BDC334

Biogeography & Global Ecology

Topic 4
The Multivariate Nature of Ecological Data

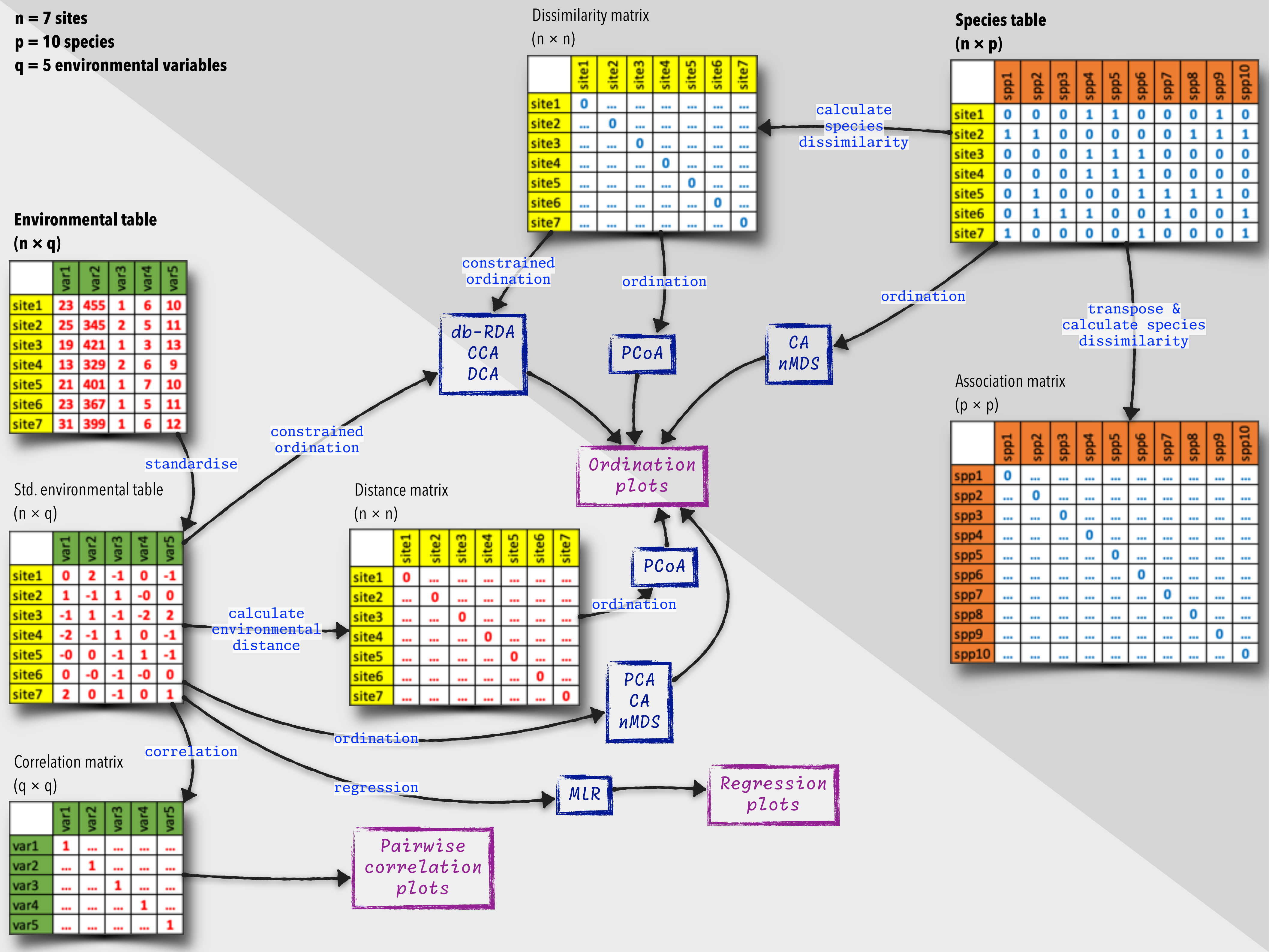
Diversity can consider either,

- whether species are present or absent; this kind of data is called **presence-absence** data
    - this kind of data is binary (i.e. a species is there, or it is not there), or
- it can include aspects of how much (biomass, abundance, % cover) of each of the species that is present
    - we will call this kind of data **abundance** data.

site	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
site_A	1	1	1	2	1	10
site_B	1	2	1	1	2	1
site_C	4	4	5	4	5	4
site_D	10	11	10	10	10	11
site_E	0	0	0	0	1	1
site_F	0	0	0	0	1	10
site_G	1	1	1	1	1	1
site_H	10	10	10	10	10	10

convert to presence-absence

site	sp_A	sp_B	sp_C	sp_D	sp_E	sp_F
site_A	1	1	1	1	1	1
site_B	1	1	1	1	1	1
site_C	1	1	1	1	1	1
site_D	1	1	1	1	1	1
site_E	0	0	0	0	1	1
site_F	0	0	0	0	1	1
site_G	1	1	1	1	1	1
site_H	1	1	1	1	1	1



