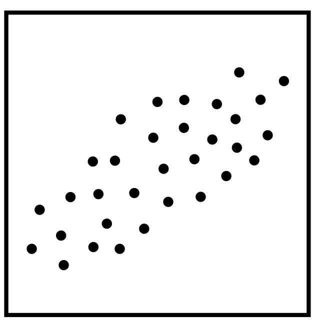
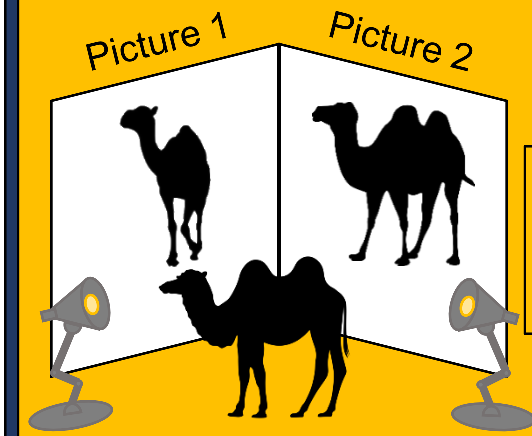
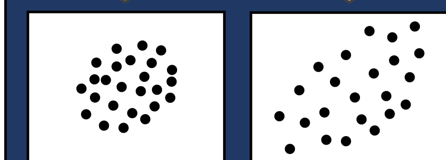


Image quality



PCA

- Accurately reproduces cloud shape
- Idea of distance between idividuals
- Distances as close as possible



Quantify quality

Low dispersion → Poor-fitted picture

High dispersion → Well-fitted picture

Notion of dipersion ≡ Variability ≡ Inertia (several dimensions)

High dispersion ≡ High variability ≡ Well-fitted picture

Objective

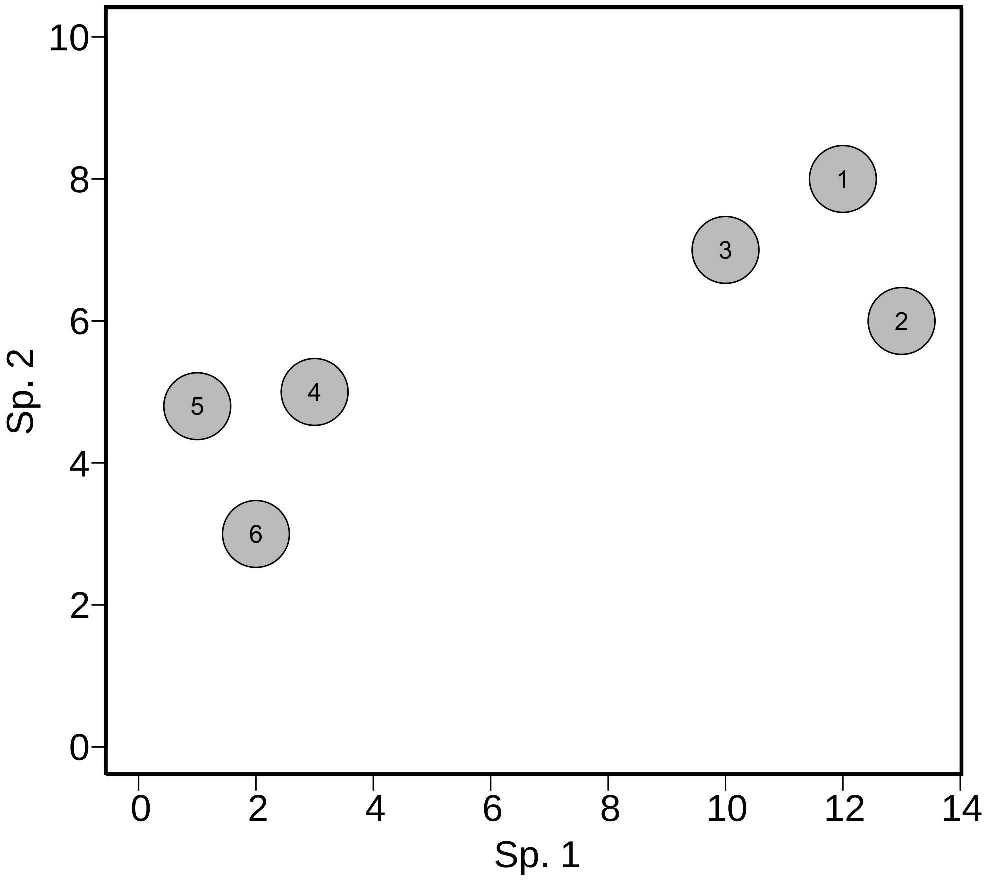
Search for a subspace that best summarizes the data (i.e., principal components)

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)

We start by plotting the data …

	Sp. 1	Sp. 2
Lake 1	12	8
Lake 2	13	6
Lake 3	10	7
Lake 4	3	5
Lake 5	1	4.8
Lake 6	2	3

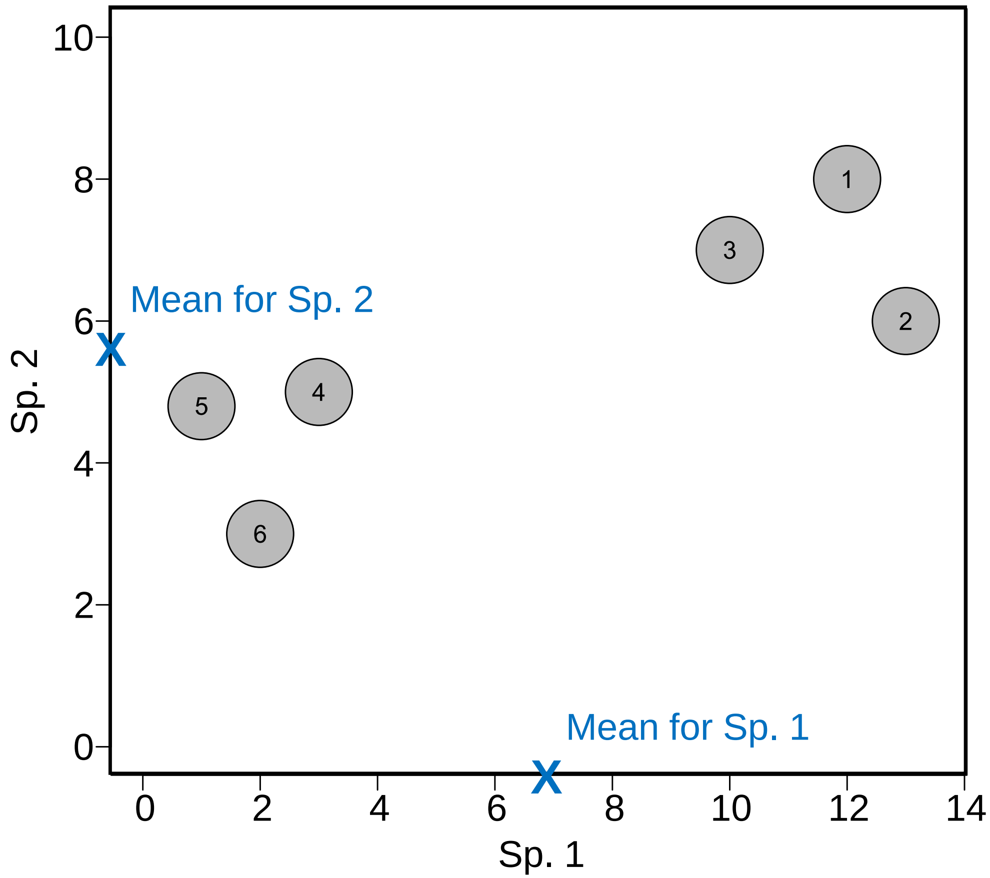


Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)

	Sp. 1	Sp. 2
Lake 1	12	8
Lake 2	13	6
Lake 3	10	7
Lake 4	3	5
Lake 5	1	4.8
Lake 6	2	3

We start by plotting the data …



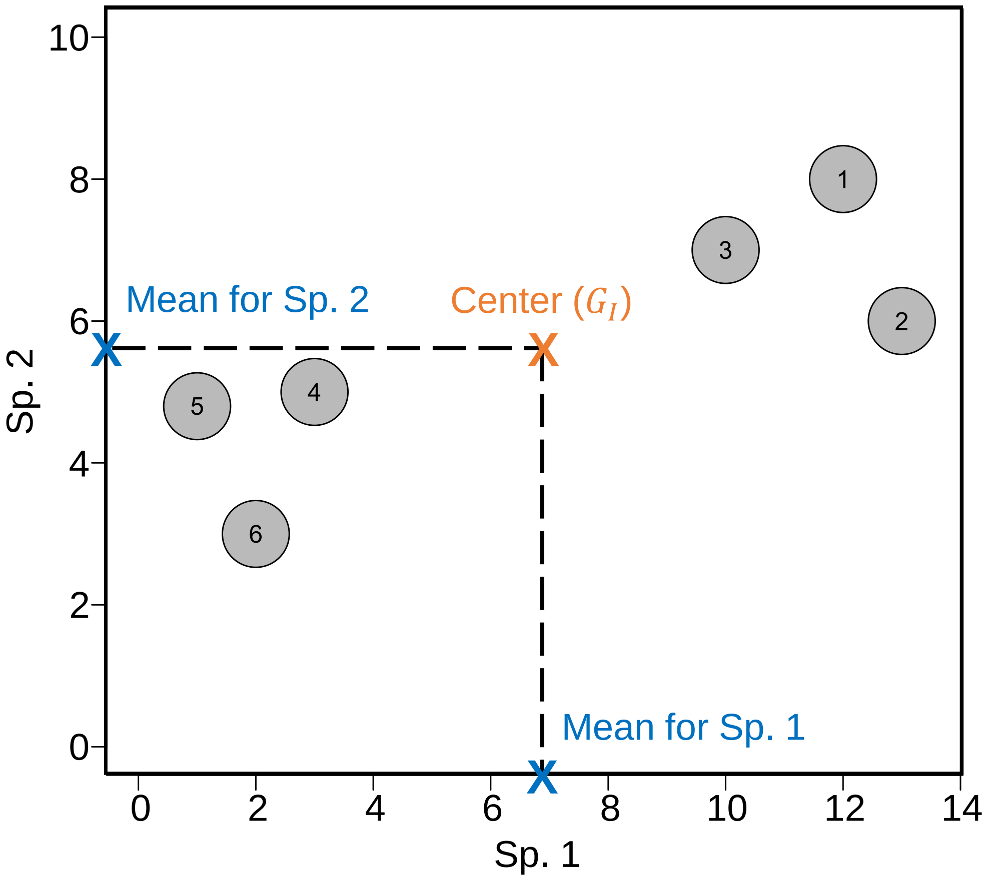
Then we calculate the average values for Sp. 1 and Sp. 2 …

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)

We start by plotting the data …

	Sp. 1	Sp. 2
Lake 1	12	8
Lake 2	13	6
Lake 3	10	7
Lake 4	3	5
Lake 5	1	4.8
Lake 6	2	3

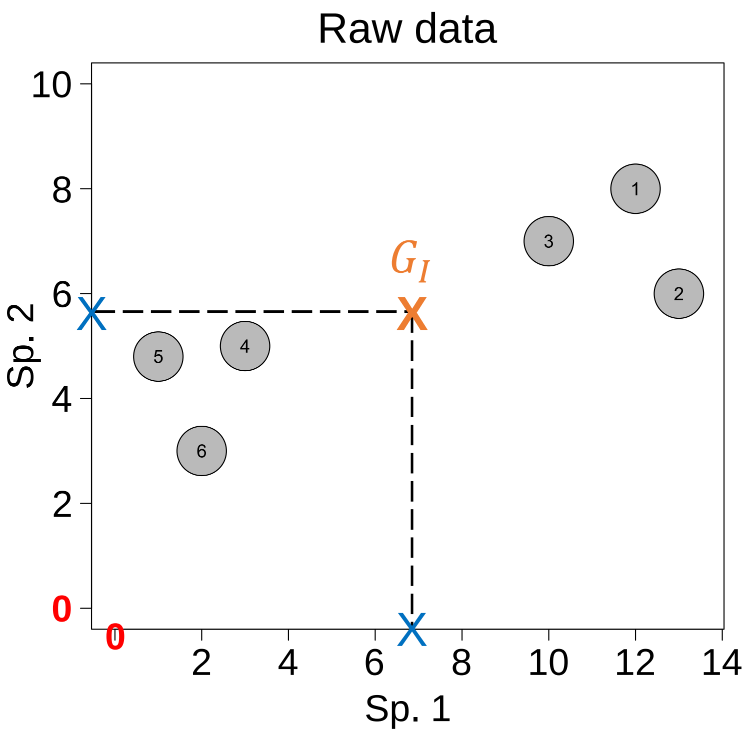


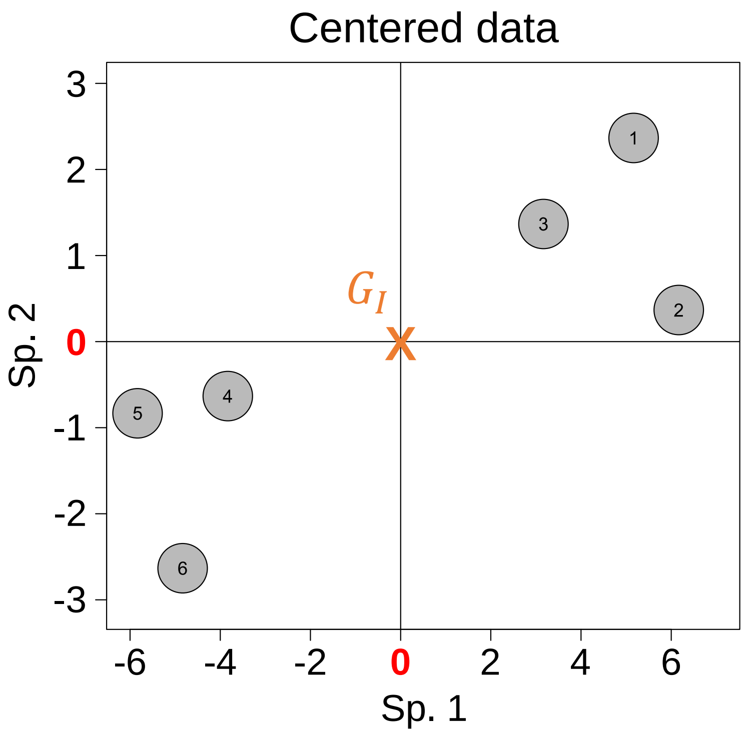
Then we calculate the average values for Sp. 1 and Sp. 2 …

With the average values, we can calculate the center of the data.

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)

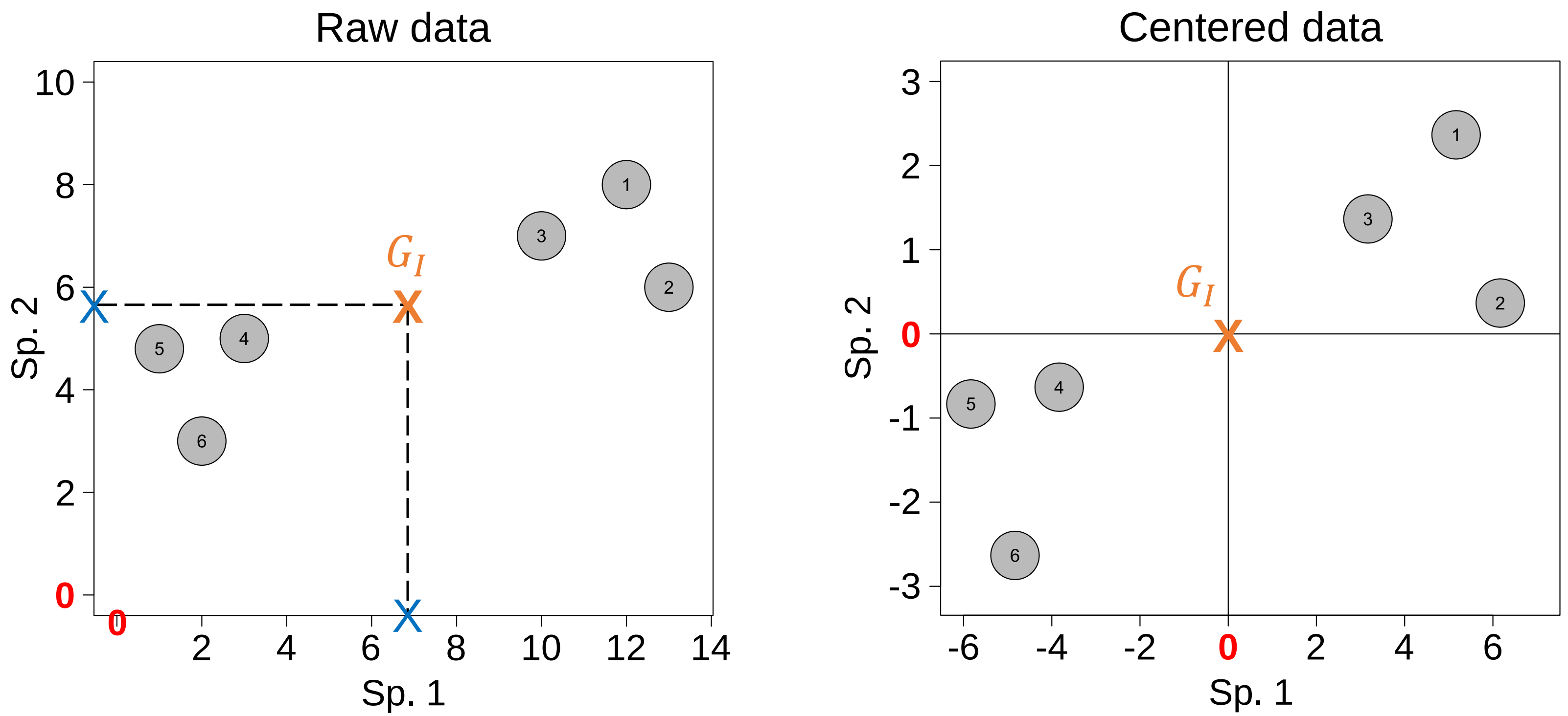




Then we centered the data … ⟶ $y_{ik} = x_{ik} - \overline{x}_{ik}$

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)

