

7. Worksheet: Diversity Synthesis

Madison Brown; Z620: Quantitative Biodiversity, Indiana University

18 February, 2025

OVERVIEW

In this worksheet, you will conduct exercises that reinforce fundamental concepts of biodiversity. First, you will construct a site-by-species matrix by sampling confectionery taxa from a source community. Second, you will make a preference-profile matrix, reflecting each student's favorite 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`).

QUANTITATIVE CONFECTIONOLOGY

We will construct a site-by-species matrix using confectionery taxa (i.e., jelly beans). The instructors have created a **source community** with known abundance (N) and richness (S). Like a real biological community, the species abundances are unevenly distributed such that a few jelly bean types are common while most are rare. Each student will sample the source community and bin their jelly beans into operational taxonomic units (OTUs).

SAMPLING PROTOCOL: SITE-BY-SPECIES MATRIX

1. From the well-mixed source community, each student should take one Dixie Cup full of individuals.
2. At your desk, sort the jelly beans into different types (i.e., OTUs), and quantify the abundance of each OTU.
3. Working with other students, merge data into a site-by-species matrix with dimensions equal to the number of students (rows) and taxa (columns)
4. Create a worksheet (e.g., Google sheet) and share the site-by-species matrix with the class.

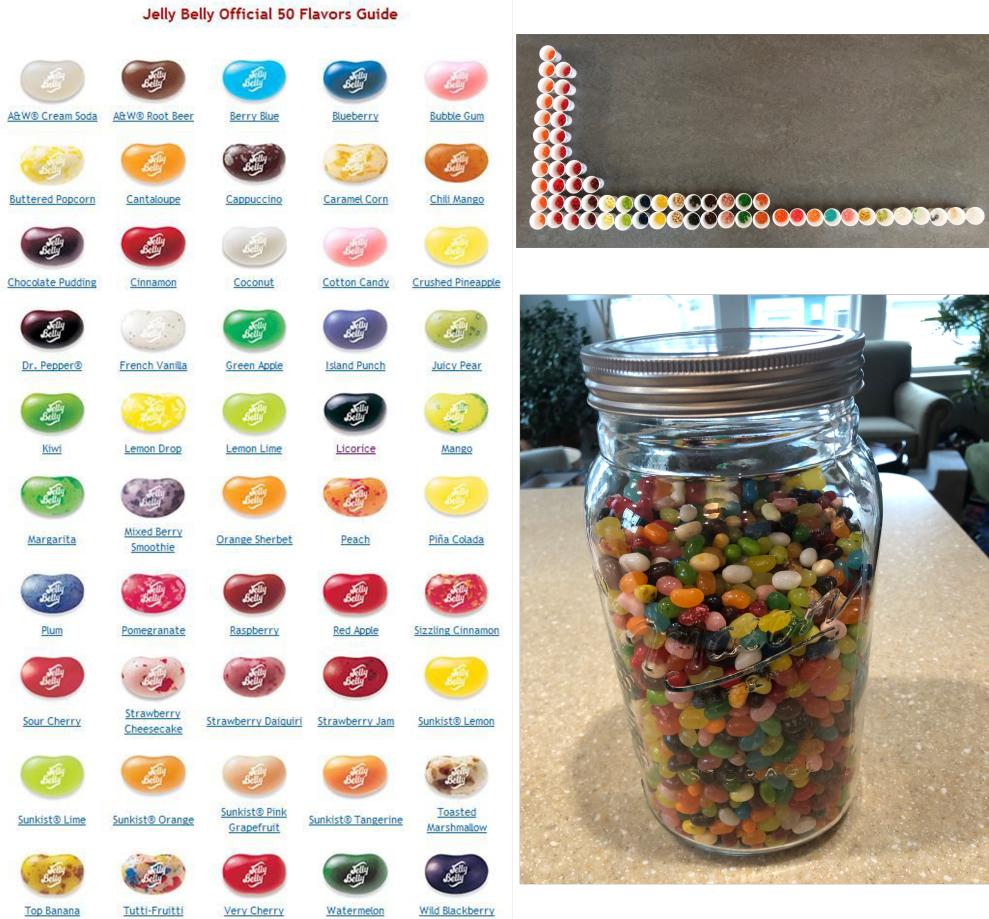


Figure 1: **Left:** taxonomic key, **Top right:** rank abundance distribution, **Bottom right:** source community

SAMPLING PROTOCOL: PREFERENCE-PROFILE MATRIX

1. With your individual sample only, each student should choose their top 5-10 preferred taxa based on flavor, color, sheen, etc.
2. Working with other students, merge data into preference-profile incidence matrix where 1 = preferred and 0 = non-preferred taxa.
3. Create a worksheet (e.g., Google sheet) and share the preference-profile matrix with the class.

