# 5. Worksheet: Alpha Diversity

Erica Nadolski; Z620: Quantitative Biodiversity, Indiana University

24 January, 2023

## OVERVIEW

In this exercise, we will explore aspects of local or site-specific diversity, also known as alpha ($\alpha$) diversity. First we will quantify two of the fundamental components of ($\alpha$) 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. In the Markdown version of this document in your cloned repo, change "Student Name" on line 3 (above) to your name.
2. Complete as much of the worksheet as possible during 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. Answer questions in the worksheet. 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 ">". You should notice that the answer is highlighted in green by RStudio (color may vary if you changed the editor theme).
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. For the assignment portion of the worksheet, 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, submit the completed exercise by creating a **pull request** via GitHub. Your pull request should include this file `AlphaDiversity_Worskheet.Rmd` and the PDF output of `Knitr` (`AlphaDiversity_Worskheet.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 `5.AlphaDiversity` folder, and 4) Load the `vegan` R package (be sure to install first if you haven't already).

```
options(repos = list(CRAN="http://cran.rstudio.com/"))
getwd()
```

```
## [1] "/Users/ericanadolski/GitHub/QB2023_Nadolski/2.Worksheets/5.AlphaDiversity"
```

```
setwd("/Users/ericanadolski/GitHub/QB2023_Nadolski/2.Worksheets/5.AlphaDiversity")
install.packages("vegan")
```

```
##
## The downloaded binary packages are in
##  /var/folders/7j/ntpmcppd2gb8_5v7mlyv0h840000gn/T//RtmpZCJhRm/downloaded_packages
```

```
require("vegan")
```

```
## Loading required package: vegan

## Loading required package: permute

## Loading required package: lattice

## This is vegan 2.6-4
```

## 2) LOADING DATA

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

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

```
## 'data.frame':    50 obs. of  225 variables:
##  - attr(*, "original.names")= chr [1:225] "Abarema.macradenium" "Acacia.melanoceras" "Acalypha.divers
```

## 3) SPECIES RICHNESS

**Species richness (S)** refers to 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` of the BCI data set, and
3. Compare the output of your function to the output of the `specnumber()` function in `vegan`.

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

```
##  1
## 93
```

```
specnumber(BCI[1,])
```

```
##  1
## 93
```

```
S.obs(BCI[1:4,])
```

```
##  1  2  3  4
## 93 84 90 94
```

***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 four sites (i.e., rows) of the BCI matrix?

> ***Answer 1***: Yes, our DIY function returns the same richness value as the vegan function. Richness for site 1 is 93 species, site 2 is 84 species, site 3 is 90 species, 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 - (rowSums(x==1) / rowSums(x))
}
coverage <- cbind(C(BCI[1:50,]))
max(coverage)
```

```
## [1] 0.9468504
```

```
min(coverage)
```

```
## [1] 0.8705882
```

```
mean(coverage)
```

```
## [1] 0.9182232
```

***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` was represented by singletons?
d. Make some observations about coverage at the BCI plots.

> ***Answer 2a***: Good's coverage metric is basically a percentage of how much coverage of the total biodiversity you captured so the values can be between 0 and 1.

***Answer 2b***: If n1 equaled N then the calculation would be C=(1 - 1)=0 so it would be a very low coverage (it doesn't make sense to say the coverage was actually 0 but it is impossible to tell how little coverage you achieved).

***Answer 2c***: 7% were singletons

***Answer 2d***: The coverage for the BCI sites ranged from 87% to 94%, with a mean of 92%, which seems consistent and high overall, which will benefit future analyses and comparisons.

### Estimated richness

In the R code chunk below, do the following:

1. Load the microbial dataset (located in the `5.AlphaDiversity/data` folder),
2. Transform and transpose the data as needed (see handout),
3. Create a new vector (`soilbac1`) by indexing the bacterial OTU abundances of any site in the dataset,
4. Calculate the observed richness at that particular site, and
5. Calculate coverage of that 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,]
S.obs(soilbac1)
```

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

```
C(soilbac1)
```

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

```
sum(soilbac1[,1:11310])
```

```
## [1] 2025
```

***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***: 2025 total sequences

***Answer 3b***: 1074 OTUs

***Answer 3c***: coverage of soilbac1 was 0.65, so about 30% lower than BCI; this makes sense though as it is much easier to count medium-size trees as they did in BCI than to expect that DNA extractions would recover as high of a coverage from a bacterial sample from the environment.

