

5. Worksheet: Alpha Diversity

Madison Brown; Z620: Quantitative Biodiversity, Indiana University

27 January, 2025

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. 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 **Week-2/** folder folder, and 4) Load the **vegan** R package (be sure to install first if you have not already).

```
rm(list = ls())

getwd()

## [1] "/cloud/project/QB2025_Brown/Week2-Alpha"
setwd("/cloud/project/QB2025_Brown/Week2-Alpha")

library(vegan)

## Loading required package: permute
## Loading required package: lattice
```

```
## This is vegan 2.6-8
```

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)
```

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

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

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

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

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

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

```
## 1  
## 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 four sites (i.e., rows) of the BCI matrix?

Answer 1: Yes, `specnumber()` and `S.obs` do return the same values. The species richness of sites 1 through 4 is 93, 84, 90, and 94, respectively.

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))
}
```

C(BCI)

```
##      1      2      3      4      5      6      7      8
## 0.9308036 0.9287356 0.9200864 0.9468504 0.9287129 0.9174757 0.9326923 0.9443155
##      9     10     11     12     13     14     15     16
## 0.9095355 0.9275362 0.9152120 0.9071038 0.9242054 0.9132420 0.9350649 0.9267735
##     17     18     19     20     21     22     23     24
## 0.8950131 0.9193084 0.8891455 0.9114219 0.8946078 0.9066986 0.8705882 0.9030612
##     25     26     27     28     29     30     31     32
## 0.9095023 0.9115479 0.9088729 0.9198966 0.8983516 0.9221053 0.9382423 0.9411765
##     33     34     35     36     37     38     39     40
## 0.9220183 0.9239374 0.9267887 0.9186047 0.9379310 0.9306488 0.9268868 0.9386503
##     41     42     43     44     45     46     47     48
## 0.8880597 0.9299517 0.9140049 0.9168704 0.9234234 0.9348837 0.8847059 0.9228916
##     49     50
## 0.9086651 0.9143519
```

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: The range of values will be between 0 and 1.

Answer 2b: This would result in a Good's Coverage value of 0 which means that no intervals in your data had the ideal value that you are wanting to estimate.

Answer 2c: By using the numbers found from the formulas above and putting them into the Good's Coverage formula, when solving for n_1 I got 6.44. This means that there are about 6 singleton species in site 1.

Answer 2d: Based on the formula above, many of the Good's Coverage values are above .90 and the few that are not are still relatively high, with the lowest being at 0.87. This means that in all of the sites, a very large proportion of the intervals being used are capturing the true parameter.

Estimated richness

In the R code chunk below, do the following:

1. Load the microbial dataset (located in the `Week-2/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
```

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

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

Answer 3a: 50 sequences were recovered.

Answer 3b: The observed richness of `soilbac1` is 1074

Answer 3c: The coverage for the KBS sample is much lower (.6479471) than the coverage for the BCI sample (.9308036). This means that in the KBS sample, there is a smaller proportion of the intervals being used that are capturing the true parameter in comparison to the BCI sample.

Richness estimators

In the R code chunk below, do the following:

- Write a function to calculate **Chao1**,
- Write a function to calculate **Chao2**,
- Write a function to calculate **ACE**, and
- Use these functions to estimate richness at `site1` and `soilbac1`.

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

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

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

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

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

```
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.chao2(1, BCI)

##          1
## 104.6053

S.chao2(1, soilbac.t)

##      T1_1
## 21055.39

S.ace <- function(x = "", thresh = 10){
  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)
}

S.ace(soilbac1)

## [1] 4465.983

S.ace(BCI[1,])

## [1] 159.3404

```

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: Chao estimators are used to observe richness and the observed number of singletons and doubletons. Chao1 examines the richness of a single site while Chao2 is used to examine richness across multiple sites. Chao1 is referred to as an abundance-based estimator while Chao2 is an incidence-based estimator. On the other hand, ACE estimators look at the abundance of other rare species. Due to these two estimators analyzing different things, they will not necessarily give consistent results. You could conduct an Ace and Chao estimator on your data set, but it likely would not give the same result. The one I would choose would differ based on my dataset and what I am wanting to analyze. If I had a dataset with a high number of singletons and doubletons and smaller sample sizes, I would likely use a Chao estimator. On the other hand, if I had a very large sample size with a large amount of expected variability, I would likely use an ACE estimator.

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.

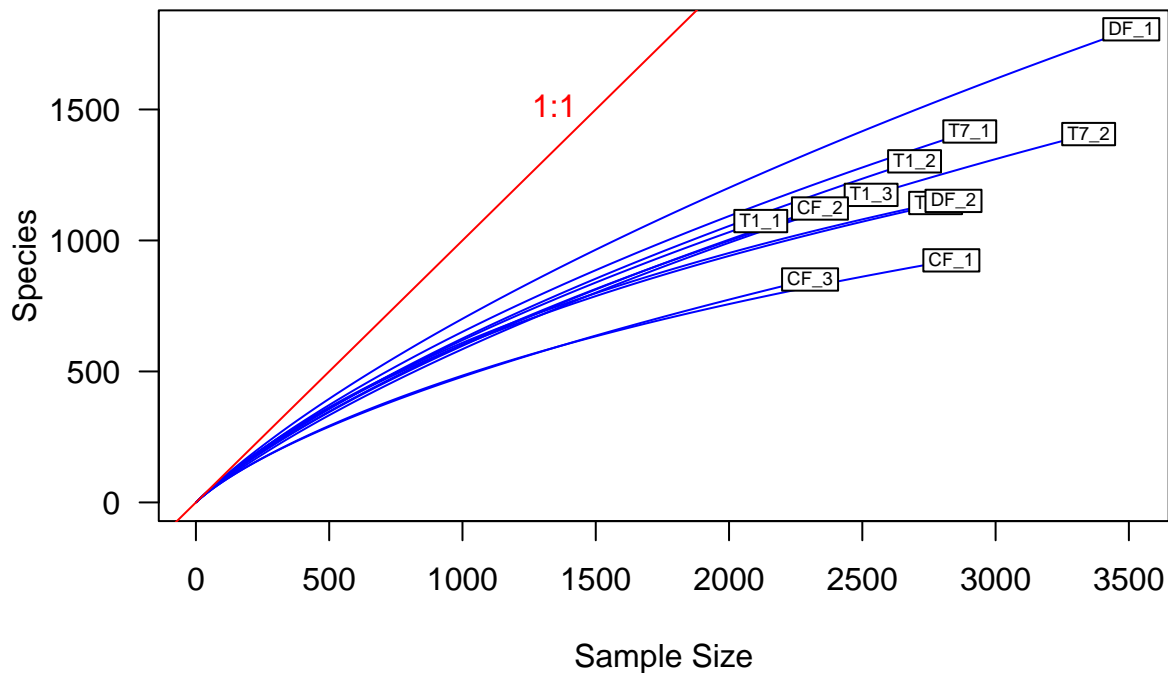
```
soilbac.S <- S.obs(soilbac.t)

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')
```



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.

