Week 2 Exercise: Local diversity

Z620: Quantitative Biodiversity, Indiana University

November 8, 2014

Overview

In this exercise, we will use R and RStudio to explore aspects of local or site-specific diversity, also known as alpha (α) diversity. We will use the tropical forest dataset from Barro-Colorado Island (BCI), a 1,560-hectare island in the center of the Panama Canal that is administered by the Smithsonian. After setting up our working environment, we will explore the BCI dataset and ways to quantify and examine the α-diversity within individual plots of tropical trees.

RETRIEVE AND SET YOUR WORKING DIRECTORY

```
## [1] "/Users/lisalocey/GitHub/Quantitative_Biodiversity/Assignments/Week2"
```

INSTALL PACKAGES

Ecologists have developed many packages for conducting quantitative analyses in R. You can use the 'help' function to learn more about package installation and R add-ons. In this class, we will often use the package called 'vegan', which contains tools commonly used in ecological research, including analyses of (α)-diversity. To install vegan into your R environment type: install.packages("vegan") into your RStudio console.

```r
require("vegan") #You can also type: library("vegan"). The difference being that "require"
```

```
## Loading required package: vegan
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.0-10
```

```r
#loads additional packages (dependencies) that vegan requires for some analyses.
```

LOADING BCI FROM VEGAN

Conveniently, the vegan library contains a version of the BCI dataset, which is obtained by censusing the trees of 50 1-hectare plots. More information on the BCI forest dataset can be found here: http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci/.

```r
data(BCI) # Loading the BCI dataset
```

EXPLORING THE SITE-BY-SPECIES MATRIX

In vegan, the BCI data are organized into a **site-by-species matrix**, that is, a data table where the abundances of all 225 tree species (columns) are given for each of the 50 sites (rows). Consequently, this table should have 50 rows and 225 columns. Let's verify this using the dimension or "dim" function:

```r
dim(BCI)  # Prints the dimensions of the site (row) by species (column) matrix
```

```
## [1]  50 225
```

The **site-by-species matrix** is one of the most basic data structures used by ecologists and contains the abundances (or presence-absence) of all species found among a given number of sites, e.g., quadrats, transects, etc. With the BCI site-by-species matrix loaded, let's print the abundances of 4 species found in the first 8 sites.

```r
BCI[1:8, 14:17] # print abundances for sites (rows) 1 to 8 and for species (columns) 14 to 17
```

```
##   Andira.inermis Annona.spraguei Apeiba.aspera Apeiba.tibourbou
## 1              0               1            13                2
## 2              0               0            12                0
## 3              0               1             6                1
## 4              0               0             3                1
## 5              1               0             4                0
## 6              1               0            10                0
## 7              0               0             5                0
## 8              0               1             4                1
```

Here, we can see that Cabbage Bark (*Andira inermis*) is absent from six of the eight sites and is only found as a single individual in two sites. On ther other hand, *Apeiba aspera* (locally known as Monkey Comb) is found at all eight sites and is relatively much more abundant than any of the other three species that we selected.

EXPLORING LOCAL (α) DIVERSITY

So far, you have been introduced to an important data structure, the site-by-species matrix. You have learned to print its dimensions and to print the abundances of particular species (columns) found in specific sites (rows). Now, let's focus on a specific site by selecting the first row and assigning it to a variable called "Site1":

```r
Site1 <- BCI[1,] # assign the first row (site) to the variable Site1
dim(Site1) # print the dimensions of Site1 to the screen
```

```
## [1]   1 225
```

You can see that Site1 has one row (it is only 1 site) and 225 columns. But remember that the BCI site-by-species matrix also accounts for absences, i.e., with zeros. Let's find out how many species actually occupied Site1 when it was censused:

```r
S1 <- specnumber(Site1) # Find the number of species in Site1 and assign it to a varible S1
cat('There are', S1, 'species at Site1') # "cat()" concatenates strings and numbers
```

```
## There are 93 species at Site1
```

Now we know that the recorded species richness (usually denoted as S) of Site1 is 93. Local species richness is the most basic aspect of diversity and is, in fact, what we usually refer to as α-diversity. Beyond knowing the number of species, we are often interested in knowing how many individuals were found among the, say, 93 species.

```r
N1 <- sum(Site1) # find the number of individuals at Site1 and assign it to a variable N1
cat('There are', N1, 'individuals among', S1, 'species at Site1')
```

```
## There are 448 individuals among 93 species at Site1
```

