Section: Diversity analysis

Diversity indices

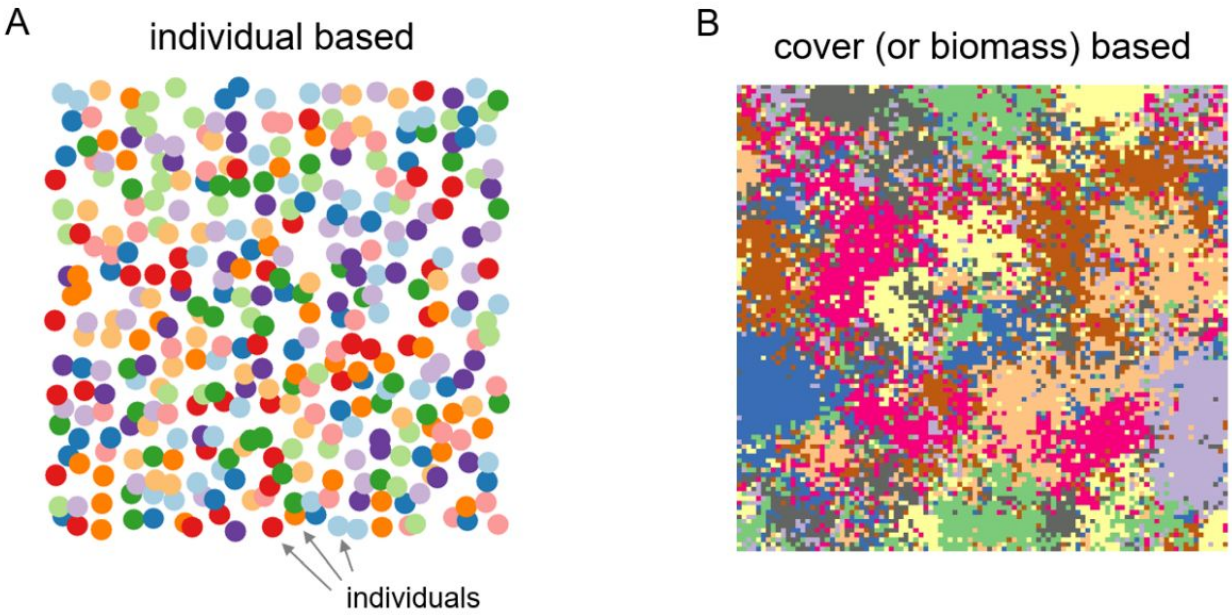
Theory R functions Examples

This section will overview commonly used indices measuring the diversity of ecological communities (species richness, Shannon index, Simpson index). We will also introduce the measures of evenness, the concept of the effective number of species and the general framework of Hill numbers.