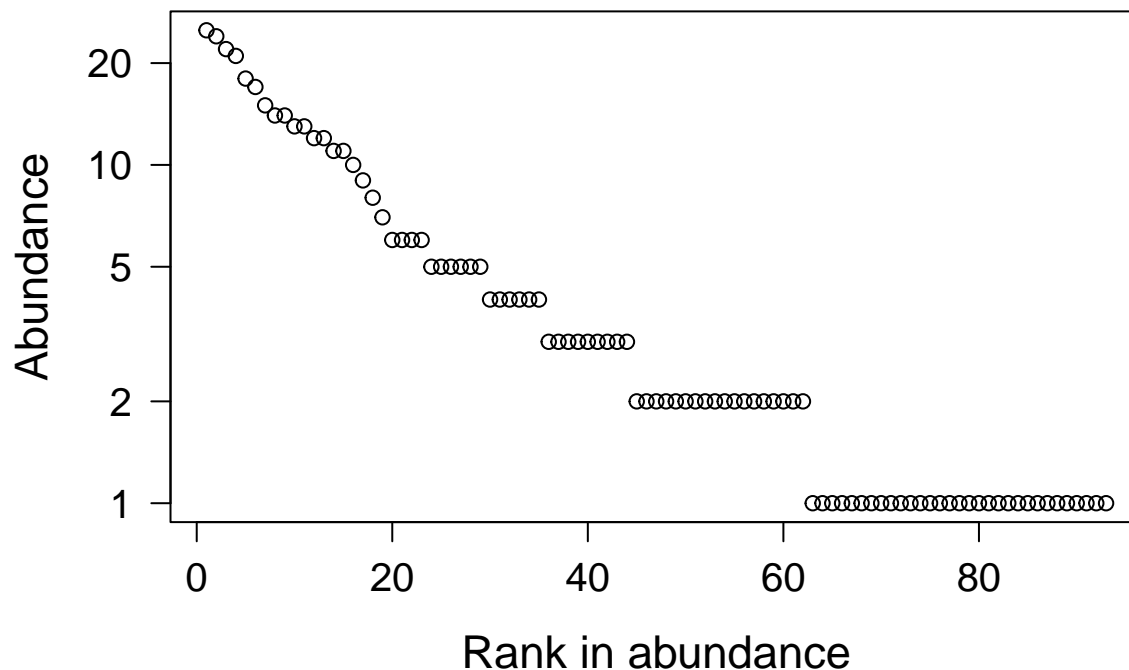
In the R code chunk below, do the following:

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)))
```



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: By using a log-scaled axis, it prevents one, or several, species from consuming the entire graph. When you have a large sample size, it can be difficult to graph the richness of all of these species in a clear manner. Using a log-scaled axis prevents extremely abundant species from consuming the graph and preventing you from seeing the less abundant species.

Now that we have visualized unevenness, 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: Using Simpson's Evenness, the value for `site1` was ~0.42 and with Smith and Wilson's Evenness Index, the value for `site1` was ~0.51. These values are relatively close to each other, yet there is still almost a 0.1 difference. This is likely due to the fact that Simpson's Evenness is influenced by large numbers; in this case that would be referring to the most abundant species. Smith and Wilson's Evenness Index was formulated to not allow highly abundant species to have a larger influence. In this case, the difference is not drastic and you can infer that `site1` has a

relatively moderate evenness. This means that there are some species who are more abundant than others, but it is not highly skewing the data.

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)
}

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)
}

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

diversity(site1, "inv")

## [1] 39.41555
```

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

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

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

Answer 7: Fisher's alpha is much more focused on rare species. Shannon's Diversity is used when you are wanting to analyze species richness and evenness. Simpson's Diversity focuses on dominant species and the value you get will be greater impacted by the presence of dominant species. Overall, if you are interested in analyzing further the presence of rare species in an area, Fisher's alpha may be more ideal to use because Shannon and Simpson's Diversity do not focus as much on rare species. Also, Fishers alpha takes into account sampling error because it estimates diversity as opposed to calculating diversity. Therefore, it takes into account the likelihood that you are not able to observe every unique individual in a community.

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 .

```
C1 <- data.frame(t(rep(1, 500))); colnames(C1) <- paste("sp", 1:500)
C2 <- data.frame(t(rep(1, 250))); colnames(C2) <- paste("sp", 1:250)
```

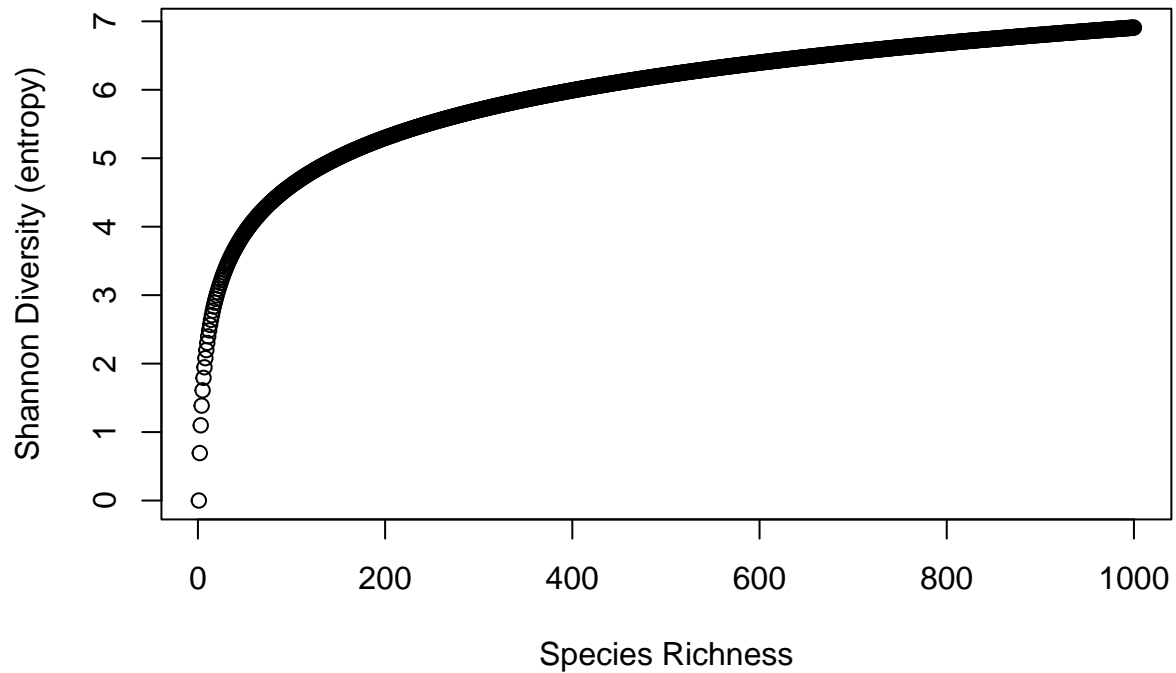
```
H1 <- diversity(C1, index = "shannon")
H2 <- diversity(C2, index = "shannon")
H1; H2
```

```
## [1] 6.214608
```

```
## [1] 5.521461
```

```
H_all <- matrix(ncol = 2, nrow = 1000)
for(i in 1:1000) {
  C <- data.frame(t(rep(1, i)))
  colnames(C) = paste("sp", 1:i)
  H_all[i,1] <- i
  H_all[i,2] <- diversity(C, index = "shannon")
}
```

```
plot(H_all[,1], H_all[,2], xlab = "Species Richness", ylab = "Shannon Diversity (entropy)")
```