Dissimilarity matrix (n × n)

	site1	site2	site3	site4	site5	site6	site7
site1	0	...	...	...	...	...	...
site2	...	0	...	...	...	...	...
site3	...	...	0	...	...	...	...
site4	...	...	...	0	...	...	...
site5	...	...	...	...	0	...	...
site6	...	...	...	...	...	0	...
site7	...	...	...	...	...	...	0

Species table (n × p)

	spp1	spp2	spp3	spp4	spp5	spp6	spp7	spp8	spp9	spp10
site1	0	0	0	1	1	0	0	0	1	0
site2	1	1	0	0	0	0	0	1	1	1
site3	0	0	0	1	1	1	0	0	0	0
site4	0	0	0	1	1	1	0	0	0	0
site5	0	1	0	0	0	1	1	1	1	0
site6	0	1	1	1	0	0	1	0	0	1
site7	1	0	0	0	0	1	0	0	0	1

Environmental table (n × q)

	var1	var2	var3	var4	var5
site1	23	455	1	6	10
site2	25	345	2	5	11
site3	19	421	1	3	13
site4	13	329	2	6	9
site5	21	401	1	7	10
site6	23	367	1	5	11
site7	31	399	1	6	12

Association matrix (p × p)

	spp1	spp2	spp3	spp4	spp5	spp6	spp7	spp8	spp9	spp10
spp1	0	...	...	...	...	...	...	...	...	...
spp2	...	0	...	...	...	...	...	...	...	...
spp3	...	...	0	...	...	...	...	...	...	...
spp4	...	...	...	0	...	...	...	...	...	...
spp5	...	...	...	...	0	...	...	...	...	...
spp6	...	...	...	...	...	0	...	...	...	...
spp7	...	...	...	...	...	...	0	...	...	...
spp8	...	...	...	...	...	...	...	0	...	...
spp9	...	...	...	...	...	...	...	...	0	...
spp10	...	...	...	...	...	...	...	...	...	0

Std. environmental table (n × q)

	var1	var2	var3	var4	var5
site1	0	2	-1	0	-1
site2	1	-1	1	-0	0
site3	-1	1	-1	-2	2
site4	-2	-1	1	0	-1
site5	-0	0	-1	1	-1
site6	0	-0	-1	-0	0
site7	2	0	-1	0	1

Distance matrix (n × n)

	site1	site2	site3	site4	site5	site6	site7
site1	0	...	...	...	...	...	...
site2	...	0	...	...	...	...	...
site3	...	...	0	...	...	...	...
site4	...	...	...	0	...	...	...
site5	...	...	...	...	0	...	...
site6	...	...	...	...	...	0	...
site7	...	...	...	...	...	...	0

Correlation matrix (q × q)

	var1	var2	var3	var4	var5
var1	1	...	...	...	...
var2	...	1	...	...	...
var3	...	...	1	...	...
var4	...	...	...	1	...
var5	...	...	...	...	1

Distance matrices

- how similar sites (plot or quadrats or transects) are to each other is shown by distance matrices
- they are calculated from data tables (species table or environment table) by applying dissimilarity or distance calculations of some indices:
  - e.g. Euclidian distances for environmental data
- the result is a matrix of pairwise differences (or distances) or similarities in a metric that relates to the ecological distance between all sites, or the community composition (as synthesised by the chosen index)

Similarity and dissimilarity

- sites sharing a similar species composition are ecologically similar
  - ‘composition’ a function of species richness and abundance
    - *i.e.* high similarity / low dissimilarity
- how similar sites are depends on…
  - measurable environmental differences that influence species composition, or
  - it can be due to unmeasured influences, or
  - it can also simply be ‘noise’
- it is the ecologist’s role to figure out what influences the similarity / dissimilarity among sites
- they are grouped with a special class of matrix, *i.e.* the distance matrix

Distance matrix for environmental data

- Euclidian distance is "*the 'ordinary' straight-line distance between two points in Euclidean space*" (*i.e.* in its simplest form a planar area such as a graph with *x*- and *y*-axes)
- in 2D and 3D, gives cartesian distance between points on a plane (*x*, *y*), in a volume (*x*, *y*, *z*), or higher dimensions
- conforms to our physical concept of distance
  - *e.g.* short geographic distances between points on a map
  - (loses accuracy over large distances, as Earth's surface is not on a plane but on a sphere… correct by using great circle distances, *e.g.* use the Haversine formula)
- calculated using the Pythagorean theorem
  - differences are squared, so single large differences become very important
  - this is not useful for species data

Distance matrices: properties

- the matrices are square and symmetrical
- as many rows and columns as the number of sites (*i.e.* rows) in the original species or environment table
- the diagonals are zero (a site is the same as itself, so it has zero dissimilarity), or one if it is a similarity matrix
- the table can be read directly, and each cell represents the species or ecological difference between a pair of sites
- all information of the species ID (and maybe also abundance) at a site is lost, as this info is condensed into one metric, the dissimilarity metric (or similarity metric)

Distance matrices: Euclidian distance with the Pythagorean Theorem for environmental data

Two dimensions [edit]

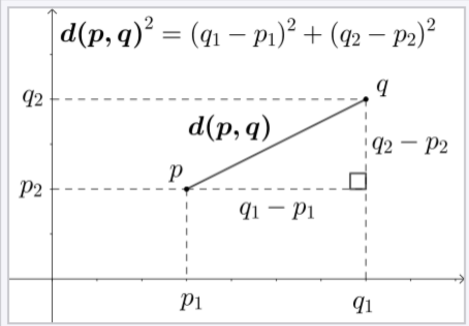
In the Euclidean plane, if $\mathbf{p} = (p_1, p_2)$ and $\mathbf{q} = (q_1, q_2)$ then the distance is given by

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}.$$

This is equivalent to the Pythagorean theorem.