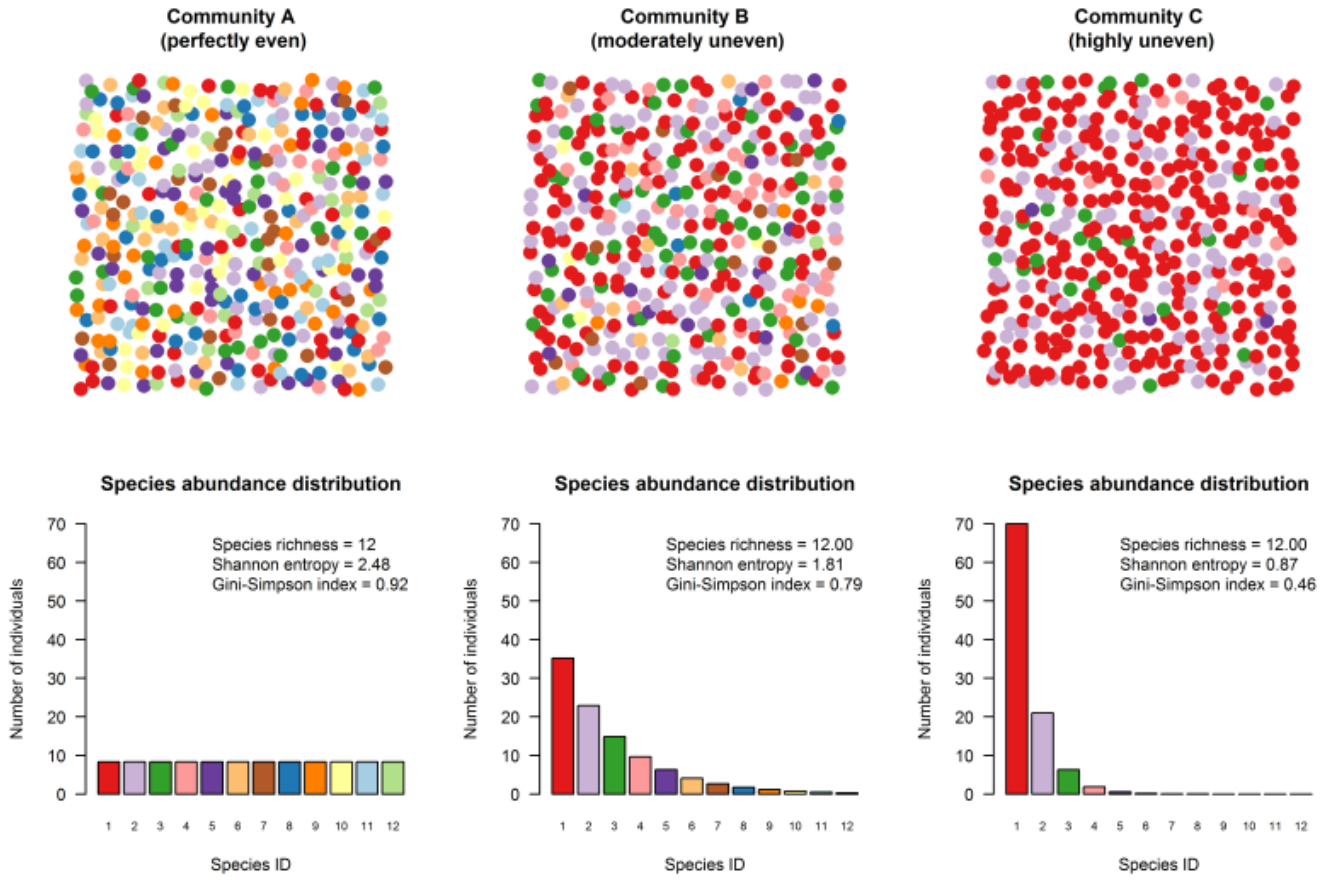
Relative proportion of species in a community

As already mentioned in the Diversity analysis section, diversity consists of two components – species richness, i.e. the number of species in the community, and evenness, i.e. the fact that some species in the community are common and others are rare. An important component of most diversity indices is a value expressing the relative proportion of the species *i* in the community, $\mathbf{p}_i$. This value can be calculated as the absolute (raw) abundance or dominance of the species *i* in the community, divided by the sum of all species abundances/dominances in that community. The sum of relative abundances for all species in the community equals to unity.

There are two main ways in which the *quantity* of a species in a community can be expressed (see figure below): by counting the number of individuals for each species (abundance) or by estimating the biomass of each species, namely it's cover or volume (dominance). The abundance of individuals can be counted for organisms for which individuals are easy to be recognized (most animals, for plants mostly trees), while the dominance of species (cover or volume of their biomass) is usually estimated for those organisms where boundaries between individuals are not clear (mainly plant species that can be clonal). In most cases, the dominance ("biomass-based") values can be treated in the same way as the numbers of individuals and can be imagined as "bits of biomass" randomly collected within a fixed area of the community.



Below is an example of three communities with the same species richness (12 species) but different evenness, from perfectly even (A), moderately uneven (B) and highly uneven (C). In the following examples, we will focus on individual-based species abundances, while acknowledging that the same formulas also work for biomass-based species dominance.



The upper row represents a snapshot of each community, where each circle is an individual, and each colour is a different species. The bar plots in the lower row represent the shape of species abundance distribution after 100 individuals were randomly taken from each community; the colour of each bar corresponds with the colour of species, and the height of the bar represents the abundance of species, i.e. how many individuals within the selected set belong to that species; species are ordered from the highest to lowest abundance. The legend within each barplot shows relevant values of diversity indices (species richness, Shannon index and Gini-Simpson index).

Diversity indices

Diversity indices reviewed below differ from each other in the weight they put on either species richness or evenness.

