7. Worksheet: Diversity Synthesis

Lauren Albert; Z620: Quantitative Biodiversity, Indiana University

14 February, 2023

OVERVIEW

In this worksheet, you will conduct exercises that reinforce fundamental concepts of biodiversity. Specifically, you will construct a a site-by-species matrix by sampling confectionery taxa. With this primary data structure, you will then answer questions and generate figures using tools from previous weeks, along with wrangling techniques that we learned about in class.

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. Refer to previous handouts to help with developing of questions and writing of code.
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 `7.DiversitySynthesis_Worskheet.Rmd` and the PDF output of `Knitr` (`DiversitySynthesis_Worskheet.pdf`).

CONFECTIONARY EXERCISE GOALS

We will construct a site-by-species matrix using confectionery taxa (i.e, gummies). The instructors have created distinct **sources communities** that vary in the composition of gummy taxa with even and uneven communities. It might be fun to consider them as distinct geographical regions experiencing different environmental regimes, or different experimental units under different treatments. Each student will sample a source community and then use a taxonomic key to identify gummies and their abundances.

In the end, students will use the site-by-species matrix to:

1) explore their sampling efforts and their effects on species richness using **coverage** and **rarefaction** concept,

2) measure **alpha diversity** for each sub-sample collated from data with their peers from the same source community,

3) examine **beta diversity** between each source community using the data generated across each source community, and

4) use **data wrangling** tools they have learned during the class to accomplish the above goals.

SAMPLING PROTOCOL TO CONSTRUCT A SITE-BY-SPECIES MATRIX

1. Instructors will assign you to sample confectionery taxa from one of the two designated source community bucket (A and B).
2. After randomly sampling one unit (imagine as an equal biomass) from the source community, each student will count the total number of individuals (N), identify the taxa using the species key and quantify the abundance of each taxon.
3. Work with other students in your group to assemble data into a site-by-species matrix on the white board. One person needs to create a .csv or .txt file and share your group's site-by-species matrix with the class using GitHub. Make sure that you include a sample identifier (student name) and what community you sampled from.

GROUP BRAINSTORM

In smaller groups, take 15 minutes to brainstorm questions, code, statistical tests, and "fantasy figures" using the site-by-species matrix the class generated.

1. Using this data, explore how well your sampling effort was. You can use rarefaction and coverage tools you have learned earlier.
2. Investigate alpha diversity based on the methods you have learned in the rest of the handout and accompanying worksheet. For example, you can measure richness, Shannon diversity and Simpson index. You can also convert them to effective number of species using the Hill numbers concept.
3. Measure beta diversity using ordination and multivariate statistical methods. For example, you can create a PCoA plot, based on Bray-Curtis dissimilarity, of sites and communities using different shape and color codes. Use Permanova to test if there are differences between communities.

DATA ANALYSIS

1) Sampling coverage and rarefaction curves

Question 1: Using this data, explore how well your sampling effort was. Compare your sampling efforts with other groups. Do you think that your samples cover the actual diversity found in each source community? You can use rarefaction and coverage tools you have learned earlier.

Answer 1: Use the space below to generate a rarefaction curve/sample coverage based on the data we collected in class for each community. Make sure to annotate your code using # symbols so others (including instructors) understand what you have done and why you have done it.

```
#clearing environment, setting wd, loading packages
rm(list = ls())
getwd()
```

```
## [1] "/Users/laurenalbert/GitHub/QB2023_Albert/2.Worksheets/7.DiversitySynthesis"
```

```
library(vegan)
```

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

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

```
## This is vegan 2.6-4
```