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 Week5-Confection/ 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/Week5-Confection"
setwd("/cloud/project/QB2025_Brown/Week5-Confection")
library(vegan)

## Loading required package: permute
## Loading required package: lattice
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4    v readr     2.1.5
## vforcats   1.0.0    v stringr   1.5.1
## v ggplot2   3.5.1    v tibble    3.2.1
## v lubridate 1.9.4    v tidyrr    1.3.1
## v purrr    1.0.4

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)
library(dplyr)
library(broom)
```

DATA ANALYSIS

Question 1: In the space below, generate a rarefaction plot for all samples of the source community. Based on these results, discuss how individual vs. collective sampling efforts capture the diversity of the source community.

```
library(readr)
Jellybean_SbyS <- read_csv("/cloud/project/QB2025_Brown/Week5-Confection/Jelly belly data! - SbyS.csv")

## Rows: 12 Columns: 52
## -- Column specification -----
## Delimiter: ","
## chr (1): Name
## dbl (51): A & W Cream Soda, A & W Root Beer, Berry Blue, Blueberry, Bubble G...
```

```

## 
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

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

s.obs(Jellybean_SbyS)

## [1] 21 24 30 35 37 23 26 17 24 24 30 27
Jellybean_obs <- s.obs(Jellybean_SbyS)

str(Jellybean_SbyS)

## #> #> spc_tbl_ [12 x 52] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## #> $ Name : chr [1:12] "Bryan Guevara" "Madison Brown" "Jaeyoung" "Jocelyn" ...
## #> $ A & W Cream Soda : num [1:12] 1 2 4 1 1 3 2 1 0 0 ...
## #> $ A & W Root Beer : num [1:12] 0 0 2 0 1 3 0 0 1 0 ...
## #> $ Berry Blue : num [1:12] 4 4 2 2 1 1 0 2 2 1 ...
## #> $ Blueberry : num [1:12] 10 4 5 3 5 4 2 2 2 4 ...
## #> $ Bubble Gum : num [1:12] 4 2 1 3 4 1 0 1 2 1 ...
## #> $ Buttered Popcorn : num [1:12] 0 1 1 0 0 0 0 0 1 0 ...
## #> $ Cantaloupe : num [1:12] 0 0 0 1 1 0 2 0 0 0 ...
## #> $ Cappuccino : num [1:12] 0 0 0 0 0 0 0 0 0 0 ...
## #> $ Caramel Corn : num [1:12] 1 0 1 3 2 0 2 0 0 1 ...
## #> $ Chili Mango : num [1:12] 0 0 0 0 0 0 0 0 0 0 ...
## #> $ Chocolate Pudding : num [1:12] 0 0 0 1 0 0 1 0 0 0 ...
## #> $ Cinnamon : num [1:12] 0 2 0 0 1 4 0 2 4 1 ...
## #> $ Coconut : num [1:12] 5 3 1 1 4 4 1 4 3 4 ...
## #> $ Cotton Candy : num [1:12] 1 1 2 1 1 1 2 0 3 0 ...
## #> $ Crushed Pineapple : num [1:12] 0 0 0 2 0 0 6 0 0 0 ...
## #> $ Dr. Pepper : num [1:12] 2 0 0 1 1 0 0 0 0 0 ...
## #> $ French Vanilla : num [1:12] 2 0 3 0 1 0 3 0 0 0 ...
## #> $ Green Apple : num [1:12] 1 2 1 1 1 3 2 0 2 1 ...
## #> $ Island Punch : num [1:12] 0 0 0 0 2 0 0 0 0 0 ...
## #> $ Juicy Pear : num [1:12] 4 0 5 6 4 6 0 1 0 2 ...
## #> $ Kiwi : num [1:12] 4 5 5 2 1 2 2 1 2 3 ...
## #> $ Lemon Drop : num [1:12] 6 2 0 1 2 0 0 0 10 0 ...
## #> $ Lemon Lime : num [1:12] 0 0 0 1 2 0 2 0 0 1 ...
## #> $ Licorice : num [1:12] 0 0 0 1 0 0 2 0 1 0 ...
## #> $ Mango : num [1:12] 0 0 0 0 0 0 0 5 1 0 ...
## #> $ Margarita : num [1:12] 0 7 1 0 2 1 2 0 4 0 ...
## #> $ Mixed Berry Smoothie : num [1:12] 0 0 0 1 2 0 0 0 0 0 ...
## #> $ Orange Sherbet : num [1:12] 0 0 0 1 1 0 2 6 2 0 ...
## #> $ Peach : num [1:12] 1 1 5 1 2 3 0 4 6 1 ...
## #> $ Piña Colada : num [1:12] 0 2 0 2 1 2 0 2 1 3 ...
## #> $ Plum : num [1:12] 0 0 4 1 0 0 0 0 1 1 ...
## #> $ Pomegranate : num [1:12] 1 1 0 0 3 0 1 0 0 0 ...
## #> $ Raspberry : num [1:12] 11 7 10 5 0 11 5 8 0 1 ...
## #> $ Red Apple : num [1:12] 1 0 1 3 3 0 3 0 0 0 ...
## #> $ Sizzling Cinnamon : num [1:12] 3 0 1 1 0 0 0 0 0 0 ...
## #> $ Sour Cherry : num [1:12] 0 3 0 1 1 0 2 2 12 0 ...
## #> $ Strawberry Cheesecake : num [1:12] 0 0 4 1 0 0 2 0 2 0 ...

```

