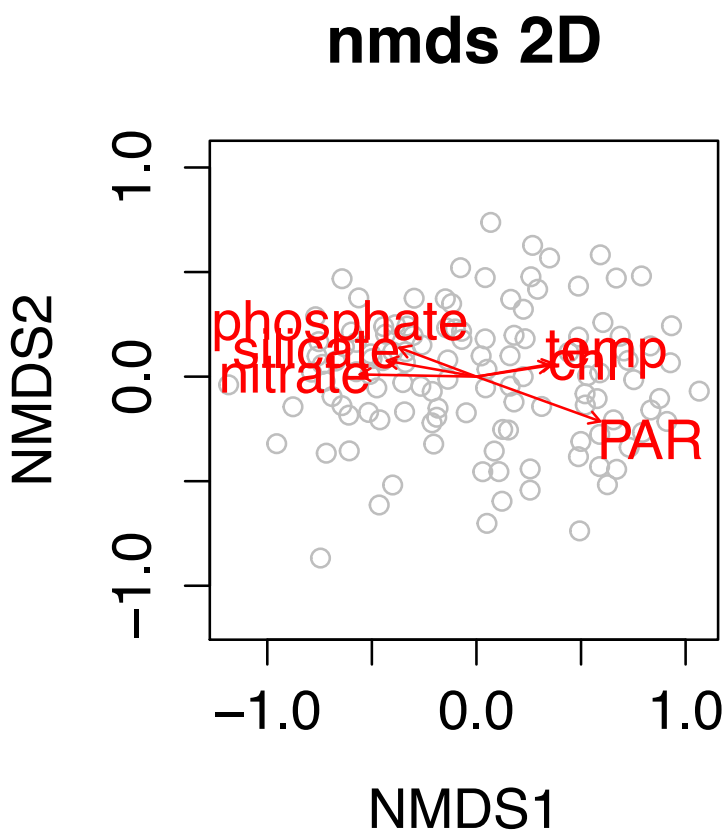
Comparing continuous environmental correlates with an ordination

```
fit = envfit(ord ~ nitrate + phosphate + silicate + PAR + chl + temp, data = e
nviro, na.rm = T)
ordiplot(ord, display = "sites", type = 'n', main = 'nmds 2D')
points(ord, col = 'grey')
plot(fit, col = 'red', arrow.mul = 0.7)
```

Important: this is not a model, just uses a formula for convenience

Comparing continuous environmental correlates with an ordination



- The vectors show the direction of maximum correlation,
  between the ordination scores and the environmental values for those samples

- Get an axis of stratification again: but now the ordination is based on species data, not environmental data

```
fit

##
## ***VECTORS
##
##             NMDS1  NMDS2   r2 Pr(>r)
## nitrate   -0.999 -0.049 0.66  0.001 ***
## phosphate -0.999 -0.050 0.33  0.001 ***
## silicate  -0.963 -0.271 0.42  0.001 ***
## PAR        1.000 -0.030 0.80  0.001 ***
## chl        0.821  0.570 0.31  0.001 ***
## temp       0.799 -0.601 0.33  0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 999
##
## 21 observations deleted due to missingness
```

envfit() also give correlation statitistics and p-values

These are from **permutation tests**

Randomly assign the environmental data to the samples. 1000 times. How often is the correlation this large?

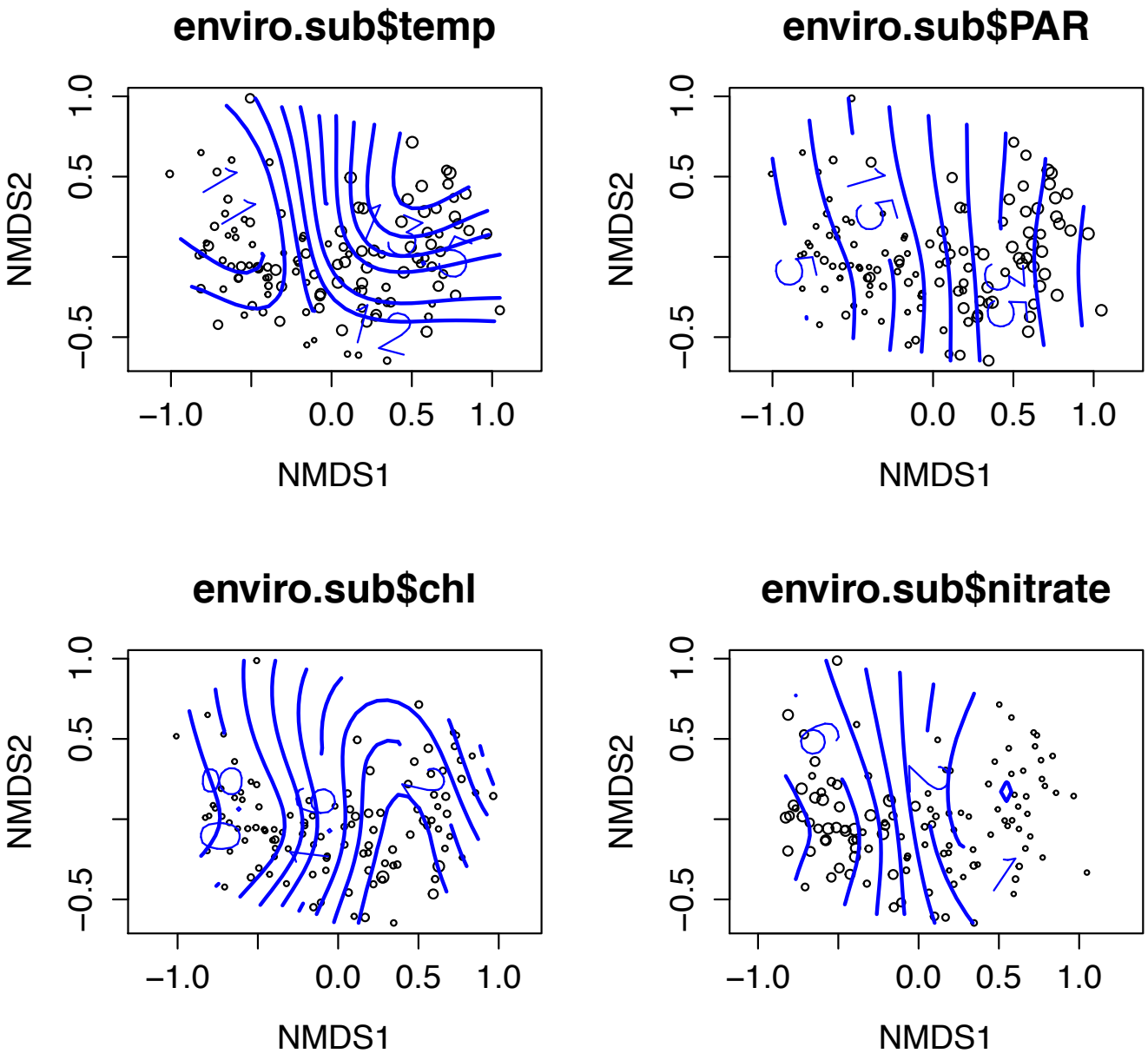
Warning: hidden assumptions. E.g. independence of samples.

