Assignment: Local (alpha) Diversity

Savannah Bennett; Z620: Quantitative Biodiversity, Indiana University

23 January, 2017

OVERVIEW

In this exercise, we will explore aspects of local or site-specific diversity, also known as alpha (α) diversity. First we will quantify two of the fundamental components of (α) diversity: **richness** and **evenness**. From there, we will then discuss ways to integrate richness and evenness, which will include univariate metrics of diversity along with an investigation of the **species abundance distribution (SAD)**.

Directions:

1. Change "Student Name" on line 3 (above) with your name.
2. Complete as much of the exercise as possible during class; what you do not complete in class will need to be done on your own outside of class.
3. Use the handout as a guide; it contains a more complete description of data sets along with the proper scripting needed to carry out the exercise.
4. Be sure to **answer the questions** in this exercise document; they also correspond to the handout. Space for your answer is provided in this document and indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">".
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. For homework, follow the directions at the bottom of this file.
7. When you are done, **Knit** the text and code into a PDF file.
8. After Knitting, please submit the completed exercise by creating a **pull request** via GitHub. Your pull request should include this file `alpha_assignment.Rmd` and the PDF output of `Knitr` (`alpha_assignment.pdf`).

1) R SETUP

In the R code chunk below, please provide the code to: 1) Clear your R environment, 2) Print your current working directory, 3) Set your working directory to your `/Week2-Alpha` folder, and 4) Load the `vegan` R package (be sure to install if needed).

```
rm(list=ls())
getwd()
```

```
## [1] "C:/Users/Savannah/GitHub/QB2017_Bennett/Week2-Alpha"
```

```
setwd("/Users/Savannah/GitHub/QB2017_Bennett/Week2-Alpha")
require("vegan")
```

```
## Loading required package: vegan
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.4-2
```

2) LOADING DATA

In the R code chunk below, do the following: 1) Load your dataset, and 2) Display the structure of the dataset (if the structure is long, use `max.level=0` to show just basic information).

```
data(BCI)
str(BCI, max.level=0)
```

```
## 'data.frame':    50 obs. of  225 variables:
##   [list output truncated]
##  - attr(*, "original.names")= chr  "Abarema.macradenium" "Acacia.melanoceras" "Acalypha.diversifolia
```

3) SPECIES RICHNESS

Species richness (S) is simply the number of species in a system or the number of species observed in a sample.

Observed Richness

In the R code chunk below, do the following:

1. Write a function called `S.obs` to calculate observed richness
2. Use your function to determine the number of species in `site1`, and
3. Compare the output of your function to the output of the `specnumber()` function in vegan.

```
# S.obs <- function(   ){
#   rowSums(     ) *
#   }

S.obs <- function(x=""){rowSums(x>0*1)}
S.obs(BCI)
```

```
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##  93  84  90  94 101  85  82  88  90  94  87  84  93  98  93  93  93  89
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
## 109 100  99  91  99  95 105  91  99  85  86  97  77  88  86  92  83  92
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50
##  88  82  84  80 102  87  86  81  81  86 102  91  91  93
```

```
specnumber(BCI)
```

```
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##  93  84  90  94 101  85  82  88  90  94  87  84  93  98  93  93  93  89
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
## 109 100  99  91  99  95 105  91  99  85  86  97  77  88  86  92  83  92
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50
##  88  82  84  80 102  87  86  81  81  86 102  91  91  93
```

Question 1: Does `specnumber()` from `vegan` return the same value for observed richness in `site1` as our function `S.obs`? What is the species richness of the first 4 sites (i.e., rows) of the BCI matrix?

> ***Answer 1***:Yes, 'specnumber()' from 'vegan' and 'S.obs' both provide richness values of 93 species for site 1. The richness of site 2 is 84, the richness of site 3 is 90, and the richness of site 4 is 94 species.

Coverage. How Well Did You Sample Your Site?

In the R code chunk below, do the following:

1. Write a function to calculate Good's Coverage, and
2. Use that function to calculate coverage for all sites in the BCI matrix.

```
C <- function(x= ""){1-(sum(x==1)/rowSums(x))}
C(BCI[1,])
```

```
##         1
## 0.9308036
```

```
D <- function(x= ""){rowSums(x==1)/length(BCI[1,])}
D(BCI[1,])
```

```
##         1
## 0.1377778
```

```
E <- function(x= ""){rowSums(x==1)/length(BCI)}
E(BCI)
```

```
##         1         2         3         4         5         6         7
## 0.1377778 0.1377778 0.1644444 0.1200000 0.1600000 0.1511111 0.1244444
##         8         9        10        11        12        13        14
## 0.1066667 0.1644444 0.1555556 0.1511111 0.1511111 0.1377778 0.1688889
##        15        16        17        18        19        20        21
## 0.1333333 0.1422222 0.1777778 0.1244444 0.2133333 0.1688889 0.1911111
##        22        23        24        25        26        27        28
## 0.1733333 0.1955556 0.1688889 0.1777778 0.1600000 0.1688889 0.1377778
##        29        30        31        32        33        34        35
## 0.1644444 0.1644444 0.1155556 0.1200000 0.1511111 0.1511111 0.1955556
##        36        37        38        39        40        41        42
## 0.1555556 0.1200000 0.1377778 0.1377778 0.1333333 0.2000000 0.1288889
##        43        44        45        46        47        48        49
## 0.1555556 0.1511111 0.1511111 0.1244444 0.2177778 0.1422222 0.1733333
##        50
## 0.1644444
```

Question 2: Answer the following questions about coverage:

a. What is the range of values that can be generated by Good's Coverage?
b. What would we conclude from Good's Coverage if n_i equaled N?
c. What portion of taxa in `site1` were represented by singletons?
d. Make some observations about coverage at the BCI plots.

Answer 2a:Values generated by Good's Coverage range from 0-1.

Answer 2b:This would indicate that all of the taxa are singletons.

Answer 2c:13.77% of taxa from site 1 are represented by singletons.

Answer 2d:The Good's Coverage values indicate that coverage is relatively high across most of the sampling sites. Around 12-21% of the taxa at these sites are represented by singletons. Therefore, the majority of taxa at these sites were not singletons.