Alternatively, it follows from (*2*) that if the polar coordinates of the point **p** are (r_1, θ_1) and those of **q** are (r_2, θ_2), then the distance between the points is

$$\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_1 - \theta_2)}.$$



Euclidean distance in R^2

Three dimensions [edit]

In three-dimensional Euclidean space, the distance is

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

n dimensions [edit]

In general, for an *n*-dimensional space, the distance is

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}.$$

https://en.wikipedia.org/wiki/Euclidean_distance

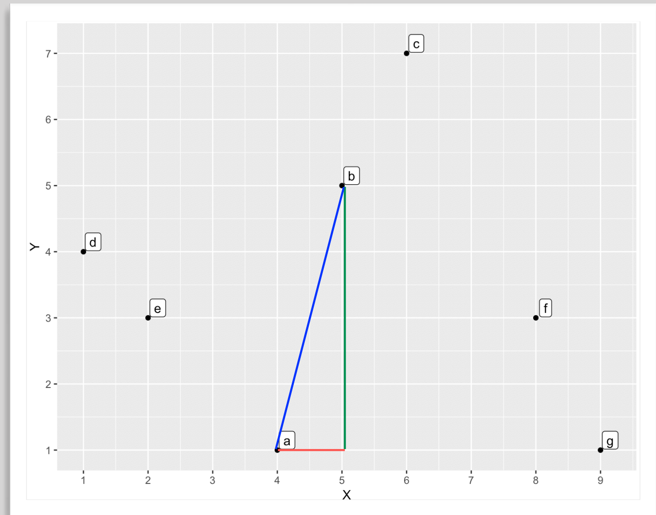
e.g. example with position (such as geographic) coordinates…

Raw data

site	x	y
a	4	1
b	5	5
c	6	6
d	1	4
e	2	3
f	8	3
g	9	1

Euclidian distances

	a	b	c	d	e	f
b	4.123106					
c	6.324555	2.236068				
d	4.242641	4.123106	5.830952			
e	2.828427	3.605551	5.656854	1.414214		
f	4.472136	3.605551	4.472136	7.071068	6.000000	
g	5.000000	5.656854	6.708204	8.544004	7.280110	2.236068



$$d(a, b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}$$

e.g. example with 3D position coordinates (a.k.a. dimensions)…

Raw data

site	x	y	z
a	4	1	3
b	5	5	5
c	6	6	4
d	1	4	9
e	2	3	8
f	8	3	1
g	9	1	5

Euclidian distances

	1	2	3	4	5	6
2	4.582576					
3	5.477226	1.732051				
4	7.348469	5.744563	7.348469			
5	5.744563	4.690416	6.403124	1.732051		
6	4.898979	5.385165	4.690416	10.677078	9.219544	
7	5.385165	5.656854	5.916080	9.433981	7.874008	4.582576

$$d(a, b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2}$$

e.g. example with environmental 'dimensions'…

a dimensionless number

Raw data

site	temperature	depth	light
a	4	1	3
b	5	5	5
c	6	6	4
d	1	4	9
e	2	3	8
f	8	3	1
g	9	1	5

Euclidian distances

```
R> ex.xyz.euc <- vegdist(ex.xyz[,2:4], method = "euclidian")
R> ex.xyz.euc
          1         2         3         4         5         6
2  4.582576
3  5.477226  1.732051
4  7.348469  5.744563  7.348469
5  5.744563  4.690416  6.403124  1.732051
6  4.898979  5.385165  4.690416 10.677078  9.219544
7  5.385165  5.656854  5.916080  9.433981  7.874008  4.582576
```

$$d(a, b) = \sqrt{(a_{temp} - b_{temp})^2 + (a_{depth} - b_{depth})^2 + (a_{light} - b_{light})^2}$$

e.g. example with higher dimension environmental data…

Raw data

	pH	O2	temp	depth
a	7.1	6.5	12.1	1.1
b	7.5	5.5	12.3	1.3
c	7.6	5.7	11.9	1.5
d	7.0	5.4	11.8	1.6
e	7.1	6.3	12.0	1.8
f	7.2	6.3	12.1	1.9
g	6.9	6.1	12.2	2.2

(transformation)