Do environmental variables map nonlinearly onto the ordination?

Can use a GAM, with the **environment as the response** and the ordination scores as a **2D smoother**

e.g., nitrate ~ s(axis1, axis2)

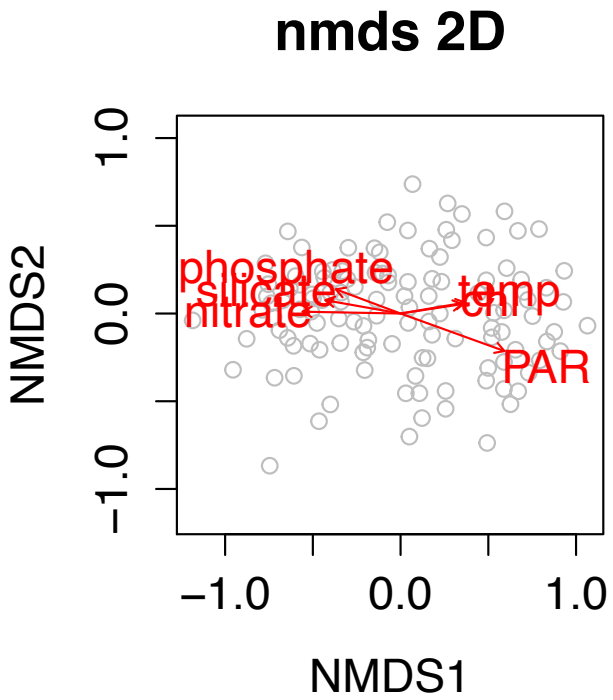
```
par(mfrow = c(2,2))
ordisurf(ord4, enviro.sub$temp, bubble = TRUE, labcex = 1.5, col = 'blue', lwd
 = 2)

ordisurf(ord4, enviro.sub$PAR, bubble = TRUE, labcex = 1.5, col = 'blue', lwd
= 2)

ordisurf(ord4, enviro.sub$chl, bubble = TRUE, labcex = 1.5, col = 'blue', lwd
= 2)

ordisurf(ord4, enviro.sub$nitrate, bubble = TRUE, labcex = 1.5, col = 'blue',
lwd = 2)
```


enviro.sub$temp
enviro.sub$PAR
enviro.sub$chl
enviro.sub$nitrate
NMDS1
NMDS2
−1.0
0.0
0.5
1.0
−0.5

Relating multivariate data to predictors

This kind of approach is good for seeing if dominant axes of composition are associated with the environment

But if you really want to directly test the role of some correlates, need a method designed for this