Estimated Richness

In the R code chunk below, do the following:

1. Load the microbial dataset (located in the `/Week2-Alpha/data` folder),
2. Transform and transpose the data as needed (see handout),
3. Create a vector (`soilbac1`) with the bacterial OTU abundances at any site in the dataset,
4. Calculate the observed richness at that particular site, and
5. Calculate the coverage at that particular site

```
soilbac <- read.table("data/soilbac.txt", sep = "\t", header = TRUE, row.names = 1)
soilbac.t <- as.data.frame(t(soilbac))
soilbac1 <- soilbac.t[1,]
R <- function(x=""){rowSums(x>0*1)}
R(soilbac1[1,])
```

```
## T1_1
## 1074
```

```
E <- function(x= ""){1-(sum(x==1)/rowSums(x))}
E(soilbac1[1,])
```

```
##      T1_1
## 0.6479471
```

```
#Question 3a
rowSums(soilbac1[1,])
```

```
## T1_1
## 2119
```

```
#Question 3c
H <- function(x= ""){1-(sum(x==1)/rowSums(x))}
H(BCI[1,])
```

```
##         1
## 0.9308036
```

Question 3: Answer the following questions about the soil bacterial dataset.

a. How many sequences did we recover from the sample `soilbac1`, i.e. N?
b. What is the observed richness of `soilbac1`?
c. How does coverage compare between the BCI sample (`site1`) and the KBS sample (`soilbac1`)?

Answer 3a:We recovered a total of 2119 sequences.

Answer 3b:The observed richness of soilbac1 is 1,074.

Answer 3c:The coverage is higher in the BCI sample (0.93) compared to the KBS sample (0.64).This indicates that there are more singletons in the KBS sample than the BCI sample.

Richness Estimators

In the R code chunk below, do the following:

1. Write a function to calculate **Chao1**,
2. Write a function to calculate **Chao2**,
3. Write a function to calculate **ACE**, and
4. Use these functions to estimate richness at both `site1` and `soilbac1`.

```
S.chao1 <- function(x = ""){ S.obs(x) + (sum(x == 1)^2) / (2 * sum(x == 2)) }

S.chao2 <-function(site ="",SbyS =""){
SbyS = as.data.frame(SbyS)
x =SbyS[site, ]
SbyS.pa <-(SbyS >0) *1
Q1 = sum(colSums(SbyS.pa) ==1)
Q2 = sum(colSums(SbyS.pa) ==2)
S.chao2 = S.obs(x) +(Q1^2)/(2*Q2)
return(S.chao2)
}

S.ace <-function(x ="",thresh =""){
x <-x[x>0]
S.abund <- length(which(x >thresh))
S.rare <- length(which(x <=thresh))
singlt <- length(which(x ==1))
N.rare <- sum(x[which(x <=thresh)])
C.ace <-1-(singlt /N.rare)
i <- c(1:thresh)
count <-function(i, y){
length(y[y ==i])
}
a.1<- sapply(i, count, x)
f.1<-(i *(i -1)) *a.1
G.ace <-(S.rare/C.ace)*(sum(f.1)/(N.rare*(N.rare-1)))
S.ace <-S.abund +(S.rare/C.ace) +(singlt/C.ace) * max(G.ace,0)
return(S.ace)
}

#Chao1 Site 1
S.chao1(BCI[1,])
```

```
##        1
## 119.6944
```

```
#Chao Soilbac1
S.chao1(soilbac1[1,])
```

```
##     T1_1
## 2628.514
```

```
#Chao2 Site 1
S.chao2(1,BCI)
```

```
##        1
## 104.6053
```

```
#Chao2 Soilbac1
S.chao2("T1_1",soilbac.t)
```

```
##     T1_1
## 21055.39
```

```
#ACE Site 1
S.aceBCI <- S.ace(BCI[1,],10)
S.aceBCI
```

```
## [1] 159.3404
```

```
#ACE Soilbac1
S.acesoilbac1 <- S.ace(soilbac1,10)
S.acesoilbac1
```

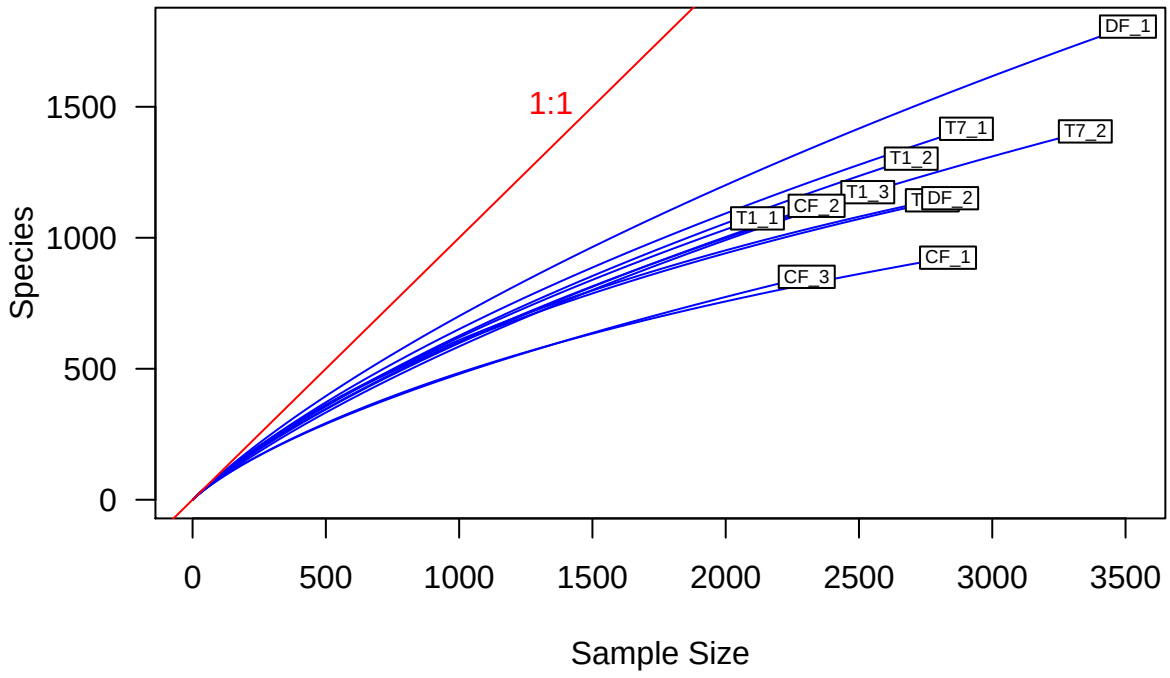
```
## [1] 4465.983
```

Rarefaction

In the R code chunk below, please do the following:

1. Calculate observed richness for all samples in `soilbac`,
2. Determine the size of the smallest sample,
3. Use the `rarefy()` function to rarefy each sample to this level,
4. Plot the rarefaction results, and
5. Add the 1:1 line and label.

```
soilbac.S <- S.obs(soilbac.t)
min.N <- min(rowSums(soilbac.t))
S.rarefy <- rarefy(x =soilbac.t,sample =min.N,se =TRUE)
rarecurve(x =soilbac.t,step =20,col ="blue",cex =0.6,las=1)
abline(0,1,col = 'red')
text(1500,1500,"1:1",pos =2,col = 'red')
```



Question 4: What is the difference between ACE and the Chao estimators?

Answer 4:ACE utilizes a threshold to examine the abundance of rare species, or species with fewer than ten individuals. Chao1 estimates coverage based on singletons and doubletons. In other words, Chao1 considers singletons and doubletons rare, while ACE considers rare as ten or fewer individuals.

4) SPECIES EVENNESS

Here, we consider how abundance varies among species, that is, **species evenness**.

Visualizing Evenness: The Rank Abundance Curve (RAC)

One of the most common ways to visualize evenness is in a **rank-abundance curve** (sometime referred to as a rank-abundance distribution or Whittaker plot). An RAC can be constructed by ranking species from the most abundant to the least abundant without respect to species labels (and hence no worries about 'ties' in abundance).

In the R code chunk below, do the following:

1. Write a function to construct a RAC,
2. Be sure your function removes species that have zero abundances,
3. Order the vector (RAC) from greatest (most abundant) to least (least abundant), and
4. Return the ranked vector

```
RAC <-function(x =""){
x = as.vector(x)
x.ab =x[x >0]
x.ab.ranked =x.ab[order(x.ab,decreasing =TRUE)]
return(x.ab.ranked)
}
```

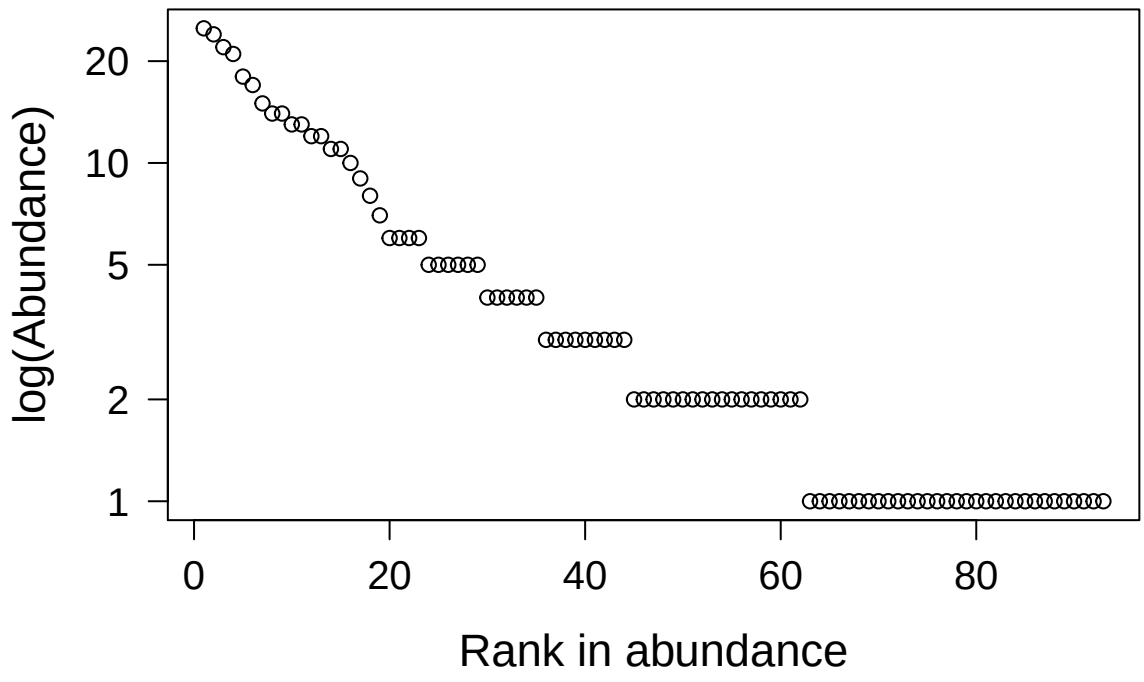
Now, let's examine the RAC for `site1` of the BCI data set.

In the R code chunk below, do the following:

1. Create a sequence of ranks and plot the RAC with natural-log-transformed abundances,
2. Label the x-axis "Rank in abundance" and the y-axis "log(abundance)"

```
plot.new()
site1 <-BCI[1, ]

rac <- RAC(x =site1)
ranks <- as.vector(seq(1, length(rac)))
opar <- par(no.readonly =TRUE)
par(mar = c(5.1,5.1,4.1,2.1))
plot(ranks, log(rac),type = 'p',axes =F,
xlab ="Rank in abundance",ylab ="log(Abundance)",
las =1,cex.lab =1.4,cex.axis =1.25)
box()
axis(side =1,labels =T,cex.axis =1.25)
axis(side =2,las =1,cex.axis =1.25,
labels = c(1,2,5,10,20),at = log(c(1,2,5,10,20)))
```



Question 5: What effect does visualizing species abundance data on a log-scaled axis have on how we interpret evenness in the RAC?

> ***Answer 5***:The log-scale axis makes it easier to compare species abundances when there is a large range in abundance values across species. It essentially scales and normalizes the data such that abundances can be easily compared.

Now that we have visualized unevennes, it is time to quantify it using Simpson's evenness ($E_{1/D}$) and Smith and Wilson's evenness index (E_{var}).