**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 `site1` and `soilbac1`.

```r
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 #convert SbyS to presence/absence
  Q1 = sum(colSums(SbyS.pa) ==1) #spp. observed once
  Q2 = sum(colSums(SbyS.pa) ==2) #spp. observed twice
  S.chao2 = S.obs(x) + (Q1^2)/(2*Q2)
  return(S.chao2)
}
S.ace <- function(x = "", thresh=10){
  x <- x[x>0]                           #excludes zero-abundance taxa
  S.abund <- length(which(x > thresh)) #richness of abundant taxa
  S.rare <- length(which(x <= thresh)) #richness of rare taxa
  singlt <- length(which(x == 1)) #number of singletons
  N.rare <- sum(x[which(x <= thresh)]) #abundance of rare taxa
  C.ace <- 1 - (singlt/N.rare) #coverage (proportion non-singleton rare taxa)
  i <- c(1:thresh)            #threshold abundance range
  count <- function(i,y){     #counter to go through i range
    length(y[y==i])
  }
  a.1 <- sapply(i, count, x) #number of individuals in richness i richness classes
  f.1 <- (i * (i - 1)) * a.1 #k(k-1)kf sensu Gotelli
  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)
}

S.chao1(BCI[1,])
```

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

```r
S.chao1(soilbac1)
```

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

```
S.chao2("1", BCI)
```

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

```
S.chao2("T1_1",soilbac.t)
```

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

```
S.ace(BCI[1,],thresh=10)
```

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

```
S.ace(soilbac1,thresh=10)
```

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

***Question 4***: What is the difference between ACE and the Chao estimators? Do the estimators give consistent results? Which one would you choose to use and why?

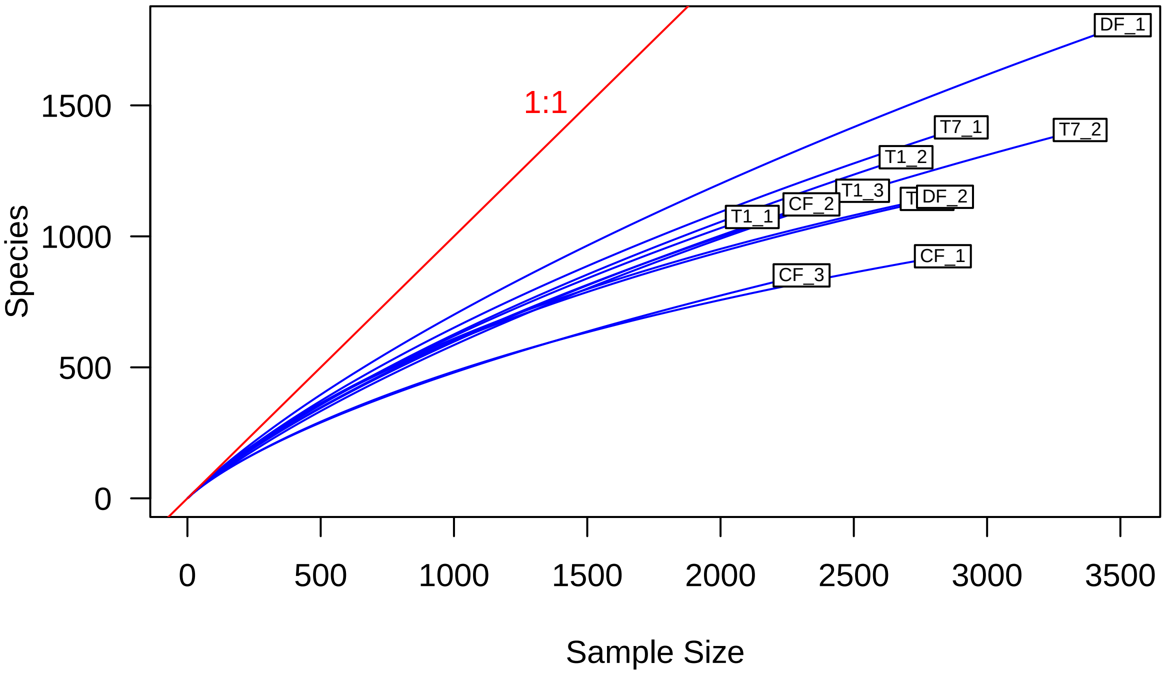
> ***Answer 4***: For the BCI site, Chao1 gave estimated richness of ~120 spp, Chao2: 104, and ACE: 159. For the soilbac data, Chao1: 2629 spp, Chao2: 21055, and ACE: 4466. The estimators give more consistent results for datasets with fewer rare taxa (BCI), but for the bacterial data the estimates differed by an order of magnitude.I would probably choose to use Chao1 since it seems the simplest to interpret :-)