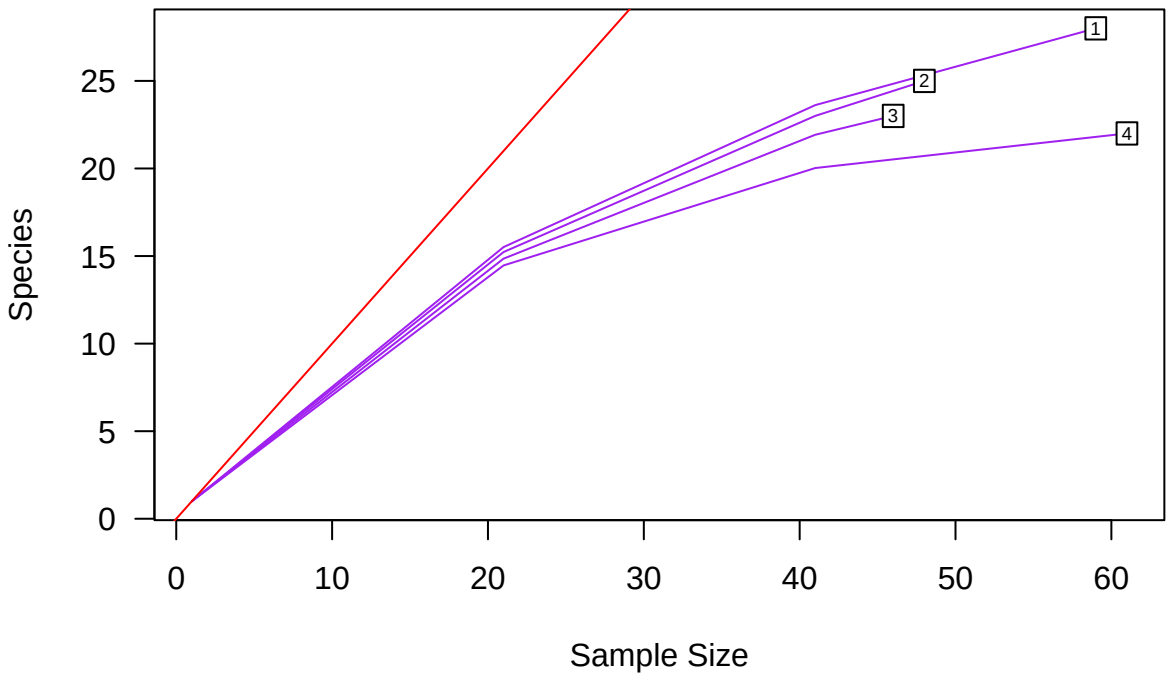
```
#import data
gummy <- read.csv("QB Data Wrangling Lab SP23.csv")
#remove NAs and subset out into site by species matrix
gummy <- na.omit(gummy)
gummy <- gummy[,3:32]

#need to change all columns from integer to numeric
gummy[1:30] = lapply(gummy[1:30], FUN = function(y){as.numeric(y)})

#subset by community A and community B
gummyA <- gummy[1:4,]
gummyB <- gummy[5:8,]

#write a function to calculate observed richness
S.obs <- function(x =""){
rowSums(x > 0) * 1
}

#calculate observed richness of each community
S.obs(gummyA)
```

```
##  1  2  3  4
## 28 25 23 22
```

