Cluster analysis

A **cluster** is a group of objects that, within a larger group, do not have a random or uniform occurrence, and their mutual distance or dissimilarity is smaller than the distance or dissimilarity with objects that belong to other clusters.
The **centroid** of a cluster is a hypothetical (not necessarily existing) element whose coordinates in the feature space are given by the average values of the coordinates of individual objects.

Cluster analysis

Method of cluster creation: agglomerative methods – divisive methods

Cluster arrangement: hierarchical methods – non-hierarchical methods

Cluster overlap: non-overlapping or overlapping clusters (fuzzy clustering)

Clustering procedure: sequential methods – simultaneous methods

SAHN category clustering methods:
(a) methods based on minimizing the distance between clusters
(b) methods based on optimizing the homogeneity of clusters according to a certain criterion

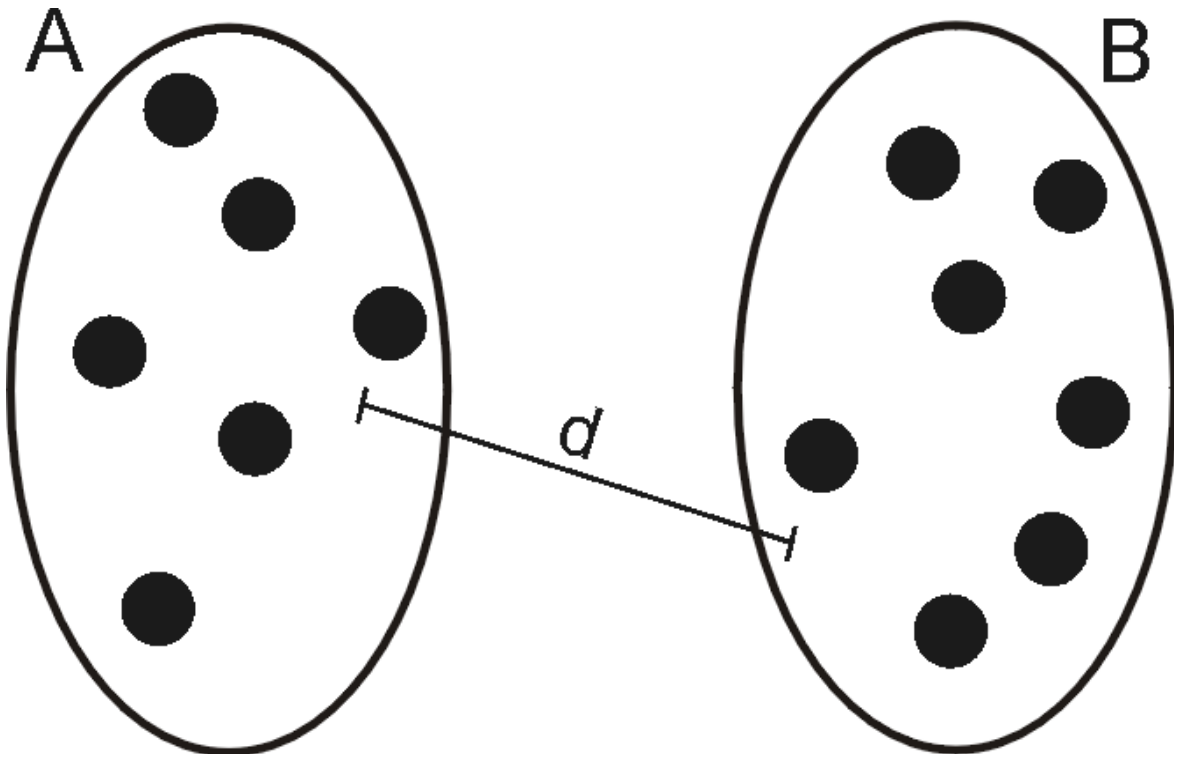
Examples of coefficients used in cluster analyses:

(a) for binary data – Jaccard coefficient, Sørensen coefficient, simple matching coefficient, Euclidean distance, chordal distance;

(b) for mixed data – Gower coefficient, distance for mixed data;

(c) for quantitative data – Euclidean distance, Manhattan metric, chordal distance.

Single linkage (the nearest neighbor method)



$D_1 =$

	1	**2**	**3**	**4**	**5**
1	0.0	1.0	7.0	4.0	12.0
2	1.0	0.0	2.0	5.0	9.0
3	7.0	2.0	0.0	8.0	3.0
4	4.0	5.0	8.0	0.0	6.0
5	12.0	9.0	3.0	6.0	0.0

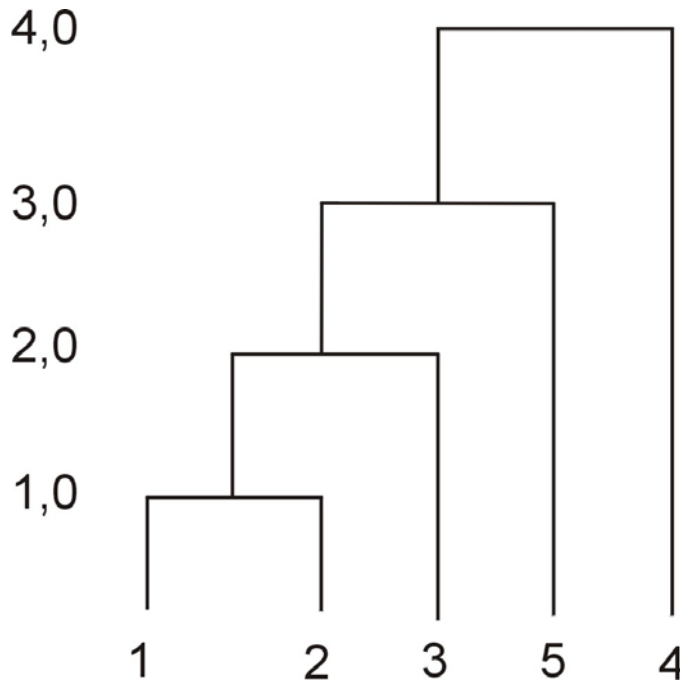
$d_{(1, 2)3} = \min \{d_{1, 3}, d_{2, 3}\} = d_{2, 3} = 2.0$

$d_{(1, 2)4} = \min \{d_{1, 4}, d_{2, 4}\} = d_{1, 4} = 4.0$

$d_{(1, 2)5} = \min \{d_{1, 5}, d_{2, 5}\} = d_{2, 5} = 9.0$

$D_2 =$

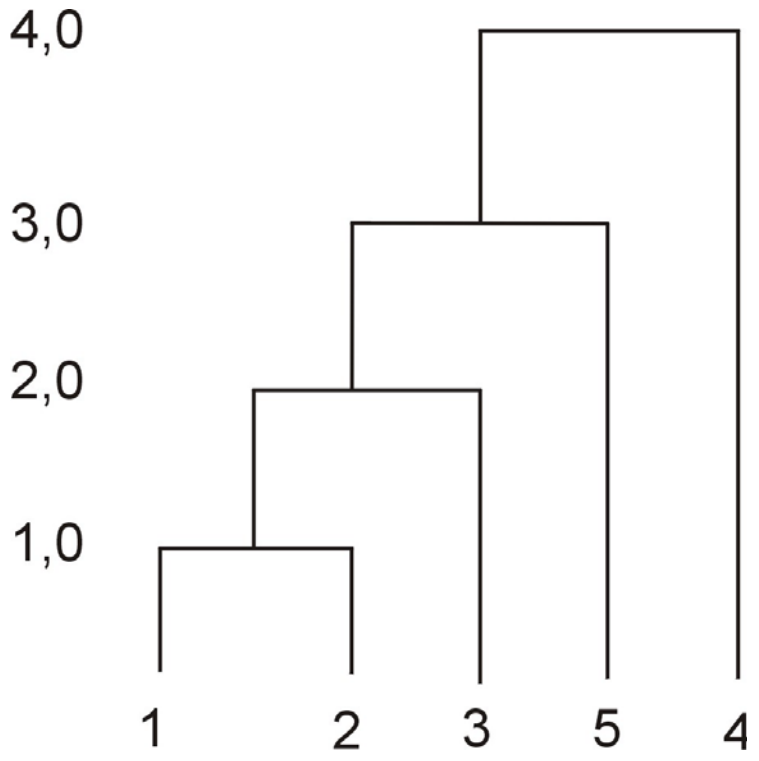
	(1, 2)	**3**	**4**	**5**
(1, 2)	0.0	2.0	4.0	9.0
3	2.0	0.0	8.0	3.0
4	4.0	8.0	0.0	6.0
5	9.0	3.0	6.0	0.0



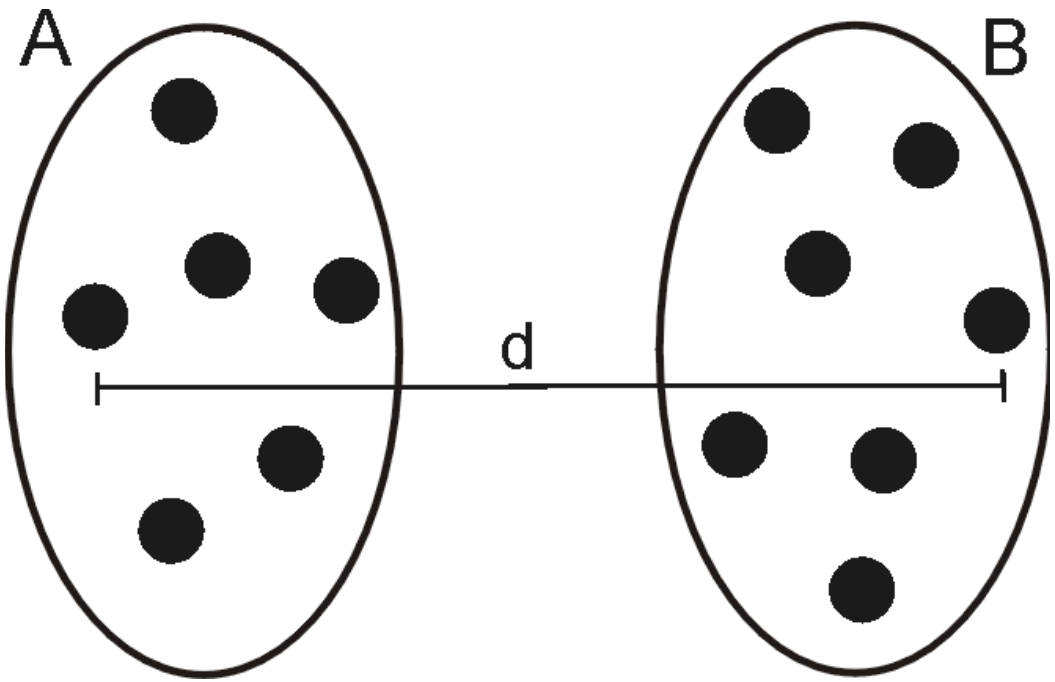
$d_{(1, 2, 3)4} = \min \{d_{(1, 2)\,4}, d_{3, 4}\} = d_{(1, 2)\,4} = 4.0$

$d_{(1, 2, 3)5} = \min \{d_{(1, 2)\,5}, d_{3, 5}\} = d_{3, 5} = 3.0$

		(1, 2, 3)	**4**	**5**
	(1, 2, 3)	0.0	4.0	3.0
$D_3 =$	**4**	4.0	0.0	6.0
	5	3.0	6.0	0.0



Complete linkage (the furthest neighbor method)



$D_1 =$

	1	**2**	**3**	**4**	**5**
1	0.0	1.0	7.0	4.0	12.0
2	1.0	0.0	2.0	5.0	9.0
3	7.0	2.0	0.0	8.0	3.0
4	4.0	5.0	8.0	0.0	6.0
5	12.0	9.0	3.0	6.0	0.0

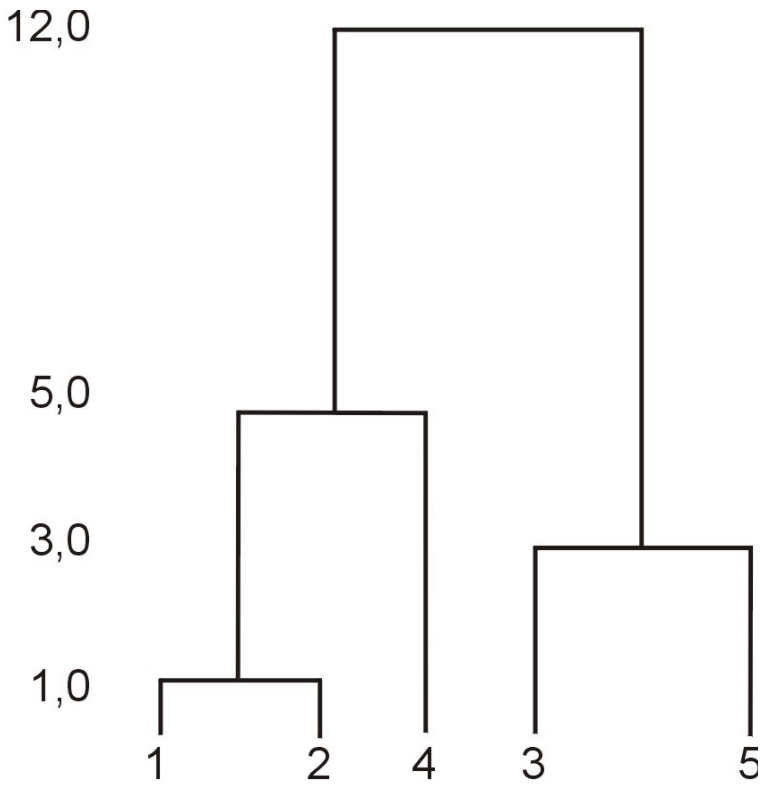
$d_{(1,2)3} = \max\{d_{1,3}, d_{2,3}\} = d_{1,3} = 7.0$