Species richness (denoted as *S* here) is the most intuitive and natural index of diversity, and I bet that it is used the most frequently in studies dealing with diversity. However, it is also the most sensitive to the difference in sampling effort since it weights all species equally independent from their relative abundances, i.e. rare species count equally to common species although they are more likely to be undetected.

Shannon index (or Shannon entropy[2)], Shannon-Wiener or (incorrectly) Shannon-Weaver; denoted as H, H' or H_{Sh}) considers both species richness and evenness. The index is derived from information theory and represents the uncertainty with which we can predict which species will be one randomly selected individual in the community. If a community contains only one species, the uncertainty is zero since we are sure that the randomly chosen individual will belong to that one species only. The more species the community contains, the higher the uncertainty; in a diverse community, we are unlikely to guess which species will be the randomly chosen individual. However, if the community has many

Shannon index

$$H = -\sum_{i=1}^{S} p_i \log p_i \quad H_{\max} = \log S$$ [1)]

where
S = species richness,
p_i = relative abundance of species i,
log = usually natural logarithm (i.e.

species, but only one (or a few) prevails (many individuals of one or a few species), uncertainty will not be so high since we have a high probability that the randomly selected individual will be the most abundant species. This is why the Shannon index increases with richness and evenness, and it puts more weight on richness than on evenness.

log_e or *ln*)

In real ecological data, values of *H* are usually between 1.5-3.5 (the units are *bits* of information); note that the absolute value of the *H* depends on the base of the logarithm used for the calculation (usually $\log_e$, where *e* = 2.718). The maximum value of *H* index (H_{max}) for the community of given richness occurs at a situation where it is perfectly even (all species have the same relative proportion).

Simpson index (also Simpson concentration index, denoted as *D*, H_S or *λ*) also considers both richness and evenness, but compared to Shannon, it is more influenced by evenness than richness. It represents the probability that two randomly selected individuals will be of the same species. Since this probability decreases with increasing species richness of the community, the Simpson index also decreases with richness, which is not too intuitive. For that reason, more meaningful is to use the **Gini-Simpson** index, which is simply the 1-Simpson index, and which increases with the increasing richness of the community (it is identical to Hurlbert's *probability of interspecific encounter*, PIE).

The values of *D* are in the range between 0 and 1, and the units are probability. When the species richness of the community exceeds 10, the values of the Simpson index are mostly influenced by evenness.

Simpson index

$$D = \sum_{i=1}^{S} {p_i}^2$$

$$D_{\max} = \frac{1}{S}$$ [3)]