Standarised data

	pH	O2	temp	depth
a	-0.3872983	1.2156767	0.2494233	-1.41749621
b	1.1618950	-1.0842522	1.4133987	-0.88114629
c	1.5491933	-0.6242664	-0.9145521	-0.34479637
d	-0.7745967	-1.3142450	-1.4965398	-0.07662142
e	-0.3872983	0.7556909	-0.3325644	0.45972850
f	0.0000000	0.7556909	0.2494233	0.72790346
g	-1.1618950	0.2957051	0.8314110	1.53242833

Euclidian distances

	a	b	c	d	e	f
b	4.123106					
c	6.324555	2.236068				
d	4.242641	4.123106	5.830952			
e	2.828427	3.605551	5.656854	1.414214		
f	4.472136	3.605551	4.472136	7.071068	6.000000	
g	5.000000	5.656854	6.708204	8.544004	7.280110	2.236068

Distance matrix for species data

- instead of using the Pythagorean Theorem to calculate ‘distances’ between species, we use
  - **Bray-Curtis** or **Jaccard** index for the case where data are abundances
  - Jaccard for presence-absence data — this is called the **Sørensen** dissimilarity index
- instead of having columns with measurements of environmental variables we have species abundance or presence-absence data
- Dissimilarity = 1 - Similarity
  - dissimilarity: the indices go from 0 (sites are identical) to 1 (sites are completely dissimilar)
  - similarity: the indices go from 0 (sites are completely dissimilar) to 1 (sites are identical)
- Qualitative indices (e.g. applied to presence-absence data) give more weight to rare species because the weights assigned to rare and common species are the same (1 in both instances).
- Quantitative indices give more weight to common species, which have more numerical variation between plots and these ‘weights’ feature more strongly in the calculation of indices.

Distance matrices:
example with real environmental data (Doubs River data)

	dfs	alt	slo	flo	pH	har	pho	nit	amm	oxy	bod
	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.3	934	48	0.84	7.9	45	0.01	0.2	0	12.2	2.7
2	2.2	932	3	1	8	40	0.02	0.2	0.1	10.3	1.9
3	10.2	914	3.7	1.8	8.3	52	0.05	0.22	0.05	10.5	3.5
4	18.5	854	3.2	2.53	8	72	0.1	0.21	0	11	1.3
5	21.5	849	2.3	2.64	8.1	84	0.38	0.52	0.2	8	6.2
6	32.4	846	3.2	2.86	7.9	60	0.2	0.15	0	10.2	5.3
7	36.8	841	6.6	4	8.1	88	0.07	0.15	0	11.1	2.2
8	70.5	752	1.2	4.8	8	90	0.3	0.82	0.12	7.2	5.2
9	99	617	9.9	10	7.7	82	0.06	0.75	0.01	10	4.3
10	123.	483	4.1	19.9	8.1	96	0.3	1.6	0	11.5	2.7

… with 19 more rows

	Cogo	Satr	Phph	Babl	Thth	Teso	Chna	Pato	Lele	Sqce	Baba	Albi	Gogo	Eslu	Pefl	Rham	Legi	Scer	Cyca
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	0	3	0	0	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	0	0
3	0	5	5	5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4	0	4	5	5	0	0	0	0	0	1	0	0	1	2	2	0	0	0	0
5	0	2	3	2	0	0	0	0	5	2	0	0	2	4	4	0	0	2	0
6	0	3	4	5	0	0	0	0	1	2	0	0	1	1	1	0	0	0	0
7	0	5	4	5	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
8	0	0	1	3	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
9	0	1	4	4	0	0	0	0	2	2	0	0	1	0	0	0	0	0	0
10	1	3	4	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0

… with 19 more rows, and 8 more variables: Titi <int>, Abbr <int>, Icme <int>, Gyce <int>, Ruru <int>,
Blbj <int>, Alal <int>, Anan <int>

Environmental table

	dfs	alt	slo	flo	pH	har	pho	nit	amm	oxy	bod
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
7	36.8	841	6.6	4.00	8.1	88	0.07	0.15	0.00	11.1	2.2
8	70.5	752	1.2	4.80	8.0	90	0.30	0.82	0.12	7.2	5.2
9	99.0	617	9.9	10.00	7.7	82	0.06	0.75	0.01	10.0	4.3
10	123.4	483	4.1	19.90	8.1	96	0.30	1.60	0.00	11.5	2.7
11	132.4	477	1.6	20.00	7.9	86	0.04	0.50	0.00	12.2	3.0
12	143.6	450	2.1	21.10	8.1	98	0.06	0.52	0.00	12.4	2.4
13	152.2	434	1.2	21.20	8.3	98	0.27	1.23	0.00	12.3	3.8
14	164.5	415	0.5	23.00	8.6	86	0.40	1.00	0.00	11.7	2.1
15	185.9	375	2.0	16.10	8.0	88	0.20	2.00	0.05	10.3	2.7
16	198.5	349	0.5	24.30	8.0	92	0.20	2.50	0.20	10.2	4.6
17	211.0	333	0.8	25.00	8.0	90	0.50	2.20	0.20	10.3	2.8
18	224.6	310	0.5	25.90	8.1	84	0.60	2.20	0.15	10.6	3.3
19	247.7	286	0.8	26.80	8.0	86	0.30	3.00	0.30	10.3	2.8
20	282.1	262	1.0	27.20	7.9	85	0.20	2.20	0.10	9.0	4.1
21	294.0	254	1.4	27.90	8.1	88	0.20	1.62	0.07	9.1	4.8
22	304.3	246	1.2	28.80	8.1	97	2.60	3.50	1.15	6.3	16.4
23	314.7	241	0.3	29.76	8.0	99	1.40	2.50	0.60	5.2	12.3
24	327.8	231	0.5	38.70	7.9	100	4.22	6.20	1.80	4.1	16.7
25	356.9	214	0.5	39.10	7.9	94	1.43	3.00	0.30	6.2	8.9
26	373.2	206	1.2	39.60	8.1	90	0.58	3.00	0.26	7.2	6.3
27	394.7	195	0.3	43.20	8.3	100	0.74	4.00	0.30	8.1	4.5
28	422.0	183	0.6	67.70	7.8	110	0.45	1.62	0.10	9.0	4.2
29	453.0	172	0.2	69.00	8.2	109	0.65	1.60	0.10	8.2	4.4