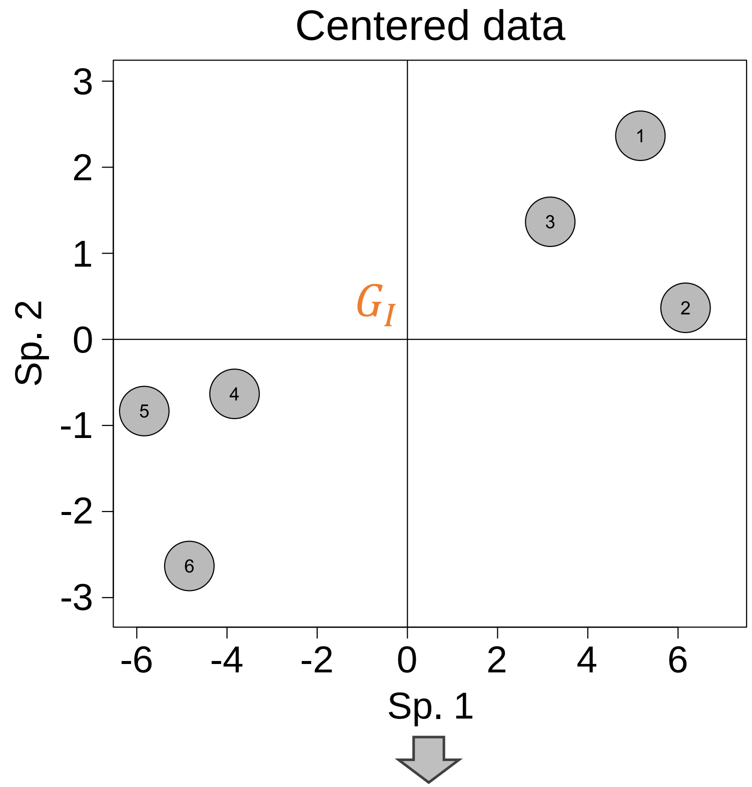


Then we centered the data … ⟶ $y_{ik} = x_{ik} - \overline{x}_{ik}$

NOTE 1: Centered ≡ Translation of cloud ⟶ No change in shape of cloud

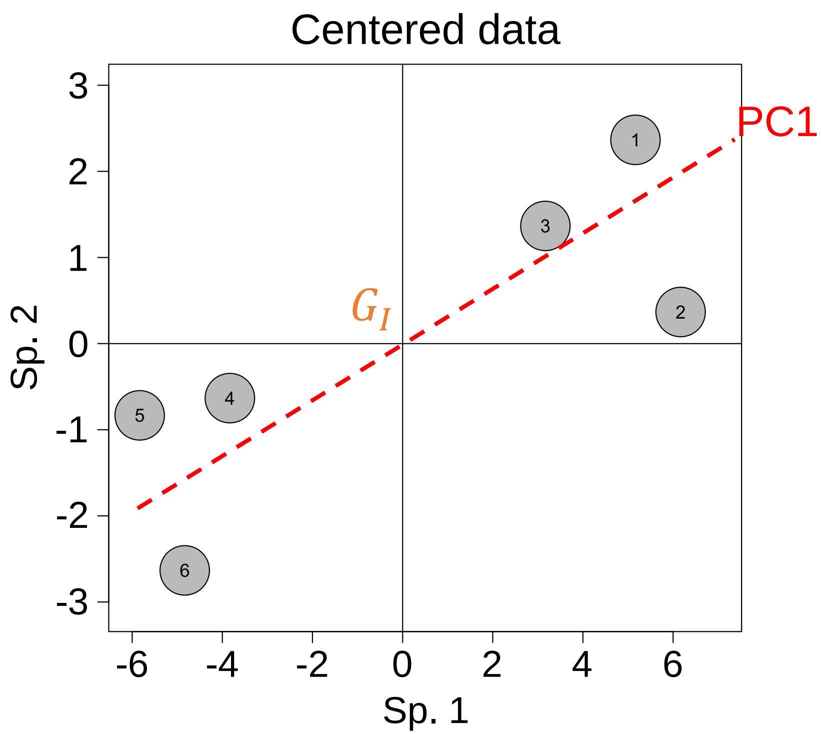
Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)



Subspace providing the best approximate picture of the cloud of individuals

Bringing out a structure – PCA – study of individuals (N^I)

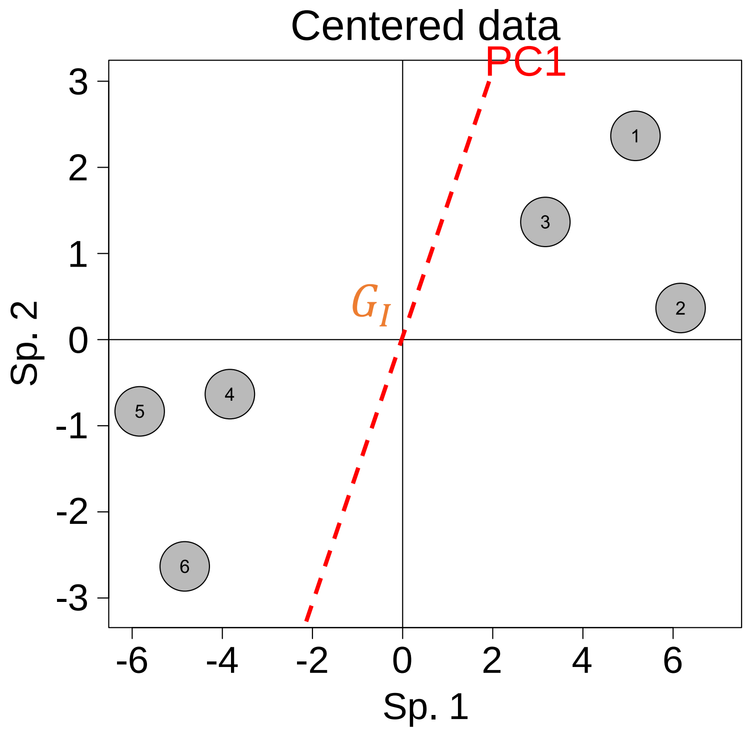


Subspace providing the best approximate picture of the cloud of individuals

- Find the first axis (PC1) goes through the origin (0, 0) in the K-dimensional variable space

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)

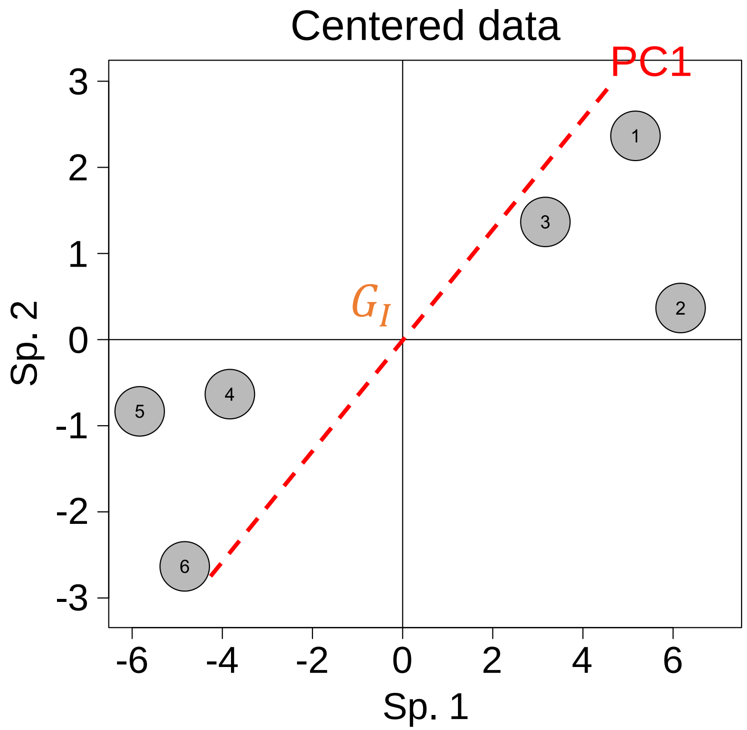


Subspace providing the best approximate picture of the cloud of individuals

- Find the first axis (PC1) goes through the origin (0, 0) in the K-dimensional variable space
- Best possible fit to the data, in the least-squares sense

Funded by
the European Union

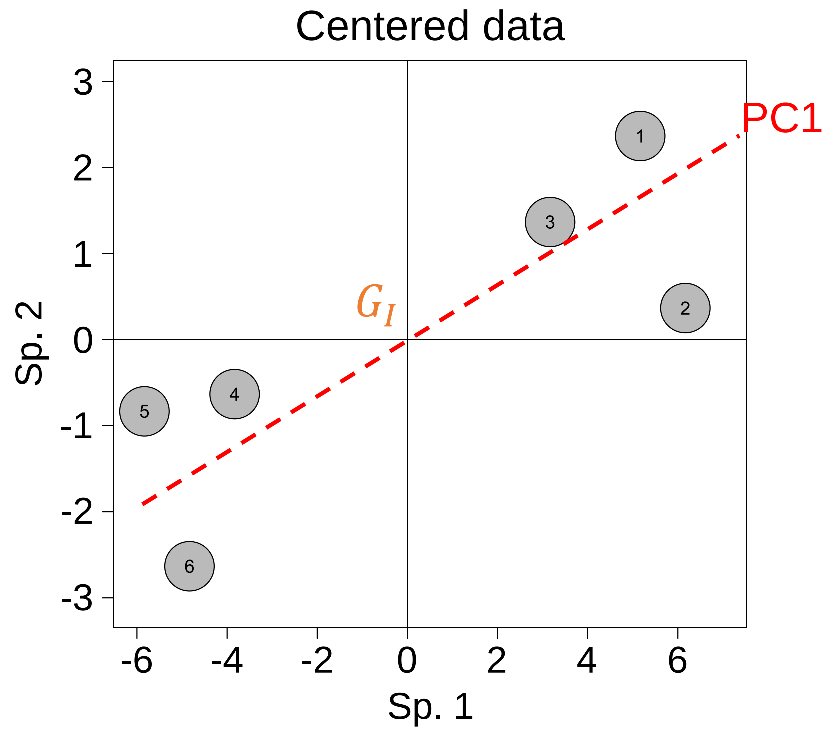
Bringing out a structure – PCA – study of individuals (N^I)



Subspace providing the best approximate picture of the cloud of individuals

- Find the first axis (PC1) goes through the origin (0, 0) in the K-dimensional variable space
- Best possible fit to the data, in the least-squares sense

Bringing out a structure – PCA – study of individuals (N^I)

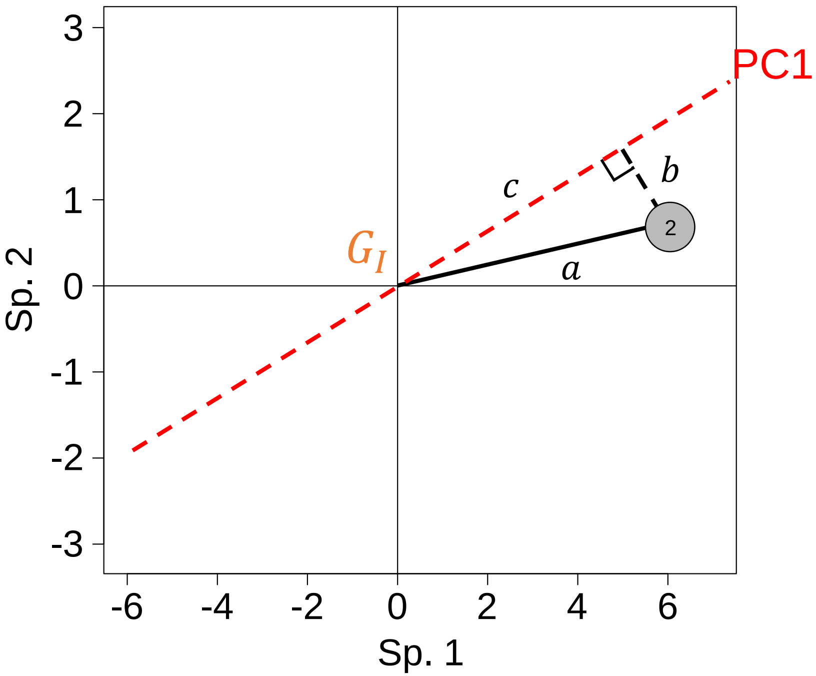


Subspace providing the best approximate picture of the cloud of individuals

- Find the first axis (PC1) goes through the origin (0, 0) in the K-dimensional variable space
- Best possible fit to the data, in the least-squares sense

Funded by the European Union

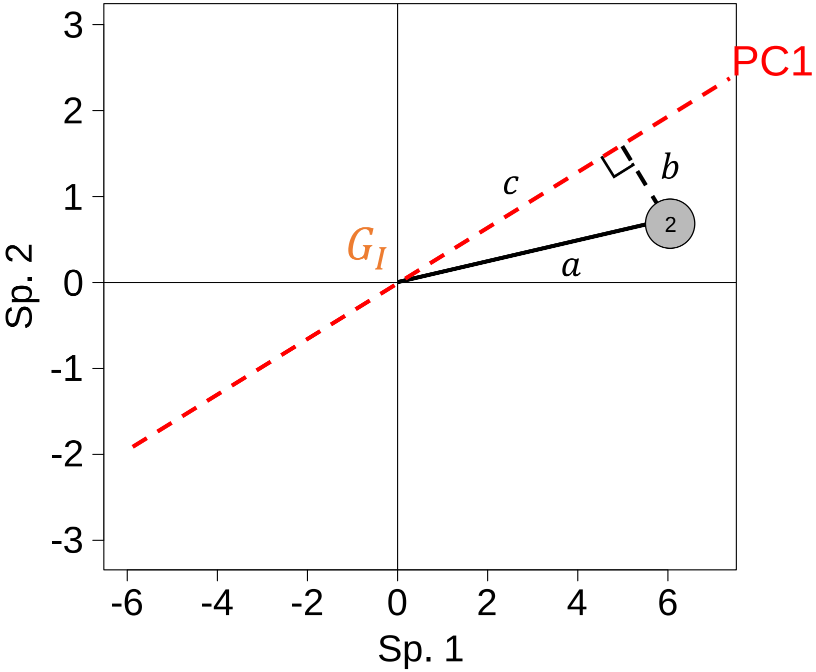
Bringing out a structure – PCA – study of individuals (N^I)



- $a \longrightarrow$ distance from the point to the origin
- $b \longrightarrow$ distance from the point to its projection on the axis
- $c \longrightarrow$ distance from the projection to the origin

Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)

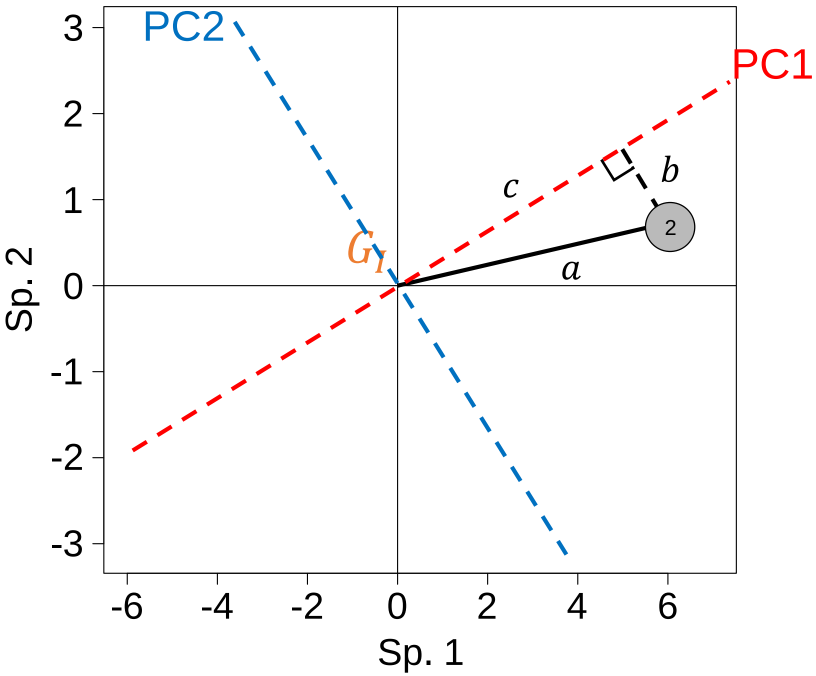


- $a \longrightarrow$ constant distance
- b is small <=> c is large

  $a^2 = b^2 + c^2$ [Pythagoras]
- Minimizing b (or b^2) ≡ Maximizing c (or c^2)
- $I_{total} = I_{PC1} + I_{residual}$

Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)