Constrained ordination

- Finding axes in ordination space that correlate with predictors

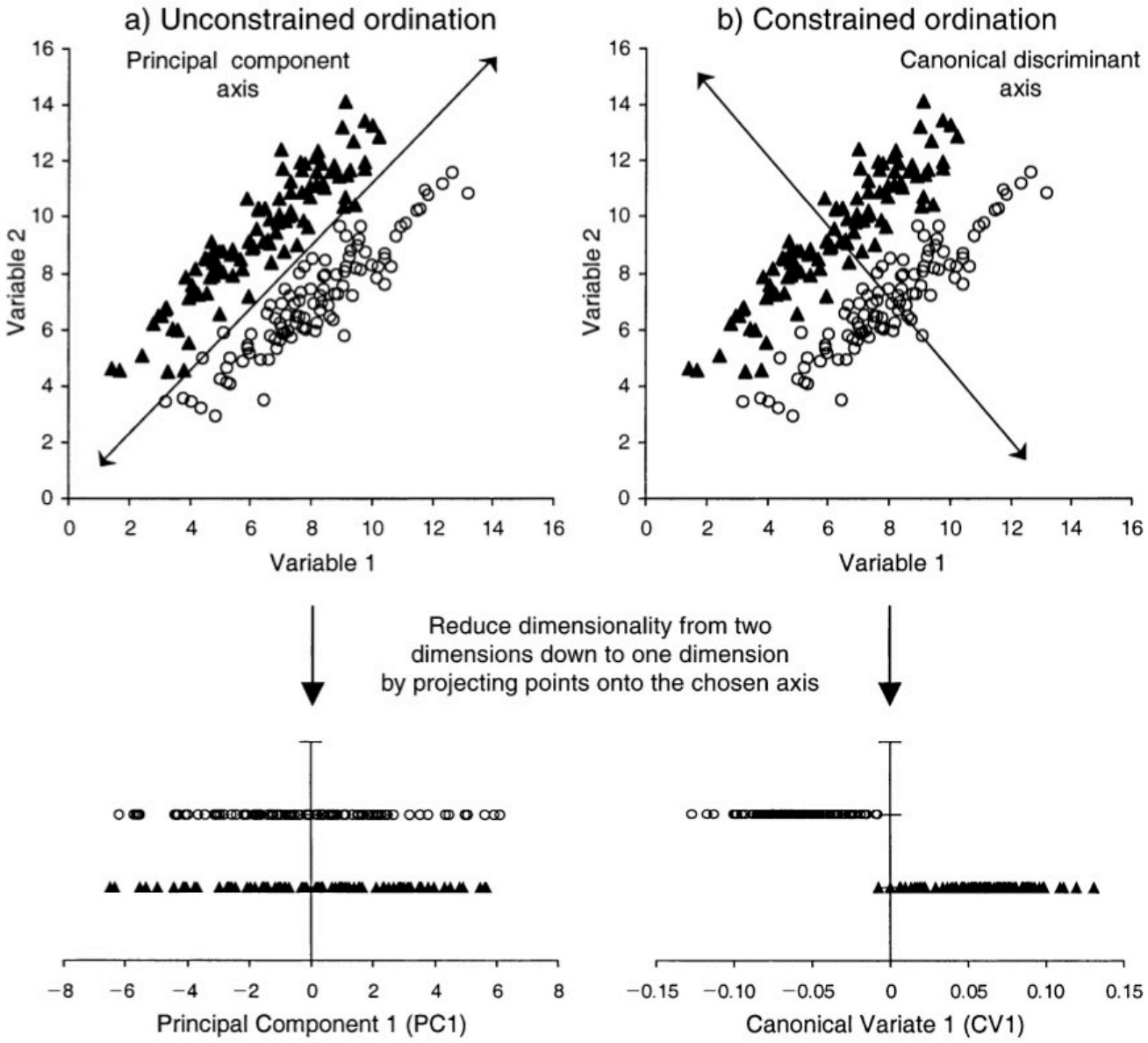


FIG. 1. Visual comparison of the method used to reduce dimensions in (a) an unconstrained and (b) a constrained ordination procedure. Data were simulated from a multivariate normal distribution with the two groups having different centroids (6, 9) and (9, 7), but both variables had a standard deviation of 2, and the correlation between the two variables was 0.9. Note the difference in scale between the first canonical axis (CV1) and the first principal component (PC1).

Constrained ordination

- Finding axes in ordination space that correlate with predictors
- There are different methods: canonical discriminant analysis, canonical correlation analysis, redundancy analysis, canonical analysis of principal coordinates, etc.
- The basic principle is the same: find the axes in the multivariate response along which the environment explains as much variation as possible
- This more directly answers the question of whether the environment can explain composition
- Because the goals are different, **unconstrained** ordination may not reveal the same patterns

Canonical analysis of principal coordinates (CAP)

Incorporates predictors into PCoA

1) Take your data on species' abundances, and turn it into a dissimilarity matrix

2) Take your dissimilarity matrix and do a PCoA on it

- Now we've gone from **dissimilarity** to **distances** in ordination space.
- Can use these distances to see how much variation is accounted for by predictors

3) Find axis in predictor space that explains the most variation in community ordination space

- Then find a second orthogonal axis that explains the most remaining variation, etc.

```r
species.data = wisconsin(sqrt(species.data))

cap = capscale(species.data ~ month, data = enviro, dist = "bray")
cap
```

```
## Call: capscale(formula = species.data ~ month, data = enviro,
## distance = "bray")
##
##               Inertia Proportion Rank
## Total          20.620
## Real Total     24.350      1.000
## Constrained     8.360      0.343   11
## Unconstrained  16.000      0.657   53
## Imaginary      -3.730              50
## Inertia is squared Bray distance
##
## Eigenvalues for constrained axes:
##  CAP1  CAP2  CAP3  CAP4  CAP5  CAP6  CAP7  CAP8  CAP9 CAP10 CAP11
##  5.50  1.07  0.49  0.45  0.21  0.18  0.16  0.12  0.08  0.06  0.04
##
## Eigenvalues for unconstrained axes:
##  MDS1  MDS2  MDS3  MDS4  MDS5  MDS6  MDS7  MDS8
## 1.832 1.717 1.284 1.124 0.872 0.805 0.724 0.624
## (Showed only 8 of all 53 unconstrained eigenvalues)
```

```
## Accumulated constrained eigenvalues
## Importance of components:
##                          CAP1  CAP2   CAP3   CAP4   CAP5   CAP6  CAP7   CAP8
## Eigenvalue             5.499 1.069 0.4924 0.4503 0.2078 0.1796 0.159 0.1164
## Proportion Explained   0.658 0.128 0.0589 0.0539 0.0249 0.0215 0.019 0.0139
## Cumulative Proportion 0.658 0.786 0.8451 0.8990 0.9238 0.9453 0.964 0.9782
##                           CAP9   CAP10   CAP11
## Eigenvalue             0.08005 0.06490 0.03698
## Proportion Explained   0.00958 0.00777 0.00443
## Cumulative Proportion 0.98781 0.99557 1.00000
##
```

