# Characterizing insect communities using measures of diversity

ENTMLGY 6707 Entomological Techniques and Data Analysis

Learning objectives

1) Distinguish among different diversity metrics

2) Understand how multiple dimensions of diversity contribute to characterizing communities

3) Interpret the outputs of different metrics

Traditional approach has focused on taxonomy

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

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

n = 2

90 x 90 m plots

Harvard Forest

Established in 1907

Harvard University's 4000 acre laboratory & classroom

Long Term Ecological Research site since 1988

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

Long format

	year	date	block	plot	treatment	moose.cage	trap.type	trap.num	genus	species	code	abundance
1	2003	2003-06-01	Ridge	4	Logged	NA	bait	1 hour	Aphaenogaster	fulva	aphful	2
2	2003	2003-06-01	Ridge	4	Logged	NA	bait	1 hour	Aphaenogaster	picea	aphpic	2
3	2003	2003-06-01	Ridge	4	Logged	NA	bait	1 hour	Temnothorax	longispinosus	temlon	1
4	2003	2003-06-01	Ridge	4	Logged	NA	bait	15 minutes	Aphaenogaster	picea	aphpic	2
5	2003	2003-06-01	Ridge	4	Logged	NA	hand	NA	Aphaenogaster	picea	aphpic	1
6	2003	2003-06-01	Ridge	4	Logged	NA	hand	NA	Camponotus	pennsylvanicus	campen	1
7	2003	2003-06-01	Ridge	4	Logged	NA	hand	NA	Lasius	americanus	lasame	1
8	2003	2003-06-01	Ridge	4	Logged	NA	hand	NA	Myrmica	punctiventris	myrpun	1
9	2003	2003-06-01	Ridge	4	Logged	NA	hand	NA	Temnothorax	longispinosus	temlon	1
10	2003	2003-06-01	Ridge	4	Logged	NA	litter	1	Aphaenogaster	fulva	aphful	1
11	2003	2003-06-01	Ridge	4	Logged	NA	litter	1	Aphaenogaster	picea	aphpic	1
12	2003	2003-06-01	Ridge	4	Logged	NA	litter	2	Aphaenogaster	picea	aphpic	1
13	2003	2003-06-01	Ridge	4	Logged	NA	litter	2	Temnothorax	longispinosus	temlon	2
14	2003	2003-06-01	Ridge	4	Logged	NA	litter	3	Aphaenogaster	picea	aphpic	1
15	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	1	Aphaenogaster	picea	aphpic	1
16	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	15	Aphaenogaster	picea	aphpic	1
17	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	17	Aphaenogaster	picea	aphpic	2
18	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	19	Aphaenogaster	picea	aphpic	1
19	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	2	Aphaenogaster	picea	aphpic	1
20	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	20	Temnothorax	longispinosus	temlon	1
21	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	22	Aphaenogaster	picea	aphpic	1
22	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	23	Aphaenogaster	picea	aphpic	1
23	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	24	Aphaenogaster	picea	aphpic	1
24	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	5	Aphaenogaster	picea	aphpic	2
25	2003	2003-06-01	Ridge	4	Logged	NA	pitfall	7	Aphaenogaster	picea	aphpic	1
26	2003	2003-06-01	Ridge	5	Girdled	NA	bait	1 hour	Aphaenogaster	fulva	aphful	1

Wide format

2006

	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	

Abundance: the number of individuals of each distinct type

Descriptive information about a community is valuable

Species dominance

1175 individuals

15 species

Aphaenogaster picea (64.5%)

Aphaenogaster fulva (12.7%)

Abundance: the number of individuals of each distinct type

Indicator species analyses

Relationship between species abundance and environmental classification of sites

1. Reflect the biotic or abiotic state of the environment
2. Provide evidence for the impacts of environmental change
3. Predict the diversity of other species, taxa, or communities within an area

