6. Worksheet: Among Site (Beta) Diversity – Part 1

Student Name; Z620: Quantitative Biodiversity, Indiana University

05 February, 2025

OVERVIEW

In this worksheet, we move beyond the investigation of within-site α-diversity. We will explore β-diversity, which is defined as the diversity that occurs among sites. This requires that we examine the compositional similarity of assemblages that vary in space or time.

After completing this exercise you will know how to:

1. formally quantify β-diversity
2. visualize β-diversity with heatmaps, cluster analysis, and ordination
3. test hypotheses about β-diversity using multivariate statistics

Directions:

1. In the Markdown version of this document in your cloned repo, change "Student Name" on line 3 (above) with 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 examples of proper scripting needed to carry out the exercises.
4. Answer questions in the worksheet. Space for your answers is provided in this document and is 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 (**6.BetaDiversity_1_Worksheet.Rmd**) with all code blocks filled out and questions answered) and the PDF output of Knitr (**6.BetaDiversity_1_Worksheet.pdf**).

The completed exercise is due on **Wednesday, February 5th, 2025 before 12:00 PM (noon)**.

1) R SETUP

Typically, the first thing you will do in either an R script or an RMarkdown file is setup your environment. This includes things such as setting the working directory and loading any packages that you will need.

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 Week3-Beta/ folder folder, and
4) Load the vegan R package (be sure to install first if you have not already).

```
getwd()
```

```
## [1] "C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week3-Beta"
```

```
library(vegan)
```

```
## Warning: package ’vegan’ was built under R version 4.4.2
```

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

```
## Warning: package ’permute’ was built under R version 4.4.2
```

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

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

```
library(ade4)
```

```
## Warning: package ’ade4’ was built under R version 4.4.2
```

2) LOADING DATA

Load dataset

In the R code chunk below, do the following:

1. load the doubs dataset from the ade4 package, and
2. explore the structure of the dataset.

```
# note, please do not print the dataset when submitting
data(doubs)

str(doubs)
```

```
## List of 4
##  $ env    :’data.frame’: 30 obs. of  11 variables:
##   ..$ dfs: num [1:30] 3 22 102 185 215 324 268 491 705 990 ...
##   ..$ alt: num [1:30] 934 932 914 854 849 846 841 792 752 617 ...
##   ..$ slo: num [1:30] 6.18 3.43 3.64 3.5 3.18 ...
##   ..$ flo: num [1:30] 84 100 180 253 264 286 400 130 480 1000 ...
##   ..$ pH : num [1:30] 79 80 83 80 81 79 81 81 80 77 ...
##   ..$ har: num [1:30] 45 40 52 72 84 60 88 94 90 82 ...
##   ..$ pho: num [1:30] 1 2 5 10 38 20 7 20 30 6 ...
##   ..$ nit: num [1:30] 20 20 22 21 52 15 15 41 82 75 ...
##   ..$ amm: num [1:30] 0 10 5 0 20 0 0 12 12 1 ...
##   ..$ oxy: num [1:30] 122 103 105 110 80 102 111 70 72 100 ...
##   ..$ bdo: num [1:30] 27 19 35 13 62 53 22 81 52 43 ...
##  $ fish   :’data.frame’: 30 obs. of  27 variables:
##   ..$ Cogo: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Satr: num [1:30] 3 5 5 4 2 3 5 0 0 1 ...
##   ..$ Phph: num [1:30] 0 4 5 5 3 4 4 0 1 4 ...
##   ..$ Neba: num [1:30] 0 3 5 5 2 5 5 0 3 4 ...
##   ..$ Thth: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Teso: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Chna: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Chto: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Lele: num [1:30] 0 0 0 0 5 1 1 0 0 2 ...
##   ..$ Lece: num [1:30] 0 0 0 1 2 2 1 0 5 2 ...
##   ..$ Baba: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Spbi: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Gogo: num [1:30] 0 0 0 1 2 1 0 0 0 1 ...
##   ..$ Eslu: num [1:30] 0 0 1 2 4 1 0 0 0 0 ...
##   ..$ Pefl: num [1:30] 0 0 0 2 4 1 0 0 0 0 ...
##   ..$ Rham: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Legi: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Scer: num [1:30] 0 0 0 0 2 0 0 0 0 0 ...
##   ..$ Cyca: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Titi: num [1:30] 0 0 0 1 3 2 0 0 1 0 ...
##   ..$ Abbr: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Icme: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Acce: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Ruru: num [1:30] 0 0 0 0 5 1 0 0 4 0 ...
##   ..$ Blbj: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Alal: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ Anan: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##  $ xy     :’data.frame’: 30 obs. of  2 variables:
##   ..$ x: num [1:30] 88 94 102 100 106 112 114 110 136 168 ...
##   ..$ y: num [1:30] 7 14 18 28 39 51 61 76 100 112 ...
##  $ species:’data.frame’: 27 obs. of  4 variables:
##   ..$ Scientific: chr [1:27] "Cottus gobio" "Salmo trutta fario" "Phoxinus phoxinus" "Nemacheilus ba
##   ..$ French    : chr [1:27] "chabot" "truite fario" "vairon" "loche franche" ...
##   ..$ English   : chr [1:27] "european bullhead" "brown trout" "minnow" "stone loach" ...
##   ..$ code      : Factor w/ 27 levels "Abbr","Acce",..: 9 22 19 17 26 25 7 8 16 14 ...
```

```r
env<-doubs$env

fish<-doubs$fish


str(env)
```

```
## ’data.frame’:    30 obs. of  11 variables:
##  $ dfs: num  3 22 102 185 215 324 268 491 705 990 ...
##  $ alt: num  934 932 914 854 849 846 841 792 752 617 ...
##  $ slo: num  6.18 3.43 3.64 3.5 3.18 ...
##  $ flo: num  84 100 180 253 264 286 400 130 480 1000 ...
##  $ pH : num  79 80 83 80 81 79 81 81 80 77 ...
##  $ har: num  45 40 52 72 84 60 88 94 90 82 ...
##  $ pho: num  1 2 5 10 38 20 7 20 30 6 ...
##  $ nit: num  20 20 22 21 52 15 15 41 82 75 ...
##  $ amm: num  0 10 5 0 20 0 0 12 12 1 ...
##  $ oxy: num  122 103 105 110 80 102 111 70 72 100 ...
##  $ bdo: num  27 19 35 13 62 53 22 81 52 43 ...
```

```
str(fish)
```

```
## 'data.frame':    30 obs. of  27 variables:
##  $ Cogo: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Satr: num  3 5 5 4 2 3 5 0 0 1 ...
##  $ Phph: num  0 4 5 5 3 4 4 0 1 4 ...
##  $ Neba: num  0 3 5 5 2 5 5 0 3 4 ...
##  $ Thth: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Teso: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Chna: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Chto: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Lele: num  0 0 0 0 5 1 1 0 0 2 ...
##  $ Lece: num  0 0 0 1 2 2 1 0 5 2 ...
##  $ Baba: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Spbi: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Gogo: num  0 0 0 1 2 1 0 0 0 1 ...
##  $ Eslu: num  0 0 1 2 4 1 0 0 0 0 ...
##  $ Pefl: num  0 0 0 2 4 1 0 0 0 0 ...
##  $ Rham: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Legi: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Scer: num  0 0 0 0 2 0 0 0 0 0 ...
##  $ Cyca: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Titi: num  0 0 0 1 3 2 0 0 1 0 ...
##  $ Abbr: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Icme: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Acce: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Ruru: num  0 0 0 0 5 1 0 0 4 0 ...
##  $ Blbj: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Alal: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Anan: num  0 0 0 0 0 0 0 0 0 0 ...
```

```
ncol(fish)
```

```
## [1] 27
```

Question 1: Describe some of the attributes of the doubs dataset.

a. How many objects are in doubs?
b. How many fish species are there in the doubs dataset?
c. How many sites are in the doubs dataset?

> ***Answer 1a***: There are four dataframes in the list called Doubs.
> ***Answer 1b***: 27 ***Answer 1c***: 30

Visualizing the Doubs River Dataset

Question 2: Answer the following questions based on the spatial patterns of richness (i.e., α-diversity) and Brown Trout (*Salmo trutta*) abundance in the Doubs River.

a. How does fish richness vary along the sampled reach of the Doubs River?
b. How does Brown Trout (*Salmo trutta*) abundance vary along the sampled reach of the Doubs River?
c. What do these patterns say about the limitations of using richness when examining patterns of biodiversity?

> ***Answer 2a***: From downsteam to upstream we see that there is a decrease in richness. ***Answer 2b***:Meanwhile we see an increase in trout abundance from Downstream to upstream ***Answer 2c***: The current patterns suggest that trout abudance in increasing despite there bring a decrease in species richess suggesting increase in dominance of the trout from Downstream to upstream.