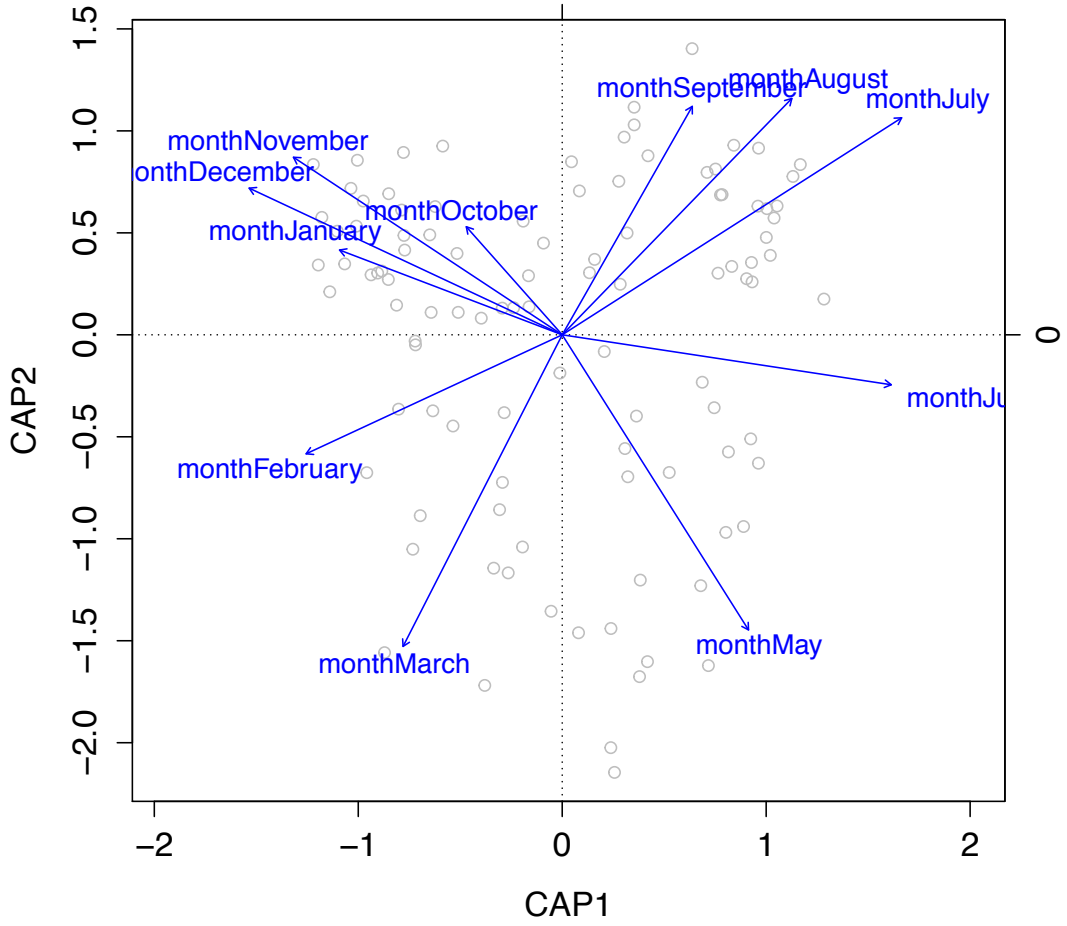
You need 11 axes to code a 12-level factor

Then these axes get rotated to find the axis that explains as much of the response as possible, Etc.

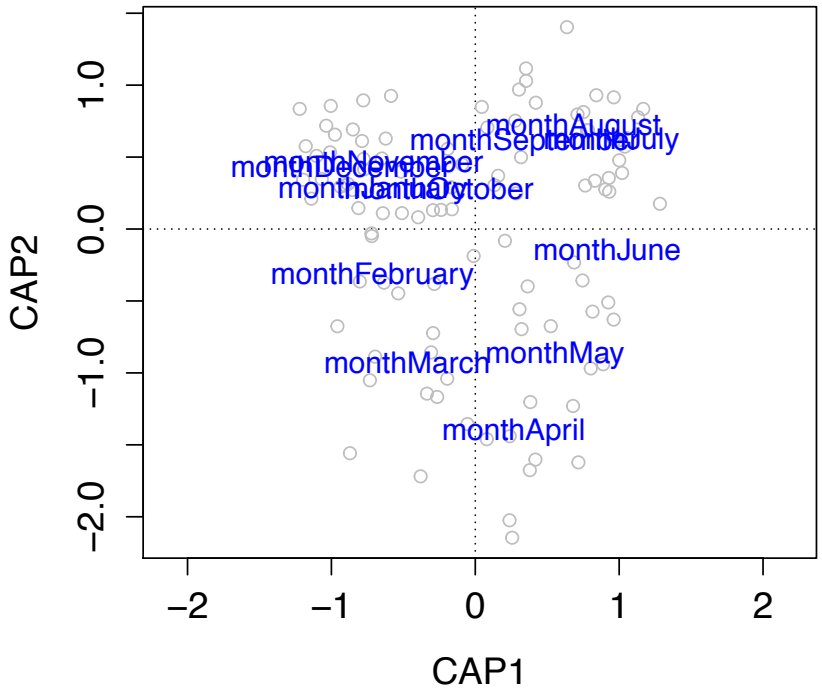
The predictor can only explain so much variation, in total

The ‘residual’ variation in community distance is also ordinated, using PCoA axes.

```r
plot(cap, display = "sites", type = 'n')
points(cap, display = "sites", col = 'grey')
text(cap, display = 'bp', col = 'blue')
```