```

## $ Strawberry Daiquiri      : num [1:12] 0 0 0 0 0 0 0 0 0 2 ...
## $ Strawberry Jam           : num [1:12] 0 0 0 0 1 0 1 0 5 10 ...
## $ Sunkist Lemon            : num [1:12] 0 2 4 0 3 9 2 0 0 3 ...
## $ Sunkist Lime              : num [1:12] 0 0 0 0 0 2 0 0 0 0 ...
## $ Sunkist Orange             : num [1:12] 0 4 8 0 1 6 0 3 0 3 ...
## $ Sunkist Pink Grapefruit: num [1:12] 0 0 0 3 0 0 0 0 0 1 ...
## $ Sunkist Tangerine          : num [1:12] 0 4 1 1 4 0 0 0 1 1 ...
## $ Toasted Marshmallow        : num [1:12] 0 2 1 0 2 1 2 0 1 1 ...
## $ Top Banana                 : num [1:12] 0 1 1 2 1 1 0 0 0 1 ...
## $ Tutti-Fruitti              : num [1:12] 0 0 1 1 0 0 0 0 0 1 ...
## $ Very Cherry                : num [1:12] 1 0 1 0 2 0 0 0 0 0 ...
## $ Watermelon                  : num [1:12] 0 0 0 3 2 3 1 0 0 0 ...
## $ Wild blackberry             : num [1:12] 2 1 3 1 1 3 1 1 0 0 ...
## $ Blue Rasberry               : num [1:12] 0 0 1 3 1 0 0 0 0 0 ...
## - attr(*, "spec")=
##   .. cols(
##     .. Name = col_character(),
##     .. `A & W Cream Soda` = col_double(),
##     .. `A & W Root Beer` = col_double(),
##     .. `Berry Blue` = col_double(),
##     .. Blueberry = col_double(),
##     .. `Bubble Gum` = col_double(),
##     .. `Buttered Popcorn` = col_double(),
##     .. Cantaloupe = col_double(),
##     .. Cappuccino = col_double(),
##     .. `Caramel Corn` = col_double(),
##     .. `Chili Mango` = col_double(),
##     .. `Chocolate Pudding` = col_double(),
##     .. Cinnamon = col_double(),
##     .. Coconut = col_double(),
##     .. `Cotton Candy` = col_double(),
##     .. `Crushed Pineapple` = col_double(),
##     .. `Dr. Pepper` = col_double(),
##     .. `French Vanilla` = col_double(),
##     .. `Green Apple` = col_double(),
##     .. `Island Punch` = col_double(),
##     .. `Juicy Pear` = col_double(),
##     .. Kiwi = col_double(),
##     .. `Lemon Drop` = col_double(),
##     .. `Lemon Lime` = col_double(),
##     .. Licorice = col_double(),
##     .. Mango = col_double(),
##     .. Margarita = col_double(),
##     .. `Mixed Berry Smoothie` = col_double(),
##     .. `Orange Sherbet` = col_double(),
##     .. Peach = col_double(),
##     .. `Piña Colada` = col_double(),
##     .. Plum = col_double(),
##     .. Pomegranate = col_double(),
##     .. Raspberry = col_double(),
##     .. `Red Apple` = col_double(),
##     .. `Sizzling Cinnamon` = col_double(),
##     .. `Sour Cherry` = col_double(),
##     .. `Strawberry Cheesecake` = col_double(),

```

```

## .. `Strawberry Daiquiri` = col_double(),
## .. `Strawberry Jam` = col_double(),
## .. `Sunkist Lemon` = col_double(),
## .. `Sunkist Lime` = col_double(),
## .. `Sunkist Orange` = col_double(),
## .. `Sunkist Pink Grapefruit` = col_double(),
## .. `Sunkist Tangerine` = col_double(),
## .. `Toasted Marshmallow` = col_double(),
## .. `Top Banana` = col_double(),
## .. `Tutti-Fruitti` = col_double(),
## .. `Very Cherry` = col_double(),
## .. Watermelon = col_double(),
## .. `Wild blackberry` = col_double(),
## .. `Blue Raspberry` = col_double()
## ...
## - attr(*, "problems")=<externalptr>
Jellybean_SbyS <- as.data.frame(lapply(Jellybean_SbyS, as.numeric))