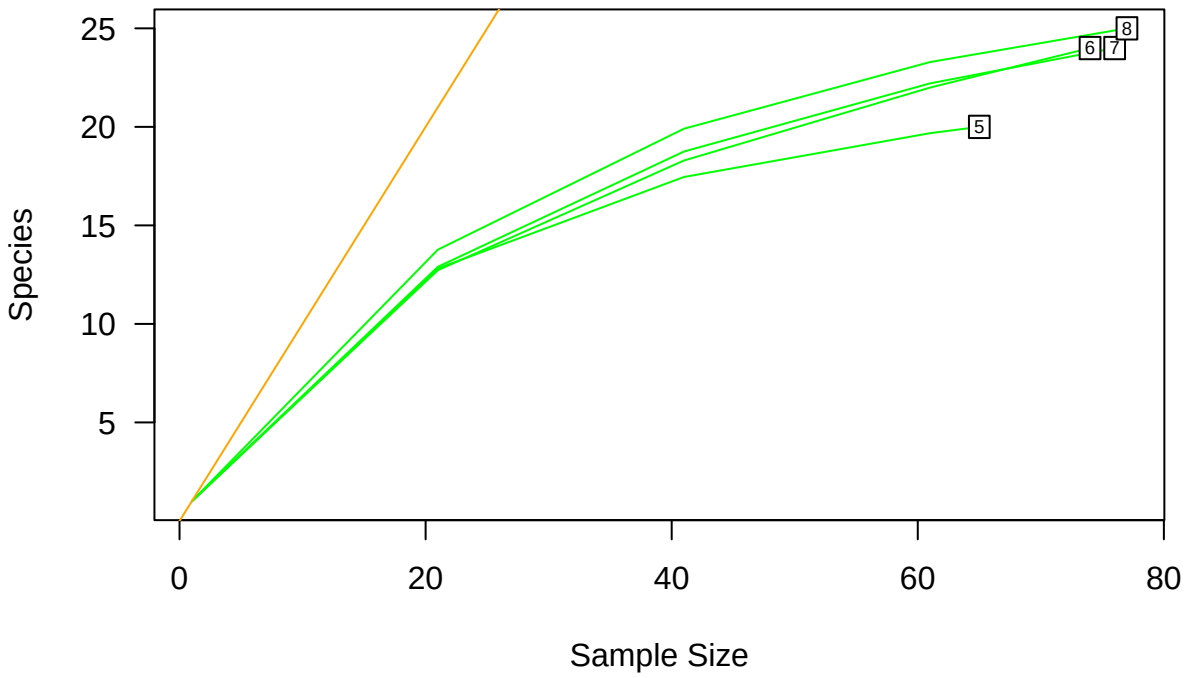
```
S.obs(gummyB)
```

```
##  5  6  7  8
## 20 24 24 25
```

```
#determine size of the smallest sample in each community
min.A <- min(rowSums(gummyA))
min.B <- min(rowSums(gummyB))

#use rarefy function to rarefy each community to this level
A.rarefy <- rarefy(x = gummyA, sample = min.A, se = TRUE)
B.rarefy <- rarefy(x = gummyB, sample = min.B, se = TRUE)

#plot rarefaction curve and add a 1:1 line
#Community A
rarecurveA <- rarecurve(x = gummyA, step = 20, col = "purple", cex = 0.6, las = 1)
abline(0, 1, col = 'red')
text(1500, 1500, "1:1", pos = 2, col = 'red')
```




```
#Community B
rarecurveB <- rarecurve(x = gummyB, step = 20, col = "green", cex = 0.6, las = 1)
abline(0, 1, col = 'orange')
text(1500, 1500, "1:1", pos = 2, col = 'orange')
```




```
#Calculate coverage for each community
#function for Good's Coverage
GoodCoverage <- function(x = ""){
  1 - (rowSums(x == 1)/rowSums(x))
}

GoodCoverage(gummyA)
```

```
##         1         2         3         4
## 0.8135593 0.7291667 0.8043478 0.9508197
```

```
GoodCoverage(gummyB)
```

```
##         5         6         7         8
## 0.9230769 0.8513514 0.8947368 0.9090909
```

2) Alpha diversity

Question 2: Compare alpha diversity measures within sites and among communities. You can calculate and plot richness, Shannon diversity, and Simpson index. You can also convert these indices to effective number of species using the Hill numbers concept by generating a diversity profile, which will make comparisons easier across sites.

What is the variation among the samples in your group and between the communities of other groups for the alpha diversity indices? Generate a hypothesis around the diversity metrics you chose and test your hypothesis. Interpret your findings.

Answer 2a - Analysis: Use the space below for code that is being used to analyze your data and test your hypotheses on your chosen alpha diversity tool. Make sure to annotate your code using # symbols so others (including instructors) understand what you have done and why you have done it.

```
#writing function to calculate Chao 1
chao1 <- function(x = ""){
  S.obs(x) + (sum(x == 1)^2) / (2 * sum(x == 2))

}

#using chao1 because max abundance of a species is 16, ACE places threshold at 10
#calculate richness within each site in community A
communityAchao <- chao1(gummyA)

chao1(gummyA[1,])
```

```
##    1
## 33.5
```

```
chao1(gummyA[2,])
```

```
##    2
## 41.9
```

```
chao1(gummyA[3,])
```

```
##     3
## 27.05
```

```
chao1(gummyA[4,])
```

```
##       4
## 22.5625
```

```
#calculate richness within each site in Community B
communityBchao <- chao1(gummyB)
chao1(gummyB[1,])
```

```
##      5
## 23.125
```

```
chao1(gummyB[2,])
```

```
##     6
## 54.25
```

```
chao1(gummyB[3,])
```

```
##    7
## 30.4
```

```
chao1(gummyB[4,])
```

```
##    8
## 29.9
```

```
#writing function to calculate Shannon diversity
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)
}

#calculating Shannon diversity at the four sites within community A

ShanH(gummyA[1,])
```

```
## [1] 3.166063
```

```
ShanH(gummyA[2,])
```

```
## [1] 3.053443
```

```
ShanH(gummyA[3,])
```

```
## [1] 2.975329
```

```
ShanH(gummyA[4,])
```

```
## [1] 2.998721
```

```
ShanH(gummyB[1,])
```

```
## [1] 2.787795
```

```
ShanH(gummyB[2,])
```

```
## [1] 2.834947
```

```
ShanH(gummyB[3,])
```