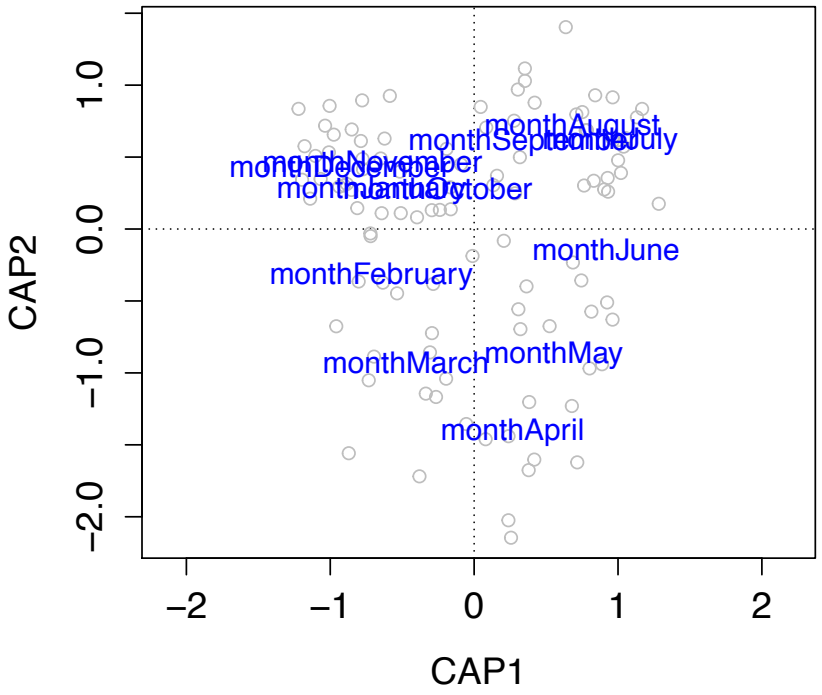


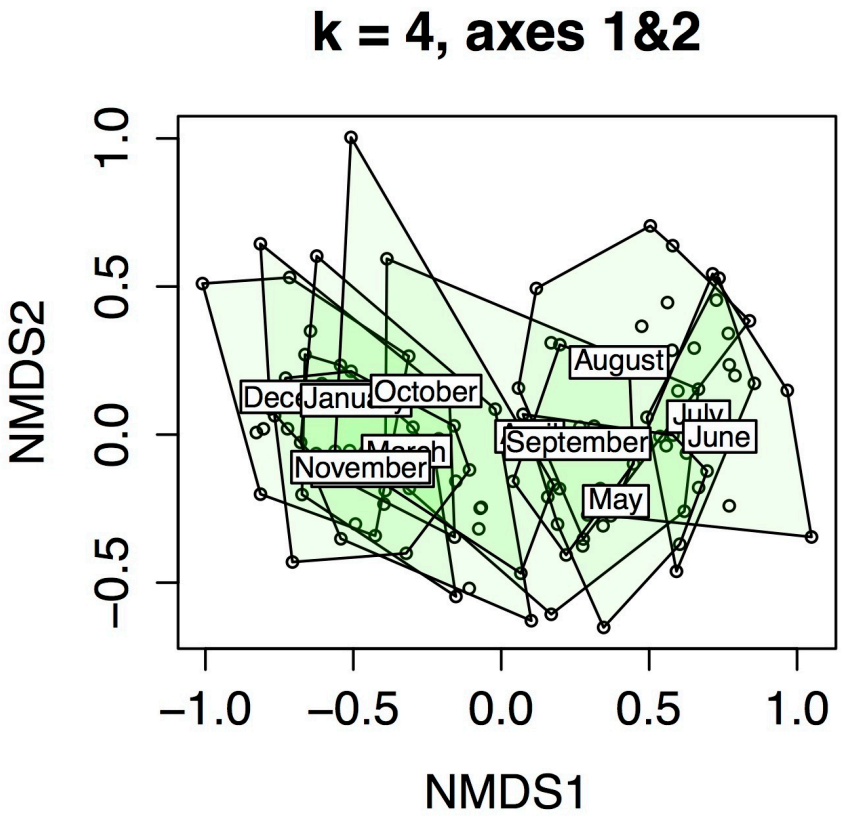
```
plot(cap, display = "sites", type = 'n')
points(cap, display = "sites", col = 'grey')
text(cap, display = 'cn', col = 'blue')
```



Can plot factor levels as **centroids** instead

```r
plot(cap, display = "sites", type = 'n')
points(cap, display = "sites", col = 'grey')
text(cap, display = 'cn', col = 'blue')
```

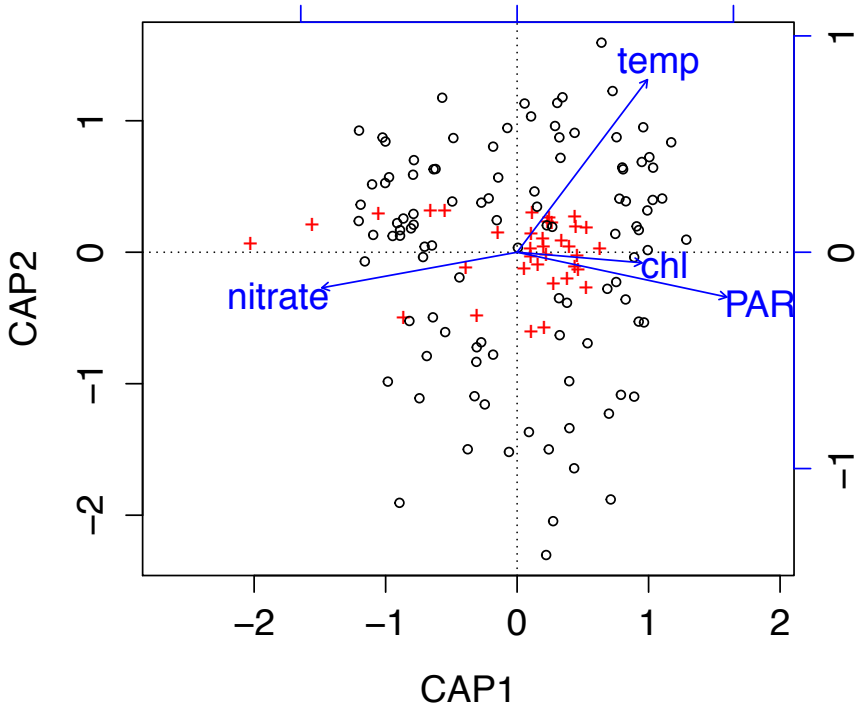




CAP gives clearer patterns by month than NMDS, especially for the first half of the year

This makes sense because it is isolating the variation explained by this predictor

```r
cap = capscale(species.data ~ nitrate + chl + temp + PAR, data = enviro, dist
= "bray")
plot(cap)
```