Can be used for conservation monitoring purposes

Abundance: the number of individuals of each distinct type

Indicator species analyses

```
> indval <- multipatt(ants3[,6:20], ants3[,4], duleg = TRUE, control = how(nperm=999))
> summary(indval, indvalcomp = TRUE)


 Multilevel pattern analysis
 ---------------------------

 Association function: IndVal.g
 Significance level (alpha): 0.05

 Total number of species: 15
 Selected number of species: 3
 Number of species associated to 1 group: 3
 Number of species associated to 2 groups: 0
 Number of species associated to 3 groups: 0

 List of species associated to each combination:

 Group HardwoodControl  #sps.  3
                A       B  stat p.value
myrpun  0.9760 1.0000 0.988   0.001 ***
forsub1 0.9192 0.5000 0.678   0.013 *
lasame  0.9192 0.5000 0.678   0.015 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

R package indicspecies

Indicator value index:

A = specificity; given that the species is found, probability that the site belongs to the target group

B = sensitivity; probability of finding the species in sites belonging to the site group

Myrmica punctiventris
Formica subaenescens
Lasius americanus

De Caceres and Legendre 2009

Species richness

How many species (or genera, families, etc)?

Absolute species richness: Count all species. Possible with some taxa, like large mammals. Very challenging with insects!

Estimate species richness by collecting samples
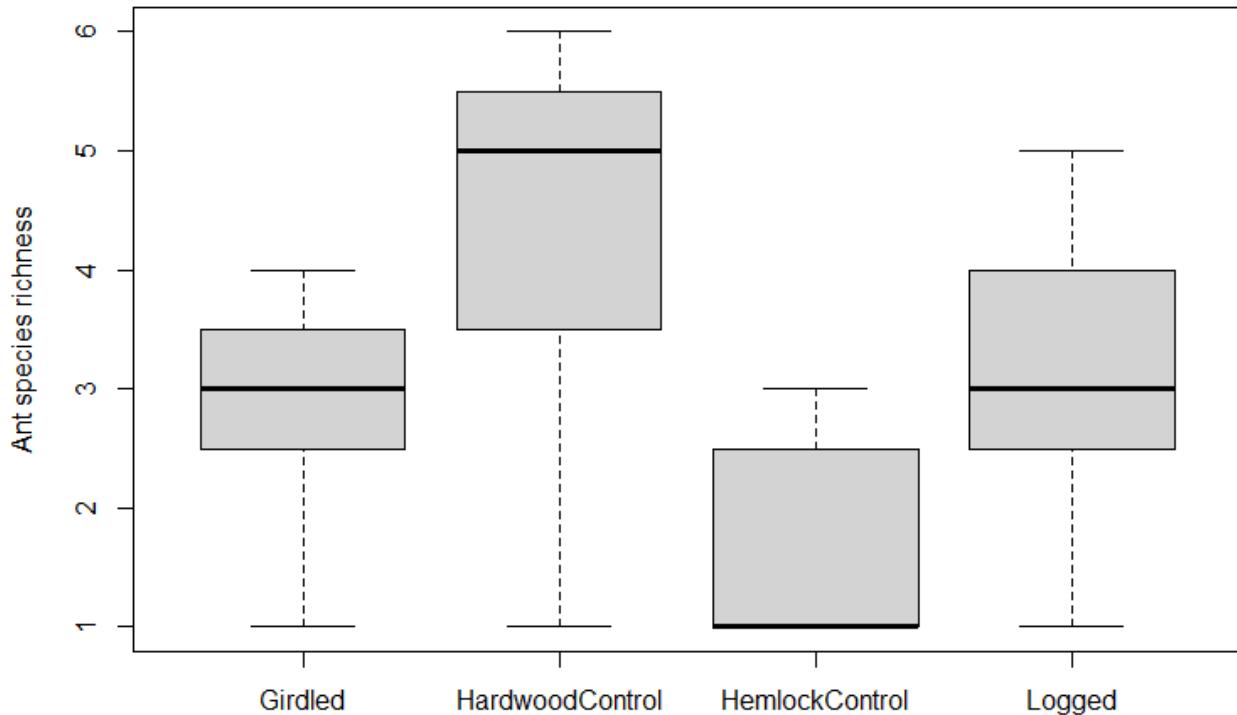
1. **Numerical richness:** number of species per specified number of individuals, biomass, or collection traps that do not yield density measure (such as pan traps, sticky traps, pitfall traps)

2. **Density based richness:** number of species per area or unit (number of species per meter square, species per sweep or per flower, etc.)

Richness: the number of distinct types

Count the number of species collected per site, trap, etc.