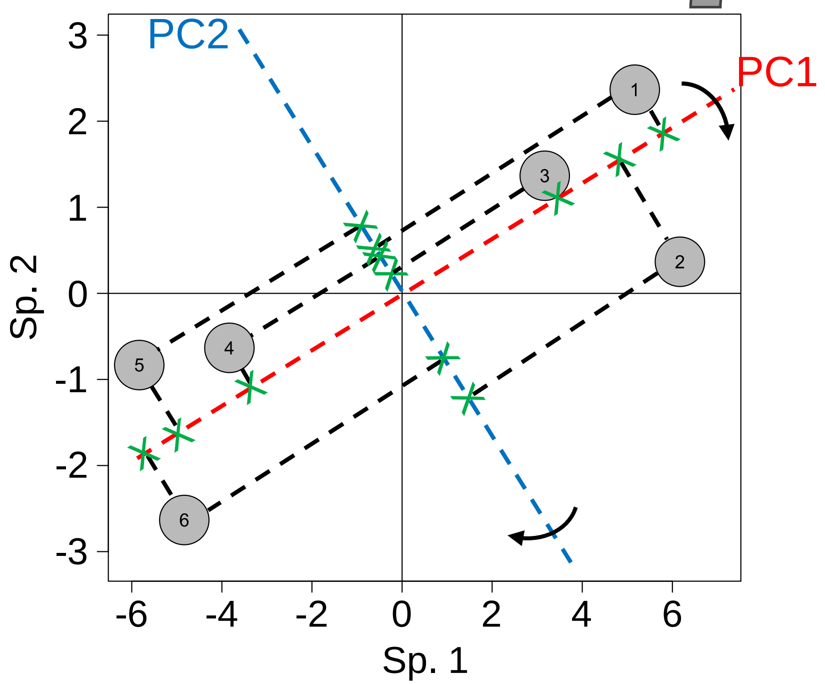
- We look for PC2 ⊥ PC1

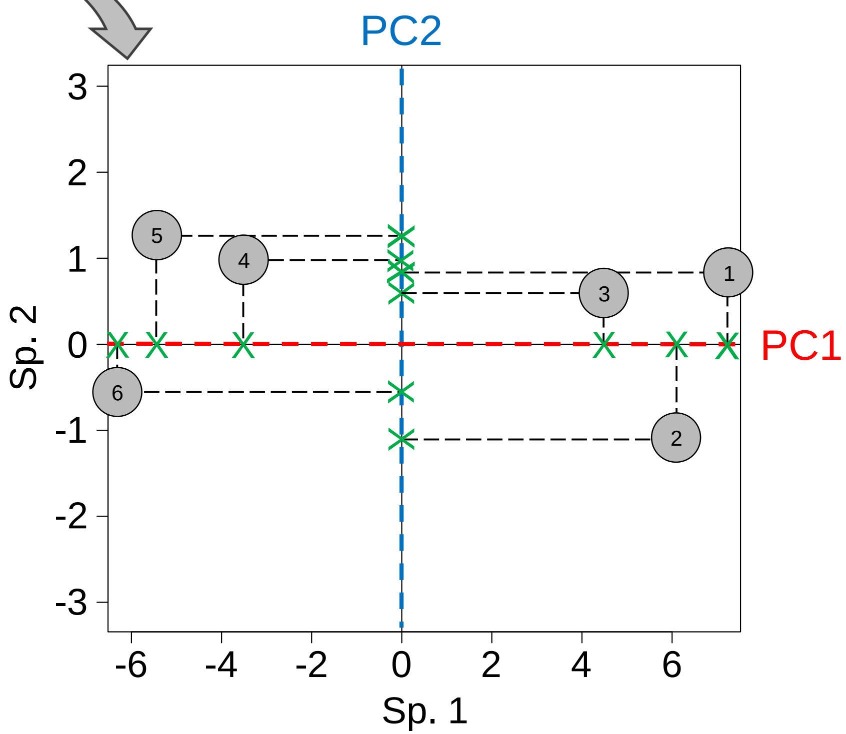
  ⊥ (orthogonal) ≡ no correlation ≡ independent
- PC1 + PC2 ≡ best projection plane for the cloud of individuals
- Look for a 3th axis and sequentially the others (⊥)
- $I_{total} = I_{PC1} + \underbrace{I_{PC2} + I_{PC3} + \cdots + I_{PCK}}_{I_{residual}}$

Funded by the European Union

Bringing out a structure – PCA – study of individuals (N^I)

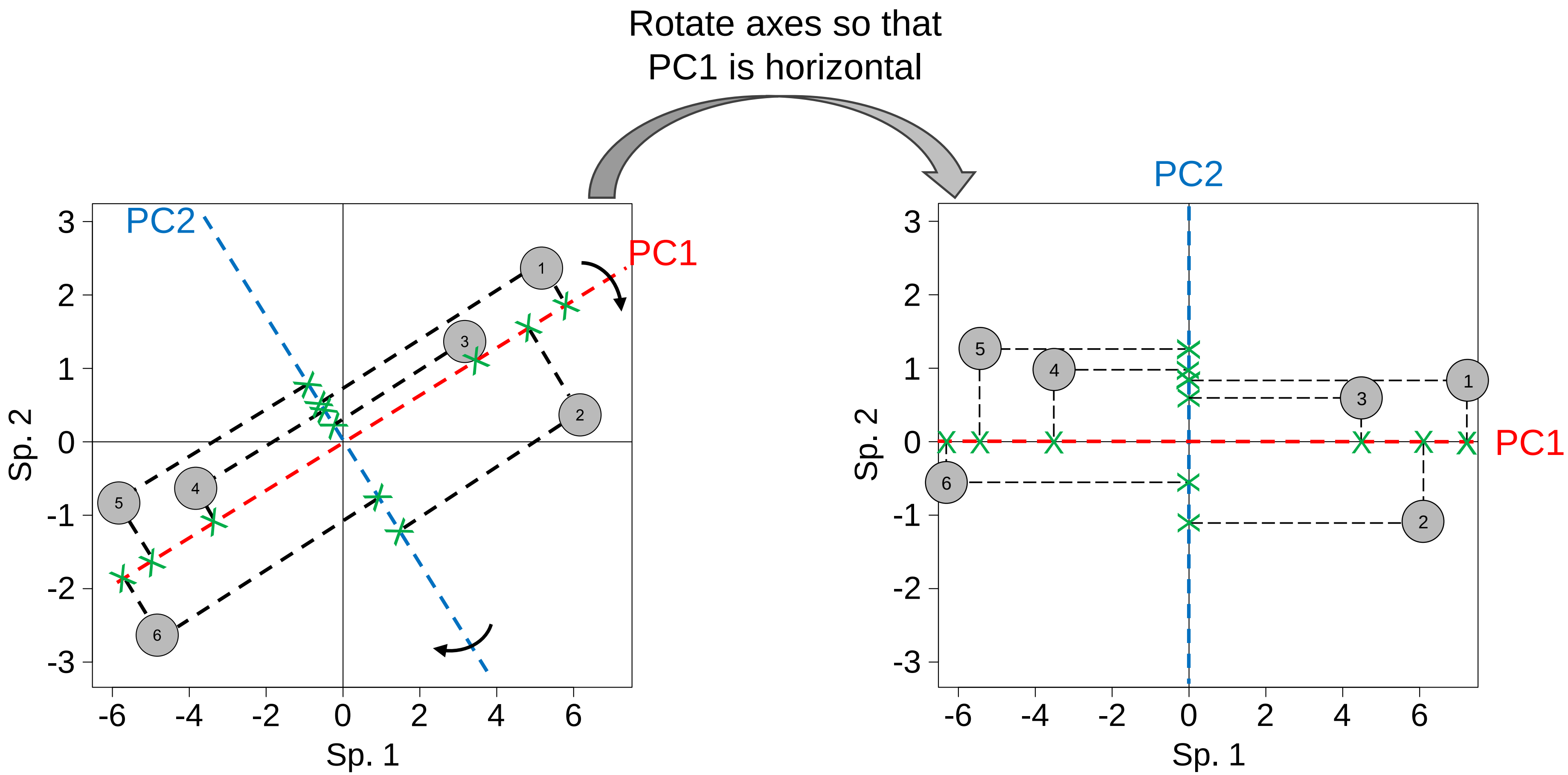
Rotate axes so that
PC1 is horizontal





Funded by
the European Union

Bringing out a structure – PCA – study of individuals (N^I)



Great, we made a PCA

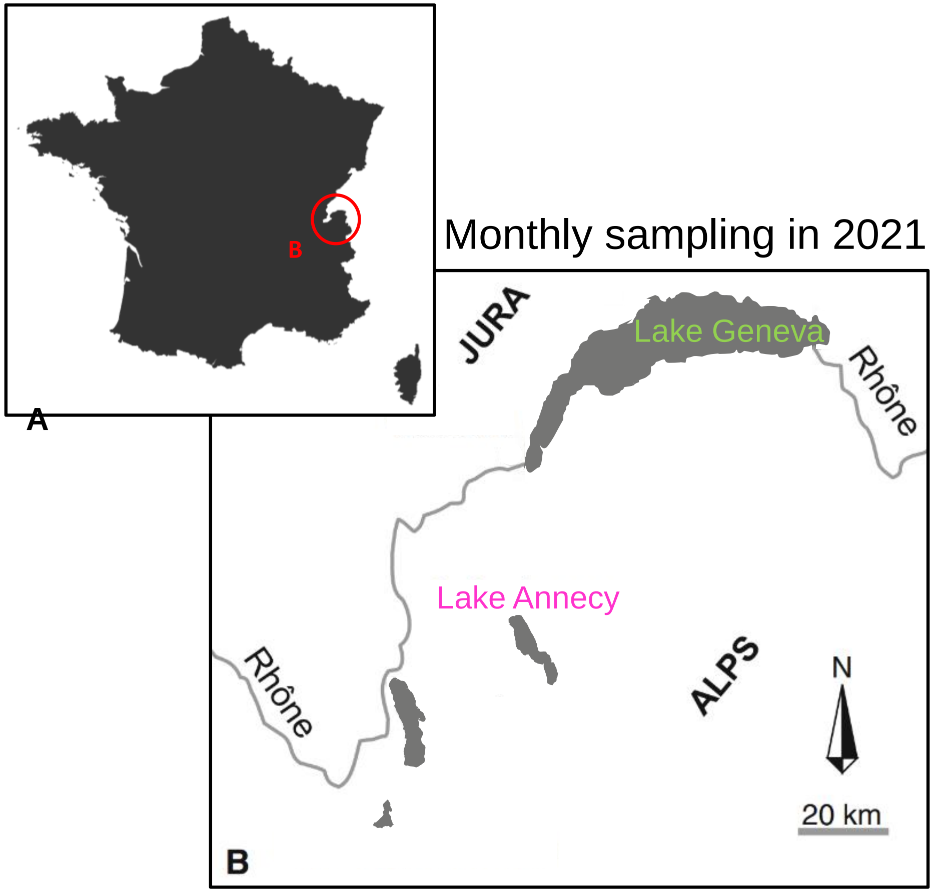
Funded by
the European Union

Bringing out a structure – PCA

Case of study: Spatio-temporal monitoring of microalgae in lakes Geneva and Annecy

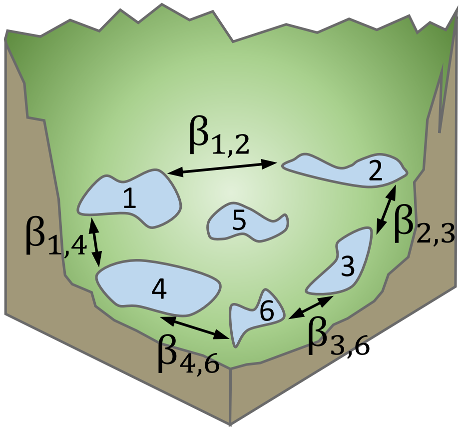
Practical course on R

Monthly sampling in 2021



Funded by the European Union

Understanding the structure



ASVs

Env.

Sites

Sites

Measuring the variation among commuties

Distance matrix

Aitchison distance

Bringing out a structure

Principal Component Analysis (PCA)	Correspondence Analysis (CA)	Principal Coordinate Analysis (PCoA)

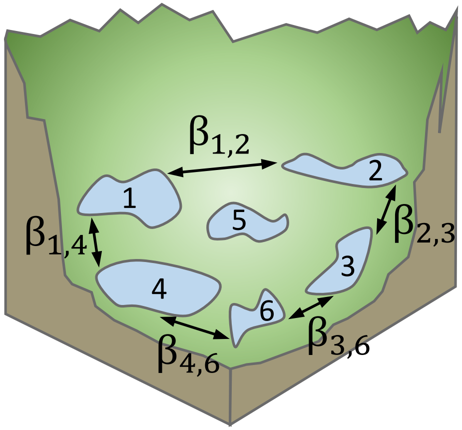
Explaining the structure with external variables

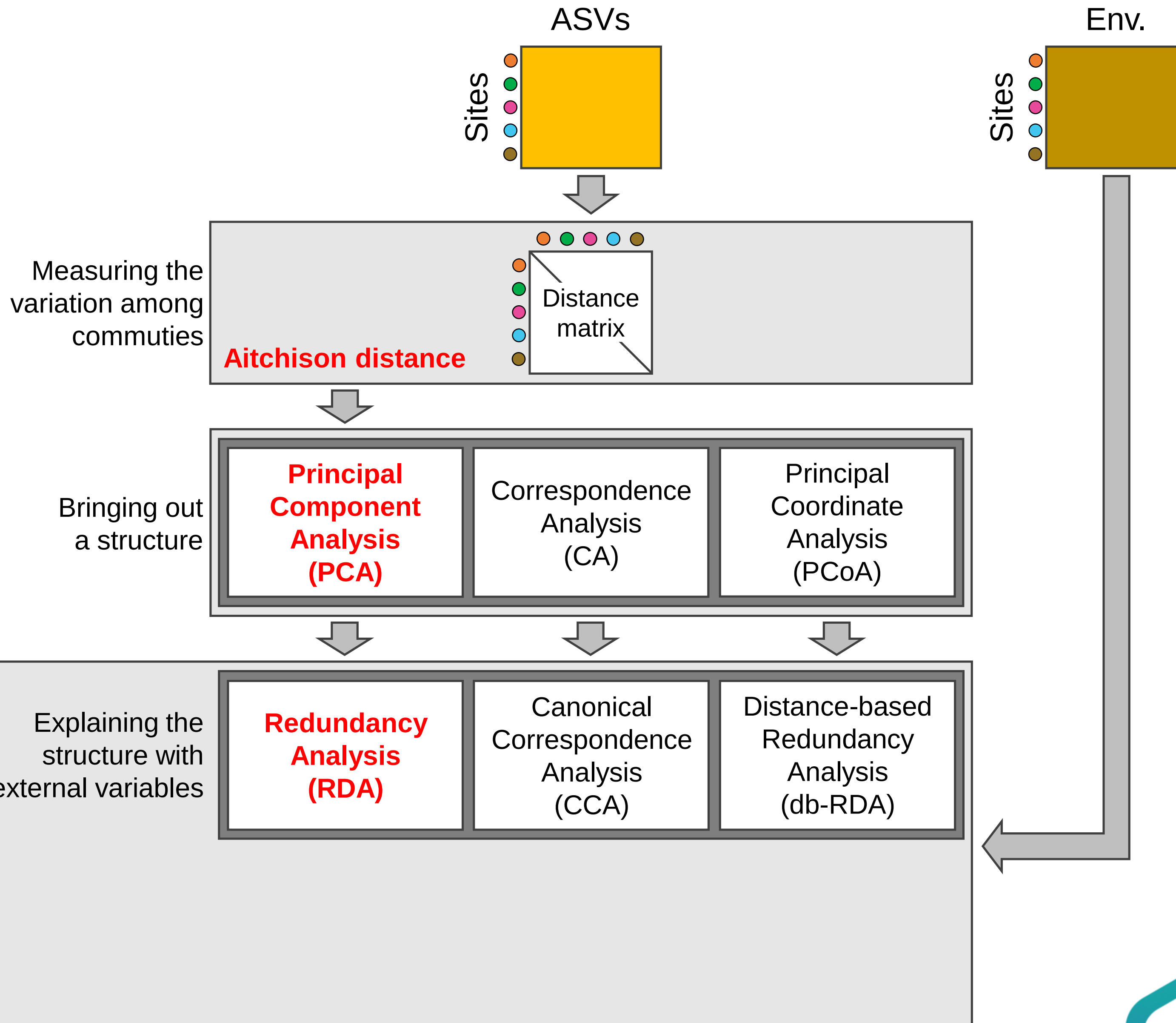
Redundancy Analysis (RDA)	Canonical Correspondence Analysis (CCA)	Distance-based Redundancy Analysis (db-RDA)

Understanding the structure

Funded by the European Union

Understanding the structure

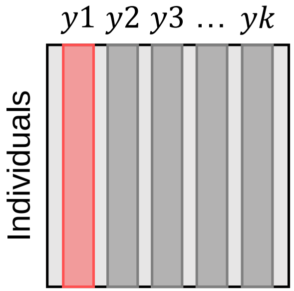




Funded by the European Union

Understanding the structure – RDA

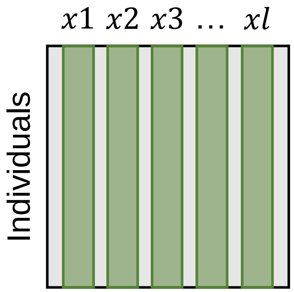
Data table $\boldsymbol{Y}$
(response variables)



Individuals: lakes

Response variables: species

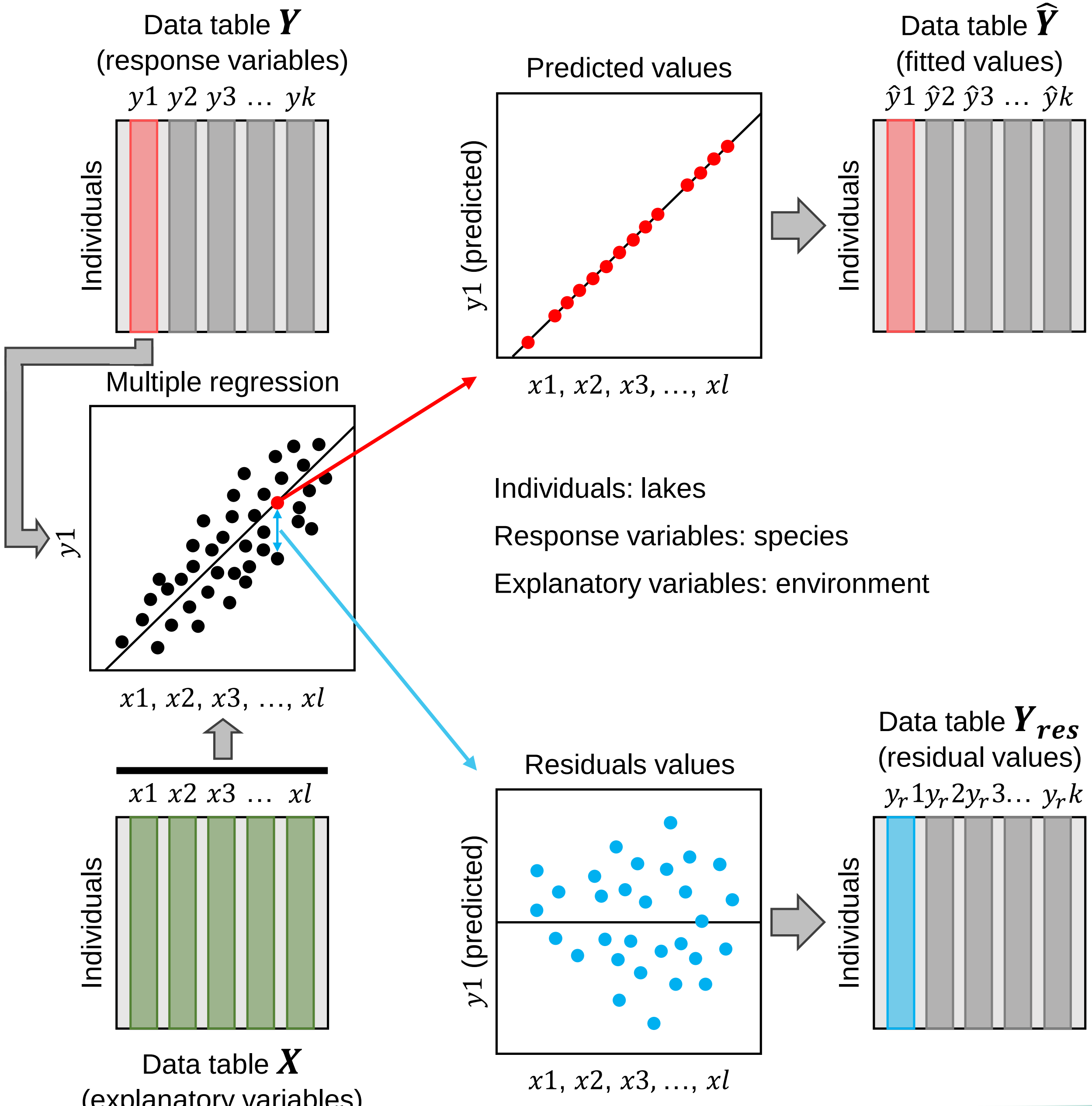
Explanatory variables: environment



Data table $\boldsymbol{X}$
(explanatory variables)

