CONSTRAINED ORDINATION

Constrained ordinations

- constrained ordination adds a level of statistical testing (next topic)

- also called **direct gradient analysis** or **canonical ordination**

- uses an explanatory matrix to explain the patterns (or variability) in the response matrix

- (A) ...two matrices; explanatory grouping variable(s) (**env.**; **qualitative, ordinal**), and a multivariate response (**spp.**)

  - Discriminant Analysis (**DA**)

  - ✗ Analysis of Similarities (**ANOSIM**)

  - ✔ Permutational Analysis of Variance (**PERMANOVA**)

  - Mantel test

  - etc...

- (B) ...two matrices; a multivariate explanatory variables (**env.; continuous, categorical, or ordinal covariate(s)**), and a multivariate response (**spp.**)

  - Redundancy Discriminant Analysis (**RDA**) – 'extension' of PCA

  - ✔ Distance-based redundancy analysis (**db-RDA**)

  - ✔ Canonical Correspondence Analysis (**CCA**) – extension of CA

  - Canonical Analysis of Principal Coordinates (**CAP**) – extension of PCoA

Constrained ordinations

- explanatory variables sometimes called external variables
- used in an **hypothesis-driven** setting
- CCA and db-RDA combine classical ordination (CA and PCA, respectively) with multiple regression
- the significance of the ordinations is assessed via permutation (bootstrapping) tests, e.g. as in a PERMANOVA

Multivariate Techniques

Obs	Group	X-set				Y-set			
1	A	a_{11}	a_{12}	a_{13}	$... a_{1p}$	b_{11}	b_{12}	b_{13}	$... b_{1m}$
2	A	a_{21}	a_{22}	a_{23}	$... a_{2p}$	b_{21}	b_{22}	b_{23}	$... b_{2m}$
3	A	a_{31}	a_{32}	a_{33}	$... a_{3p}$	b_{31}	b_{32}	b_{33}	$... b_{3m}$
.	.	.	.	.	....	.	.	.	....
.	.	.	.	.	....	.	.	.	....
n	A	a_{n1}	a_{n2}	a_{n3}	$... a_{np}$	b_{n1}	b_{n2}	b_{n3}	$... b_{nm}$
n+1	C	c_{11}	c_{12}	c_{13}	$... c_{1p}$				
n+2	C	c_{21}	c_{22}	c_{23}	$... c_{2p}$				
n+3	C	c_{31}	c_{32}	c_{33}	$... c_{3p}$				
.	.	.	.	.	....				
.	.	.	.	.	....				
N	C	c_{n1}	c_{n2}	c_{n3}	$... c_{np}$				

factors

(A) Constrained ordination
Discriminant Analysis (DA)
Analysis of Similarities (ANOSIM)
Permutational ANOVA (PERMANOVA)
Mantel test
etc…

Multivariate Techniques

Obs	Group	X-set				Y-set			
1	A	a_{11}	a_{12}	a_{13}	... a_{1p}	b_{11}	b_{12}	b_{13}	... b_{1m}
2	A	a_{21}	a_{22}	a_{23}	... a_{2p}	b_{21}	b_{22}	b_{23}	... b_{2m}
3	A	a_{31}	a_{32}	a_{33}	... a_{3p}	b_{31}	b_{32}	b_{33}	... b_{3m}
.	.	.	.	.	... .	.	.	.	... .
.	.	.	.	.	... .	.	.	.	... .
n	A	a_{n1}	a_{n2}	a_{n3}	... a_{np}	b_{n1}	b_{n2}	b_{n3}	... b_{nm}
n+1	C	c_{11}	c_{12}	c_{13}	... c_{1p}				
n+2	C	c_{21}	c_{22}	c_{23}	... c_{2p}				
n+3	C	c_{31}	c_{32}	c_{33}	... c_{3p}				
.	.	.	.	.	... .				
.	.	.	.	.	... .				
N	C	c_{n1}	c_{n2}	c_{n3}	... c_{np}				

continuous

(B) Constrained ordination
Redundancy Discriminant Analysis (RDA)
Canonical Correspondence Analysis (CCA)
Distance-based redundancy analysis (db-RDA)
Canonical analysis of principal coordinates (CAP)

Redundancy Analysis (RDA)

- see https://sites.google.com/site/mb3gustame/constrained-analyses/rda and https://www.davidzeleny.net/anadat-r/doku.php/en:rda_cca
- a method that combines PCA with multiple regression
- it is therefore a 'linear' method, and the same constraints inherent in a PCA are present here too
- in essence, it is a multivariate (multiple response) multiple linear regression, followed by a PCA of the table of fitted values
- it works on centered response data (species, matrix **Y**) and standardised explanatory data (env., etc., matrix **X**)
- **X** conditions the weights (eigenvalues), the orthogonality, and the direction of the ordination axes
- an extremely useful and powerful method available to ecologists
- additional detail in 6.3.1. of Numerical Ecology with R