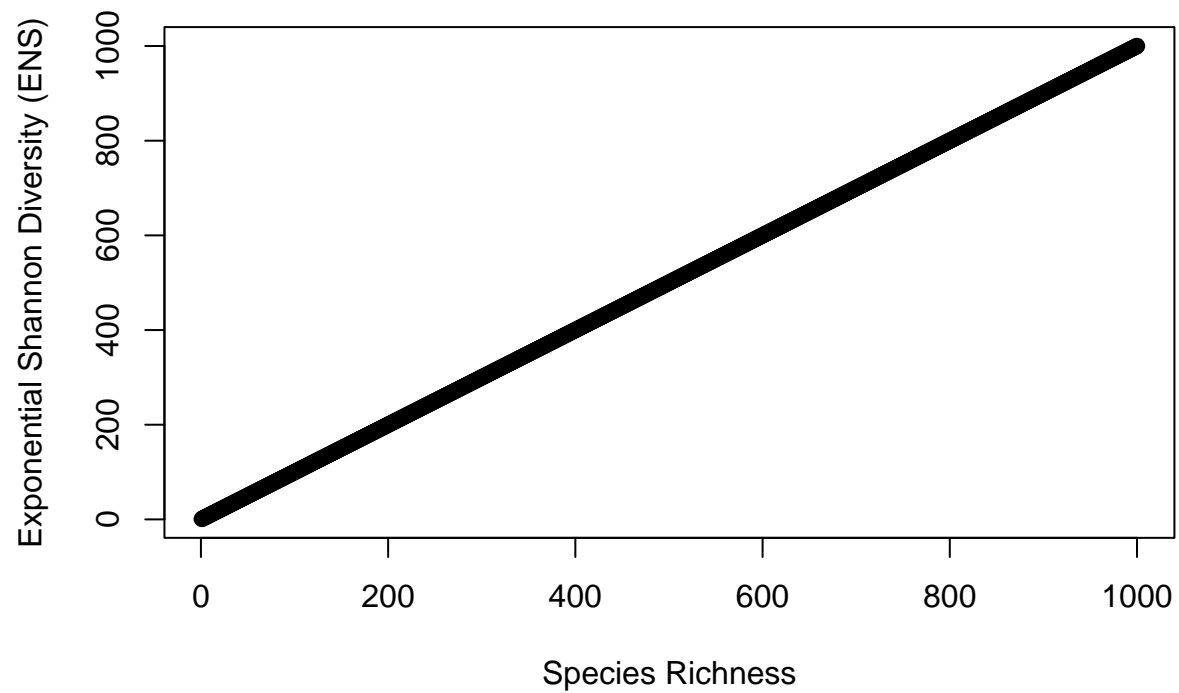
```
H1_Hill <- exp(diversity(C1,index = "shannon"))
H2_Hill <- exp(diversity(C2,index = "shannon"))
H1_Hill; H2_Hill
```

```
## [1] 500
```

```
## [1] 250
```

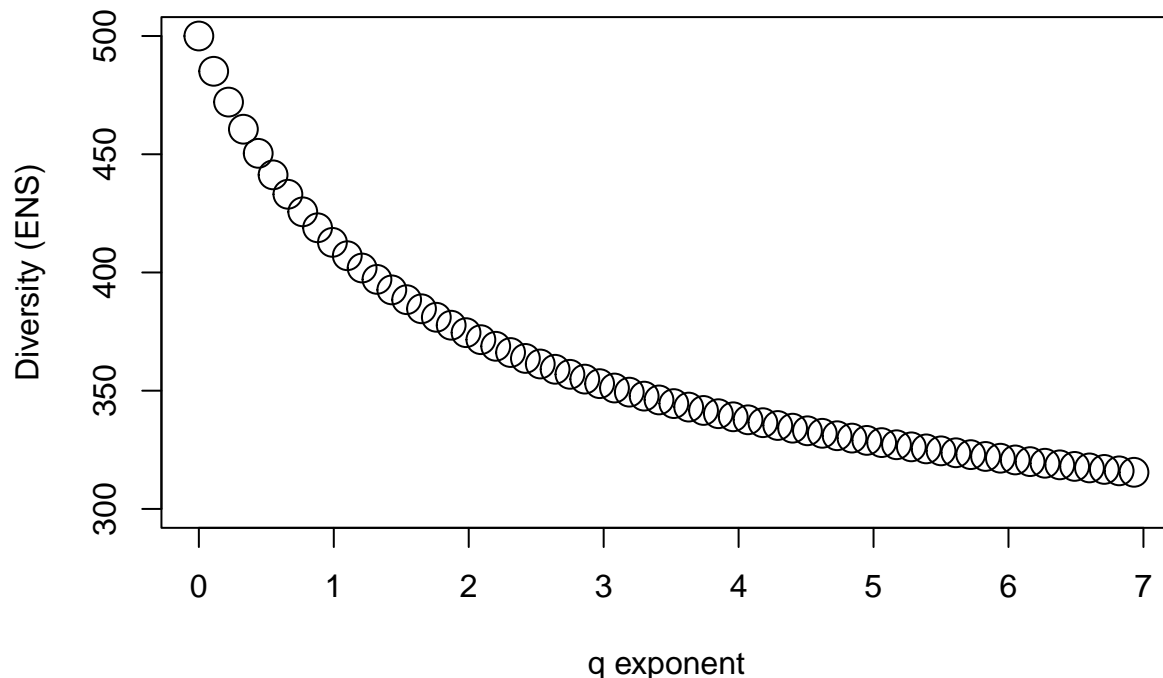
```
H_all_Hill <- matrix(ncol = 2,nrow = 1000)
for(i in 1:1000) {
  C <- data.frame(t(rep(1,i)))
  colnames(C) = paste("sp", 1:i)
  H_all_Hill[i,1] <- i
  H_all_Hill[i,2] <- exp(diversity(C, index = "shannon"))
}
```

```
plot(H_all_Hill[,1], H_all_Hill[,2], xlab = "Species Richness", ylab = "Exponential Shannon Diversity (")
```



```
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-q))))))
}

set.seed(42)
C3 <- data.frame(t(sample(1:1000, 500))); colnames(C3) = paste("sp", 1:500)
C3_profile <- profile(C3)
plot(C3_profile[,1], C3_profile[,2], ylim=c(300,500), cex = 2,
     xlab = "q exponent", ylab = "Diversity (ENS)")
```



```
library(vegan)

site1 <- BCI[1, ]

site1 <- site1[site1 > 0]

hill_0 <- specnumber(site1)

hill_1 <- exp(diversity(site1, index = "shannon"))

hill_2 <- 1 / diversity(site1, index = "simpson")

cat("Hill number for q = 0 (Richness):", hill_0, "\n")

## Hill number for q = 0 (Richness): 93

cat("Hill number for q = 1 (Exponential Shannon's entropy):", hill_1, "\n")

## Hill number for q = 1 (Exponential Shannon's entropy): 55.6127

cat("Hill number for q = 2 (Inverse Simpson's diversity):", hill_2, "\n")

## Hill number for q = 2 (Inverse Simpson's diversity): 1.026031
```

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 .