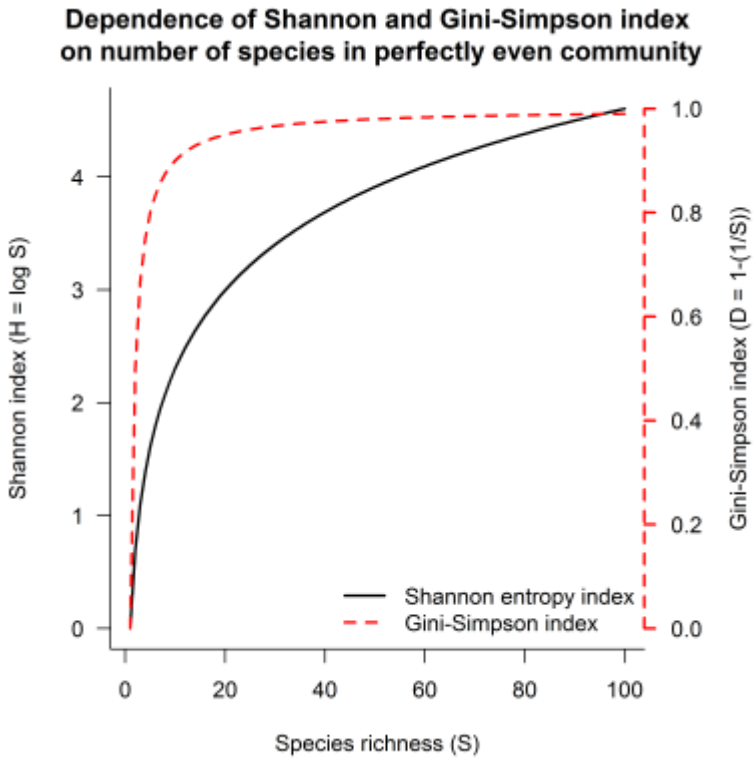
Gini-Simpson index

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

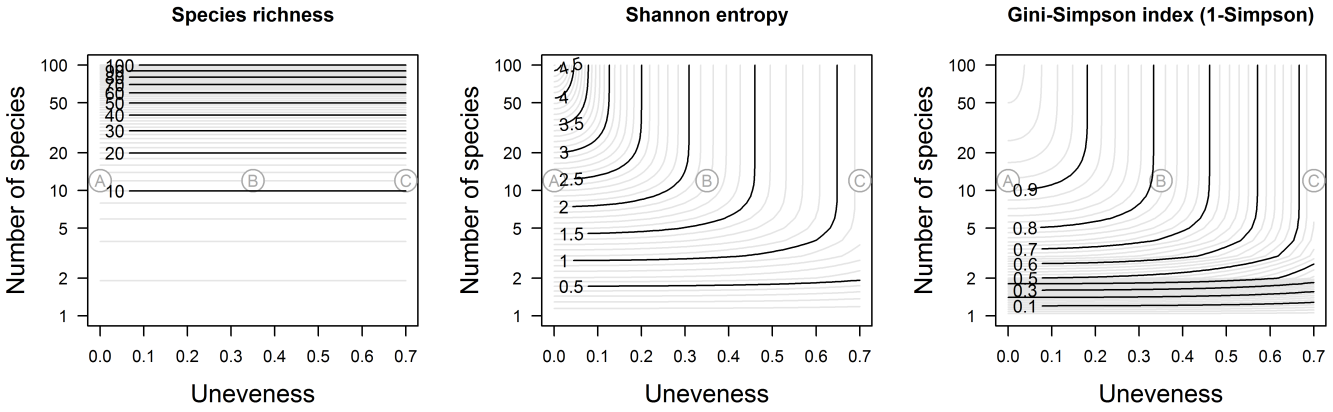
where
S = species richness,
p_i = relative abundance of species *i*,

Comparison of species richness, Shannon index and Simpson index

Below are two examples comparing species richness, Shannon entropy and Gini-Simpson concentration index to each other. In the case of perfectly even communities, the Shannon and Gini-Simpson index (when projected against the species richness of the same community) increases non-linearly with the number of species in the community; the Gini-Simpson index increases faster than Shannon entropy. This relationship also illustrates that the Gini-Simpson index changes very fast in low species richness values (0.5 for *S* = 2, 0.67 for *S* = 3, 0.75 for *S* = 4, ... 0.9 for *S* = 10), and with richness over 10, it changes much slower (0.95 for *S* = 20 and 0.99 for *S* = 100).



If projected against species richness and evenness, the patterns for each of the three indices have rather different shapes (see figure below; evenness is displayed in an inverted manner, as unevenness). The labels A, B and C within each figure represent scenarios for perfectly even community (A), moderately uneven (B) and highly uneven (C).



Evenness

Evenness is a synthetic measure describing the pattern of relative species abundances in a community. There are many ways how evenness can be calculated; here, I mention just two common ones, one derived from Shannon and the other from the Simpson index.

Shannon's evenness

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

Shannon's evenness (also called Pielou's *J*) is calculated as a ratio of the Shannon index calculated from the real community (with *S* species and p_1, p_{2i}, p_3....p_i relative species abundances), and maximum Shannon index for the community with the same richness H_{max} (i.e. with *S* species all having $p_1 = p_2 = p_i = 1/S$). The value is 1 in case all species have the same relative abundances and decreases with increasing differences in species relative

abundances in the community.

Simpson's evenness

$$equitability = \frac{\frac{1}{D}}{S} = \frac{1}{DS}$$

Simpson's evenness (called also *equitability*) is calculated from Simpson's effective number of species divided by the observed number of species. The effective number of species (ENS) is the number of equally abundant species which would need to be in a community so that it has the same Simpson's index as the one really calculated (more about the concept of *effective number of species* below). In the case of Simpson's *D*, the effective number of species is simply 1/*D*.

