# Characterizing insect communities using multivariate analyses

ENTMLGY 6707 Entomological Techniques and Data Analysis

# Learning objectives

1) Distinguish among different ordination techniques

2) Understand how to apply these methods to your data

3) Interpret the outputs of the ordination analyses

# Traditional taxonomic approach to study insect diversity

Measure unique "types" and compare among samples or communities

Type: taxonomic unit (species, genera, families, orders)

1. **Abundance** measures the number of individuals of each distinct type

2. **Richness** measures the number of distinct types

3. **Evenness** measures the relative abundance of each type

4. **Diversity** measures the number of types and their evenness

# Advantages of multivariate ordination methods

- Analysis of multiple environmental factors on many species simultaneously

- Represents sample and species relationships in a low-dimensional space along (ideally) important and interpretable environmental gradients

- Capable of handling noisy and redundant data

- Accommodates sparse data (i.e., large portion of the entries consist of zeros) because most species are infrequent

# Types of ordination techniques

R package vegan

Indirect gradient analysis (aka unconstrained ordination)

- Utilizes only the species x sample matrix

- Any environmental data are used after the analysis to aid with interpretation

**Nonmetric Multidimensional Scaling and Principal Component Analysis*

# Types of ordination techniques

R package vegan

Indirect gradient analysis (aka unconstrained ordination)

- Utilizes only the species x sample matrix

- Any environmental data are used after the analysis to aid with interpretation

**Nonmetric Multidimensional Scaling and Principal Component Analysis*

Direct gradient analysis (aka constrained ordination)

- Utilizes environmental data in addition to a species x sample matrix

- Assess whether species composition is related to measured environmental data

**Canonical Correspondence Analysis and Redundancy Analysis*

1. Hemlock girdled
2. Hemlock logged
3. Hemlock control
4. Hardwood control

n = 2

90 x 90 m plots


**Methods in Ecology and Evolution**

British Ecological Society

*Methods in Ecology and Evolution* 2010, **1**, 168–179 doi: 10.1111/j.2041-210X.2010.00025.x

## Experimentally testing the role of foundation species in forests: the Harvard Forest Hemlock Removal Experiment

**Aaron M. Ellison*, Audrey A. Barker-Plotkin, David R. Foster and David A. Orwig**

*Harvard Forest, Harvard University, 324 North Main Street, Petersham, MA 01366, USA*


1. Hemlock girdled
2. Hemlock logged
3. Hemlock control
4. Hardwood control

n = 2

90 x 90 m plots


**Environmental data**:
Shrub and herbaceous species in the understory

# Harvard Forest Data Archive

HF106

## Understory Vegetation in Hemlock Removal Experiment at Harvard Forest since 2003

Related Publications

### Data

- hf106-01: species codes (preview)
- hf106-02: o-layer and substrate (preview)
- hf106-03: shrub and herb cover (preview)
- hf106-04: seedling count and cover (preview)
- hf106-05: saplings (preview)
- hf106-06: species list (preview)
- hf106-07: sapling heights (preview)

# Wide format

2014 & 2015

| | year | block | plot | treatment | trap.type | aphful | aphpic | camchr | camher | camnea | camnov | campen | crelin | forarg | forase | fordol |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2006 | Ridge | 4 | Logged | bait | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | |
| 2 | 2006 | Ridge | 4 | Logged | hand | 0 | 2 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | |
| 3 | 2006 | Ridge | 4 | Logged | litter | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | 2006 | Ridge | 4 | Logged | pitfall | 0 | 11 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 5 | 2006 | Ridge | 5 | Girdled | bait | 2 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | |
| 6 | 2006 | Ridge | 5 | Girdled | hand | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 7 | 2006 | Ridge | 5 | Girdled | litter | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | 2006 | Ridge | 5 | Girdled | pitfall | 4 | 19 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | |
| 9 | 2006 | Ridge | 6 | HemlockControl | bait | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 10 | 2006 | Ridge | 6 | HemlockControl | hand | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | 2006 | Ridge | 6 | HemlockControl | litter | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | 2006 | Ridge | 6 | HemlockControl | pitfall | 1 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | 2006 | Ridge | 7 | HardwoodControl | bait | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | |
| 14 | 2006 | Ridge | 7 | HardwoodControl | hand | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 15 | 2006 | Ridge | 7 | HardwoodControl | litter | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 16 | 2006 | Ridge | 7 | HardwoodControl | pitfall | 0 | 18 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | |
| 17 | 2006 | Valley | 1 | Girdled | bait | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 18 | 2006 | Valley | 1 | Girdled | hand | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 19 | 2006 | Valley | 1 | Girdled | litter | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 20 | 2006 | Valley | 1 | Girdled | pitfall | 1 | 10 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | |
| 21 | 2006 | Valley | 2 | Logged | bait | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 22 | 2006 | Valley | 2 | Logged | hand | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | |
| 23 | 2006 | Valley | 2 | Logged | pitfall | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 24 | 2006 | Valley | 3 | HemlockControl | bait | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 25 | 2006 | Valley | 3 | HemlockControl | hand | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 26 | 2006 | Valley | 3 | HemlockControl | pitfall | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 27 | 2006 | Valley | 8 | HardwoodControl | bait | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 28 | 2006 | Valley | 8 | HardwoodControl | hand | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 29 | 2006 | Valley | 8 | HardwoodControl | litter | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

# Nonmetric Multidimensional Scaling (NMDS)

*Used to assess differences in species composition among sites, treatments, etc.*

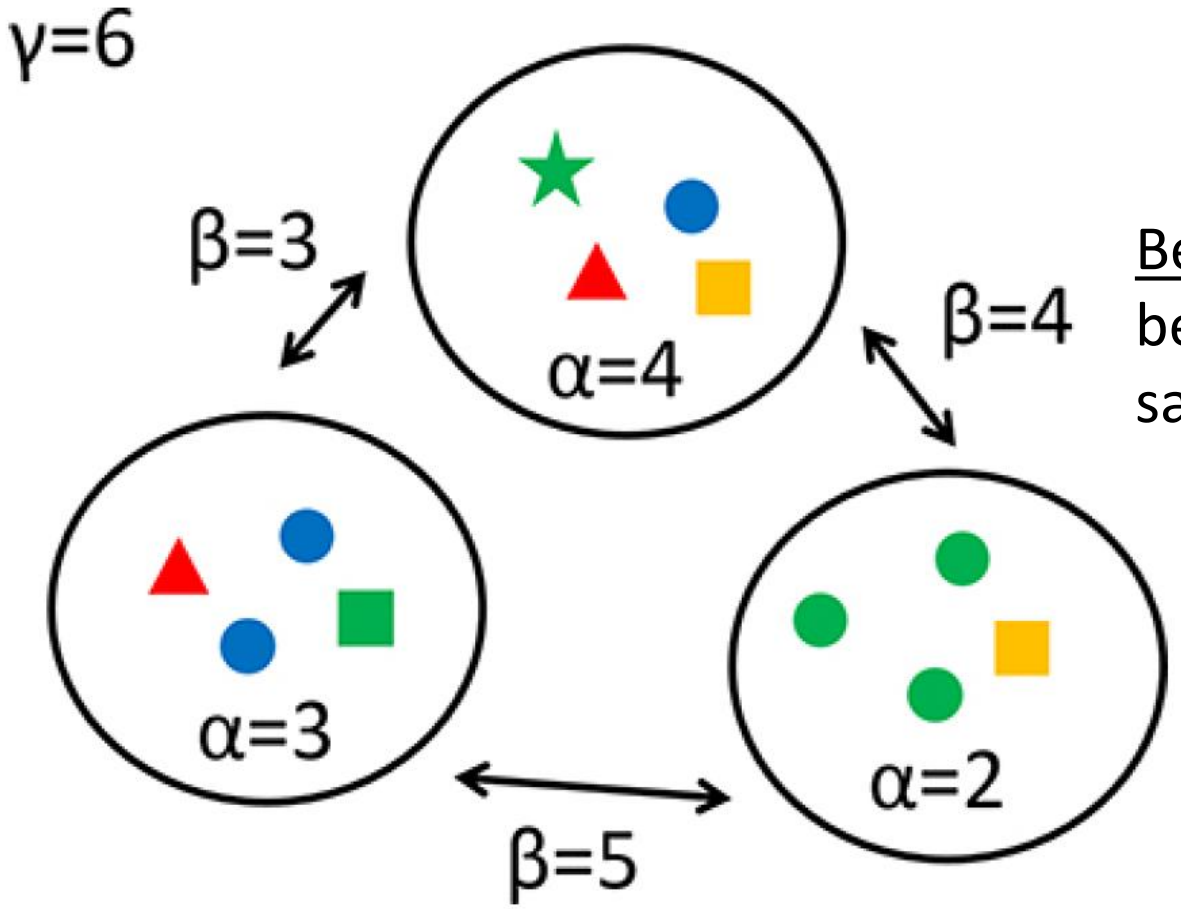
Represents (as well as possible) the ordering relationships among sites in species space

Based on a distance (dissimilarity) matrix as a measure of beta-diversity

Iterative method that maximizes the rank order correlations between the distances in the dissimilarity matrix and the distances in low-dimensional space

Robust technique – no assumptions of normality or linear relationships among variables.

Gamma diversity ($\gamma$) = total number of species across all habitats within a landscape




Beta diversity ($\beta$) = diversity between habitat patches, samples, or sites.

Alpha diversity ($\alpha$) = diversity within a particular habitat patch, sample, or site.

**Figure 5.** Illustration of the concept of alpha, beta, and gamma diversity. Colored symbols represent species, circles represent habitat patches, and the large rectangle represents the landscape.

Sirami 2016

# Beta diversity metrics

Beta diversity ($\beta$) = diversity between habitat patches, samples, or sites.

Introduced by R.H. Whittaker in 1960

- "The extent of change in community composition, or degree of community differentiation, in relation to a complex-gradient of environment"

$$\beta_W = (\gamma - \alpha)/\alpha$$

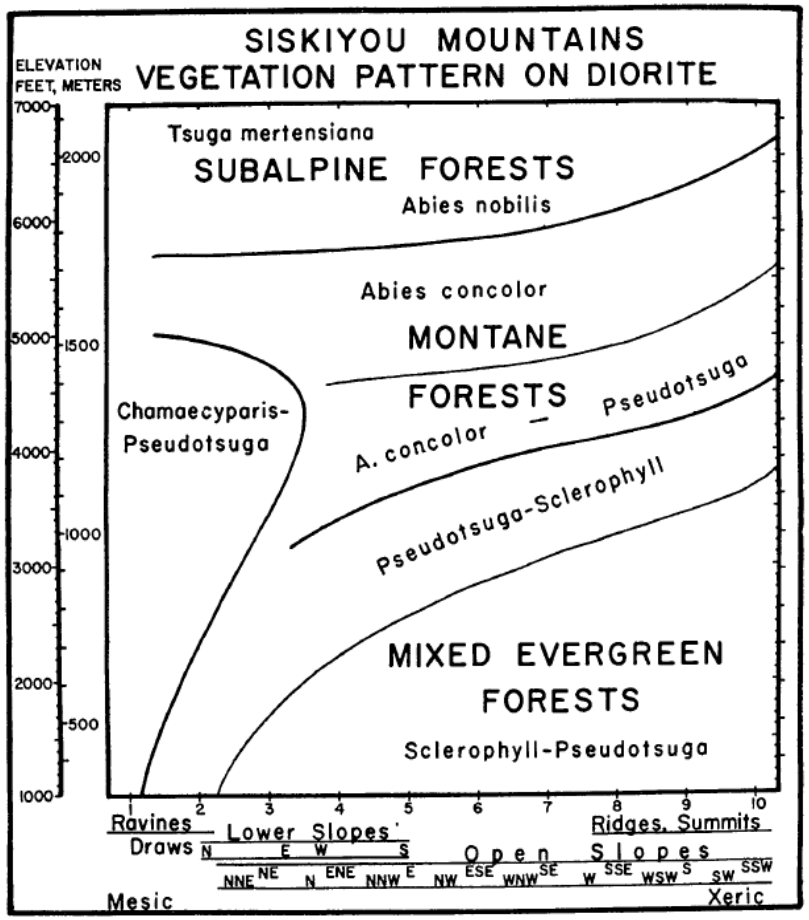



FIG. 11. Mosaic chart of vegetation on quartz diorite, central Siskiyou Mountains, Oregon.