Answer 8a: The Hill number for $q = 0$ (Richness) is 93. The Hill number for $q = 1$ (Exponential Shannon's entropy) is 55.6127. The Hill number for $q = 2$ (Inverse Simpson's diversity) is 1.026031

Answer 8b: Based on the values above, it can be inferred that rare species do not have much of an affect on the community. For instance, the inverse Simpson's diversity of 1.02 is relatively low which means that there are a few very dominant species in the community. Therefore, most of the community abundance stems from those dominant species and not the rare species. Having a

species richness of 93 and an exponential Shannon's entropy of ~55 means that there is a relatively high amount of species present, yet the abundance is only moderately distributed. This means that there are still some rare species present, but not in a way that overwhelmingly impacts the community.

##7) 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.

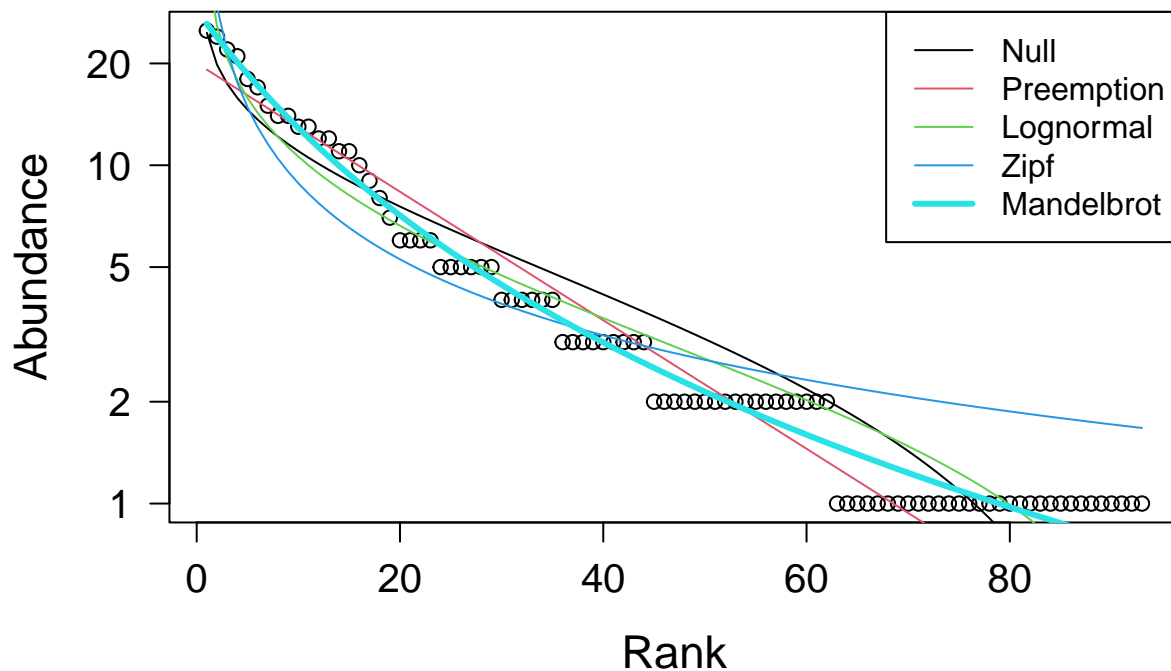
In the R code chunk below, please do the following:

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)
```

```
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: Based on the plot above, you can see that the Mandelbrot model best fits the rank abundance curve. The thick blue line most closely aligns with the plotted data from the rank abundance curve. It also begins to have a slight tail towards the right end of the graph, similar to the rank abundance curve. The other models do have some overlap, yet the Mandelbrot model overlaps the most. **Answer 9b:** Based on the Rank Abundance Curve and the Mandelbrot model, we can infer that our ecological community has many rare species. The models are slightly skewed in the bottom right tail and this is due to a large number of species having a very small abundance (rare species). Since the community has a large number of rare species, we can infer that the community is relatively healthy and abundant in resources since it can support many rare species. This could be seen as a negative thing because it would be relatively easy to lose much of the biodiversity in the area with so many species in low abundance. However, the graph does show that there are many species in high abundance which further supports that the population seems to be in stable condition.

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 assumes that there is a direct relationship between the total abundance and total resources that can be preempted. Therefore, as total abundance increases, so does the amount of resources that can be preempted. **Answer 10b:** A straight line in the niche preemption model likely means that resources are divided evenly. This would mean that species are not competing over resources. This further supports the statement above that this is a relatively stable and healthy community.

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: When a model has more parameters, it typically makes it better suited to fit a data set. However, too many parameters can lead to overfitting which does not accurately describe your data. Ideally, you want your model to represent your data well, but to not be too complex where overfitting occurs. Tools like the AIC and BIC are commonly used to determine which model works best for your data.

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.

```
site_1_data <- BCI[1, ]

N <- sum(site_1_data)

simpson_D <- sum((site_1_data / N)^2)

simpson_evenness <- 1 - simpson_D

simpson_inverse <- 1 / simpson_D

cat("Simpson's D:", simpson_D, "\n")

## Simpson's D: 0.0253707
```

```
cat("Simpson's Evenness (1 - D):", simpson_evenness, "\n")
```

```
## Simpson's Evenness (1 - D): 0.9746293
```

```
cat("Simpson's Inverse (1 / D):", simpson_inverse, "\n")
```

```
## Simpson's Inverse (1 / D): 39.41555
```

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.

```
site_1_data <- BCI[1, ]
```