Simpson's evenness ($E_{1/D}$)

In the R code chunk below, do the following:

1. Write the function to calculate $E_{1/D}$, and
2. Calculate $E_{1/D}$ for `site1`.

```
SimpE <-function(x =""){
S <- S.obs(x)
x = as.data.frame(x)
D <- diversity(x,"inv")
E <-(D)/S
return(E)
}

site1 <-BCI[1, ]
SimpE(site1)
```

```
##         1
## 0.4238232
```

Smith and Wilson's evenness index (E_{var})

In the R code chunk below, please do the following:

1. Write the function to calculate E_{var},
2. Calculate E_{var} for `site1`, and
3. Compare $E_{1/D}$ and E_{var}.

```
Evar <-function(x){
x <- as.vector(x[x >0])
1-(2/pi)*atan(var(log(x)))
}

Evar(site1)
```

```
## [1] 0.5067211
```

Question 6: Compare estimates of evenness for `site1` of BCI using $E_{1/D}$ and E_{var}. Do they agree? If so, why? If not, why? What can you infer from the results.

> ***Answer 6***:Simpson's index yielded an evenness value of 0.4238, while Smith and Wilson's evenness index is 0.5067. Simpson's index is more sensitive to differences in the most abundant species, which could explain the discrepancy in these values. The Smith and Wilson's index is log transformed, which reduces this bias. One can infer that this data set has at least one species with relatively high abundance values based on the fact that the index values differed. Therefore, the Smith and Wilson's Index may be a better representation of species evenness.

5) INTEGRATING RICHNESS AND EVENNESS: DIVERSITY METRICS

So far, we have introduced two primary aspects of diversity, i.e., richness and evenness. Here, we will use popular indices to estimate diversity, which explicitly incorporate richness and evenness We will write our own diversity functions and compare them against the functions in `vegan`.

Shannon's diversity (a.k.a., Shannon's entropy)

In the R code chunk below, please do the following:

1. Provide the code for calculating H' (Shannon's diversity),
2. Compare this estimate with the output of `vegan`'s diversity function using method = "shannon".

```
ShanH <-function(x =""){
H =0
for (n_i in x){
if(n_i >0) {
p =n_i / sum(x)
H =H -p*log(p)
}
}
return(H)
}
ShanH(site1)
```

```
## [1] 4.018412
```

```
diversity(site1, index="shannon")
```

```
## [1] 4.018412
```

Simpson's diversity (or dominance)

In the R code chunk below, please do the following:

1. Provide the code for calculating D (Simpson's diversity),
2. Calculate both the inverse (1/D) and 1 - D,
3. Compare this estimate with the output of `vegan's` diversity function using method = "simp".

```
SimpD <-function(x =""){
D =0
N = sum(x)
for (n_i in x){
D =D +(n_i^2)/(N^2)
}
return(D)
}
SimpD(site1)
```

```
## [1] 0.0253707
```

```
D.inv <-1/SimpD(site1)
D.inv
```

```
## [1] 39.41555
```

```
D.sub <-1-SimpD(site1)
D.sub
```

```
## [1] 0.9746293
```

```
diversity(site1,"inv")
```

```
## [1] 39.41555
```

```
diversity(site1,"simp")
```

```
## [1] 0.9746293
```

Question 7: Compare estimates of evenness for `site1` of BCI using $E_{H'}$ and E_{var}. Do they agree? If so, why? If not, why? What can you infer from the results.

> ***Answer 7***:These results do not agree because the Shannon index value was 4.01 while the Smith and Wilson's index value was 0.506. The Shannon index incorporates both richness and evenness. Smith and Wilson's index only measures evenness. Thus, the Shannonand Smith and Wilson's indices are different tests; they are not taking the same measurements.

Fisher's α

In the R code chunk below, please do the following:

1. Provide the code for calculating Fisher's α,

2. Calculate Fisher's $\boldsymbol{\alpha}$ for `site1` of BCI.

```
rac <- as.vector(site1[site1>0])
invD <- diversity(rac, "inv")
invD
```

```
## [1] 39.41555
```

```
Fisher <- fisher.alpha(rac)
Fisher
```

```
## [1] 35.67297
```

Question 8: How is Fisher's $\boldsymbol{\alpha}$ different from $E_{H'}$ and E_{var}? What does Fisher's $\boldsymbol{\alpha}$ take into account that $E_{H'}$ and E_{var} do not?

> ***Answer 8***:Fisher's $\boldsymbol{\alpha}$ estimates diversity as opposed to calculating a diversity metric like the Shannon index. Like the Shannon index, it differs from the Smith and Wilson's index in that it incorporates richness and evenness, while the Smith and Wilson's index measures evenness. In this way, Fisher's $\boldsymbol{\alpha}$ accounts for sampling error.

6) MOVING BEYOND UNIVARIATE METRICS OF α DIVERSITY

The diversity metrics that we just learned about attempt to integrate richness and evenness into a single, univariate metric. Although useful, information is invariably lost in this process. If we go back to the rank-abundance curve, we can retrieve additional information – and in some cases – make inferences about the processes influencing the structure of an ecological system.

Species abundance models

The RAC is a simple data structure that is both a vector of abundances. It is also a row in the site-by-species matrix (minus the zeros, i.e., absences).

Predicting the form of the RAC is the first test that any biodiversity theory must pass and there are no less than 20 models that have attempted to explain the uneven form of the RAC across ecological systems.