Understanding the structure – RDA

Step – 1

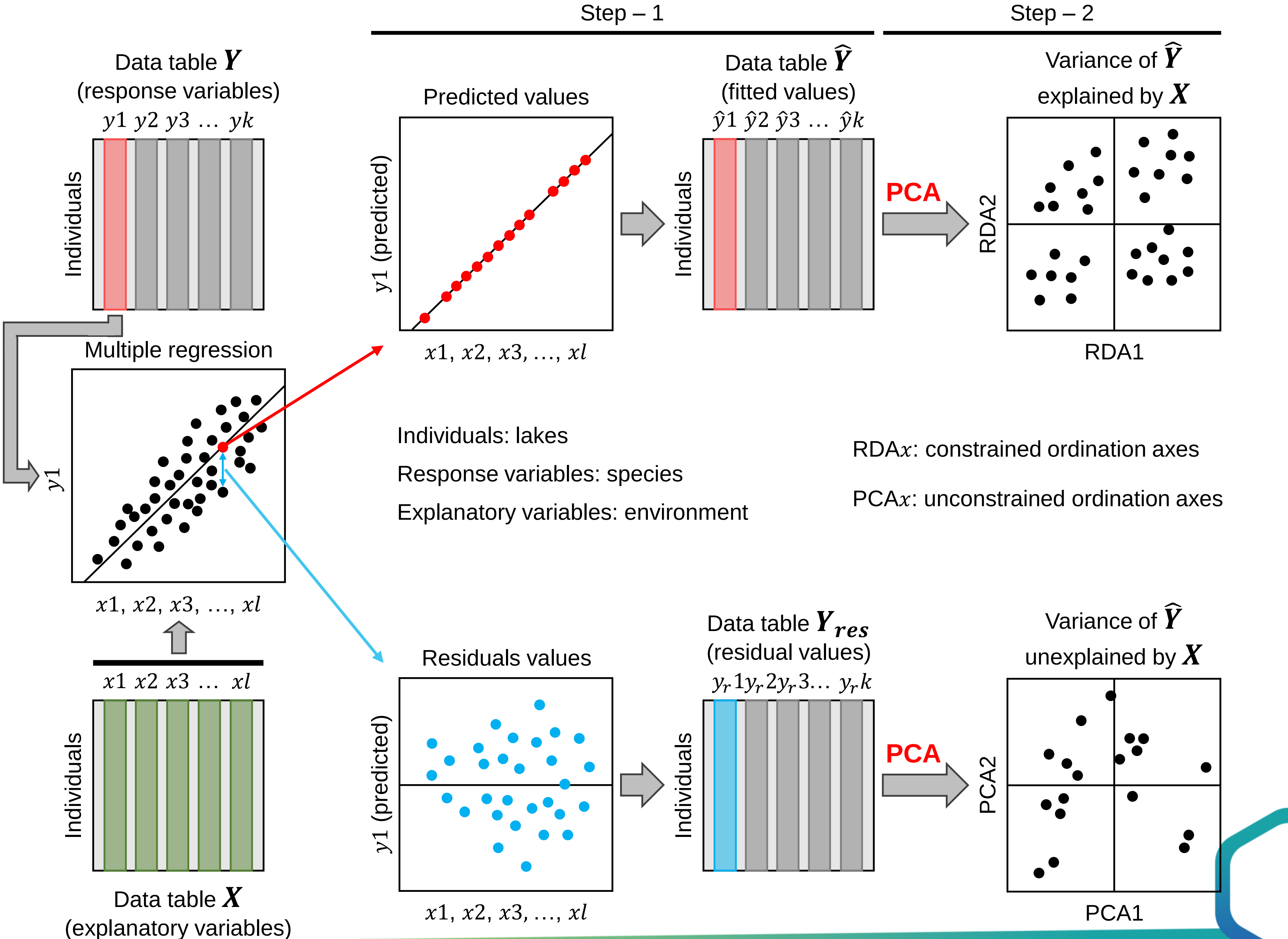


Individuals: lakes

Response variables: species

Explanatory variables: environment

Understanding the structure – RDA



Individuals: lakes

Response variables: species

Explanatory variables: environment

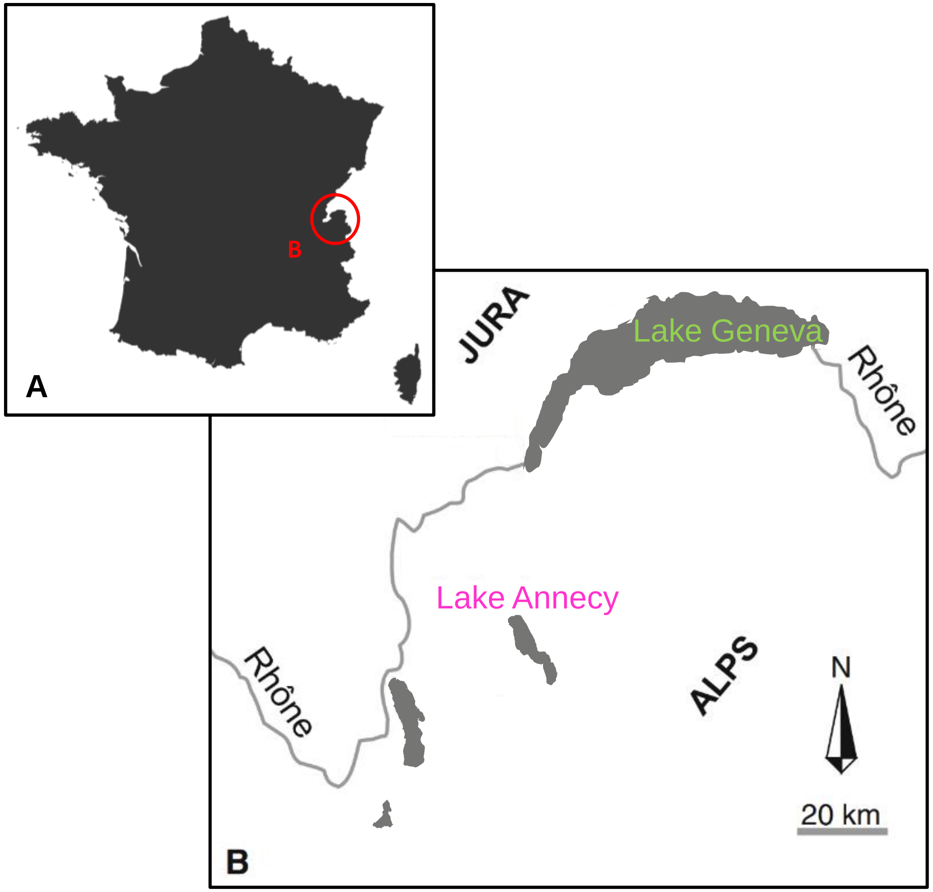
RDAx: constrained ordination axes

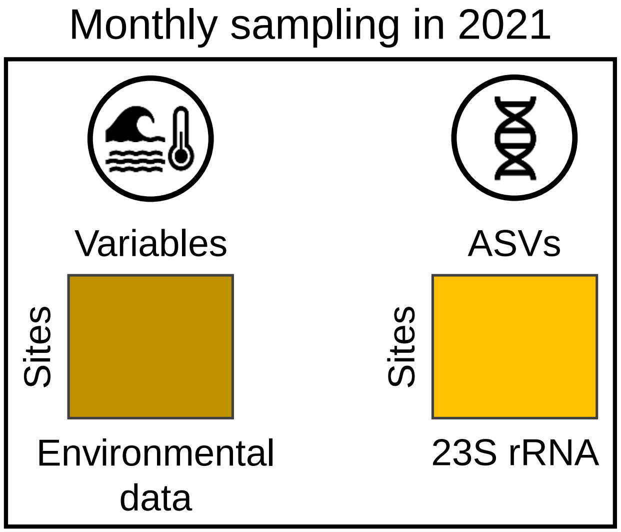
PCAx: unconstrained ordination axes

Understanding the structure – RDA

Case of study: Spatio-temporal monitoring of lakes Geneva and Annecy

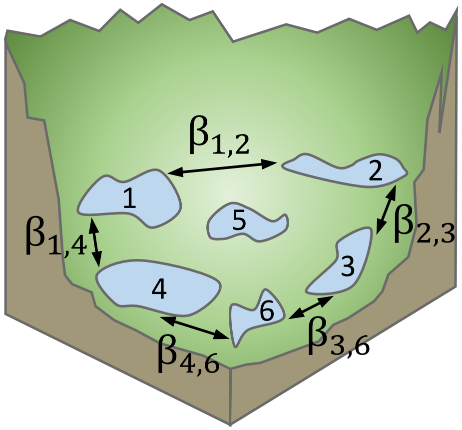
Practical course on R





Funded by the European Union

Understanding the structure



ASVs

Sites

Env.

Sites

Measuring the variation among commuties

Distance matrix

Aitchison distance

Bringing out a structure

Principal Component Analysis (PCA)

Correspondence Analysis (CA)

Principal Coordinate Analysis (PCoA)

Understanding the structure

Explaining the structure with external variables

Redundancy Analysis (RDA)

Canonical Correspondence Analysis (CCA)

Distance-based Redundancy Analysis (db-RDA)

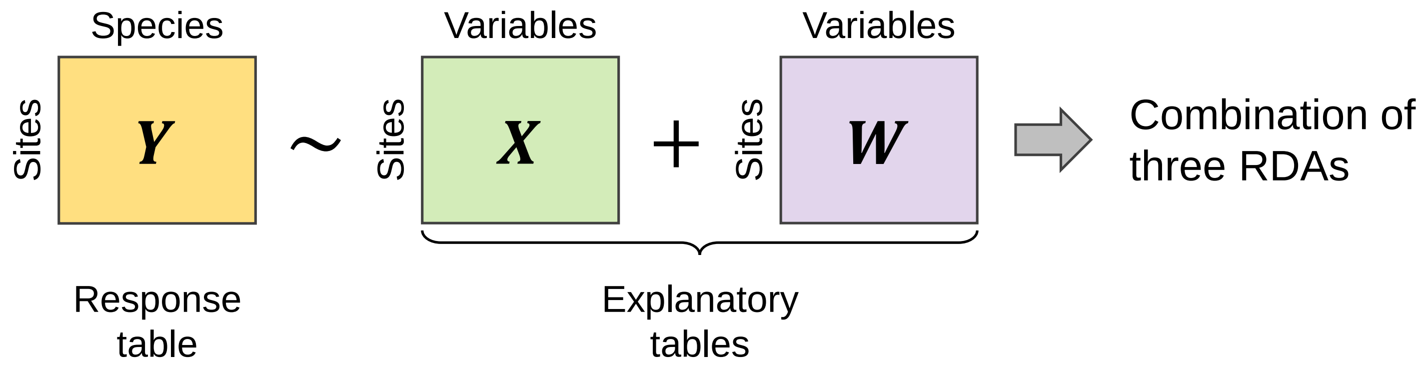
Quantify the effect of different groups of variables

Variation partitioning

Funded by the European Union

Understanding the structure – Variation partitioning

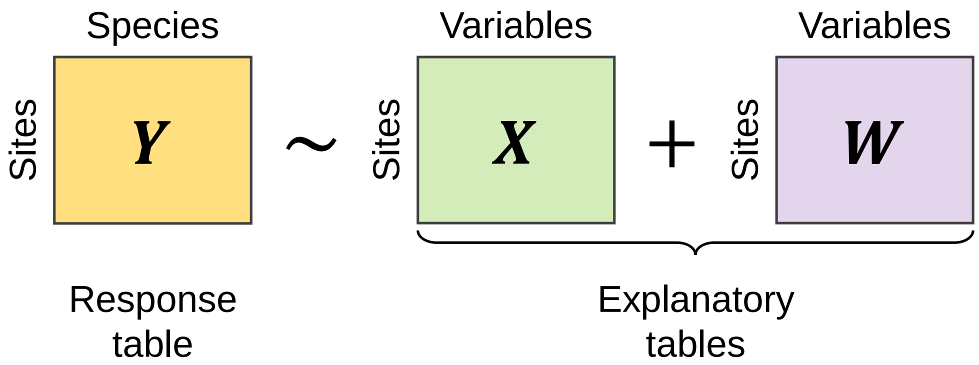
Variation partitioning:
Quantify the relative effects of different groups
of explanatory variables on the response variables



Funded by
the European Union

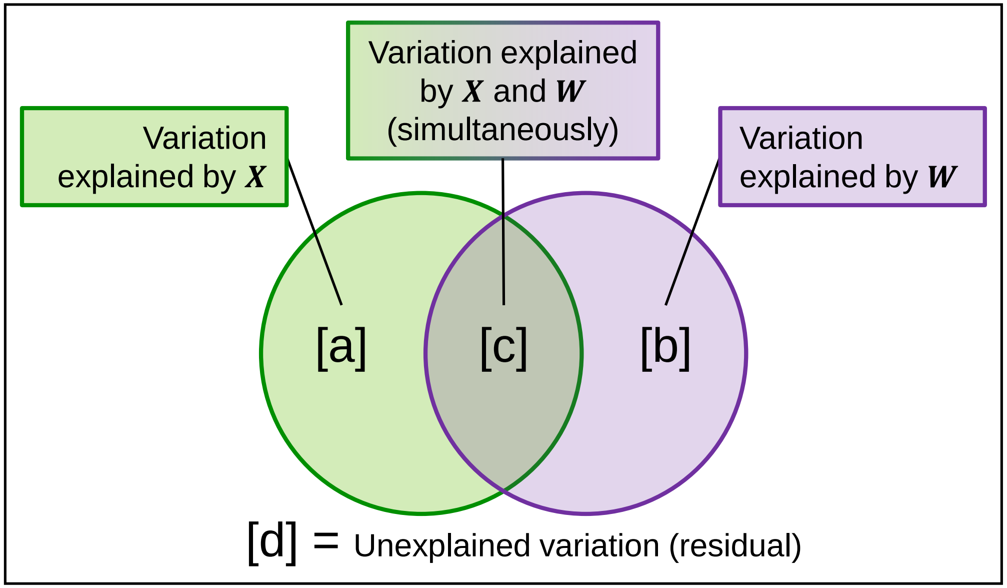
Understanding the structure – Variation partitioning

Variation partitioning:
Quantify the relative effects of different groups
of explanatory variables on the response variables



Combination of three RDAs

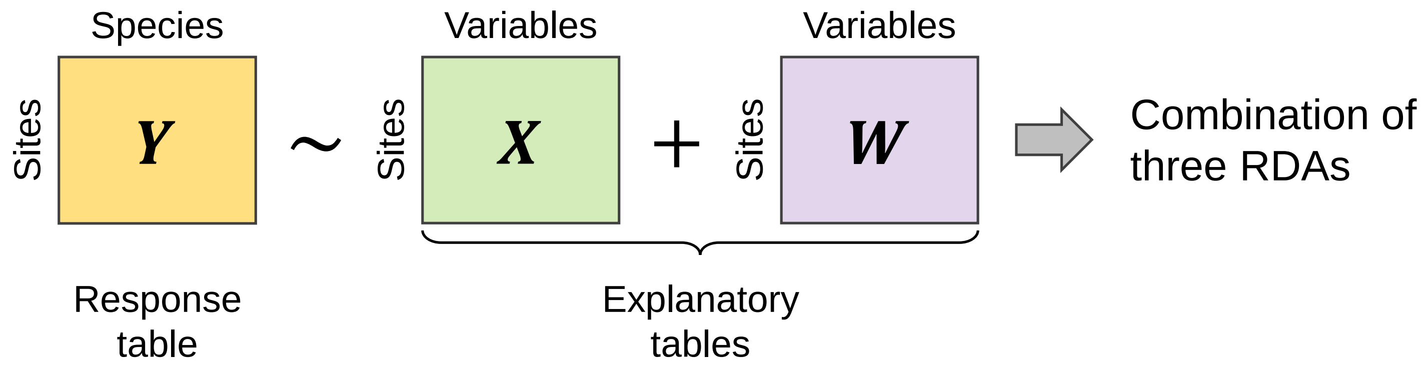
Variation in table Y =



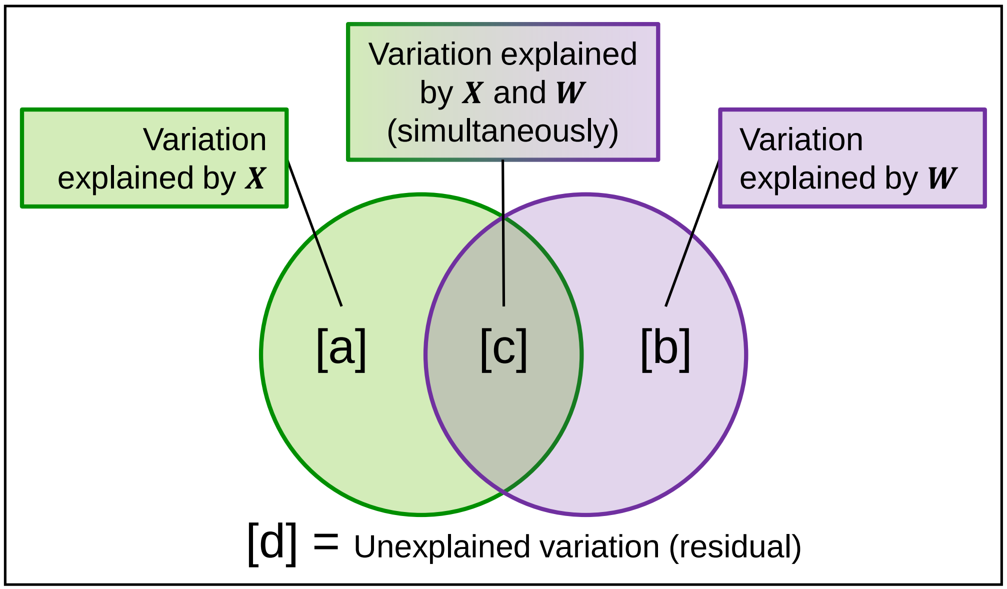
Funded by the European Union

Understanding the structure – Variation partitioning

Variation partitioning:
Quantify the relative effects of different groups
of explanatory variables on the response variables