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

```
R> # Hellinger-transform the species dataset
R> # Suited to species abundance data. It gives low weights to species
R> # with low counts and many zeros. The transformation comprises dividing
R> # each value in a data matrix by its row sum, and taking the square root
R> # of the quotient.
R> spe.hel <- as.tibble(decostand(spe, "hellinger"))
R> spe.hel
# A tibble: 29 x 27
   Cogo  Satr  Phph  Babl  Thth  Teso  Chna  Pato  Lele  Sqce  Baba  Albi  Gogo  Eslu  Pefl  Rham  Legi  Scer  Cyca
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1     0 1     0     0         0     0     0 0     0         0         0 0 0     0     0         0 0 0         0
2     0 0.645 0.577 0.5       0     0     0 0     0         0         0 0 0     0     0         0 0 0         0
3     0 0.559 0.559 0.559     0     0     0 0     0         0         0 0 0     0.25  0         0 0 0         0
4     0 0.436 0.488 0.488     0     0     0 0     0         0.218     0 0 0.218 0.309 0.309     0 0 0         0
5     0 0.243 0.297 0.243     0     0     0 0     0.383     0.243     0 0 0.243 0.343 0.343     0 0 0.243     0
6     0 0.378 0.436 0.488     0     0     0 0     0.218     0.309     0 0 0.218 0.218 0.218     0 0 0         0
7     0 0.559 0.5   0.559     0     0     0 0     0.25      0.25      0 0 0     0     0         0 0 0         0
8     0 0     0.267 0.463     0     0     0 0     0         0.598     0 0 0     0     0         0 0 0         0
9     0 0.267 0.535 0.535     0     0     0 0     0.378     0.378     0 0 0.267 0     0         0 0 0         0
10 0.302 0.522 0.603 0.302 0.302     0     0 0     0         0.302     0 0 0     0     0         0 0 0         0
# … with 19 more rows, and 8 more variables: Titi <dbl>, Abbr <dbl>, Icme <dbl>, Gyce <dbl>, Ruru <dbl>, Blbj <dbl>,
#   Alal <dbl>, Anan <dbl>
```

```
R> # Remove empty site 8
R> env <- env[-8, ]
R> env
# A tibble: 29 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   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
```

```
R> # Remove the 'dfs' variable from the env dataset
R> env <- decostand(env[, -1], method = "standardize")
R> env
          alt         slo          flo          pH          har
1  1.71767530  5.04893671 -1.228203403 -0.8400173 -2.38806559 -0.6297
2  1.71025775 -0.06029282 -1.219304073 -0.2734940 -2.68052294 -0.6184
3  1.64349976  0.01918408 -1.174807424  1.4260758 -1.97862529 -0.5847
4  1.42097314 -0.03758514 -1.134204232 -0.2734940 -0.80879586 -0.5285
5  1.40242926 -0.13976973 -1.128085943  0.2930293 -0.10689821 -0.2136
6  1.39130293 -0.03758514 -1.115849365 -0.8400173 -1.51069352 -0.4160
7  1.37275904  0.34844554 -1.052441640  0.2930293  0.12706768 -0.5622
8  1.04267789 -0.26466200 -1.007944992 -0.2734940  0.24405062 -0.3036
9  0.54199300  0.72312237 -0.718716775 -1.9730638 -0.22388115 -0.5734
10 0.04501688  0.06459945 -0.168070747  0.2930293  0.59499945 -0.3036
```

```
# RDA of Hellinger-transformed fish data, constrained
# by all env. vars. in the environmental dataset (sans d.f.s.)
spe.rda <- rda(spe.hel ~ ., env)
```

```
> sum(spe.rda$CA$eig)
[1] 0.1371389
```

```
> sum(spe.rda$CCA$eig)
[1] 0.3653713
```

```
> (spe.rda <- rda(spe.hel ~ ., env2)) # Observe the shortcut
Call: rda(formula = spe.hel ~ alt + slo + flo + pH + har + pho + nit + amm + oxy + bod, data = env2)

              Inertia Proportion Rank
Total          0.5025     1.0000
Constrained    0.3654     0.7271    12
Unconstrained  0.1371     0.2729    16
Inertia is variance
```

Eigenvalues for constrained axes: RDA*n* eigenvalues measure amount of variance explained by the RDA

```
    RDA1    RDA2    RDA3    RDA4    RDA5    RDA6    RDA7    RDA8    RDA9   RDA10   RDA11   RDA12
 0.22808 0.05370 0.03212 0.02321 0.00870 0.00722 0.00487 0.00292 0.00214 0.00116 0.00091 0.00034
```