### 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.

```
bac.Richness <- S.obs(soilbac.t[,1:13310])
sample.sizes <- rowSums(soilbac.t)
min.N <- min(sample.sizes)
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")
```



## 4) SPECIES EVNENNESS

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

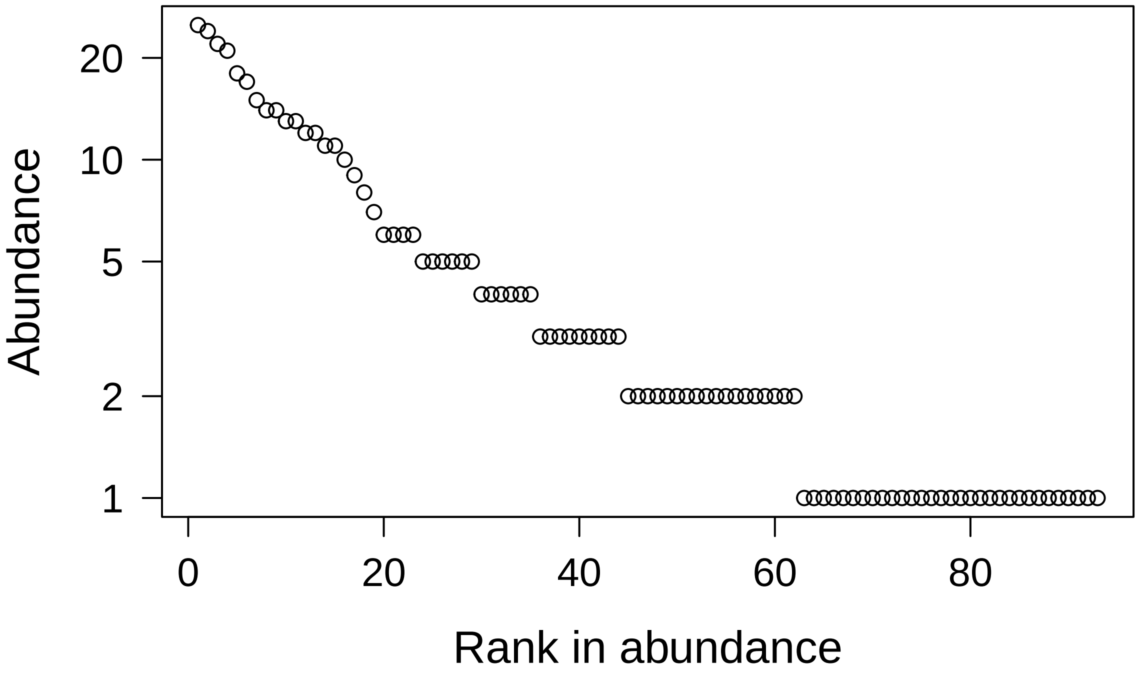
```
RAC <- function(x="")+{
  x.ab = x[x > 0]
  x.ab.ranked = x.ab[order(x.ab, decreasing=TRUE)]
  as.data.frame(lapply(x.ab.ranked, unlist))
  return(x.ab.ranked)
}
```

Now, let us 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="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)))
```



```
par <- opar
```

***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***: Log scale is useful for visualizing data with variance on orders of magnitude. Log scale diminishes the orders of magnitude of the variance, so it makes evenness appear higher/greater than it actually is.

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)
}

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) # error: C stack usage  7953896 is too close to the limit
```

```
## [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***: I would say they are largely in agreement, they differ only by 0.08. Wilson's estimate is slightly higher, likely because this method transforms to log abundances to lessen a bias toward the estimate being influenced by the largest abundances in the data set.

## 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)
}

1/SimpD(site1)
```

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

```
1-SimpD(site1)
```

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

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

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

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

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

### Fisher's $\alpha$

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

1. Provide the code for calculating Fisher's $\alpha$,
2. Calculate Fisher's $\alpha$ for `site1` of BCI.

```
?fisher.alpha()

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

fisher.alpha(rac)
```

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

***Question 7***: How is Fisher's $\alpha$ different from $E_{H'}$ and $E_{var}$? What does Fisher's $\alpha$ take into account that $E_{H'}$ and $E_{var}$ do not?