Now we know the recorded species richness and recorded total abundance (usually denoted as N) of Site1 in the BCI site-by-species matrix. N and S are two of the most common pieces of data we obtain when sampling ecological communities and assemblages. Finally, we are beginning to quantify our data!
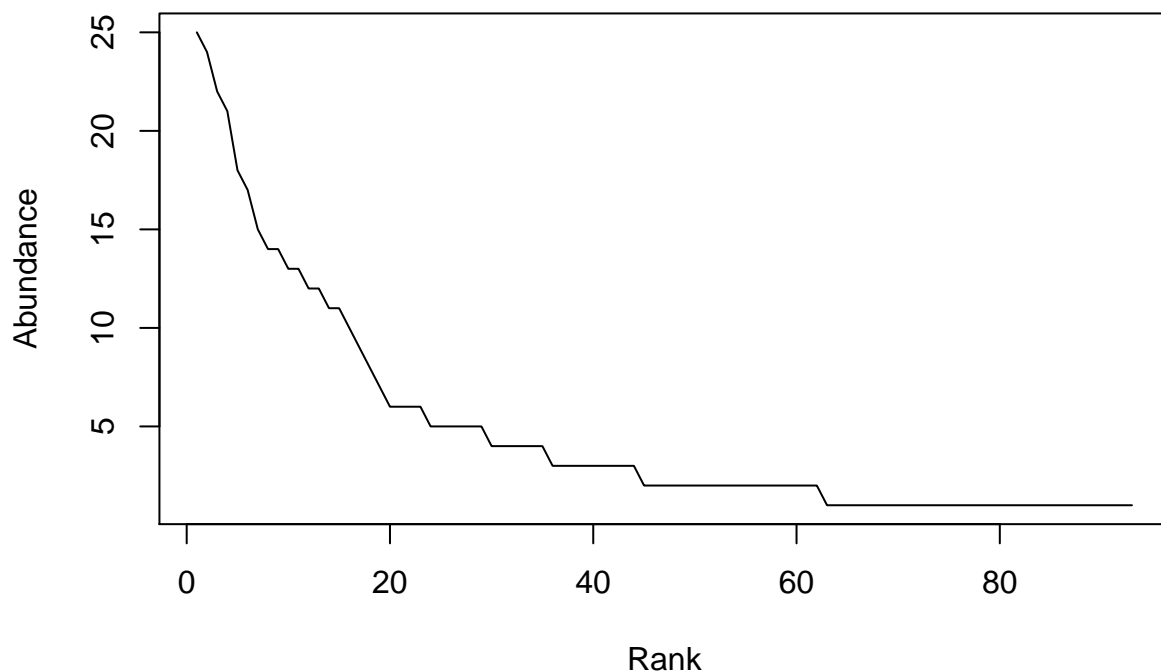
The species abundance distribution

A next natural step in exploring aspects of local diversity is simply to ask how abundance varies among species. One of the most common ways to visually explore this relationship, commonly known as the species abundance distribution (SAD) is simply to rank the species from greatest to least, that is, as a rank-abundance curve (RAC). Let's plot the RAC for Site1.

Let's begin by first removing the zeros from Site1. This removes the species that, while in the BCI site-by-species matrix, were not found at Site1.

```r
Site1 <- Site1[ !Site1 %in% c(0) ] # Remove all zeros from Site1 and reassign the new vector
```

Now, we can plot the RAC, accounting only for the species that were found there.

```r
RAC <- Site1[order(Site1, decreasing=TRUE)] # Rank the taxa by abundance (y-values)
ranks <- seq(1, ncol(RAC)) # Create a sequence of S ranks (x-values)
plot(ranks, RAC, type='l', xlab="Rank", ylab="Abundance") # plot the RAC
```

Looking at the RAC for Site1, we can see that abundance is distributed very unevenly among species. In fact, few species have more than 10 individuals and most species have less than 5. This sort of uneven distribution of abundance is one of the most ubiquitous patterns in ecology and has provoked a long history of study, theories, and explanations. Now, you might ask whether all sites in the BCI site-by-species matrix have similarly uneven RACs. But, before moving on to other sites, let's explore aspects of local diversity at Site1 more deeply, that is, by quantifying evenness, diversity, dominance, and rarity at Site1.