Species table

	Cogo	Satr	Phph	Babl	Thth	Teso	Chna	Pato	Lele	Sqce	Baba	Albi	Gogo	Eslu	Pefl	Rl											
1	0	3	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											
3	0	5	5	5	0	0	0	0	0	0	0	0	0	1	0	0											
4	0	4	5	5	0	0	0	0	0	1	0	0	1	2	2	0											
5	0	2	3	2	0	0	0	0	5	2	0	0	2	4	4	0											
6	0	3	4	5	0	0	0	0	1	2	0	0	1	1	1	0											
7	0	5	4	5	0	0	0	0	1	1	0	0	0	0	0	0											
8	0	0	1	3	0	0	0	0	0	5	0	0	0	0	0	0											
9	0	1	4	4	0	0	0	0	2	2	0	0	1	0	0	0											
10	1	3	4	1	1	0	0	0	0	1	0	0	0	0	0	0											
11	2	5	4	4	2	0	0	0	0	1	0	0	0	0	0	0											
12	2	5	5	2	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	3	5	5	4	4	3	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
14	3	4	4	5	2	4	0	0	3	3	2	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0
15	2	3	3	5	0	5	0	4	5	2	2	1	2	1	1	0	1	0	1	1	0	0	0	1	0	0	0
16	1	2	4	4	1	2	1	4	3	2	3	4	1	1	2	1	1	0	1	1	0	0	0	2	0	2	1
17	1	1	3	3	1	1	1	3	2	3	3	3	2	1	3	2	1	0	1	1	0	0	1	2	0	2	1
18	0	0	3	5	0	1	2	3	2	1	2	2	4	1	1	2	1	1	1	2	1	0	1	5	1	3	1
19	0	0	1	2	0	0	2	2	2	3	4	3	4	2	2	3	2	2	1	4	1	0	2	5	2	5	2
20	0	0	1	1	0	0	2	2	2	2	4	2	5	3	3	3	2	2	2	4	3	1	3	5	3	5	2
21	0	0	0	1	0	0	3	2	3	4	5	1	5	3	4	3	3	2	3	4	4	2	4	5	4	5	2
22	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0
23	0	0	0	0	0	0	1	0	0	2	0	0	1	0	0	0	1	0	0	0	0	0	2	2	1	5	0
24	0	0	0	0	0	0	0	0	1	1	0	0	2	1	0	0	0	1	0	0	0	0	1	1	0	3	0
25	0	0	0	1	0	0	1	0	1	2	2	1	3	2	1	2	2	1	1	3	2	1	4	4	2	5	2
26	0	0	0	1	0	0	1	1	2	3	4	1	4	4	1	3	3	1	2	5	3	2	5	5	4	5	3
27	0	0	0	1	0	0	1	1	2	4	3	1	4	3	2	4	4	2	4	4	3	3	5	5	5	5	4
28	0	1	1	1	1	1	2	2	3	4	5	3	5	5	4	5	5	2	3	3	4	4	5	5	4	5	4
29	0	0	0	0	0	0	1	2	3	3	3	5	5	4	5	5	3	5	5	5	5	5	5	5	5	5	5

Raw data

	dfs	alt	slo	flo	pH	har	pho	nit	amm	oxy	bod
	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.3	934	48	0.84	7.9	45	0.01	0.2	0	12.2	2.7
2	2.2	932	3	1	8	40	0.02	0.2	0.1	10.3	1.9
3	10.2	914	3.7	1.8	8.3	52	0.05	0.22	0.05	10.5	3.5
4	18.5	854	3.2	2.53	8	72	0.1	0.21	0	11	1.3
5	21.5	849	2.3	2.64	8.1	84	0.38	0.52	0.2	8	6.2
6	32.4	846	3.2	2.86	7.9	60	0.2	0.15	0	10.2	5.3
7	36.8	841	6.6	4	8.1	88	0.07	0.15	0	11.1	2.2
8	70.5	752	1.2	4.8	8	90	0.3	0.82	0.12	7.2	5.2
9	99	617	9.9	10	7.7	82	0.06	0.75	0.01	10	4.3
10	123.	483	4.1	19.9	8.1	96	0.3	1.6	0	11.5	2.7