> ***Answer 7***: Fisher's alpha is an estimate of diversity, not simply a calculation, so it takes sampling error into account, which Shannon's and Simpson's do not.

## 6) HILL NUMBERS

Remember that we have learned about the advantages of Hill Numbers to measure and compare diversity among samples. We also learned to explore the effects of rare species in a community by examining diversity for a series of exponents $q$.

***Question 8***: Using `site1` of BCI and `vegan` package, a) calculate Hill numbers for $q$ exponent 0, 1 and 2 (richness, exponential Shannon's entropy, and inverse Simpson's diversity). b) Interpret the effect of rare species in your community based on the response of diversity to increasing exponent $q$.

```
q.zero <- specnumber(site1)
q.one <- diversity(site1, "shannon")
q.two <- diversity(site1, "invsimpson")

q.zero; q.one; q.two
```

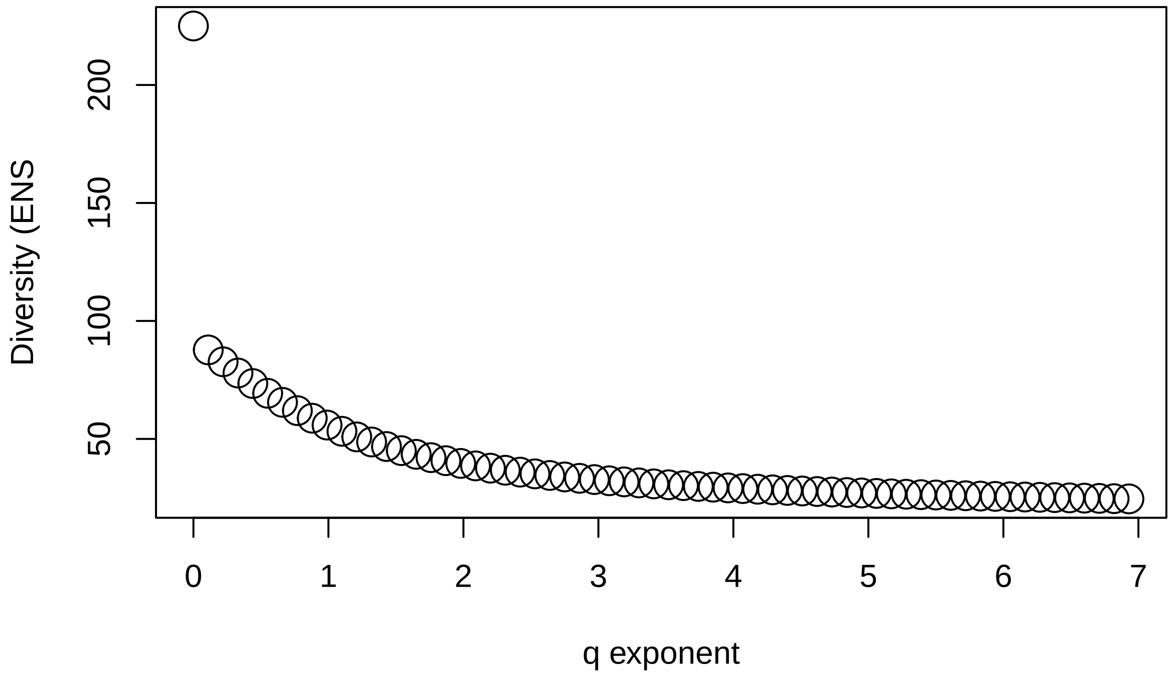
```
##  1
## 93
```

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

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

```
profile <- function(C){
  cbind(seq(0, 7, by=0.11),
        unlist(lapply(seq(0, 7, by=0.11), function(q) sum(apply(C, 1, function(x) (x/sum(x))^q))^(1/(1-
}
```

```
site1.profile <- profile(site1)
plot.new()
plot(site1.profile[,1], site1.profile[,2], cex=2,
     xlab="q exponent", ylab="Diversity (ENS")
```



***Answer 8a***: q^0 or richness was 93, q^1 or Shannon's entropy was 4.01, and q^2 or inverse Simpson's was 39.42. ***Answer 8b***: The profile of diversity plotted against increasing q exponents shows that richness is fairly high, but diversity begins to drop as abundant species are favored more; this means that the sample is composed of a mix of both abundant and rare species.