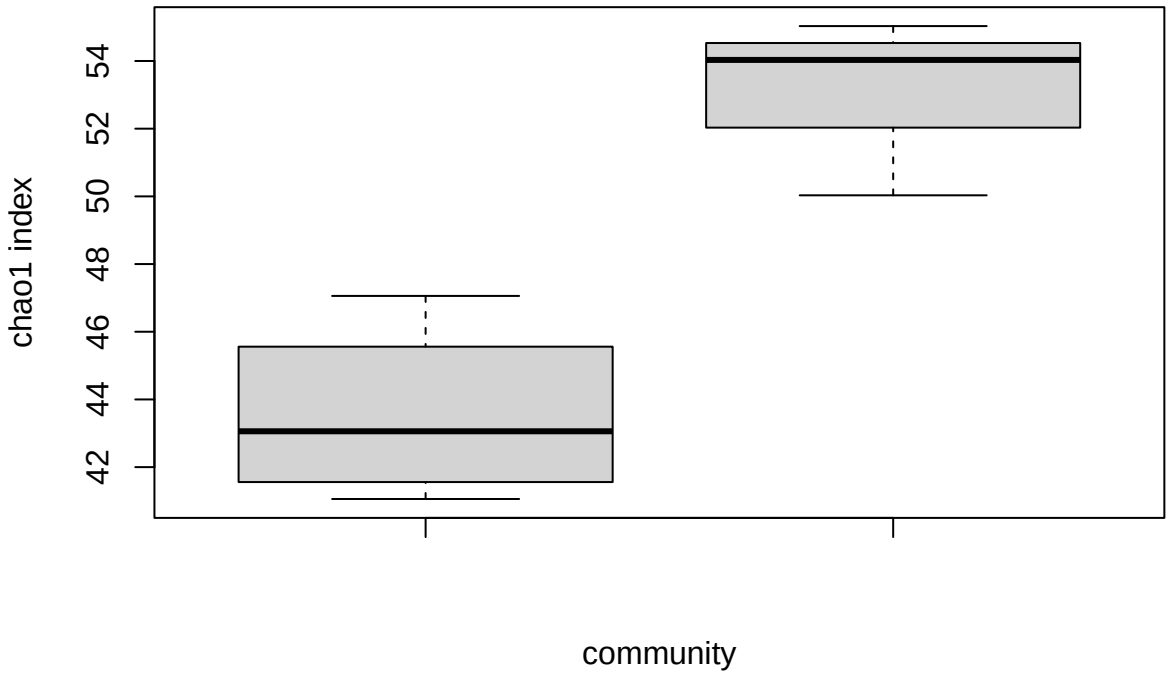
```
## [1] 2.836906
```

```
ShanH(gummyB[4,])
```

```
## [1] 2.978193
```

Answer 2b - Plot: With your analysis, create one (and only one, although it can have multiple panels) *publication-quality* figure.

```
boxplot(communityAchao, communityBchao, xlab = "community", ylab = "chao1 index")
```



Answer 2c - Interpret results: Write an informative yet succinct (~5 sentences) caption that creates a "stand-alone" figure. Take a peek at figures and figure captions in a paper published in your favorite journal for inspiration. >A boxplot describing the Chao1 index of community A and community B. Within site chao1 indexes (not shown) demonstrate that sites in community B had more species with high abundance compared to community A, which had fewer species with high abundance (more singletons and doubletons). Overall, the metric of alpha diversity suggest community B had greater diversity than community A when estimating the number of singleton and doubletons in the community.