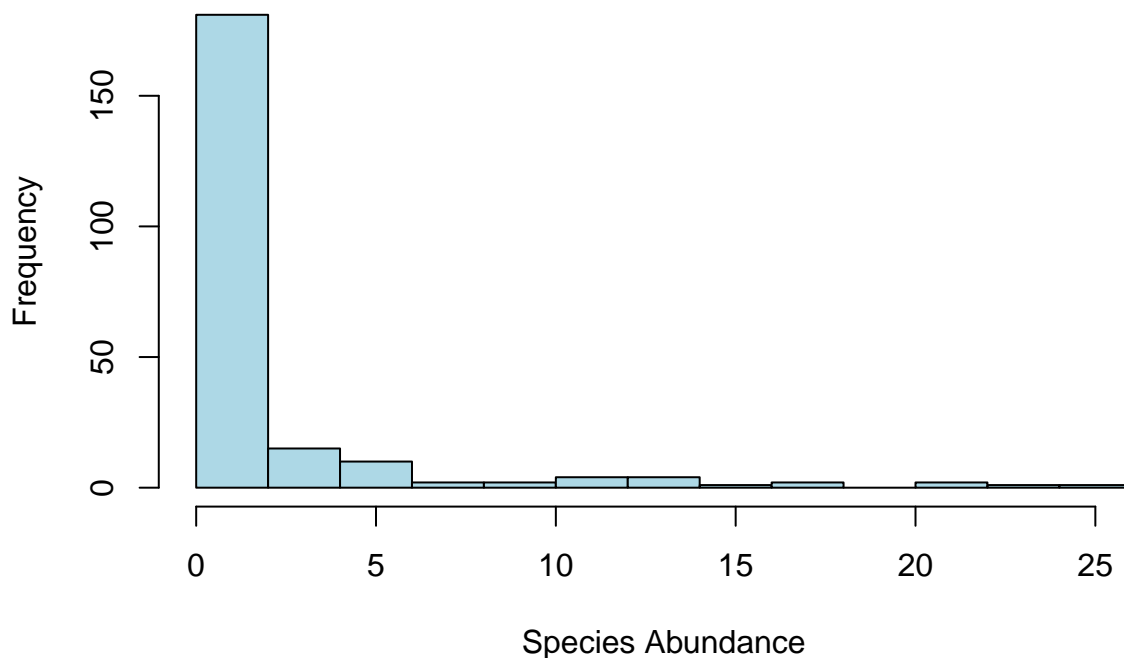
```
site_1_data_numeric <- as.numeric(site_1_data)
```

```
str(site_1_data_numeric)
```

```
## num [1:225] 0 0 0 0 0 0 2 0 0 0 ...
```

```
hist(site_1_data_numeric,  
      main = "Frequency Distribution for Site 1",  
      xlab = "Species Abundance",  
      ylab = "Frequency",  
      col = "lightblue",  
      border = "black")
```

Frequency Distribution for Site 1



Answer 2: > The histogram above shows that the data is heavily right skewed. From this you can infer that

there are a few species that are extremely abundant and then many species with very low abundance. This also means that there is extremely low evenness in this site.

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.

```
fish_data <- read.csv("/cloud/project/QB2025_Brown/Fish_Dataset.csv")
```

```
num_sites <- length(unique(fish_data$SiteID))
print(num_sites)
```

```
## [1] 2753
```

```
fish_data1 <- fish_data[,23:658]
```

```
length(fish_data[,23:658])
```

```
## [1] 636
```

```
site_species <- length(fish_data[,23:658])
```

```
colnames(fish_data1)
```

## [1] "Acantharchus.pomotis"	"Acanthogobius.flavimanus"
## [3] "Acipenser.brevirostrum"	"Acipenser.fulvescens"
## [5] "Acrocheilus.alutaceus"	"Agosia.chrysogaster"
## [7] "Alosa.aestivalis"	"Alosa.alabamae"
## [9] "Alosa.chrysochloris"	"Alosa.mediocris"
## [11] "Alosa.pseudoharengus"	"Alosa.sapidissima"
## [13] "Ambloplites.ariomus"	"Ambloplites.cavifrons"
## [15] "Ambloplites.constellatus"	"Ambloplites.rupestris"
## [17] "Ameiurus.brunneus"	"Ameiurus.catus"
## [19] "Ameiurus.melas"	"Ameiurus.natalis"
## [21] "Ameiurus.nebulosus"	"Ameiurus.platycephalus"
## [23] "Ameiurus.serracanthus"	"Amia.calva"
## [25] "Ammocrypta.beanii"	"Ammocrypta.clara"
## [27] "Ammocrypta.pellucida"	"Ammocrypta.vivax"
## [29] "Anchoa.mitchilli"	"Anguilla.rostrata"
## [31] "Aphredoderus.sayanus"	"Aplodinotus.grunniens"
## [33] "Archoplites.interruptus"	"Archosargus.probatoccephalus"
## [35] "Ascaphus.truei"	"Astyanax.mexicanus"
## [37] "Atractosteus.spatula"	"Awaous.banana"
## [39] "Brevoortia.patronus"	"Brevoortia.tyrannus"
## [41] "Campostoma.anomalum"	"Campostoma.oligolepis"
## [43] "Campostoma.pauciradii"	"Carassius.auratus"
## [45] "Carpiodes.carpio"	"Carpiodes.cyprinus"
## [47] "Carpiodes.velifer"	"Catostomus.ardens"
## [49] "Catostomus.catostomus"	"Catostomus.clarkii"
## [51] "Catostomus.columbianus"	"Catostomus.commerstonii"
## [53] "Catostomus.discobolus"	"Catostomus.insignis"
## [55] "Catostomus.latipinnis"	"Catostomus.macrocheilus"
## [57] "Catostomus occidentalis"	"Catostomus.platyrrhynchus"
## [59] "Catostomus.plebeius"	"Catostomus.rimiculus"
## [61] "Catostomus.santaanae"	"Catostomus.snyderi"
## [63] "Catostomus.tahoensis"	"Centrarchus.macropterus"
## [65] "Centropomus.parallelus"	"Centropomus.undecimalis"
## [67] "Chaenobryttus.gulosus"	"Channa.argus"