$d_{(1,2)4} = \max\{d_{1,4}, d_{2,4}\} = d_{2,4} = 5.0$

$d_{(1,2)5} = \max\{d_{1,5}, d_{2,5}\} = d_{1,5} = 12.0$

$D_2 =$

	(1, 2)	**3**	**4**	**5**
(1, 2)	0.0	7.0	5.0	12.0
3	7.0	0.0	8.0	3.0
4	5.0	8.0	0.0	6.0
5	12.0	3.0	6.0	0.0

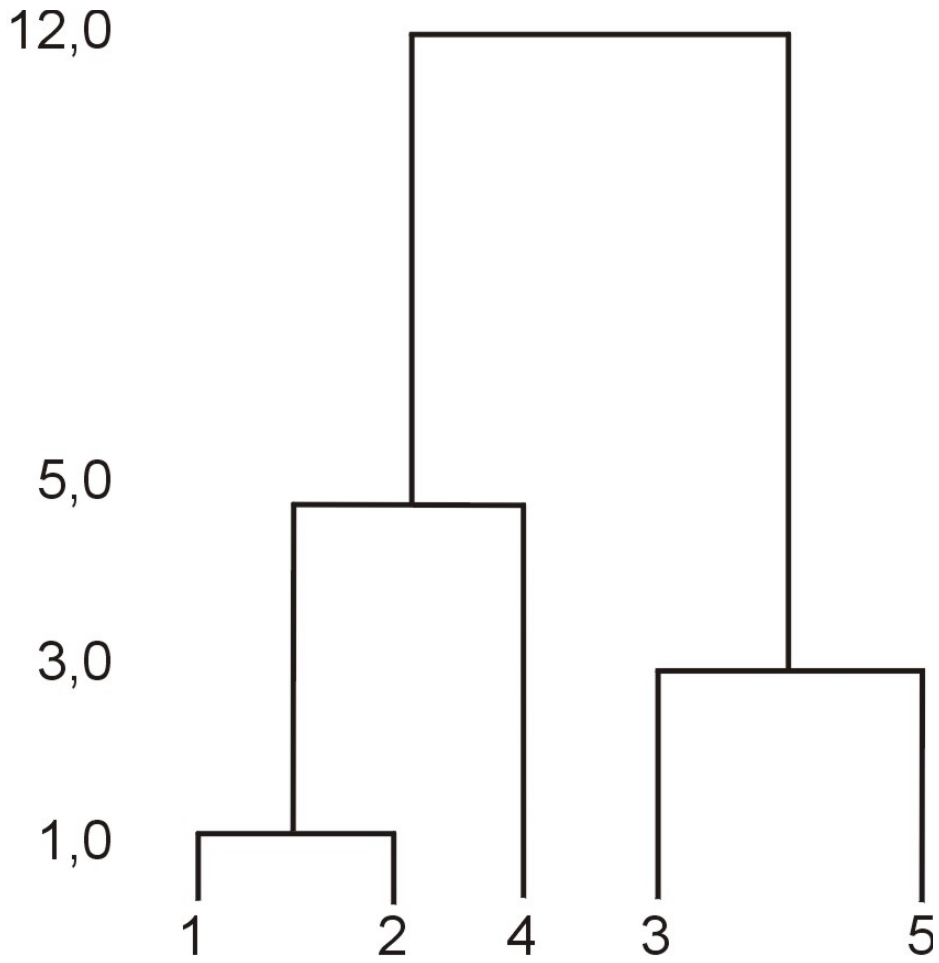


$d_{(1,\ 2)(3,\ 5)} = \max\{d_{(1,\ 2)\ 3}, d_{(1,\ 2)\ 5}\} = d_{(1,2),\ 5} = 12.0$

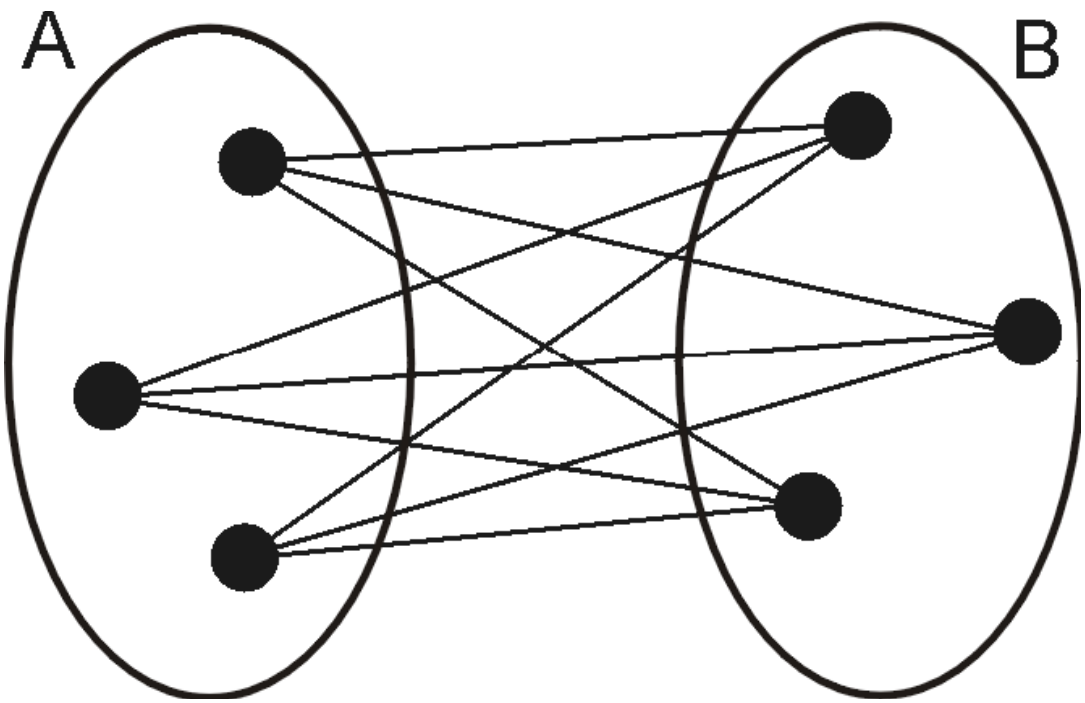
$d_{(3,\ 5)4} = \max\{d_{3,\ 4}, d_{3,\ 5}\} = d_{3,\ 4} = 8.0$

$D_3 =$

	(1, 2)	**(3, 5)**	**4**
(1, 2)	0.0	12.0	5.0
(3, 5)	12.0	0.0	8.0
4	5.0	8.0	0.0



Average linkage (UPGMA – unweighted pair-group method using arithmetic averages)



$D_1 =$

	1	**2**	**3**	**4**	**5**
1	0.0	1.0	7.0	4.0	12.0
2	1.0	0.0	2.0	5.0	9.0
3	7.0	2.0	0.0	8.0	3.0
4	4.0	5.0	8.0	0.0	6.0
5	12.0	9.0	3.0	6.0	0.0

$d_{(1, 2)3} = 1/2\ (d_{1, 3} + d_{2, 3}) = 4.5$

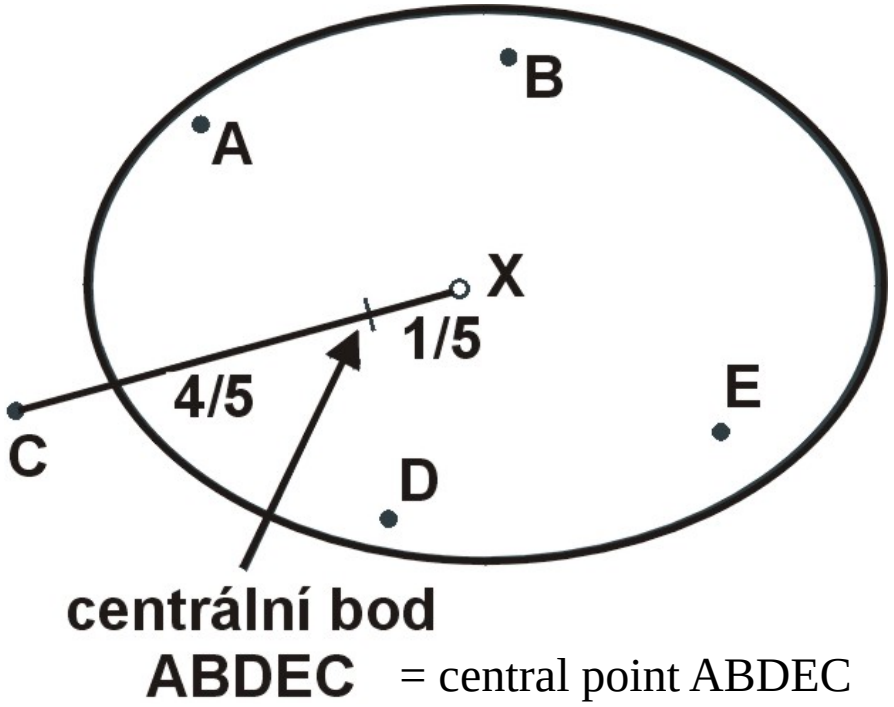
$d_{(1, 2)4} = 1/2\ (d_{1, 4} + d_{2, 4}) = 4.5$

$d_{(1, 2)5} = 1/2\ (d_{1, 5} + d_{2, 5}) = 10.5$

$D_2 =$

	(1, 2)	**3**	**4**	**5**
(1, 2)	0.0	4.5	4.5	10.5
3	4.5	0.0	8.0	3.0
4	4.5	8.0	0.0	6.0
5	10.5	3.0	6.0	0.0

Centroid method (UPGMC – unweighted pair-group method using centroids, Gower's method)



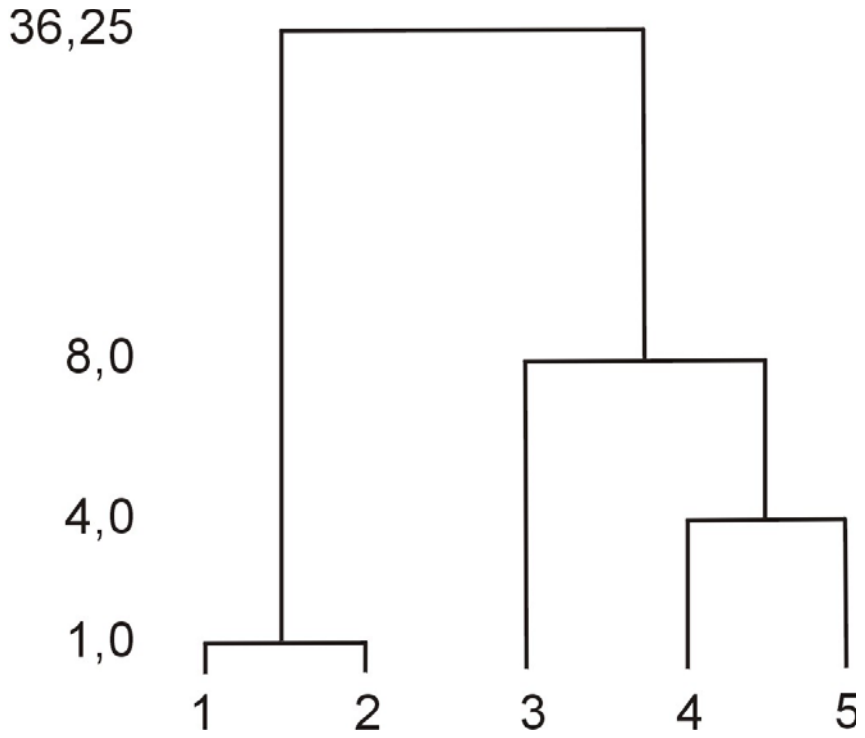
		Character (variable)	
		1	**2**
	1	1.0	1.0
	2	1.0	2.0
OTU	**3**	6.0	3.0
	4	8.0	2.0
	5	8.0	0.0

coefficient – the square of the Euclidean distance

		1	**2**	**3**	**4**	**5**
	1	0,0	1,0	29,0	50,0	50,0
	2	1,0	0,0	26,0	49,0	53,0
D_1 =	**3**	29,0	26,0	0,0	5,0	13,0
	4	50,0	49,0	5,0	0,0	4,0
	5	50,0	53,0	13,0	4,0	0,0

Character (variable)