Effective numbers of species (ENS)

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)

Lou Jost (2002) argued that to call Shannon and Simpson (or Ginni-Simpson, respectively) indices as *diversity* is misleading since diversity should be measured in intuitive units of *species*, while each of the two indices has different units (Shannon has *bits* and Simpson *probability*)[4]. This problem can be overcome by introducing the concept of the `effective number of species` (ENS, MacArthur 1965), i.e., the number of species in an equivalent community (i.e., one with the same diversity index as the community in question) composed of equally abundant species. In the case of a perfectly even community, ENS is equal to species richness, while for uneven communities, ENS is always smaller than species richness. Each of the indices above can be converted into the effective number of species following a simple formula.

Hill numbers

Hill numbers

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

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

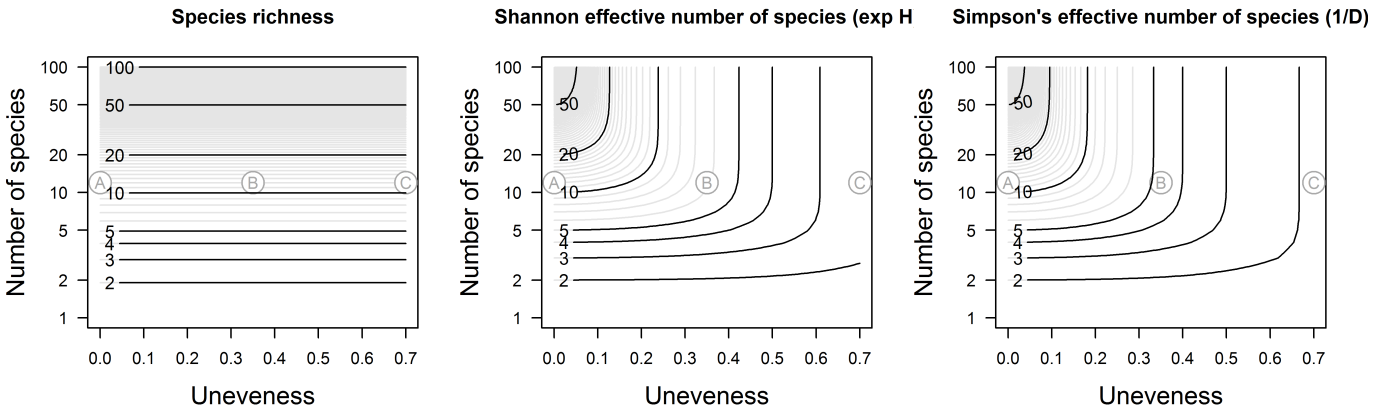
$${}^0D = S$$ (species richness)