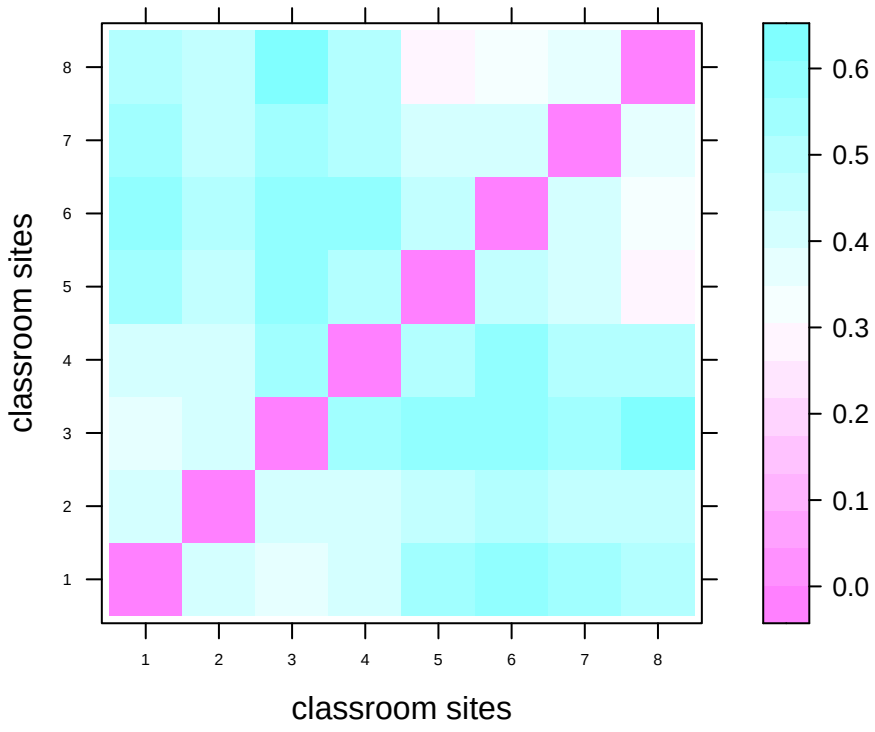
3) Beta diversity

Question 3: Measure beta diversity using ordination and multivariate statistics methods. You can create a PCoA plot, based on Bray-Curtis dissimilarity, of sites and communities using different shape and color codes. Then, you can use a Permanova to test if there are differences between communities. Generate a hypothesis around your chosen analysis and test your hypothesis. Interpret your findings.

Can you detect compositional differences between each source community sampled?