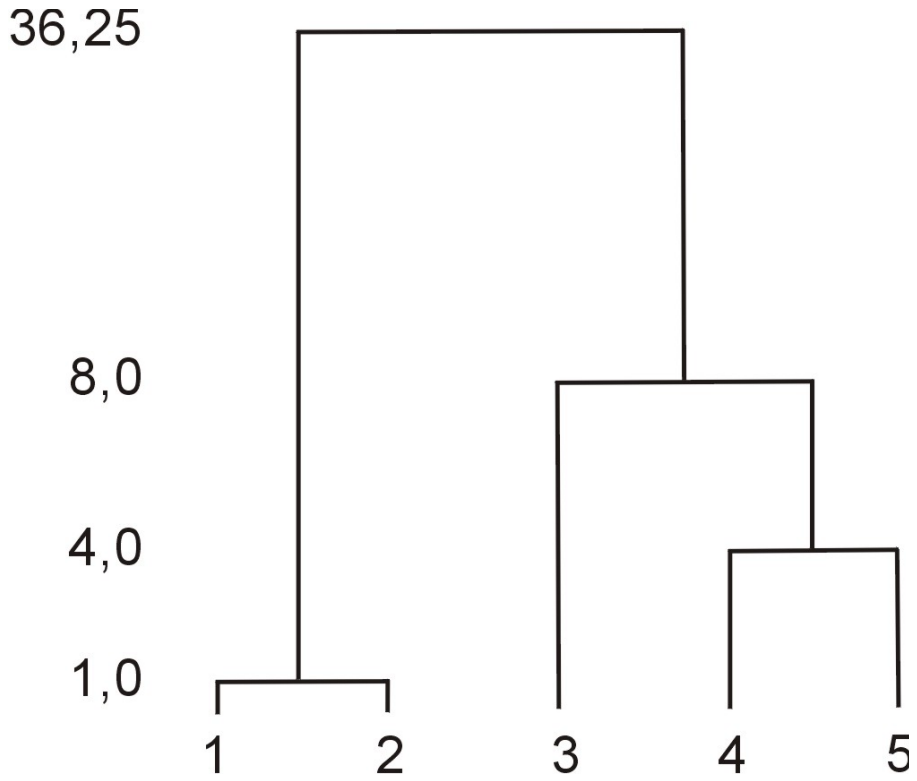
	1	**2**
(1, 2)	1.0	1.5
3	6.0	3.0
4	8.0	2.0
5	8.0	0.0



D_2 =

	(1, 2)	**3**	**4**	**5**
(1, 2)	0.0	27.25	49.25	51.25
3	27.25	0.0	5.0	13.0
4	49.25	5.0	0.0	4.0
5	51.25	13.0	4.0	0.0

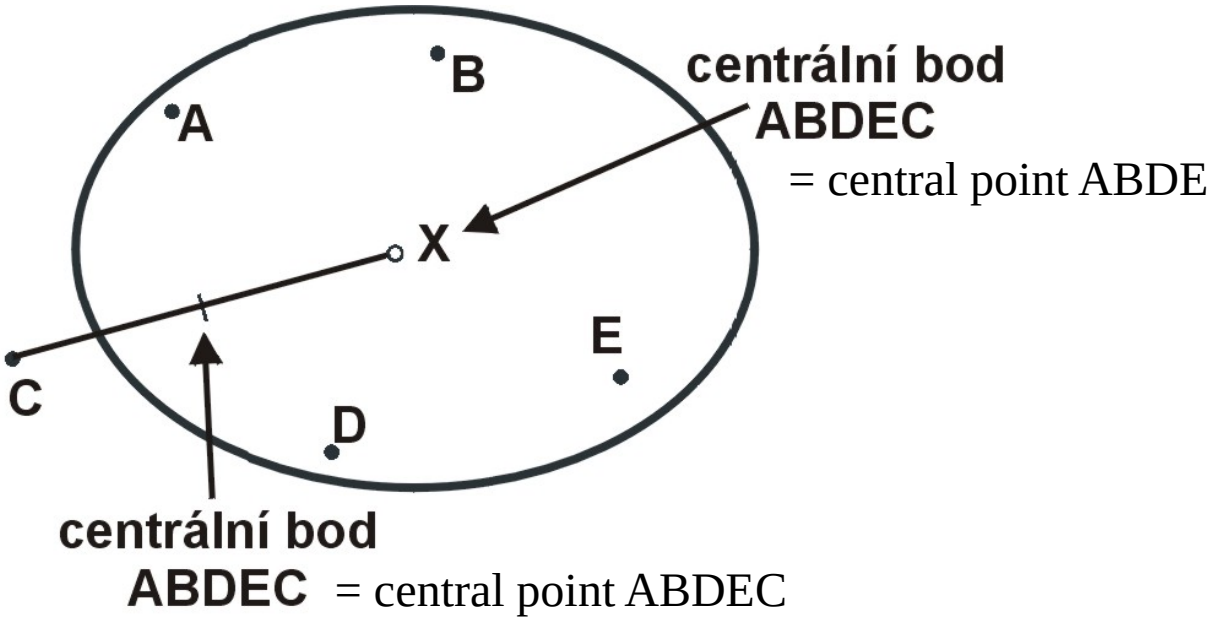
	znak	
	1	**2**
(1, 2)	1.0	1.5
3	6.0	3.0
(4, 5)	8.0	1.0



$D_3 =$

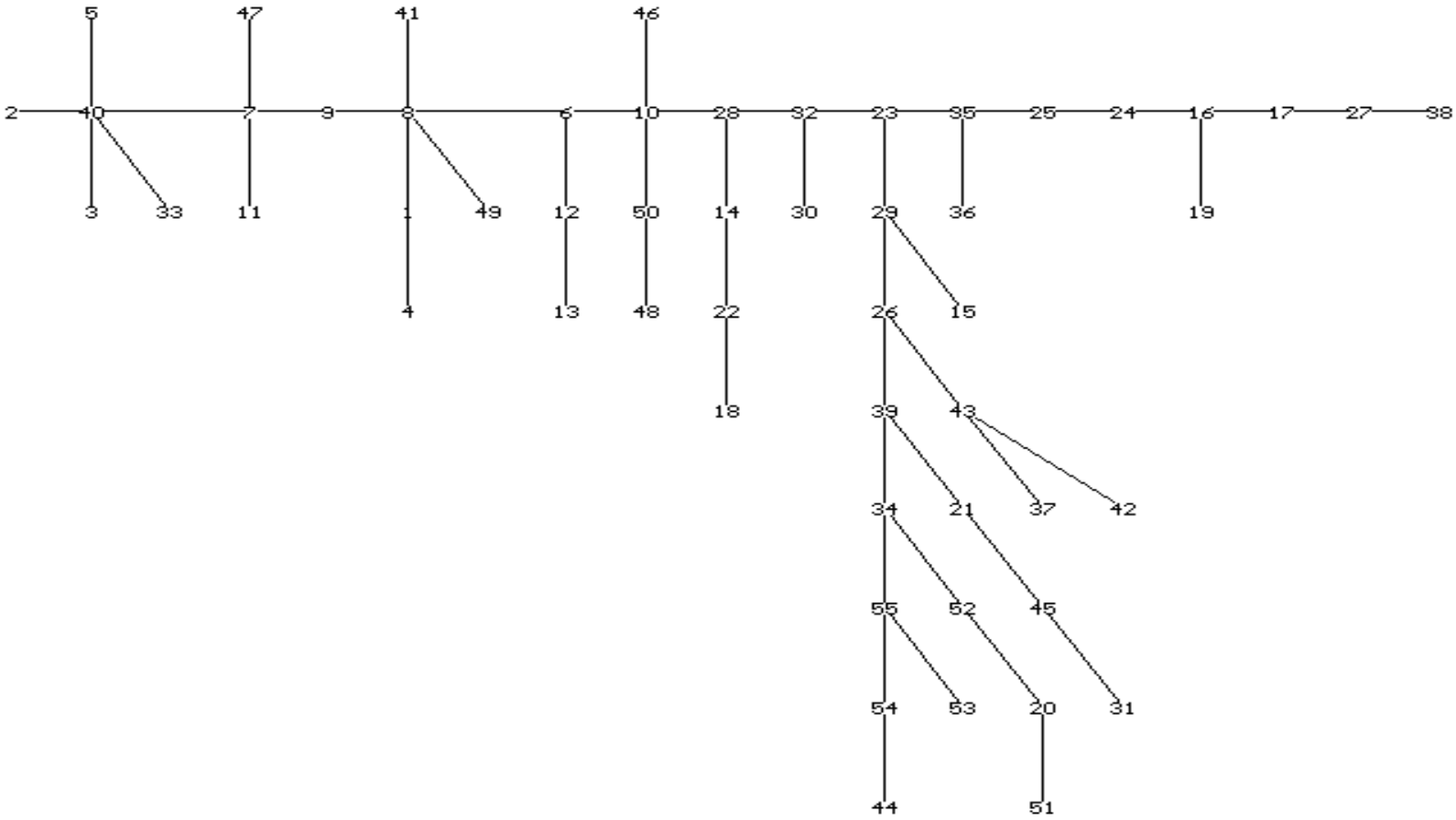
	(1, 2)	**3**	**(4, 5)**
(1, 2)	0.0	27.25	49.25
3	27.25	0.0	8.0
(4, 5)	49.25	8.0	0.0

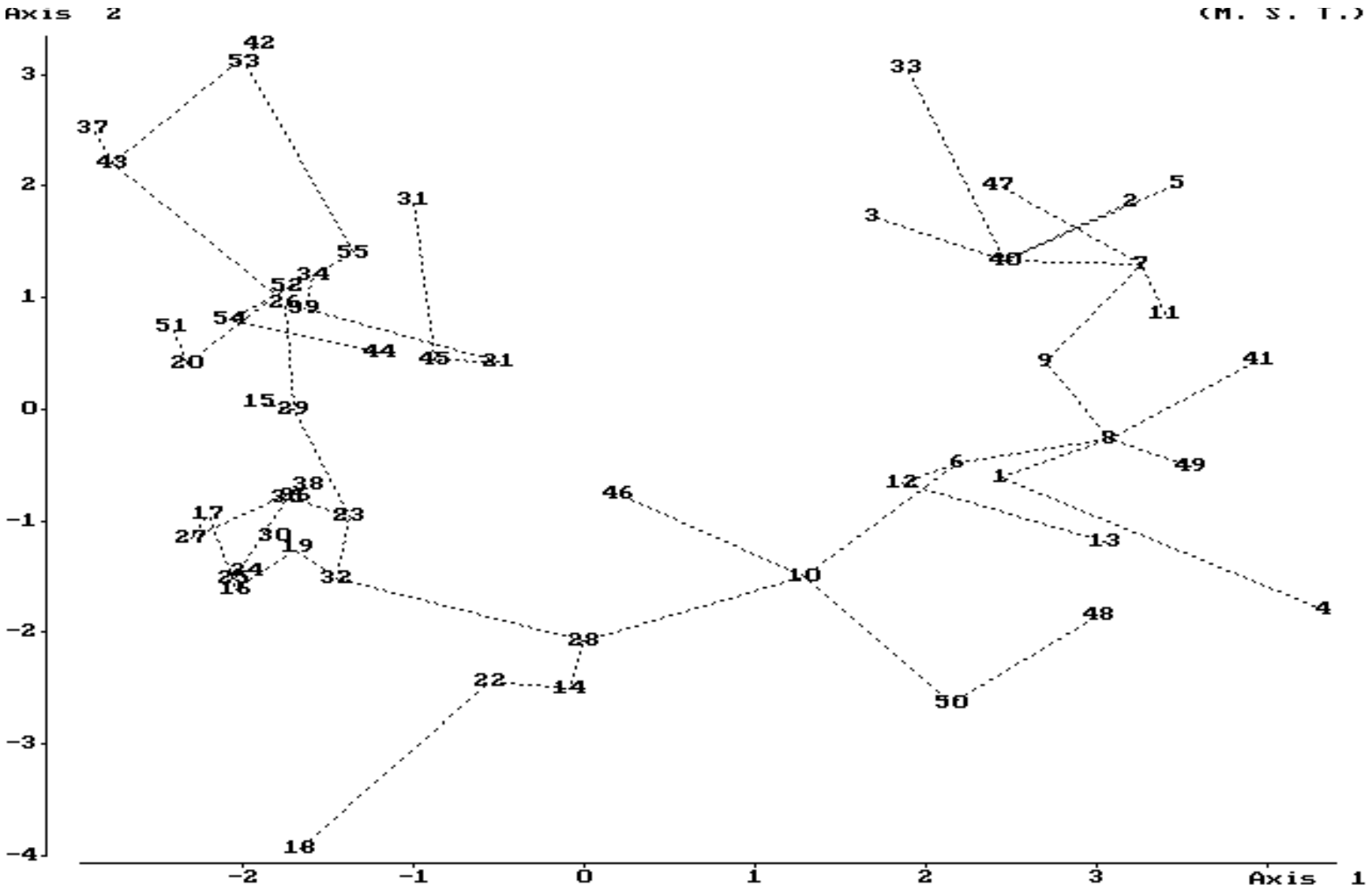
Median method (WPGMC – weighted pair-group method using centroids, weighted centroid clustering)



Minimum spanning tree

A graph that connects all objects in such a way that no loops or circles occur, and at the same time, the sum of the lengths of the connections between the nodes (objects) is minimal.