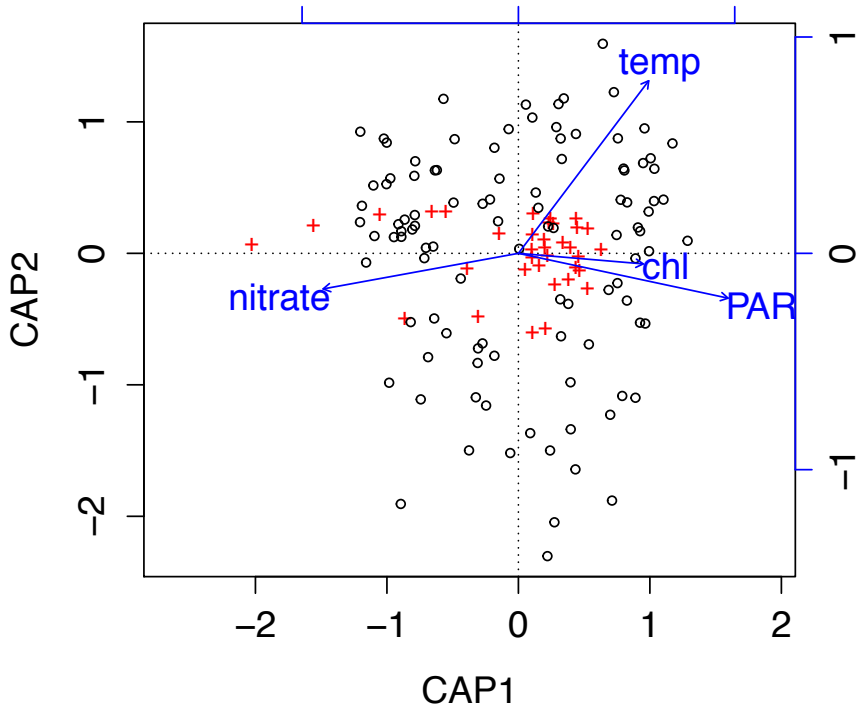
Again, an axis of stratification

Explains 23% of community composition

A little more just due to temperature?

```r
summary(cap)$cont
```

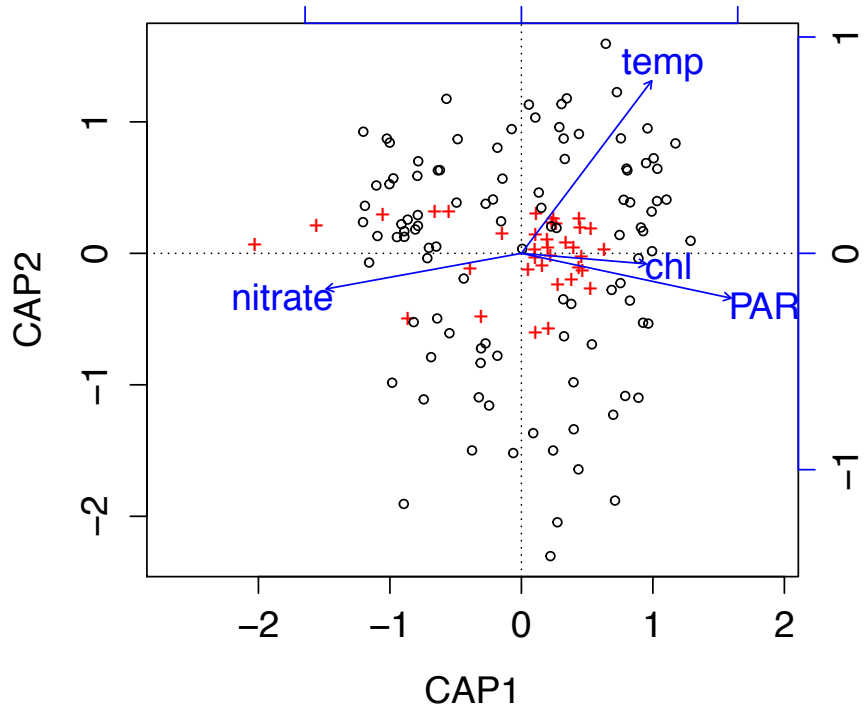
```
## Importance of components:
##                          CAP1   CAP2  CAP3   CAP4  MDS1   MDS2   MDS3   MDS4
## Eigenvalue             5.497 0.9916 0.463 0.2826 1.826 1.6673 1.2979 1.1232
## Proportion Explained   0.226 0.0407 0.019 0.0116 0.075 0.0685 0.0533 0.0461
## Cumulative Proportion 0.226 0.2665 0.285 0.2971 0.372 0.4405 0.4939 0.5400
...
```



Note: CAP is not evaluating the predictors **individually**

It is looking for **linear combinations of predictors** that explain the multivariate response

In a sense, it accounts for the fact that predictors may be correlated




```
summary(cap)$biplot
```

```
##             CAP1     CAP2     CAP3     CAP4 MDS1 MDS2
## nitrate -0.9029 -0.16461  0.20422 -0.34066    0    0
## chl      0.5769 -0.04886 -0.71824 -0.38586    0    0
## temp     0.6023  0.79755  0.01816  0.02903    0    0
## PAR      0.9691 -0.20721  0.12495 -0.04724    0    0
```

Permutation tests

Like PCA, PCoA, NMDS, etc:

Canonical analysis of principal coordinates is not really a **model**

But we still want to know if these associations are 'significant'

Can assess with **permutation tests**:

- Reshuffle the environmental data randomly, to break any association
- Do it 1000 times to get a null distribution for a test statistic – **pseudo-F statistic**

- As with envfit(), permutation tests have **assumptions** that are often broken
- Can include a blocking variable in the permutations (only permute within blocks)

Test the predictors as a whole

```
anova(cap)
```

```
## Permutation test for capscale under reduced model
## Permutation: free
## Number of permutations: 999
##
## Model: capscale(formula = species.data ~ nitrate + chl + temp + PAR, data =
enviro, distance = "bray")
##          Df Variance    F Pr(>F)
## Model     4     7.23 10.5  0.001 ***
## Residual 99    17.12
```

Test the individual axes

```
anova(cap, by = "axis")
```

```
## Permutation test for capscale under reduced model
## Marginal tests for axes
## Permutation: free
## Number of permutations: 999
##
## Model: capscale(formula = species.data ~ nitrate + chl + temp + PAR, data =
 enviro, distance = "bray")
##          Df Variance     F Pr(>F)
## CAP1      1     5.50 31.80  0.001 ***
## CAP2      1     0.99  5.74  0.001 ***
## CAP3      1     0.46  2.68  0.001 ***
## CAP4      1     0.28  1.63  0.025 *
## Residual 99    17.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test the individual variables

