

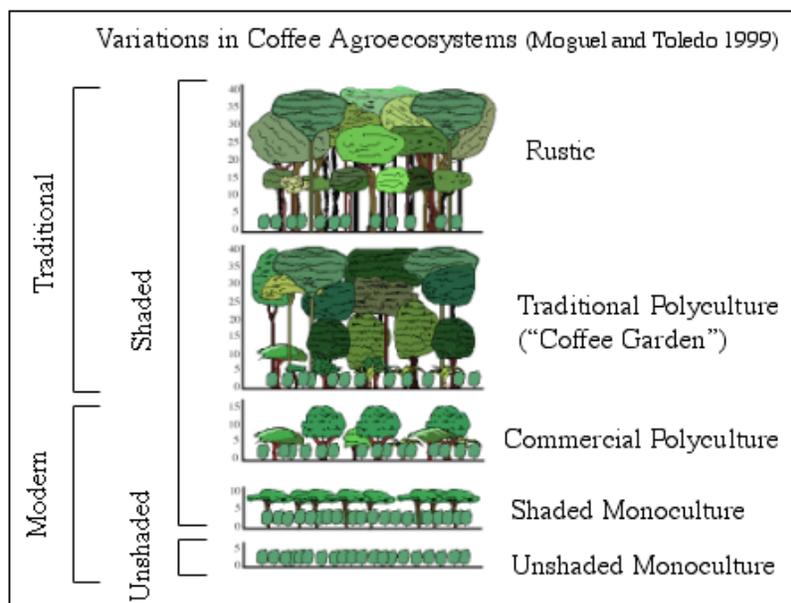
Measuring biological diversity in R

Dan Flynn, Brenda Lin, and Daniel Bunker

18 March 2008
Columbia University

Understanding biodiversity and the ecosystem processes that come with said biodiversity can help inform policy toward sustainable development. Land use changes have affected the habitat and feeding resources of many taxa, thereby changing species composition and abundances. This change in land use and species distribution may also lead to changes in ecosystem services and functions that are necessary for human well-being.

In this lab, we will look at changes within bird species diversity as land is converted from forest to more intensive coffee agriculture. Coffee is a crop that is grown within a large variation of shade cover. On the one extreme is traditional rustic coffee, which is planted under a forest canopy, and on the other extreme is intensive coffee agriculture, which has little or no shade cover (Figure 1). There is a large range of shade systems in between these two extremes. The conversion of coffee farms from traditional rustic systems to intensive, unshaded monocultures is presumed to have large effects on bird biodiversity because of the loss of vegetation diversity and structure in sun monocultures. This may be especially important for bird migration between North and South America because shade trees within traditional coffee farms have served as potential sources of habitat and food during previous migration, and declines in certain bird species have been blamed on the gradual removal of shade trees in Central American coffee systems.



The data you will be using in this lab looks at the bird species found in forest, rustic coffee, polyculture, and shaded monoculture habitats. Along with the bird data, you will be given descriptive vegetation data for each of these habitats. The data comes from a study looking at shade coffee certification and its effectiveness to offer potential conservation significance for bird species. For more information, please see:

Mas, A.H. and T.V. Dietsch. 2004. Linking shade coffee certification to biodiversity conservation: butterflies and birds in Chiapas, Mexico. *Ecological Application* 14(3): 642-654.

The data comes from a region in Southern Chiapas, Mexico rich in coffee farms. Many farms are managed in more than one manner, and intensively managed farms can be next to traditionally managed farms or even forest. Because birds have large ranges, landscape rather than local level changes in management may be more important for bird diversity. It may be important to note that farms in group A lie in a separate watershed from farms in group B.



Using these data, we will explore various questions regarding biodiversity measurements across managed landscapes. We will use the R statistical software package that you downloaded this past week to compare species richness and evenness among the field sites, to compare similarity of species among habitat types, and to gain an understanding of how land management can affect diversity.

Now let's get started!