## [69]	"Chologaster.cornuta"	"Chrosomus.cumberlandensis"
## [71]	"Chrosomus.eos"	"Chrosomus.erythrogaster"
## [73]	"Chrosomus.neogaeus"	"Chrosomus.oreas"
## [75]	"Chrosomus.tennesseensis"	"Cichlasoma.cyanoguttatum"
## [77]	"Cichlasoma.urophthalmus"	"Citharichthys.spilopterus"
## [79]	"Clarias.batrachus"	"Clinostomus.elongatus"
## [81]	"Clinostomus.funduloides"	"Conger.conger"
## [83]	"Coregonus.clupeaformis"	"Cottus.aleuticus"
## [85]	"Cottus.asper"	"Cottus.bairdii"
## [87]	"Cottus.beldingii"	"Cottus.caeruleomentum"
## [89]	"Cottus.carolinae"	"Cottus.chattahoochee"
## [91]	"Cottus.cognatus"	"Cottus.confusus"
## [93]	"Cottus.girardi"	"Cottus.gulosus"
## [95]	"Cottus.hubbsi"	"Cottus.hypselurus"
## [97]	"Cottus.kanawhae"	"Cottus.klamathensis"
## [99]	"Cottus.leiopomus"	"Cottus.perplexus"
## [101]	"Cottus.rhotheus"	"Cottus.tenuis"
## [103]	"Couesius.plumbeus"	"Crystallaria.asprella"
## [105]	"Ctenogobius.shufeldti"	"Ctenopharyngodon.idella"
## [107]	"Culaea.inconstans"	"Cycleptus.elongatus"
## [109]	"Cycleptus.meridionalis"	"Cyprinella.analostana"
## [111]	"Cyprinella.callisema"	"Cyprinella.callistia"
## [113]	"Cyprinella.camura"	"Cyprinella.chloristia"
## [115]	"Cyprinella.formosa"	"Cyprinella.galactura"
## [117]	"Cyprinella.gibbsi"	"Cyprinella.leedsi"
## [119]	"Cyprinella.lepida"	"Cyprinella.lutrensis"
## [121]	"Cyprinella.nivea"	"Cyprinella.pyrrhomelas"
## [123]	"Cyprinella.spiloptera"	"Cyprinella.trichroistia"
## [125]	"Cyprinella.venusta"	"Cyprinella.whipplei"
## [127]	"Cyprinella.xaenura"	"Cyprinodon.rubrofluviatilis"
## [129]	"Cyprinodon.variegatus"	"Cyprinus.carpio"
## [131]	"Dicamptodon.tenebrosus"	"Dionda.nigrotaeniata"
## [133]	"Dionda.serena"	"Dormitator.maculatus"
## [135]	"Dorosoma.cephedianum"	"Dorosoma.petenense"
## [137]	"Elassoma.evergladei"	"Elassoma.zonatum"
## [139]	"Elops.saurus"	"Elotris.amblyopsis"
## [141]	"Enneacanthus.chaetodon"	"Enneacanthus.gloriosus"
## [143]	"Enneacanthus.obesus"	"Entosphenus.lethophagus"
## [145]	"Entosphenus.tridentatus"	"Erimonax.monachus"
## [147]	"Erimystax.dissimilis"	"Erimystax.harryi"
## [149]	"Erimystax.insignis"	"Erimystax.x.punctatus"
## [151]	"Erimyzon.oblongus"	"Erimyzon.sucetta"
## [153]	"Erimyzon.tenuis"	"Esox.americanus"
## [155]	"Esox.americanus.americanus"	"Esox.americanus.vermiculatus"
## [157]	"Esox.lucius"	"Esox.masquinongy"
## [159]	"Esox.niger"	"Etheostoma.acuticeps"
## [161]	"Etheostoma.aquali"	"Etheostoma.artesiaie"
## [163]	"Etheostoma.asprigene"	"Etheostoma.autumnale"
## [165]	"Etheostoma.baileyi"	"Etheostoma.barrenense"
## [167]	"Etheostoma.bellum"	"Etheostoma.blennioides"
## [169]	"Etheostoma.blennius"	"Etheostoma.boschungii"
## [171]	"Etheostoma.caeruleum"	"Etheostoma.camurum"
## [173]	"Etheostoma.chlorobranchium"	"Etheostoma.chlorosomum"
## [175]	"Etheostoma.chuckwachatte"	"Etheostoma.cinereum"

## [177]	"Etheostoma.collettei"	"Etheostoma.collis"
## [179]	"Etheostoma.colorosum"	"Etheostoma.coosae"
## [181]	"Etheostoma.cragini"	"Etheostoma.crossopterus"
## [183]	"Etheostoma.derivativum"	"Etheostoma.ditrema"
## [185]	"Etheostoma.duryi"	"Etheostoma.edwini"
## [187]	"Etheostoma.euzonum"	"Etheostoma.exile"
## [189]	"Etheostoma.flabellare"	"Etheostoma.flavum"
## [191]	"Etheostoma.fragi"	"Etheostoma.fricksium"
## [193]	"Etheostoma.fusiforme"	"Etheostoma.gracile"
## [195]	"Etheostoma.gutselli"	"Etheostoma.histrion"
## [197]	"Etheostoma.hopkinsi"	"Etheostoma.inscriptum"
## [199]	"Etheostoma.jessiae"	"Etheostoma.jordani"
## [201]	"Etheostoma.juliae"	"Etheostoma.kantuckeense"
## [203]	"Etheostoma.kennicottii"	"Etheostoma.lachneri"
## [205]	"Etheostoma.lepidum"	"Etheostoma.longimanum"
## [207]	"Etheostoma.luteovinctum"	"Etheostoma.lynceum"
## [209]	"Etheostoma.maculatum"	"Etheostoma.mariae"
## [211]	"Etheostoma.mihileze"	"Etheostoma.neopterus"
## [213]	"Etheostoma.nigripinne"	"Etheostoma.nigrum"
## [215]	"Etheostoma.occidentale"	"Etheostoma.olmstedii"
## [217]	"Etheostoma.oophylax"	"Etheostoma.osburni"
## [219]	"Etheostoma.parvipinne"	"Etheostoma.podostemone"
## [221]	"Etheostoma.proeliare"	"Etheostoma.punctulatum"
## [223]	"Etheostoma.radiosum"	"Etheostoma.rafinesquei"
## [225]	"Etheostoma.ramseyi"	"Etheostoma.rufilineatum"
## [227]	"Etheostoma.rupestre"	"Etheostoma.sanguifluum"
## [229]	"Etheostoma.scottii"	"Etheostoma.serrifer"
## [231]	"Etheostoma.simoterum"	"Etheostoma.spectabile"
## [233]	"Etheostoma.squamiceps"	"Etheostoma.stigmaeum"
## [235]	"Etheostoma.swaini"	"Etheostoma.swannanoa"
## [237]	"Etheostoma.tallapoosae"	"Etheostoma.tetrazonum"
## [239]	"Etheostoma.thalassinum"	"Etheostoma.tippecanoe"
## [241]	"Etheostoma.variatum"	"Etheostoma.virgatum"
## [243]	"Etheostoma.vitreum"	"Etheostoma.vulneratum"
## [245]	"Etheostoma.whipplei"	"Etheostoma.zonale"
## [247]	"Etheostoma.zonistium"	"Eucinostomus.argenteus"
## [249]	"Exoglossum.laurae"	"Exoglossum.maxillingua"
## [251]	"Forbesichthys.agassizii"	"Fundulus.catenatus"
## [253]	"Fundulus.chrysotus"	"Fundulus.diaphanus"
## [255]	"Fundulus.dispar"	"Fundulus.euryzonus"
## [257]	"Fundulus.grandis"	"Fundulus.heteroclitus"
## [259]	"Fundulus.kansae"	"Fundulus.lineolatus"
## [261]	"Fundulus.notatus"	"Fundulus.olivaceus"
## [263]	"Fundulus.rathbuni"	"Fundulus.sciadicus"
## [265]	"Fundulus.seminolis"	"Fundulus.stellifer"
## [267]	"Fundulus.stellifera"	"Fundulus.zebrinus"
## [269]	"Gambusia.affinis"	"Gambusia.holbrooki"
## [271]	"Gasterosteus.aculeatus"	"Gerres.cinereus"
## [273]	"Gila.atriaria"	"Gila.bicolor"
## [275]	"Gila.cypha"	"Gila.elegans"
## [277]	"Gila.orcuttii"	"Gila.pandora"
## [279]	"Gila.robusta"	"Gila.seminuda"
## [281]	"Gobioides.broussonnetii"	"Gobiomorus.dormitor"
## [283]	"Gobiosoma.bosc"	"Hemitremia.flammea"

