Correlation matrix for environmental matrix

	env1	env2	env3	env4	env5	env6	env7	env8	env9	env10	env11	env12
env1	1	0.087451	0.113264	0.225049	-0.13835	0.338366	-0.01485	0.166309	-0.11046	0.088327	-0.41099	-0.19944
env2	0.087451	1	0.13723	-0.27979	0.062584	0.175709	0.282526	-0.03127	0.105821	0.995779	0.013042	-0.30027
env3	0.113264	0.13723	1	0.000589	-0.09107	0.161244	0.154152	-0.00599	-0.01457	0.14354	0.262487	-0.03718
env4	0.225049	-0.27979	0.000589	1	0.051554	0.258213	-0.2701	-0.29359	0.034171	-0.3064	-0.36954	0.054425
env5	-0.13835	0.062584	-0.09107	0.051554	1	0.256326	-0.21568	0.173568	-0.02543	0.05572	0.050751	0.132103
env6	0.338366	0.175709	0.161244	0.258213	0.256326	1	-0.24099	0.224421	0.163937	0.148964	-0.07426	-0.05844
env7	-0.01485	0.282526	0.154152	-0.2701	-0.21568	-0.24099	1	-0.19038	0.191563	0.286013	-0.06477	-0.13396
env8	0.166309	-0.03127	-0.00599	-0.29359	0.173568	0.224421	-0.19038	1	-0.07767	-0.01752	0.232991	-0.12359
env9	-0.11046	0.105821	-0.01457	0.034171	-0.02543	0.163937	0.191563	-0.07767	1	0.058181	0.077117	0.019898
env10	0.088327	0.995779	0.14354	-0.3064	0.05572	0.148964	0.286013	-0.01752	0.058181	1	0.027291	-0.31873
env11	-0.41099	0.013042	0.262487	-0.36954	0.050751	-0.07426	-0.06477	0.232991	0.077117	0.027291	1	0.060384
env12	-0.19944	-0.30027	-0.03718	0.054425	0.132103	-0.05844	-0.13396	-0.12359	0.019898	-0.31873	0.060384	1

vif.cca (Variance Inflation Scores)

```
> vif.cca(base_cca)
  spatialb   spatialc       env1       env2       env3       env4       env5       env6       env7       env8       env9
  8.903666  19.974361   3.008475 332.023395   1.757367   3.599143   2.427301   2.187329   9.626137   3.031498   2.572406
     env10      env11      env12
350.060286   2.176076   2.399144
```

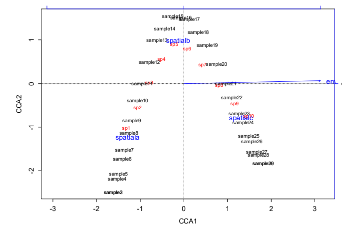
Eliminate variable env10:

```
spatialb  spatialc      env1      env2      env3      env4      env5      env6      env7      env8      env9     env11
7.920615 15.860032  2.974134  2.383930  1.756746  3.264919  1.845888  2.136770  8.893480  2.700565  1.586309  1.900621
   env12
1.855973
```

Eliminate variable spatial:

```
 env1     env2     env3     env4     env5     env6     env7     env8     env9    env11    env12
1.675300 1.605842 1.684828 1.789000 1.468596 1.889358 1.792075 2.267421 1.530498 1.642532 1.702221
```

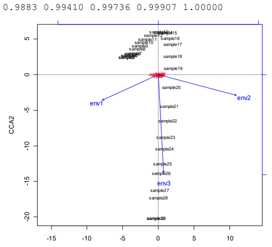
```
cca(formula = community ~ spatial + env7, data = envdata)

Partitioning of mean squared contingency coefficient:
              Inertia Proportion
Total          1.7795     1.0000
Constrained    1.2399     0.6967
Unconstrained  0.5396     0.3033

Eigenvalues, and their contribution to the mean squared contingency coefficient

Importance of components:
                        CCA1   CCA2   CCA3    CA1     CA2     CA3     CA4     CA5      CA6      CA7     CA8      CA9
Eigenvalue            0.8204 0.4049 0.0146 0.2735 0.13900 0.07275 0.02382 0.01248 0.009799 0.004215 0.00275 0.001338
Proportion Explained  0.4610 0.2275 0.0082 0.1537 0.07811 0.04088 0.01339 0.00701 0.005510 0.002370 0.00155 0.000750
Cumulative Proportion 0.4610 0.6885 0.6967 0.8504 0.92855 0.96943 0.98282 0.98983 0.995330 0.997700 0.99925 1.000000

Accumulated constrained eigenvalues
Importance of components:
                        CCA1   CCA2    CCA3
Eigenvalue            0.8204 0.4049 0.01460
Proportion Explained  0.6616 0.3266 0.01177
Cumulative Proportion 0.6616 0.9882 1.00000
```




```
> vif.cca(model_fit)
spatialb spatialc     env7
3.401162 8.427027 5.309164
```