```
> ### Calculate species richness
> library(vegan)
> specnumber(ants3[,6:20]) # calculate species richness
 [1] 4 3 1 3 4 3 1 3 1 3 1 2 4 5 3 5 3 3 2 4 5 4 2 1 3 1 5 6 1 6
> ants3$sp.rich <- specnumber(ants3[,6:20]) # add it as a column in the data set
> boxplot(sp.rich ~ treatment, data = ants3,
+        xlab = "", ylab = "Ant species richness")
> |
```

Richness: the number of distinct types
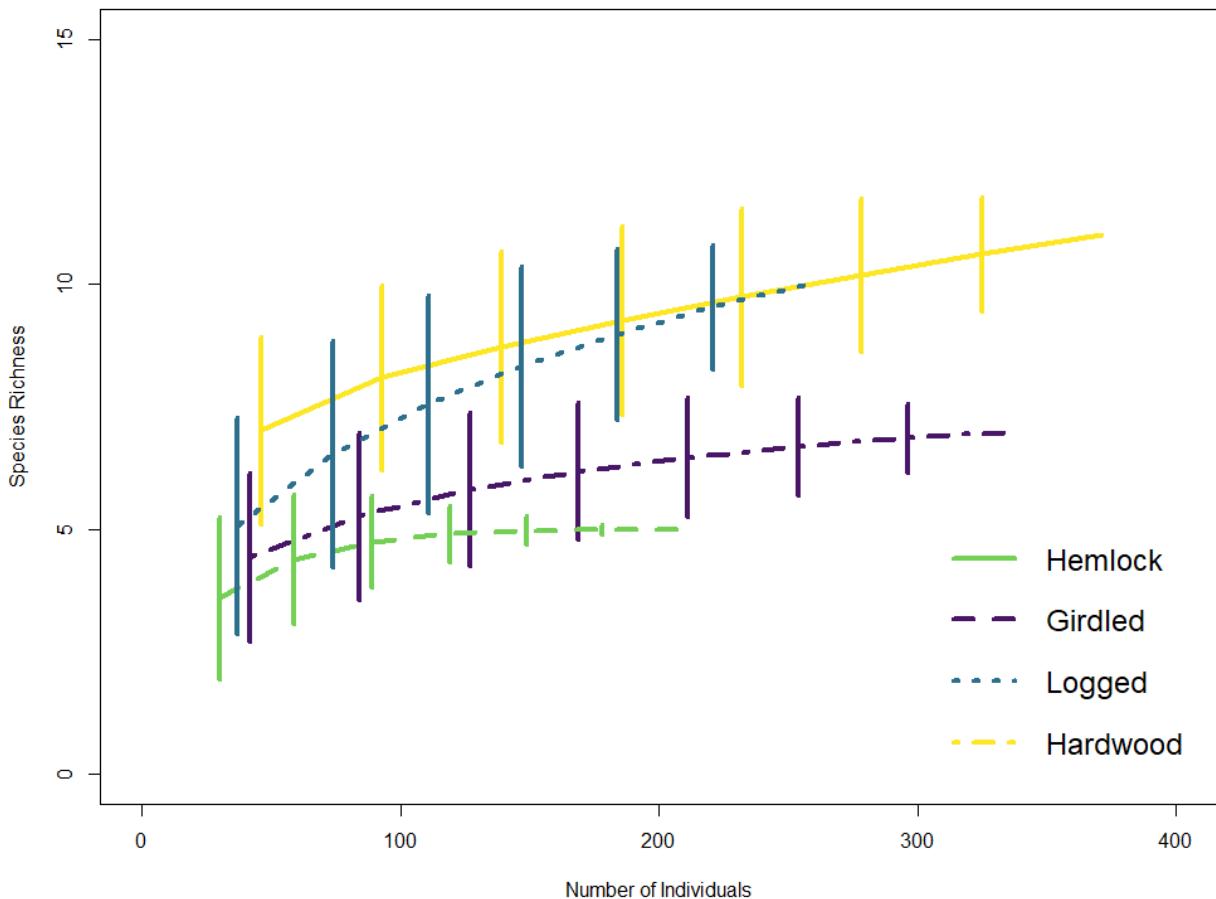
Estimate species richness with accumulation curves

R package
vegan

```
> sp.girdled <- specaccum(girdled[6:20], method = "rarefaction", permutations = 100, gamma = "jack2")
> sp.hardwood <- specaccum(hardwood[6:20], method = "rarefaction", permutations = 100, gamma = "jack2")
> sp.hemlock <- specaccum(hemlock[6:20], method = "rarefaction", permutations = 100, gamma = "jack2")
> sp.logged <- specaccum(logged[6:20], method = "rarefaction", permutations = 100, gamma = "jack2")
> |
```

Estimate species richness by
standardizing based on:
     1) Sites
     2) Individuals

Permutation procedure that
subsamples the data

Richness: the number of distinct types

Calculate first- and second-order jackknife estimates of species richness