## [285] "Herichthys.cyanoguttatum"	"Hesperoleucus.symmetricus"
## [287] "Heterandria.formosa"	"Hiodon.alosoides"
## [289] "Hiodon.tergatus"	"Hoplosternum.littorale"
## [291] "Hybognathus.amarus"	"Hybognathus.argyritis"
## [293] "Hybognathus.hankinsoni"	"Hybognathus.hayi"
## [295] "Hybognathus.nuchalis"	"Hybognathus.placitus"
## [297] "Hybognathus.regius"	"Hybopsis.amblops"
## [299] "Hybopsis.amnis"	"Hybopsis.dorsalis"
## [301] "Hybopsis.hypsinothos"	"Hybopsis.lineapunctata"
## [303] "Hybopsis.rubrifrons"	"Hybopsis.winchelli"
## [305] "Hybopsis.zanema"	"Hypentelium.etowanum"
## [307] "Hypentelium.nigricans"	"Hypentelium.roanokense"
## [309] "Hypophthalmichthys.molitrix"	"Hypophthalmichthys.nobilis"
## [311] "Hypostomus.plecostomus"	"Hysterothorax.traskii"
## [313] "Ichthyomyzon.bdelium"	"Ichthyomyzon.castaneus"
## [315] "Ichthyomyzon.fossor"	"Ichthyomyzon.gagei"
## [317] "Ichthyomyzon.greeleyi"	"Ichthyomyzon.unicuspis"
## [319] "Ictalurus.furcatus"	"Ictalurus.punctatus"
## [321] "Ictiobus.bubalus"	"Ictiobus.cyprinellus"
## [323] "Ictiobus.niger"	"Iotichthys.phlegethontis"
## [325] "Labidesthes.sicculus"	"Lampetra.aepyptera"
## [327] "Lethenteron.appendix"	"Lampetra.richardsoni"
## [329] "Lampetra.tridentata"	"Lavinia.exilicauda"
## [331] "Leiostomus.xanthurus"	"Lepidomeda.copei"
## [333] "Lepidomeda.mollispinis"	"Lepidomeda.vittata"
## [335] "Lepisosteus.oculatus"	"Lepisosteus.osseus"
## [337] "Lepisosteus.platostomus"	"Lepisosteus.platyrrhynchus"
## [339] "Lepomis.auritus"	"Lepomis.cyanellus"
## [341] "Lepomis.gibbosus"	"Lepomis.gulosus"
## [343] "Lepomis.humilis"	"Lepomis.macrochirus"
## [345] "Lepomis.marginatus"	"Lepomis.megalotis"
## [347] "Lepomis.microlophus"	"Lepomis.miniatus"
## [349] "Lepomis.punctatus"	"Lepomis.symmetricus"
## [351] "Lota.lota"	"Lucania.goodei"
## [353] "Lucania.parva"	"Luxilus.albeolus"
## [355] "Luxilus.cardinalis"	"Luxilus.cerasinus"
## [357] "Luxilus.chrysocephalus"	"Luxilus.coccogenis"
## [359] "Luxilus.cornutus"	"Luxilus.pilsbryi"
## [361] "Luxilus.zonatus"	"Luxilus.zonistius"
## [363] "Lythrurus.ardens"	"Lythrurus.atrapiculus"
## [365] "Lythrurus.bellus"	"Lythrurus.fasciolaris"
## [367] "Lythrurus.fumeus"	"Lythrurus.lirus"
## [369] "Lythrurus.roseipinnis"	"Lythrurus.umbratilis"
## [371] "Macrhybopsis.aestivalis"	"Macrhybopsis.australis"
## [373] "Macrhybopsis.hystoma"	"Macrhybopsis.marconis"
## [375] "Macrhybopsis.meeki"	"Macrhybopsis.storeriana"
## [377] "Margariscus.margarita"	"Margariscus.nachtriebi"
## [379] "Membras.martinica"	"Menidia.audens"
## [381] "Menidia.beryllina"	"Menidia.menidia"
## [383] "Microgadus.tomcod"	"Microgobius.gulosus"
## [385] "Micropterus.cataractae"	"Micropterus.coosae"
## [387] "Micropterus.dolomieu"	"Micropterus.notius"
## [389] "Micropterus.punctulatus"	"Micropterus.salmoides"
## [391] "Micropterus.treculii"	"Minytrema.melanops"

## [393]	"Misgurnus.anguillicaudatus"	"Morone.americana"
## [395]	"Morone.chrysops"	"Morone.mississippiensis"
## [397]	"Morone.saxatilis"	"Moxostoma.anisurum"
## [399]	"Moxostoma.breviceps"	"Moxostoma.carinatum"
## [401]	"Moxostoma.cervinum"	"Moxostoma.collapsum"
## [403]	"Moxostoma.congestum"	"Moxostoma.duquesnii"
## [405]	"Moxostoma.erythrurum"	"Moxostoma.lachneri"
## [407]	"Moxostoma.macrolepidotum"	"Moxostoma.pappillosum"
## [409]	"Moxostoma.pisolabrum"	"Moxostoma.poecilurum"
## [411]	"Moxostoma.robustum"	"Moxostoma.rupiscartes"
## [413]	"Moxostoma.valenciennesi"	"Mugil.cephalus"
## [415]	"Mugil.curema"	"Mylocheilus.caurinus"
## [417]	"Mylopharodon.conocephalus"	"Mylopharyngodon.piceus"
## [419]	"Myrophis.punctatus"	"Neogobius.melanostomus"
## [421]	"Nocomis.asper"	"Nocomis.biguttatus"
## [423]	"Nocomis.effusus"	"Nocomis.leptocephalus"
## [425]	"Nocomis.micropogon"	"Nocomis.platyrrhynchus"
## [427]	"Nocomis.raneyi"	"Notemigonus.crysoleucas"
## [429]	"Notropis.alborus"	"Notropis.altipinnis"
## [431]	"Notropis.amabilis"	"Notropis.ammophilus"
## [433]	"Notropis.amoenus"	"Notropis.amplamala"
## [435]	"Notropis.ariommus"	"Notropis.atherinoides"
## [437]	"Notropis.atrocaudalis"	"Notropis.baileyi"
## [439]	"Notropis.bairdi"	"Notropis.bifrenatus"
## [441]	"Notropis.blennius"	"Notropis.boops"
## [443]	"Notropis.braytoni"	"Notropis.buccatus"
## [445]	"Notropis.buccula"	"Notropis.buchanani"
## [447]	"Notropis.cahabae"	"Notropis.candidus"
## [449]	"Notropis.chalybaeus"	"Notropis.chiliticus"
## [451]	"Notropis.chlorocephalus"	"Notropis.chrosomus"
## [453]	"Notropis.cummingsae"	"Notropis.edwardraneyi"
## [455]	"Notropis.girardi"	"Notropis.greenei"
## [457]	"Notropis.harperi"	"Notropis.heterodon"
## [459]	"Notropis.heterolepis"	"Notropis.hudsonius"
## [461]	"Notropis.hypsilepis"	"Notropis.leuciodus"
## [463]	"Notropis.longirostris"	"Notropis.lutipinnis"
## [465]	"Notropis.maculatus"	"Notropis.micropteryx"
## [467]	"Notropis.nubilus"	"Notropis.oxyrhynchus"
## [469]	"Notropis.ozarcanus"	"Notropis.percobromus"
## [471]	"Notropis.perpallidus"	"Notropis.petersoni"
## [473]	"Notropis.photogenis"	"Notropis.potteri"
## [475]	"Notropis.procne"	"Notropis.rafinesquei"
## [477]	"Notropis.rubellus"	"Notropis.rubricroceus"
## [479]	"Notropis.sabinae"	"Notropis.scabriceps"
## [481]	"Notropis.scepticus"	"Notropis.semperasper"
## [483]	"Notropis.shumardi"	"Notropis.simus"
## [485]	"Notropis.spectrunculus"	"Notropis.stilbius"
## [487]	"Notropis.stramineus"	"Notropis.suttkusi"
## [489]	"Notropis.telescopus"	"Notropis.texanus"
## [491]	"Notropis.topeka"	"Notropis.uranoscopus"
## [493]	"Notropis.volucellus"	"Notropis.wickliffi"
## [495]	"Notropis.xaenoccephalus"	"Noturus.albater"
## [497]	"Noturus.elegans"	"Noturus.eleutherus"
## [499]	"Noturus.exilis"	"Noturus.flavater"