\# … with 19 more rows

(transformation)

Standarised data

	dfs	alt	slo	flo	pH	har	pho	nit	amm	oxy	bod
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	-1.38	1.72	5.05	-1.23	-0.84	-2.39	-0.63	-1.06	-0.55	1.24	-0.59
2	-1.37	1.71	-0.06	-1.22	-0.27	-2.68	-0.62	-1.06	-0.290	0.37	-0.8
3	-1.31	1.64	0.02	-1.17	1.43	-1.98	-0.580	-1.04	-0.42	0.47	-0.39
4	-1.25	1.42	-0.04	-1.13	-0.27	-0.81	-0.53	-1.05	-0.55	0.69	-0.95
5	-1.23	1.4	-0.14	-1.13	0.290	-0.11	-0.21	-0.83	-0.03	-0.67	0.3
6	-1.15	1.39	-0.04	-1.12	-0.84	-1.51	-0.42	-1.09	-0.55	0.33	0.07
7	-1.12	1.37	0.35	-1.05	0.290	0.13	-0.56	-1.09	-0.55	0.74	-0.72
8	-0.88	1.04	-0.26	-1.01	-0.27	0.24	-0.3	-0.62	-0.24	-1.03	0.05
9	-0.68	0.54	0.72	-0.72	-1.97	-0.22	-0.570	-0.67	-0.53	0.24	-0.18
10	-0.5	0.05	0.06	-0.17	0.290	0.59	-0.3	-0.07	-0.55	0.92	-0.59

\# … with 19 more rows

Euclidian distances

	\`1\`	\`2\`	\`3\`	\`4\`	\`5\`	\`6\`	\`7\`	\`8\`	\`9\`	\`10\`	\`11\`	\`12\`	\`13\`	\`14\`	\`15\`	\`16\`
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0	0.01	0.08	0.3	0.32	0.33	0.35	0.68	1.18	1.67	1.7	1.8	1.86	1.93	2.08	2.17
2	0.01	0	0.07	0.290	0.31	0.32	0.34	0.67	1.17	1.66	1.69	1.79	1.85	1.92	2.07	2.16
3	0.08	0.07	0	0.22	0.24	0.25	0.27	0.6	1.1	1.59	1.62	1.72	1.78	1.85	2	2.09
4	0.3	0.290	0.22	0	0.02	0.03	0.05	0.38	0.88	1.37	1.4	1.5	1.56	1.63	1.78	1.87
5	0.32	0.31	0.24	0.02	0	0.01	0.03	0.36	0.86	1.35	1.38	1.48	1.54	1.61	1.76	1.85
6	0.33	0.32	0.25	0.03	0.01	0	0.02	0.35	0.85	1.34	1.37	1.47	1.53	1.6	1.75	1.84
7	0.35	0.34	0.27	0.05	0.03	0.02	0	0.33	0.83	1.32	1.35	1.45	1.51	1.58	1.73	1.82
8	0.68	0.67	0.6	0.38	0.36	0.35	0.33	0	0.5	0.99	1.02	1.12	1.18	1.25	1.4	1.49
9	1.18	1.17	1.1	0.88	0.86	0.85	0.83	0.5	0	0.49	0.52	0.62	0.68	0.75	0.9	0.99
10	1.67	1.66	1.59	1.37	1.35	1.34	1.32	0.99	0.49	0	0.03	0.13	0.19	0.26	0.41	0.5

\# … with 19 more rows, and 13 more variables: \`17\` <dbl>, \`18\` <dbl>, \`19\` <dbl>, \`20\` <dbl>,
\#   \`21\` <dbl>, \`22\` <dbl>, \`23\` <dbl>, \`24\` <dbl>, \`25\` <dbl>, \`26\` <dbl>, \`27\` <dbl>, \`28\` <dbl>,
\#   \`29\` <dbl>

