



Monte Carlo Approaches to Statistics

Carri LeRoy, PhD

Lab II, 2265

360-867-5483

LeRoyC@evergreen.edu

Monte Carlo & Resampling

- Computer intensive (-)

Monte Carlo & Resampling

- Computer intensive (-)
- Relatively new (because of the above) (-)

Monte Carlo & Resampling

- Computer intensive (-)
- Relatively new (because of the above) (-)
- Relatively easy to understand (+)

Monte Carlo & Resampling

- Computer intensive (-)
- Relatively new (because of the above) (-)
- Relatively easy to understand (+)
- Adaptable and applicable to many circumstances (+)

Monte Carlo & Resampling

- Computer intensive (-)
- Relatively new (because of the above) (-)
- Relatively easy to understand (+)
- Adaptable and applicable to many circumstances (+)
- results can change due to random re-sampling (-)

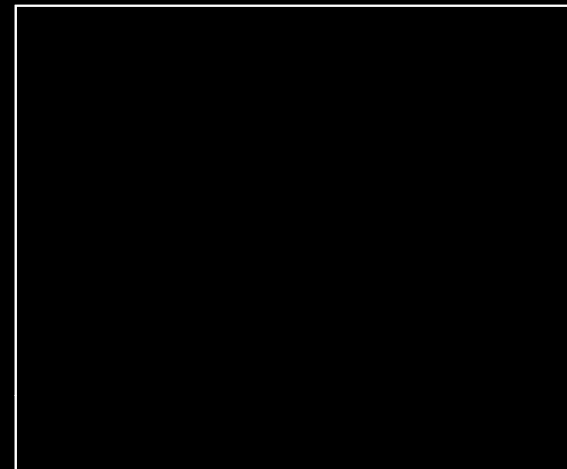
How does it work?

Re-sampling vs. Bootstrapping

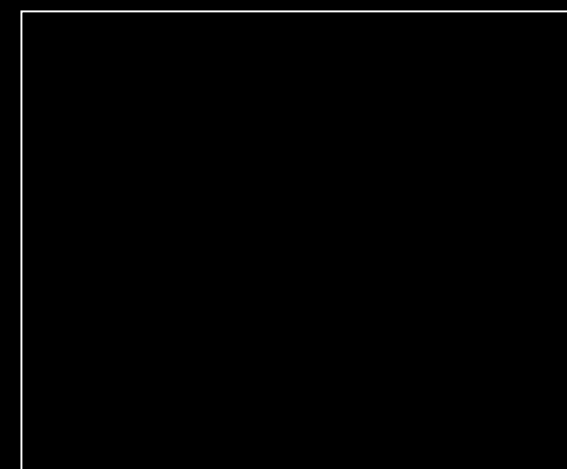
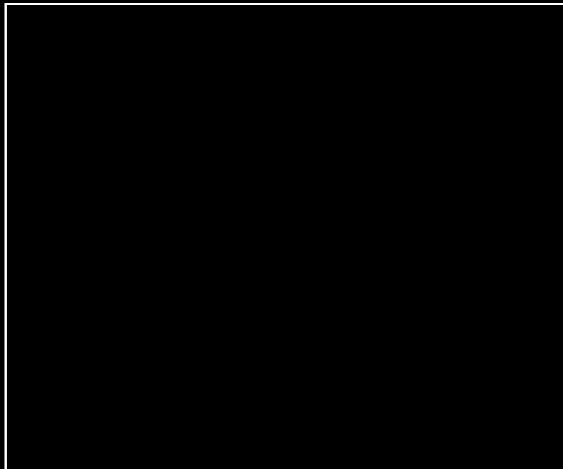
- Re-sampling w/o replacement = shuffling
- Re-sampling w/ replacement = bootstrapping
- Some analyses will require one or the other...most simple analyses require shuffling

4 Basic Analyses

Continuous Y



Categorical Y

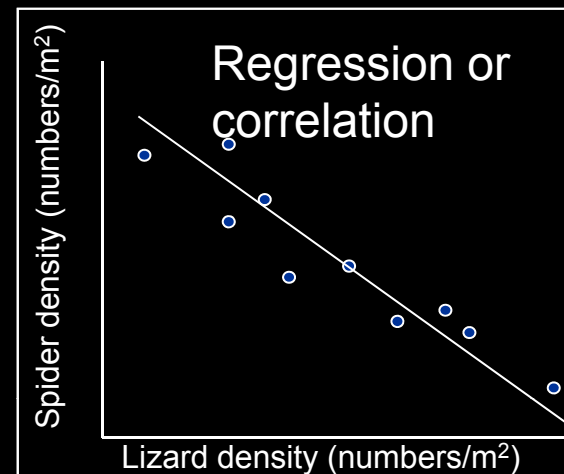
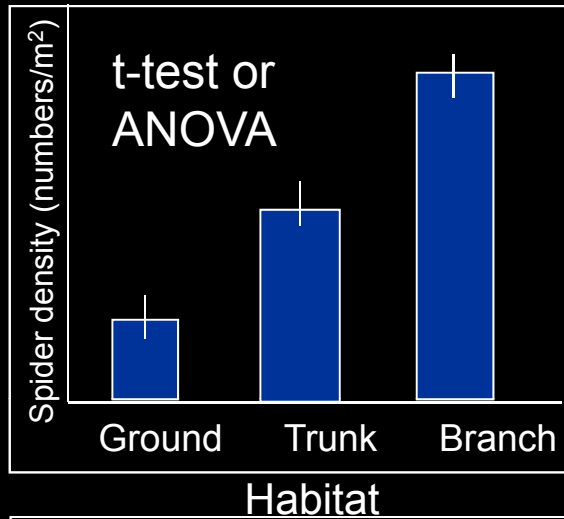