CCA with first three variables only

```
Call:
cca(formula = community ~ env1 + env2 + env3, data = envdata)

Partitioning of mean squared contingency coefficient:
              Inertia Proportion
Total          1.7795     1.0000
Constrained    0.2047     0.1150
Unconstrained  1.5748     0.8850

Eigenvalues, and their contribution to the mean squared contingency coefficient

Importance of components:
              CCA1    CCA2    CCA3   CA1   CA2   CA3    CA4    CA5     CA6     CA7     CA8     CA9
Eigenvalue   0.184 0.01449 0.00631 0.776 0.409 0.249 0.0912 0.0297 0.01028 0.00580 0.00304 0.00165
Prop Explain 0.103 0.00814 0.00354 0.436 0.230 0.140 0.0513 0.0167 0.00578 0.00326 0.00171 0.00093
Cum. Prop.   0.103 0.11150 0.11505 0.551 0.781 0.920 0.9716 0.9883 0.99410 0.99736 0.99907 1.00000

Accumulated constrained eigenvalues
Importance of components:
                       CCA1   CCA2    CCA3
Eigenvalue            0.184 0.0145 0.00631
Proportion Explained  0.898 0.0707 0.03081
Cumulative Proportion 0.898 0.9692 1.00000
```



Stepwise model fitting

```
Start: community ~ 1

          Df    AIC       F N.Perm Pr(>F)
+ spatial  2 22.297 21.5441    199  0.005 **
+ env7     1 30.811 22.9271    199  0.005 **
+ env5     1 43.932  4.7571    199  0.005 **
+ env8     1 46.807  1.7596     99  0.160
+ env4     1 46.893  1.6743     99  0.170
+ env2     1 46.724  1.8425     99  0.240
+ env6     1 47.059  1.5103     99  0.260
+ env9     1 47.478  1.1019     99  0.370
+ env1     1 47.599  0.9845     99  0.450
+ env12    1 47.737  0.8519     99  0.480
+ env11    1 48.442  0.1830     99  0.880
+ env3     1 48.398  0.2239     99  0.940
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Step: community ~ spatial

          Df    AIC      F N.Perm Pr(>F)
- spatial  2 46.638 21.544     99   0.01 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        Df    AIC      F N.Perm Pr(>F)
+ env7   1 18.035 6.0248    199  0.005 **
+ env6   1 22.388 1.7011     99  0.180
+ env5   1 22.849 1.2791     99  0.370
+ env8   1 23.053 1.0954     99  0.390
+ env2   1 23.159 1.0004     99  0.420
+ env11  1 23.274 0.8970     99  0.470
+ env4   1 23.386 0.7976     99  0.470
+ env9   1 23.560 0.6429     99  0.670
+ env12  1 23.771 0.4569     99  0.810
+ env1   1 23.975 0.2788     99  0.920
+ env3   1 24.002 0.2553     99  0.930
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Step: community ~ spatial + env7
```

Note that exact output changes each time you run it (P values permutational).

Adjust Pin option to 0.2 and Pout to 0.3:

```
Step: community ~ spatial + env7

          Df    AIC      F N.Perm Pr(>F)
- env7     1 22.297 6.0248     99   0.01 **
- spatial  2 30.811 9.7913     99   0.01 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        Df    AIC      F N.Perm Pr(>F)
+ env5   1 17.963 1.7776     99   0.09 .
+ env6   1 18.574 1.2405    199   0.29
+ env11  1 18.860 0.9925    299   0.29
+ env8   1 18.908 0.9511     99   0.42
+ env4   1 18.904 0.9546     99   0.51
+ env12  1 19.412 0.5216     99   0.71
+ env2   1 19.572 0.3863     99   0.80
+ env1   1 19.672 0.3029     99   0.82
+ env3   1 19.749 0.2379     99   0.91
+ env9   1 19.867 0.1396     99   0.97
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Step: community ~ spatial + env7 + env5

          Df    AIC      F N.Perm Pr(>F)
- env5     1 18.035 1.7776     99   0.14
- env7     1 22.849 6.4326     99   0.01 **
- spatial  2 30.686 9.3612     99   0.01 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        Df    AIC      F N.Perm Pr(>F)
+ env6   1 18.117 1.5115    999  0.219
+ env8   1 18.640 1.0738    199  0.305
+ env4   1 18.749 0.9830     99  0.350
+ env11  1 18.637 1.0766     99  0.390
+ env12  1 19.275 0.5525     99  0.710
+ env1   1 19.393 0.4564     99  0.740
+ env2   1 19.427 0.4292     99  0.770
+ env3   1 19.561 0.3209     99  0.830
+ env9   1 19.794 0.1344     99  0.970
```