## Warning in lapply(Jellybean_SbyS, as.numeric): NAs introduced by coercion
Jellybean_SbyS[is.na(Jellybean_SbyS)] <- 0

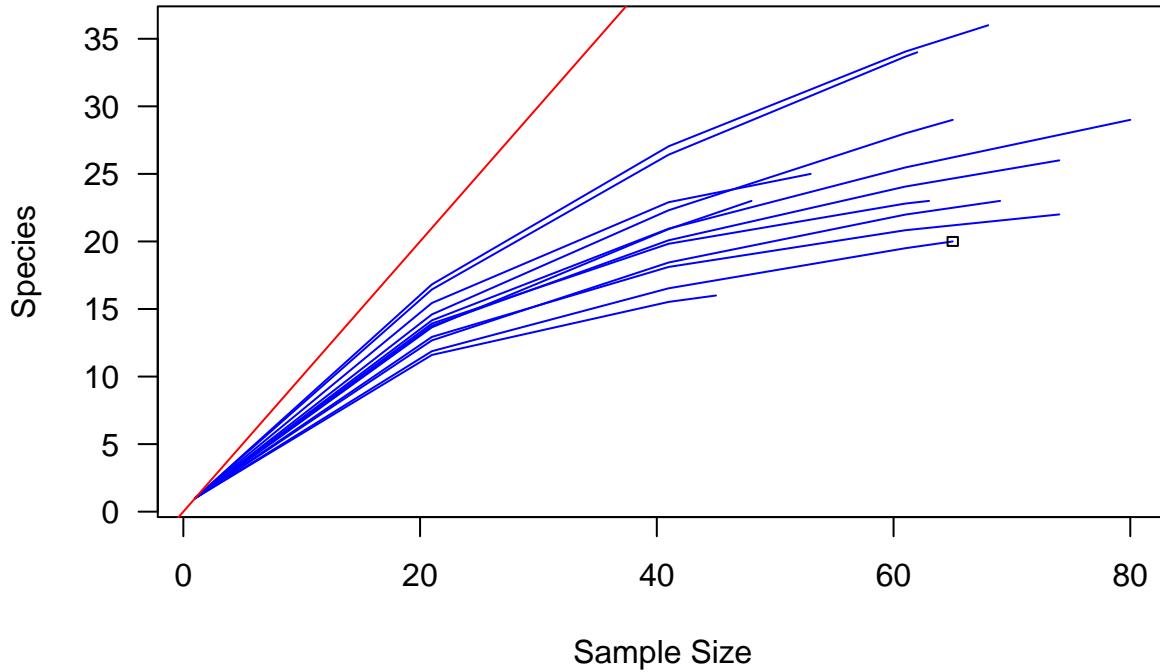
Min.N <- min(rowSums(Jellybean_SbyS))

s.rarefy <- rarefy(x = Jellybean_SbyS, sample = Min.N, se = TRUE)

rarecurve(x = Jellybean_SbyS, step = 20, col = "blue", cex = 0.6, las = 1)

abline(0, 1, col = 'red')
text(1500, 1500, "1:1", pos = 2, col = 'red')

```



Answer 1: Individual sampling efforts are not the most ideal because it is highly unlikely that you will be able to sample every species within an area and if that is your goal, it is very hard to

do. Additionally, if you analyze everyone's samples individually, you are not able to determine large trends or data consistency. If you utilize community sampling efforts, you can determine patterns and organize and graph all of the collected data to better visualize it. Community sampling efforts can also be more ideal, but it is likely that someone will sample something that you or someone else may have missed.

Question 2: Starting with the site-by-species matrix, visualize beta diversity. In the code chunk below, conduct principal coordinates analyses (PCoA) using both an abundance- and incidence-based resemblance matrix. Plot the sample scores in species space using different colors, symbols, or labels. Which "species" are contributing the patterns in the ordinations? How does the choice of resemblance matrix affect your interpretation?