Axis 2
(M. S. T.)
Axis 1
3
2
1
0
-1
-2
-3
-4
-2
-1
0
1
2
3
42
53
37
43
55
31
34
52
26
29
51
54
44
45
41
20
15
38
17
23
27
30
19
32
16
46
10
28
22
14
18
33
47
5
2
3
40
7
11
9
41
8
49
6
1
12
13
4
48
50

General notes/comments on clustering methods

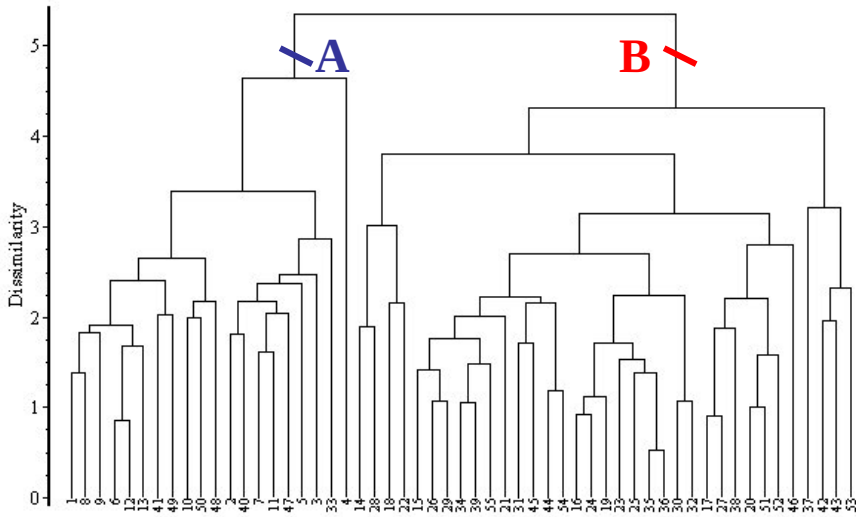
If the data do not have a completely unambiguous and clear structure (e.g., more or less randomly distributed objects), it is likely that using different clustering techniques will yield different results.

If various clustering techniques provide identical or similar results from the same dataset, this to some extent confirms the structure contained within the data (although clustering methods are hypothesis-generating procedures and are not meant for hypothesis testing).

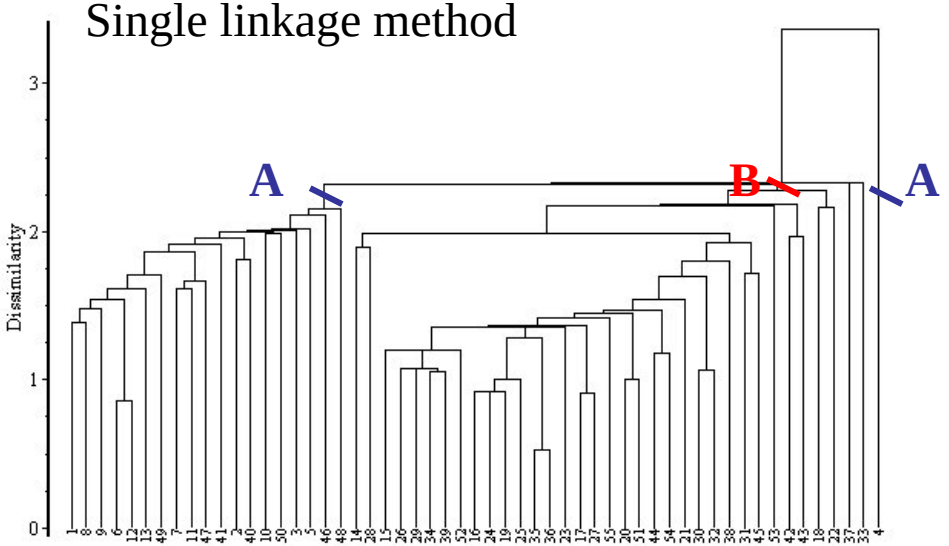
Many clustering techniques are sensitive to the presence of outliers (highly atypical cases). Before performing the actual cluster analysis, it is advisable to use some method for their detection, such as principal components analysis. Significant outliers are usually excluded from further analyses.

Cluster analyses are generally not suitable for data describing clinal variability of traits (cline = variability of a trait dependent on an environmental gradient).

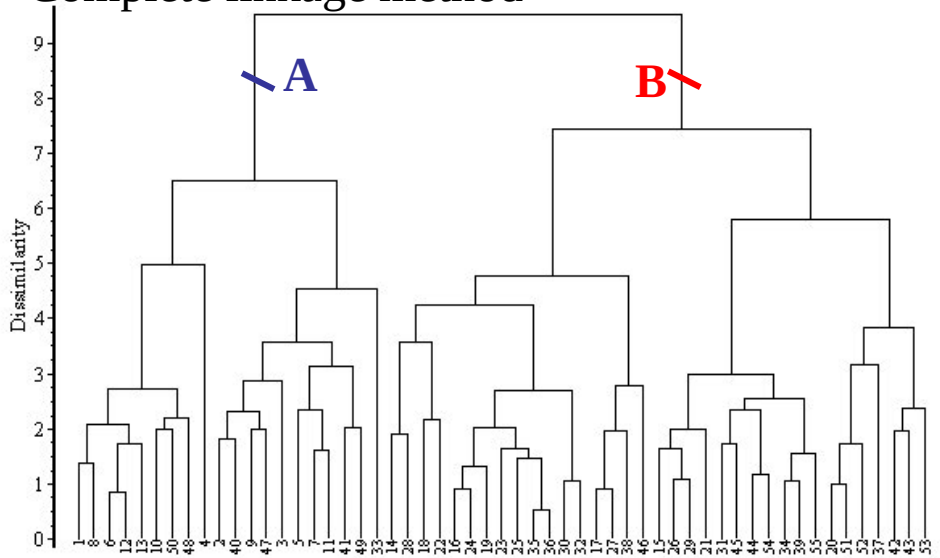
Average linkage method



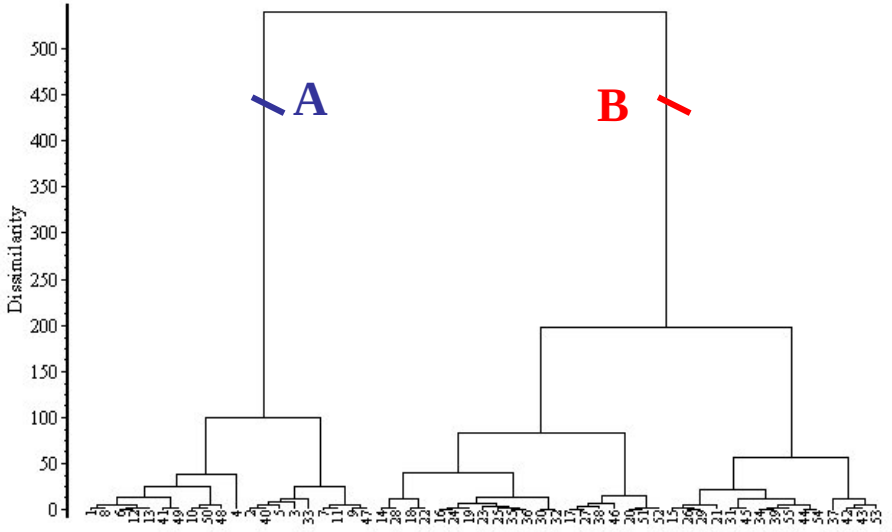
Single linkage method



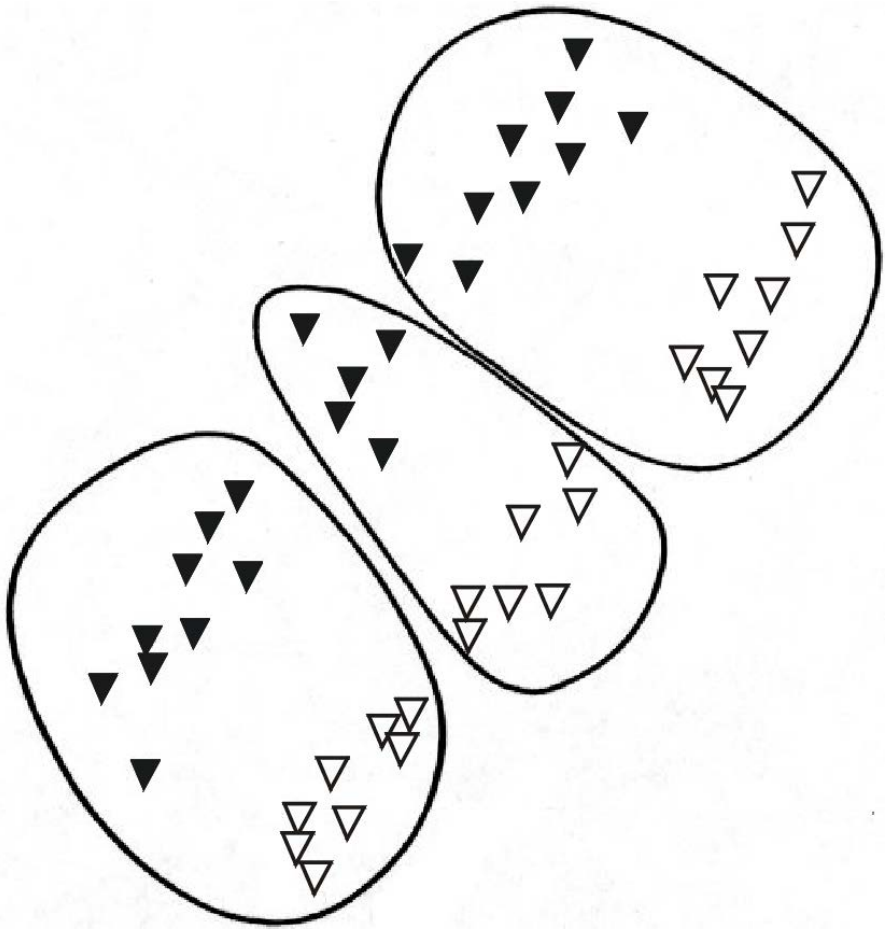
Complete linkage method



Ward's method

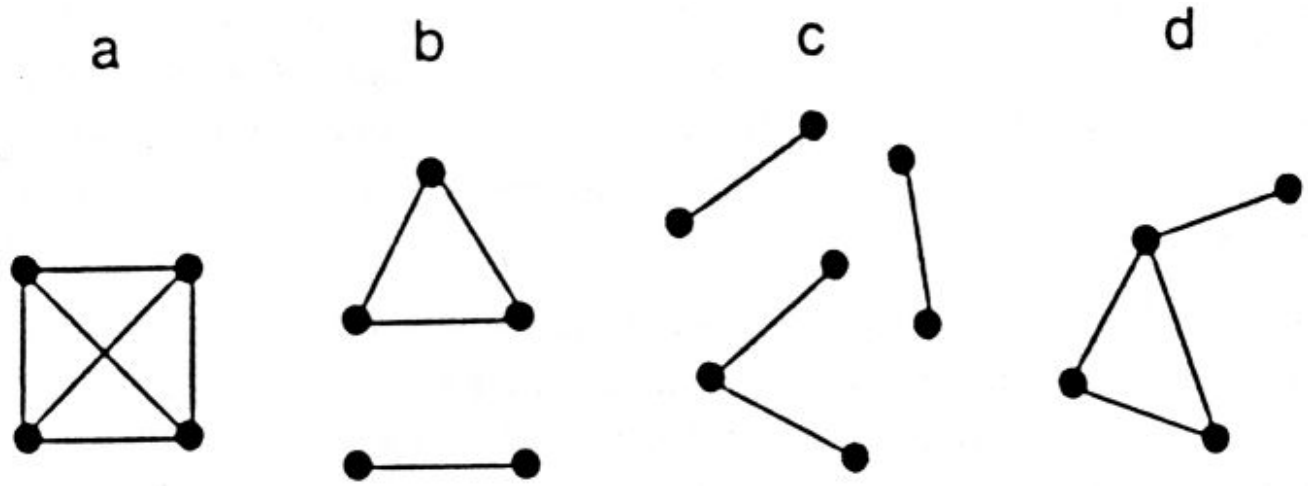


Clear support is for two taxa, further groupings reflect differences in clustering algorithms.



The single linkage method would, as a result of the chaining effect, group the filled triangles into one cluster and the empty triangles into another, whereas Ward's method and the average linkage method would produce groups bounded by lines (according to Everitt & Dunn 1983).