Categorical X

Continuous X

4 Basic Analyses

Continuous Y



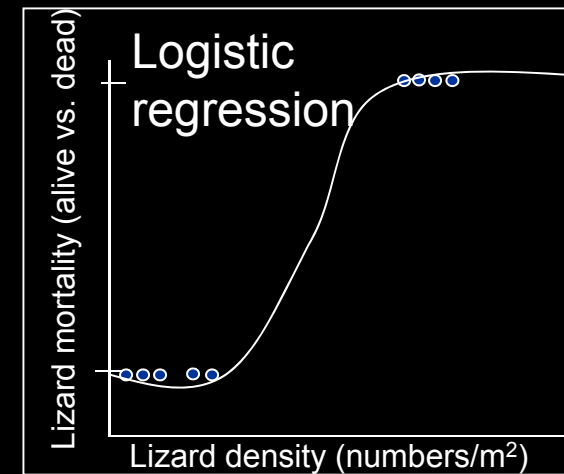
Categorical Y

Contingency table

Lizard mortality (alive vs. dead)	17	12	9
	4	8	15
	Ground	Trunk	Branch

Habitat

Categorical X



Continuous X

But others include:
ANCOVA
MANOVA
Multivariate
Modeling



Community Analysis using PC-ORD®



Communities

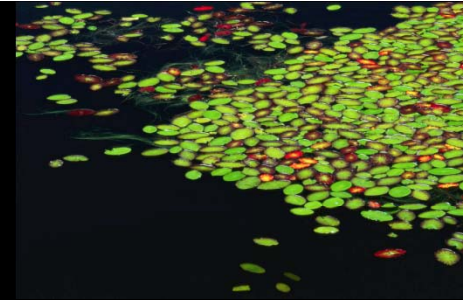
Communities

- Assemblages of plants, insects, microbes, birds
- Multi-trophic communities including producers, consumers & predators
- Assemblages of chemicals, environ. variables, genetic markers
- Assemblages of taxonomic attributes, traits
- Assemblages of values, opinions
- Any multivariate dataset



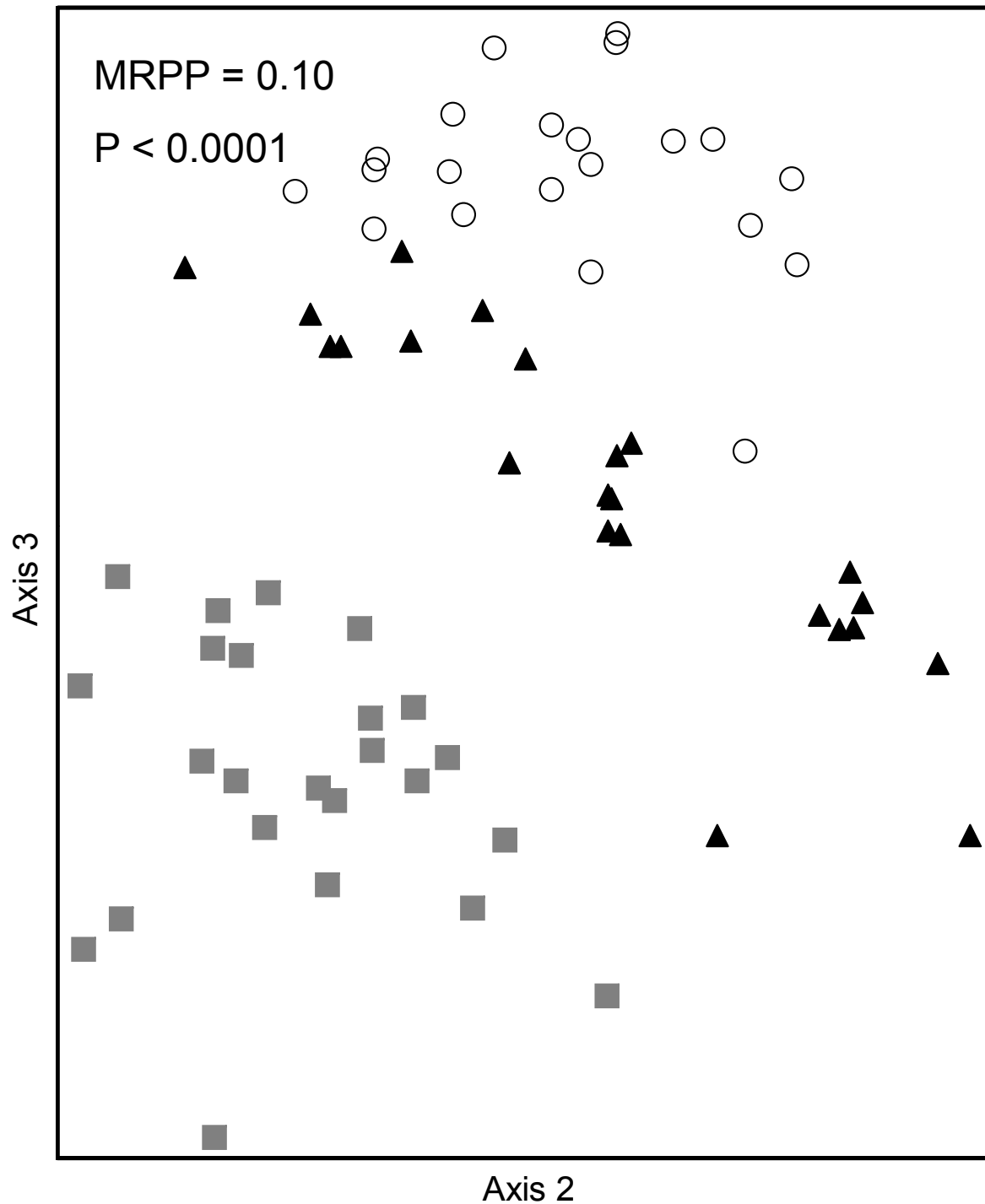
NMS Ordination: Possible uses

- Community ecology:
 - Main matrix: species abundance data
 - Second matrix: environmental data
- Chemical attributes:
 - Main matrix: a suite of chemical traits per sample
 - Second matrix: environmental data -soil samples



A Brief Introduction to Ordination

- Community data is inherently *multivariate data*, with a series of samples, each containing different abundances of taxa (or chemical concentrations; many variables)
- Ordination means to condense complicated data sets down to a few dimensions for visual inspection, and analysis



Stream

- Fossil Creek
- Oak Creek
- ▲ Wet Beaver Creek

Each symbol represents an entire community

There are 85 species in this dataset

Ordination reduced it to 2 so we can visualize community differences

How does it work?