The full matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	2	20	80	85	88	93	142	182	317	451	457	484	500	519	559	585	601	624	648	672	680	688	693	703	720	728	739	751	762
2	2	0	18	78	83	86	91	140	180	315	449	455	482	498	517	557	583	599	622	646	670	678	686	691	701	718	726	737	749	760
3	20	18	0	60	65	68	73	122	162	297	431	437	464	480	499	539	565	581	604	628	652	660	668	673	683	700	708	719	731	742
4	80	78	60	0	5	8	13	62	102	237	371	377	404	420	439	479	505	521	544	568	592	600	608	613	623	640	648	659	671	682
5	85	83	65	5	0	3	8	57	97	232	366	372	399	415	434	474	500	516	539	563	587	595	603	608	618	635	643	654	666	677
6	88	86	68	8	3	0	5	54	94	229	363	369	396	412	431	471	497	513	536	560	584	592	600	605	615	632	640	651	663	674
7	93	91	73	13	8	5	0	49	89	224	358	364	391	407	426	466	492	508	531	555	579	587	595	600	610	627	635	646	658	669
8	142	140	122	62	57	54	49	0	40	175	309	315	342	358	377	417	443	459	482	506	530	538	546	551	561	578	586	597	609	620
9	182	180	162	102	97	94	89	40	0	135	269	275	302	318	337	377	403	419	442	466	490	498	506	511	521	538	546	557	569	580
10	317	315	297	237	232	229	224	175	135	0	134	140	167	183	202	242	268	284	307	331	355	363	371	376	386	403	411	422	434	445
11	451	449	431	371	366	363	358	309	269	134	0	6	33	49	68	108	134	150	173	197	221	229	237	242	252	269	277	288	300	311
12	457	455	437	377	372	369	364	315	275	140	6	0	27	43	62	102	128	144	167	191	215	223	231	236	246	263	271	282	294	305
13	484	482	464	404	399	396	391	342	302	167	33	27	0	16	35	75	101	117	140	164	188	196	204	209	219	236	244	255	267	278
14	500	498	480	420	415	412	407	358	318	183	49	43	16	0	19	59	85	101	124	148	172	180	188	193	203	220	228	239	251	262
15	519	517	499	439	434	431	426	377	337	202	68	62	35	19	0	40	66	82	105	129	153	161	169	174	184	201	209	220	232	243
16	559	557	539	479	474	471	466	417	377	242	108	102	75	59	40	0	26	42	65	89	113	121	129	134	144	161	169	180	192	203
17	585	583	565	505	500	497	492	443	403	268	134	128	101	85	66	26	0	16	39	63	87	95	103	108	118	135	143	154	166	177
18	601	599	581	521	516	513	508	459	419	284	150	144	117	101	82	42	16	0	23	47	71	79	87	92	102	119	127	138	150	161
19	624	622	604	544	539	536	531	482	442	307	173	167	140	124	105	65	39	23	0	24	48	56	64	69	79	96	104	115	127	138
20	648	646	628	568	563	560	555	506	466	331	197	191	164	148	129	89	63	47	24	0	24	32	40	45	55	72	80	91	103	114
21	672	670	652	592	587	584	579	530	490	355	221	215	188	172	153	113	87	71	48	24	0	8	16	21	31	48	56	67	79	90
22	680	678	660	600	595	592	587	538	498	363	229	223	196	180	161	121	95	79	56	32	8	0	8	13	23	40	48	59	71	82
23	688	686	668	608	603	600	595	546	506	371	237	231	204	188	169	129	103	87	64	40	16	8	0	5	15	32	40	51	63	74
24	693	691	673	613	608	605	600	551	511	376	242	236	209	193	174	134	108	92	69	45	21	13	5	0	10	27	35	46	58	69
25	703	701	683	623	618	615	610	561	521	386	252	246	219	203	184	144	118	102	79	55	31	23	15	10	0	17	25	36	48	59
26	720	718	700	640	635	632	627	578	538	403	269	263	236	220	201	161	135	119	96	72	48	40	32	27	17	0	8	19	31	42
27	728	726	708	648	643	640	635	586	546	411	277	271	244	228	209	169	143	127	104	80	56	48	40	35	25	8	0	11	23	34
28	739	737	719	659	654	651	646	597	557	422	288	282	255	239	220	180	154	138	115	91	67	59	51	46	36	19	11	0	12	23
29	751	749	731	671	666	663	658	609	569	434	300	294	267	251	232	192	166	150	127	103	79	71	63	58	48	31	23	12	0	11
30	762	760	742	682	677	674	669	620	580	445	311	305	278	262	243	203	177	161	138	114	90	82	74	69	59	42	34	23	11	0

Distance matrices: example with real species data (Doubs River data)

- use Bray-Curtis for the case where data are abundances
- use Jaccard (with `binary = TRUE`) for presence/absence data
- many more in **vegan**; see `?vegdist`

Raw data

```
> spe
# A tibble: 29 x 27
    Cogo  Satr  Phph  Babl  Thth  Teso  Chna  Pato  Lele  Sqce  Baba  Albi
   <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
 1     0     3     0     0     0     0     0     0     0     0     0     0
 2     0     5     4     3     0     0     0     0     0     0     0     0
 3     0     5     5     5     0     0     0     0     0     0     0     0
 4     0     4     5     5     0     0     0     0     0     1     0     0
 5     0     2     3     2     0     0     0     0     5     2     0     0
 6     0     3     4     5     0     0     0     0     1     2     0     0
 7     0     5     4     5     0     0     0     0     1     1     0     0
 8     0     0     1     3     0     0     0     0     0     5     0     0
 9     0     1     4     4     0     0     0     0     2     2     0     0
10     1     3     4     1     1     0     0     0     0     1     0     0
# … with 19 more rows, and 15 more variables: Gogo <int>, Eslu <int>,
#   Pefl <int>, Rham <int>, Legi <int>, Scer <int>, Cyca <int>,
#   Titi <int>, Abbr <int>, Icme <int>, Gyce <int>, Ruru <int>,
#   Blbj <int>, Alal <int>, Anan <int>
```