Ties

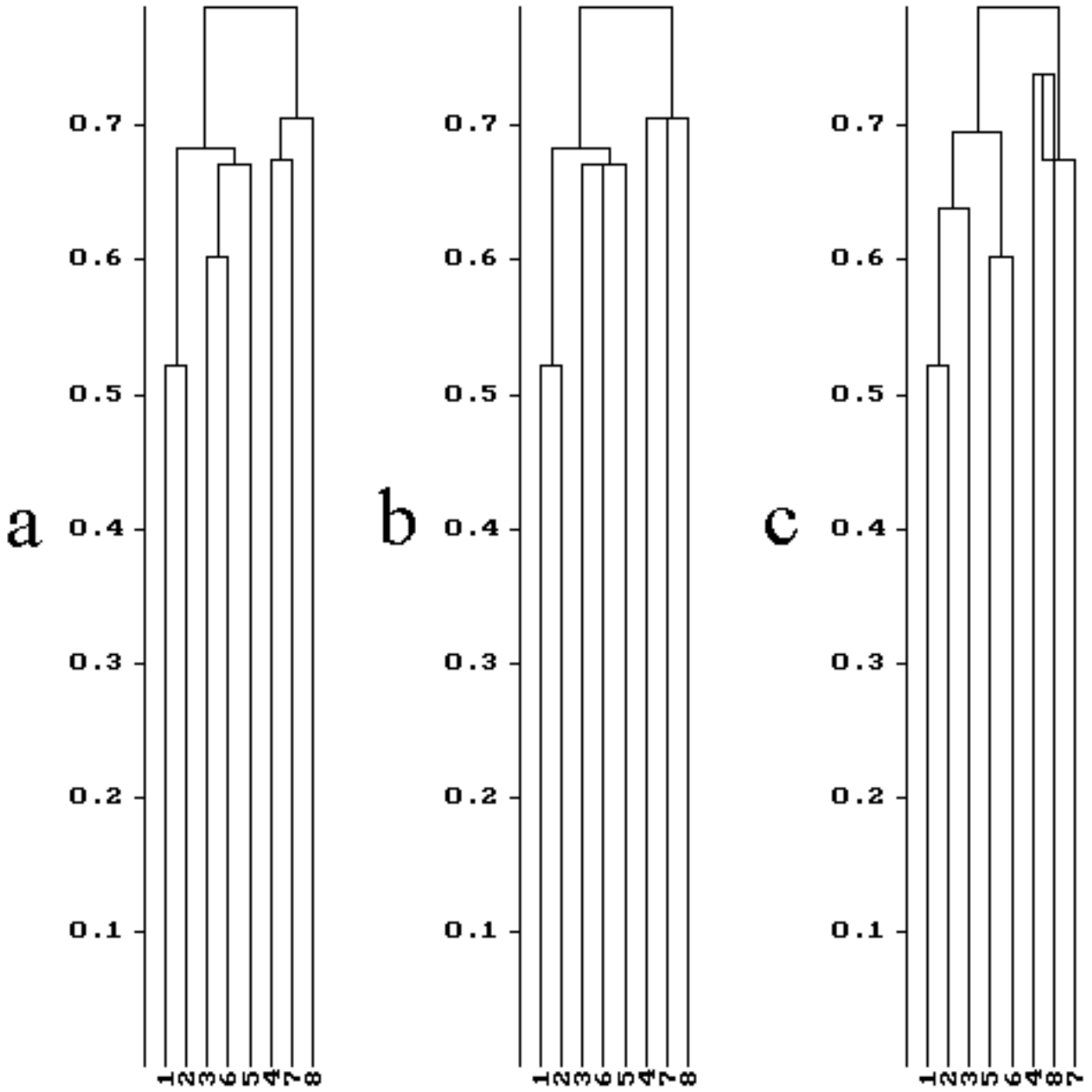


a – the graph is complete, b – the graph is disconnected and all isolated components are complete, c – the graph is disconnected and at least one component is not complete, d – the graph is connected but not complete.

(a) „*silent mode (arbitrary)*“

(b) „*single linkage*“

(c)„*suboptimal fusions*“

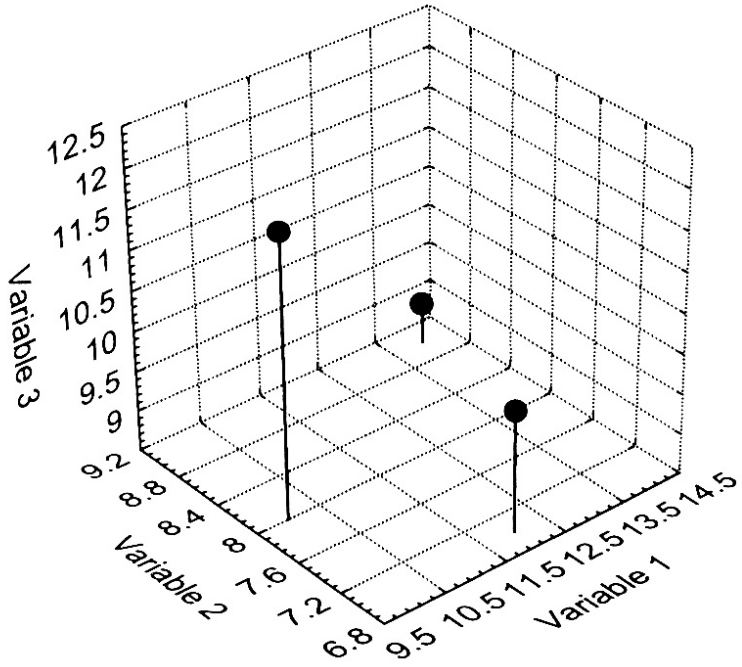


Ordination methods

Objects characterized by *p* variables can be imagined as points in a *p*-dimensional space, where each dimension represents the values of one variable.

If we work with only two or three variables, it is possible to observe the relationships between objects. their distances, and clustering without three-dimensional graph.

A larger number of variables => necessity to reduce their number with the least possible loss of information.



Ordination methods

Principal components analysis (PCA)

Principal coordinates analysis (PCoA)

Non-metric multidimensional scaling (NMDS)

Useful information about ordination methods can be found on the website

http://ordination.okstate.edu/

OSU ECOLOGY

Ordination Methods for Ecologists

JUMP TO: | OVERVIEW | Ordination Topics | Ordination Software Links | Ordination Glossary | Other Ordination Links | Ordination Listserv | OSU Botany

Members of the Laboratory for Innovative Biodiversity Research And Analysis (LIBRA) are often available to engage in consulting activities for particular projects, or to offer short courses on ordination methods and the use of CANOCO. For more information, contact Mike Palmer at mike.palmer@okstate.edu.

Ordination is a widely-used family of methods which attempts to reveal the relationships between ecological communities. For definitions, go HERE.

This ordination web page is designed to address some of the most frequently asked questions about ordination. It is my intention to gear this page towards the student and the practitioner rather than the ordination specialist, so please contact me if the jargon is unintelligible!

The ecological literature is filled with papers describing, contrasting, and modifying existing ordination techniques. Then why is an ordination web page needed? My main motivation is based upon the following observation: many of us, when we start to use ordination methods, make the same simple mistakes. If we are good scientists, we will learn from our own mistakes. But wouldn't it save a lot of time if we could also learn from other people's mistakes?

It turns out that there are a number of concerning ordination, as well as a number of "tricks of the trade"

Ordination Topics

General and Reference

- Overview of ordination methods
- A Glossary for terms used in Ordination
- Milestones in the history of Ordination
- Ordination terminology: some confusions
- The ideal ordination method
- Recommendations for ordination: a key
- Suggested references for self-education
- Hypothesis-driven and Exploratory Analysis

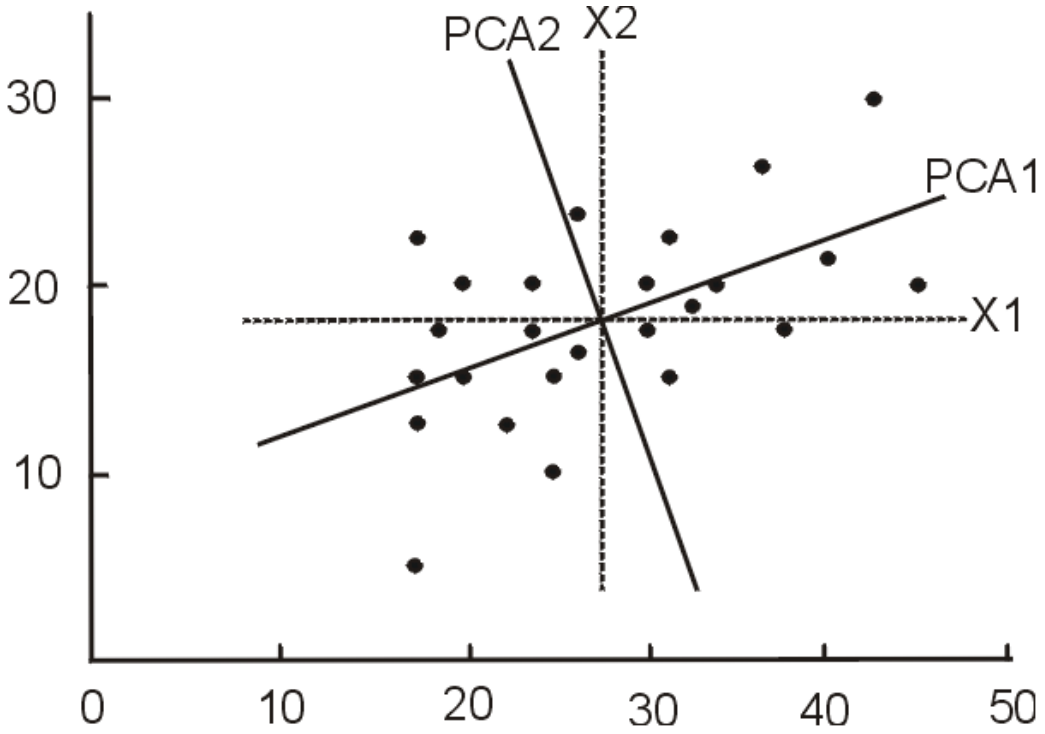
Indirect Gradient Analysis

- Distance-based ordination methods
- Eigenanalysis-based ordination methods
- Principal Components Analysis
- Correspondence Analysis
- Detrended Correspondence Analysis

https://ordination.okstate.edu/

PCA – principal component(s) analysis

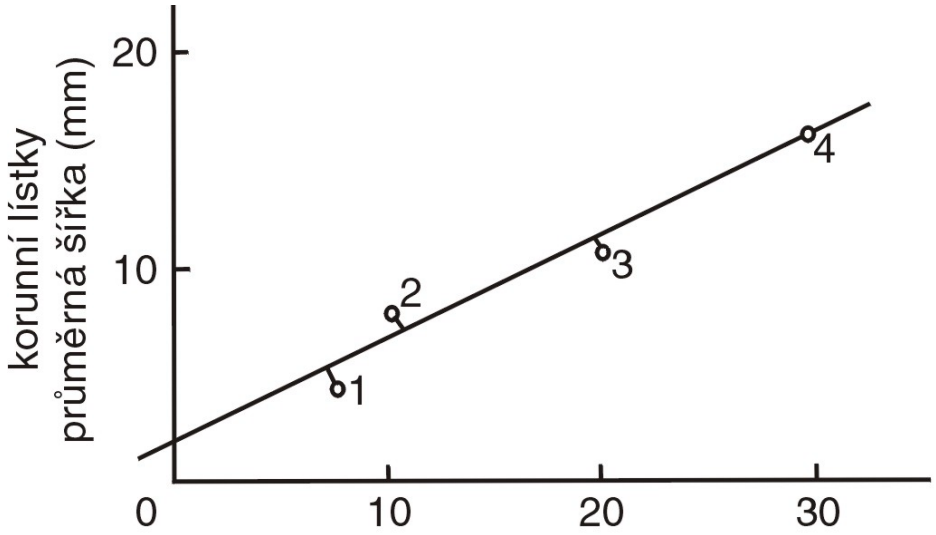
Replaces the original set of observed variables with a set of new (hypothetical) variables that are mutually uncorrelated, so that the first new axis (the first principal component, PC1, the first new variable) is directed in the direction of the greatest variability among objects, and the second axis (the second principal component, PC2, the second new variable) is directed in the direction of the greatest variability, which is perpendicular to the direction of the first component, and so on..



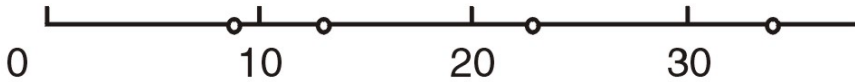
Geometric interpretation of PCA (podle Dunn & Everitt 1982):

OTU	**1**	**2**	**3**	**4**
Average length of petals (mm)	8	10	20	30
Average width of petals (mm)	4	9	11	18

=Average width of petals (mm)



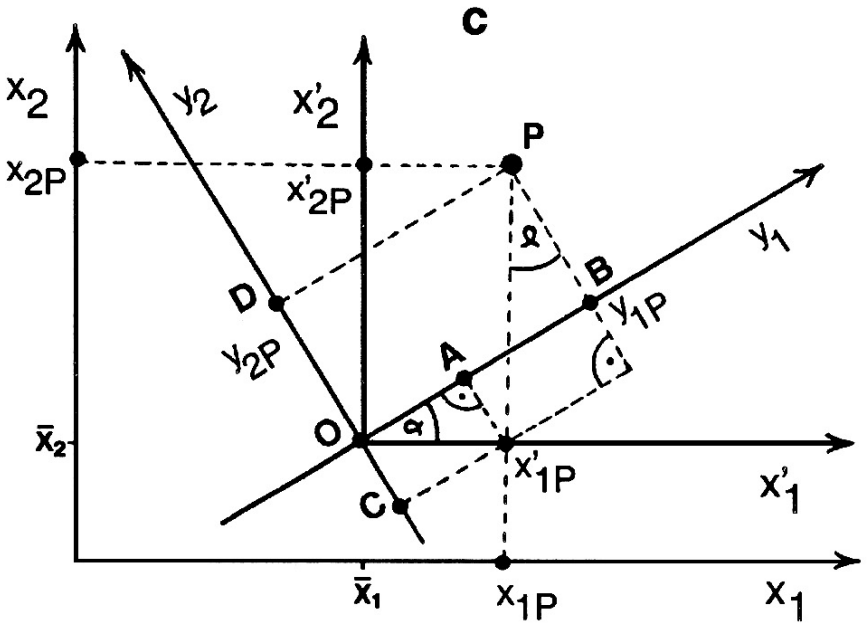
=Average length of petals (mm)



= first principal component

PCA – principal component(s) analysis

It is based on eigenanalysis of symmetric matrices (correlation, covariance matrices)



Objective of PCA: determining the angles between the original and new axes of the coordinate system, and the coordinates of objects in the new coordinate system.