```
# Abundance Based

bean.db <- vegdist(Jellybean_SbyS, method = "bray", upper = TRUE, diag = TRUE)

bean.pcoa <- cmdscale(bean.db, eig = TRUE, k = 3)

Explainvar1 <- round(bean.pcoa$eig[1] / sum(bean.pcoa$eig), 3) * 100
Explainvar2 <- round(bean.pcoa$eig[2] / sum(bean.pcoa$eig), 3) * 100
Explainvar3 <- round(bean.pcoa$eig[3] / sum(bean.pcoa$eig), 3) * 100
Sum.eig <- sum(Explainvar1, Explainvar2, Explainvar3)

head(row.names(bean.pcoa$points))

## NULL

rownames(bean.pcoa$points) <- 1:nrow(bean.pcoa$points)

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

plot(bean.pcoa$points[,1], bean.pcoa$points[,2],
     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,
     xlim = range(bean.pcoa$points[,1]) * 1.1,
     ylim = range(bean.pcoa$points[,2]) * 1.1)

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(bean.pcoa$points[,1], bean.pcoa$points[,2],
       pch = 19, cex = 2.5, bg = "gray", col = "gray")
text(bean.pcoa$points[,1], bean.pcoa$points[,2],
     labels = row.names(bean.pcoa$points), cex = 1.0, pos = NULL)

JellyREL <- Jellybean_SbyS
for(i in 1:nrow(Jellybean_SbyS)){
  JellyREL[i, ] = Jellybean_SbyS[i, ] / sum(Jellybean_SbyS[i, ])
}

library(vegan)
```

```

`add.spec.scores.class` <-
  function(ordi,comm,method="cor.scores",multi=1,Rscale=F,scaling="1") {
    ordiscores <- scores(ordi,display="sites")
    n <- ncol(comm)
    p <- ncol(ordiscores)
    specscores <- array(NA,dim=c(n,p))
    rownames(specscores) <- colnames(comm)
    colnames(specscores) <- colnames(ordiscores)
    if (method == "cor.scores") {
      for (i in 1:n) {
        for (j in 1:p) {specscores[i,j] <- cor(comm[,i],ordiscores[,j],method="pearson")}
      }
    }
    if (method == "wa.scores") {specscores <- wascores(ordiscores,comm)}
    if (method == "pcoa.scores") {
      rownames(ordiscores) <- rownames(comm)
      eigenv <- ordi$eig
      accounted <- sum(eigenv)
      tot <- 2*(accounted/ordi$GOF[2])-(accounted/ordi$GOF[1])
      eigen.var <- eigenv/(nrow(comm)-1)
      neg <- length(eigenv[eigenv<0])
      pos <- length(eigenv[eigenv>0])
      tot <- tot/(nrow(comm)-1)
      eigen.percen <- 100*eigen.var/tot
      eigen.cumpercen <- cumsum(eigen.percen)
      constant <- ((nrow(comm)-1)*tot)^0.25
      ordiscores <- ordiscores * (nrow(comm)-1)^-0.5 * tot^-0.5 * constant
      p1 <- min(p, pos)
      for (i in 1:n) {
        for (j in 1:p1) {
          specscores[i,j] <- cor(comm[,i],ordiscores[,j])*sd(comm[,i])/sd(ordiscores[,j])
          if(is.na(specscores[i,j])) {specscores[i,j]<-0}
        }
      }
      if (Rscale==T && scaling=="2") {
        percen <- eigen.var/tot
        percen <- percen^0.5
        ordiscores <- sweep(ordiscores,2,percen,"/")
        specscores <- sweep(specscores,2,percen,"*")
      }
      if (Rscale==F) {
        specscores <- specscores / constant
        ordiscores <- ordi$points
      }
      ordi$points <- ordiscores
      ordi$eig <- eigen.var
      ordi$eig.percen <- eigen.percen
      ordi$eig.cumpercen <- eigen.cumpercen
      ordi$eigen.total <- tot
      ordi$R.constant <- constant
      ordi$Rscale <- Rscale
      ordi$scaling <- scaling
    }
  }

```