Eigenvalues for unconstrained axes: PC*n* eigenvalues measure amount of variance represented by the residual axes, *i.e.* **not** explained by the RDA

```
     PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10    PC11    PC12    PC13    PC14    PC15    PC16
 0.04581 0.02814 0.01528 0.01399 0.00984 0.00767 0.00420 0.00331 0.00276 0.00202 0.00175 0.00099 0.00059 0.00047 0.00021 0.00010
```

```
> # Proportion of variation explained by RDA1 is therefore
> spe.rda$CCA$eig[1] / (sum(spe.rda$CCA$eig) + sum(spe.rda$CA$eig))
     RDA1
0.4538882
```

```
> summary(spe.rda)      # Scaling 2 (default)

Call:
rda(formula = spe.hel ~ alt + slo + flo + pH + har + pho + nit +      amm + oxy + bod, data = env2)

Partitioning of variance:
          Inertia Proportion
Total        0.5025    1.0000
Constrained  0.3654    0.7271
Unconstrained 0.1371   0.2729

Eigenvalues, and their contribution to the variance

Importance of components:
                    RDA1    RDA2    RDA3    RDA4    RDA5    RDA6    RDA7    RDA8    RDA9    RDA10    RDA11     RDA12      PC1      PC2
Eigenvalue          0.2281  0.0537  0.03212 0.02321 0.008699 0.007218 0.004869 0.002924 0.002141 0.001160 0.000914 0.0003405 0.04581  0.02814
Proportion Explained 0.4539 0.1069  0.06392 0.04618 0.017311 0.014363 0.009689 0.005819 0.004261 0.002308 0.001819 0.0006776 0.09116  0.05601
Cumulative Proportion 0.4539 0.5607 0.62466 0.67084 0.688155 0.702518 0.712207 0.718027 0.722287 0.724596 0.726415 0.7270922 0.81825 0.87425
                     PC3     PC4     PC5     PC6     PC7     PC8     PC9     PC10     PC11     PC12      PC13      PC14      PC15
Eigenvalue           0.01528 0.01399 0.009841 0.007674 0.004201 0.003311 0.002761 0.002017 0.001752 0.0009873 0.0005919 0.0004674 0.0002128
Proportion Explained 0.03042 0.02784 0.019583 0.015271 0.008361 0.006588 0.005494 0.004013 0.003486 0.0019647 0.0011779 0.0009301 0.0004234
Cumulative Proportion 0.90467 0.93251 0.952090 0.967361 0.975722 0.982310 0.987804 0.991818 0.995304 0.9972687 0.9984467 0.9993767 0.9998002
                     PC16
Eigenvalue           0.0001004
Proportion Explained 0.0001998
Cumulative Proportion 1.0000000

Accumulated constrained eigenvalues
Importance of components:
                    RDA1    RDA2    RDA3    RDA4    RDA5    RDA6    RDA7    RDA8    RDA9    RDA10    RDA11     RDA12
Eigenvalue          0.2281  0.0537  0.03212 0.02321 0.008699 0.007218 0.004869 0.002924 0.002141 0.001160 0.000914 0.0003405
Proportion Explained 0.6243 0.1470  0.08791 0.06351 0.023808 0.019755 0.013326 0.008003 0.005860 0.003175 0.002502 0.0009320
Cumulative Proportion 0.6243 0.7712 0.85913 0.92264 0.946448 0.966202 0.979528 0.987532 0.993391 0.996566 0.999068 1.0000000

Scaling 2 for species and site scores
* Species are scaled proportional to eigenvalues
* Sites are unscaled: weighted dispersion equal on all dimensions
* General scaling constant of scores:  1.93676
```