```
> # girdled
> jack1(girdled[6:20], taxa.row = FALSE, abund = TRUE)
>[1] 7.997041
> jack2(girdled[6:20], taxa.row = FALSE, abund = TRUE)
>[1] 7.991115
```

Jackknife (1st order)

$$S_P = S_O + a_1 (N-1)/N$$

Jackknife (2nd order)

$$S_P = S_O + a_1 (2N-3)/N - a_2 (N-2)^2 / N (N-1)$$

Treatment	Observed species richness	First-order jackknife estimate	First-order %	Second-order jackknife estimate	Second-order %
Hemlock	5	5	100	5	100
Girdled	7	7.9	88.6	7.9	88.6
Logged	10	12.9	77.5	13.9	71.9
Hardwood	11	13.9	79.1	15.9	69.1

Diversity: the number of types and their evenness

Many metrics that are used to calculate diversity

1) Shannon index

2) Simpson index

3) Effective number of species

4) Hill numbers

Measures of species diversity

1. Shannon – Wiener Index (also called Shannon Index, Shannon Weaver)

$$H' = -\sum_{i=1}^{n} p_i \ln p_i$$

$p_i = n_i / N$, where n_i is the abundance of the ith species and N is the total abundance.

H' increases as diversity increases
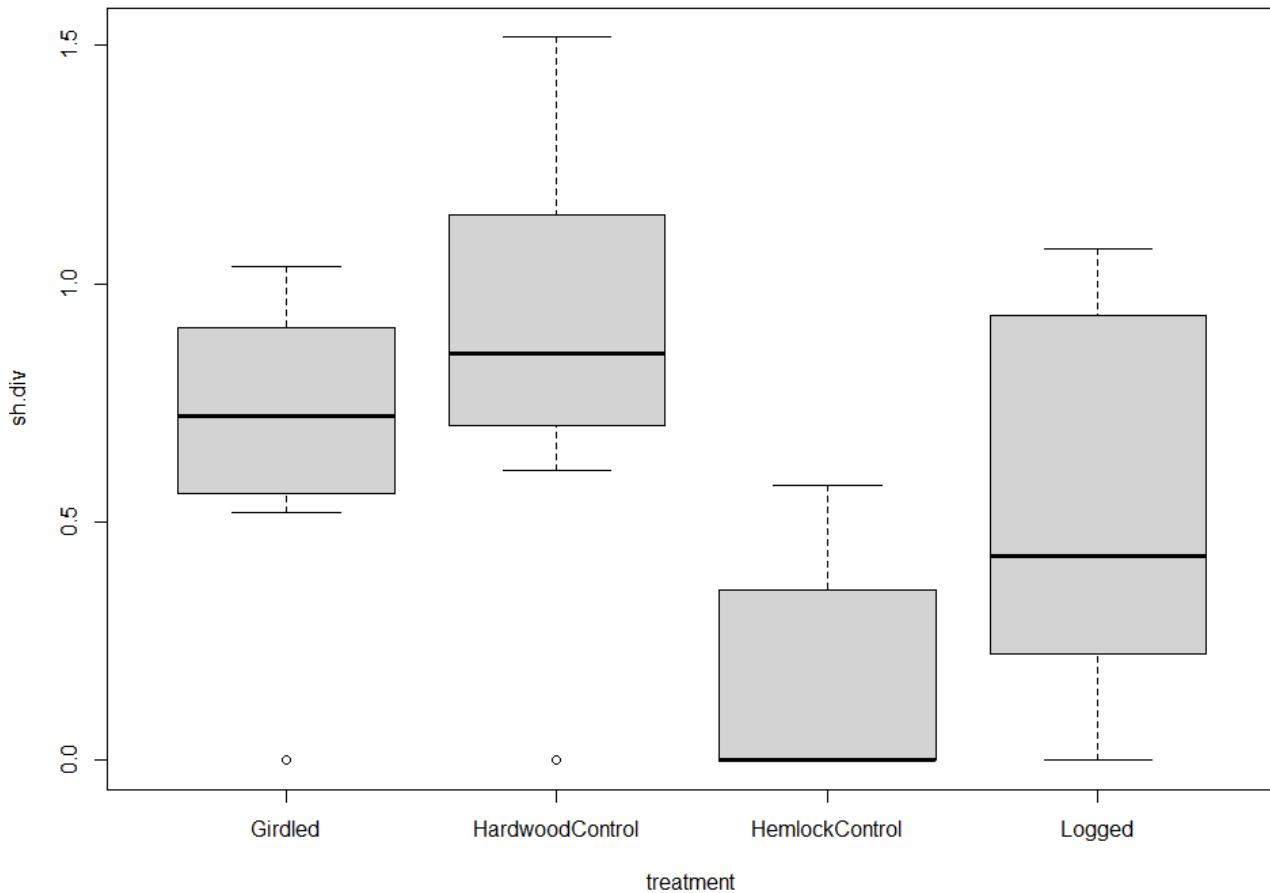
You can compare H' values among sites you sample using the same method and equal effort.

Measures of species diversity

1. Shannon – Wiener Index

```
> # Shannon index
> diversity(ants3[,6:20], index = "shannon")
 [1] 0.4290551 1.0296530 0.0000000 0.2530673 0.9132809 0.6019240 0.0000000 0.9012346 0.0000000 0.5181862
[11] 0.0000000 0.1985152 0.6082961 0.7948526 0.9502705 0.8778248 0.5189139 0.8058275 0.6365142 1.0369939
[21] 0.8367674 1.0720170 0.1914441 0.0000000 0.5763821 0.0000000 0.8278890 1.5169130 0.0000000 1.3418353
> ants3$sh.div <- diversity(ants3[,6:20], index = "shannon")
```

Measures of species diversity

2. Simpsons Diversity Index

Simpson (1949) gave the probability of any two individuals drawn at random from an infinitely large community belonging to the same species as:

$$D = \sum \frac{n_i(n_i - 1)}{N(N - 1)}$$

Gini-Simpson index

$$GS = 1\text{-}D$$

Where ni= the number of individuals in the ith species and N=the total number of individuals.

Note: *D* represents the probability of two individuals belonging to the same species, so diversity actually decreases as *D* increases.

Measures of species diversity

Gini-Simpson index

$$GS = 1 - D$$

R package vegan
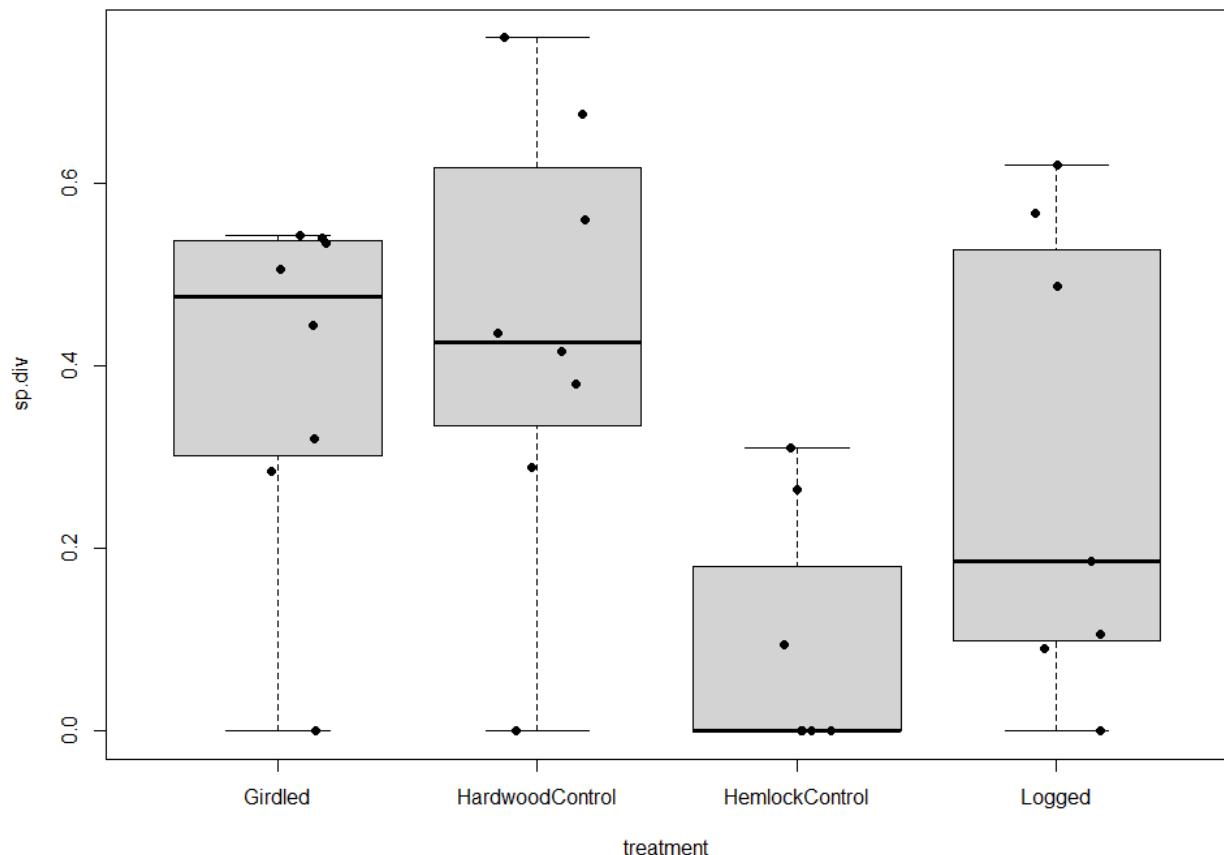
2. Simpsons Diversity Index