- Between every pair of samples – calculate a “similarity” value
- Create a similarity matrix
- Plot the samples in space to represent the similarity
- Do this many times and pick the best solution
- Determine whether good groups exist

Distance measures: Which?

- Euclidean: p-dimensional version of the pythagorean theorem.
 - Poor for ecological data – good for environmental data or chemical data
- Sørensen's (Bray-Curtis): measured as percent similarity measured in “city-block-space.”
 - Great for ecological data – either abundance or presence/absence

See Faith et al. 1987. *Vegetatio*, 69:57-68

Example Similarity Matrix

	Sample 1	Sample 2	Sample 3
Sample 1	1.00	.78	.56
Sample 2	.78	1.00	.21
Sample 3	.56	.21	1.00

Raw Data – into PC-Ord

MAIN MATRIX

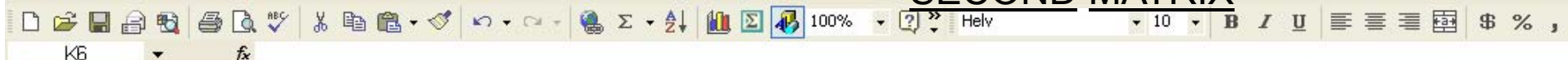
main_matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	68 SAMPLES																
2	85 SPECIES																
3		Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
4		HYDRAC	HYDRA2	PISIDIU	PISID2	AGABUS	DYTISCU	HULEECH	HULEEAD	ELMIDB	GYRINUS	PELTODY	SPHAERI	SEPHE	DOPEPO	CERATOR	CHIRONI
5	ALL 1	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	100
6	ALL 2	0	0	2	3	0	0	0	0	1	0	0	0	0	1	0	288
7	ALL 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	136
8	ALL 4	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	26
9	ALL 5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	772
10	ALL 6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4
11	ALL 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	109
12	ALL 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	822
13	ALL 9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36
14	ALL 10	0	0	0	0	0	0	1	0	0	0	0	0	0	2	7	186
15	ALL 11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4
16	ALL 12	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	28
17	ALL 13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
18	ALL 14	0	0	0	0	0	0	7	0	0	0	0	0	0	0	6	27
19	ALL 15	1	0	0	0	0	0	4	0	0	0	0	0	0	0	1	556
20	ALL 16	0	0	0	2	0	0	0	0	1	0	0	0	0	2	2	156
21	ALL 17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	9
22	ALL 18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
23	ALL 19	4	0	1	0	0	0	6	15	0	0	0	0	0	0	40	529
24	ALL 20	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	8
25	ALL 21	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	17
26	ALL 22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	37
27	ALL 23	1	0	0	0	0	0	2	1	0	0	0	0	0	0	5	116
28	ALL 24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27
29	ALL 25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
30	ALL 26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10
31	ALL 27	0	0	1	0	0	0	2	0	0	0	0	0	0	0	19	47
32	ALL 28	1	0	0	0	0	0	0	3	0	0	0	0	0	0	51	529
33	ALL 29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Species

Samples or plots

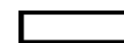
SECOND MATRIX



second_matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		68 SAMPLES													
2		8 PARAM													
3		C	Q	C	C	C	C	Q	Q						
4		ID tag	Ln AFDM	Species	Stream	Harvest	Rebar	Invert richness	ABUND						
5	ALL 1	366	1.3751	1	2	3	1	9	166						
6	ALL 2	367	1.0802	1	2	5	1	12	454						
7	ALL 3	368	1.4930	1	3	1	1	4	146						
8	ALL 4	370	1.3099	1	2	3	2	9	72						
9	ALL 5	371	1.3561	1	3	3	1	9	866						
10	ALL 6	373	1.4892	1	3	1	2	2	5						
11	ALL 7	374	1.4614	1	3	1	3	3	112						
12	ALL 8	375	1.3597	1	3	3	2	9	930						
13	ALL 9	377	1.4565	1	3	1	4	2	37						
14	ALL 10	378	1.1025	1	3	5	1	23	349						
15	ALL 11	379	1.4195	1	1	1	1	7	19						
16	ALL 12	381	1.4502	1	1	1	2	9	67						
17	ALL 13	383	1.4484	1	1	1	3	8	57						
18	ALL 14	385	1.3810	1	1	3	1	14	86						
19	ALL 15	387	0.9228	1	3	5	2	24	906						
20	ALL 16	388	1.0361	1	2	5	2	16	371						
21	ALL 17	389	1.3197	1	2	3	3	6	21						
22	ALL 18	390	0.0000	1	2	5	3	7	56						
23	ALL 19	394	1.0081	1	1	5	1	22	704						
24	ALL 20	395	1.1828	1	2	3	4	7	77						
25	ALL 21	396	0.0106	1	2	5	4	8	48						
26	ALL 22	397	1.3799	1	1	3	2	12	86						
27	ALL 23	398	0.9146	1	3	5	3	26	264						
28	ALL 24	399	1.4879	1	2	1	1	6	39						
29	ALL 25	400	1.4418	1	1	1	4	5	34						
30	ALL 26	405	1.4782	1	1	1	5	3	26						
31	ALL 27	406	1.4873	1	1	3	3	15	125						
32	ALL 28	407	1.1462	1	1	3	4	24	714						
33	ALL 29	409	0.0000	1	2	5	5	5	30						

Groups or parameters



Samples or plots



Ready

Some important data set issues

- Main matrix (samples & species)
- Second matrix (groups & other variables)
- Won't allow either a column or row of zeros
- Empty cells must be labeled as zero "0"
- Sample names cannot be numbers (must start with a letter, ALL1, ALL2, ALL3)
- Import .xls files – choose which sheet
- The program is very picky about data – be patient