- the coordinates of the tips of the vectors representing the response variables in the biplots (or triplots); they depend on the scaling chosen

```
Species scores

          RDA1      RDA2      RDA3      RDA4      RDA5      RDA6
Cogo  0.13386   0.11619  -0.238205  0.018531   0.043161  -0.029728
Satr  0.64240   0.06654   0.123649  0.181606  -0.009584   0.029785
Phph  0.47477   0.07009  -0.010153 -0.115349  -0.045312  -0.030034
Babl  0.36260   0.06966   0.041311 -0.190563  -0.046944   0.006446
Thth  0.13081   0.10707  -0.239273  0.043512   0.065818   0.003468
Teso  0.06590   0.12471  -0.216916 -0.004228   0.021732  -0.004195
Chna -0.17416   0.06774  -0.008397 -0.016402  -0.079764   0.044700
Pato -0.12679   0.16050  -0.035692 -0.016060  -0.089638  -0.001905
Lele -0.07961   0.04200   0.007677 -0.059155  -0.033512  -0.121451
Sqce -0.10903  -0.17555  -0.090012 -0.168382   0.019417   0.008753
Baba -0.18525   0.21152  -0.073101 -0.006901  -0.012999   0.040480
Albi -0.16059   0.15514  -0.014286 -0.002463  -0.060684   0.011025
Gogo -0.20538   0.02482  -0.007965 -0.017743  -0.049209  -0.096232
Eslu -0.10731   0.02857   0.090028  0.012339   0.075334  -0.057091
Pefl -0.09159   0.10509   0.070433 -0.057397   0.014063  -0.009913
Rham -0.20905   0.16004   0.025478  0.012081  -0.011453   0.022031
Legi -0.22798   0.11121   0.018794 -0.009470  -0.027464   0.024518
Scer -0.16101   0.01353   0.041554  0.032367   0.054060  -0.094579
Cyca -0.17381   0.14903   0.022239  0.009535   0.005020  -0.005397
Titi -0.14025   0.10636   0.078279 -0.122633   0.054114   0.031262
Abbr -0.18596   0.12223   0.053808  0.026142   0.043877   0.014596
Icme -0.14631   0.08899   0.061818  0.034746   0.083492   0.004417
Gyce -0.30883   0.01606   0.039351  0.029260  -0.011189  -0.052403
Ruru -0.31982  -0.16596  -0.018193 -0.115462   0.054397   0.064760
Blbj -0.23901   0.09089   0.051566  0.010197   0.006795   0.036516
Alal -0.43215  -0.22643  -0.108131  0.138837  -0.083787   0.008447
Anan -0.19440   0.14152   0.033624  0.017384   0.008122   0.017637
```

```
Site scores (weighted sums of species scores)

          RDA1       RDA2      RDA3       RDA4       RDA5       RDA6
row1    0.40149  -0.154133   0.55506   1.601005   0.193044   0.916850
row2    0.53522  -0.025131   0.43393   0.294832  -0.518997   0.458849
row3    0.49429  -0.014617   0.49415   0.169258  -0.246061   0.163409
row4    0.33451   0.001188   0.51644  -0.320793   0.089569  -0.219820
row5    0.02794  -0.194202   0.44626  -0.559104   0.855973  -1.115731
row6    0.24422  -0.130758   0.41397  -0.696109   0.182246  -0.273498
row7    0.46589  -0.126054   0.31692  -0.137640  -0.549192  -0.061715
row8    0.03660  -0.605090  -0.06989  -1.261027   0.668699   1.165074
row9    0.31380  -0.198797   0.10795  -0.635022  -0.742828  -0.990209
row10   0.48116  -0.039713  -0.37858   0.181771   0.220972   0.254669
row11   0.49162   0.014144  -0.37992   0.162929   0.223012   0.324824
row12   0.49848   0.212240  -0.67436   0.518502   0.399406   0.221762
row13   0.38203   0.229406  -0.75795   0.223281   0.515085  -0.139590
row14   0.28740   0.218549  -0.71897  -0.211161   0.175136  -0.553069
row15   0.09131   0.400116  -0.34437  -0.376171  -0.366402  -0.575447
row16  -0.05303   0.423927  -0.41000  -0.188493  -0.725181   0.151547
row17  -0.14182   0.385867  -0.36804  -0.217143  -0.643463  -0.001315
row18  -0.28203   0.275496  -0.01868  -0.371433  -0.692317  -0.062410
row19  -0.39681   0.209508   0.11549  -0.177920  -0.387207   0.048563
row20  -0.42850   0.278334   0.22003  -0.005965  -0.027098  -0.042218