```
> # Simpson index
> diversity(ants3[,6:20], index = "simpson")
 [1] 0.18625000 0.62000000 0.00000000 0.10648148 0.54330886 0.32031250 0.00000000 0.53515414 0.00000000
[10] 0.26500000 0.00000000 0.09500000 0.28843750 0.37984173 0.56000000 0.43681519 0.28458050 0.50661626
[19] 0.44444444 0.53979239 0.48692042 0.56804734 0.09070295 0.00000000 0.30991736 0.00000000 0.41632653
[28] 0.75968809 0.00000000 0.67645308
> ants3$sp.div <- diversity(ants3[,6:20], index = "simpson")
```

Measures of species diversity

3. Effective number of species (ENS)

Shannon and Simpson diversity indices are misleading because they are not measured in units of species

Shannon: *bits*
Simpson: *probability*

Each index should be converted into the effective number of species

In a perfectly even community, ENS is equal to species richness, and always smaller than species richness in uneven communities

Entropy and diversity

Lou Jost, Baños, Tungurahua, Ecuador (loujost@yahoo.com).

Effective number of species

- for species richness = S
- for Shannon index = e^H (exponential of Shannon entropy index)
- for Simpson index = $1/D$ (reciprocal of Simpson concentration index)

Measures of species diversity
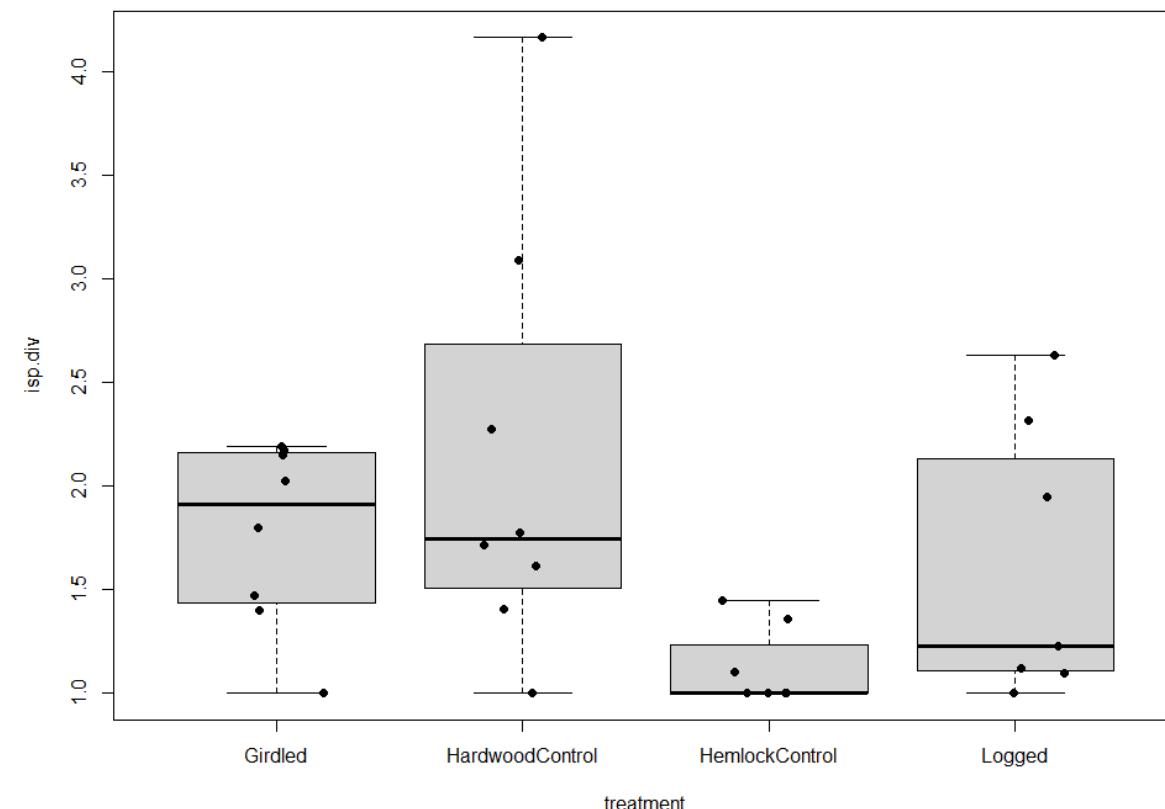
3. Effective numbers of species (ENS)

Effective number of species

- for Simpson index = $1/D$ (reciprocal of Simpson concentration index)

R package vegan