1. Summary statistics

- Under “Summary” – basic community analysis
 - Means, std. dev., sums, min/max
 - Richness
 - H’ (Shannon’s)
 - Evenness
 - D (Simpsons)

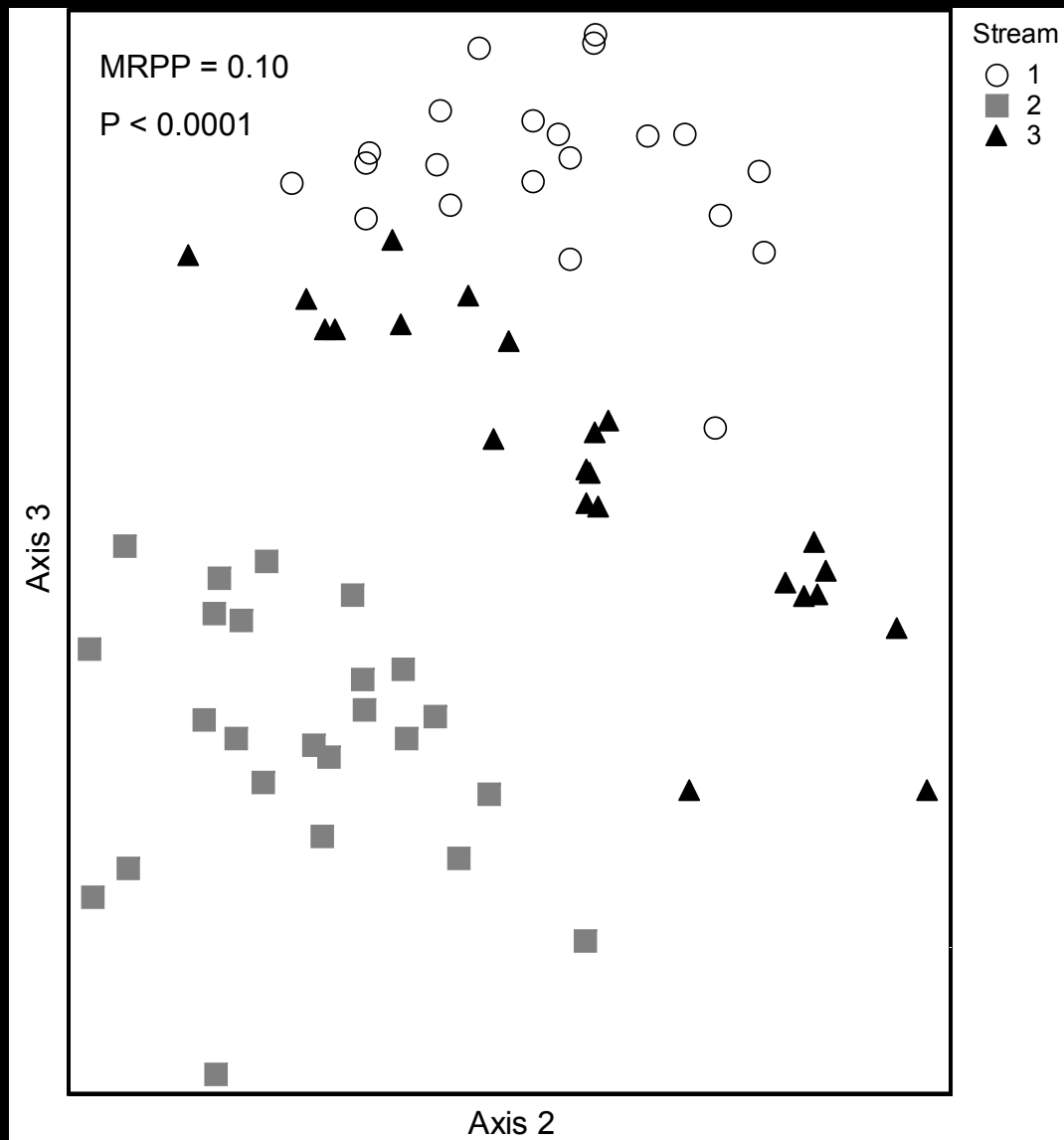
The screenshot displays the PC-ORD software interface. The main window shows a data matrix with 68 samples and 85 species. The 'Summary' menu is open, displaying a 'Result - RESULT.TXT' window with the following data:

***** Data Summarization *****
 PC-ORD, Version 4.02
 23 Aug 2006, 14:29

summary

Summary of		68 SAMPLES		N = 85 SPECIES						
Num.	Name	Mean	Stand.Dev.	Sum	Minimum	Maximum	S	E	H	D'
1	ALL 1	1.953	11.285	166.000	0.000	100.000	9	0.610	1.340	0.6000
2	ALL 2	5.341	32.719	454.000	0.000	288.000	12	0.470	1.168	0.5520
3	ALL 3	1.718	14.755	146.000	0.000	136.000	4	0.227	0.314	0.1303
4	ALL 4	0.847	3.865	72.000	0.000	26.000	9	0.750	1.648	0.7461
5	ALL 5	10.188	83.943	866.000	0.000	772.000	9	0.204	0.448	0.1990
6	ALL 6	0.059	0.446	5.000	0.000	4.000	2	0.722	0.500	0.3200
7	ALL 7	1.318	11.821	112.000	0.000	109.000	3	0.128	0.140	0.0525
8	ALL 8	10.941	89.493	930.000	0.000	833.000	9	0.316	0.478	0.3104
9	ALL 9	0.435	3.905	37.000	0.000	36.000	2	0.179	0.124	0.0526
10	ALL 10	4.106	22.116	349.000	0.000	186.000	23	0.512	1.605	0.6509
11	ALL 11	0.224	0.993	19.000	0.000	8.000	7	0.862	1.677	0.7590
12	ALL 12	0.788	4.282	67.000	0.000	28.000	9	0.599	1.316	0.6451
13	ALL 13	0.671	3.234	57.000	0.000	20.000	7	0.735	1.429	0.7178
14	ALL 14	1.129	4.044	96.000	0.000	27.000	14	0.796	2.101	0.8392
15	ALL 15	10.659	62.801	906.000	0.000	556.000	24	0.433	1.375	0.5846
16	ALL 16	4.365	20.904	371.000	0.000	156.000	16	0.556	1.540	0.7216
17	ALL 17	0.247	1.164	21.000	0.000	9.000	6	0.842	1.509	0.7302
18	ALL 18	0.659	2.885	56.000	0.000	20.000	7	0.826	1.607	0.7653
19	ALL 19	8.282	57.446	704.000	0.000	529.000	21	0.394	1.199	0.4289
20	ALL 20	0.906	5.541	77.000	0.000	49.000	7	0.595	1.157	0.5532
21	ALL 21	0.565	2.495	48.000	0.000	17.000	8	0.790	1.642	0.7613
22	ALL 22	1.012	4.542	86.000	0.000	31.000	11	0.723	1.734	0.7539
23	ALL 23	3.106	13.445	264.000	0.000	118.000	26	0.662	2.156	0.7704
24	ALL 24	0.459	2.644	39.000	0.000	23.000	6	0.691	1.238	0.6022
25	ALL 25	0.400	2.300	34.000	0.000	16.000	5	0.680	1.094	0.6038