```

## Cinnamon          0.97495 -0.22243 0.5550  0.033 *
## Coconut           0.71298 -0.70119 0.3318  0.163
## Cotton.Candy     -0.10558  0.99441 0.3877  0.116
## Crushed.Pineapple -0.51309  0.85833 0.6290  0.012 *
## Dr..Pepper        -0.97379 -0.22745 0.4077  0.102
## French.Vanilla    -0.94613  0.32378 0.5200  0.038 *
## Green.Apple       -0.87564  0.48296 0.1057  0.632
## Island.Punch      -0.44836  0.89385 0.0764  0.926
## Juicy.Pear         -0.65475 -0.75585 0.3442  0.157
## Kiwi               0.53679 -0.84372 0.0327  0.845
## Lemon.Drop         0.79666  0.60443 0.2860  0.226
## Lemon.Lime         -0.29156  0.95655 0.5125  0.040 *
## Licorice            -0.42274  0.90625 0.6862  0.008 **
## Mango              0.86334 -0.50462 0.2175  0.342
## Margarita          0.48440  0.87485 0.1717  0.410
## Mixed.Berry.SMOOTHIE -0.62910  0.77733 0.1936  0.403
## Orange.Sherbet     0.98649  0.16380 0.1848  0.421
## Peach              0.81267 -0.58272 0.2392  0.302
## Piña.Colada        0.92603 -0.37745 0.3147  0.179
## Plum                0.46063 -0.88759 0.0233  0.915
## Pomegranate        -0.38122  0.92448 0.1989  0.354
## Raspberry           -0.55868 -0.82938 0.5298  0.041 *
## Red.Apple           -0.82207  0.56939 0.8138  0.001 ***
## Sizzling.Cinnamon   -0.82505 -0.56506 0.3291  0.174
## Sour.Cherry         0.78012  0.62563 0.7361  0.004 **
## Strawberry.Cheesecake -0.85184  0.52380 0.1708  0.423
## Strawberry.Daiquiri 0.65912 -0.75204 0.0653  0.819
## Strawberry.Jam      0.98391  0.17866 0.2272  0.279
## Sunkist.Lemon        0.29528 -0.95541 0.1041  0.582
## Sunkist.Lime          -0.76039 -0.64947 0.1947  0.297
## Sunkist.Orange        0.47286 -0.88114 0.5270  0.036 *
## Sunkist.Pink.Grapefruit -0.98730  0.15887 0.1319  0.523
## Sunkist.Tangerine     0.20534  0.97869 0.0362  0.853
## Toasted.Marsmallow    -0.01500  0.99989 0.2469  0.281
## Top.Banana            0.99314  0.11693 0.0365  0.840
## Tutti.Fruitti         -0.53303 -0.84610 0.0162  0.917
## Very.Cherry           -0.99388  0.11042 0.0229  0.897
## Watermelon            -0.82304  0.56799 0.1632  0.446
## Wild.blackberry       -0.40508 -0.91428 0.2208  0.315
## Blue.Rasberry         -0.99896  0.04563 0.2574  0.279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 999
# Incidence-Based

bean.dj <- vegdist(Jellybean_SbyS, method = "jaccard", binary = TRUE)

beaninc.pcoa <- cmdscale(bean.dj, eig = TRUE, k = 3)

Explainvar1inc <- round(beaninc.pcoa$eig[1] / sum(beaninc.pcoa$eig), 3) * 100
Explainvar2inc <- round(beaninc.pcoa$eig[2] / sum(beaninc.pcoa$eig), 3) * 100
Explainvar3inc <- round(beaninc.pcoa$eig[3] / sum(beaninc.pcoa$eig), 3) * 100

```

```

Suminc.eig <- sum(Explainvar1inc, Explainvar2inc, Explainvar3inc)

head(row.names(beaninc.pcoa$points))

## NULL

rownames(beaninc.pcoa$points) <- 1:nrow(beaninc.pcoa$points)

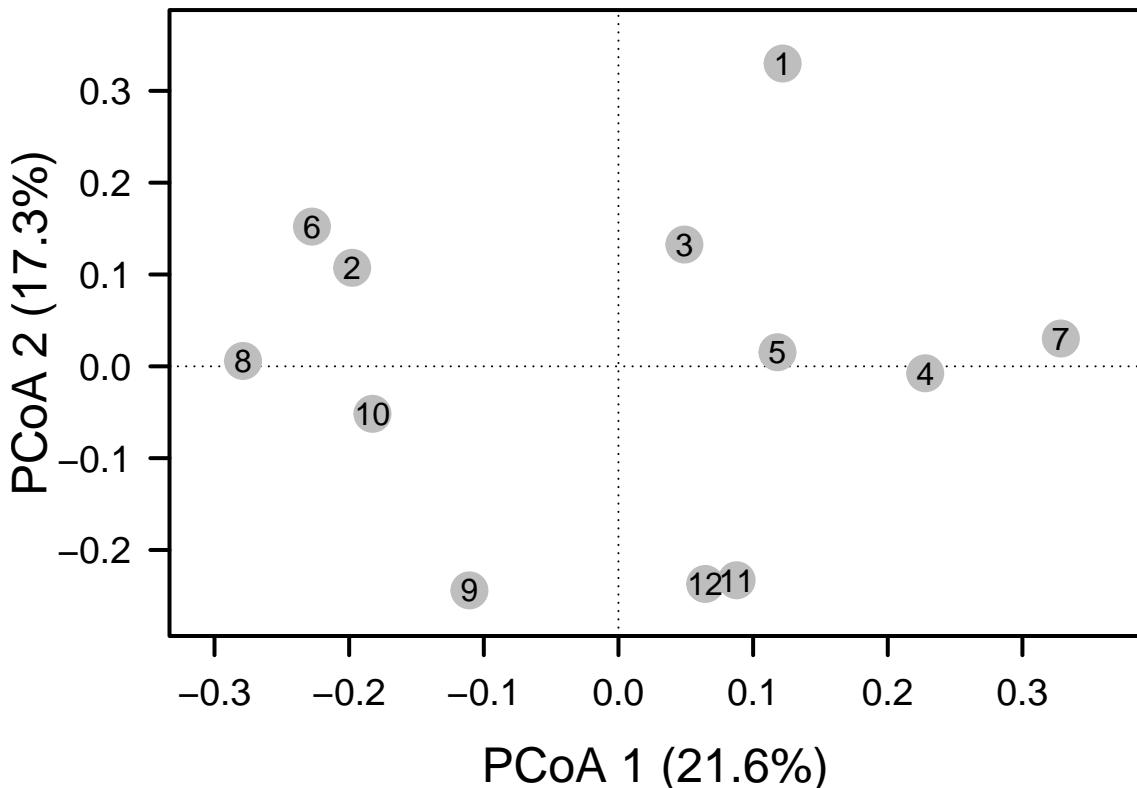
par(mar = c(5,5,1,2) + 0.1)