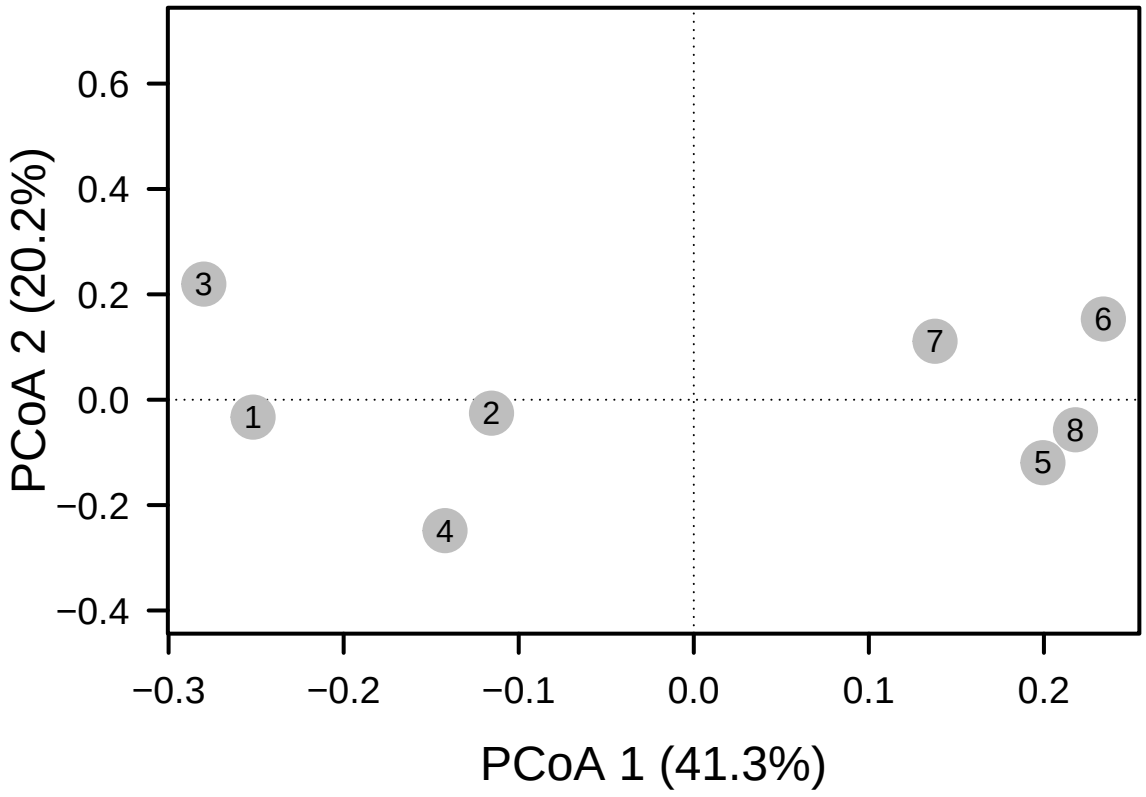
Answer 3a - Analysis: Use the space below for code that is being used to analyze your data and test your hypotheses on your chosen beta diversity tool. Make sure to annotate your code using # symbols so others (including instructors) understand what you have done and why you have done it.

```
library(lattice)

#calculating Bray-Curtis dissimilarity
gummy.db <- vegdist(gummy, method = "bray")

levelplot(as.matrix(gummy.db), aspect = "iso", xlab = "classroom sites", ylab = "classroom sites", scal
```




```
#Creating PCoA plot
#first perform a Principal Coordinates Analysis to visualize beta-diversity
gummy.pcoa <- cmdscale(gummy.db, eig = TRUE, k = 3)

#then calculate the variation explained by the first three axes in your ordination
explainvar1 <- round(gummy.pcoa$eig[1] / sum(gummy.pcoa$eig), 3) * 100
explainvar2 <- round(gummy.pcoa$eig[2] / sum(gummy.pcoa$eig), 3) * 100
explainvar3 <- round(gummy.pcoa$eig[3] / sum(gummy.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)

par(mar = c(5,5,1,2) + 0.1)

plot(gummy.pcoa$point[ ,1], gummy.pcoa$points[ ,2], ylim = c(-0.4, 0.7),
     xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
```

```
     ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5,
     cex.axis = 1.2, axes = FALSE
     )

axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)

points(gummy.pcoa$points[,1], gummy.pcoa$points[,2],
       pch = 19, cex = 3, bg = "gray", col = "gray")
text(gummy.pcoa$points[ ,1], gummy.pcoa$points[ ,2],
     labels = row.names(gummy.pcoa$points))
```




```
#creating factor of communities to test with PERMANOVA
gummycommunities<- c(rep("A",4), rep("B",4))
#permanova
#adonis(gummy ~ gummycommunities,  method = "bray", permutations = 999)

#want to see if there are indicator species of each community
library(indicspecies)
#species-site group associations
indval <- multipatt(gummy, cluster = gummycommunities, func = "IndVal.g", control = how(nperm = 999))
summary(indval)
```

```
##
##  Multilevel pattern analysis
##  ---------------------------
##
##  Association function: IndVal.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 30
##  Selected number of species: 2
##  Number of species associated to 1 group: 2
##
##  List of species associated to each combination:
##
##  Group A  #sps.  1
##       stat p.value
## SP28 0.913   0.041 *
##
##  Group B  #sps.  1
##      stat p.value
## SP1 0.963   0.041 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
gummy.rel <- decostand(gummy, method = "total")
phi <- multipatt(gummy.rel, cluster = gummycommunities, func = "r.g", control = how(nperm = 999))
summary(phi)
```

```
##
##  Multilevel pattern analysis
##  ---------------------------
##
##  Association function: r.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 30
##  Selected number of species: 4
##  Number of species associated to 1 group: 4
##
##  List of species associated to each combination:
##
##  Group A  #sps.  3
##       stat p.value
## SP28 0.888   0.034 *
## SP27 0.832   0.034 *
## SP24 0.778   0.034 *
##
##  Group B  #sps.  1
##      stat p.value
## SP1 0.926   0.034 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer 3b - Plot: With your analysis, create one (and only one, although it can have multiple panels) *publication-quality* figure.

Answer 3c - Interpret results: Write an informative yet succinct (~5 sentences) caption that creates a "stand-alone" figure. Take a peek at figures and figure captions in a paper published in your favorite journal for inspiration. >I would like to try and make a cluster plot to demonstrate that community A and community B had few species that only appeared in each of the respective communities. I ran the species-site association to confirm that there were species that were indicative of each community, as the PCoA plot was little informative, creating clusters of the four sites within each community.

SUBMITTING YOUR ASSIGNMENT

Use Knitr to create a PDF of your completed 7.DiversitySynthesis_Worksheet.Rmd document, push it to GitHub, and create a pull request. Please make sure your updated repo includes both the pdf and RMarkdown files.

Unless otherwise noted, this assignment is due on **Wednesday, February 15th, 2023 at 12:00 PM (noon)**.