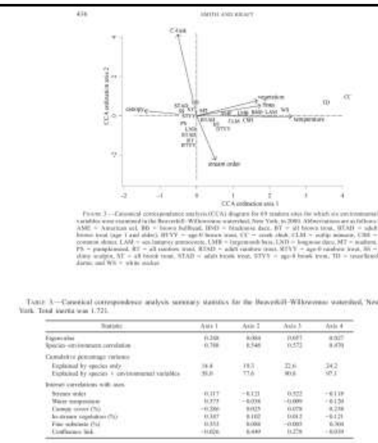
Tests of significance

```
> anova(model_fit)
Permutation test for cca under reduced model

Model: cca(formula = community ~ spatial + env7 + env5, data = envdata)
         Df   Chisq       F N.Perm Pr(>F)
Model     4  1.2771 15.2511    199  0.005 **
Residual 24  0.5024
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> anova(model_fit, by = "terms", perm = 1000)
Permutation test for cca under reduced model
Terms added sequentially (first to last)

Model: cca(formula = community ~ spatial + env7 + env5, data = envdata)
         Df   Chisq       F N.Perm Pr(>F)
spatial   2  1.1099 26.5073     99   0.01 **
env7      1  0.1301  6.2121     99   0.01 **
env5      1  0.0372  1.7776     99   0.21
Residual 24  0.5024
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> anova(model_fit, by = "axis", perm = 1000)
Model: cca(formula = community ~ spatial + env7 + env5, data = envdata)
         Df   Chisq       F N.Perm Pr(>F)
CCA1      1  0.8242 39.3697    199  0.005 **
CCA2      1  0.4139 19.7723    199  0.005 **
CCA3      1  0.0327  1.5635     99  0.160
CCA4      1  0.0063  0.2988     99  0.860
Residual 24  0.5024
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What to Present

The forward selection procedure for the CCA resulted in the retention of six variables as significant contributors to variation in the ordination: percent fine substrate, stream order, C-link, percent canopy cover, water temperature, and percent in-stream vegetation. The CCA produced four axes that together accounted for 24.2% of the total variance in fish species abundances among sites (Table 3). Eigenvalues, which range between 0 and 1, measure the importance of each axis. The first ordination axis accounted for 14.4% of the variance of the species data, whereas the second axis accounted for 4.9% of this variance; we did not attempt to interpret the third and fourth ordination axes (Table 3). The entire ordination accounted for more variation than expected by chance (Monte Carlo permutation tests, $N = 1000$, $P = 0.001$), indicating a significant relationship between species abundance and the environmental variables.

The Monte Carlo permutation tests indicated that the following factors were important in constructing CCA ordination axis one: percent fine substrate, percent in-stream vegetation, percent canopy cover, and water temperature (Table 3). Percent fine substrate, percent in-stream vegetation, and water temperature were positively correlated with the first ordination axis, whereas canopy cover was negatively correlated. C-link and stream order were the two factors important to CCA ordination axis two. C-link was positively correlated with the second ordination axis, and stream order showed a negative correlation.



Redundancy Analysis (RDA)

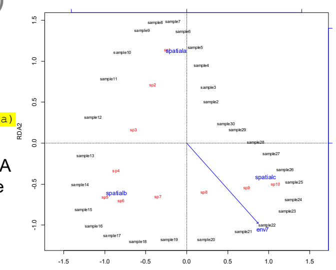
- CCA – aligned with CA, species relationships assumed to be unimodal
  - A CCA with no constraints = CA
- RDA – aligned with PCA, species relationships assumed to be linear
  - An RDA with no constraints = PCA
- CCA is more common for the same reasons PCA not usually used for species data.

Redundancy Analysis (RDA)

- Function rda (vegan package) works similarly to cca.