##7) MOVING BEYOND UNIVARIATE METRICS OF $\alpha$ 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.

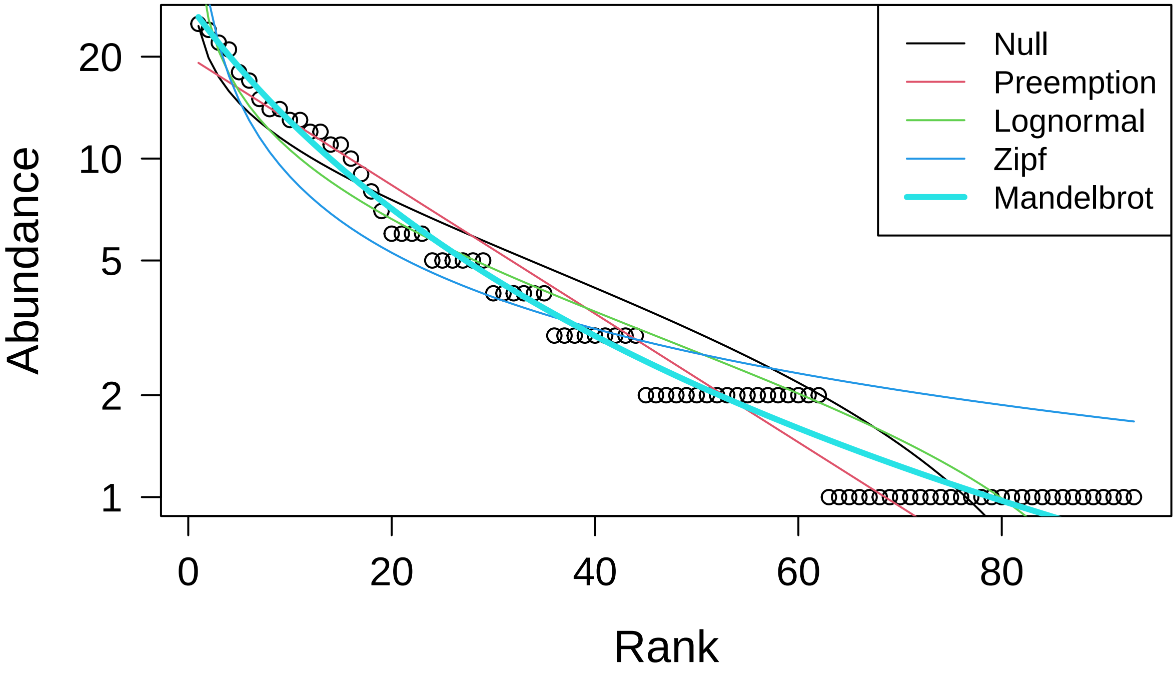
In the R code chunk below, please do the following:

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 Mandelbrot model has the lowest AIC and BIC values and visually fits the data the best. ***Answer 9b***: The Mandelbrot model can be inferred to point toward a model of late succession of a site: pioneer species had lower 'costs' to entry, therefore they are able to achieve high abundances, while species that arrive later during succession have more extensive 'costs' in terms of pre-existing conditions and pre-existing competitors leading to a long tail in the RAC of rare species (Bastow Journal of Vegetation Science 1991)

***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***: The preemption model appears to assume that each additional species will "preempt" or take over half of the remaining niche space that is left in an environment. ***Answer 10b***: The line is straight because in this model, there is a linear relationship between the each additional species added to the richness, and the abundance they can achieve via the resources they take up. The first most abundant species preempts the highest proportion, then the second most abundant takes up the next greatest proportion of resources, and so on.

***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***: Parameterization of a model is key to its interpretation, because the more parameters a model has, the better it will fit the data because it will pick up on even small influences of 'noise' in the data and fit to those idiosyncrasies. However, an over-fitted model with more parameters than could really be justified for that data set will likely not be broadly applicable to future data or be useful for predictions because it is relying on more randomness in the data than the main ecological factors that are actually influencing the variation.

## 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.

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.