row21  -0.46552   0.251911   0.22773   0.040193   0.152844   0.032242
row22  -0.28125  -1.145589  -0.50524   0.300043  -0.003393   1.157109
row23  -0.40896  -0.752956  -0.26777   0.428883  -0.647205   0.643149
row24  -0.35207  -0.770295  -0.12183   0.459218   0.079646  -1.726015
row25  -0.46923   0.094061   0.23046   0.107850   0.310149   0.132668
row26  -0.47021   0.213629   0.24872   0.084194   0.331473   0.125540
row27  -0.47266   0.234036   0.27037   0.105748   0.381225   0.093844
row28  -0.37456   0.393349   0.10404   0.202063   0.282836   0.021890
row29  -0.48931   0.321574   0.31409   0.278210   0.488026  -0.150951
```

```
Biplot scores for constraining variables

                  RDA1       RDA2       RDA3       RDA4       RDA5       RDA6
alt             0.8239  -0.203027   0.46599   -0.16932   0.003151   0.10413
slomoderate    -0.3592  -0.008707  -0.21727   -0.18287   0.158087   0.50092
slosteep        0.2791   0.155999  -0.06881    0.01921   0.176534  -0.15468
slovery_steep   0.6129  -0.148526   0.45392    0.03640  -0.190954  -0.04712
flo            -0.7769   0.255048  -0.17494    0.30980   0.227429  -0.11932
pH              0.1023   0.178385  -0.30143    0.03941   0.298243   0.04856
har            -0.5722   0.044958  -0.56049   -0.14839   0.275283  -0.24521
pho            -0.4930  -0.650420  -0.19865    0.29287   0.056281  -0.39347
nit            -0.7755  -0.203755  -0.23289    0.19739  -0.078644  -0.35079
amm            -0.4744  -0.687510  -0.16641    0.28475  -0.051257  -0.33858
oxy             0.7632   0.575392  -0.16436    0.08018  -0.136461   0.13748
bod            -0.5171  -0.791730  -0.15644    0.22067   0.075935  -0.09106


Centroids for factor constraints

                  RDA1       RDA2       RDA3       RDA4       RDA5       RDA6
slolow         -0.2800   0.005549  -0.09029    0.07610  -0.07882  -0.18390
slomoderate    -0.2093  -0.005073  -0.12660   -0.10656   0.09212   0.29188
slosteep        0.1965   0.109846  -0.04845    0.01353   0.12431  -0.10892
slovery_steep   0.3908  -0.094698   0.28941    0.02321  -0.12175  -0.03004
```

```
> # Canonical coefficients from the rda object
> coef(spe.rda)
                     RDA1          RDA2          RDA3          RDA4          RDA5          RDA6          RDA7          RDA8          RDA9
alt           0.0004483347  7.795777e-05  0.0005188756  0.0003875664  0.001857466 -6.331179e-05 -0.001359974  0.001120126 -0.0002517503
slomoderate  -0.0123140760 -1.655649e-02  0.0160736225 -0.0278161005  0.276687997  1.310961e-01 -0.022999100  0.018861788 -0.3109131626
slosteep      0.0480170930  4.905556e-02  0.1023432587  0.1349154913  0.394849998 -1.795415e-01  0.046031368  0.123198624  0.0970066893
slovery_steep 0.0181630025 -5.708251e-02  0.2326204779  0.1005535075  0.037825142 -1.741339e-01  0.516814197  0.069166271 -0.2260049338
flo          -0.0014041126  4.456720e-03  0.0089169975  0.0164695330  0.013331778  2.705617e-03 -0.002458359  0.010400061 -0.0006270792
pH            0.0189943698 -3.167179e-02 -0.0480329994  0.1142667244  0.413629926  1.092387e-01  0.141057854 -0.436371199 -0.0219077691
har           0.0025569166 -1.952059e-03 -0.0065931826 -0.0093613897  0.005218302 -6.097975e-03  0.002166368  0.010548543 -0.0006706737
pho           0.1033308237  4.578896e-02 -0.0998238552 -0.1049627245  0.423624076 -3.693234e-01  0.037569158 -0.700230790 -0.2878854128
nit          -0.0123768964  1.045028e-01  0.0622488395  0.0772807715  0.234375571 -3.544812e-02 -0.241069952  0.128366902 -0.0685703803