Variation in table $\boldsymbol{Y}$ =

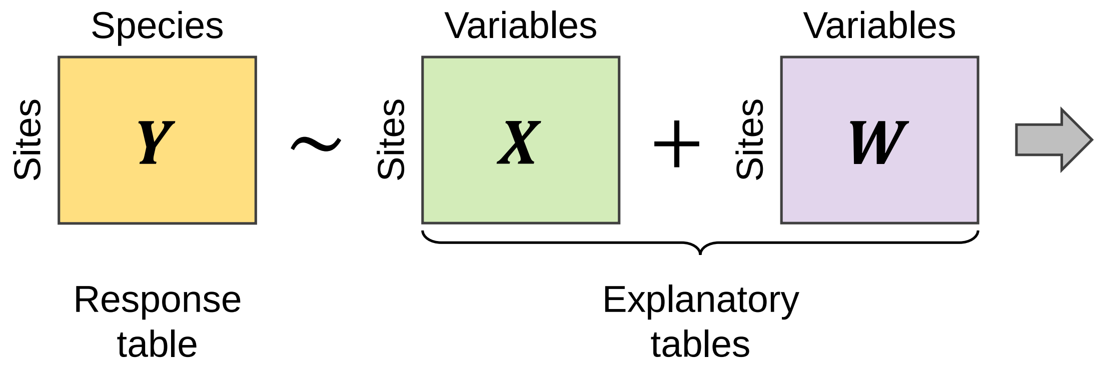


1. RDA of $\boldsymbol{Y}$ by $\boldsymbol{X}$: [a + c]
2. RDA of $\boldsymbol{Y}$ by $\boldsymbol{W}$: [b + c]
3. RDA of $\boldsymbol{Y}$ by $\boldsymbol{X}$ and $\boldsymbol{W}$: [a + b + c]

Funded by
the European Union

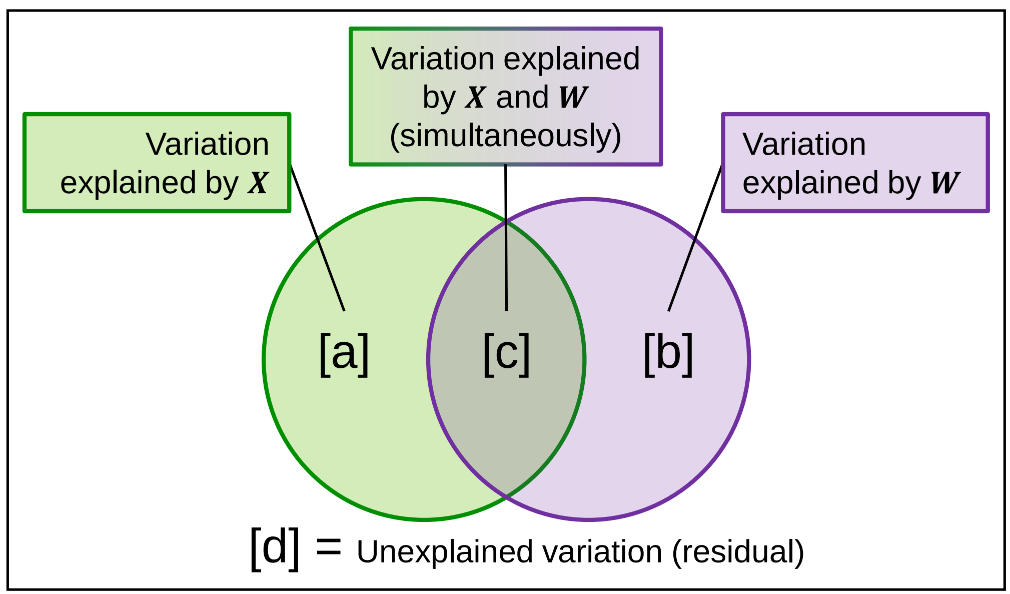
Understanding the structure – Variation partitioning

Variation partitioning:
Quantify the relative effects of different groups
of explanatory variables on the response variables



Combination of three RDAs

Variation in table Y =



1. RDA of Y by X: [a + c]
2. RDA of Y by W: [b + c]
3. RDA of Y by X and W: [a + b + c]

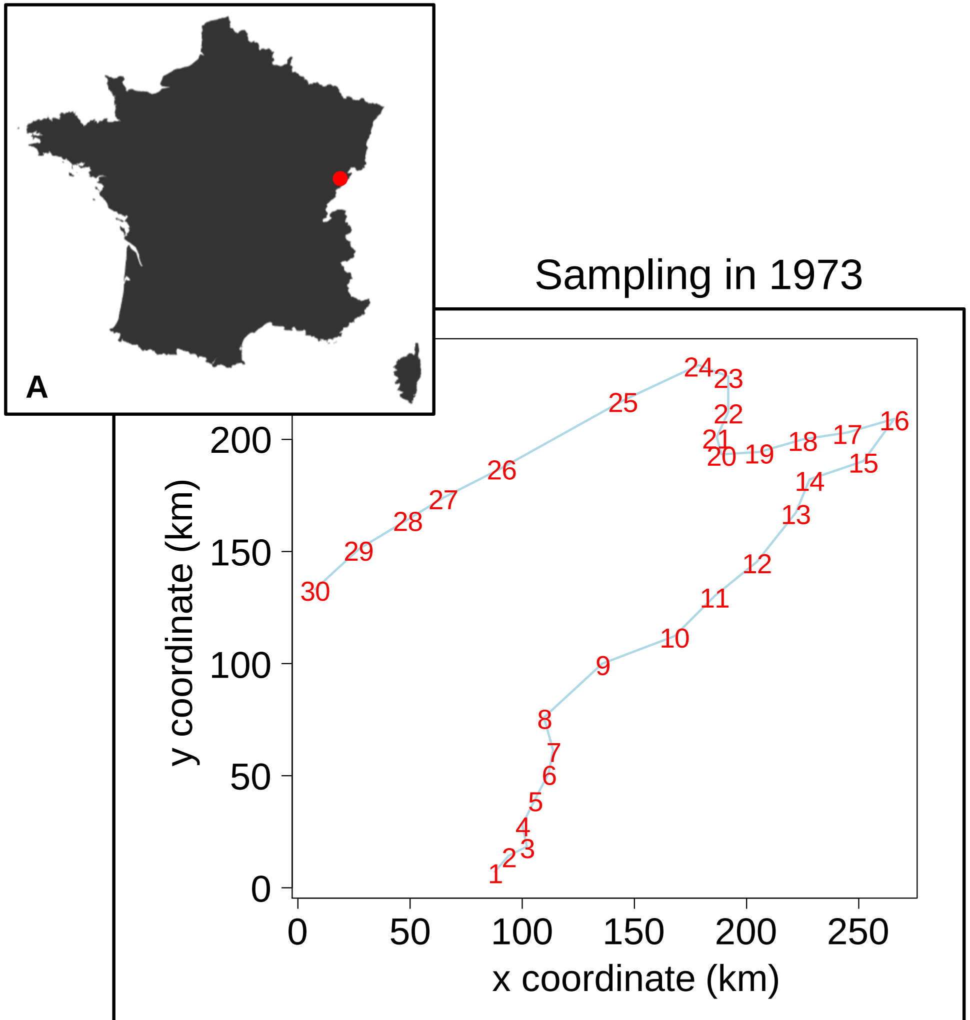
4. [a] = [a + b + c] – [b + c]
5. [b] = [a + b + c] – [a + c]
6. [d] = 1 – [a + b + c]
7. [c] = [a + c] – [a] = [b + c] – [b]

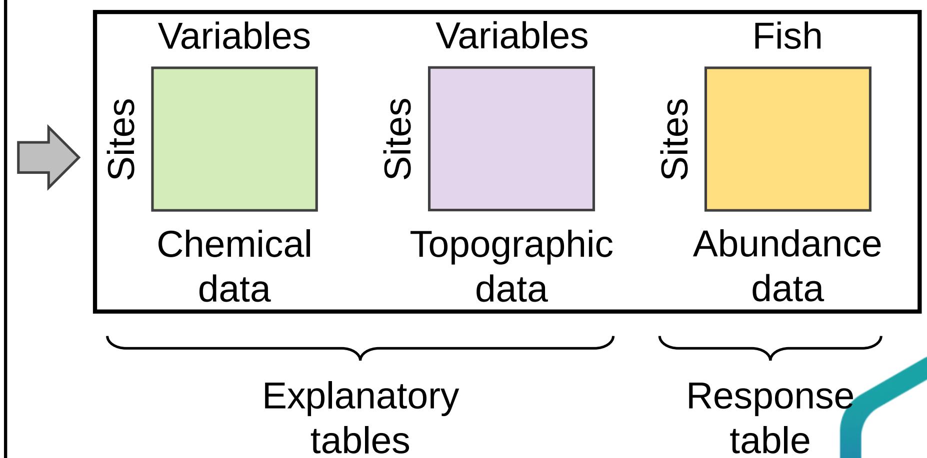
Funded by
the European Union

Understanding the structure – Variation partitioning

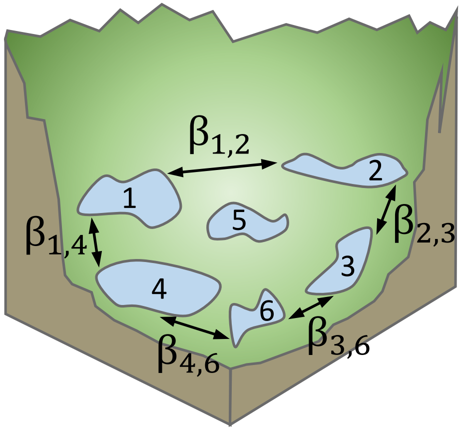
Case of study: Data from Verneaux (1973) → Doubs river (France), fish communities

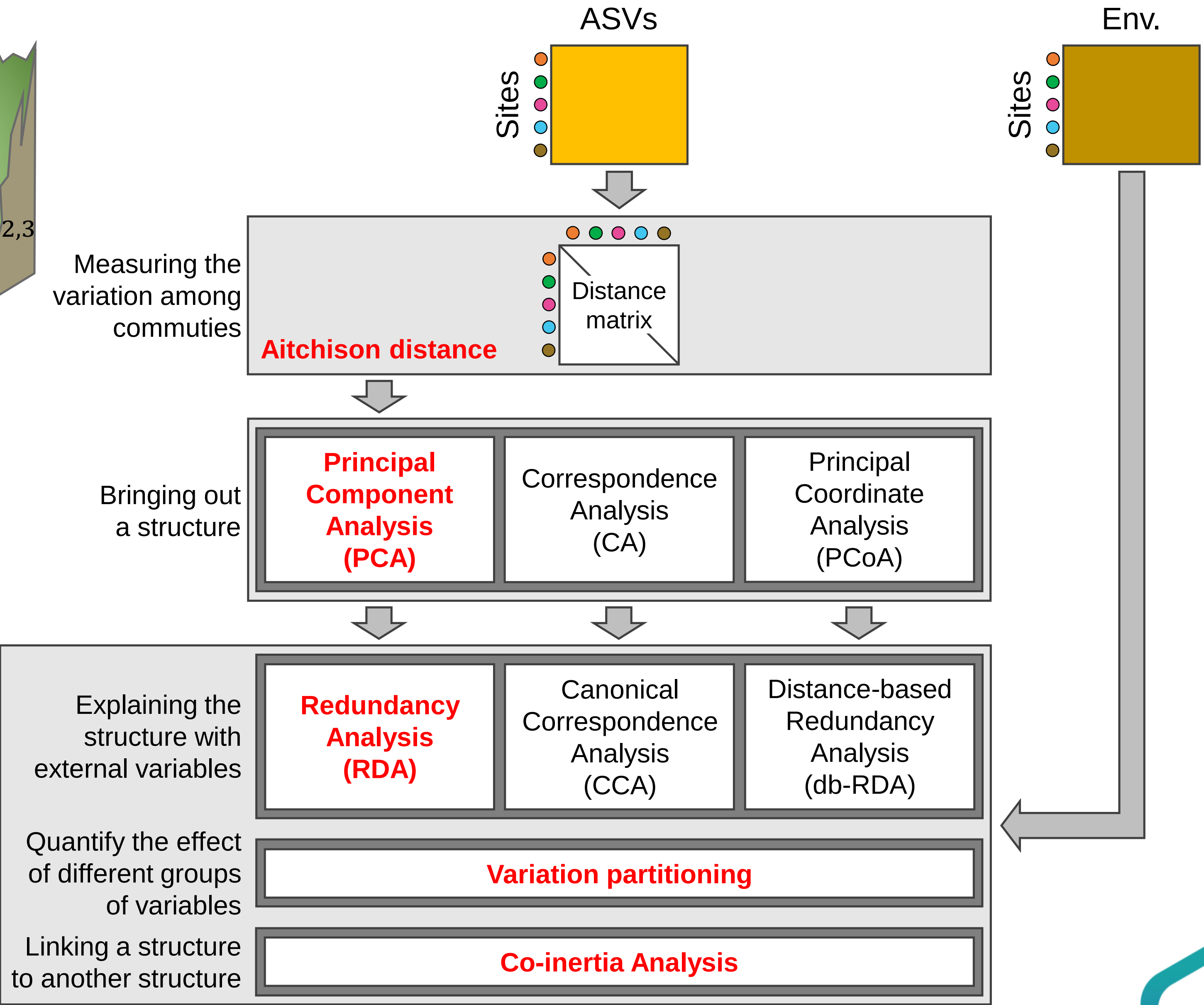
Practical course on R





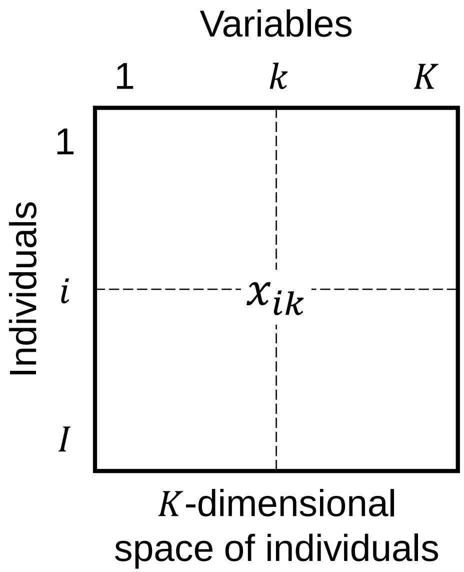
Understanding the structure



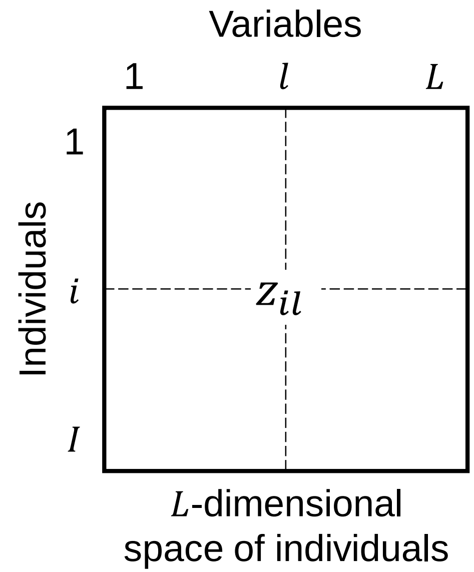


Funded by the European Union

Understanding the structure – Co-inertia analysis

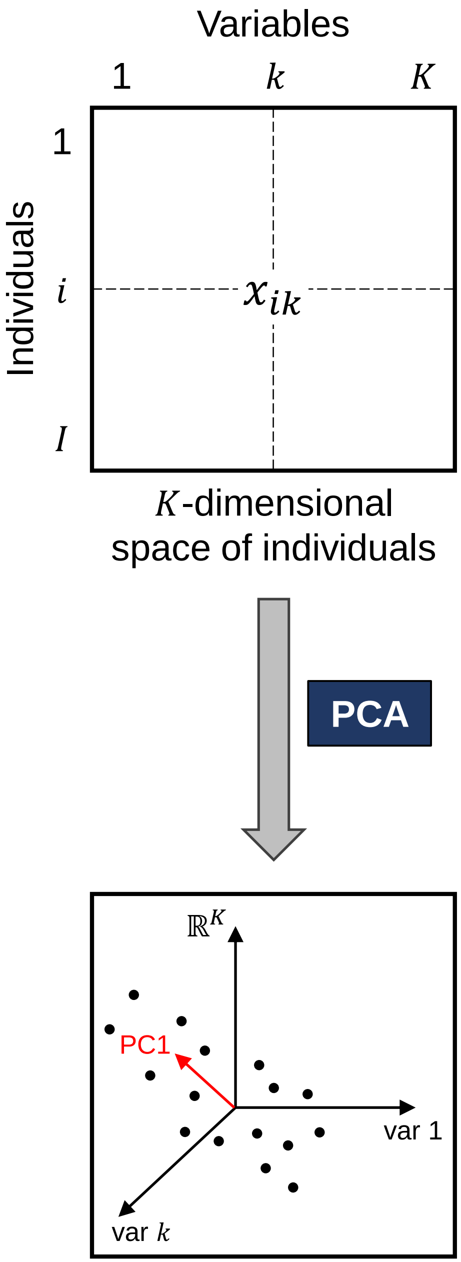


Co-inertia analysis:
Enables finding common structures
between two set of variables