Installing packages

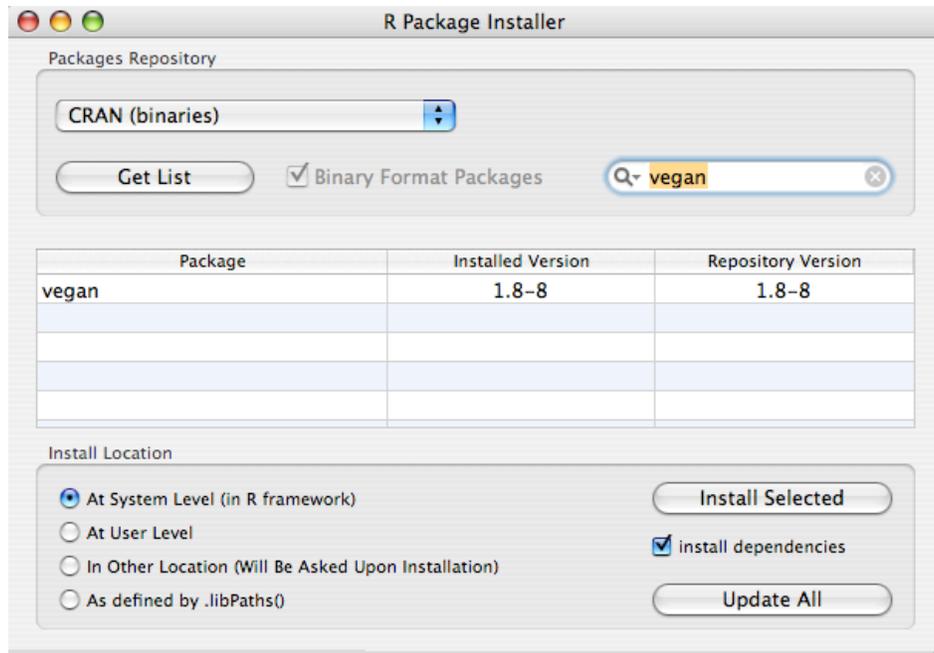
In order to do the next steps, we will use a *package* of functions which a group of ecologists have written for R. Packages are sets of tools usually written for one particular purpose, and are maintained by users. The package that we will use is called "vegan", for "vegetation analysis". It has many functions related to diversity measurement and community ecology data.

The process of installing a package is slightly different for Mac and PC users. Please follow the appropriate instructions below:

Mac

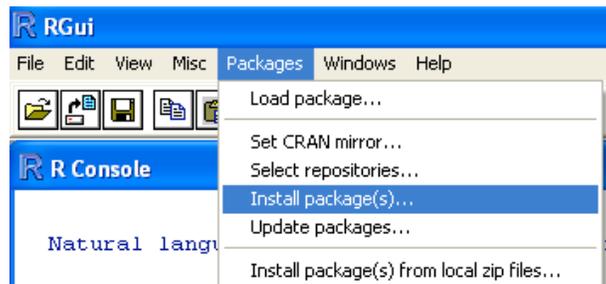
Open up R, and go to Packages & Data > Package Installer.

In the search box, type in "vegan", and click "Get list". After the search is complete, select vegan and "Install Selected".



PC

Open up R and go to Packages > Install packages.



The next window that comes up asks you to select the appropriate "mirror"; these are just servers with copies of all the associated R software. Choose any mirror; ones closer to NYC will be slightly faster.

Search for vegan, and click install.

Diversity indices

There are many different ways to count up diversity in biological communities. The simplest way is to just add up the number of species present; this is referred to as "species richness", or just "S". But this method means that two very different communities might be identical, since it ignores both the *identity* of the species and their *abundances*. For example, examine the table below. Using species richness, both communities are the same; $S = 5$ for both.

Species	Community I	Community II
A	10	5
B	1	5
C	1	5
D	1	5
E	1	5

We want to also consider the differences in abundance of each species. For this, we will use the **Shannon-Wiener diversity index**, or H' . This index is borrowed from information science, and is calculated as follows:

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

Where p_i is the *relative abundance* of species i , S is the total number of species present and \ln is the natural log. Let's use R to calculate H' for the two communities in the example above.

First, let's type in the community data:

```
communityI <- c(10, 1, 1, 1, 1)
communityII <- c(5, 5, 5, 5, 5)
```

Now we need to calculate the relative abundance of each species. This is just the proportion of individuals in the community that belong to each species. So we add up all the individuals using:

```
sum(communityI)
```

And then figure out the proportion that each species is of this total:

```
pI <- communityI/sum(communityI)
```

Check to see what the result looks like by typing pI . Finally, we multiply this by its log, and add up the result to get our H' value:

```
H.I <- -sum(pI*log(pI))
```

Calculate the H' of community II in exactly the same way, replacing I with II.

```
pII <- communityII/sum(communityII)
H.II <- -sum(pII*log(pII))
```

Now we look at the results: $H.I = 0.99$ and $H.II = 1.61$; $H.II$ is almost twice as large as $H.I$!

Re-typing each one of these lines of code for each community gets boring very quickly. This is where the package "vegan" comes in handy. We can automatically calculate the Shannon-Wiener index (and many others) using the function `diversity()`.

First put both of the communities together into a single object using rbind ("row bind"):

```
communities <- rbind(communityI,communityII)
```

Load the vegan package:

```
library(vegan)
```

And calculate the H' values:

```
diversity(communities)
```

Now that we know how to do the calculations, let's look at the bird community data from the coffee systems in Mexico.