amm          -0.1088564269 -4.418756e-01  0.0062834673  0.0542213439 -1.813295370  4.650524e-03  0.280777588  1.065125427  0.3119031905
oxy           0.0686523744  1.987930e-02 -0.0895382002  0.1200081482  0.031939548  3.879211e-02 -0.058254055  0.061392921 -0.0196091542
bod           0.0108347803 -2.687024e-02 -0.0254010855  0.0744748837  0.067062942  9.276105e-02 -0.019974123  0.047982742  0.0359019363
                     RDA10         RDA11         RDA12
alt           0.001189022  0.0006783471  0.0009454291
slomoderate  -0.277963100  0.0402707169 -0.2974017594
slosteep     -0.445926115  0.2456799680 -0.3472509351
slovery_steep -0.588753472  0.2475356694 -0.1844733035
flo           0.004413371 -0.0022711471  0.0064860542
pH           -0.902424704  0.0732392948  0.5765506530
har           0.003336283  0.0007591452  0.0062067627
pho           0.247283885 -0.0018071075 -0.6300023622
nit           0.113704621  0.3978086000  0.0936975559
amm          -1.224213088 -1.5939301235  0.8974262542
oxy           0.089825552  0.0623806783  0.0257096795
bod           0.065688349  0.1112209162  0.0402215529
> # Unadjusted R^2 retrieved from the rda object
> (R2 <- RsquareAdj(spe.rda)$r.squared)
[1] 0.7270922
> # Adjusted R^2 retrieved from the rda object
> (R2adj <- RsquareAdj(spe.rda)$adj.r.squared)
[1] 0.5224114
```

```
R> anova(spe.rda) # ... yes!
Permutation test for rda under reduced model
Permutation: free
Number of permutations: 999

Model: rda(formula = spe.hel ~ alt + slo + flo + pH + har + pho + nit + amm + oxy + bod, data = env)
         Df Variance      F Pr(>F)
Model    10  0.36889 4.9695  0.001 ***
Residual 18  0.13362
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R>
R> # The adjusted R2 --- the variance explained by the constrained axes:
R> RsquareAdj(spe.rda)$adj.r.squared
[1] 0.5863788
```

```
R> # Variance explained by full model:
R> sum(spe.rda$CCA$eig) / spe.rda$tot.chi * 100
[1] 73.41007
R>
R> # Which axes are significant?
R> anova(spe.rda, by = "axis")
Permutation test for rda under reduced model
Forward tests for axes
Permutation: free
Number of permutations: 999

Model: rda(formula = spe.hel ~ alt + slo + flo + pH + har + pho + nit + amm + oxy + bod, data = env)
         Df Variance       F Pr(>F)
RDA1      1 0.228027 30.7183  0.001 ***
RDA2      1 0.054418  7.3309  0.001 ***
RDA3      1 0.033819  4.5559  0.049 *
RDA4      1 0.030069  4.0507  0.094 .
RDA5      1 0.007488  1.0087  1.000
RDA6      1 0.005652  0.7615  1.000
RDA7      1 0.004427  0.5963  1.000
RDA8      1 0.002816  0.3794  1.000
RDA9      1 0.001383  0.1863  1.000
RDA10     1 0.000794  0.1069  1.000
Residual 18 0.133617
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R>
R> # Which terms are significant?
R> anova(spe.rda, by = "terms")
Permutation test for rda under reduced model
Terms added sequentially (first to last)
Permutation: free
Number of permutations: 999

Model: rda(formula = spe.hel ~ alt + slo + flo + pH + har + pho + nit + amm + oxy + bod, data = env)
         Df Variance       F Pr(>F)
alt       1 0.164859 22.2087  0.001 ***
slo       1 0.028371  3.8219  0.004 **
flo       1 0.015431  2.0788  0.085 .
pH        1 0.011575  1.5593  0.157
har       1 0.017701  2.3846  0.066 .
pho       1 0.041819  5.6335  0.001 ***
nit       1 0.018911  2.5476  0.065 .
amm       1 0.007005  0.9437  0.426
oxy       1 0.058402  7.8675  0.001 ***
bod       1 0.004819  0.6492  0.628
Residual 18 0.133617
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
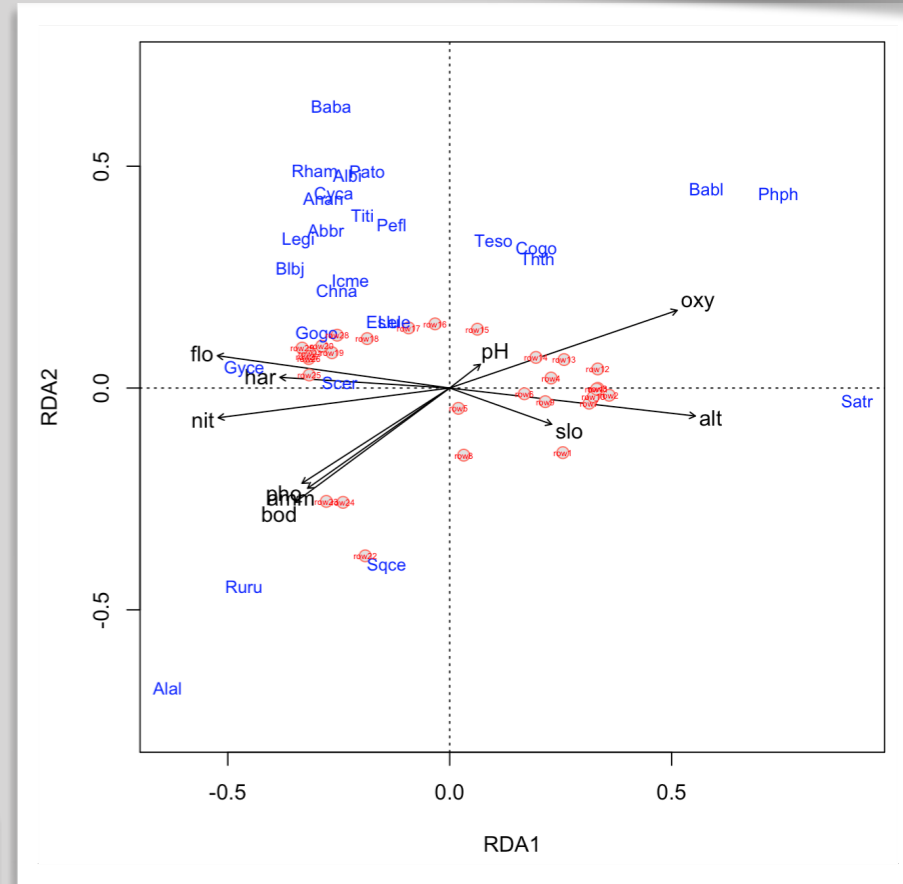
```r