Whittaker, R.H., 1960. Vegetation of the Siskiyou mountains, Oregon and California. *Ecological monographs*, *30*(3).

# Measuring the compositional (dis)similarity among communities

Incidence-based or abundance-based dissimilarity matrix

- Sorensen (Bray-Curtis)
- Jaccard
- Euclidean

Create a matrix that compares the number of shared species to the number of unique species among two sites (pairwise)

Whittaker formula: $\beta_W = (b + c) / (2 * a + b + c)$

a = number of species shared between two samples
b = number of species unique to sample 1
c = number of species unique to sample 2

# Jaccard dissimilarity matrix

vegdist()

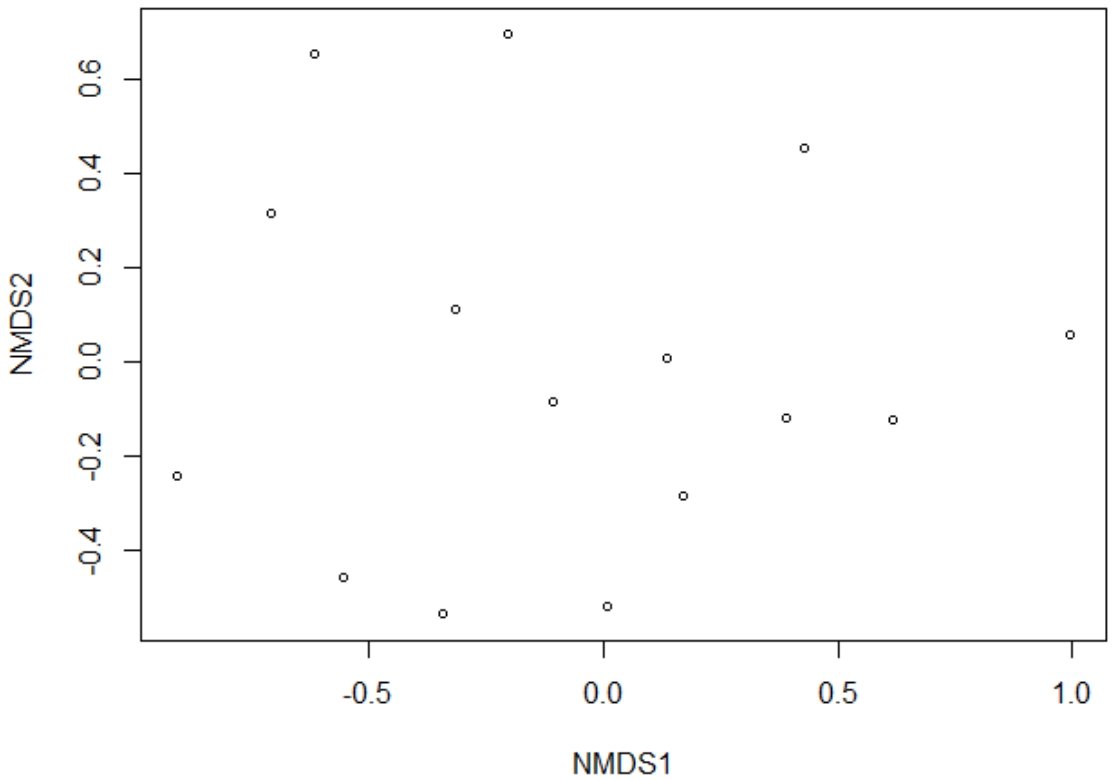
```
> dis.matrix.pa <- vegdist(ants4[,1:30], method = "jaccard")
> dis.matrix.pa
           1         2         3         4         5         6         7         8         9        10        11        12        13        14        15
2  0.6000000
3  0.8888889 0.6666667
4  0.2500000 0.5000000 0.8571429
5  0.7000000 0.6250000 0.6000000 0.7777778
6  0.5714286 0.7142857 0.9230769 0.6153846 0.6923077
7  0.6666667 0.3333333 0.5000000 0.5714286 0.5000000 0.7692308
8  0.5000000 0.5454545 0.8000000 0.5454545 0.5000000 0.6250000 0.6000000
9  0.6842105 0.7222222 0.8823529 0.7222222 0.7058824 0.6818182 0.7647059 0.5789474
10 0.7142857 0.5454545 0.8000000 0.6666667 0.8461538 0.8421053 0.7272727 0.6666667 0.5789474
11 0.8888889 0.6666667 0.0000000 0.8571429 0.6000000 0.9230769 0.5000000 0.8000000 0.8823529 0.8000000
12 0.5555556 0.4285714 0.6000000 0.4285714 0.7500000 0.7857143 0.5000000 0.5000000 0.7058824 0.5000000 0.6000000
13 0.6363636 0.5555556 0.7142857 0.5555556 0.6666667 0.6428571 0.6250000 0.5833333 0.7368421 0.6923077 0.7142857 0.5000000
14 0.6666667 0.6250000 0.8750000 0.7058824 0.6875000 0.5263158 0.7500000 0.5555556 0.5652174 0.5555556 0.8750000 0.6875000 0.7222222
15 0.7777778 0.5000000 0.3333333 0.7142857 0.6666667 0.8461538 0.2500000 0.7000000 0.8235294 0.7000000 0.3333333 0.4000000 0.5714286 0.8125000
16 0.6363636 0.3750000 0.7142857 0.5555556 0.5000000 0.5384615 0.4285714 0.4545455 0.5882353 0.6923077 0.7142857 0.5000000 0.4444444 0.5625000 0.5714286
> 
```

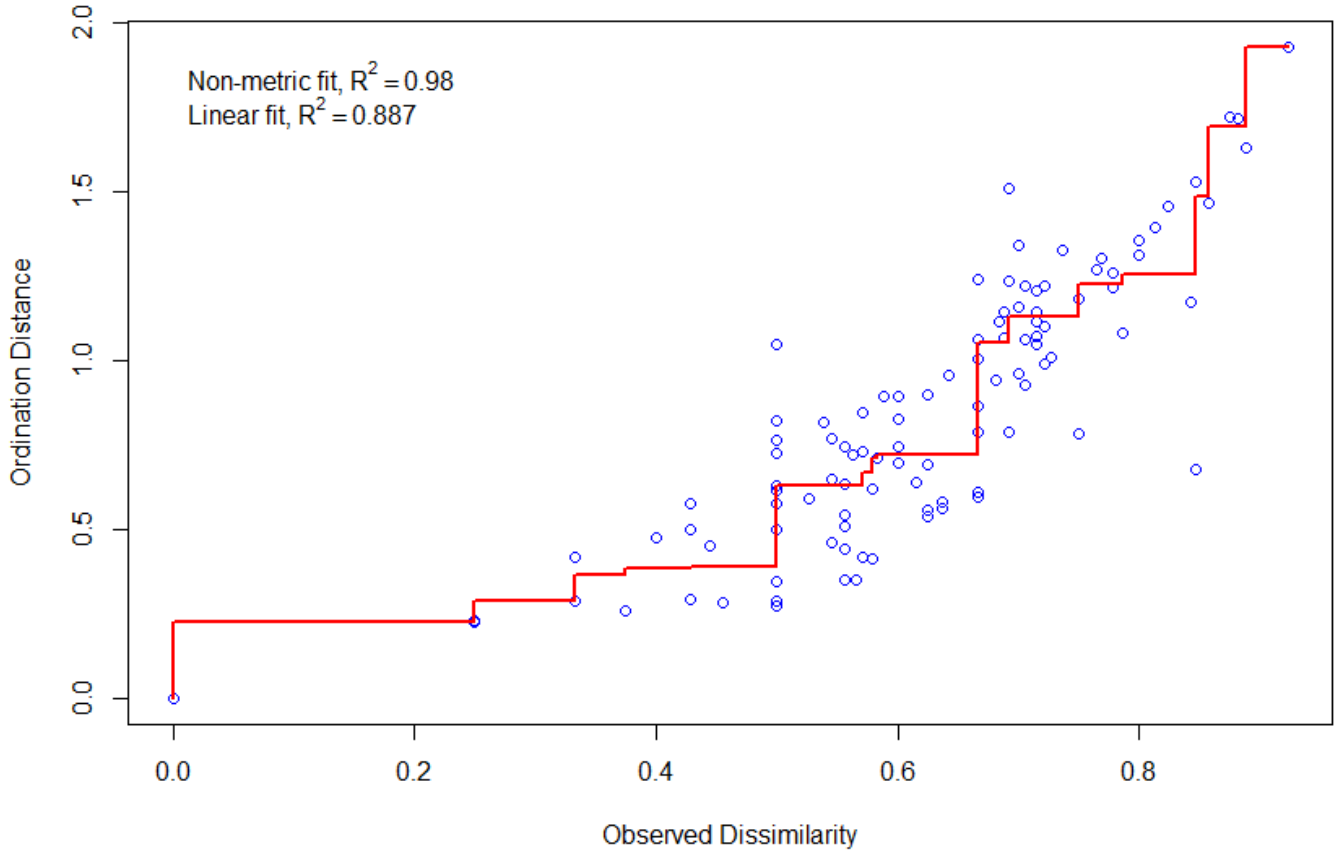
# Run a non-metric multidimensional scaling (NMDS) model

```
> nmds.ants.pa <- metaMDS(dis.matrix.pa, trymax = 500, autotransform = TRUE, k = 2)
Run 0 stress 0.1459161
Run 1 stress 0.1401423
... New best solution
... Procrustes: rmse 0.1568134  max resid 0.3461763
Run 2 stress 0.1401423
... New best solution
... Procrustes: rmse 7.163306e-06  max resid 1.843059e-05
... Similar to previous best
Run 3 stress 0.1401423
... Procrustes: rmse 3.603234e-06  max resid 9.665341e-06
... Similar to previous best
Run 4 stress 0.1543155
Run 5 stress 0.1418973
Run 6 stress 0.1454809
Run 7 stress 0.1401423
... New best solution
... Procrustes: rmse 4.744046e-06  max resid 1.131643e-05
... Similar to previous best
Run 8 stress 0.1543156
Run 9 stress 0.1854143
Run 10 stress 0.1513608
Run 11 stress 0.1543156
Run 12 stress 0.145916
Run 13 stress 0.1510867
Run 14 stress 0.1401423
... New best solution
... Procrustes: rmse 5.499765e-06  max resid 1.425497e-05
... Similar to previous best
Run 15 stress 0.1401423
... Procrustes: rmse 1.779235e-06  max resid 3.274561e-06
... Similar to previous best
Run 16 stress 0.1560107
Run 17 stress 0.1560108
Run 18 stress 0.1401423
... Procrustes: rmse 1.11992e-05  max resid 2.74026e-05
... Similar to previous best
Run 19 stress 0.1459161
Run 20 stress 0.1459162
*** Best solution repeated 3 times
```

