TOPIC 6

CORRELATIONS AND ASSOCIATIONS

Correlations and associations

- **correlations between environmental variables**

- e.g. **associations between species** (info on co-varying species) based on dissimilarities,
- associations are usually found on a **transposed** species table

Correlations: environment tables

- we apply one of the correlation function as we have seen before – use the `cor()` function, and select either Pearson's product-moment, Spearman's rho, or Kendall's tau (or in case of the latter, recode to ordinal and apply a Pearson's correlation if used together with numeric or double precision data)
- no need to standardise as one would do for the calculation of Euclidian distances
- it may be necessary to apply various transformations to the to get a linear response

Correlation: environmental variables

Doubs River data: Environment table

```
> env
# A tibble: 30 x 11
     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  49.1   792   2.5  1.3    8.1    94  0.2   0.41  0.12   7     8.1
 9  70.5   752   1.2  4.8    8      90  0.3   0.82  0.12   7.2   5.2
10  99     617   9.9 10      7.7    82  0.06  0.75  0.01  10     4.3
# ... with 20 more rows
```

(some transformation might help)

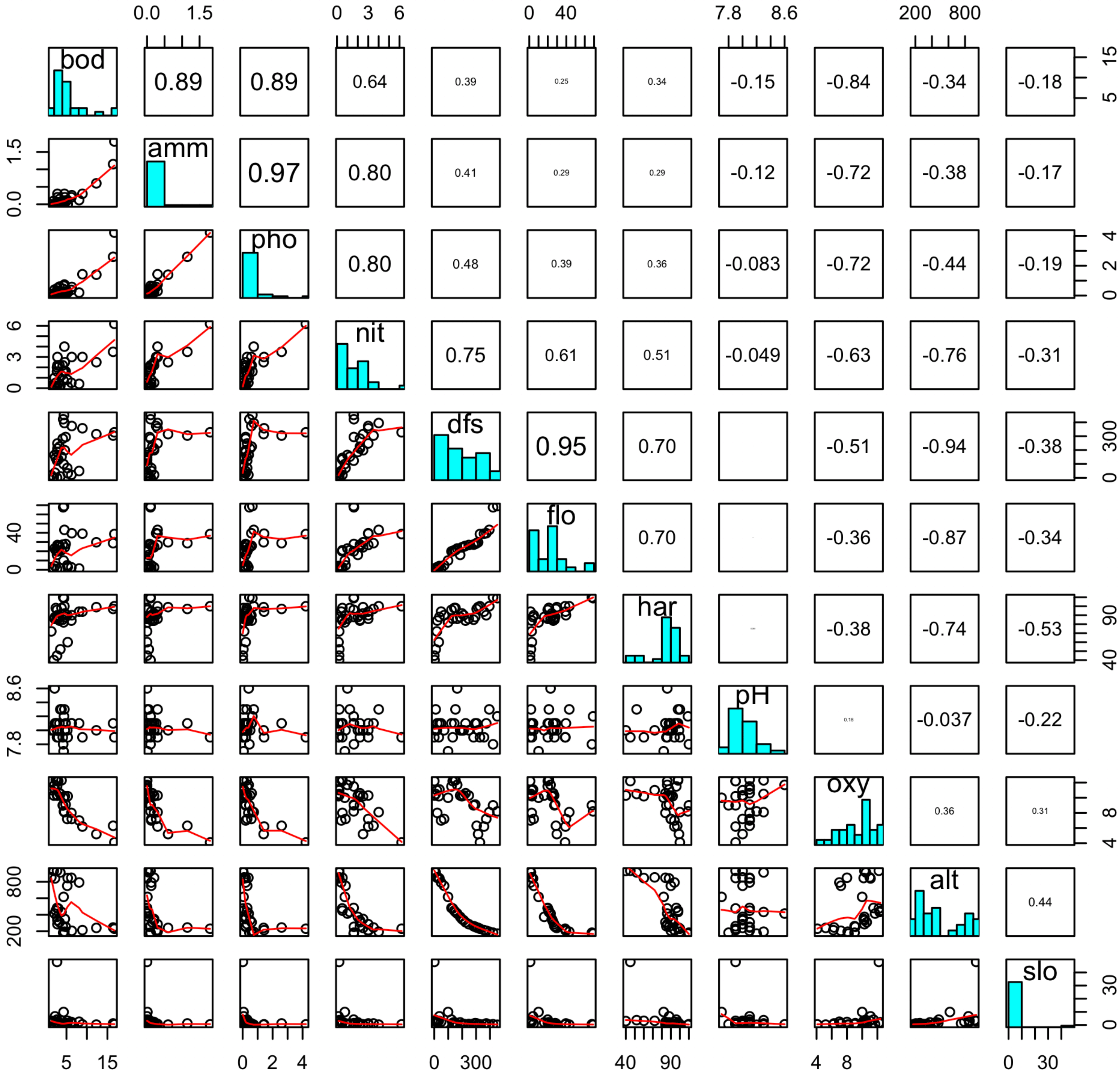
Pearson correlation coefficients

```
> env.pearson <- cor(env)        # default method = "pearson"
> round(env.pearson, 2)
      dfs   alt   slo   flo    pH   har   pho   nit   amm   oxy   bod
dfs  1.00 -0.94 -0.39  0.95  0.02  0.73  0.47  0.74  0.41 -0.57  0.43
alt -0.94  1.00  0.46 -0.86 -0.05 -0.79 -0.44 -0.75 -0.38  0.42 -0.38
slo -0.39  0.46  1.00 -0.36 -0.22 -0.53 -0.20 -0.31 -0.17  0.31 -0.17
flo  0.95 -0.86 -0.36  1.00  0.03  0.74  0.38  0.59  0.29 -0.42  0.30
pH   0.02 -0.05 -0.22  0.03  1.00  0.08 -0.08 -0.04 -0.12  0.19 -0.16
har  0.73 -0.79 -0.53  0.74  0.08  1.00  0.37  0.53  0.30 -0.37  0.34
pho  0.47 -0.44 -0.20  0.38 -0.08  0.37  1.00  0.80  0.97 -0.76  0.91
nit  0.74 -0.75 -0.31  0.59 -0.04  0.53  0.80  1.00  0.80 -0.69  0.68
amm  0.41 -0.38 -0.17  0.29 -0.12  0.30  0.97  0.80  1.00 -0.75  0.90
oxy -0.57  0.42  0.31 -0.42  0.19 -0.37 -0.76 -0.69 -0.75  1.00 -0.84
bod  0.43 -0.38 -0.17  0.30 -0.16  0.34  0.91  0.68  0.90 -0.84  1.00
```