# Another way - build from scratch:
## better control -- remember to set scaling etc. identically
plot(spe.rda, type = "n", scaling = "sites", choices = 1:2)
text(spe.rda, dis = "cn", scaling = "sites", choices = 1:2)
points(spe.rda, pch = 21, col = "salmon", bg = "grey90", cex = 1.2, scaling = "sites", choices = 1:2)
text(spe.rda, "species", col = "blue", cex = 0.8, scaling = "sites", choices = 1:2)
text(spe.rda, "sites", col = "red", cex = 0.4, scaling = "sites", choices = 1:2)
```
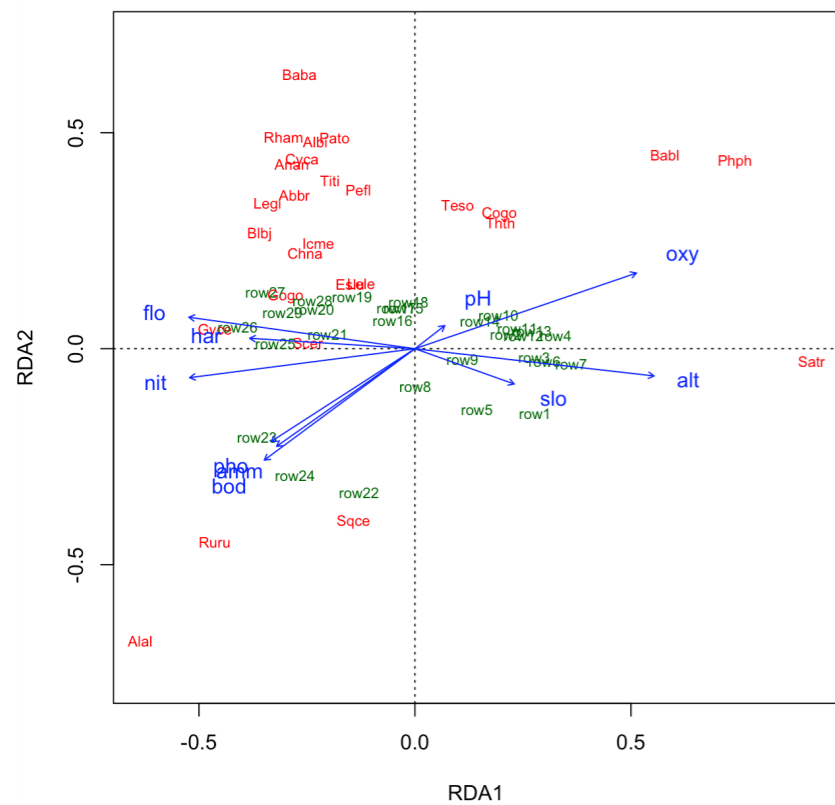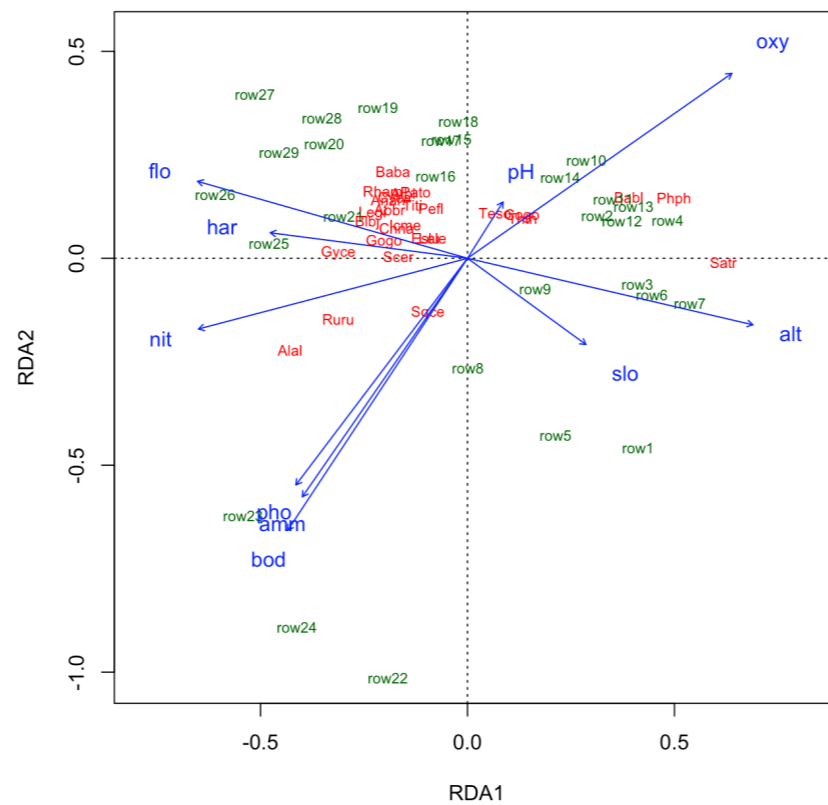
```r
# Make some plots:
plot(spe.rda, scaling = 1, display = c("sp", "lc", "cn"),
     main = "Site scaling (scaling 1)")
plot(spe.rda, scaling = 2, display = c("sp", "lc", "cn"),
     main = "Species scaling (scaling 2)")
```

Canonical Correspondence Analysis (CCA)

- see https://sites.google.com/site/mb3gustame/constrained-analyses/cca and https://www.davidzeleny.net/anadat-r/doku.php/en:rda_cca
- CCA is the merger between CA and multiple regression, therefore also based on χ^2-metric (dissimilarities)
- whereas CA maximises the correlation between species scores and sample scores, in CCAs the sample scores are constrained to be linear combinations of explanatory variables
- because of the 'constraint', eigenvalues in CCA will be lower than in CA
- the link to the unimodal model is clear—if a combination of environmental variables is strongly related to species composition, CCA will create a reduced axis from these variables that makes the species response curves most distinct
- second and higher axes will also maximise the dispersion (remaining inertia) of species, subject to the constraints that these higher axes are linear combinations of the explanatory variables, and that they are orthogonal (i.e. do not explain what has already been explained) to all previous axis

Canonical Correspondence Analysis (CCA)

- there are as many constrained (reduced) axes as there are explanatory variables
- the total 'explained inertia' is the sum of the eigenvalues of the constrained axes
- the remaining axes are unconstrained, and can be considered 'residual'
- the total inertia in the species data is the sum of eigenvalues of the constrained and the unconstrained axes, and is equivalent to the sum of eigenvalues, or total inertia, of CA
- thus, explained inertia, compared to total inertia, can be used as a measure of how well species composition is explained by the variables
- unfortunately, a strict measure of 'goodness of fit' for CCA is elusive, because the arch effect itself has some inertia associated with it–and it is not always clear whether this inertia belongs in the 'explained' or 'unexplained' portion

Canonical Correspondence Analysis (CCA)

- CCA benefits from the advantages of multiple regression, including:
    - it is possible that patterns result from the combination of several explanatory variables; these patterns would not be observable if explanatory variables are considered separately
    - it is possible to test hypotheses (though in CCA, hypothesis testing is based on randomisation procedures rather than distributional assumptions
    - explanatory variables can be of many types (e.g. continuous, ratio scale, nominal) and do not need to meet distributional assumptions

Canonical Correspondence Analysis (CCA)

- Some caveats that apply to multiple regression also apply here:
  - in observational studies, not always possible to infer direct causation
  - the independent effects of highly correlated (**collinear**) variables are difficult to disentangle—however, CCA (and regression) can test the null hypothesis that such variables are completely redundant