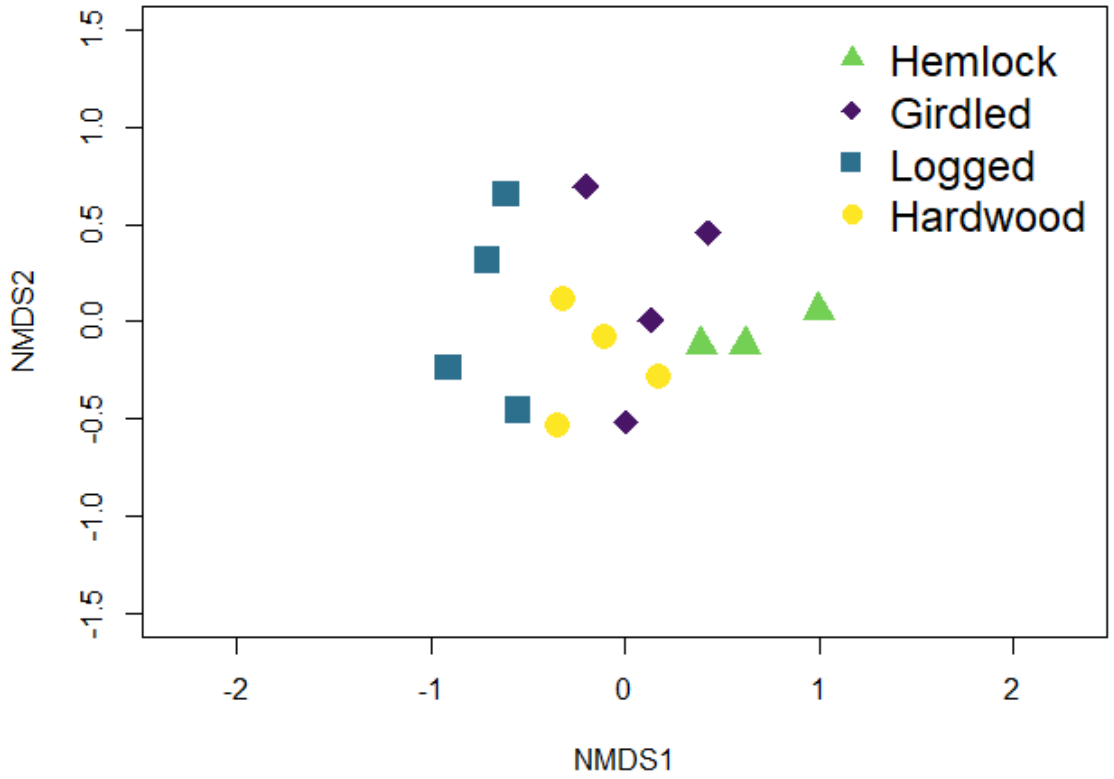
```
> nmds.ants.pa # stress is quality of fit

Call:
metaMDS(comm = dis.matrix.pa, k = 2, trymax = 500, autotransform = TRUE)

global Multidimensional Scaling using monoMDS

Data:     dis.matrix.pa
Distance: jaccard

Dimensions: 2
Stress:     0.1401423
Stress type 1, weak ties
Best solution was repeated 3 times in 20 tries
The best solution was from try 14 (random start)
Scaling: centring, PC rotation, halfchange scaling
Species: scores missing
```

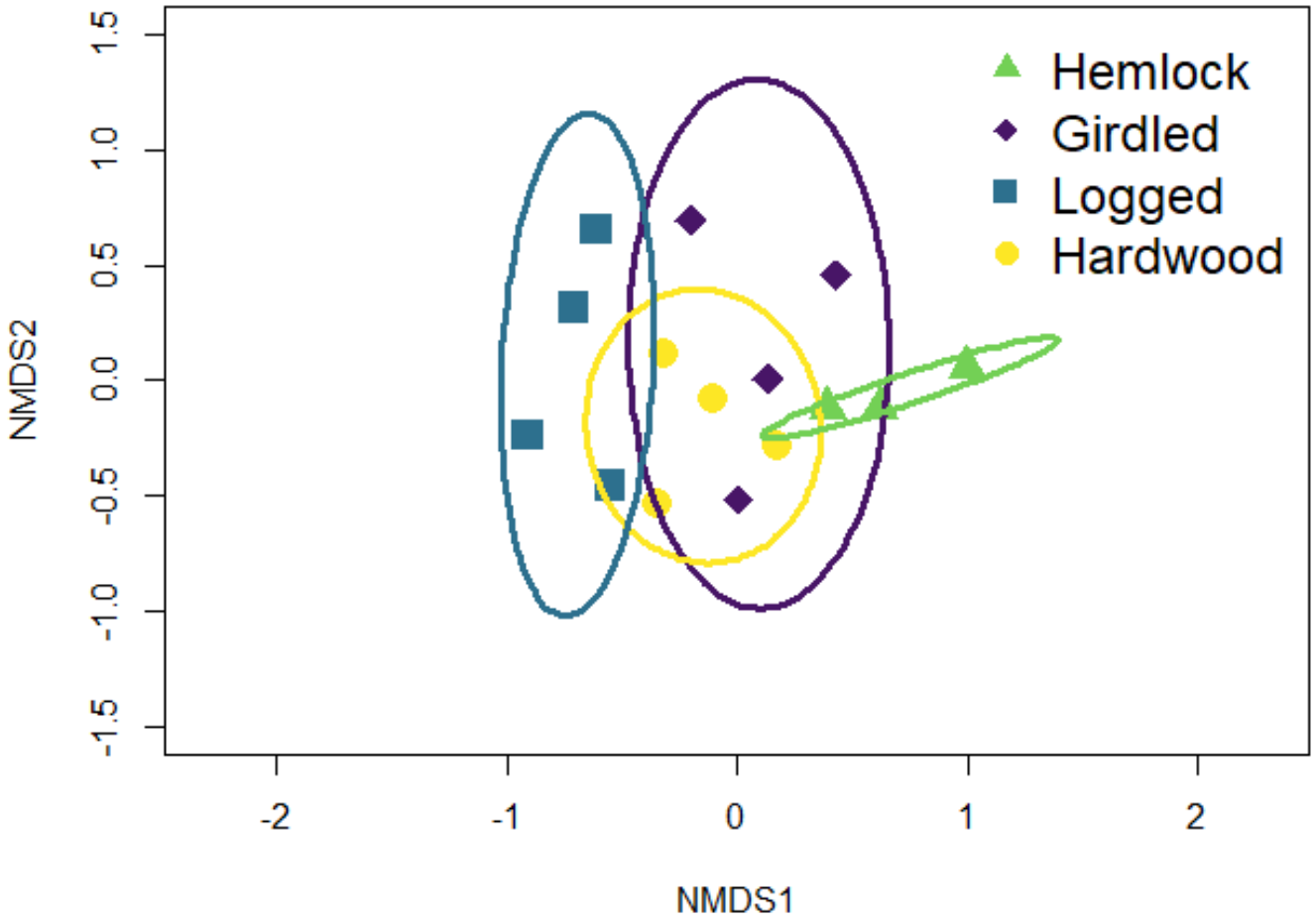
# Run a non-metric multidimensional scaling (NMDS) model



# Visualize with a non-metric multidimensional scaling (NMDS) plot



# NMDS often paired with PERMANOVA and BETADISPER

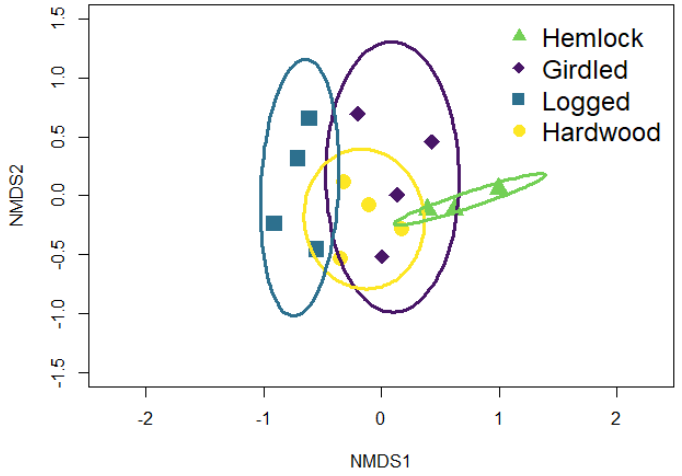
**Permutational multivariate analysis of variance (PERMANOVA)**

- tests whether the group centroid of communities differs among a categorical grouping factor in multivariate space

**Homogeneity of multivariate group dispersion (BETADISPER)**

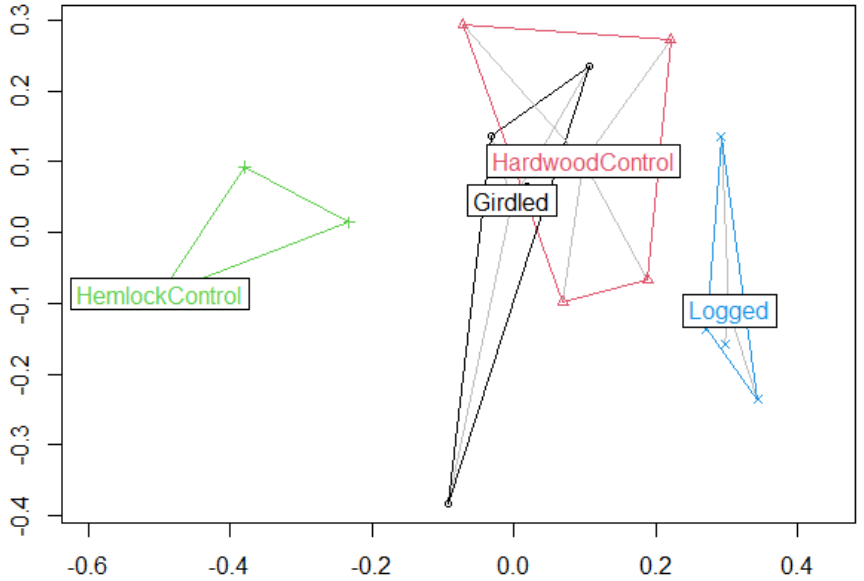
- tests whether the dispersion of a categorical grouping factor from its spatial medial is different between groups.
- Multivariate analogue of Levene's test for homogeneity of variances