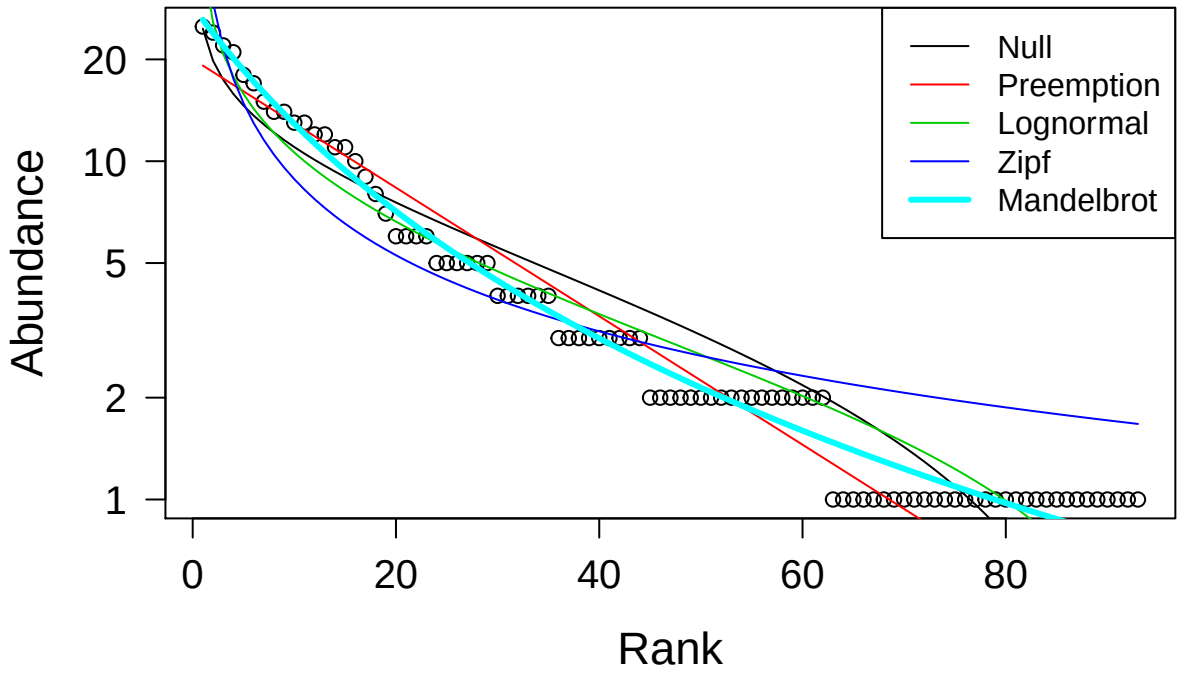
In the R code chunk below, please do the following:

1. Use the `radfit()` function in the `vegan` package to fit the predictions of various species abundance models to the RAC of `site1` in BCI,
2. Display the results of the `radfit()` function, and
3. Plot the results of the `radfit()` function using the code provided in the handout.

```
RACresults <- radfit(site1)
RACresults
```

```
##
## RAD models, family poisson
## No. of species 93, total abundance 448
##
##            par1      par2     par3    Deviance AIC      BIC
## Null                                   39.5261 315.4362 315.4362
## Preemption  0.042797                    21.8939 299.8041 302.3367
## Lognormal   1.0687    1.0186           25.1528 305.0629 310.1281
## Zipf        0.11033  -0.74705          61.0465 340.9567 346.0219
## Mandelbrot  100.52   -2.312    24.084   4.2271 286.1372 293.7350
```

```
plot.new()
plot(RACresults,las =1,cex.lab =1.4,cex.axis =1.25)
```



Question 9: Answer the following questions about the rank abundance curves: a) Based on the output of `radfit()` and plotting above, discuss which model best fits our rank-abundance curve for `site1`? b) Can we make any inferences about the forces, processes, and/or mechanisms influencing the structure of our system, e.g., an ecological community?

> ***Answer 9a***:The Zipf-Mandelbrot model best fits our rank-abundance curve for site1 because it has the lowest AIC value (286.13) compared to those of the other models.
> ***Answer 9b***:From this model, we cannot make any accurate inferences about the factors shaping community structure in this system. According to this model, the abundance of a species is inversely proportional to its rank in abundance, but we cannot make assumptions about what specifically is structuring this community.

Question 10: Answer the following questions about the preemption model: a. What does the preemption model assume about the relationship between total abundance (N) and total resources that can be preempted? b. Why does the niche preemption model look like a straight line in the RAD plot?

> ***Answer 10a***:This model assumes that the total abundance, N, is a function of the total resources that can be used. In other words, the total resources available dictates the number of abundance of organisms that can exist. It also assumes that each species removes a constant proportion of resources from the total avialable resources, and all resources will be used.
>
> ***Answer 10b***:The niche preemption model is a straight line in a RAD plot, while the other models yield curved lines. This is probably because the line for the niche preemption model is on a log scale.

Question 11: Why is it important to account for the number of parameters a model uses when judging how well it explains a given set of data?

> ***Answer 11***:The number of parameters can help determine whether a model can be used to explain a particular data set. More specifically, a greater number of parameters can provide more assurance that a model fits/does not fit a data set.

SYNTHESIS

1. As stated by Magurran (2004) the $D = \sum p_i^2$ derivation of Simpson's Diversity only applies to communities of infinite size. For anything but an infinitely large community, Simpson's Diversity index is calculated as $D = \sum \frac{n_i(n_i-1)}{N(N-1)}$. Assuming a finite community, calculate Simpson's D, 1 - D, and Simpson's inverse (i.e. 1/D) for `site 1` of the BCI site-by-species matrix.

```
SimpD <-function(x =""){
D =0
N = sum(x)
for (n_i in x){
D =D +(n_i^2)/(N^2)
}
return(D)
}
SimpD(site1)
```

```
## [1] 0.0253707
```

```
D.inv <-1/SimpD(site1)
D.inv
```

```
## [1] 39.41555
```

```
D.sub <-1-SimpD(site1)
D.sub
```

```
## [1] 0.9746293
```

```
diversity(site1,"inv")
```

```
## [1] 39.41555
```

```
diversity(site1,"simp")
```

```
## [1] 0.9746293
```