no need to transpose or standardise as the `cor()` function does this internally

“comparison among ordinal variables, or among quantitative variables that may be monotonically but not linearly related, can be achieved using rank correlation coefficients like Spearman’s ρ (rho) or Kendall’s τ (tau)”

Pearson Correlation Matrix



Associations: species tables

- we apply one of the dissimilarity matrices
  - we do not derive Euclidian distances from species data, nor do we determine pairwise covariances or correlations (this is fine with the environmental data)
- it may be necessary to apply various transformations to the species data, e.g. when there are a few rare species
- transformation options are provided by `decostand()`; see section 3.5 in Numerical Ecology with R

Associations: species presence-absence

Doubs River data: 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
```

also available are the Sørensen and Ochiai coefficients

Table 3.1 Commonly-used distance and dissimilarity functions in Q mode available in R packages. The symbol ⇒ means that applying the function designed for quantitative data to presence–absence data produces the same result as computing the corresponding function designed for presence–absence data

Quantitative data		Presence–absence data
Community composition data		

Ruzicka dissimilarity
`vegdist(.,“jac”)`
	⇒	Jaccard dissimilarity
		`vegdist(.,”jac”,`
		`binary=TRUE)`
		`dist.binary(.,method=1)`

Hellinger distance `decostand(.,“hel”)` followed by `vegdist(.,“euc”)`	⇒	Ochiai dissimilarity `dist.binary(.,method=7)`
Chord distance `decostand(.,“norm”)` followed by `vegdist(.,“euc”)`	⇒	Ochiai dissimilarity `dist.binary(.,method=7)`
Bray–Curtis dissimilarity `vegdist(.,“bray”)`	⇒	Sørensen dissimilarity `dist.binary(.,method=5)`
Chi-square distance `decostand(.,“chi.square”)`		Chi-square distance (idem)
Canberra distance `vegdist(.,“canberra”)`		
Other variables, mixed physical units		
Standardized variables: Euclidean distance `vegdist(.,“euc”)`		Standardized variables: Simple matching coefficient `dist.binary(.,method=2)`
Non-standardized variables: Gower distance `daisy(.,“gower”)`		

Correlations and associations

- the square association and correlation matrices are generally only used as intermediate steps in our workflow, and are not usually scrutinised directly
- however, meaningful information is already present in these matrices, and it is beneficial to be able to read them
- it is definitely necessary to understand how they are calculated
- the next step in the workflow takes the distance, dissimilarity, association, and/or correlation matrices and applies the multivariate analyses on them
- we will continue with the distance and dissimilarity matrices