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

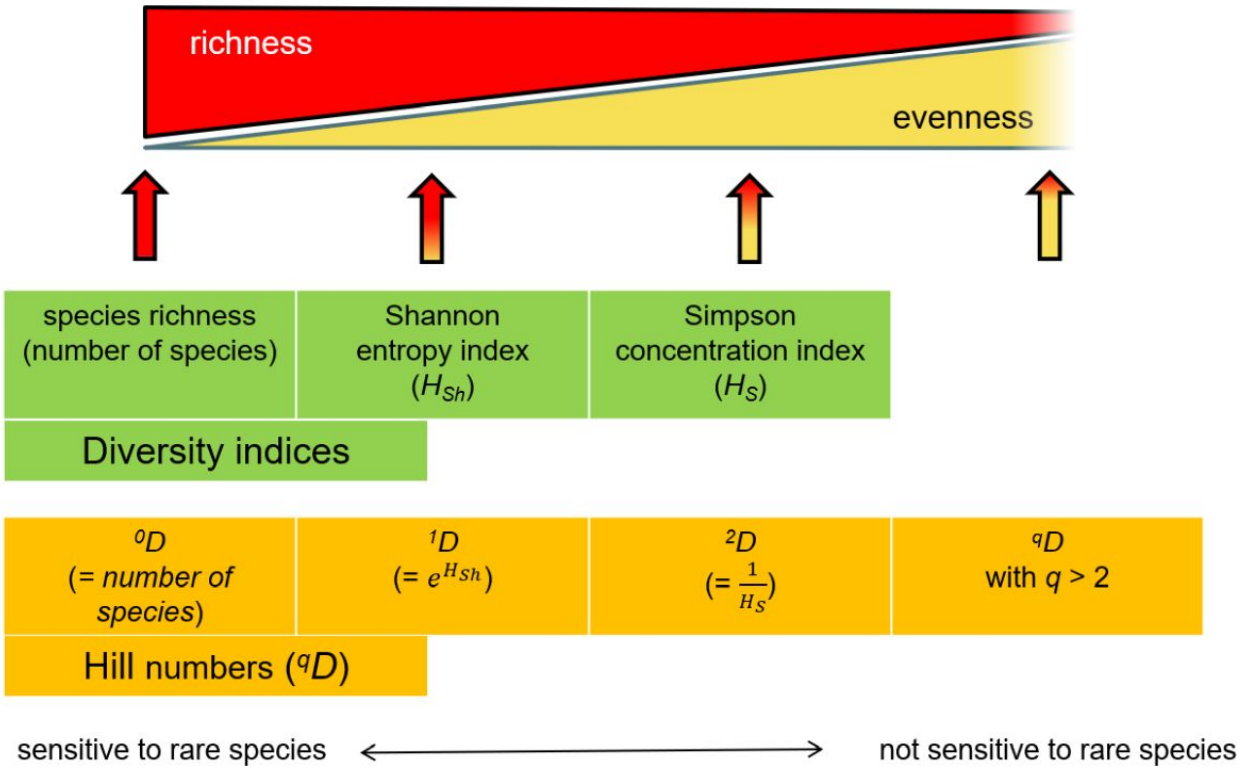
$${}^2D = \frac{1}{D}$$ (reciprocal of Simpson index)

Mark Hill (British scientist also known for introducing detrended correspondence analysis (DCA) and TWINSPAN, and for recalibrating Ellenberg species indicator values for Britain) realized that species richness, Shannon entropy, and Simpson's concentration index are all members of the same family of diversity indices, later called Hill numbers. Individual Hill numbers differ by the parameter q, which quantifies how much the measure discounts rare species when calculating diversity. Hill number for q = 0 is simply **species richness**, counting all species equally without considering their relative abundance. For q = 1, it is **Shannon diversity**[5], i.e. ENS derived from Shannon entropy, where each individual is counted equally and each species is weighted in proportion to its abundance, focusing on common and abundant species. For q = 2, it is **Simpson diversity**, i.e., ENS for Simpson concentration index, which disproportionately favours individuals of abundant species and represents the number of very abundant species. For q > 0, indices discount rare species, while for q < 0, the indices discount common species and focus on the number of rare species (which is usually not ecologically meaningful).

The dependence of species richness, Shannon diversity (effective number of species based on Shannon entropy index), and Simpson's diversity (effective number of species based on Simpson's index) on (un)evenness and diversity is illustrated below (labels A, B, and C represent even, moderately uneven, and highly uneven communities, respectively).