2. Ordination



Non-metric Multidimensional Scaling (NMDS or NMS)

- Well-suited to non-normal data
- Based on rank distances – tends to linearize the relationships
- Iterative search for ranking and placing samples in k dimensions (axes) that minimizes the “stress” of the configuration

Non-metric Multidimensional Scaling (NMDS or NMS)

- “stress” is a measure of how well the dissimilarity of the original data matrix is represented by the new k-dimensional solution (where k is usually 1-4).
- The original data matrix has p = # of species dimensions
- p -dimensions \rightarrow k-dimensions is stressful!

First things first...to decrease the effect of overabundant species on your results...

Relativization by Species Maximum

PC-ORD

File Edit **Modify Data** Summary Ordination Graph Groups Window Options Help

Main

Standardizations...
Relativizations...
Beals Smoothing
Multiply Matrices
Transpose Main Matrix
Delete Rows...
Delete Columns...
Multiply Or Add A Constant
Shuffle Data

General Relativization
Relativization By Maximum
Adjust To Mean
Adjust To Standard Deviate
Binary W/Respect To Mean
Rank Adjustment
Binary W/Respect To Median
Variates By Ubiquity Row/Col
Info Function Of Ubiquity

68				
85				
ALL 1				
ALL 2				
ALL 3				
ALL 4	0			
ALL 5	2	0	0	
ALL 6	0	0	0	
ALL 7	0	0	0	
ALL 8	0	0	0	
ALL 9	0	0	0	
ALL 10	0	0	0	
ALL 11	0	0	0	

Second - SECOND_MATRIX.WK1

68	SAMPLES			
8	PARAM			
	C	Q	C	C
	ID tag	Ln AFDM	Species	Stream
ALL 1	366	1.3751	1	2
ALL 2	367	1.08023	1	2
ALL 3	368	1.49298	1	3
ALL 4	370	1.30993	1	2
ALL 5	371	1.35609	1	3
ALL 6	373	1.48915	1	3
ALL 7	374	1.46144	1	3
ALL 8	375	1.35973	1	3
ALL 9	377	1.45651	1	3
ALL 10	378	1.10248	1	3
ALL 11	379	1.41948	1	1

Main:MAIN_MATRIX.WK1
Second:SECOND_MATRIX.WK1
Row:
Col:
Result:

NAU WebMail...
PC-ORD_2004
Control Panel
CAL-Resampling
Microsoft Word
Microsoft Po...
Microsoft Excel
PC-ORD

Main - MAIN_MATRIX.WK1

68	SAMPLES			
85	SPECIES			
	Q	Q	Q	
	HYDRACA	HYDRA2	PISIDIU	
ALL 1	0	0	0	
ALL 2	0	0	2	
ALL 3	0	0	0	
ALL 4	0	0	0	
ALL 5	2	0	0	
ALL 6	0	0	0	
ALL 7	0	0	0	
ALL 8	0	0	0	
ALL 9	0	0	0	
ALL 10	0	0	0	
ALL 11	0	0	0	

Second - SECOND_MATRIX.WK1

68	SAMPLES			
8	PARAM			
	C	Q	C	C
	ID tag	Ln AFDM	Species	Str
ALL 1	366	1.3751	1	2
ALL 2	367	1.08023	1	2
ALL 3	368	1.49298	1	3
ALL 4	370	1.30993	1	2
ALL 5	371	1.35609	1	3
ALL 6	373	1.48915	1	3
ALL 7	374	1.46144	1	3
ALL 8	375	1.35973	1	3
ALL 9	377	1.45651	1	3
ALL 10	378	1.10248	1	3
ALL 11	379	1.41948	1	1

Relativization By Maximum

$$b = X_{ij}/X_{maxj}$$

- ☒ Rows: SAMPLES
☐ Columns: SPECIES

OK

Cancel

Help

Main:MAIN_MATRIX.WK1

Second:SECOND_MATRIX.WK1

Row:

Col:

Result:

C:\Program Files\PCORD4\TEMP.WK1

68	SAMPLES							
85	SPECIES							
	Q	Q	Q	Q	Q	Q	Q	Q
	HYDRACA	HYDRA2	PISIDIU	PISID2	AGABUS	DYTISCU	HULEECH	HULEEAD
ALL 1	0	0	0	0	0	0	0	0
ALL 2	0	0	0.00694444	0.0104166	0	0	0	0
ALL 3	0	0	0	0	0	0	0	0
ALL 4	0	0	0	0	0	0	0	0
ALL 5	0.00259067	0	0	0	0	0	0	0
ALL 6	0	0	0	0	0	0	0	0
ALL 7	0	0	0	0	0	0	0	0
ALL 8	0	0	0	0	0	0	0	0
ALL 9	0	0	0	0	0	0	0	0
ALL 10	0	0	0	0	0	0	0.00537634	0
ALL 11	0	0	0				0	0
ALL 12	0	0	0				0.0357142	0

Transformed Matrix

Do you wish to use TEMP.WK1 as the new Main Matrix?