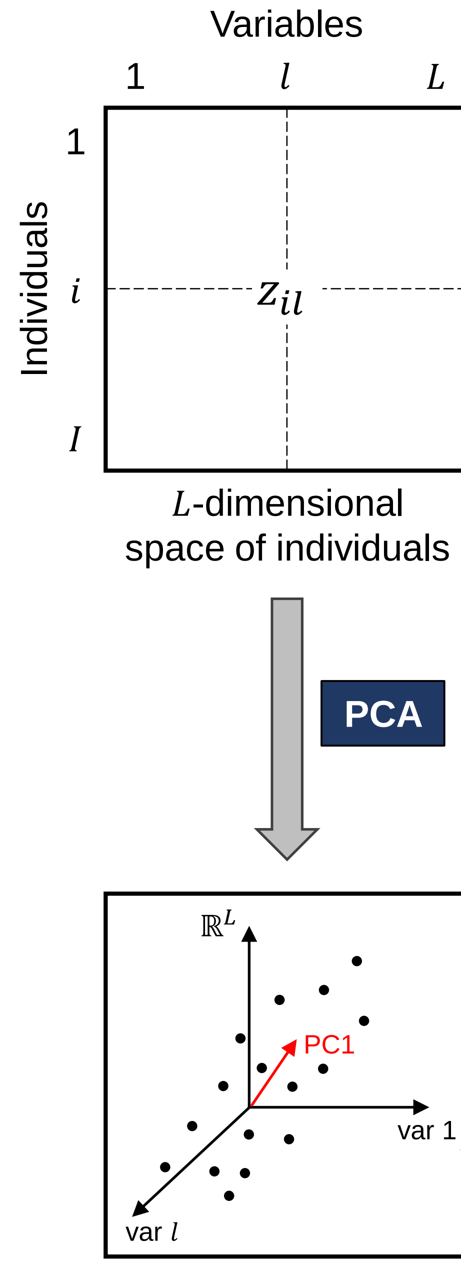


Funded by
the European Union

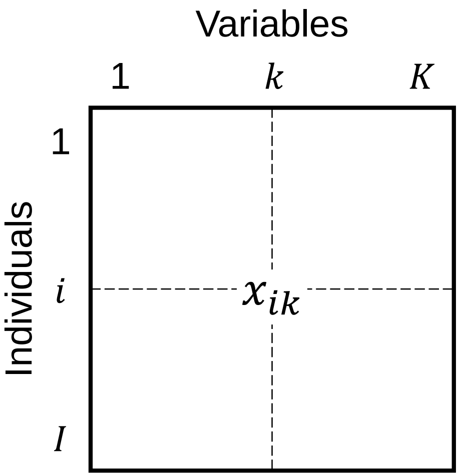
Understanding the structure – Co-inertia analysis



Co-inertia analysis:
Enables finding common structures between two set of variables

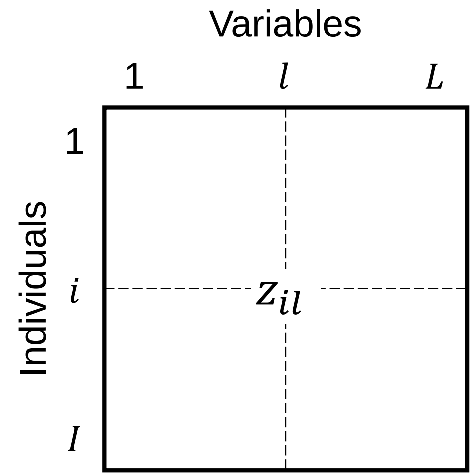


Understanding the structure – Co-inertia analysis



K-dimensional space of individuals

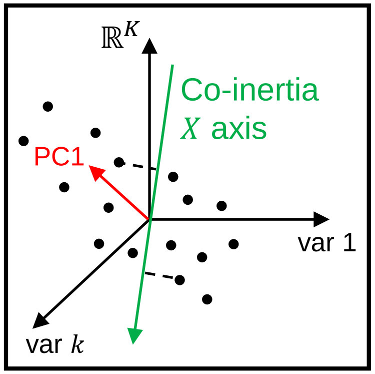
Co-inertia analysis:
Enables finding common structures between two set of variables



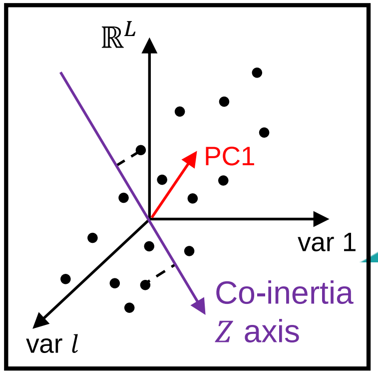
L-dimensional space of individuals

PCA

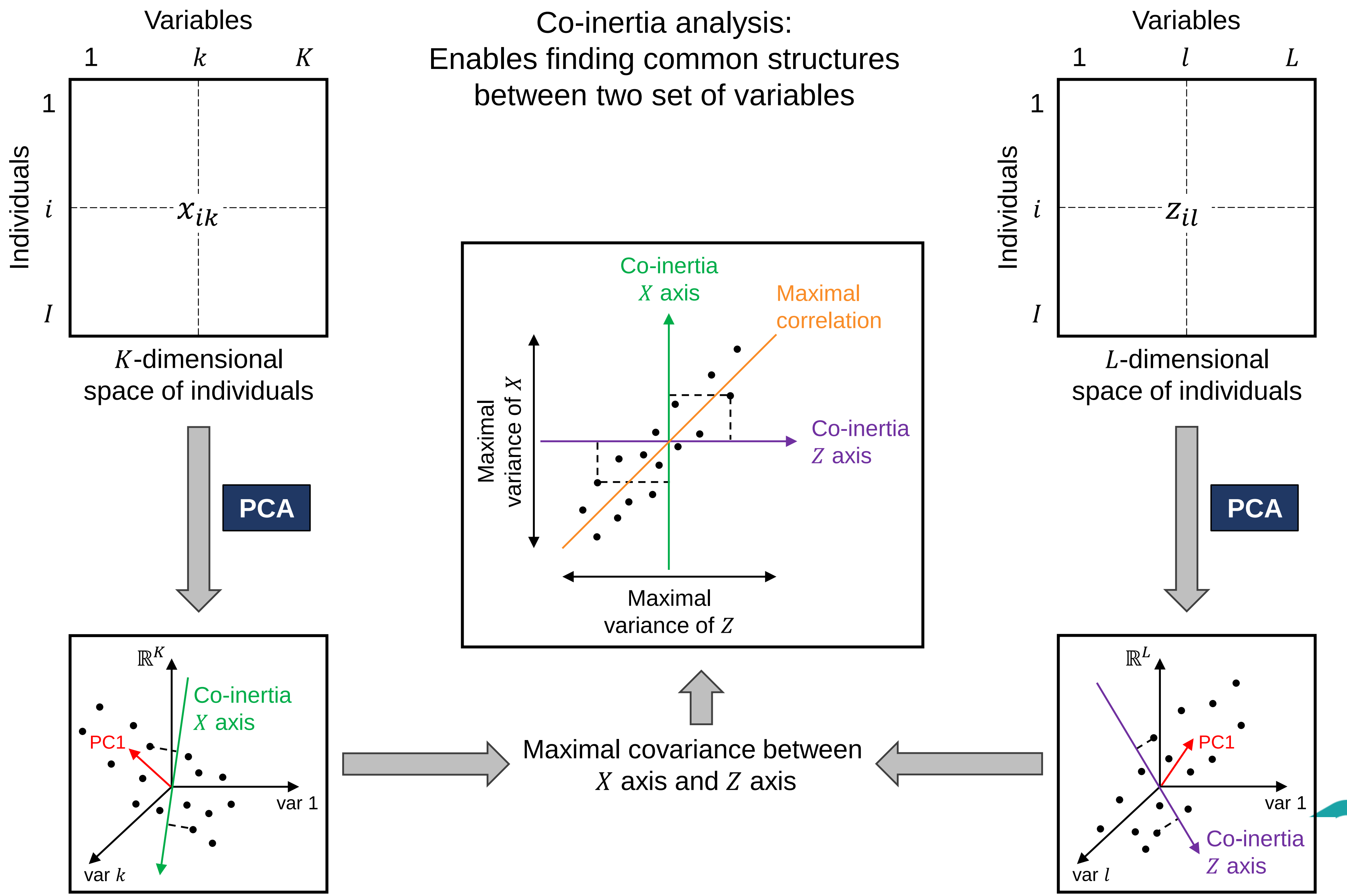
PCA



Maximal covariance between X axis and Z axis

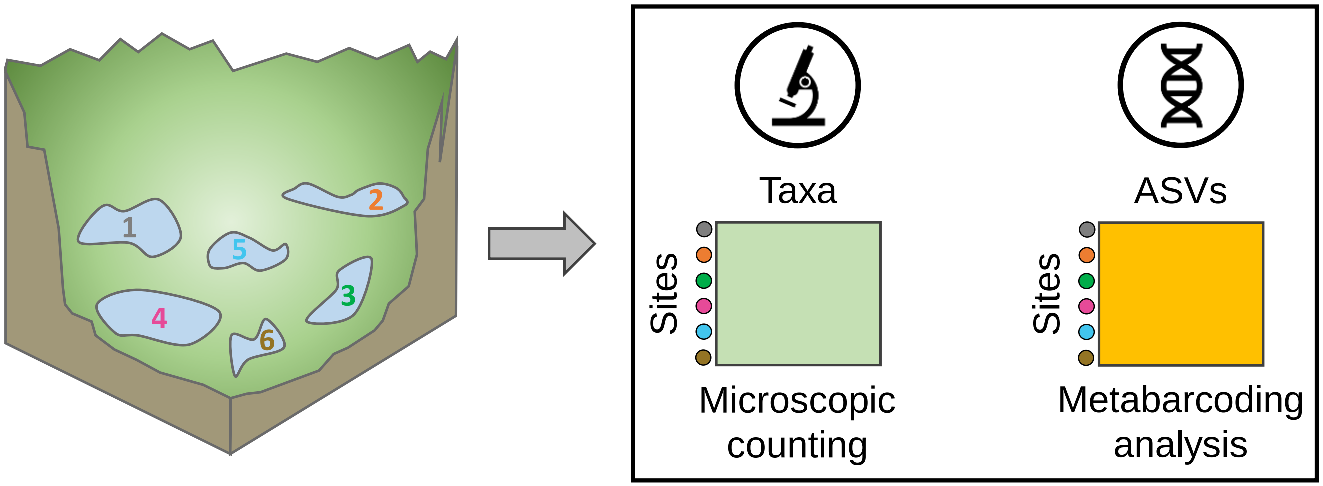


Understanding the structure – Co-inertia analysis



Understanding the structure – Co-inertia analysis

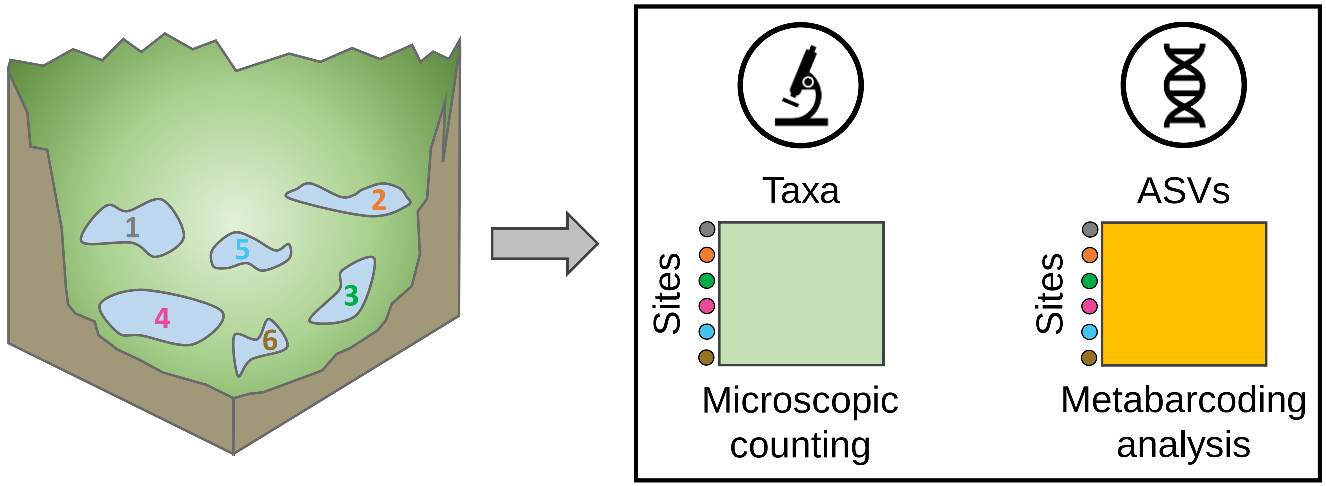
Co-inertia analysis:
Enables finding common structures between two set of variables



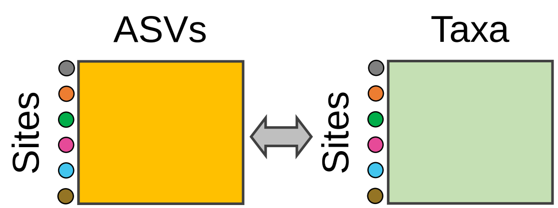
Funded by the European Union

Understanding the structure – Co-inertia analysis

Co-inertia analysis:
Enables finding common structures between two set of variables

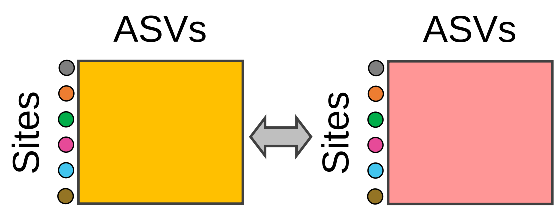


1. Concordance between eDNA and microscopy data



e.g., Nicolosi et al. (2023)
STOTEN, *in review*

2. Congruence between two biological communities



e.g., Alric et al. (2020)
Mol. Ecol. Res.

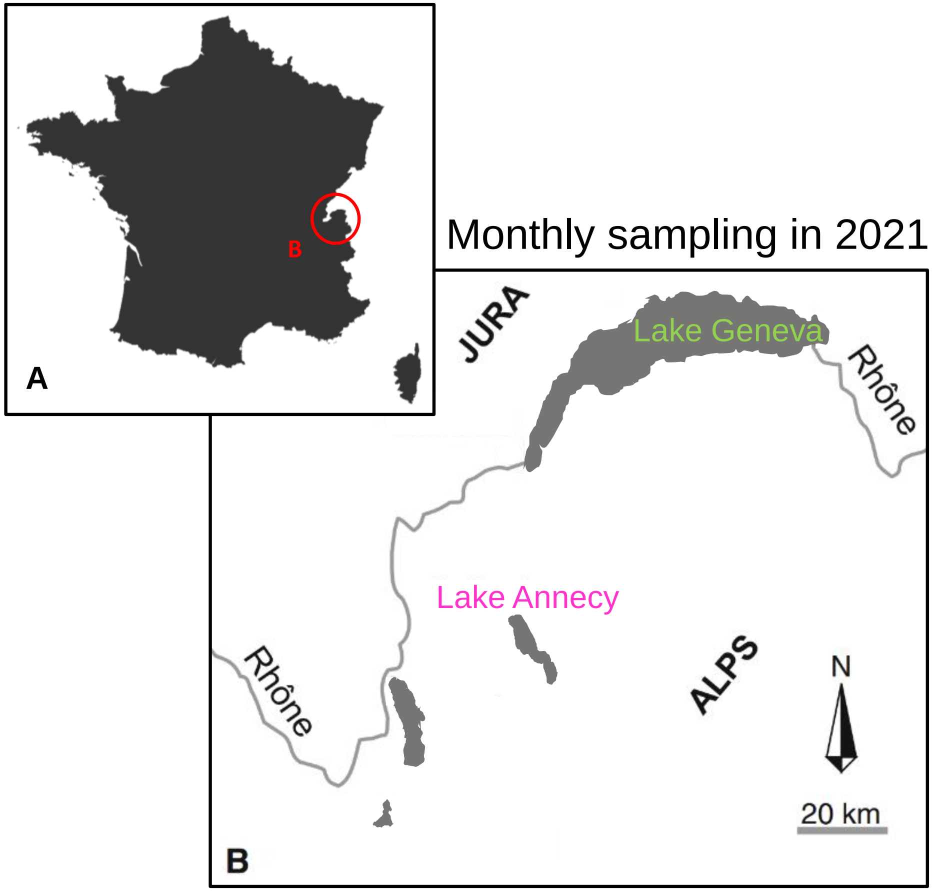
Funded by the European Union

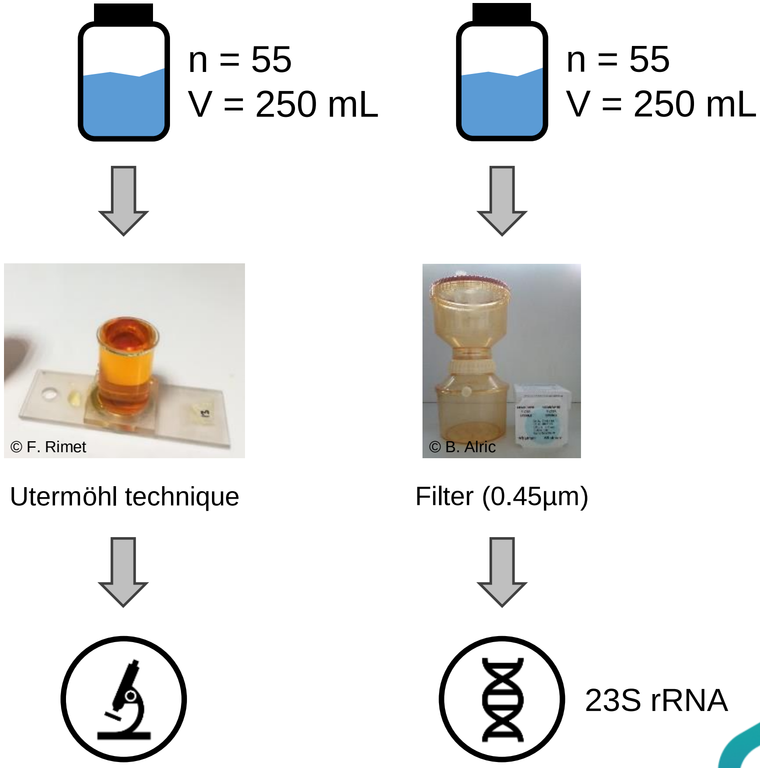
Understanding the structure – Co-inertia analysis