```
rda(community ~ spatial + env7, envdata)
```

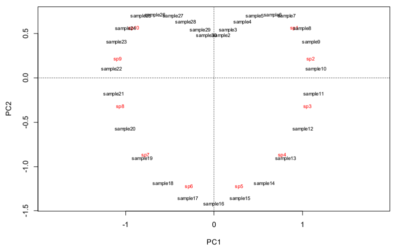
- Total inertia differs because PCA begins with variance/covariance matrix, not X^2




```
 Inertia Proportion
Total          10.000     1.0000
Constrained     5.533     0.5533
Unconstrained   4.467     0.4467

Eigenvalues, and their contribution to the correlations

Importance of components:
                       RDA1   RDA2    RDA3    PC1     PC2     PC3     PC4     PC5     PC6     PC7
Eigenvalue            2.9290 2.5034 0.10053 2.6564 0.92156 0.46490 0.25343 0.09547 0.02959 0.02058
Proportion Explained  0.2929 0.2503 0.01005 0.2656 0.09216 0.04649 0.02534 0.00955 0.00296 0.00206
Cumulative Proportion 0.2929 0.5432 0.55330 0.8189 0.91109 0.95758 0.98293 0.99248 0.99543 0.99749
```

Redundancy Analysis (RDA)

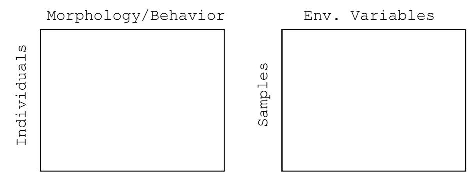
- PDA without envieonmental matrix = PCA




```
rda(X = community, scale = T)

Partitioning of correlations:
              Inertia Proportion
Total              10          1
Unconstrained      10          1

Eigenvalues, and their contribution to the correlations

Importance of components:
                        PC1    PC2    PC3     PC4     PC5     PC6     PC7     PC8     PC9     PC10
Eigenvalue           4.5669 3.2308 1.0247 0.95190 0.09795 0.05753 0.02858 0.02670 0.01022 0.004701
Proportion Explained 0.4567 0.3231 0.1025 0.09519 0.00979 0.00575 0.00286 0.00267 0.00102 0.000470
Cumulative Proportion 0.4567 0.7798 0.8822 0.97743 0.98723 0.99298 0.99584 0.99851 0.99953 1.000000
```

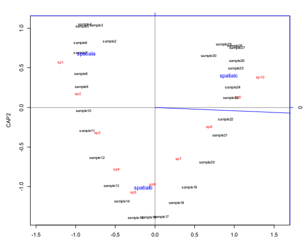
Redundancy Analysis (RDA)

- Example RDA application
  - How do various environmental variables affect morphology or behavior (plasticity question)?
- Assumption is that you know and have measured the environmental variables that are going to impact the morphology/behavior measured.



Constrained Analysis of Principal Coordinates (capscale)

- Function capscale (vegan package)
- Similar to RDA but based on non-Euclidean distances
- Function works with raw data (indicate which distance metric to use) or with a distance matrix (e.g. genetic distance)
- Same anova function tests for significance
- Various options for dealing with data standardization
- Also called "distance based redundancy analysis" in literature




```
Call:
capscale(formula = community ~ spatial + env7,
data = envdata,      distance = "bray")