plot(beaninc.pcoa$points[,1], beaninc.pcoa$points[,2],
     xlab = paste("PCoA 1 (", Explainvar1inc, "%)", sep = ""),
     ylab = paste("PCoA 2 (", Explainvar2inc, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5,
     cex.axis = 1.2, axes = FALSE,
     xlim = range(beaninc.pcoa$points[,1]) * 1.1,
     ylim = range(beaninc.pcoa$points[,2]) * 1.1)

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(beaninc.pcoa$points[,1], beaninc.pcoa$points[,2],
       pch = 19, cex = 2.5, bg = "gray", col = "gray")
text(beaninc.pcoa$points[,1], beaninc.pcoa$points[,2],
     labels = row.names(beaninc.pcoa$points), cex = 1.0, pos = NULL)

```



Answer 2: Based on the results of the permutation test, the species (flavors in this case) that are contributing to the patterns in the ordination are: sunkist orange, sour cherry, red apple, raspberry, licorice, french vanilla, crushed pineapple, cinnamon, chocolate pudding, caramel corn, and cantaloupe. All of these flavors have statistically significant p values and are therefore identified as influential species. Whether you are using an incidence based or abundance based metric, it is important to take that into consideration with how you interpret your graphs. An incidence based metric is determining if something is present or absent. In this example, it would just give you an understanding of what flavors were present or not, but not how many were present. From an ecological stand point, incidence based metrics focus on species richness. On the other hand, abundance based metrics look at how much of an individual is present. In this example, it is looking at what flavors are present the most and/or least. In an ecological viewpoint, it would be like looking at species evenness.

Question 3 Using the preference-profile matrix, determine the most popular jelly bean in the class using a control structure (e.g., for loop, if statement, function, etc).

```
Jellybean_Pref <- read_csv("/cloud/project/QB2025_Brown/Week5-Confection/Jelly belly data! - Preference

## Rows: 13 Columns: 49
## -- Column specification -----
## Delimiter: ","
## chr (1): Name
## dbl (48): A & W Cream Soda, A & W Root Beer, Berry Blue, Blueberry, Bubble G...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

library(tidyverse)

fav_counts <- Jellybean_Pref %>%
  select(-Name) %>%
  summarise(across(everything(), sum)) %>%
  pivot_longer(cols = everything(), names_to = "Flavor", values_to = "TotalVotes") %>%
  arrange(desc(TotalVotes))

most_popular <- fav_counts %>%
  slice_max(TotalVotes, n = 1)

print(most_popular)

## # A tibble: 1 x 2
##   Flavor      TotalVotes
##   <chr>        <dbl>
## 1 Berry Blue     7
```

Answer 3: Berry Blue was the most popular jellybean flavor with a total of 7 votes.

Question 4 In the code chunk below, identify the student in QB who has a preference-profile that is most like yours. Quantitatively, how similar are you to your “jelly buddy”? Visualize the preference profiles of the class by creating a cluster dendrogram. Label each terminal node (a.k.a., tip or “leaf”) with the student’s name or initials. Make some observations about the preference-profiles of the class.