1. Concordance between eDNA and microscopy data

Practical course on R

Monthly sampling in 2021

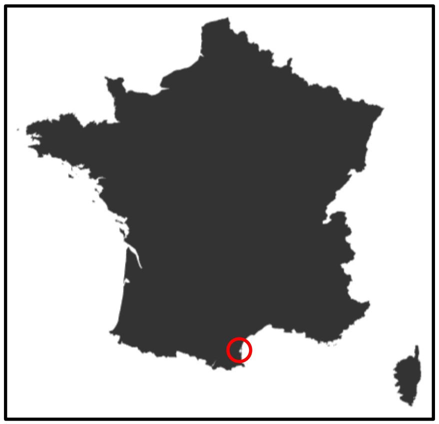


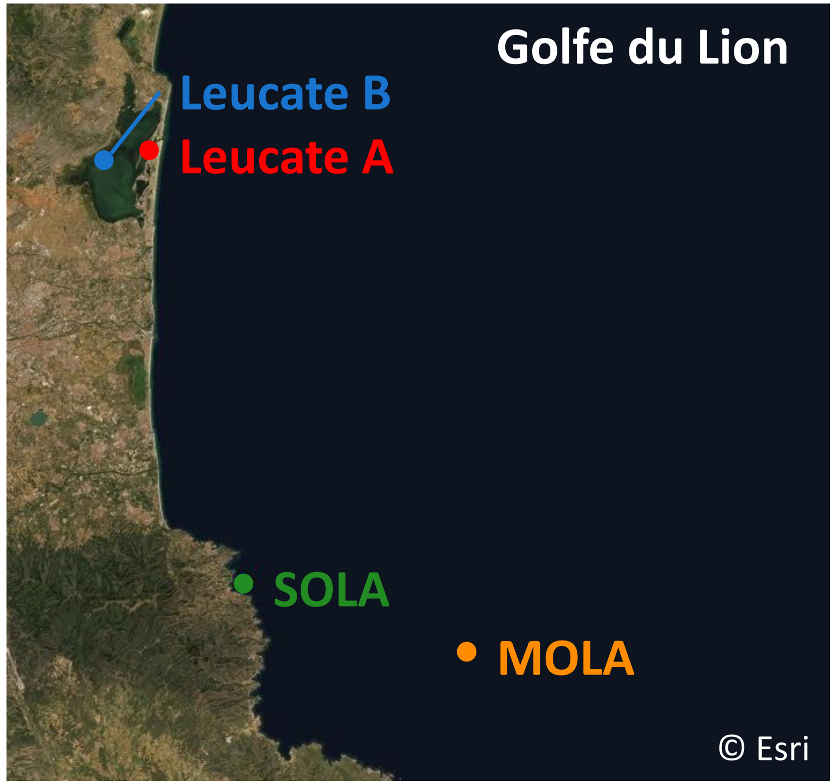


Funded by the European Union

Understanding the structure – Co-inertia analysis

2. Congruence between two biological communities
   e.g., Alric et al. Mol. Ecol. Res., 20 (2020) – Host-virus association in marine environment





Monthly monitoring over 2 years

Microalgae (Mamiellophyceae): 18S rRNA
Virus (*Prasinovirus*): *PolB*

Eutrophication gradient

Understanding the structure – Co-inertia analysis

2. Congruence between two biological communities

e.g., Alric et al. Mol. Ecol. Res., 20 (2020) – Host-virus association in marine environment

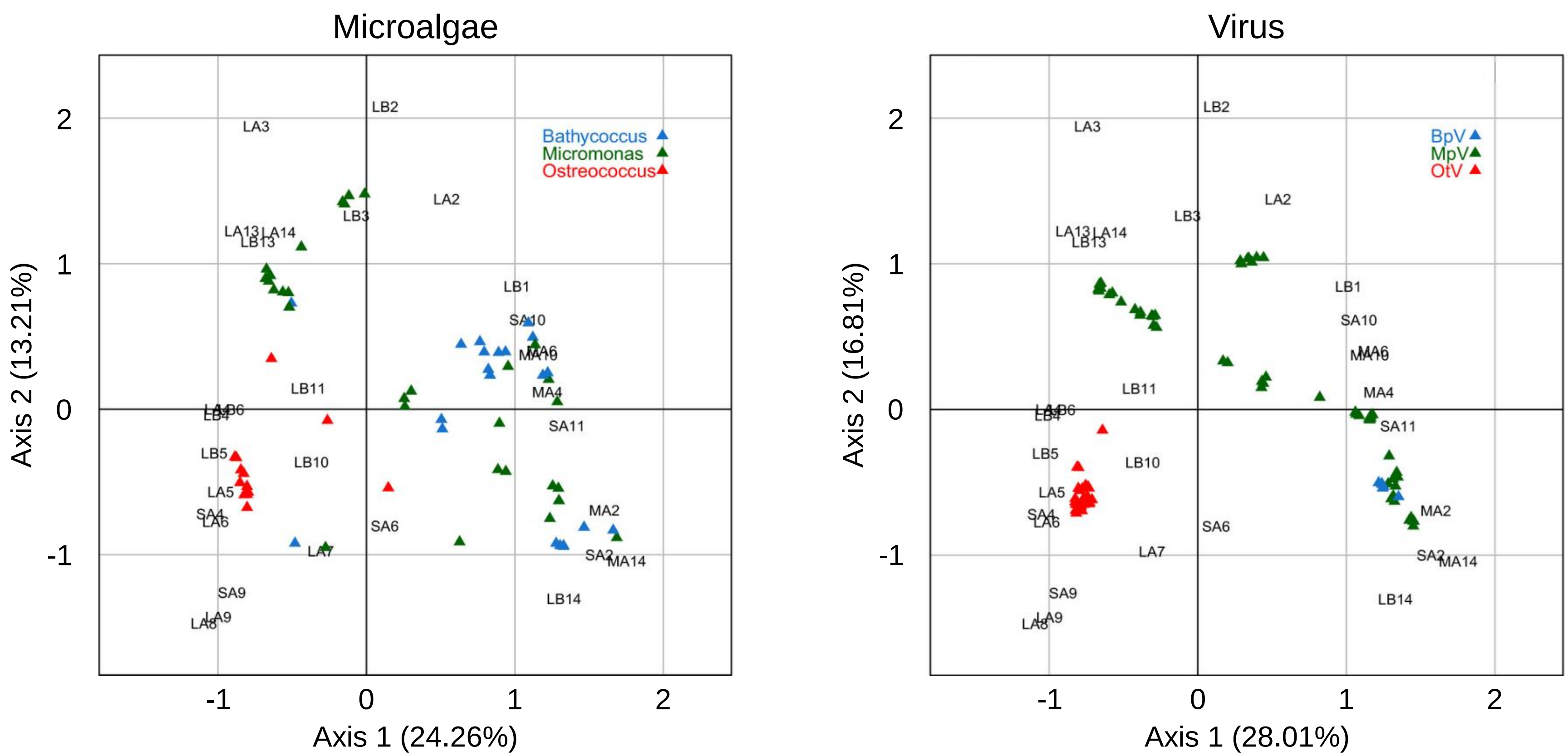
Co-correspondence Analysis (CoCA) ≡ co-inertia based on two CA

Funded by
the European Union

Understanding the structure – Co-inertia analysis

2. Congruence between two biological communities
 e.g., Alric et al. Mol. Ecol. Res., 20 (2020) – Host-virus association in marine environment

Co-correspondence Analysis (CoCA) ≡ co-inertia based on two CA
Predictive power = 32.02%, p = 0.001 (Axis 1, Axis 2)



Funded by
the European Union

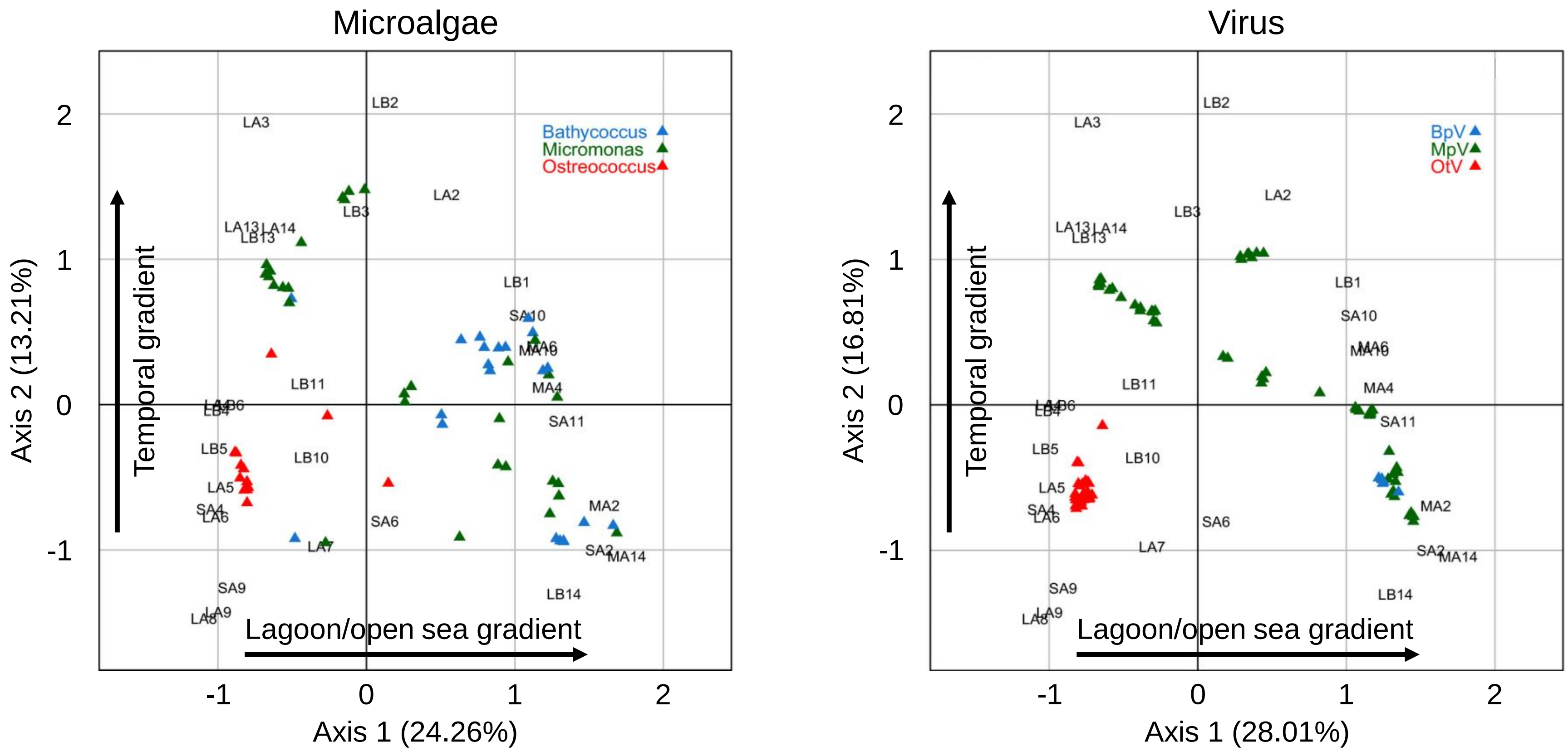
Understanding the structure – Co-inertia analysis

2. Congruence between two biological communities

e.g., Alric et al. Mol. Ecol. Res., 20 (2020) – Host-virus association in marine environment

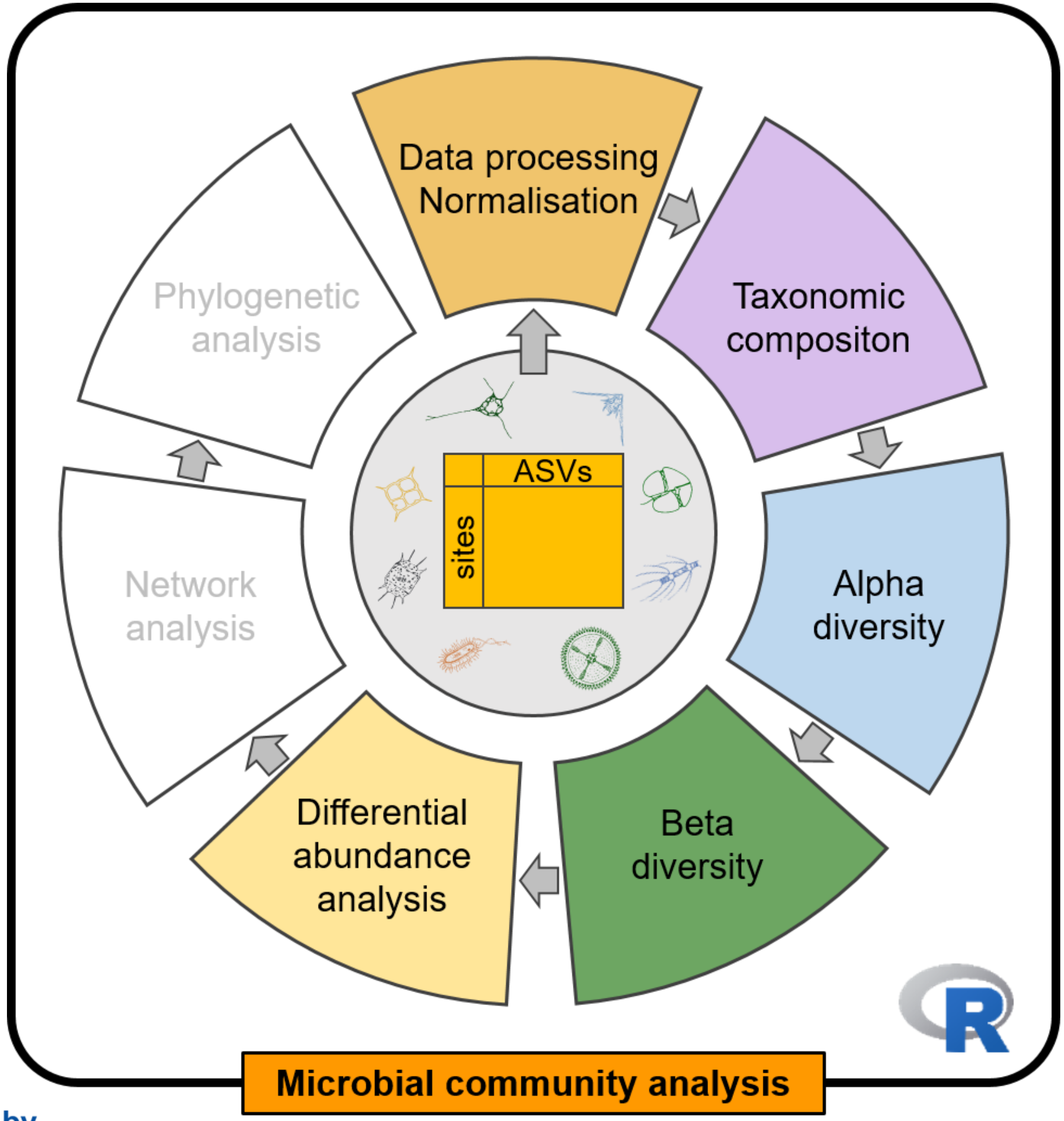
Co-correspondence Analysis (CoCA) ≡ co-inertia based on two CA

Predictive power = 32.02%, *p* = 0.001 (Axis 1, Axis 2)



Covariation of two communities along a lagoon/open-sea gradient (Axis 1) and a temporal gradient (Axis 2)