Partitioning of squared Bray distance:
              Inertia Proportion
Total           9.762     1.0000
Constrained     5.993     0.6140
Unconstrained   3.177     0.3254
```

Anderson, M.J. & Willis, T.J. (2003). Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84, 511-525.

Gradients unknown, want to describe data or imply gradients from species data.

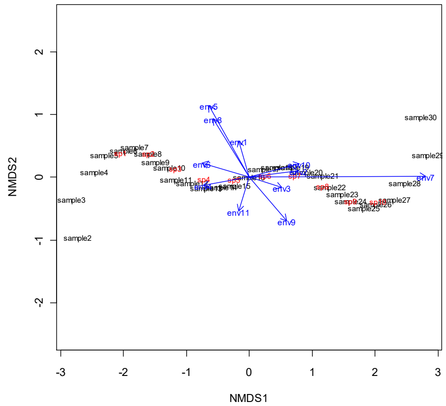
yes → Data are thought to be distinctly structured. Goal is to put in distinct groups.

yes → **Cluster** – hierarchical (agglomerative or divisive) or non hierarchial (k means).

no → Goal is to look at the relationship among samples. **Indirect gradient analysis**: NMDS or CA preferred, PCA or PCoA for non-community datasets.

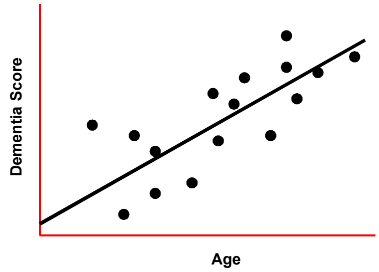
No → Important gradients are known and measured. Goal is to describe variability in species data as it relates to the measured gradients. Variability due to other unmeasured factors is ignored. **Direct gradient analysis**: CCA or RDA.

Unconstrained NMDS approach (exploratory)

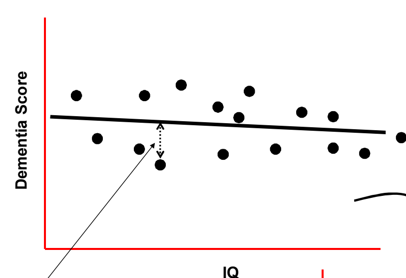


Partial CCA

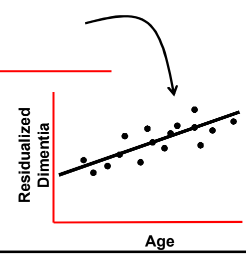
- Used when you want to control (or partial out) a variable.
  - Eg. larger spatial variable or something traditionally considered a covariate. Typically not something you're interested in quantifying as much as controlling for.
- Recall that CCA (and RDA, capscale) use multiple regression to optimize constraints to species or sample scores.
  - Partial Regression –
    - Regression coefficient of X after controlling for influence of Z on both X and Y.



Wanted – relationship between X (age) and y (degree of dementia). Numerous other variables confound the relationship, such as IQ and mean number of hours spent reading daily.



Residualized dementia score, effects of IQ have been "partialed" out.



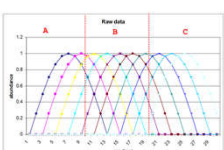
Partial CCA

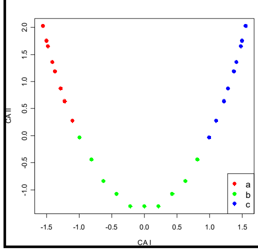
- All the same precautions apply as with CCA.
- Output will be nearly identical with the inclusion of the variance accounted for by the partialed variable.
  - Total inertial in the ordination – same as CA
  - Inertia partitioned out – comparison of the model with and without partial variable
  - Inertia constrained by variables in model
  - Unconstrained inertia

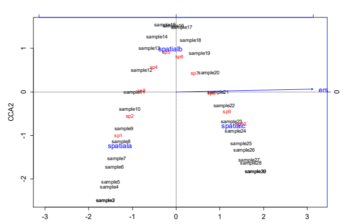
Partial CCA

- Uses the same function as CCA or RDA
- Code
  - `cca_example<-cca(community ~ env1 + env2 + Condition(spatial) ,environmental)`
- There is no significance test for the partialed variable(s), just the % variance accounted for.

Partial CCA

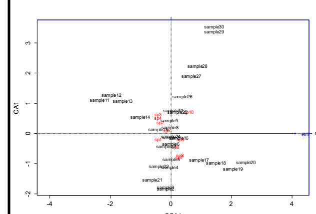






Test for influence of env7 while controlling for spatial variable.

```
cca(formula = community ~ env7 + Condition(spatial), data = envdata)

Partitioning of mean squared contingency coefficient:
              Inertia Proportion
Total          1.7795    1.00000
Conditioned    1.1099    0.62367
Constrained    0.1301    0.07308
Unconstrained  0.5396    0.30325

Eigenvalues, and their contribution to the mean squared contingency coefficient
after removing the contribution of conditiniong variables

Importance of components:
                        CCA1    CA1    CA2     CA3     CA4     CA5      CA6      CA7     CA8      CA9
Eigenvalue            0.1301 0.2735 0.1390 0.07275 0.02382 0.01248 0.009799 0.004215 0.00275 0.001338
Proportion Explained  0.1942 0.4084 0.2076 0.10862 0.03557 0.01863 0.014630 0.006290 0.00411 0.002000
Cumulative Proportion 0.1942 0.6026 0.8102 0.91877 0.95434 0.97297 0.987600 0.993890 0.99800 1.000000
```




```
> anova(partial_cca)
Permutation test for cca under reduced model

Model: cca(formula = community ~ env7 + Condition(spatial), data =
envdata)
         Df  Chisq      F N.Perm Pr(>F)
Model     1 0.1301 6.0248    199  0.005 **
Residual 25 0.5396
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(partial_cca, by = "terms", perm = 1000)
Permutation test for cca under reduced model
Terms added sequentially (first to last)

Model: cca(formula = community ~ env7 + Condition(spatial), data =
envdata)
         Df  Chisq      F N.Perm Pr(>F)
env7      1 0.1301 6.0248     99   0.01 **
Residual 25 0.5396
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```