OK

Discard

Help

Result - RESULT.TXT

***** Data Modification *****

PC-ORD, Version 4.02

23 Aug 2006, 14:30

RELATIVIZATION BY MAXIMUM FOR EACH ROW

***** Operation completed *****

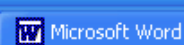
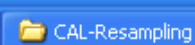
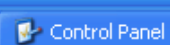
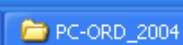
Main:MAIN_MATRIX.WK1

Second:SECOND_MATRIX.WK1

Row:

Col:

Result:

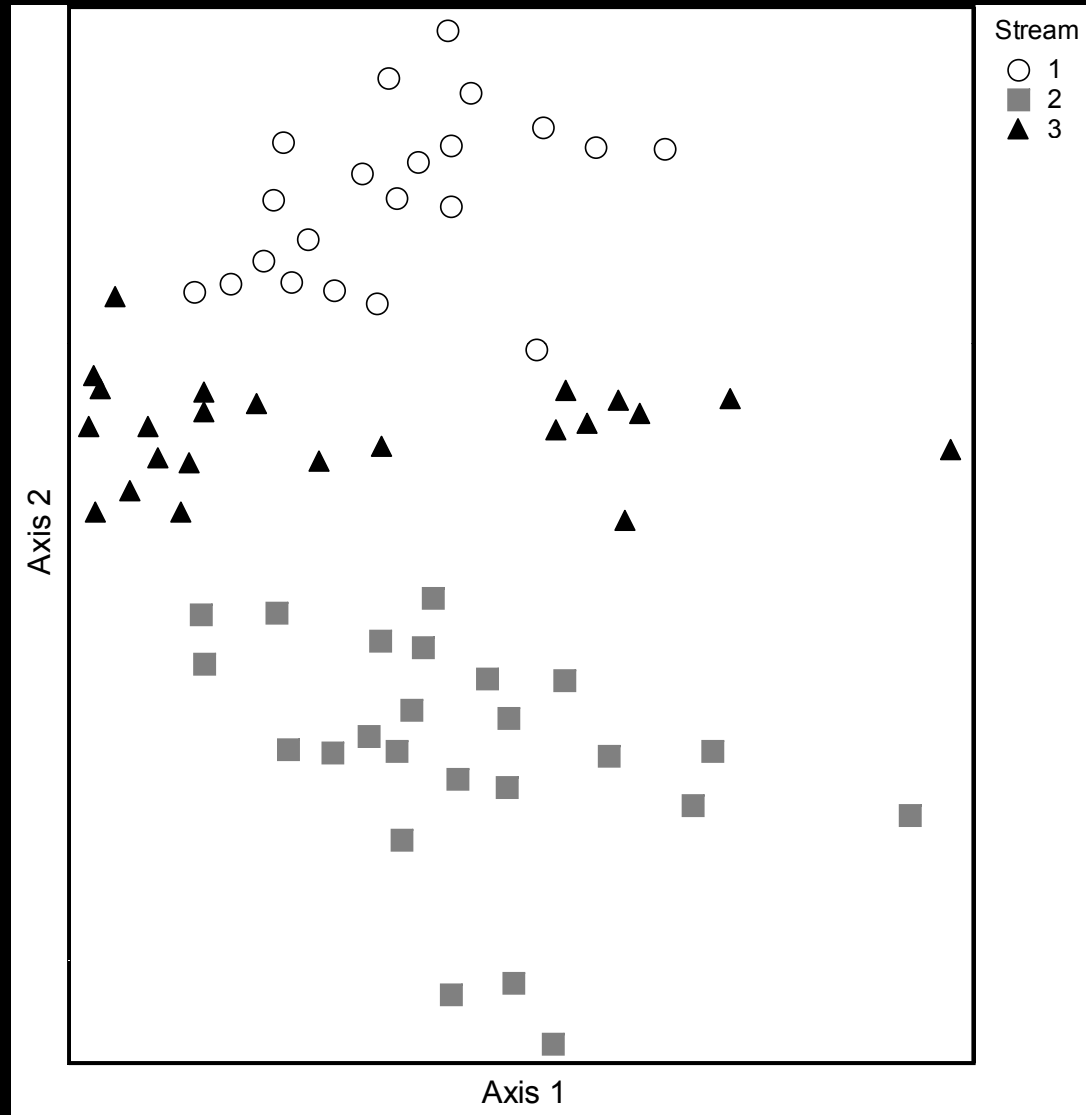


Ordination basics

Ordination / NMS: either Autopilot mode or customized.... 1) Pick distance measure, 2) set parameters, 3) set output, 4) GO!

Increase iterations if you want and here you go!