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?

```r
### 1 finite Simpsons

finite.simp <- function(x=""){
  D = 0
  N = sum(x)
  for (n_i in x){
    D = D + (n_i * (n_i - 1))/(N * (N - 1))
  }
  return(D)
}

finite.simp(site1)
```

```
## [1] 0.02319032
```

```r
1/finite.simp(site1)
```

```
## [1] 43.12145
```

```r
1-finite.simp(site1)
```

```
## [1] 0.9768097
```

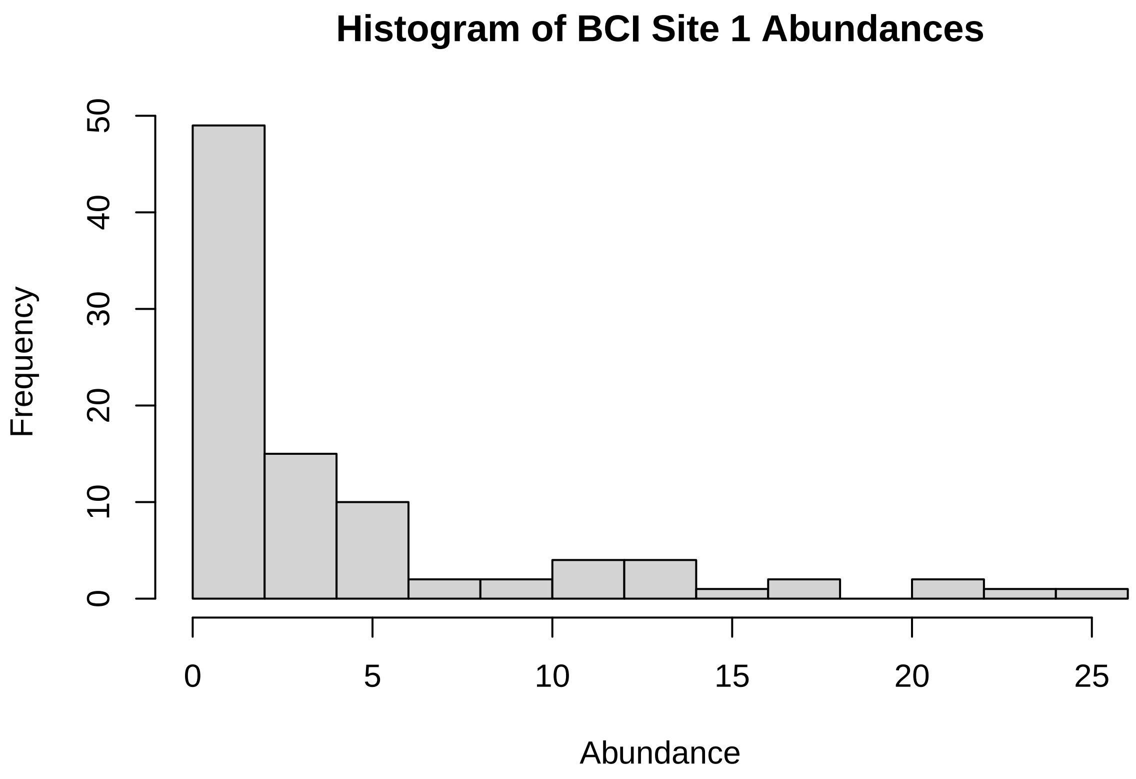
```r
### 2 histogram
?hist()

min(rac)
```

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

```r
max(rac)
```

```
## [1] 25
```

```r
breaks <- c(0,2,4,6,8,10,12,14,16,18,20,22,24,26)
hist(rac, breaks = breaks, main = paste("Histogram of BCI Site 1 Abundances"), xlab = paste("Abundance"
```



```
### Project site by species matrix

Ponds97 <- as.matrix(read.csv("/Users/ericanadolski/GitHub/QB2023_Nadolski/TeamProject/Pond97.csv", row
# raw read data matrix

Ponds.pa <- as.matrix(read.csv("/Users/ericanadolski/GitHub/QB2023_Nadolski/TeamProject/PondsPA.csv", r
# presence/absence matrix
Ponds.rel <- as.matrix(read.csv("/Users/ericanadolski/GitHub/QB2023_Nadolski/TeamProject/PondsREL.csv",
# relative abundance matrix

Ponds.env <- as.matrix(read.csv("/Users/ericanadolski/GitHub/QB2023_Nadolski/TeamProject/PondENV.csv",

nrow(Ponds.env) # number of sites
```

```
## [1] 58
```

```
ncol(Ponds97)
```

```
## [1] 34059
```

```
ponds.rac <- as.numeric(RAC(Ponds97[1,]))
length(ponds.rac)
```

```
## [1] 4311
```