```
> # Effective numbers of species - Inverse Simpson index
> diversity(ants3[,6:20], index = "invsimpson")
 [1] 1.228879 2.631579 1.000000 1.119171 2.189664 1.471264 1.000000 2.151251 1.000000 1.360544 1.000000
[12] 1.104972 1.405358 1.612492 2.272727 1.775616 1.397781 2.026820 1.800000 2.172932 1.949015 2.315068
[23] 1.099751 1.000000 1.449102 1.000000 1.713287 4.161259 1.000000 3.090742
> ants3$isp.div <- diversity(ants3[,6:20], index = "invsimpson")
```

Measures of species diversity

4. Hill numbers

Several common diversity metrics can be united into a single family of generalized entropy metrics
      - Parameterized by variable q

Hill numbers

$$^{q}D = \left(\sum_{i=1}^{S} p_i^{\,q}\right)^{\frac{1}{(1-q)}}$$

For $q = 0$, 1 and 2 (also noted as N_0, N_1 and N_2):

$$^{0}D = S \text{ (species richness)}$$

$$^{1}D = e^{H} \text{ (exponential of Shannon entropy)}$$

$$^{2}D = \frac{1}{D} \text{ (reciprocal of Simpson index)}$$

$q = 0$; species richness – counts all species equally without considering relative abundance

$q = 1$; Shannon – ENS, each species is weighted in proportion to its abundance

$q = 2$; Inverse Simpson – ENS, disproportionately favors abundant species

DIVERSITY AND EVENNESS: A UNIFYING NOTATION
AND ITS CONSEQUENCES[1]

M. O. HILL[2]

School of Plant Biology, University College of North Wales, Bangor, Caernarvonshire, Wales

Measures of species diversity

4. Hill numbers