Graph ordination →



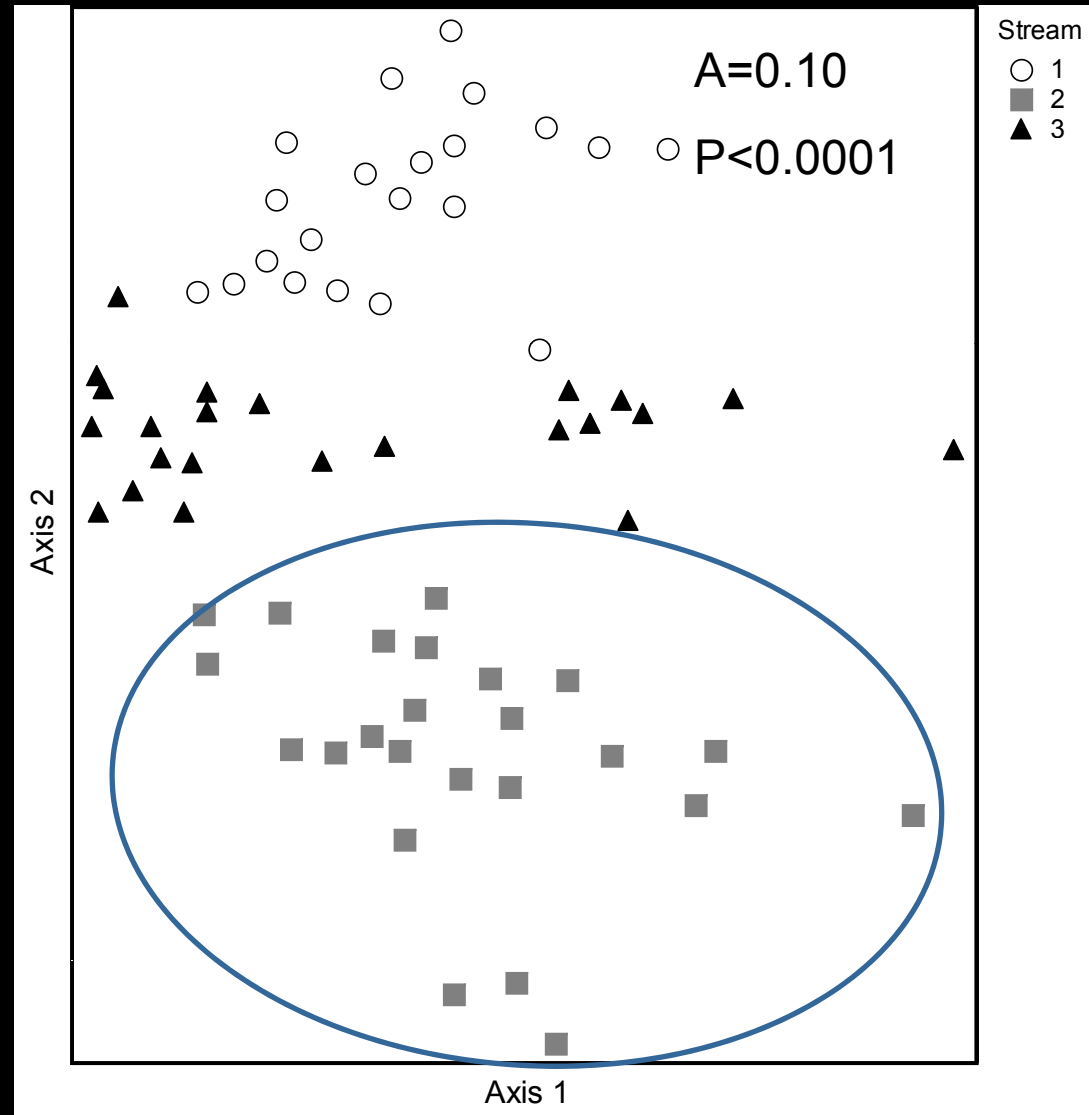
Ordination basics

Ordination / NMS: either Autopilot mode or customized.... 1) Pick distance measure, 2) set parameters, 3) set output, 4) GO!

Increase iterations if you want and here you go!

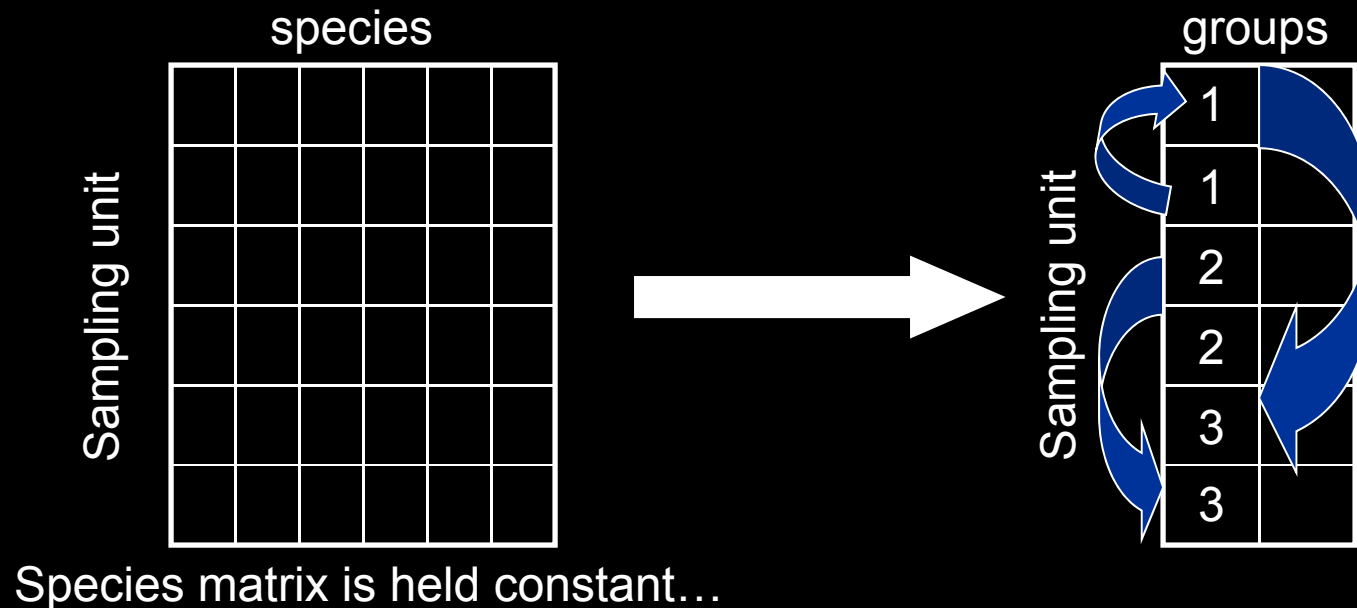
Graph ordination

Run an MRPP or PERMANOVA to go along with it



MRPP: Multi-response permutation procedure

- If planning on running MRPP or MRBP you should use the same dissimilarity measure as NMDS
- MRPP shuffles your second matrix like so...