Bray Curtis dissimilarities

```
> spe_dist <- round(vegdist(spe, method = "bray", diag = TRUE, upper = TRUE), 2)
> as.tibble(as.matrix(env_dist))
# A tibble: 29 x 29
     `1`   `2`   `3`   `4`   `5`   `6`   `7`   `8`   `9`  `10`  `11`  `12`
   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 0     0.01   0.08 0.3    0.32  0.33  0.35  0.68  1.18  1.67  1.7   1.8
 2 0.01  0      0.07 0.290  0.31  0.32  0.34  0.67  1.17  1.66  1.69  1.79
 3 0.08  0.07   0    0.22   0.24  0.25  0.27  0.6   1.1   1.59  1.62  1.72
 4 0.3   0.290  0.22 0      0.02  0.03  0.05  0.38  0.88  1.37  1.4   1.5
 5 0.32  0.31   0.24 0.02   0     0.01  0.03  0.36  0.86  1.35  1.38  1.48
 6 0.33  0.32   0.25 0.03   0.01  0     0.02  0.35  0.85  1.34  1.37  1.47
 7 0.35  0.34   0.27 0.05   0.03  0.02  0     0.33  0.83  1.32  1.35  1.45
 8 0.68  0.67   0.6  0.38   0.36  0.35  0.33  0     0.5   0.99  1.02  1.12
 9 1.18  1.17   1.1  0.88   0.86  0.85  0.83  0.5   0     0.49  0.52  0.62
10 1.67  1.66   1.59 1.37   1.35  1.34  1.32  0.99  0.49  0     0.03  0.13
# … with 19 more rows, and 17 more variables: `13` <dbl>, `14` <dbl>,
#   `15` <dbl>, `16` <dbl>, `17` <dbl>, `18` <dbl>, `19` <dbl>,
#   `20` <dbl>, `21` <dbl>, `22` <dbl>, `23` <dbl>, `24` <dbl>,
#   `25` <dbl>, `26` <dbl>, `27` <dbl>, `28` <dbl>, `29` <dbl>
```

Association matrices: example with species presence-absence

Associations: species presence-absence

Species table

```
> spe[1:10, 1:10]
# A tibble: 10 x 10
    Cogo  Satr  Phph  Babl  Thth  Teso  Chna  Pato  Lele  Sqce
   <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
 1     0     3     0     0     0     0     0     0     0     0
 2     0     5     4     3     0     0     0     0     0     0
 3     0     5     5     5     0     0     0     0     0     0
 4     0     4     5     5     0     0     0     0     0     1
 5     0     2     3     2     0     0     0     0     5     2
 6     0     3     4     5     0     0     0     0     1     2
 7     0     5     4     5     0     0     0     0     1     1
 8     0     0     1     3     0     0     0     0     0     5
 9     0     1     4     4     0     0     0     0     2     2
10     1     3     4     1     1     0     0     0     0     1
```

Transposed

```
> spe.t <- t(spe)
> spe.t[1:10, 1:10]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
Cogo    0    0    0    0    0    0    0    0    0     1
Satr    3    5    5    4    2    3    5    0    1     3
Phph    0    4    5    5    3    4    4    1    4     4
Babl    0    3    5    5    2    5    5    3    4     1
Thth    0    0    0    0    0    0    0    0    0     1
Teso    0    0    0    0    0    0    0    0    0     0
Chna    0    0    0    0    0    0    0    0    0     0
Pato    0    0    0    0    0    0    0    0    0     0
Lele    0    0    0    0    5    1    1    0    2     0
Sqce    0    0    0    1    2    2    1    5    2     1
```

Jaccard coefficient

```
> spe.t.S7 <- vegdist(spe.t, "jaccard", binary = TRUE)
> round(as.matrix(spe.t.S7)[1:10, 1:10], 2)
     Cogo Satr Phph Babl Thth Teso Chna Pato Lele Sqce
Cogo 0.00 0.53 0.60 0.67 0.22 0.40 0.89 0.81 0.82 0.73
Satr 0.53 0.00 0.24 0.36 0.53 0.61 0.88 0.83 0.65 0.55
Phph 0.60 0.24 0.00 0.17 0.60 0.60 0.77 0.71 0.54 0.39
Babl 0.67 0.36 0.17 0.00 0.67 0.67 0.62 0.60 0.38 0.25
Thth 0.22 0.53 0.60 0.67 0.00 0.40 0.82 0.81 0.82 0.73
Teso 0.40 0.61 0.60 0.67 0.40 0.00 0.75 0.64 0.70 0.73
Chna 0.89 0.88 0.77 0.62 0.82 0.75 0.00 0.23 0.42 0.52
Pato 0.81 0.83 0.71 0.60 0.81 0.64 0.23 0.00 0.39 0.56
Lele 0.82 0.65 0.54 0.38 0.82 0.70 0.42 0.39 0.00 0.28
Sqce 0.73 0.55 0.39 0.25 0.73 0.73 0.52 0.56 0.28 0.00
```

Interpretation

0—always associated with…

1—never associated with…