## [501] "Noturus.flavus"	"Noturus.funnebris"
## [503] "Noturus.gyrinus"	"Noturus.hildebrandi"
## [505] "Noturus.insignis"	"Noturus.lachneri"
## [507] "Noturus.leptacanthus"	"Noturus.miurus"
## [509] "Noturus.munitus"	"Noturus.nocturnus"
## [511] "Noturus.phaeus"	"Noturus.stigmosus"
## [513] "Oncorhynchus.clarkii"	"Oncorhynchus.clarkii.bouvieri"
## [515] "Oncorhynchus.clarkii.clarkii"	"Oncorhynchus.clarkii.henshawi"
## [517] "Oncorhynchus.clarkii.lewisi"	"Oncorhynchus.clarkii.pleuriticus"
## [519] "Oncorhynchus.clarkii.utah"	"Oncorhynchus.clarkii.virginalis"
## [521] "Oncorhynchus.kisutch"	"Oncorhynchus.mykiss"
## [523] "Oncorhynchus.mykiss.gairdnerii"	"Oncorhynchus.nerka"
## [525] "Oncorhynchus.tshawytscha"	"Opsopoeodus.emiliae"
## [527] "Oreochromis.aureus"	"Oreochromis.mossambicus"
## [529] "Orthodon.microlepidotus"	"Paralichthys.lethostigma"
## [531] "Perca.flavescens"	"Percina.aurantiaca"
## [533] "Percina.aurolineata"	"Percina.burtoni"
## [535] "Percina.caprodes"	"Percina.carbonaria"
## [537] "Percina.copelandi"	"Percina.crassa"
## [539] "Percina.cymatotaenia"	"Percina.evides"
## [541] "Percina.gymnocephala"	"Percina.kathae"
## [543] "Percina.lenticula"	"Percina.macrocephala"
## [545] "Percina.macrolepida"	"Percina.maculata"
## [547] "Percina.nasuta"	"Percina.nevisense"
## [549] "Percina.nigrofasciata"	"Percina.notogramma"
## [551] "Percina.oxyrhynchus"	"Percina.palmaris"
## [553] "Percina.pantherina"	"Percina.peltata"
## [555] "Percina.phoxocephala"	"Percina.rex"
## [557] "Percina.roanoka"	"Percina.sciera"
## [559] "Percina.shumardi"	"Percina.squamata"
## [561] "Percina.stictogaster"	"Percina.suttkusi"
## [563] "Percina.tanasi"	"Percina.uranidea"
## [565] "Percina.vigil"	"Percopsis.omiscomaycus"
## [567] "Percopsis.transmontana"	"Petromyzon.marinus"
## [569] "Phenacobius.catostomus"	"Phenacobius.crassilabrum"
## [571] "Phenacobius.mirabilis"	"Phenacobius.teretulus"
## [573] "Phenacobius.uranops"	"Phoxinus.eos"
## [575] "Phoxinus.erythrogaster"	"Phoxinus.oreas"
## [577] "Pimephales.notatus"	"Pimephales.promelas"
## [579] "Pimephales.tenellus"	"Pimephales.vigilax"
## [581] "Plagopterus.argentissimus"	"Platichthys.stellatus"
## [583] "Platygobio.gracilis"	"Poecilia.latipinna"
## [585] "Poecilia.sphenops"	"Pogonichthys.macrolepidotus"
## [587] "Polyodon.spathula"	"Pomoxis.annularis"
## [589] "Pomoxis.nigromaculatus"	"Prosopium.cylindraceum"
## [591] "Prosopium.williamsoni"	"Pteronotropis.euryzonus"
## [593] "Pteronotropis.grandipinnis"	"Pteronotropis.hypsleopterus"
## [595] "Pteronotropis.signipinnis"	"Pteronotropis.stonei"
## [597] "Pterygoplichthys.disjunctivus"	"Pterygoplichthys.multiradiatus"
## [599] "Ptychocheilus.grandis"	"Ptychocheilus.lucius"
## [601] "Ptychocheilus.oregonensis"	"Pungitius.pungitius"
## [603] "Pylodictis.olivaris"	"Rhinichthys.atratulus"
## [605] "Rhinichthys.cataractae"	"Rhinichthys.cobitis"
## [607] "Rhinichthys.falcatus"	"Rhinichthys.obtusius"

## [609] "Rhinichthys.osculus"	"Rhinichthys.umatilla"
## [611] "Richardsonius.balteatus"	"Richardsonius.egregius"
## [613] "Salmo.salar"	"Salmo.trutta"
## [615] "Salvelinus.confluentus"	"Salvelinus.fontinalis"
## [617] "Salvelinus.malma"	"Sander.canadensis"
## [619] "Sander.vitreus"	"Scaphirhynchus.platorynchus"
## [621] "Sciaenops.ocellatus"	"Semotilus.atromaculatus"
## [623] "Semotilus.corporalis"	"Semotilus.lumbee"
## [625] "Semotilus.thoreauianus"	"Strongylura.marina"
## [627] "Taricha.granulosa"	"Thaleichthys.pacificus"
## [629] "Thoburnia.atripinnis"	"Thoburnia.rhothoeca"
## [631] "Thymallus.arcticus"	"Tinca.tinca"
## [633] "Trinectes.maculatus"	"Umbra.limi"
## [635] "Umbra.pygmaea"	"Xyrauchen.texasus"

How many sites are there? > **Answer 3:** > There are 2753 fish sites.

How many species are there in the entire site-by-species matrix? > There are 636 species in the entire site-by-species matrix.

Any other interesting observations based on what you learned this week? > Within the dataset we can see that there are many fish sites as well as species within the matrix. This data can be overwhelming; however, based on information learned within the past two weeks, there are ways to organize the data efficiently. By creating a site-by-species matrix, you can go through the sites and began extropolating quantitative biodiversity measures, such as alpha diversity.

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 29th, 2025 at 12:00 PM (noon)**.