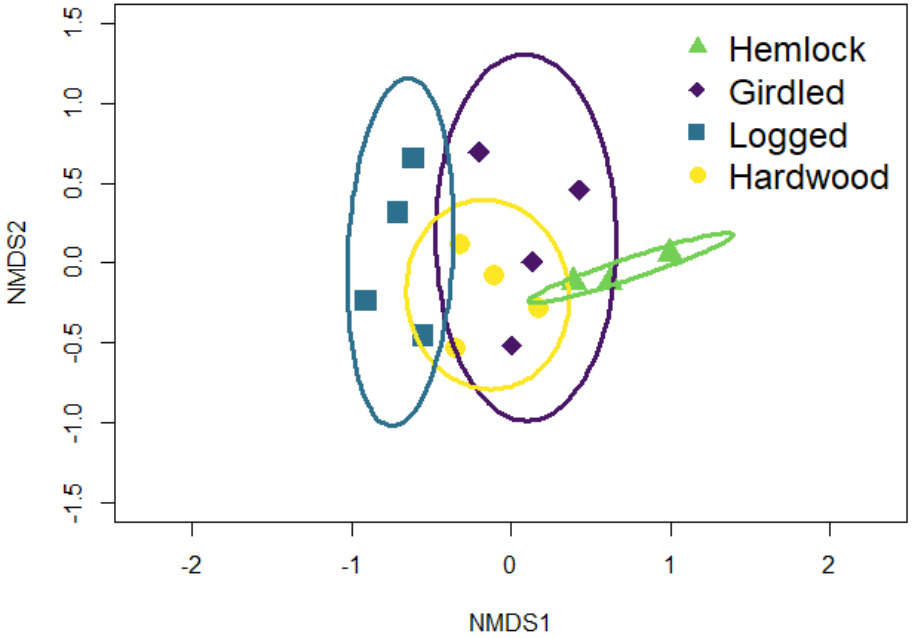
# PERMANOVA



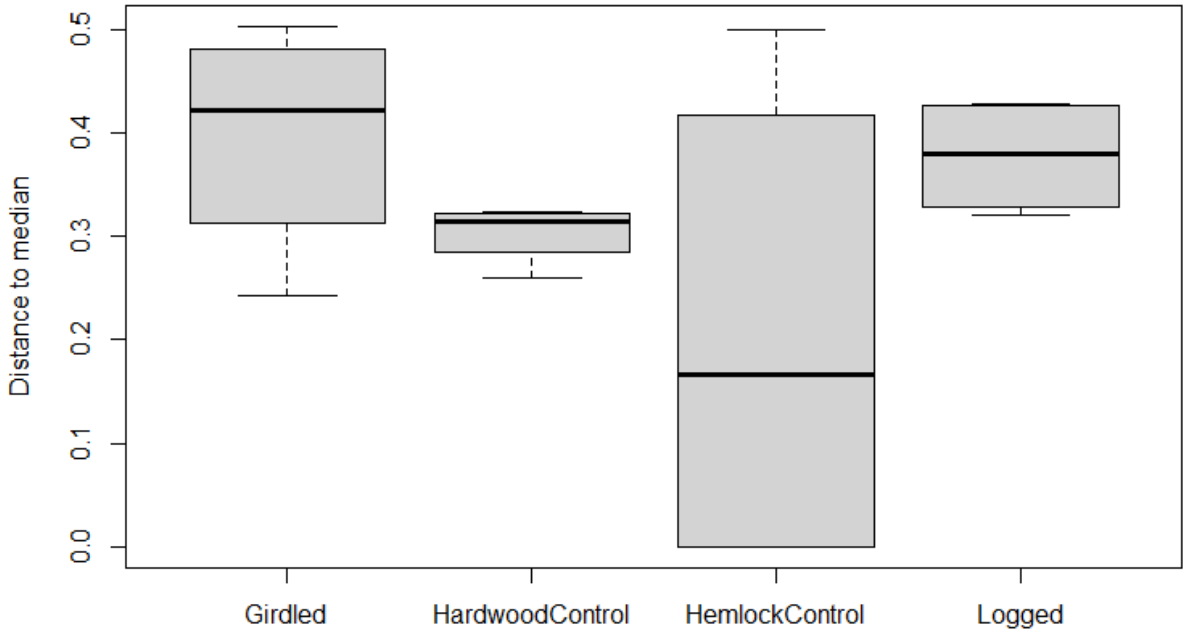

```
> adonis2(dis.matrix.pa ~ ants4$treatment, permutations = 999)
Permutation test for adonis under reduced model
Terms added sequentially (first to last)
Permutation: free
Number of permutations: 999

adonis2(formula = dis.matrix.pa ~ ants4$treatment, permutations = 999)
                Df SumOfSqs      R2      F Pr(>F)
ants4$treatment  3   1.3801 0.43335 3.0591  0.001 ***
Residual        12   1.8045 0.56665
Total           15   3.1846 1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> pairwise.adonis(dis.matrix.pa, ants4$treatment)
                          pairs Df SumsOfSqs   F.Model        R2 p.value p.adjusted sig
1               Logged vs Girdled  1 0.3164513 1.5383579 0.2040707   0.157      0.942
2         Logged vs HemlockControl  1 1.0162804 7.9057299 0.5685232   0.032      0.192
3       Logged vs HardwoodControl  1 0.2753550 1.7404940 0.2248557   0.117      0.702
4        Girdled vs HemlockControl  1 0.4099642 2.8759151 0.3240134   0.018      0.108
5      Girdled vs HardwoodControl  1 0.1401176 0.8136623 0.1194163   0.648      1.000
6 HemlockControl vs HardwoodControl  1 0.6019456 6.3330199 0.5135012   0.025      0.150
>
```

# BETADISPER




```
> anova(ants.beta.pa)
Analysis of Variance Table

Response: Distances
           Df   Sum Sq  Mean Sq F value Pr(>F)
Groups      3 0.087628 0.029209  1.4654 0.2733
Residuals 12 0.239198 0.019933
> TukeyHSD(ants.beta.pa, which = "group", conf.level = 0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = distances ~ group, data = df)

$group
                                  diff        lwr       upr     p adj
HardwoodControl-Girdled      -0.09392745 -0.3903214 0.2024665 0.7840298
HemlockControl-Girdled       -0.18850468 -0.4848986 0.1078893 0.2831197
Logged-Girdled               -0.01984796 -0.3162419 0.2765460 0.9970601
HemlockControl-HardwoodControl -0.09457723 -0.3909712 0.2018167 0.7806065
Logged-HardwoodControl        0.07407950 -0.2223145 0.3704734 0.8782710
Logged-HemlockControl         0.16865672 -0.1277372 0.4650507 0.3703064
```



# Principal Component Analysis (PCA)

*Used to assess relationships among variables along a reduced number of axes*

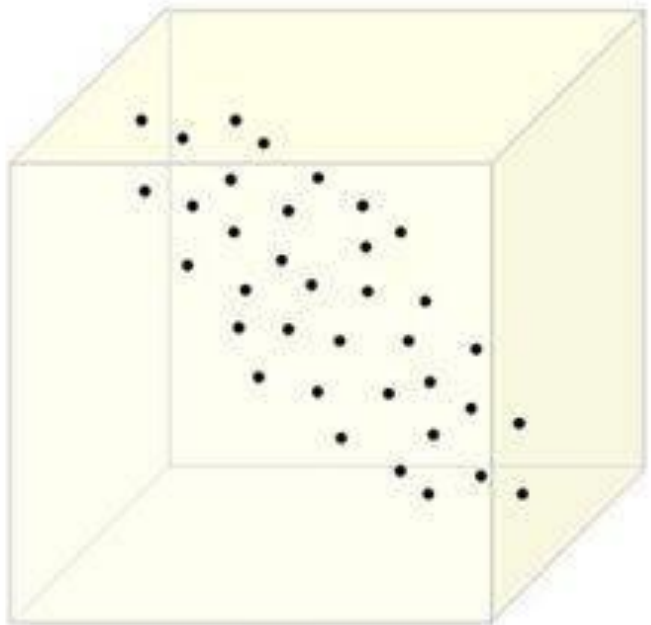
Eigenvector-based method

Eigen decomposition of a dispersion matrix (linear covariances or correlations)

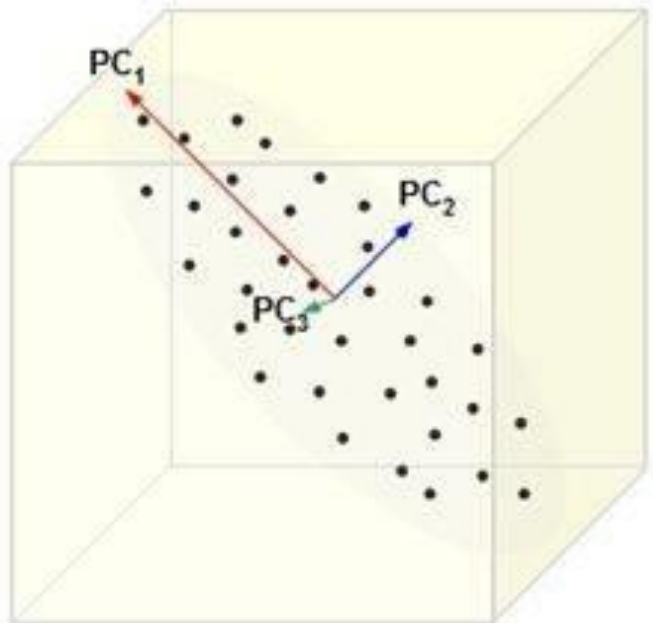
Requires raw, quantitative data

Descriptive / exploratory

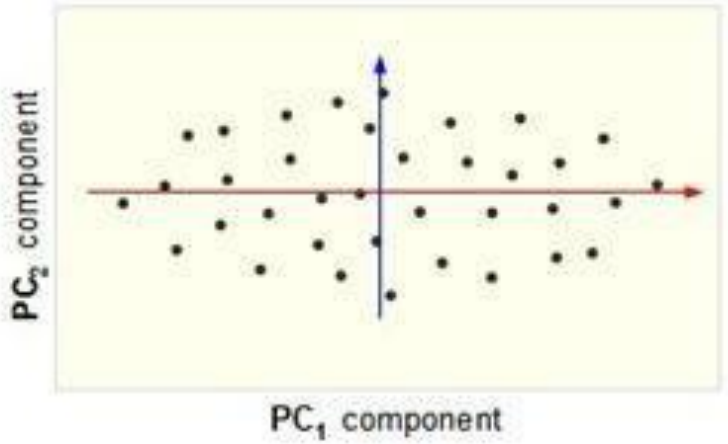
# Principal Component Analysis (PCA)