Species Evenness

Species Diversity

Dominance and Rarity

Ken hasn't touched from here on down

Calculate Shannon index. 'Margin = 1' means diversity is calculated row-wise; 'Margin = 2' means diversity is calcualted column-wise. With base = exp(1), we are esimating Shannon's index using the natural logrithm of each taxon's relative abundance using the equation; $H' = -\sum(p_i \times ln(p_i))$
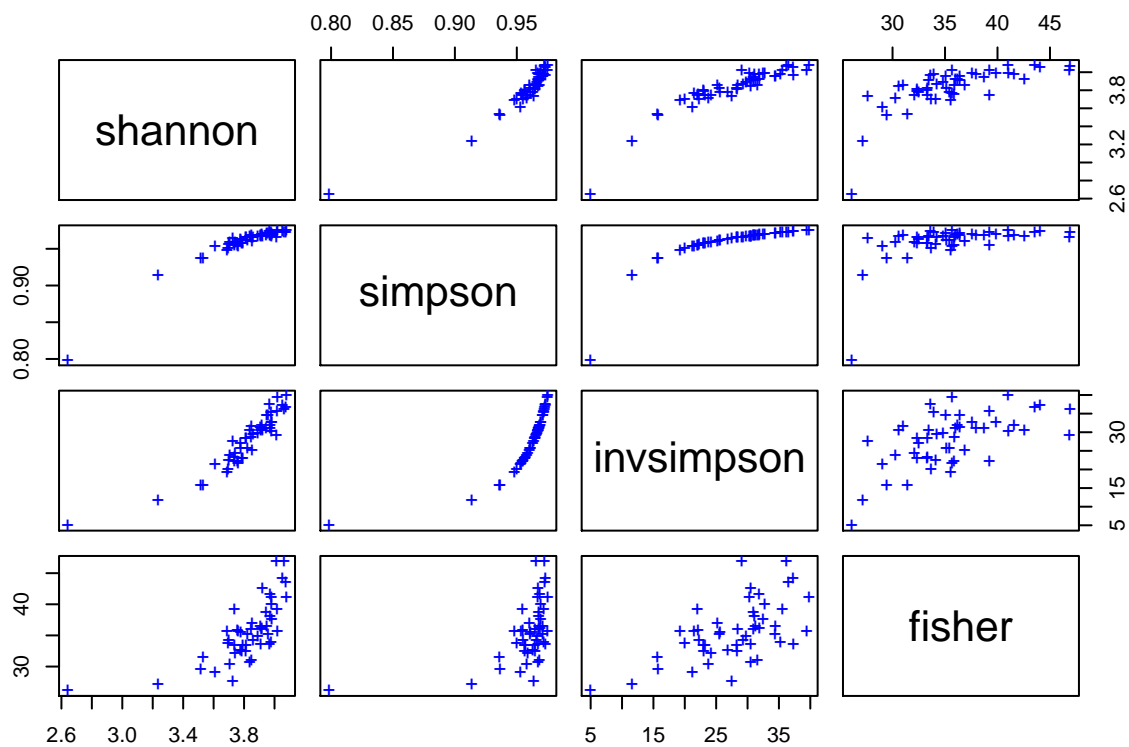
```
shannon <- diversity(BCI, index="shannon", MARGIN = 1, base=exp(1))
```

Can also calculate other diversity metrics

```
simpson <- diversity(BCI, "simpson")
invsimpson <- diversity(BCI, "inv")
fisher <- fisher.alpha(BCI)
```