```
anova(cap, by = "margin")
```

```
## Permutation test for capscale under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## Model: capscale(formula = species.data ~ nitrate + chl + temp + PAR, data =
enviro, distance = "bray")
##          Df Variance    F Pr(>F)
## nitrate   1     0.35 2.03  0.024 *
## chl       1     0.41 2.39  0.015 *
## temp      1     0.82 4.75  0.001 ***
## PAR       1     1.32 7.62  0.001 ***
## Residual 99    17.12
```

Permanova

Constrained ordination is nice for finding the best predictor axes

But we may want to test specific *a priori* predictors, e.g. an experimental treatment

Classically, multivariate responses analyzed with **MANOVA**

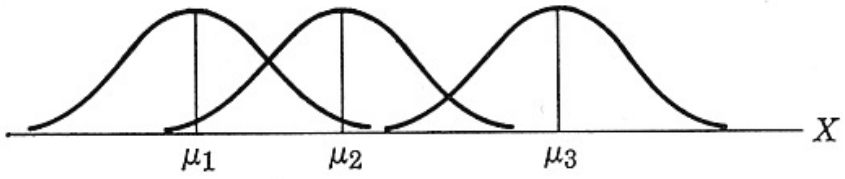


Figure 8.1. The simple anova situation, when the differences among the populations are "real."

source: Cooley & Lohnes ((1971)

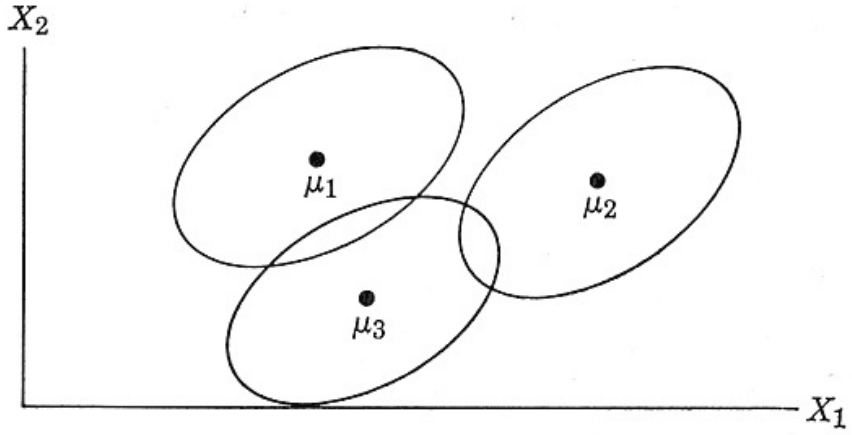


Figure 8.2. The simple manova situation, when the differences among the populations are "real."

Permanova

- Manova has a number of distribution assumptions that are not suitable for community abundance data
- Permanova tries to get around these issues with two innovations

1) Do the analysis on the **dissimilarity matrix**
- i.e., how much of the total dissimilarity is explained by treatment?

2) Test significance with **permutation tests**

In some ways, makes very few assumptions, can be applied to any data for which you can make a dissimilarity matrix to represent multivariate variation

```
adonis2(species.data ~ month + nitrate + chl + temp + PAR, data = enviro, dis
t = 'bray', by = 'margin')

## Permutation test for adonis under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = species.data ~ month + nitrate + chl + temp + PAR, data
= enviro, by = "margin", dist = "bray")
##           Df SumOfSqs      R2      F Pr(>F)
## month     11   1.9061 0.09245 1.3208  0.052 .
## nitrate    1   0.2280 0.01106 1.7377  0.073 .
## chl        1   0.4124 0.02000 3.1437  0.004 **
## temp       1   0.1470 0.00713 1.1205  0.290
## PAR        1   0.1339 0.00650 1.0209  0.376
## Residual  88  11.5452 0.55997
## Total    103  20.6176 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Important: Use adonis2 (not adonis), and set by = 'margin' to test each predictor after the others have been accounted for

In this case, only month, chl are significant, but the predictors are highly correlated, may not want to test simultaneously

The logic of permutation tests

Null hypotheses are about a **lack of a relationship**

E.g. a lack of a difference in the mean of two groups

Or a lack of correlation between two variables

Instead of assuming a particular model of the data:

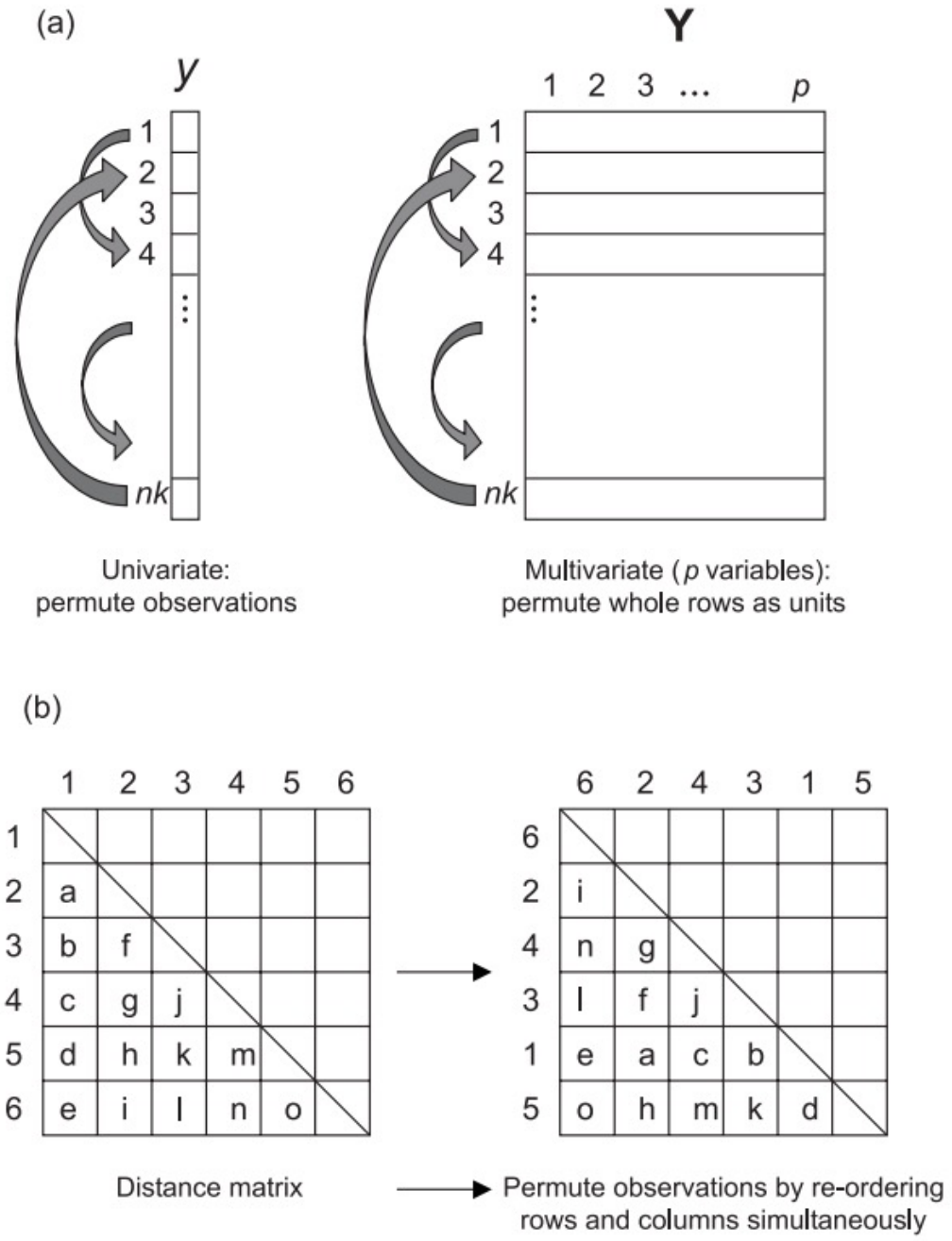
1) **Randomize the data** in a way that simulates the null hypothesis
2) Repeat this many times to create a **null distribution** of a statistic