MRPP

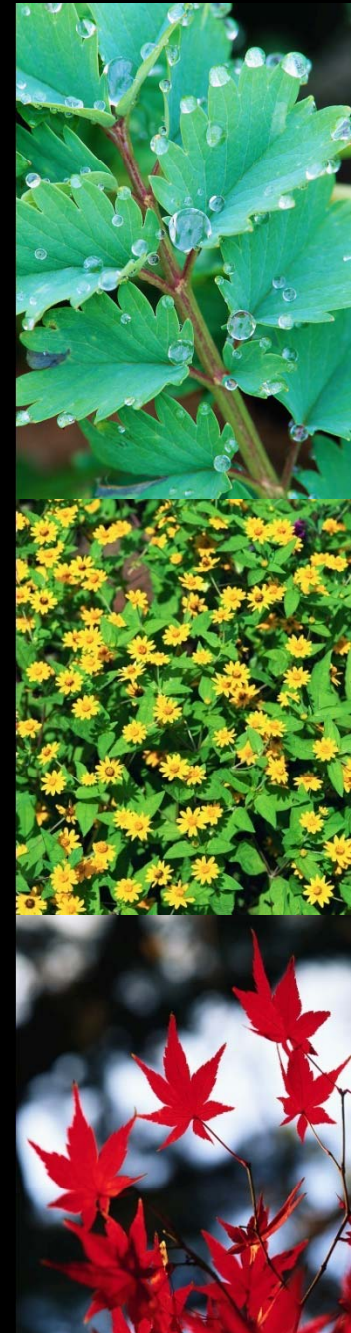
- Requires no distributional assumptions
- Assumes:
 1. Distance measure chosen adequately represents the variation of interest in data
 2. Sample units are independent

What to report?

- **The software** (because these procedures aren't available in general stats programs)
- **Distance measure & justification**
- **How groups were defined** (include the size of each group)
- **Chance-corrected within-group agreement, A**
- **P-value**

Some other special PC-ORD options...

Indicator Species Analysis
Correlations with ordination
Cluster Analysis
Mantel tests
Other types of ordination
Ordination through time



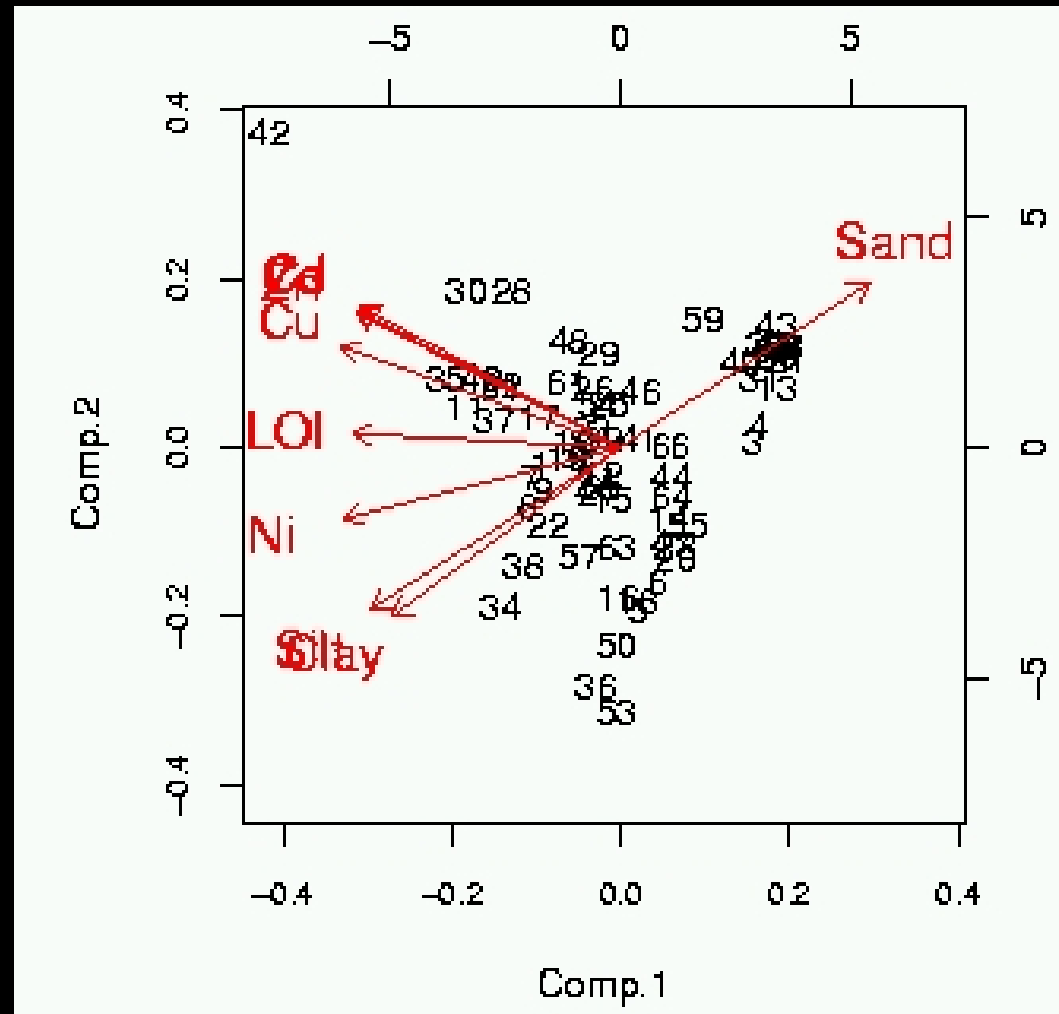
Indicator Species Analysis

Certain species “indicate” for certain conditions: old growth, wetlands, clean air, etc. – Are there any species in my dataset that “indicate” for one of my treatments? We can test for this!



Correlations with Main Matrix

Run correlations through your ordination space to look for relationships between community structure and environmental variables (pH, temp, DO, TDS, nitrate, etc.)



Questions????