Original set of p observed variables $x_1, x_2, ..., x_p$

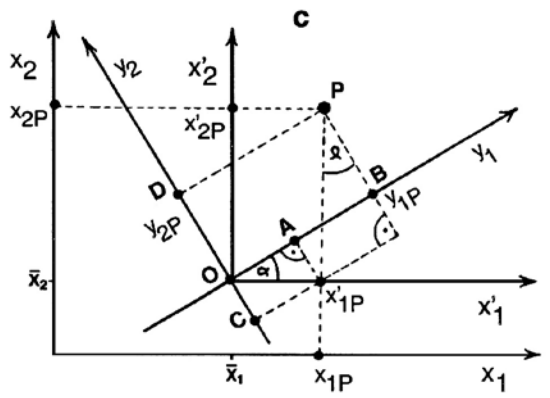
is transformed into a new set of variables $y_1, y_2, ..., y_p$

$y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p$

.

.

$y_p = a_{p1}x_1 + a_{p2}x_2 + ... + a_{pp}x_p$



Coefficients of the first principal component – vector $\boldsymbol{a}_1$

the first principal component $y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p$ expressed as vector $\boldsymbol{a}_1'\boldsymbol{x}$

Similarly $y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2p}x_p$ can be expressed as $\boldsymbol{a}_2'\boldsymbol{x}$ etc.

Components are not mutually correlated
Therefore applies: $\boldsymbol{a}_2'\boldsymbol{a}_1 = 0$

The sum of the squares of the coefficients of each linear combination equals one $\boldsymbol{a}_1'\boldsymbol{a}_1 = 1$, $\boldsymbol{a}_2'\boldsymbol{a}_2 = 1$ etc.

In general, for the *j*-th principal component, the following applies $y_j = \boldsymbol{a}_j'\boldsymbol{x}$

and this has the greatest variance under the conditions that $\boldsymbol{a}_j'\boldsymbol{a}_j = 1$ and $\boldsymbol{a}_j'\boldsymbol{a}_i = 0$, $i \neq j$.

A symmetric matrix S_{pp} (such as a covariance or correlation matrix) can be assigned *p* real eigenvalues (characteristic numbers, characteristic roots, latent roots) $\lambda_1 \ldots \lambda_p$ and *p*-column vectors of eigenvectors (characteristic vectors, characteristic vectors, latent vectors) $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$, with the condition that $S_{pp} = A_{pp} \Lambda_{pp} A_{pp}'$.

It is possible to prove that the coefficient vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_p$ are the eigenvectors of the covariance or correlation matrix; if the sum of their squares is 1 (see above $\boldsymbol{a}_1'\boldsymbol{a}_1 = 1$), the eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_p$ of this matrix can be interpreted as measures of variance captured by the components $y_1, \ldots, y_p$

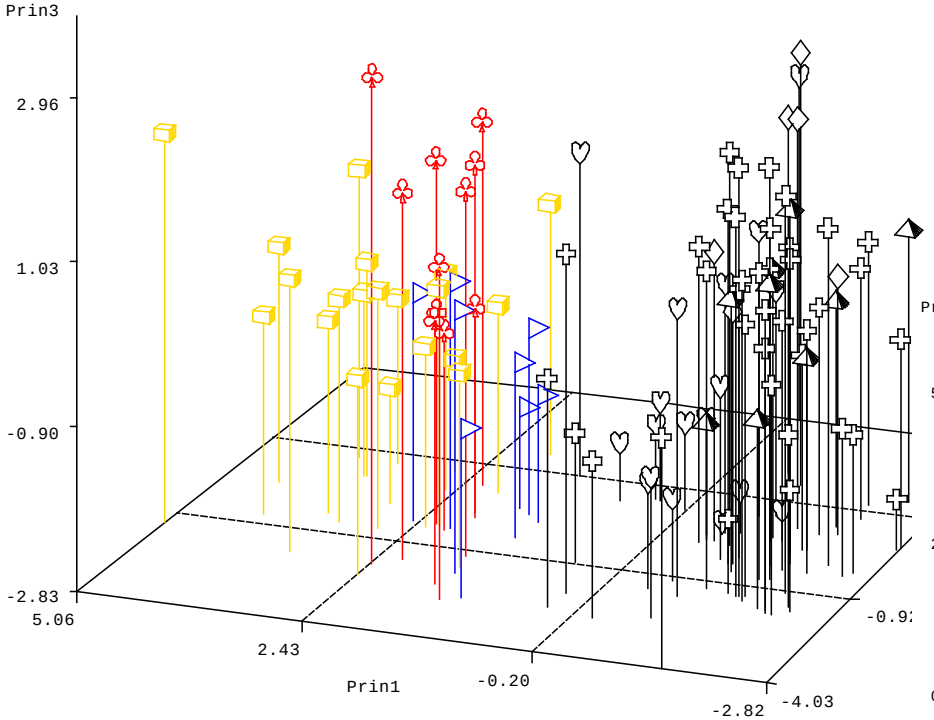
PCA – principal component(s) analysis

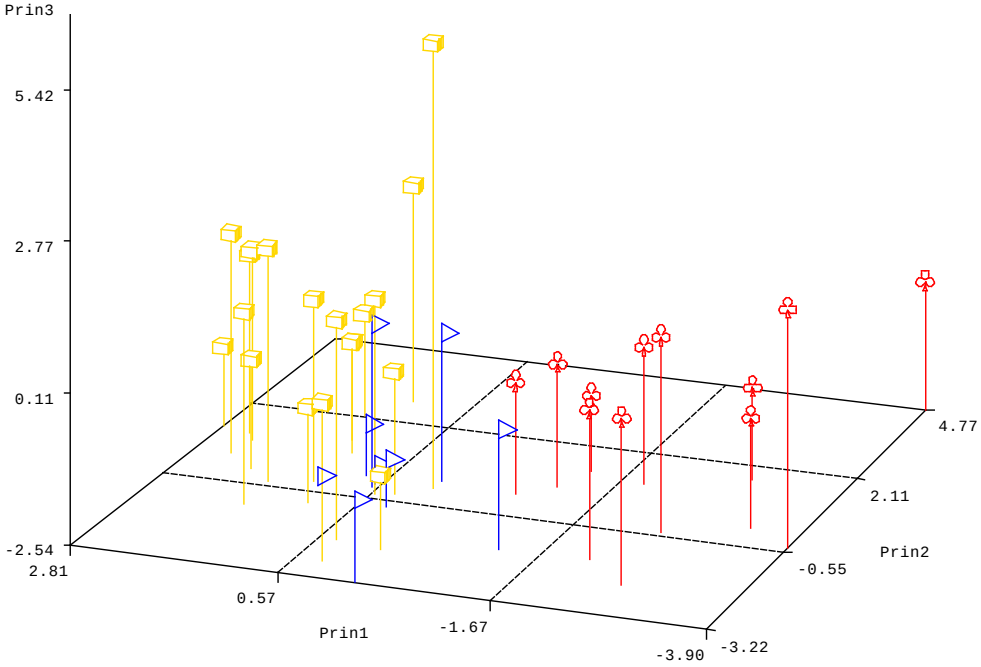
The number of objects in PCA must be at least one greater than the number of analyzed variables.

However, it is usually recommended that the number of objects approaches the square of the number of variables (this relates to the degrees of freedom).

In the case where $n \leq p$ (where n is the number of objects and pp is the number of variables), the resulting matrix (correlation or covariance) of order p has only $n - 1$ independent rows or columns. In such a case, the corresponding matrix has $p - (n - 1)$ zero eigenvalues (to position n objects based on their mutual distances, only $n - 1$ dimensions are needed).

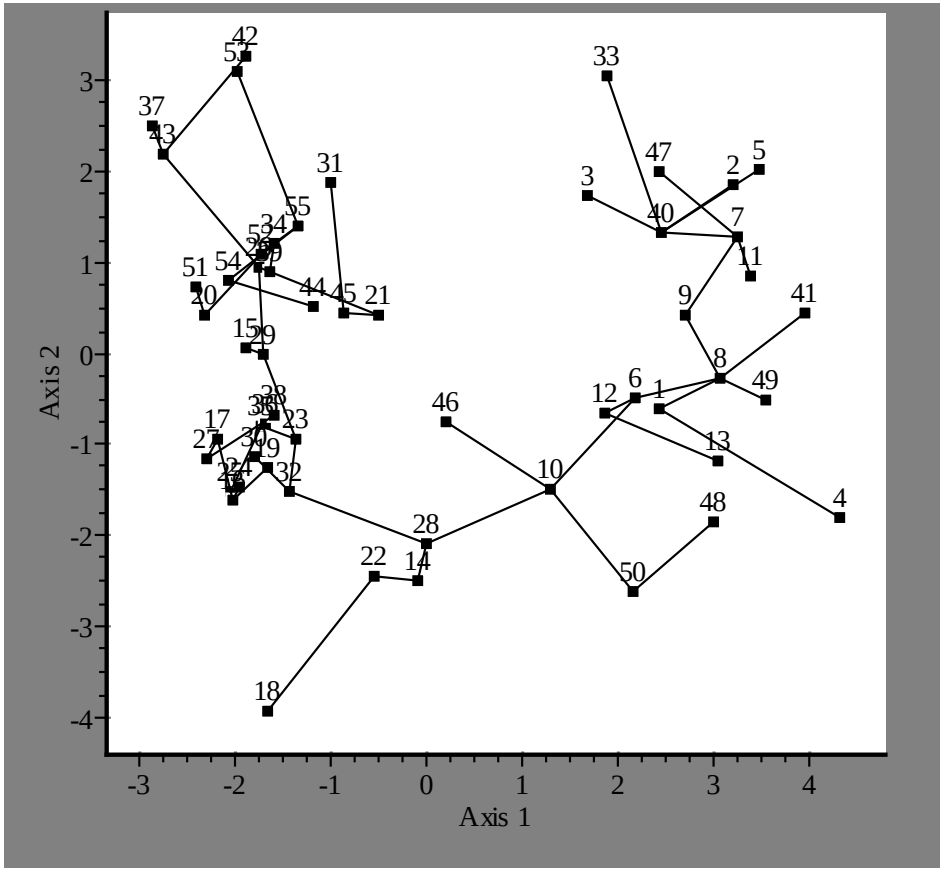
Interpretation of the ordination of objects can be complicated if the data contains a very complex structure. For example, if two basic groups (separated along the first axis) are internally divided in a more complicated way, the second, third, and further axes tend to be a compromise between the structure in both basic groups.
=> It is advisable to analyze each group (separated by the first axis) separately in further steps.