Advantage: Don't have to assume a particular distribution, or a linear relationship, etc.

Disadvantage:

- Have to be able to construct an appropriate permutation
- Still assuming data are independently distributed, come from the same distribution

The logic of permutation tests



Anderson 2001

```r
adonis2(species.data ~ month + nitrate + chl + temp + PAR, data = enviro, dis
t = 'bray', by = 'margin')

## Permutation test for adonis under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = species.data ~ month + nitrate + chl + temp + PAR, data
= enviro, by = "margin", dist = "bray")
##           Df SumOfSqs      R2      F Pr(>F)
## month     11   1.9061 0.09245 1.3208  0.052 .
## nitrate    1   0.2280 0.01106 1.7377  0.073 .
## chl        1   0.4124 0.02000 3.1437  0.004 **
## temp       1   0.1470 0.00713 1.1205  0.290
## PAR        1   0.1339 0.00650 1.0209  0.376
## Residual  88  11.5452 0.55997
## Total    103  20.6176 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Permanova also has issues

Permutation tests difficult for complex designs, other non-independence in the data

By transforming data and using dissimilarity matrix, may be affecting type I error rate and/or power, compared to modeling the raw data

Permanova also has issues

May confound **location** and **dispersion** effects: test with betadisper()

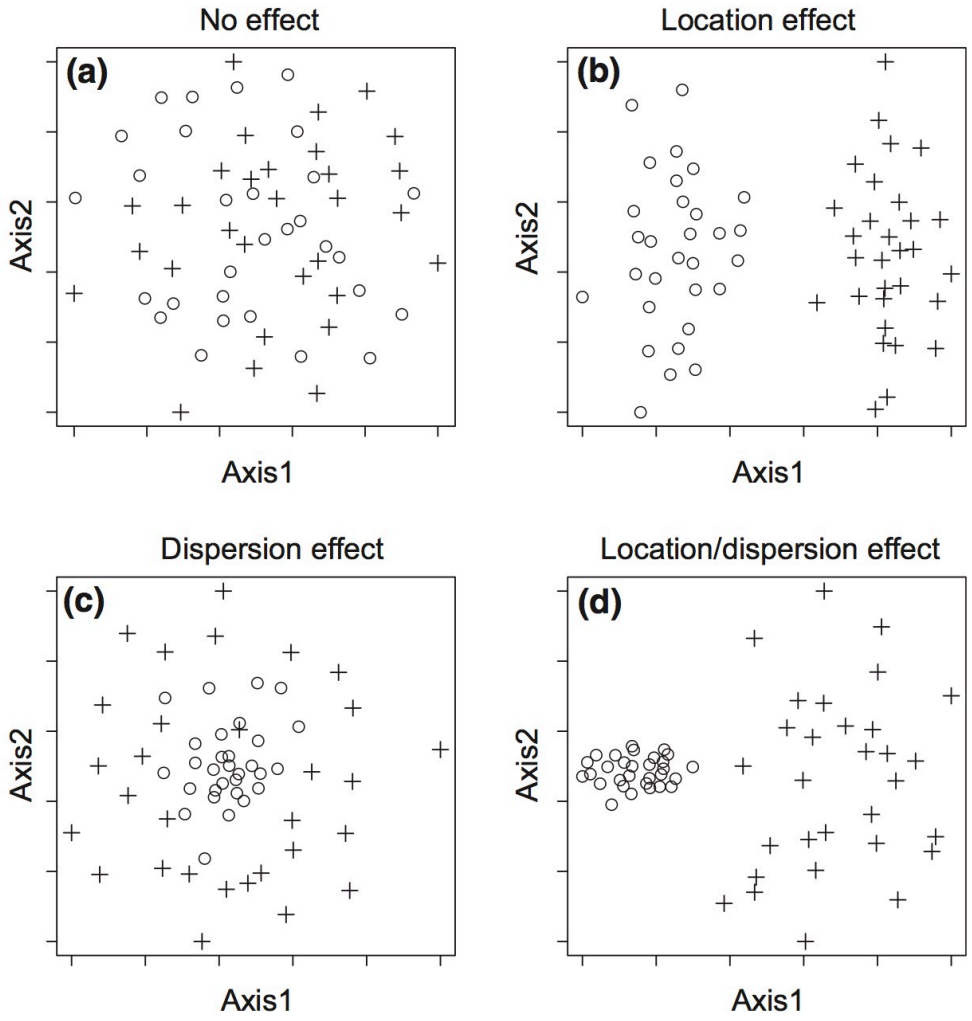


Fig. 2. A schematic diagram as in Anderson *et al.* (2008) illustrating on two axes the types of between-group effects that are often of ecological interest: (a) *no effect*; (b) *location* effect; (c) *dispersion* effect; (d) *location*/*dispersion* effect.