Statistical analysis workflow from metabarcoding data



Funded by
the European Union

Differential abundance analysis

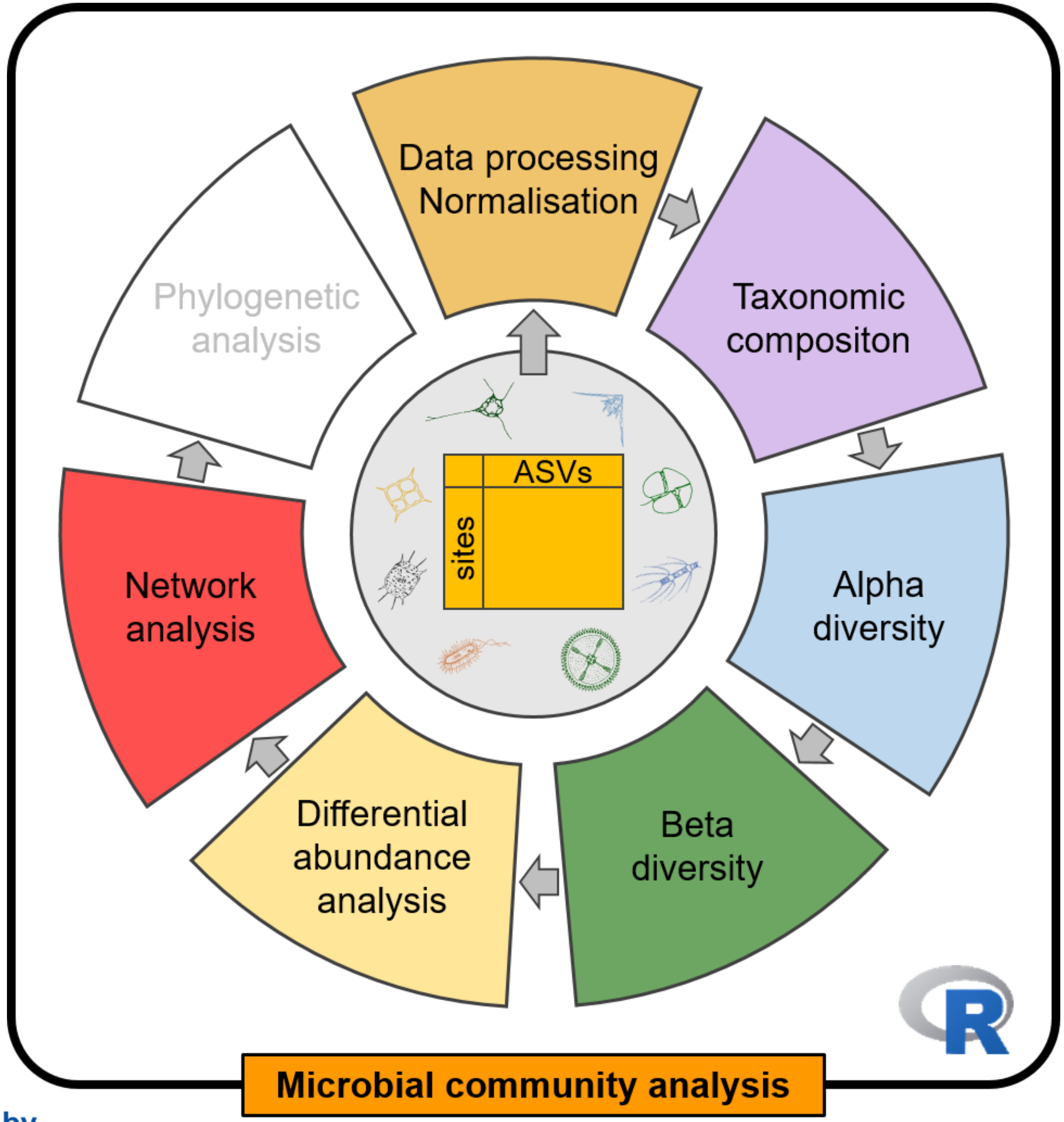
Differential abundance analysis:
Identify biomarker taxa (i.e., taxa whose relative abundance is significantly higher under given environmental conditions)

Method	Model assumption	Normalization	References	Availability	
edgeR*	Negative binomial	TMM	[42]	Bioconductor	2012
metagenomeSeq	Zero-inflated normal or log-normal	CSS	[24]	Bioconductor	2013
DESeq2*	Negative binomial	RLE	[39]	Bioconductor	2014
ANCOM	ANOVA	ALR	[40]	GitHub	2015
ZIBseq	Zero-inflated beta	TSS	[5]	CRAN	2016
ZIGDM	Zero-inflated generalized Dirichlet-multinomial	None†	[6]	CRAN	2019
corncob	Beta-binomial	None†	[41]	GitHub	2020
mixMC	PCA/sPLS-DA‡	CSS/TSS+CLR	[43]	Bioconductor	2016
maSigPro*	Generalized linear models	User specified††	[44]	Bioconductor	2014
NBME*	Negative binomial mixed effects	User specified††	[45]	CRAN	2016
MetaSplines	Gaussian + SS-ANOVA	CSS	[46]	Bioconductor	2017
MetaDprof	Gaussian + SS-ANOVA	TMM	[47]	Online	2017
MetaLonDA	Negative binomial + SS-ANOVA	TMM/CSS‡‡	[48]	CRAN	2018
NBZIMM	Negative binomial or Gaussian mixed effects	See below**	[49]	GitHub	2020

Lutz et al. Front. Appl. Math. Stats. 8 (2022)

BIOLAWEB

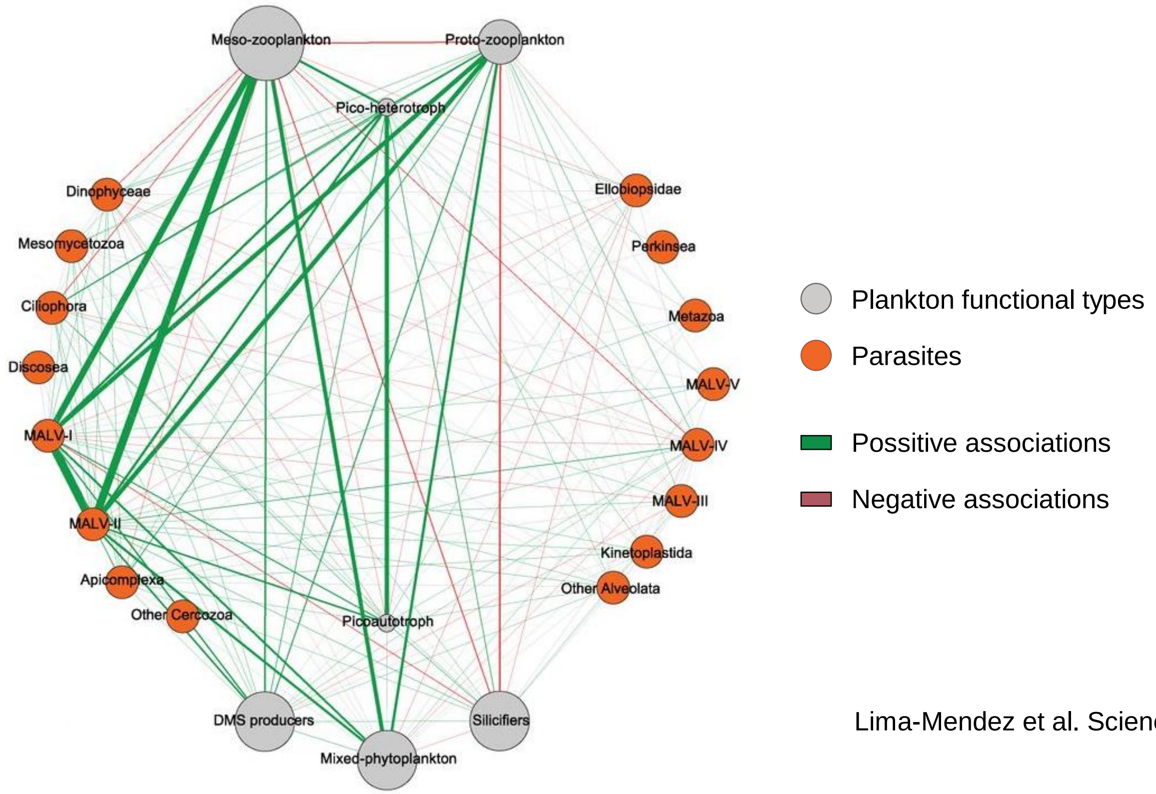
Statistical analysis workflow from metabarcoding data



Funded by
the European Union

Network analysis

Co-occurrence network:

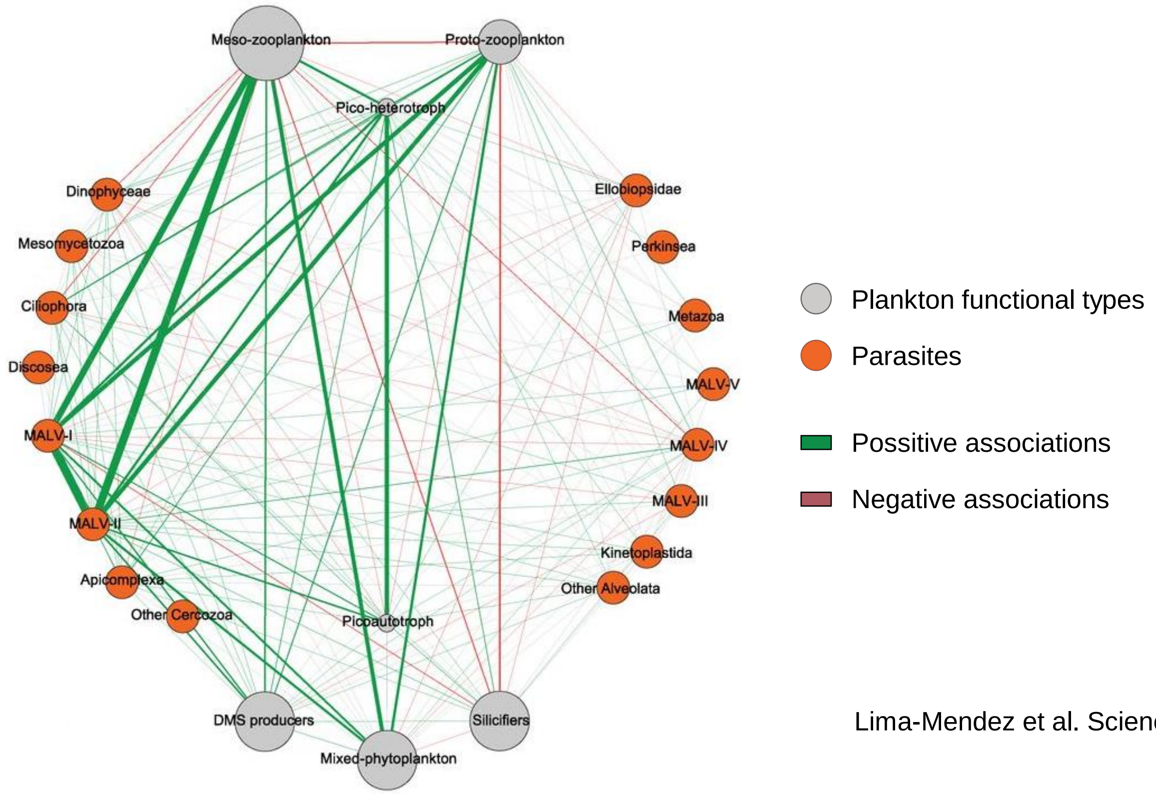


Lima-Mendez et al. Science 348 (2015)

Funded by the European Union

Network analysis

Co-occurrence network:



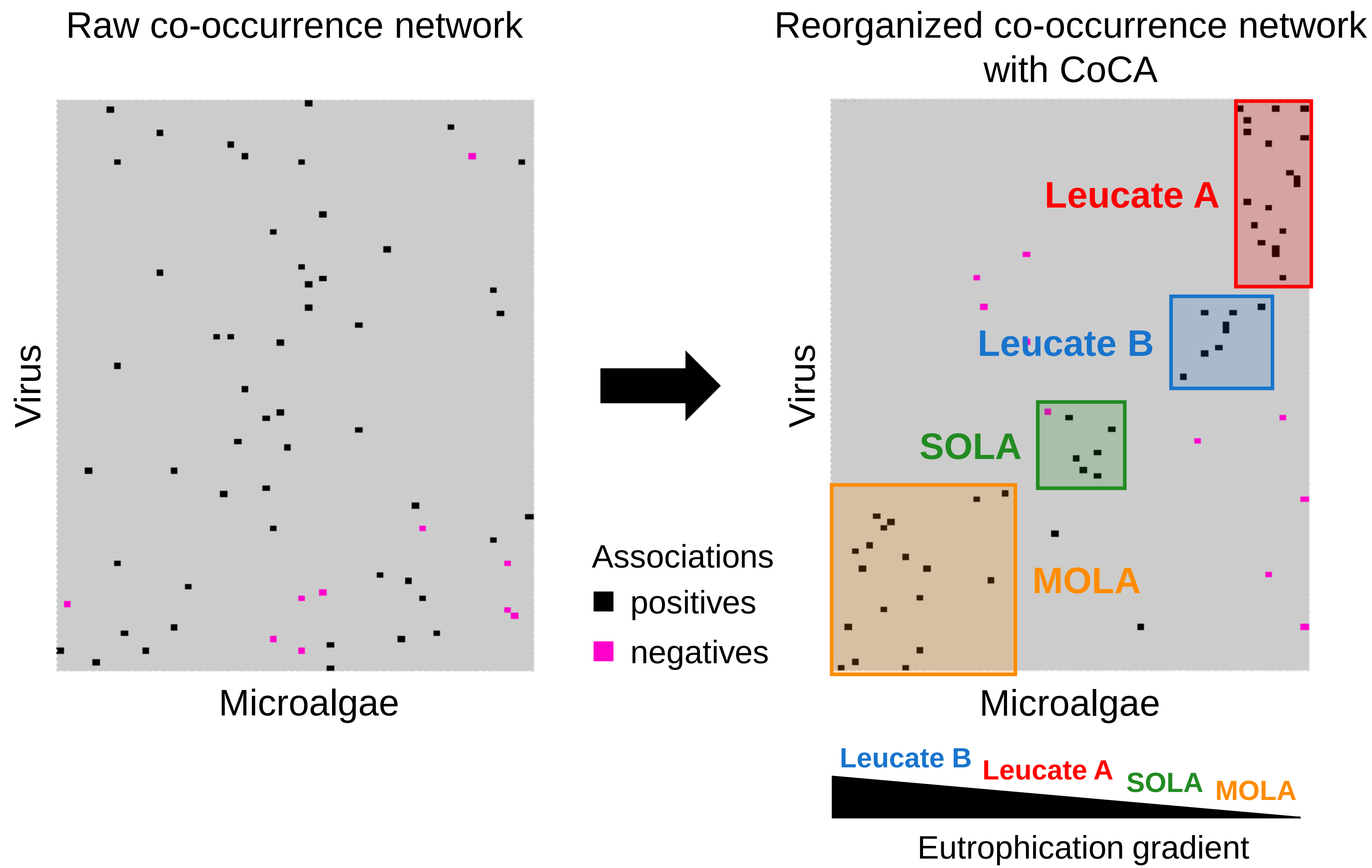
Lima-Mendez et al. Science 348 (2015)

Method	Network type	Method	Normalization	References	Application	
SparCC	Correlation	Iterative estimation of correlation	ALR	[69]	GitHub	2012
CCLasso	Correlation	Least squares with l_1 penalty	ALR	[70]	GitHub	2015
REBACCA	Correlation	Fast l_1-norm shrinkage	ALR	[71]	Online	2015
SpiecEasi	Partial correlation	Gaussian graphical model	CLR	[72]	GitHub	2015
SPRING	Partial correlation, SPR correlation	Truncated Gaussian copula model	Modified CLR	[7]	CRAN	2019
HARMONIES	Partial correlation	Gaussian graphical model	DPP	[73]	GitHub	2020

Lutz et al. Front. Appl. Math. Stats. 8 (2022)

Network analysis

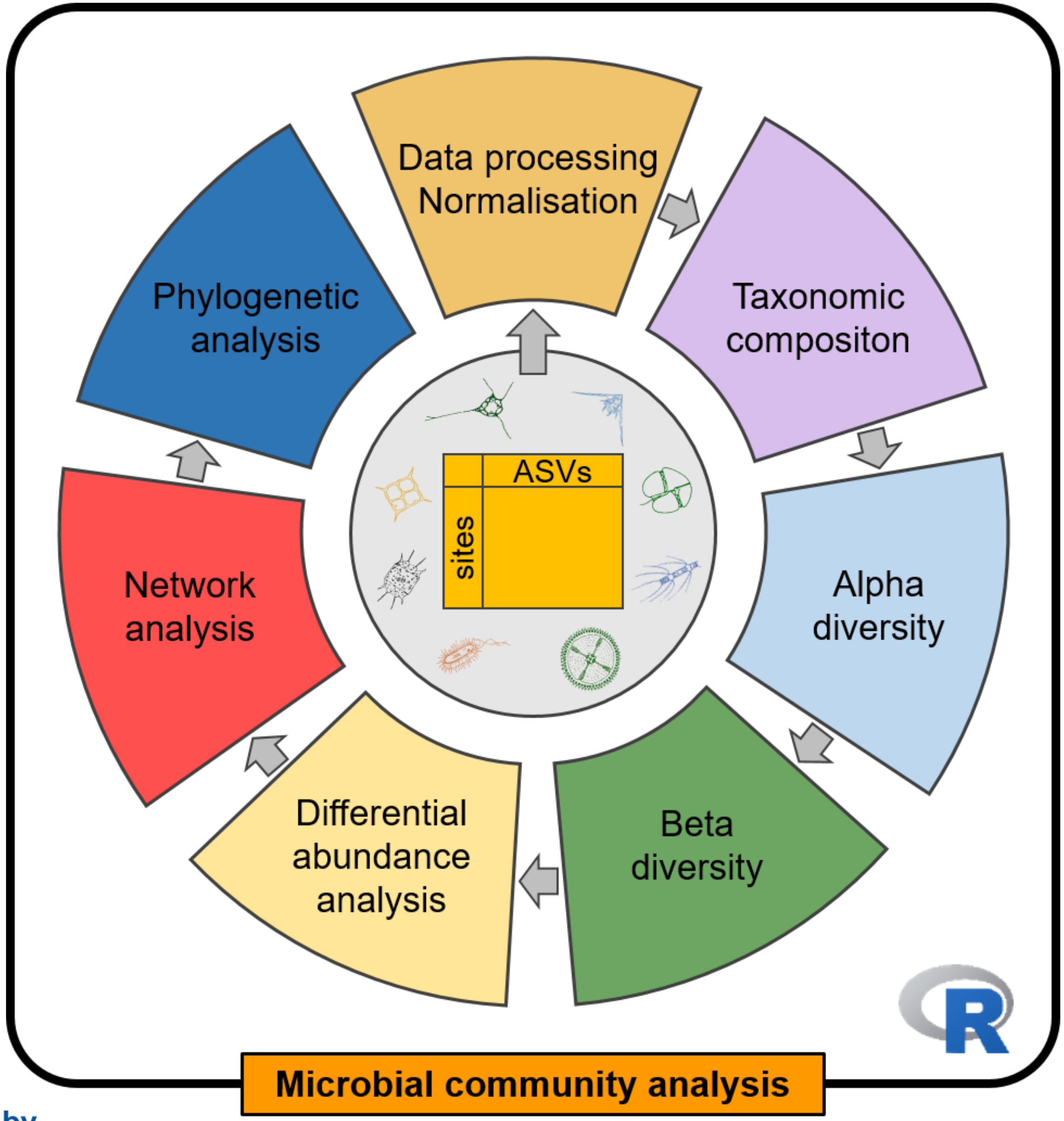
Similarity between two biological communities: Alric et al. (2020) Mol. Ecol. Res.



Spatial structuring of microalgae-virus association networks in relation to eutrophication gradient

Funded by the European Union

Statistical analysis workflow from metabarcoding data



Acknowledgement

This project has received funding from European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101079234

Thank you for your attention!

www.biolaweb.com