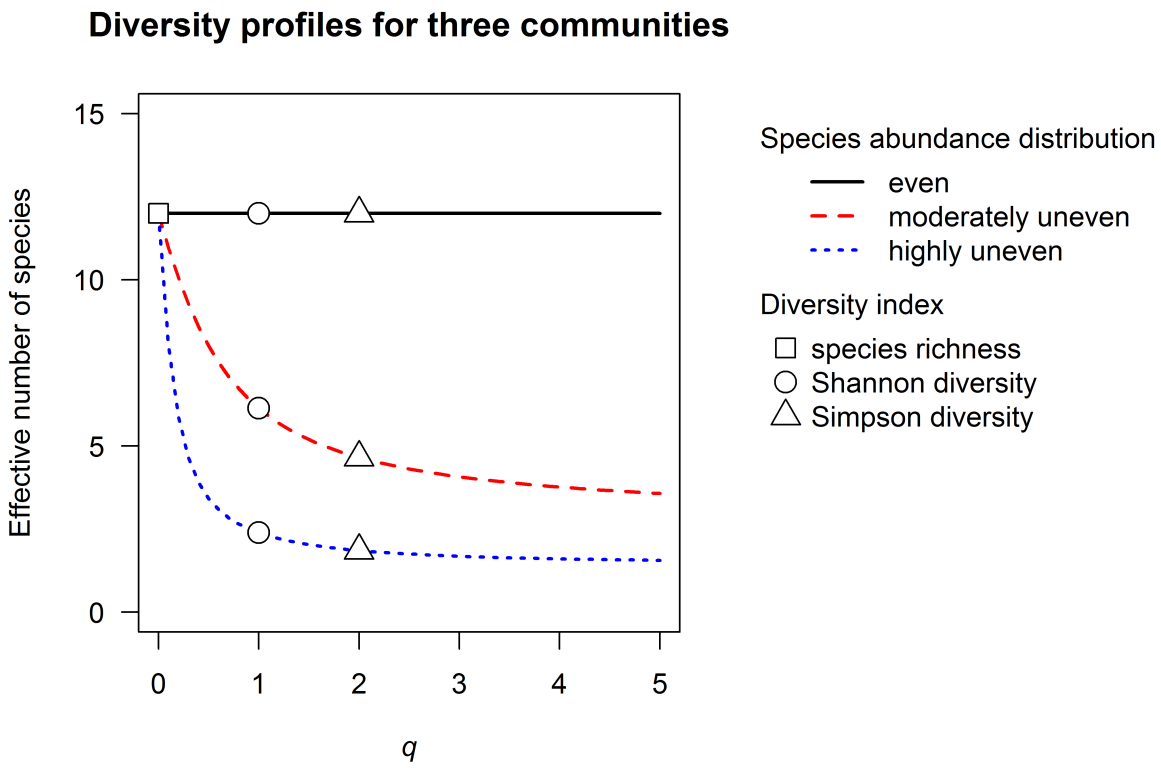


The schema below visualizes the whole concept connecting diversity indices with Hill numbers of diversity. Increasing the value of parameter *q* changes the focus given diversity metric more on evenness and less on richness by ignoring rare species and focusing on common and dominant species.



Diversity profiles

It is possible to draw the effective number of species as a function of coefficient *q* - increasing *q* decreases the impact of rare species on the measure of diversity. The value for $q = 0$ equals to species richness (in the diagram displayed by squares), for $q = 1$ equals to Shannon diversity (circles) and for $q = 2$ Simpson diversity (triangles). The shape of the diversity profile considers the differences in evenness between the three communities; the more is the community species abundance uneven, the faster the curve declines with increasing coefficient *q*. The future will show what new this form of diversity visualization brings.



Summary of values for diversity measures discussed in this chapter

	Community A (perfectly even)	**Community B (moderately uneven)**	**Community C (highly uneven)**
Species richness	12	12	12
Shannon entropy	2.48	1.81	0.87
Simpson index	0.92	0.79	0.46
Shannon evenness	1	0.73	0.35
Simpson evenness	1	0.38	0.15
Shannon diversity (^1D, N1)	12	6.14	2.39
Simpson diversity (^2D, N2)	12	4.66	1.86

1)

Why log (S)? H = $-\sum p_i \log p_i$. In case of S equally abundant species, each $p_i = 1/S$. Then, H = $-\sum 1/S \log 1/S = - S * 1/S * \log (1/S) = - \log (1/S) = - \log S^{-1} = - (- \log S) = \log S$.

2)

Entropy of the system represents the uncertainty, expected measure of surprise.

3)

Why 1/S? $D = \sum p_i^2$. In case of S equally abundant species, each $p_i = 1/S$. Then $D = \sum 1/S^2 = S*1/S^2 = S/S^2 = 1/S$.

4)

Jost (2002) argues: “The radius of a sphere is an index of its volume but is not itself the volume, and using the radius in place of the volume in engineering equations will give dangerously misleading results. This is what biologists have done with diversity indices. The most common diversity measure, the Shannon-Wiener index, is entropy, giving the uncertainty in the outcome of a sampling process... Entropies are reasonable indices of diversity, but this is no reason to claim that entropy *is* diversity.”

5)

In fact Hill's formula is not defined for $q = 1$, but it can be shown that when q approaches 1 from below or above, the index gets equal to exponential Shannon.

From:
https://anadat-r.davidzeleny.net/ - **Analysis of community ecology data in R**

Permanent link:
https://anadat-r.davidzeleny.net/doku.php/en:div-ind

Last update: **2026/05/24 17:14**