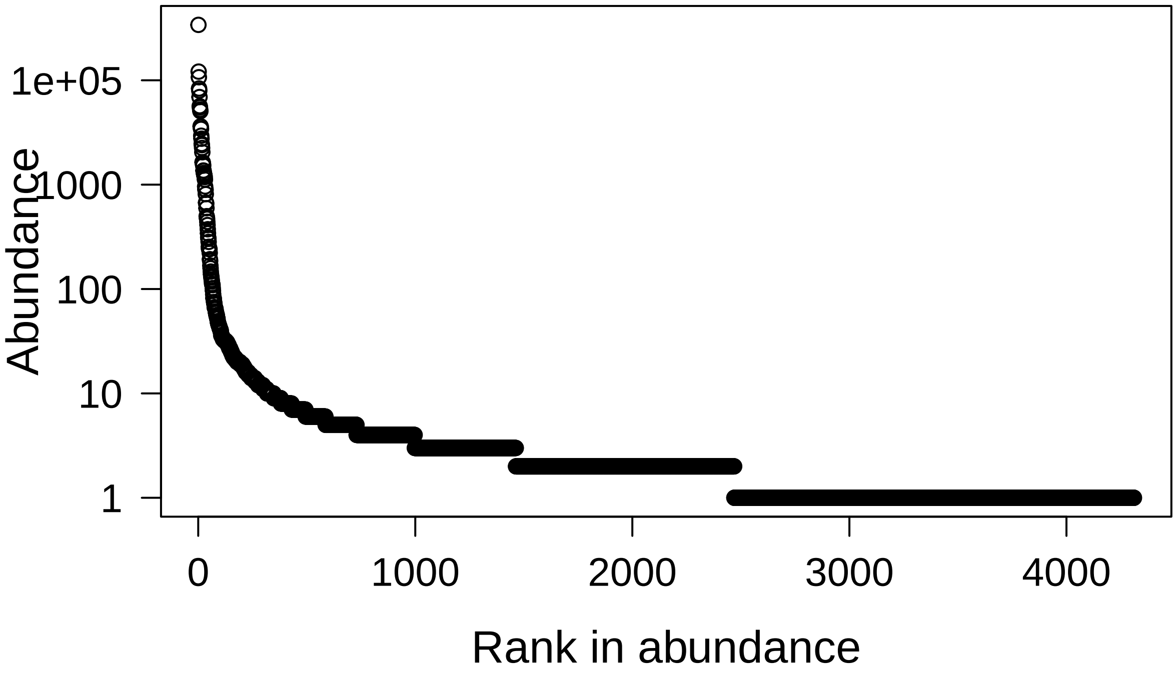
```
max(ponds.rac)
```

```
## [1] 33950
```

```
min(ponds.rac)
```

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

```
plot.new()
pond.ranks <- as.vector(seq(1, length(ponds.rac)))
plot(pond.ranks, log(ponds.rac), type="p", axes=F,
     xlab="Rank in abundance", ylab="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(1,10,100,1000,100000), at=log(c(1,10,100,1000,10000)))
```



>

***Synthesis Answer 2***: The histogram shows an inverse pattern from that seen in the scatterplot, with a high frequency of low abundances (the largest bar in the plot is that of abundances of 0-2), and a long tail of higher abundances stretching out to 25.

***Synthesis Answer 3***: We are using a Lennon lab dataset of microbial DNA and cDNA extracted from local ponds. There are 58 pond sites, and 34059 species (OTUs) across all the sites. There is high variance in abundance and evenness across all of the sites; based on exploratory rank abundance curves, there is high abundance of a few OTUs and a long tail of low-abundance OTUs.

## SUBMITTING YOUR ASSIGNMENT

Use Knitr to create a PDF of your completed 5.AlphaDiversity_Worksheet.Rmd document, push it to GitHub, and create a pull request. Please make sure your updated repo include both the pdf and RMarkdown files.

Unless otherwise noted, this assignment is due on **Wednesday, January 25th, 2023 at 12:00 PM (noon)**.