  - it is possible to 'overfit' the data as the number of variables approaches the number of samples (instead of r^2=~1, the explained inertia will equal the total inertia and the CCA solution equals the CA solution), so the solution is no longer 'constrained' by the variables
  - noise in explanatory variables will have an effect on the predicted values
  - the interpretability of the results is directly dependent on the choice and quality of the explanatory variables; and on the researcher's expert knowledge
  - although both multiple regression and CCA find the best linear combination of explanatory variables, they are not guaranteed to find the true underlying gradient (which may be related to unmeasured or unmeasurable factors), nor are they guaranteed to explain a large portion of variation in the data—again expert knowledge necessary

Canonical Correspondence Analysis (CCA)

- one of the biggest advantages of CCA lies in the intuitive nature of its ordination diagram, or triplot (you've seen this already when we did the 'envfit'); it is called a triplot because it simultaneously displays three pieces of information: samples as points, species as points, and environmental variables as arrows (or points)
- CCA triplots can get very crowded, and solutions for this include:
  - separate the parts of the triplot into biplots or scatterplots (e.g. plotting the arrows in a different panel of the same figure)
  - rescaling the arrows so that the species and sample scores are more spread out
  - only plotting the most abundant species (but remember the rare ones as knowledge of these is important in certain studies)
  - plotting only significant constraints as per PERMANOVA or some other means

- please read "Environmental variables in CCA" at http://ordination.okstate.edu

Canonical Correspondence Analysis (CCA)

- in **vegan**'s `cca()`, a chi-square transformed data matrix is subjected to weighted linear regression on constraining variables, and the fitted values are submitted to correspondence analysis performed via singular value decomposition (svd)

db-RDA

- in **vegan**'s `capscale()`, any dissimilarity matrix from `vegdist()` is subjected to weighted linear regression on constraining variables, and the fitted values are submitted to a PCoA
- the general benefits of a PCoA also apply here
- (dis)similarity matrices calculated from quantitative, semi-quantitative, qualitative, and mixed variables can be handled
- access to a wide array of dissimilarities make the method suitable to many kinds of ecological data
- since we apply species information is lost during the calculation of the dissimilarity matrix, if the original matrix of species composition matrix is available, the species scores can be added into the final ordination diagram as weighted means of site score in which they occur or as vectors fitted onto the ordination space

PERMANOVA vs. ANOSIM

- "Overall, ANOSIM and the Mantel test were very sensitive to heterogeneity in dispersions, with ANOSIM generally being more sensitive than the Mantel test. In contrast, PERMANOVA and Pillai's trace were largely unaffected by heterogeneity for balanced designs. [...]. PERMANOVA was also unaffected by differences in correlation structure. [...] PERMANOVA was generally, but not always, more powerful than the others to detect changes in community structure..."

Anderson, M. J., & Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecological Monographs, 83(4), 557–574. http://doi.org/10.1890/12-2010.1