a




b




c

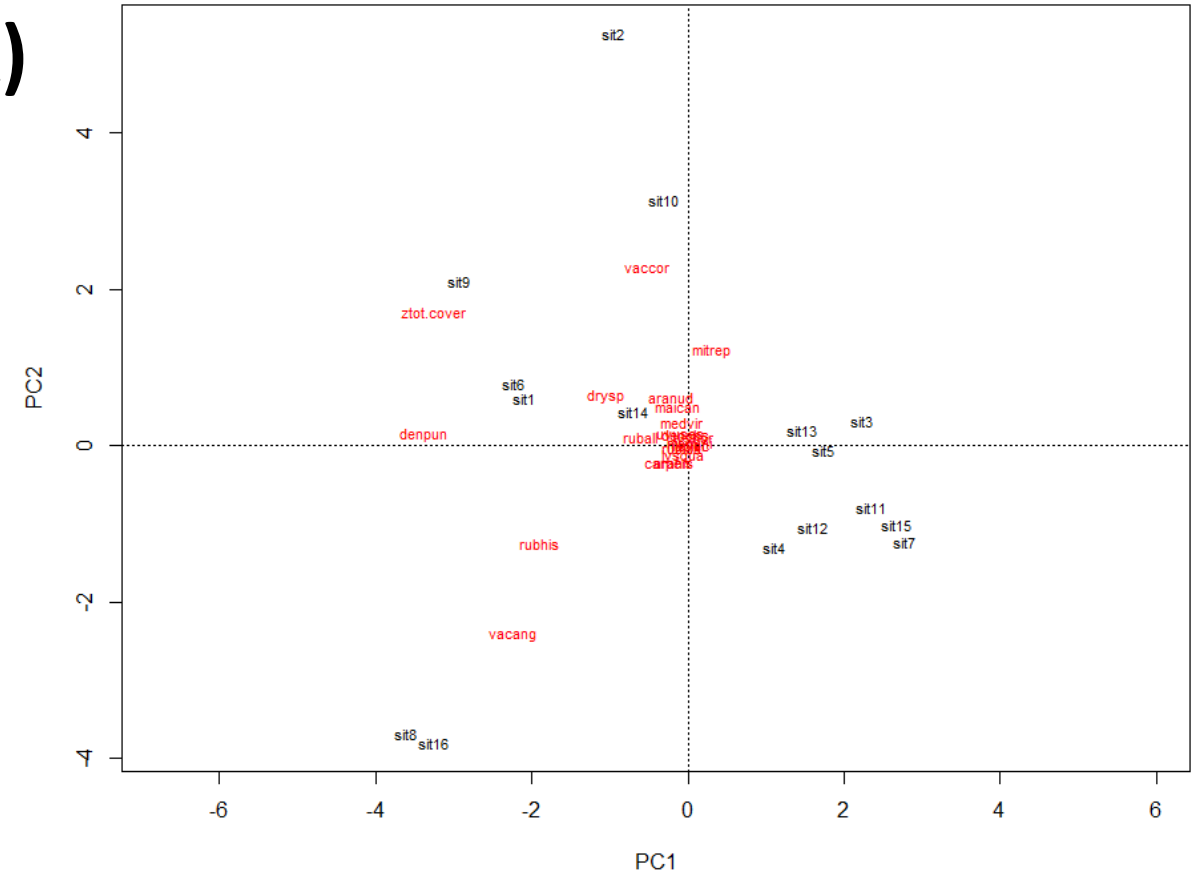
# Understory shrub and herbaceous vegetation

Wide format

2014 & 2015

| | year | block | trt | plot | amesp | anequi | arahis | aranud | aritri | berthu | braari | carpen | carsp |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2003 | ridge | girdled | 5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 2003 | ridge | hardwood | 7 | 0.0 | 0.0 | 0.0 | 52.0 | 0.0 | 0.0 | 0.0 | 6.5 | 0.0 |
| 3 | 2003 | ridge | hemlock | 6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | 2003 | ridge | logged | 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | 2003 | valley | girdled | 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 6 | 2003 | valley | hardwood | 8 | 0.0 | 0.0 | 0.0 | 22.0 | 0.0 | 0.0 | 0.0 | 18.0 | 0.0 |
| 7 | 2003 | valley | hemlock | 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 8 | 2003 | valley | logged | 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 9 | 2004 | ridge | girdled | 5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 10 | 2004 | ridge | hardwood | 7 | 0.0 | 0.0 | 0.0 | 61.0 | 0.0 | 0.0 | 0.0 | 4.0 | 0.0 |
| 11 | 2004 | ridge | hemlock | 6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 12 | 2004 | ridge | logged | 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 13 | 2004 | valley | girdled | 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 14 | 2004 | valley | hardwood | 8 | 0.0 | 0.5 | 0.0 | 35.0 | 0.0 | 0.0 | 0.0 | 15.0 | 0.0 |
| 15 | 2004 | valley | hemlock | 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 16 | 2004 | valley | logged | 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 17 | 2005 | ridge | girdled | 5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 18 | 2005 | ridge | hardwood | 7 | 0.0 | 0.0 | 0.0 | 34.5 | 0.0 | 0.0 | 0.0 | 4.5 | 1.0 |
| 19 | 2005 | ridge | hemlock | 6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

# Principal Component Analysis (PCA)

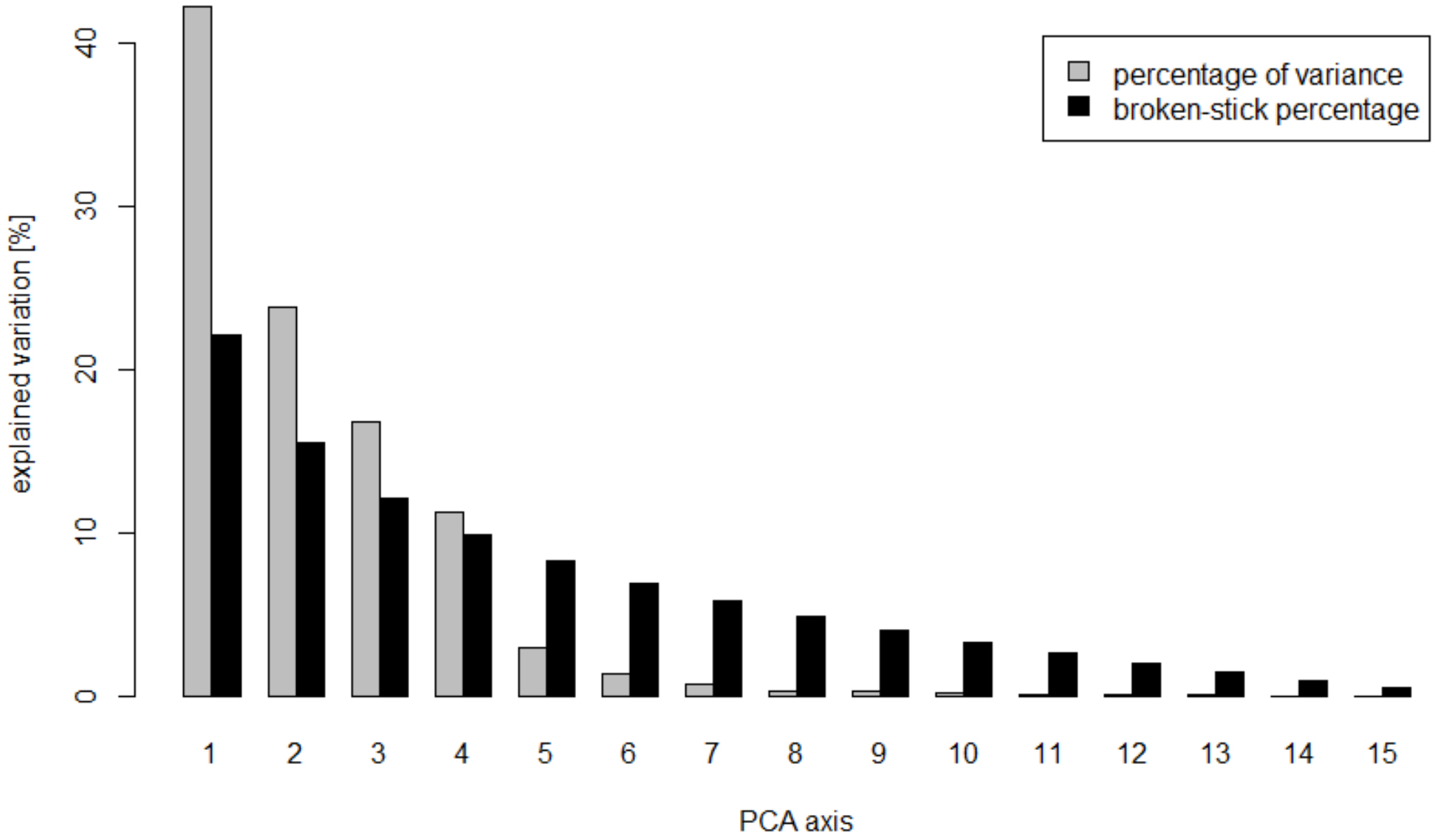



```
> herb.pca <- rda(herb3[,5:25], scale = FALSE)
> summary(herb.pca)

Call:
rda(X = herb3[, 5:25], scale = FALSE)

Partitioning of variance:
              Inertia Proportion
Total           399.4          1
Unconstrained   399.4          1

Eigenvalues, and their contribution to the variance

Importance of components:
                          PC1     PC2     PC3     PC4      PC5     PC6     PC7     PC8     PC9     PC10     PC11     PC12     PC13     PC14     PC15
Eigenvalue           168.4820 95.2045 67.1849 45.0949 11.84576 5.37022 2.68426 1.22225 1.073345 0.510973 0.404821 0.2155915 0.1125490 0.0289611 5.164e-04
Proportion Explained   0.4218  0.2383  0.1682  0.1129  0.02966 0.01344 0.00672 0.00306 0.002687 0.001279 0.001013 0.0005397 0.0002818 0.0000725 1.293e-06
Cumulative Proportion  0.4218  0.6601  0.8283  0.9412  0.97090 0.98434 0.99106 0.99412 0.996812 0.998091 0.999105 0.9996444 0.9999262 0.9999987 1.000e+00

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:  8.798011
```