Let's plot ***Pairs is not a normal plotting tool, introduce first***

```
pairs(cbind(shannon, simpson, invsimpson, fisher), pch="+", col="blue")
```
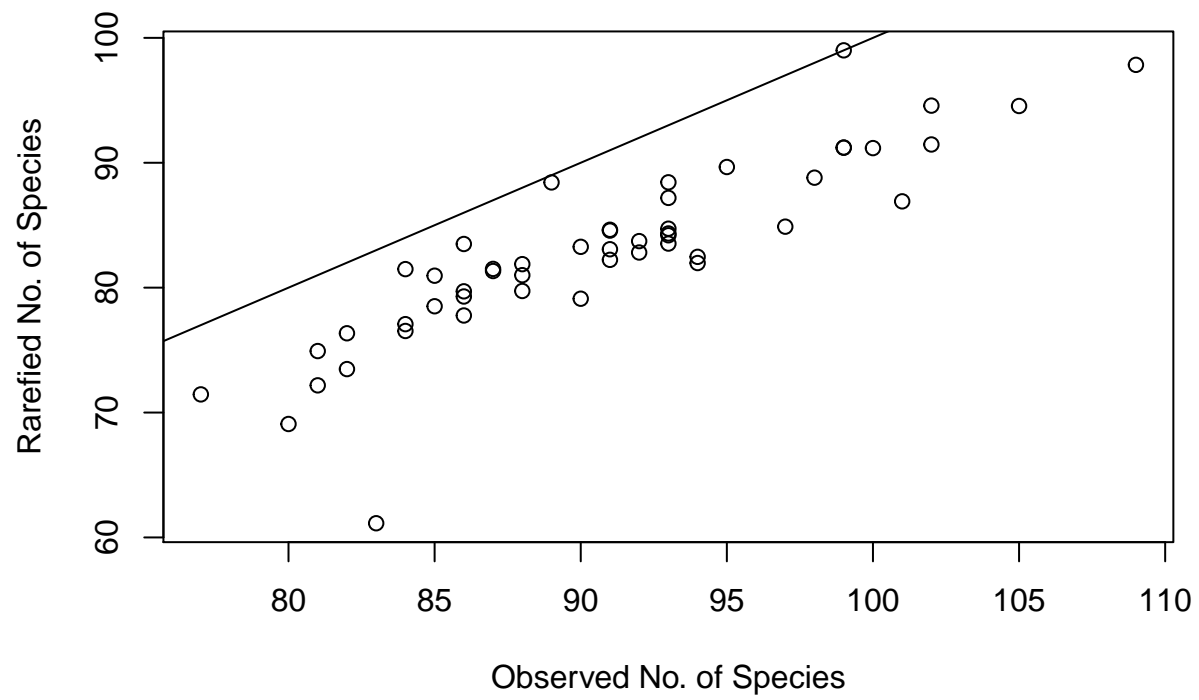
Species richness

```r
S <- specnumber(BCI)
```
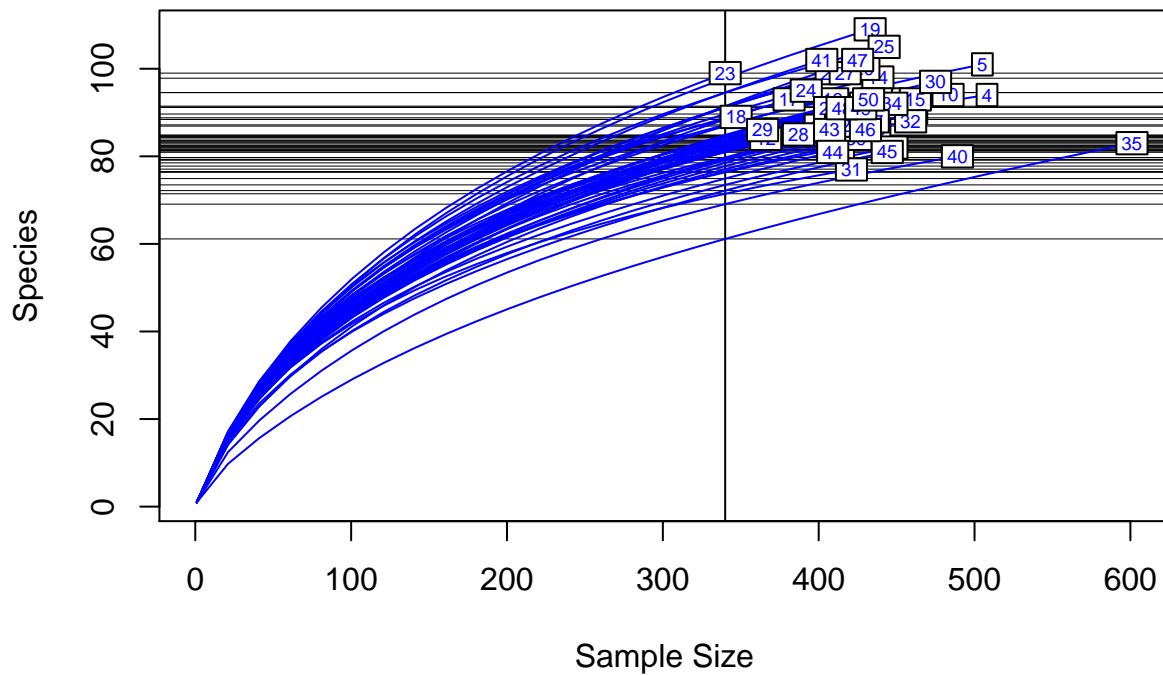
Rarefaction

```r
(raremax <- min(rowSums(BCI)))
```

```
## [1] 340
```

```r
Srare <- rarefy(BCI, raremax)
plot(S, Srare, xlab = "Observed No. of Species", ylab = "Rarefied No. of Species")
abline(0, 1)
```

```
rarecurve(BCI, step = 20, sample = raremax, col = "blue", cex = 0.6)
```

Example: http://www.jennajacobs.org/R/rarefaction.html

Calculating relative abundances

```
BCI_t <- t(BCI)

BCIrel <- BCI_t
for(i in 1:ncol(BCI_t)){
  BCIrel[,i] = BCI_t[,i] / sum(BCI_t[,i])
  }
```

What's one way to test that this worked?

```
colSums(BCIrel)
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
```