7 subspecies (defined by the degree of ploidy and range)

Although the PCA technique was originally designed for quantitative traits, it can also be used for binary and semi-quantitative traits. However, binary data tend to cause the so-called 'horseshoe effect', where objects in the plane defined by the first two components are arranged in the shape of a horseshoe.
This curvature is corrected by so-called detrended techniques, for example, in the DECORANA program (Hill 1979), but such 'straightening' is generally not used in taxonomic applications



Cardamine amara (Brassicaceae)

subsp. *amara*

subsp. *opicii*

Correlation matrix

```
RESEMBLANCE MATRIX

  ROW   1
  0.1000E+01 -0.3509E-01  0.6541E+00 -0.2760E+00 -0.2020E-01
  0.5767E+00  0.7991E+00 -0.3839E+00  0.8776E+00  0.7924E+00
  ROW   2
 -0.3509E-01  0.1000E+01  0.3224E+00  0.4878E+00  0.7738E+00
  0.2010E+00 -0.7949E-01  0.6593E-01 -0.1057E+00 -0.6612E-01
  ROW   3
  0.6541E+00  0.3224E+00  0.1000E+01 -0.9458E-01  0.3456E+00
  0.4514E+00  0.6788E+00 -0.4203E+00  0.5752E+00  0.6789E+00
  ROW   4
 -0.2760E+00  0.4878E+00 -0.9458E-01  0.1000E+01  0.6564E+00
  0.1018E+00 -0.4963E+00  0.5808E+00 -0.4945E+00 -0.5906E+00
  ROW   5
 -0.2020E-01  0.7738E+00  0.3456E+00  0.6564E+00  0.1000E+01
  0.2194E+00 -0.1818E+00  0.2400E+00 -0.1856E+00 -0.2039E+00
```

Eigenvalues

(1) NUMBER OF POSITIVE EIGENVALUES = 10

(2) SUM OF POSITIVE EIGENVALUES = 0.10000000E+02

(3) EIGENVALUES

0.5030E+01	0.2590E+01	0.1127E+01	0.3886E+00	0.3164E+00
0.1992E+00	0.1353E+00	0.1054E+00	0.6441E-01	0.4339E-01

(4) EIGENVALUES AS PERCENT

50.30	25.90	11.27	3.89	3.16
1.99	1.35	1.05	.64	.43

(5) CUMULATIVE PERCENTAGE OF EIGENVALUES

50.30	76.20	87.47	91.36	94.52
96.52	97.87	98.92	99.57	100.00

(6) SQUARE ROOTS OF EIGENVALUES

2.242671	1.609441	1.061811	.623400	.562464
.446362	.367773	.324655	.253790	.208292

Eigenvectors (direction cosines)

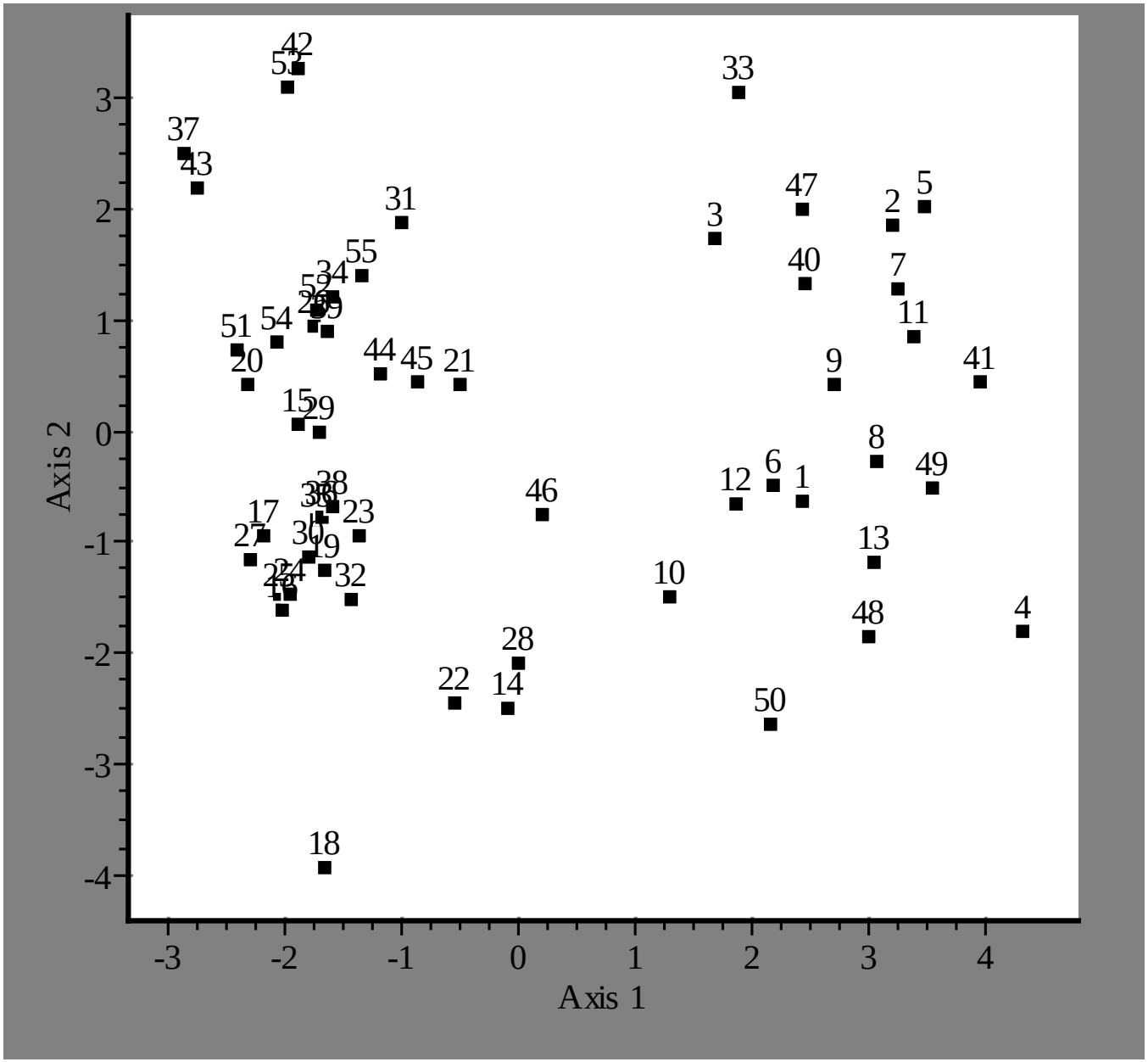
EIGENVECTORS (DIRECTION COSINES)

VECTOR 1				
.38449	-.05267	.31208	-.26483	-.09857
.16803	.42428	-.32014	.41978	.43035
VECTOR 2				
.15583	.50365	.33372	.40278	.54941
.34739	.02809	.15840	.00622	.00253
VECTOR 3				
.26530	-.38005	-.13325	.05345	-.24200
.62749	-.00397	.55644	.01315	-.04942

Component scores

COMPONENT SCORES

1	2.440	-.617	-.296
2	3.203	1.866	-.442
3	1.689	1.730	-.520
4	4.332	-1.803	1.150
5	3.485	2.018	.820
6	2.192	-.477	-.485
7	3.268	1.295	-.575
8	3.077	-.271	-.586



Ordinace objektů

Percentage of variance of variables accounted for by each component

PERCENTAGE OF VARIANCE OF VARIABLES ACCOUNTED FOR BY EACH COMPONENT

VARIABLE 1	*(šířka báze lodyhy)*		
74.352	6.290	7.935	
VARIABLE 2	*(délka nitek delších tyčinek)*		
1.395	65.706	16.284	
VARIABLE 3	*(délka kališních lístků)*		
48.986	28.849	2.002	
VARIABLE 4	*(šířka korunních lístků)*		
35.274	42.023	.322	
VARIABLE 5	*(délka korunních lístků)*		
4.887	78.187	6.603	
VARIABLE 6	*(počet květů v hlavním květenství)*		
14.201	31.260	44.392	
VARIABLE 7	*(počet lístků na lodyžních listech)*		
90.539	.204	.002	
VARIABLE 8	*(větvení lodyhy)*		
51.548	6.499	34.909	
VARIABLE 9	*(počet lodyžních listů)*		
88.628	.010	.019	
VARIABLE 10	*(nahloučení listů pod květenstvím)*		
93.148	.002	.275	

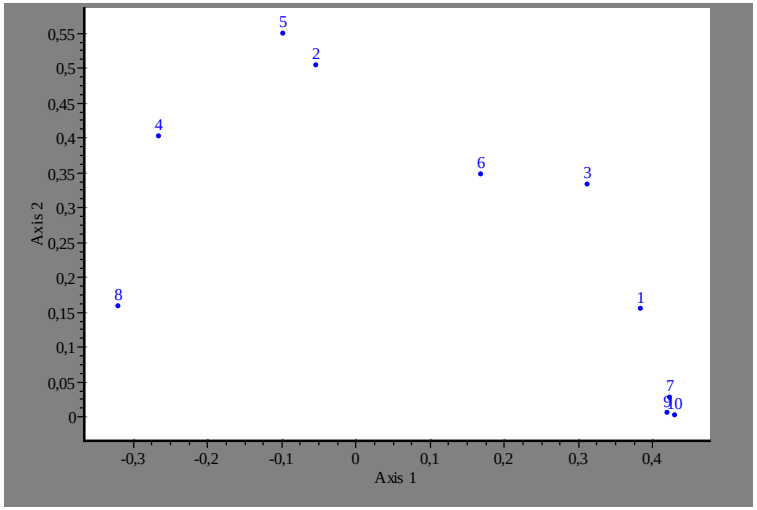
Ordination of characters/variables

„Euclidean option“ - The position of the variables expresses the positions of the corresponding variable vectors.

EIGENVECTORS AS COORDINATES OF VAR.

SCORES FOR VARIABLES

VARIABLE	1	.384	.156	.265
VARIABLE	2	-.053	.504	-.380
VARIABLE	3	.312	.334	-.133
VARIABLE	4	-.265	.403	.053
VARIABLE	5	-.099	.549	-.242
VARIABLE	6	.168	.347	.627
VARIABLE	7	.424	.028	-.004
VARIABLE	8	-.320	.158	.556
VARIABLE	9	.420	.006	.013
VARIABLE	10	.430	.003	-.049



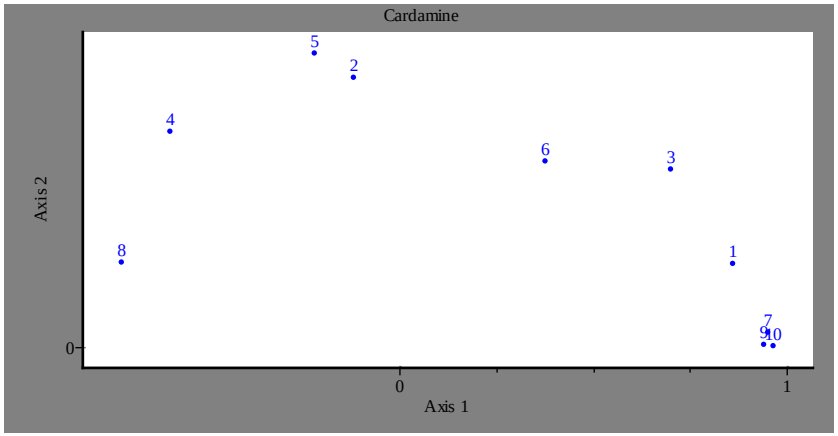
Ordination of characters/variables

„*Rohlf mixed option*“ - The position of the variables expresses the values of the correlation (or possibly covariance) of the variables with the corresponding components

CORRELATIONS OF VAR. WITH COMPONENTS

SCORES FOR VARIABLES

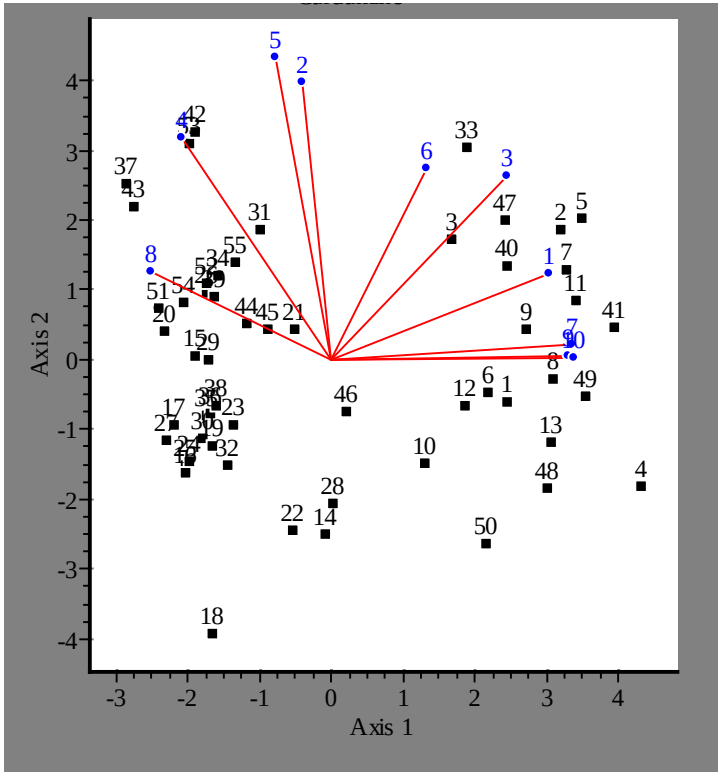
VARIABLE 1	.862	.251	.282
VARIABLE 2	-.118	.811	-.404
VARIABLE 3	.700	.537	-.141
VARIABLE 4	-.594	.648	.057
VARIABLE 5	-.221	.884	-.257
VARIABLE 6	.377	.559	.666
VARIABLE 7	.952	.045	-.004
VARIABLE 8	-.718	.255	.591
VARIABLE 9	.941	.010	.014
VARIABLE 10	.965	.004	-.052

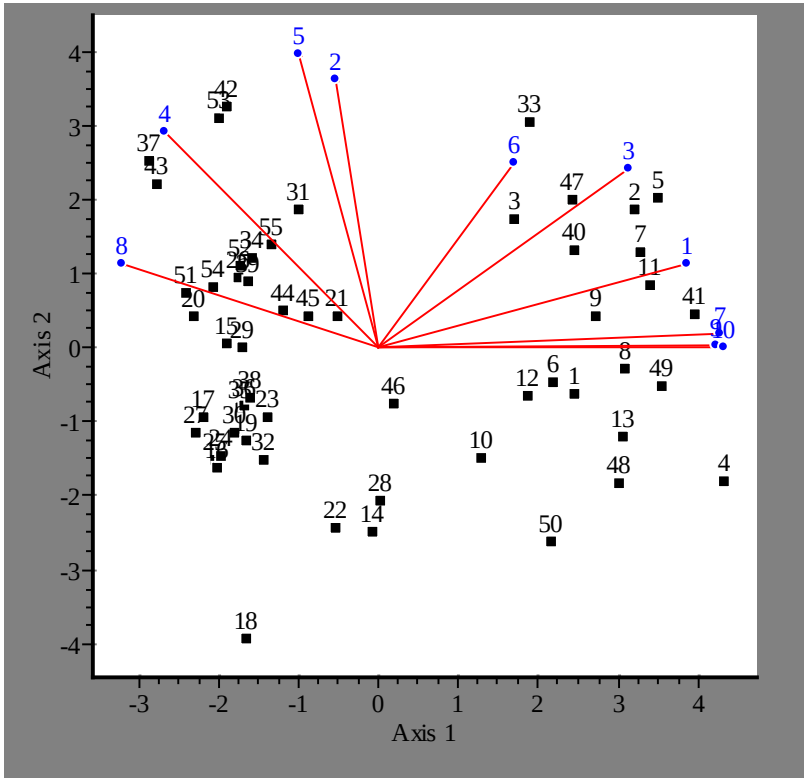


Ordination of objects and variables (biplot)

Euclidean biplot - The position of the variables expresses the positions of the corresponding variable vectors.