**Total variance**

**Axes 1 & 2 represent 66% of the variation**

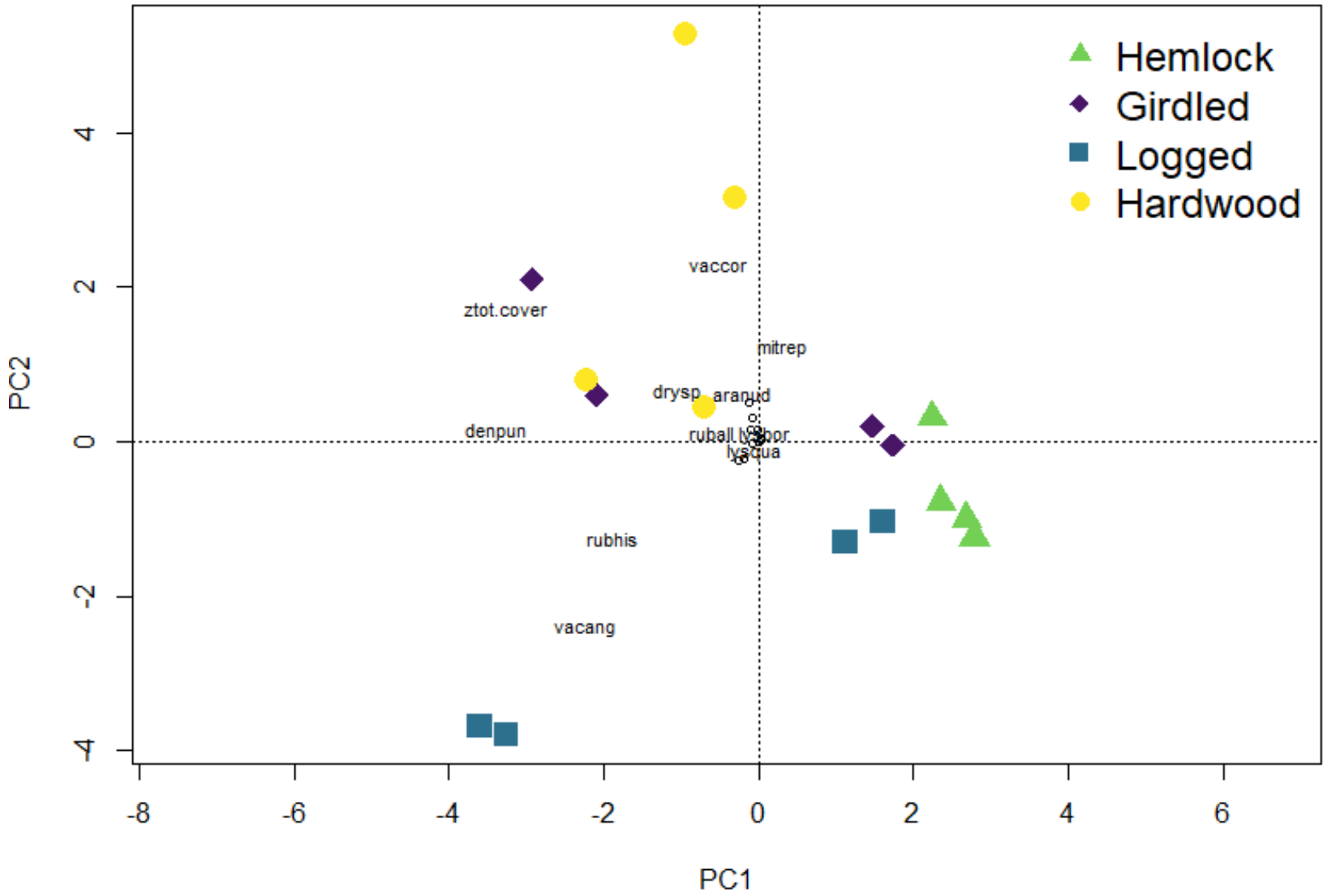
# Principal Component Analysis (PCA) - Broken Stick Method



# Principal Component Analysis (PCA)

Axis 1: 42.1%

Axis 2: 23.8%

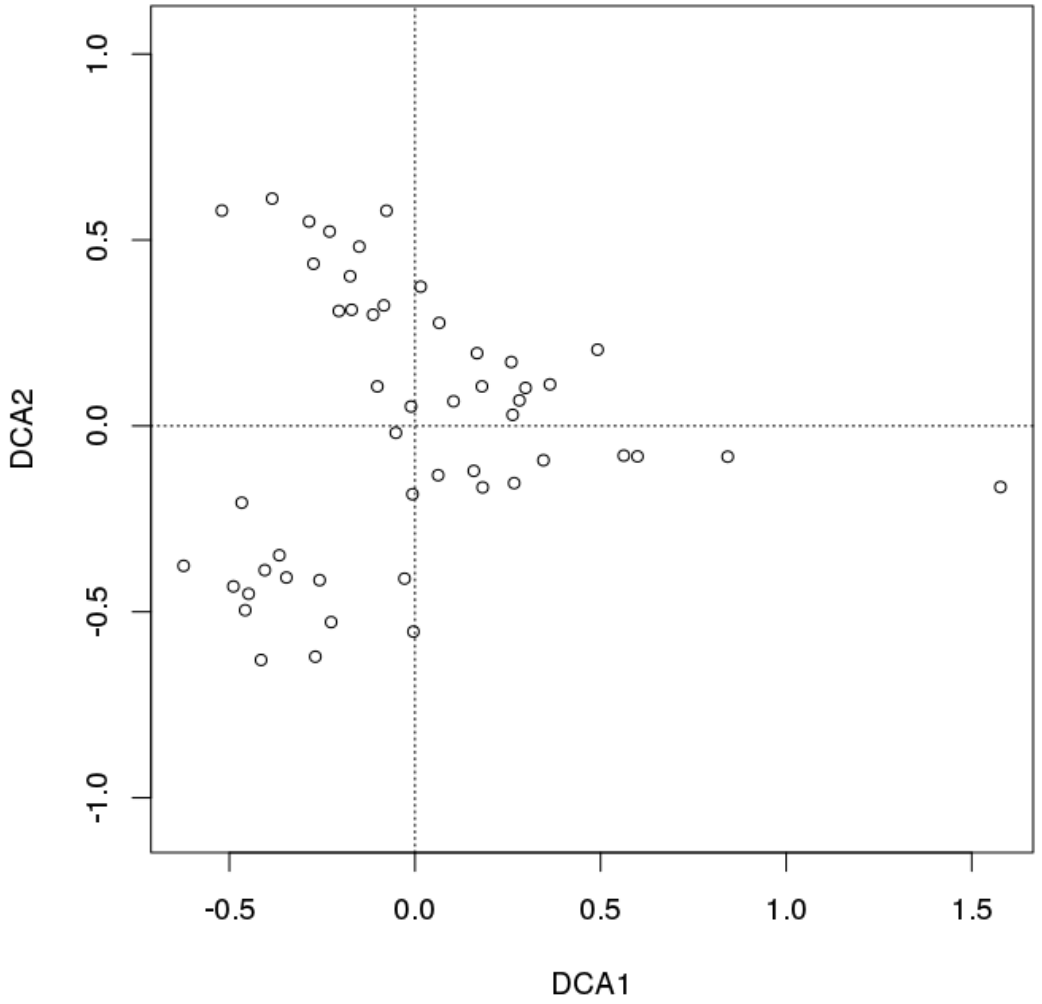


# Principal Component Analysis (PCA)

Arch or horseshoe effect – distortion in ordination diagram

1) Use sample scores on PCA axes as environmental predictor variables

2) Identify environmental variables highly correlated with PCA axes



# Types of ordination techniques

R package vegan

Indirect gradient analysis (aka unconstrained ordination)

- Utilizes only the species x sample matrix

- Any environmental data are used after the analysis to aid with interpretation

**Nonmetric Multidimensional Scaling and Principal Component Analysis*

Direct gradient analysis (aka constrained ordination)

- Utilizes environmental data in addition to a species x sample matrix

- Assess whether species composition is related to measured environmental data

**Canonical Correspondence Analysis and Redundancy Analysis*

# Canonical Correspondence Analysis and Redundancy Analysis

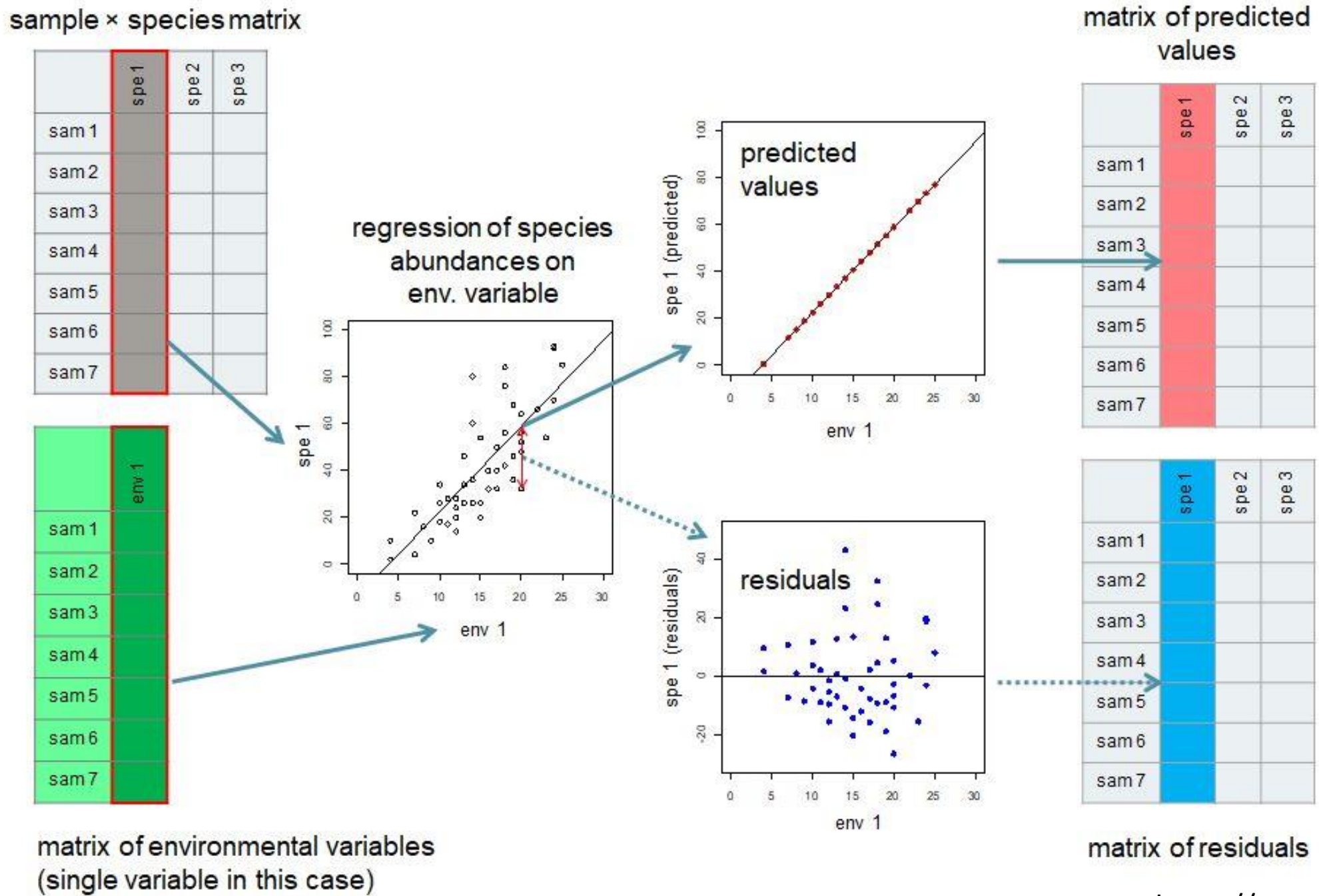
*Used to assess whether species composition is related to measured environmental variables*

Explores relationships between two matrices – response and predictor

Combines multiple regression with classical ordination

Can be used to test hypotheses via permutation tests

Linear (RDA) or unimodal (CCA)

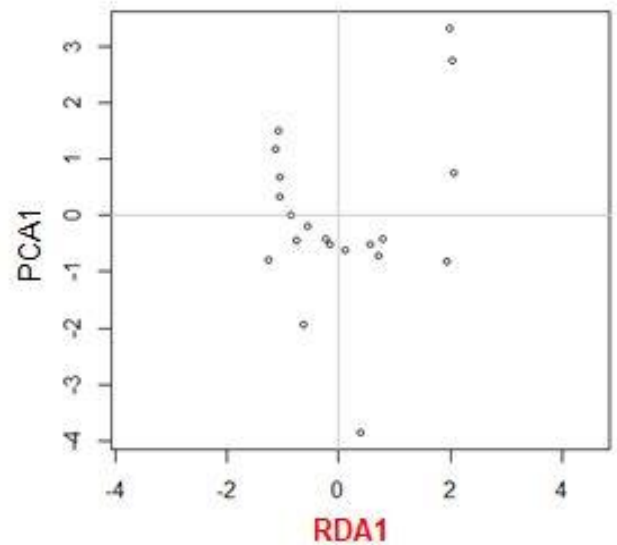



https://www.davidzeleny.net/

matrix of predicted values

| | spe 1 | spe 2 | spe 3 |
|---|---|---|---|
| sam 1 | | | |
| sam 2 | | | |
| sam 3 | | | |
| sam 4 | | | |
| sam 5 | | | |
| sam 6 | | | |
| sam 7 | | | |