Change the working directory to the folder that has the bird community data in it; see the previous handout for suggestions about how to do that. Then read your data in to R, and take a look at the data:

```
birds <- read.csv("Coffee_Bird_List.csv")
head(birds)
```

So now you see that this is a data set with seven communities and a list of birds with abundances for each species in each management type. In order to start manipulating this data, we will have to adjust the structure of the data a little. Specifically, we need to move the first column (the site names) over to the row names, and then delete that column. You can just paste the following line of code in, and check the result:

```
rownames(birds) <- birds[,1]
birds <- birds[,-1]
head(birds)
```

```
> head(birds)
  Common_name Forest_A Forest_B Rustic_A Rustic_B Polyculture_A Polyculture_B Intensive_B
1      Ada         1         0         0         0             0             0             1
2    Albert         0         0         0         0             0             3             0
3     Alice         0         0         0         2             8            10            16
4     Anna        15         4        15         1             4             0             0
5     Annie         0         0         0         0             1             3             1
6   Arthur         0         0         0         0             0             1             0
>
```

One last preparation is to transpose the data, which means flipping the rows and columns:

```
birds <- t(birds)
```

Now we're ready to use the diversity measurement functions. First look at the species richness in each site:

```
bird.species <- specnumber(birds)
bird.species
```

Now we'll use the diversity() function to calculate the Shannon-Wiener H' :

```
bird.shannon <- diversity(birds)
bird.shannon
```

We can make a very basic plot of the results like this:

```
plot(bird.shannon)
```

Or we can make it nicer using this home-made plotting function. It would be easiest to just copy and paste this into R:

```
make.sorted.plot <- function(x){
  ordered <- sort(x, T)
  plot(
    ordered,
    col = terrain.colors(10),
    xaxt = "n", pch = 16, cex = 2,
    ylim = c(min(ordered)*0.5, max(ordered)),
    xlim = c(0, length(x)+1),
    ylab = "Diversity measure", xlab = "Sites",
    main = substitute(x))

  text(ordered,
        names(ordered),
        srt = -75,
        pos = 4)
}
```

We can use this to make plots which are a little easier to interpret:

```
make.sorted.plot(bird.shannon)
make.sorted.plot(bird.species)
```

The sites clearly differ in bird diversity, whether you use species richness or the Shannon index. But what about the actual composition of bird species in these sites? Remember that the identity of species can be very important, not just the number of species.

To look at how close each pair of sites is to one another, we can use a metric like the Jaccard distance. This just summarizes how much two sets of species overlap, and is the number of species shared between two sites divided by the total number of species in both sites. The vegan package has a nice tool for this:

```
birds.jaccard <- vegdist(birds, method = "jaccard")
birds.jaccard
```

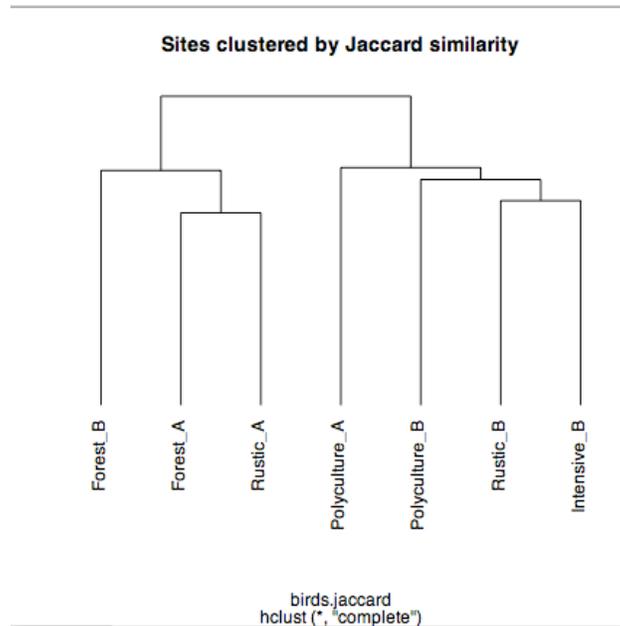
A useful way to examine this result is to build a "tree" from it. Here, a tree (actually a dendrogram) is a way to cluster the sites together based on how "far apart" they are in species composition; sites which have very similar species compositions will group together on the tree:

```
plot(
  hclust(birds.jaccard),
```

```

hang = -1,
main = "Sites clustered by Jaccard similarity",
axes = FALSE, ylab = ""
)

```



We'll use this tree to help us interpret the results of the analysis below.

How does coffee plantation management intensity affect bird diversity? Now that we have our two diversity measures for the sites, and here we can add the last piece of the puzzle, the actual information about how intensively these different sites were managed. Read in the data:

```
management <- read.csv("Coffee_Bird_Management.csv")
```

It will make it easier for us if we "attach" the data set:

```
attach(management)
```

Now we can directly look at the relationship between the Shannon diversity index values and the management intensity. Management intensity is a summary of the management system, where high values mean more agricultural inputs, lower tree cover, and fewer species of trees:

```
plot(bird.shannon ~ Management.index)
```

It looks like there is a downward trend, where more intensive management decreases bird species diversity. Is this relationship important? A linear regression will answer the question for us. Use the `lm()` command to do a regression ("linear model"):

```
regression.result <- lm(bird.shannon ~ Management.index)
summary(regression.result)
```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.05342    0.14660  20.828 4.73e-06 ***
Management.index -0.03209    0.03860  -0.831  0.444

```

The trend, while negative, is not significant ($p = 0.444$). So according to this species diversity metric, more intensive management does not decrease bird diversity.

Finally, add the linear regression to our plot using "a b-line":

```
abline(regression.result, lty = 2)
```