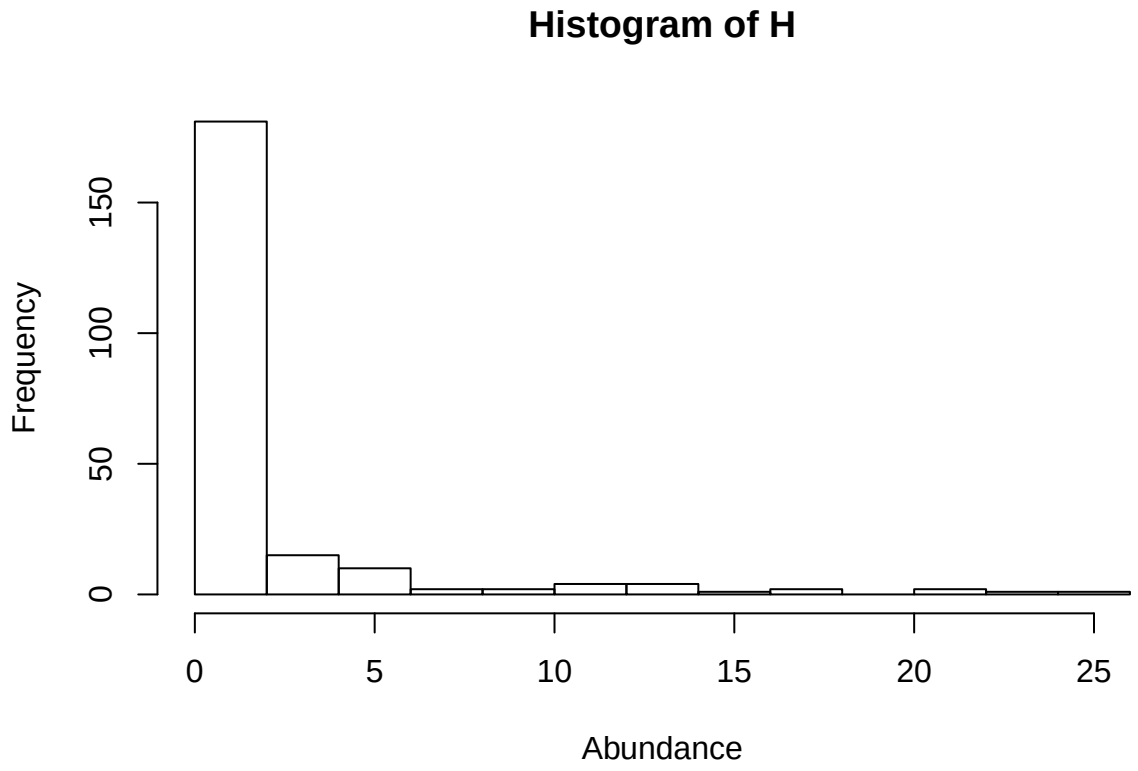
```
D.sub
```

```
## [1] 0.9746293
```

Synthesis Question 1:

2. Along with the rank-abundance curve (RAC), another way to visualize the distribution of abundance among species is with a histogram (a.k.a., frequency distribution) that shows the frequency of different abundance classes. For example, in a given sample, there may be 10 species represented by a single individual, 8 species with two individuals, 4 species with three individuals, and so on. In fact, the rank-abundance curve and the frequency distribution are the two most common ways to visualize the species-abundance distribution (SAD) and to test species abundance models and biodiversity theories. To address this homework question, use the R function **hist()** to plot the frequency distribution for `site 1` of the BCI site-by-species matrix, and describe the general pattern you see.

```
H <- t(site1)
I <- hist(H,xlab= "Abundance")
```



>Synthesis Question 2:According to the histogram, most species have lower abundance values. Very few species have abundance values higher than twenty. The majority of species have abundances lower than five.