PCA on predicted values

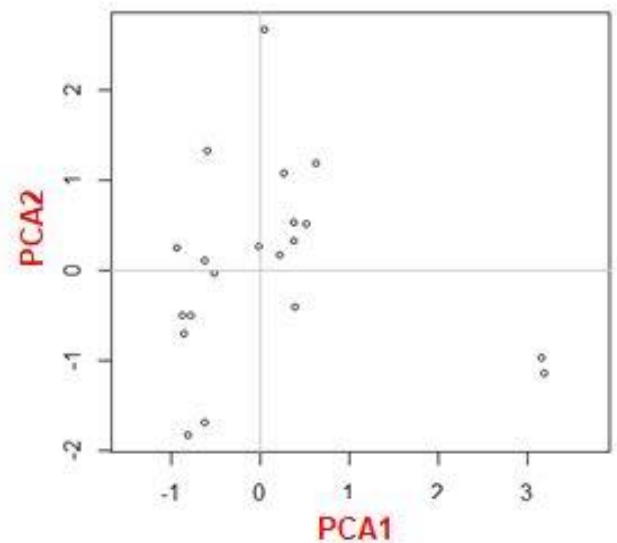



constrained ordination axes

| | spe 1 | spe 2 | spe 3 |
|---|---|---|---|
| sam 1 | | | |
| sam 2 | | | |
| sam 3 | | | |
| sam 4 | | | |
| sam 5 | | | |
| sam 6 | | | |
| sam 7 | | | |

matrix of residuals

PCA on residuals



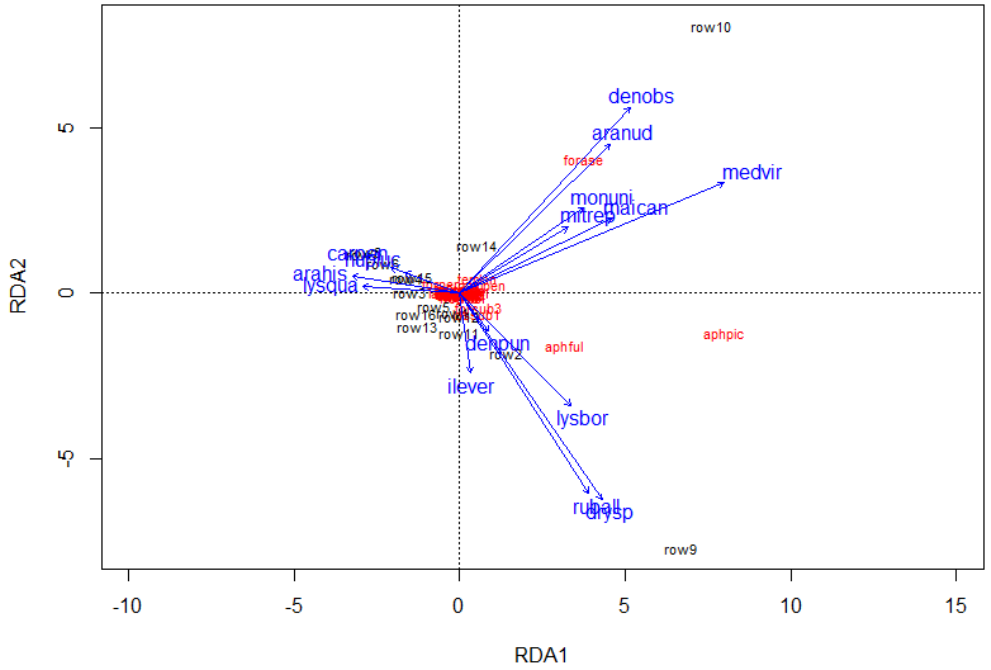

unconstrained ordination axes

https://www.davidzeleny.net

# RDA

Axis 1: 64.9%

Axis 2: 15.5%




```
> ants.rda <- rda(ants3[,6:35] ~ ., herb3[,5:25])
> ants.rda
Call: rda(formula = ants3[, 6:35] ~ arahis + aranud + carpen + denobs + denpun + drysp + hupluc + ilever + lysbor
+ lysqua + maican + medvir + mitrep + monuni + ruball + rubhis + rubida + uvuses + vacang + vaccor + ztot.cover,
data = herb3[, 5:25])

              Inertia Proportion Rank
Total            1260          1
Constrained      1260          1   15
Unconstrained       0          0    0
Inertia is variance
Some constraints or conditions were aliased because they were redundant

Eigenvalues for constrained axes:
  RDA1   RDA2   RDA3   RDA4   RDA5   RDA6   RDA7   RDA8   RDA9 RDA10 RDA11 RDA12 RDA13 RDA14 RDA15
818.0 195.4   83.1   65.5   50.6   18.9   15.9    8.4    2.3   1.2    0.3    0.3    0.1    0.0    0.0
```

# RDA – Hypothesis testing

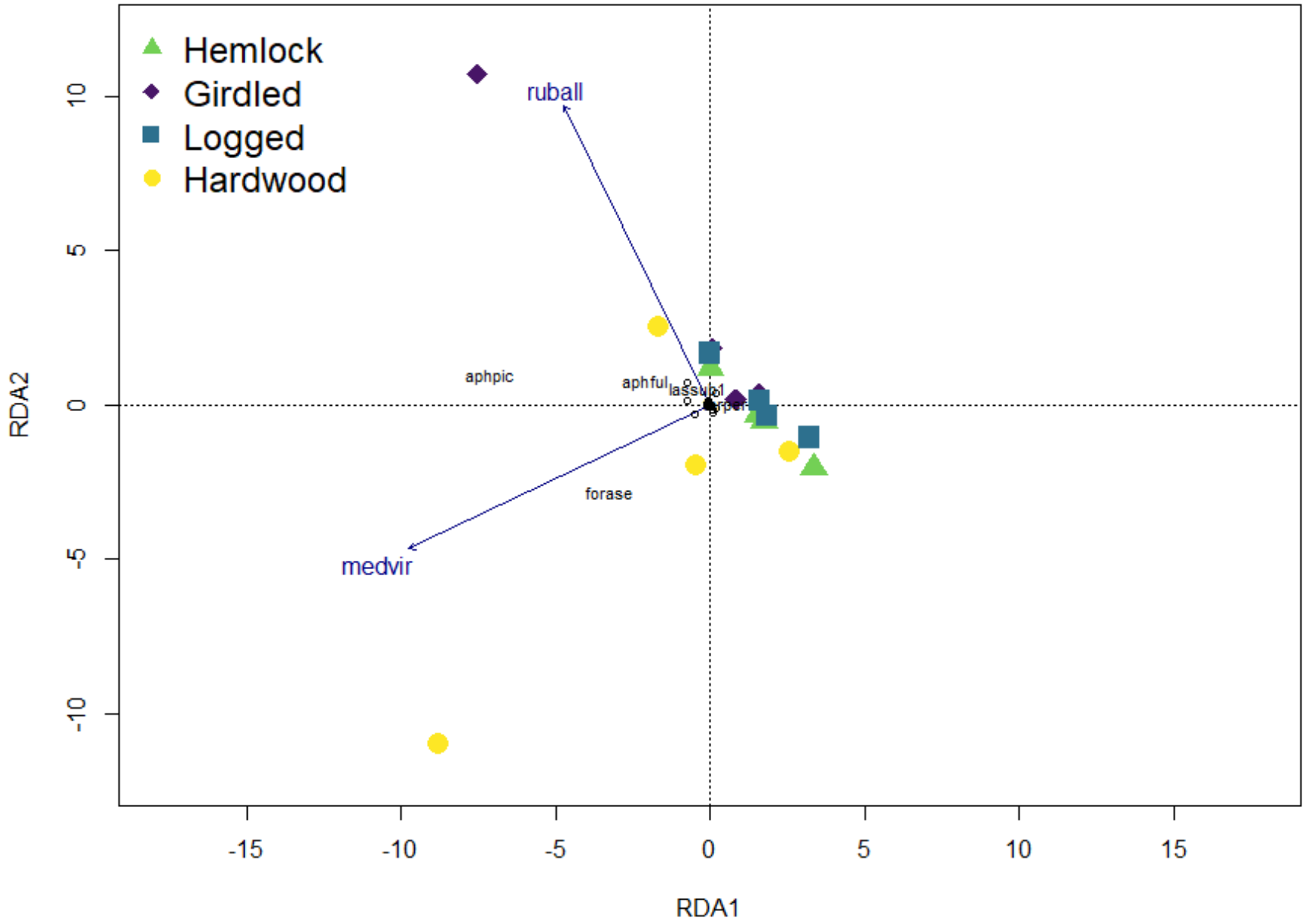
ordistep()

```
Step: ants3[, 6:35] ~ medvir + ruball

         Df    AIC      F Pr(>F)
- ruball  1 108.63  4.625  0.025 *
- medvir  1 114.47 12.390  0.010 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

             Df    AIC      F Pr(>F)
+ rubhis      1 106.05 1.3543  0.230
+ carpen      1 105.99 1.4067  0.265
+ vaccor      1 106.13 1.2869  0.285
+ monuni      1 106.50 0.9844  0.415
+ arahis      1 106.46 1.0183  0.420
+ denpun      1 106.59 0.9121  0.455
+ ztot.cover  1 106.61 0.8927  0.475
+ drysp       1 106.62 0.8895  0.495
+ lysqua      1 106.69 0.8345  0.495
+ uvuses      1 106.89 0.6741  0.650
+ denobs      1 106.83 0.7232  0.655
+ aranud      1 106.89 0.6721  0.665
+ vacang      1 106.88 0.6800  0.670
+ rubida      1 106.84 0.7152  0.675
+ lysbor      1 106.92 0.6494  0.695
+ ilever      1 107.36 0.3023  0.875
+ mitrep      1 107.25 0.3897  0.885
+ hupluc      1 107.34 0.3213  0.935
+ maican      1 107.34 0.3244  0.950
```

```
> summary(ants.rda.red)

Call:
rda(formula = ants3[, 6:35] ~ medvir + ruball, data = herb3[,      5:25])

Partitioning of variance:
              Inertia Proportion
Total          1259.9     1.0000
Constrained     715.5     0.5679
Unconstrained   544.4     0.4321
```

```
> anova(ants.rda.red, by = 'axis')
Permutation test for rda under reduced model
Forward tests for axes
Permutation: free
Number of permutations: 999

Model: rda(formula = ants3[, 6:35] ~ medvir + ruball, data = herb3[, 5:25])
          Df Variance       F Pr(>F)
RDA1       1   619.03 14.7815  0.004 **
RDA2       1    96.48  2.3039  0.061 .
Residual 13   544.42
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

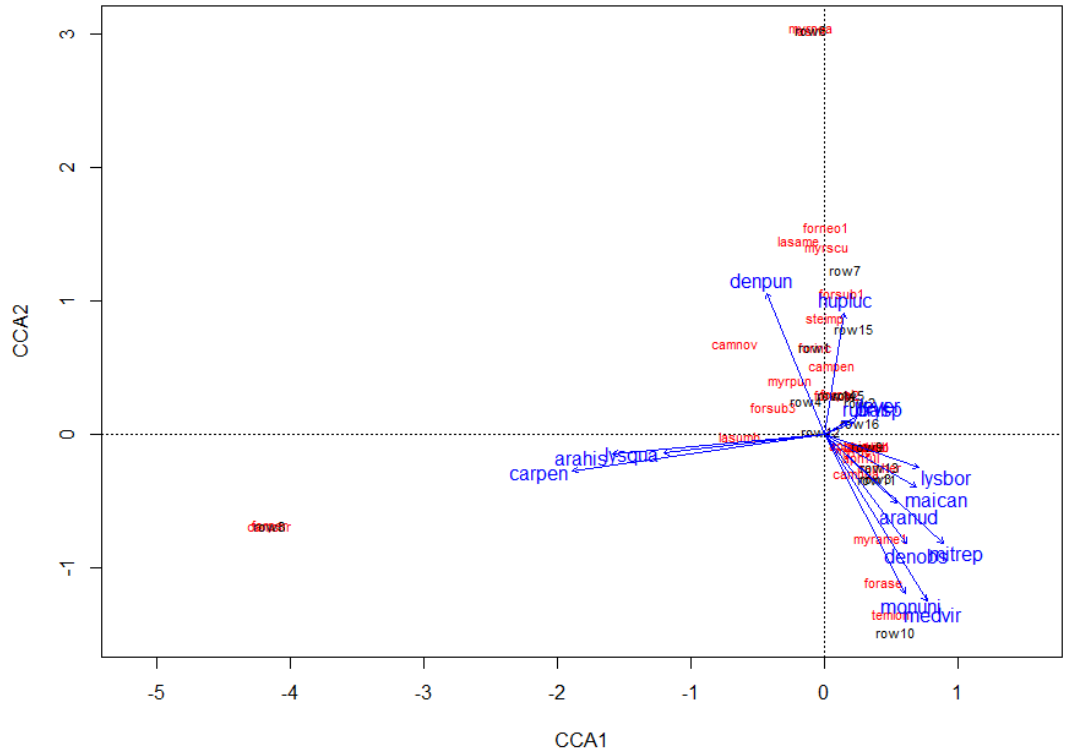
# RDA


Hemlock
Girdled
Logged
Hardwood
ruball
aphpic
aphful
lassul
forase
medvir
RDA2
RDA1
10
5
0
-5
-10
-15
-10
-5
0
5
10
15

# CCA

Axis 1: 30.3%

Axis 2: 18.3%