```
> hill_taxa(ants3[,6:20], q = 0, MARGIN = 1)
 [1] 4 3 1 3 4 3 1 3 1 3 1 2 4 5 3 5 3 3 2 4 5 4 2 1 3 1 5 6 1 6
> hill_taxa(ants3[,6:20], q = 1, MARGIN = 1)
 [1] 1.535806 2.800094 1.000000 1.287970 2.492487 1.825628 1.000000 2.462642 1.000000
[10] 1.678980 1.000000 1.219591 1.837298 2.214115 2.586409 2.405661 1.680202 2.238548
[19] 1.889882 2.820725 2.308891 2.921266 1.210997 1.000000 1.779588 1.000000 2.288483
[28] 4.558132 1.000000 3.826059
> hill_taxa(ants3[,6:20], q = 2, MARGIN = 1)
 [1] 1.228879 2.631579 1.000000 1.119171 2.189664 1.471264 1.000000 2.151251 1.000000
[10] 1.360544 1.000000 1.104972 1.405358 1.612492 2.272727 1.775616 1.397781 2.026820
[19] 1.800000 2.172932 1.949015 2.315068 1.099751 1.000000 1.449102 1.000000 1.713287
[28] 4.161259 1.000000 3.090742
```

$q = 0$; species richness
$q = 1$; Shannon
$q = 2$; Inverse Simpson
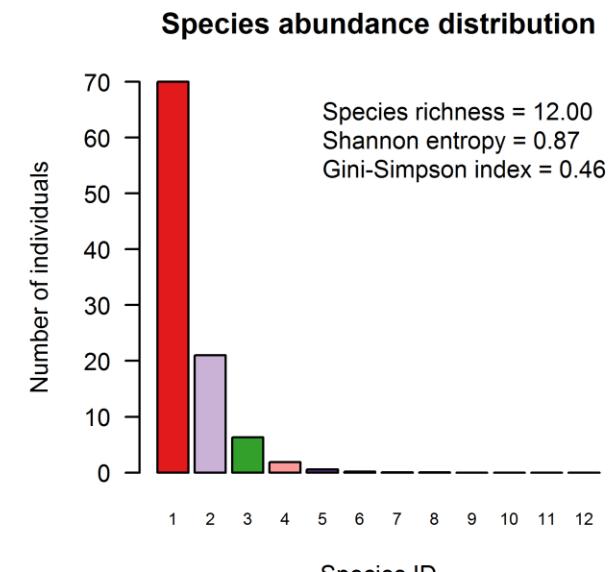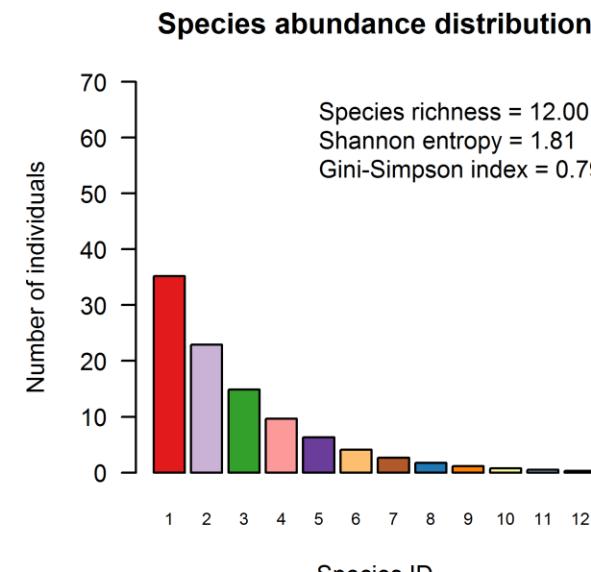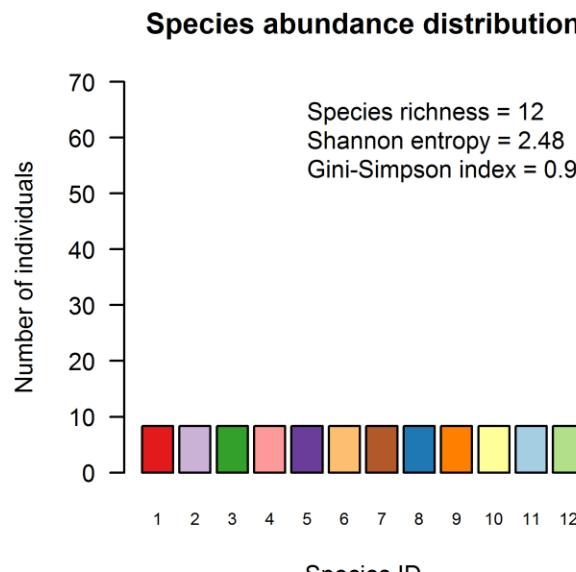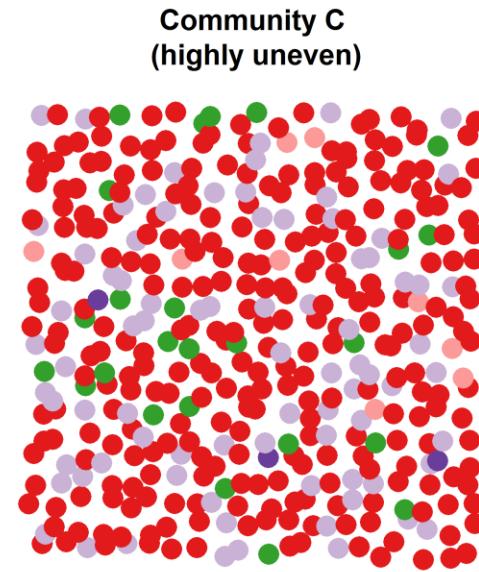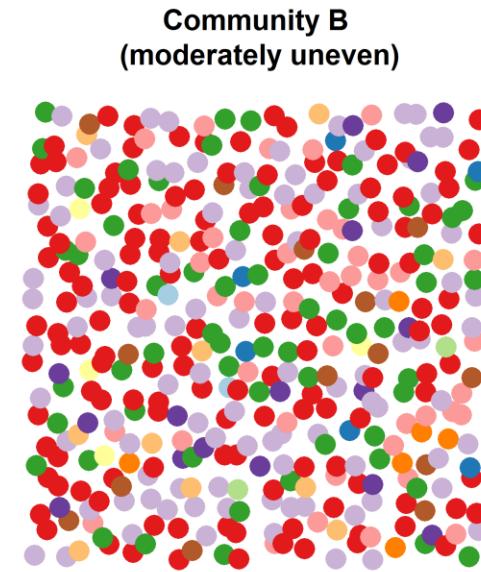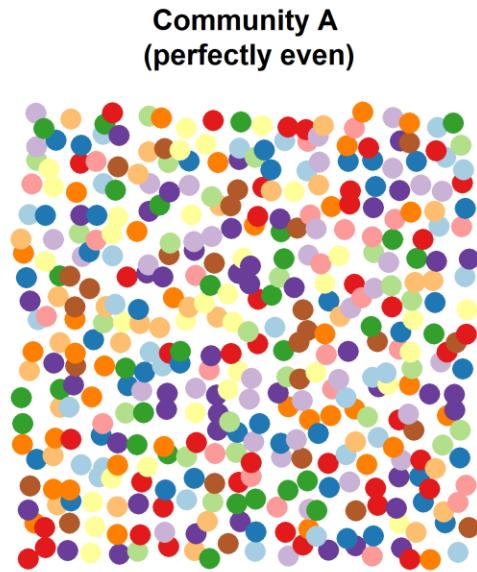
Evenness: the relative abundance of each type