3. We asked you to find a biodiversity dataset with your partner. This data could be one of your own or it could be something that you obtained from the literature. Load that dataset. How many sites are there? How many species are there in the entire site-by-species matrix? Any other interesting observations based on what you learned this week?

```
projectdata <- read.table("finalprojectdata1text.txt", sep = "\t", header = TRUE)
str(projectdata)
```

```
## 'data.frame':    97 obs. of  847 variables:
##  $ species                                : Factor w/ 7 levels "Canada P. Contorta",..: 7 7 7 7 7 7 7
##  $ Piloderma_sphaerosporum                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Phialocephala_sp_KO_2013               : num  0.182 0.012 0.112 0.094 0.08 0.04 0.03 0.031 0.052 0.
##  $ Atheliales_sp__SH212379.06FU           : num  0.006 0 0 0.008 0 0.002 0 0 0 0 ...
##  $ Rhizoscyphus_ericae                    : num  0.001 0 0 0 0.002 0 0 0.005 0 0 ...
##  $ Sistotrema_sp__SH211497.06FU           : num  0 0 0 0 0 0 0 0 0.063 0 ...
##  $ Helotiales_sp_SC3_1                    : num  0.006 0.011 0 0.028 0.088 0.001 0 0.023 0.052 0 ...
##  $ Atheliales_sp__SH220593.06FU           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Oidiodendron_sp__SH217744.06FU         : num  0 0 0 0.001 0.003 0 0 0 0 0 ...
##  $ Lachnum_virgineum                      : num  0.086 0.374 0.003 0.009 0.245 0 0.048 0.037 0.039 0
##  $ Cenococcum_sp__SH196545.06FU           : num  0 0 0 0.011 0 0.001 0.033 0.028 0.012 0.038 ...
##  $ Atheliales_sp__SH229868.06FU           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Cadophora_finlandica                   : num  0.006 0.007 0.023 0.071 0.029 0.048 0.014 0.022 0.0
```

```
##  $ Piloderma_sp_9_RT_2012                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Wilcoxina_rehmii                      : num  0.005 0.024 0.125 0.003 0.011 0.512 0.16 0.139 0.02
##  $ Tylospora_sp__SH213920.06FU           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Meliniomyces_bicolor                  : num  0.001 0.013 0.015 0.014 0.002 0 0.001 0.003 0.023 0
##  $ Mycenaceae_sp__SH193138.06FU          : num  0.235 0.001 0 0.014 0.002 0 0 0 0 0 ...
##  $ Pseudeurotium_sp__SH209207.06FU       : num  0 0 0 0.001 0 0 0 0 0 0.001 ...
##  $ Suillus_luteus                        : num  0 0 0 0 0 0 0 0.022 0.143 0.01 ...
##  $ Leptodontidium_orchidicola            : num  0 0.005 0 0 0 0.012 0 0 0 0 ...
##  $ Piloderma_sp_12_RT_2012               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Oidiodendron_sp__SH217771.06FU        : num  0.002 0 0.001 0.001 0.002 0.003 0.012 0.005 0.004 0
##  $ Xylaria_sp_UFMGCB_3872                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mycenaceae_sp__SH193137.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mycena_epipterygia                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Cryptosporiopsis_sp__SH209253.06FU    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH209196.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Atheliales_sp__SH214554.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Atheliales_sp__SH229870.06FU          : num  0 0 0 0 0.028 0 0 0 0 0 ...
##  $ Acephala_sp__SH213468.06FU            : num  0 0 0.078 0.076 0.001 0.008 0.007 0.012 0.026 0.005
##  $ Atheliaceae_sp__SH238707.06FU         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Russula_emetica                       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Tylopilus_felleus                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Phialocephala_sp__SH213468.06FU       : num  0.004 0.001 0.033 0.054 0.024 0.023 0.016 0.034 0.0
##  $ Penicillium_spinulosum                : num  0.002 0 0 0.001 0 0 0 0 0.001 0 ...
##  $ Basidiomycota_sp_JH_84                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mortierellales_sp__SH218044.06FU      : num  0 0.001 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH215664.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Herpotrichiellaceae_sp_RB_2011        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH228612.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Phialocephala_fortinii                : num  0.046 0.003 0.003 0.007 0.02 0.017 0.01 0.026 0 0.0
##  $ Oidiodendron_sp__SH217804.06FU        : num  0 0 0 0 0.003 0 0 0 0 0 ...
##  $ Russula_decolorans                    : num  0 0.005 0 0 0 0 0 0 0 0 ...
##  $ Mycena_sp_1_KO_2013                   : num  0 0.011 0 0 0 0 0 0.003 0.006 0.011 ...
##  $ Hyaloscyphaceae_sp__SH189785.06FU     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Podospora_sp_LH107                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mortierella_alpina                    : num  0 0 0 0 0 0 0 0.002 0 0 0 ...
##  $ Mycena_aetites                        : num  0.001 0.092 0.133 0.006 0.001 0 0 0.033 0.002 0 ...
##  $ Cryptosporiopsis_sp_1_KO_2013         : num  0.022 0.002 0 0 0 0 0 0 0 0 ...
##  $ Hygrophorus_piceae                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Wilcoxina_mikolae                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mycena_sp__SH194135.06FU              : num  0 0 0 0.002 0.021 0 0.152 0.01 0 0.028 ...
##  $ Helotiaceae_sp_II_GK_2010             : num  0.005 0 0 0.003 0.002 0 0 0 0 0 ...
##  $ Helotiales_sp__SH228608.06FU          : num  0 0 0 0 0.001 0 0 0 0 0 ...
##  $ Ilyonectria_pseudodestructans         : num  0 0.002 0 0 0 0.001 0 0 0 0 ...
##  $ Arachnopeziza_sp__SH230020.06FU       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Atheliaceae_sp__SH222529.06FU         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH240584.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Rickenella_sp__SH201022.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Oidiodendron_sp__SH217751.06FU        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Chalara_sp__SH209190.06FU             : num  0.01 0.001 0.053 0 0 0 0.004 0.012 0.026 0.005 ...
##  $ Oidiodendron_sp_PDKA6                 : num  0.001 0 0 0 0.001 0 0 0.004 0 0 ...
##  $ Amanita_muscaria                      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Rhizoscyphus_sp__SH207220.06FU        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Piloderma_olivaceum                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Oidiodendron_sp__SH217754.06FU        : num  0.007 0 0 0.001 0 0 0 0 0 0 ...
```

```
##  $ Mycena_albidolilacea                  : num  0.083 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH215668.06FU          : num  0.021 0 0 0 0 0 0.01 0.001 0 0.177 ...
##  $ Suillus_tomentosus                    : num  0 0 0 0.04 0.054 0 0 0 0.073 0 ...
##  $ Hebeloma_mesophaeum                   : num  0 0 0.001 0 0 0 0 0.001 0.003 0 ...
##  $ Lecanoromycetes_sp__SH197703.06FU     : num  0.015 0 0 0 0 0 0 0 0 0 ...
##  $ Piloderma_sp_A18                      : num  0 0 0 0.019 0 0 0.001 0 0 0 ...
##  $ Helotiales_sp_28_MV_2011              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Venturia_sp__SH207408.06FU            : num  0.006 0.006 0 0 0 0 0.018 0.009 0.012 0.004 ...
##  $ Cryptococcus_podzolicus               : num  0.013 0.001 0 0 0 0 0 0 0 0 ...
##  $ Thelephoraceae_sp__SH195956.06FU      : num  0 0 0 0.028 0.049 0 0 0 0 0 ...
##  $ Helotiales_sp__SH228628.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mycena_sp__SH208812.06FU              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Dothideomycetes_sp__SH229757.06FU     : num  0.017 0.002 0.006 0.006 0 0.002 0.032 0.004 0.02 0.0
##  $ Ascomycota_sp_6_RB_2011               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Rhizoscyphus_sp__SH228608.06FU        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH240585.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Cryptococcus_terricola                : num  0.004 0 0 0 0 0 0 0 0 0 ...
##  $ Auriculariales_sp__SH219570.06FU      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH196873.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Hydnotrya_cubispora                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Atheliaceae_sp__SH214560.06FU         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Piloderma_sp__SH238374.06FU           : num  0 0 0.039 0.007 0.011 0 0.003 0.022 0.001 0 ...
##  $ Lachnum_sp__SH189775.06FU             : num  0.002 0.08 0.004 0.004 0.007 0 0.001 0.003 0.002 0.0
##  $ Pezizomycetes_sp__SH234386.06FU       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH228638.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Glarea_sp__SH203829.06FU              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Trichocladium_opacum                  : num  0.002 0 0 0 0 0.003 0 0 0 0 ...
##  $ Pseudotomentella_atrofusca            : num  0.015 0 0 0 0 0 0 0 0 0 ...
##  $ Helotiales_sp__SH209213.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Microbotryomycetes_sp__SH212308.06FU  : num  0 0 0 0 0 0.003 0 0 0 0 ...
##  $ Atheliales_sp__SH212380.06FU          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Trechisporales_sp__SH216269.06FU      : num  0 0 0.019 0.004 0.043 0 0 0 0 0 ...
##   [list output truncated]
```

```
#Richness Values at North American Sites
rich <- function(x=""){rowSums(x>0*1)}
rich(projectdata[1:18, 2:847])
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18
## 63 55 52 66 59 44 59 56 56 46 49 53 71 53 55 68 57 65
```

```
#Richness Values at South American Sites
rich(projectdata[19:28, 2:847])
```

```
## 19 20 21 22 23 24 25 26 27 28
## 40 38 46 61 45 38 48 54 49 32
```

```
#Richness Values at Canadian Sites
rich(projectdata[29:45, 2:847])
```

```
## 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
## 57 51 62 67 53 61 47 60 63 43 50 54 55 30 49 52 39
```

```
#Richness Values of South American Sites where Contorta is Invasive
rich(projectdata[88:97, 2:847])
```

```
## 88 89 90 91 92 93 94 95 96 97
```

```
## 34 37 47 47 22 48 55 52 49 48
```

```
#Shannon Diversity For American Tree
diversity(projectdata[1, 2:847], index="shannon")
```

```
## [1] 2.888292
```

```
#Shannon Diversity for Southern Hemisphere Tree (non-invasive)
diversity(projectdata[19, 2:847], index="shannon")
```

```
## [1] 2.780041
```

```
#Shannon Diversity for Southern Hemisphere Tree (invasive)
diversity(projectdata[88, 2:847], index="shannon")
```

```
## [1] 2.925967
```

```
#Shannon Diversity for Canadian Tree
diversity(projectdata[29, 2:847], index="shannon")
```

```
## [1] 3.14873
```

```
#Shannon Diversity for European Tree
diversity(projectdata[46, 2:847], index="shannon")
```

```
## [1] 2.906819
```

```
#Simpson's Index for American Tree
diversity(projectdata[1, 2:847], index="simp")
```

```
## [1] 0.889704
```

```
#Simpson's Index for Southern Hemisphere Tree (non-invasive)
diversity(projectdata[19, 2:847], index="simp")
```

```
## [1] 0.9070952
```

```
#Simpson's Index for Southern Hemisphere Tree (invasive)
diversity(projectdata[88, 2:847], index="simp")
```

```
## [1] 0.9304301
```

```
#Simpson's Index for Canadian Tree
diversity(projectdata[29, 2:847], index="simp")
```

```
## [1] 0.935944
```

```
#Simpson's Index for European Tree
diversity(projectdata[46, 2:847], index="simp")
```

```
## [1] 0.88476
```

```
#Fisher's Alpha for American Tree
amtree <- projectdata[1, 2:847]
rac1 <- as.vector(amtree[amtree > 0])
invD1 <- diversity(rac1, "inv")
invD1
```

```
## [1] 9.066512
```

```
#Fisher's Alpha for Southern Hemisphere Tree (non-invasive)
shtree <- projectdata[19, 2:847]
rac2 <- as.vector(shtree[shtree > 0])
```

```
invD2 <- diversity(rac2, "inv")
invD2
```

```
## [1] 10.7637
```

```
#Fisher's Alpha for Southern Hemisphere Tree (invasive)
shtree2 <- projectdata[88, 2:847]
rac3 <- as.vector(shtree2[shtree2 > 0])
invD3 <- diversity(rac3, "inv")
invD3
```

```
## [1] 14.37402
```

```
#Fisher's Alpha for Canadian Tree
cantree <- projectdata[29, 2:847]
rac4 <- as.vector(cantree[cantree > 0])
invD4 <- diversity(rac4, "inv")
invD4
```

```
## [1] 15.61134
```

```
#Fisher's Alpha for European Tree
eurotree <- projectdata[46, 2:847]
rac5 <- as.vector(eurotree[eurotree > 0])
invD5 <- diversity(rac5, "inv")
invD5
```

```
## [1] 8.677543
```