Rohlf mixed option - The position of the variables expresses the values of the correlation (or possibly covariance) of the variables with the corresponding components





Types of PCA

Centered PCA – based on the covariance matrix of variables
The starting point of the new coordinate system (principal components) is shifted from the original starting point of the coordinate system of the original variables to the centroid of the "cloud" of ordinated objects (a hypothetical point representing the "average object"). Distances between objects in the new coordinate system remain the same as in the original system.

Standardized PCA – based on the correlation matrix of variables
It includes data standardization (compare the relationship between covariance and correlation: by dividing covariance by the product of standard deviations, we obtain the correlation coefficient). The starting point of the new coordinate system shifts to the centroid of the object cloud, and the original variables are rescaled to have unit variance. The new variables (principal components) do not have unit variance, but their variance corresponds to the respective eigenvalues (the sum of the eigenvalues equals the number of variables).

Non-centered PCA – based on the matrix of scalar products of variables (cross-products between variables)
It does not involve standardization or centering. The starting point of the new coordinate system is in the same place as in the original system. This technique is used in some ecological applications.

PCoA – principal coordinate(s) analysis (metric multidimensional scaling, classical scaling)

Arrangement of the set of objects in the new space defined by the principal coordinates (new axes).

The mutual (Euclidean) distances between objects reflect the relationships between the original objects measured by any similarity or distance coefficient.

Binary variables

Multistate qualitative variables

Mixed data

(1) Primary data matrix → secondary distance matrix → → symmetric matrix, equivalent to the correlation or covariance matrix used in PCA

(2) Calculation of eigenvalues, eigenvectors, and component scores

The coordinates in the space determined by the principal coordinates are not linearly dependent on the values of the original variables.

It can also be appropriately used when the number of variables exceeds the number of objects (e.g., in molecular data).

(1) EIGENVALUES

271.59700	139.87590	60.88202	20.98608	17.08320
10.7592	7.30412	5.69138	3.47811	2.34287
.00060	.00001	.00001	.00001	.00001
.00001	.00001	.00001	.00001	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000
.00001	.00001	.00001	.00001	.00001
.00001	.00001	.00001	.00001	.00001

(2) SUM OF POSITIVE EIGENVALUES

540.00054

(3) NUMBER OF POSITIVE EIGENVALUES

33

(4) EIGENVALUES AS PERCENT

50.30	25.90	11.27	3.89	3.16
1.99	1.35	1.05	.64	.43
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00		

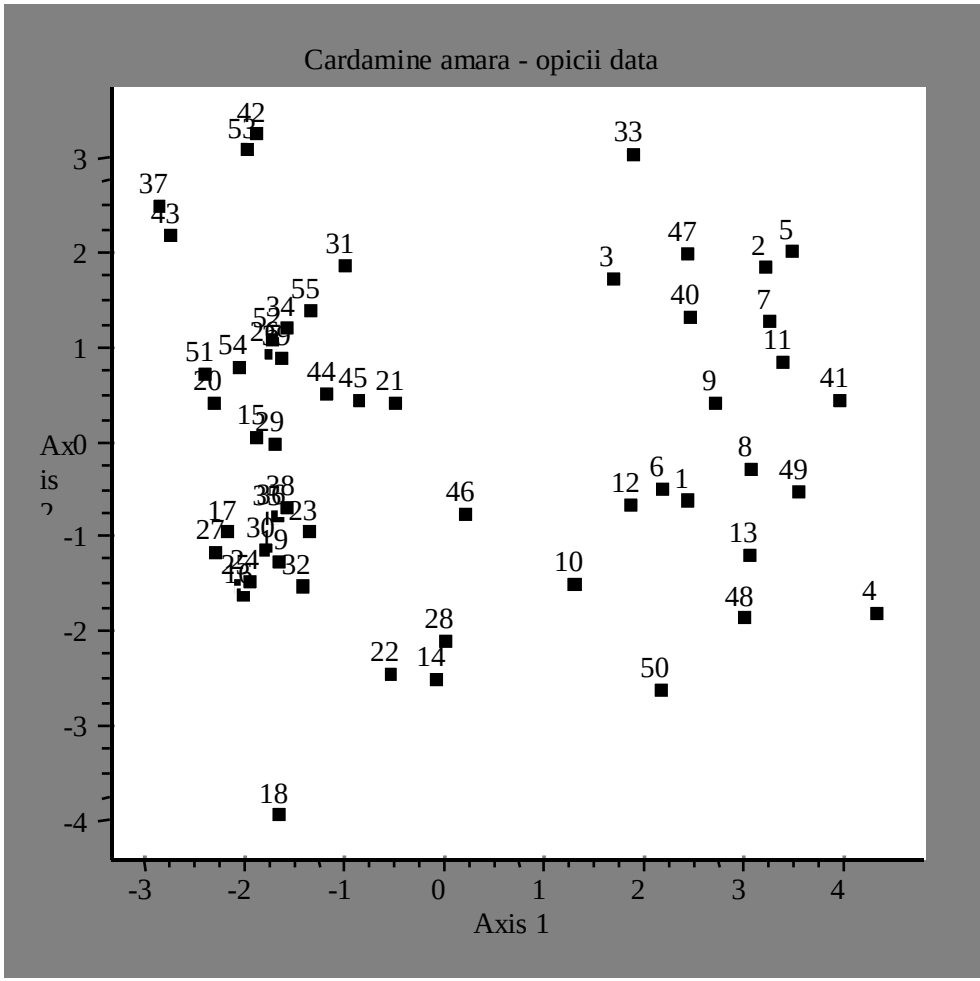
(5) CUMULATIVE PERCENTAGE OF EIGENVALUES

50.30	76.20	87.47	91.36	94.52
96.52	97.87	98.92	99.57	100.00
100.00	100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00	100.00
100.00	100.00	100.00		

COORDINATES

```
OBJECT  1
   2.44015      -.61730       .29618
OBJECT  2
   3.20319      1.86563       .44235
OBJECT  3
   1.68897      1.73017       .51992
```

(abbreviated)



NMDS – non-metric multidimensional scaling

Reducing the dimension of the original variable space

Maintaining the order of distances between objects

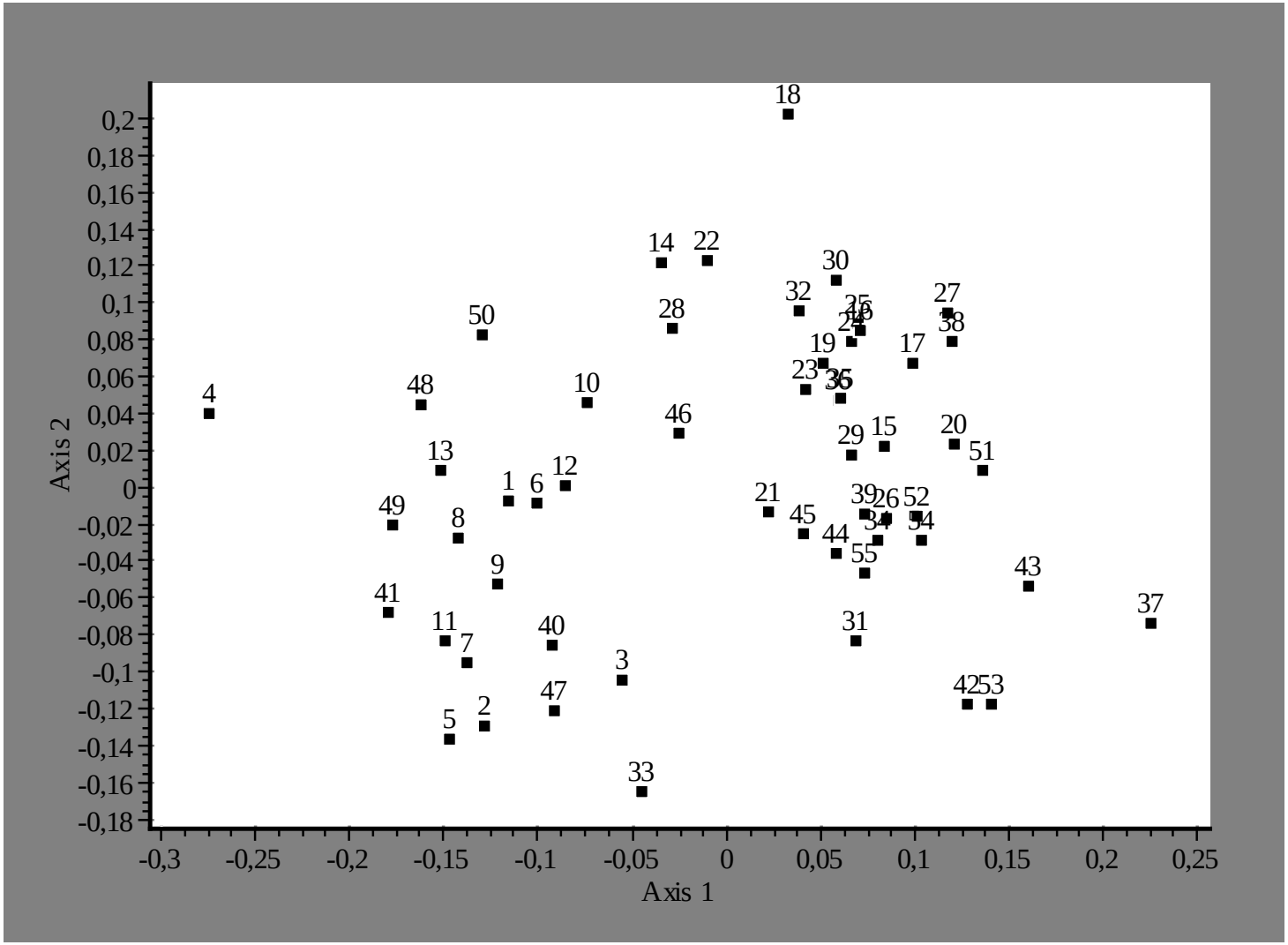
4 OTU and 6 dissimilarity values:

$$\delta_{23} < \delta_{12} < \delta_{34} < \delta_{13} < \delta_{24} < \delta_{14}$$

OTU = points in Euclidean space, their mutual distances are: d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34}

It is assumed that these distances perfectly match the observed dissimilarities if:

$$d_{23} \leq d_{12} \leq d_{34} \leq d_{13} \leq d_{24} \leq d_{14}$$


Axis 2
Axis 1
0,2
0,18
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
-0,02
-0,04
-0,06
-0,08
-0,1
-0,12
-0,14
-0,16
-0,18
-0,3
-0,25
-0,2
-0,15
-0,1
-0,05
0
0,05
0,1
0,15
0,2
0,25
18
14
22
30
32
27
28
50
38
19
17
4
48
10
46
29
15
20
13
51
12
1
6
21
39
26
52
49
8
45
34
44
9
55
43
41
37
11
40
31
7
3
47
42
53
2
5
33

Stress

A measure of the agreement of distances on the ordination diagram with the original dissimilarity values

- below 0.05 – excellent agreement
- 0.05–0.10 – satisfactory agreement
- 0.10–0.15 – acceptable agreement with reservations

```
DIMENSIONALITY USED BELOW =    5
   CHANGE VERY SMALL, FINAL STRESS =         .01672606

DIMENSIONALITY USED BELOW =    4
   CHANGE VERY SMALL, FINAL STRESS =         .02733298

DIMENSIONALITY USED BELOW =    3
   CHANGE VERY SMALL, FINAL STRESS =         .04378727

DIMENSIONALITY USED BELOW =    2
   CHANGE VERY SMALL, FINAL STRESS =         .10600522
```