```
#Finding "jelly buddy"
```

```
Jellybean_Pref[is.na(Jellybean_Pref)] <- 0

Jellybean_Pref_numeric <- Jellybean_Pref[, -1]
```

```

pref.dj <- vegdist(Jellybean_Pref_numeric, method = "jaccard", binary = TRUE)

jaccard_similarity_matrix <- 1 - as.matrix(pref.dj)

my_index <- 3
jelly_buddy <- jaccard_similarity_matrix[my_index, ]

jelly_buddy[my_index] <- NA

most_similar_person <- which.max(jelly_buddy)

print(most_similar_person)

## 7
## 7
jaccard_value <- jaccard_similarity_matrix[my_index, most_similar_person]

print(jaccard_value)

## [1] 0.5714286
"Cluster dendrogram"

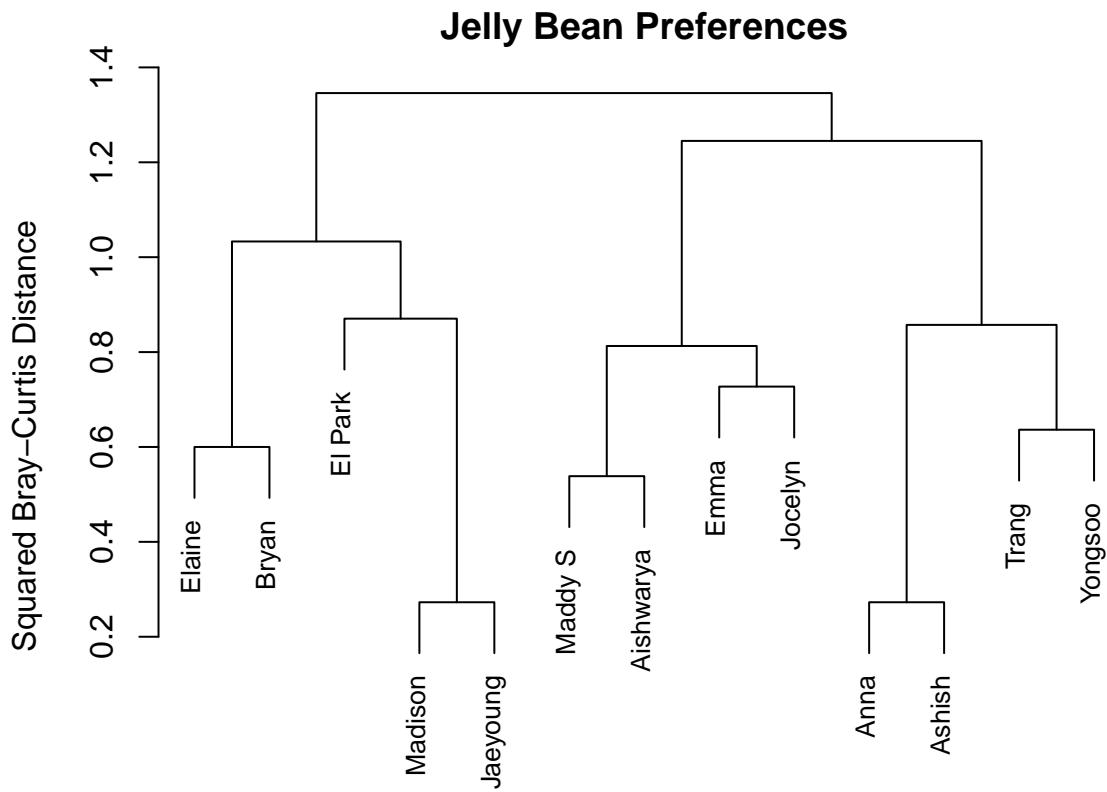
## [1] "Cluster dendrogram"
rownames(Jellybean_Pref_numeric) <- Jellybean_Pref>Name

## Warning: Setting row names on a tibble is deprecated.
beanpref.db <- vegdist(Jellybean_Pref_numeric, method = "bray", upper = TRUE, diag = TRUE)

pref.ward <- hclust(beanpref.db, method = "ward.D2")

par(mar = c(1, 5, 2, 2) + 0.1)
plot(pref.ward, main = "Jelly Bean Preferences",
     ylab = "Squared Bray-Curtis Distance", labels = rownames(Jellybean_Pref_numeric), cex = 0.8)

```



Answer 4: Student 7 has the most similar preference profile as myself. Looking back at the data set, person 7 corresponds to Jaeyoung. We also have a Jaccard similarity score of ~0.57. This means that Jaeyoung and I share approximately 57% of unique preferences. There are many students who have similar preferences. For example, Jaeyoung and myself compared to Anna and Ashish have very similar dissimilarity values; which are very low and means we have similar preferences. However, given how spread apart our two groups are, our two groups chose very different flavors. Our two groups are also the most similar when using the Bray-Curtis model compared to other class preferences. The next set of pairs have Bray-Curtis values around 0.6. This value corresponds to not having very similar preferences. Going back to Jaeyoung and myself, the pair of people who have the most different preferences than us are Trang and Yongsoo because they merged from us very early on and are very far from us on the dendrogram.

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 19th, 2025 at 12:00 PM (noon)**.