```
> ants.cca <- cca(ants3[,6:35] ~ ., herb3[,5:25])
> ants.cca
Call: cca(formula = ants3[, 6:35] ~ arahis + aranud + carpen + denobs + denpun + drysp + hupluc +
ilever + lysbor + lysqua + maican + medvir + mitrep + monuni + ruball + rubhis + rubida + uvuses +
vacang + vaccor + ztot.cover, data = herb3[, 5:25])

              Inertia Proportion Rank
Total           1.667      1.000
Constrained     1.667      1.000   15
Unconstrained   0.000      0.000    0
Inertia is scaled Chi-square
Some constraints or conditions were aliased because they were redundant

Eigenvalues for constrained axes:
  CCA1   CCA2   CCA3   CCA4   CCA5   CCA6   CCA7   CCA8   CCA9  CCA10  CCA11  CCA12  CCA13  CCA14  CCA15
0.5053 0.3058 0.2512 0.1681 0.1206 0.0951 0.0832 0.0618 0.0480 0.0095 0.0086 0.0065 0.0025 0.0003 0.0000
```

# CCA – Hypothesis testing

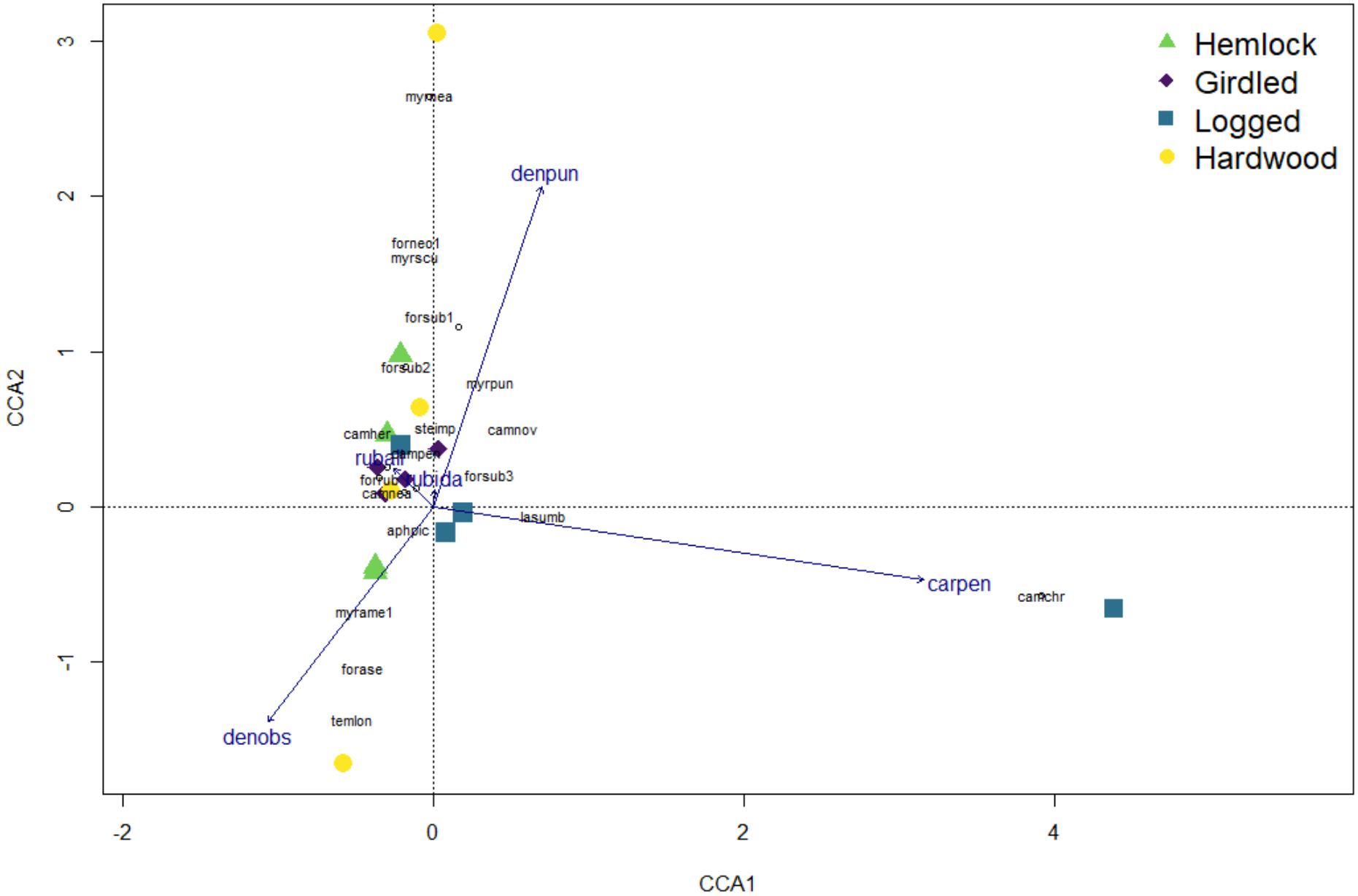
ordistep()

```
Step: ants3[, 6:35] ~ carpen + denobs + denpun + ruball + rubida

          Df     AIC      F Pr(>F)
- rubida   1 72.728 2.2035  0.095 .
- ruball   1 73.374 2.7061  0.040 *
- denobs   1 74.605 3.7220  0.020 *
- denpun   1 75.072 4.1286  0.005 **
- carpen   1 77.799 6.7536  0.005 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*'

             Df     AIC      F Pr(>F)
+ monuni      1 71.137 1.4601  0.155
+ lysqua      1 71.243 1.3906  0.260
+ aranud      1 71.395 1.2924  0.260
+ medvir      1 71.538 1.2007  0.260
+ vacang      1 71.605 1.1582  0.445
+ uvuses      1 71.866 0.9937  0.470
+ rubhis      1 71.925 0.9574  0.475
+ arahis      1 71.920 0.9602  0.495
+ lysbor      1 72.173 0.8042  0.495
+ vaccor      1 72.226 0.7717  0.550
+ drysp       1 72.126 0.8331  0.615
+ ilever      1 72.476 0.6199  0.660
+ hupluc      1 72.301 0.7260  0.680
+ ztot.cover  1 72.445 0.6386  0.755
+ mitrep      1 72.764 0.4485  0.895
+ maican      1 72.899 0.3693  0.915
```

```
> summary(ants.cca.red)

Call:
cca(formula = ants3[, 6:35] ~ carpen + denobs + denpun + ruball +      rubida, data = herb3[, 5:25])

Partitioning of scaled Chi-square:
              Inertia Proportion
Total           1.666     1.0000
Constrained     1.098     0.6586
Unconstrained   0.569     0.3414
```

```
> anova(ants.cca.red, by = 'axis')
Permutation test for cca under reduced model
Forward tests for axes
Permutation: free
Number of permutations: 999

Model: cca(formula = ants3[, 6:35] ~ carpen + denobs + denpun + ruball + rubida, data = herb3[, 5:25])
         Df ChiSquare      F Pr(>F)
CCA1      1   0.45327 7.9665  0.001 ***
CCA2      1   0.26690 4.6909  0.004 **
CCA3      1   0.21112 3.7106  0.002 **
CCA4      1   0.10005 1.7585  0.231
CCA5      1   0.06616 1.1629  0.281
Residual 10   0.56897
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

# CCA


Hemlock
Girdled
Logged
Hardwood
denpun
carpen
denobs
rubalb
rubida
myrnea
forneo1
myrscu
forsub1
forsub2
myrpun
steimp
camnov
camher
forsub3
lasumb
aphpic
myrame1
forase
temlon
camchr
CCA1
CCA2
-2
0
2
4
3
2
1
0
-1

# How do I know which to use: RDA or CCA?

Linear vs unimodal

Detrended correspondence analysis (DCA) to determine the length of the first axis

```
> DCA <- decorana(ants3[,6:35])
> DCA

Call:
decorana(veg = ants3[, 6:35])

Detrended correspondence analysis with 26 segments.
Rescaling of axes with 4 iterations.
Total inertia (scaled Chi-square): 1.6665

                        DCA1   DCA2    DCA3    DCA4
Eigenvalues           0.3538 0.1555 0.09320 0.09082
Additive Eigenvalues  0.3538 0.1551 0.08423 0.07622
Decorana values       0.5053 0.1358 0.03636 0.01418
Axis lengths          2.6486 1.5961 0.98833 0.91647
```

RDA: Axis length < 4

CCA: Axis length > 4

ter Braak and Smilauer 2002