Answer to Synthesis Question 3: There are 7 sites in my dataset, and a total of 847 species were observed. The data set my partner and I selected has fungal OTUs associated with Pinus Contorta trees at different geographic locations. The data set includes data from multiple trees within each geographic location. Total richness at the North American sites ranged from 53-71 OTUs, and richness values at the Canadian sites ranged from 30-61.Richness at the South American sites ranged from 32-61. Similarly, richness at south American sites where Contorta is invasive ranged from 34-61. Further analyses would be necessary to determine if richness was significantly different among these sites, but based on these data, it appears that in South American sites, total richness (of fungi associated with Contorta trees) does not greatly differ between sites where this tree is invasive and sites where this tree is not invasive. Further analyses of mean richness per geographic location would be needed to determine whether richness differed significantly by geographic location. I then selected one tree from each geographic location for further preliminary analyses. However, to fully explore the data and generate more reliable results, data from all trees at each location should be analyzed.I first determined the Shannon index values (for one tree from each geographic location). These values ranged from 2.7-3.1. Therefore, based on this preliminary analysis, it does not appear that diversity values differ by geographic location. The Canadian tree exhibited a slightly higher Shannon index value (3.14) than the other sites. I also determined Simpson's index values, which ranged from 0.88-0.93. Like the Shannon index values, the Simpson's index values do not appear to differ by geographic location. If possible for this type of diversity measure, tests to determine whether these values significantly differ could be performed. The Simpson's index values were relatively low, which might suggest that fungal communities associated with these trees are not dominated by a small number of species. The Canadian tree had a marginally higher Simpson's index value value than the other sites, which might suggest that it is dominated by a few taxa to a greater extent than trees from the other geographic locations. Finally, I determined the Fisher's alpha values for one tree from each tree. The Fisher's alpha values ranged from 9-15, where the Canadian tree exhibited the highest value, and the American tree exhibited the lowest value. Fisher's alpha values could provide an additional diversity measure to support/refute differences in diversity seen among the sites in the

other diversity measures. In sum, there appear to be only slight differences in fungal diversity an evenness among the different geographic locations. However, these metrics should be determined for all trees within each geographic location for a more reliable analysis.

SUBMITTING YOUR ASSIGNMENT

Use Knitr to create a PDF of your completed alpha_assignment.Rmd document, push it to GitHub, and create a pull request. Please make sure your updated repo include both the HTML and RMarkdown files.

Unless otherwise noted, this assignment is due on **Wednesday, January 25th, 2015 at 12:00 PM (noon)**.