David Zelený

Species evenness

Measure of the relative abundance of species

Commonness vs rarity of species

Pielou's *J* – calculated as a ratio of the Shannon index to the maximum Shannon index (where all species have the same relative abundance)
      - Ranges from 0 – 1

Equitability – calculated as Simpson ENS (or inverse Simpson) divided by the observed number of species

Shannon's evenness

$$J = \frac{H}{H_{max}} = \frac{H}{\log S}$$

Simpson's evenness
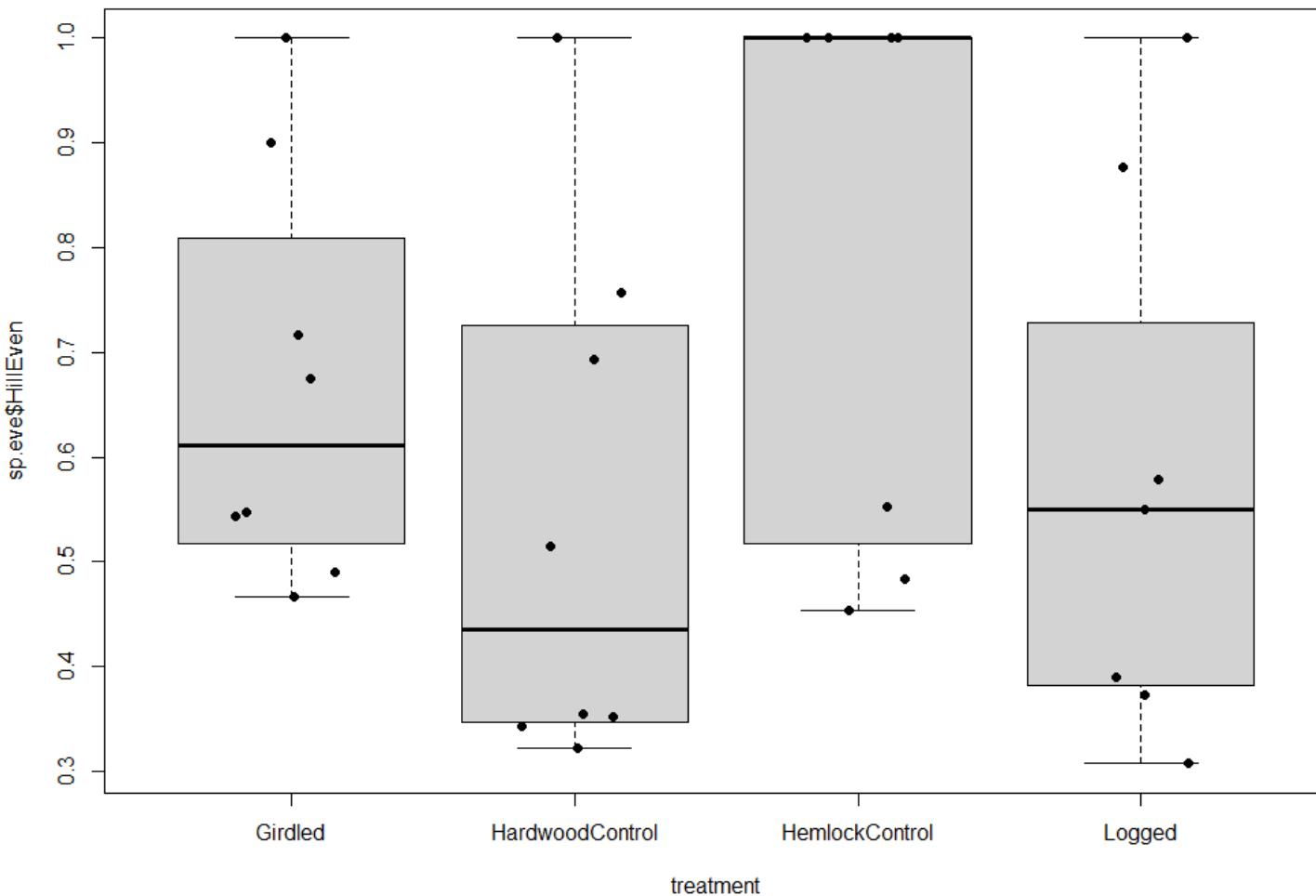
$$equitability = \frac{\frac{1}{D}}{S} = \frac{1}{DS}$$

Species evenness

Measure of the relative abundance of species

R package chemodiv

```
calcDiv(ants3[,6:20], type = "PielouEven")
calcDiv(ants3[,6:20], type = "HillEven", q = 1)
calcDiv(ants3[,6:20], type = "HillEven", q = 2)
```

Although understanding taxonomic metrics are important, there are limitations to these approaches

Context dependencies limit comparison among studies
- Spatial comparisons

Other dimensions of biological diversity that are not reflected in taxonomic approaches

1) Genetic diversity

2) Phylogenetic diversity

3) Functional diversity