3) QUANTIFYING BETA-DIVERSITY

In the R code chunk below, do the following:

1. write a function (beta.w()) to calculate Whittaker's β-diversity (i.e., β_w) that accepts a site-by-species matrix with optional arguments to specify pairwise turnover between two sites, and
2. use this function to analyze various aspects of β-diversity in the Doubs River. Here is what I have so far. I want to be able to answer the last question.

```r
beta.w <- function(site.by.species = "") {
  SbyS.pa <- decostand(site.by.species, method = "pa")  # Convert to presence-absence
  S <- ncol(SbyS.pa[, which(colSums(SbyS.pa) > 0)])    # Number of species in the region
  a.bar <- mean(specnumber(SbyS.pa))                    # Average richness at each site
  b.w <- round(S / a.bar - 1, 3)                        # Compute beta-diversity
  return(b.w)
}



beta_div <- beta.w(fish)
beta_div
```

```
## [1] 1.16
```

```r
#-------------pairwise beta-diversity------------


beta.w <- function(site.by.species = "", sitenum1 = "", sitenum2 = "", pairwise = FALSE) {

  if (pairwise == TRUE) {
    if (sitenum1 == "" | sitenum2 == "") {
      print("Error: please specify sites to compare")
      return(NA)
    }

    site1 <- site.by.species[sitenum1,]
    site2 <- site.by.species[sitenum2,]

    site1 <- subset(site1, select = site1 > 0)  # Remove absences
    site2 <- subset(site2, select = site2 > 0)  # Remove absences

    gamma <- union(colnames(site1), colnames(site2))  # Gamma richness
    S <- length(gamma)
    a.bar <- mean(c(specnumber(site1), specnumber(site2)))  # Mean sample richness

    b.w <- round(S / a.bar - 1, 3)
    return(b.w)
  }
  else {
    SbyS.pa <- decostand(site.by.species, method = "pa")  # Convert to presence-absence
    S <- ncol(SbyS.pa[, which(colSums(SbyS.pa) > 0)])     # Number of species in the region
    a.bar <- mean(specnumber(SbyS.pa))                    # Average richness at each site
    b.w <- round(S / a.bar - 1, 3)                        # Compute betadiversity
    return(b.w)
  }
}


#calculations pairwise beta-diversity


beta_A <- beta.w(fish, sitenum1 = 1, sitenum2 = 2, pairwise = TRUE)
beta_B<-beta.w(fish, sitenum1 = 1, sitenum2 = 10, pairwise = TRUE)

beta_A
```

```
## [1] 0.5
```

```r
beta_B
```

```
## [1] 0.714
```

Question 3: Using your `beta.w()` function above, answer the following questions:

a. Describe how local richness (α) and turnover (β) contribute to regional (γ) fish diversity in the Doubs.
b. Is the fish assemblage at site 1 more similar to the one at site 2 or site 10?
c. Using your understanding of the equation $\beta_w = \gamma/\alpha$, how would your interpretation of β change if we instead defined beta additively (i.e., $\beta = \gamma - \alpha$)?

> ***Answer 3a***: Since beta diversity is calculated as beta = gamma/alpha, and the value we get is 1.16 (beta-diversity). Then, gamma will be seen as 1.16 times alpha.
>
> ***Answer 3b***: Beta is inversely proportional to Alpha and hence since 1 and 2 have the lower Beta-diversity, they are the closest.
>
> ***Answer 3c***: The relationship would be such that beta would still decrease with decrease in alpha, however the shape of thge curve would be different . However, this relationship would be weird sicen gamma will be beta+alpha, which makes no sense logically speaking, why would you added "turnover rate" to a metric of diversity?

The Resemblance Matrix

In order to quantify β-diversity for more than two samples, we need to introduce a new primary ecological data structure: the **Resemblance Matrix**.

Question 4: How do incidence- and abundance-based metrics differ in their treatment of rare species?

> ***Answer 4***:

In the R code chunk below, do the following:

1. make a new object, `fish`, containing the fish abundance data for the Doubs River,
2. remove any sites where no fish were observed (i.e., rows with sum of zero),
3. construct a resemblance matrix based on Sørensen's Similarity ("fish.ds"), and
4. construct a resemblance matrix based on Bray-Curtis Distance ("fish.db").

```
fish <- doubs$fish
fish <- fish[-8, ] #Removes site 8 from data

#Calculate Jaccard
fish.dj <- vegdist(fish, method = "jaccard", binary = TRUE)
new.fish.dj <- vegdist(fish, method = "jaccard", binary = TRUE, upper = TRUE, diag = TRUE)
#Calculate Bray-Curtis
fish.db <- vegdist(fish, method = "bray")

#Calculate Sørensen
fish.ds <- vegdist(fish, method = "bray", binary = TRUE)
full.fish.ds <- vegdist(fish, method = "bray", binary = TRUE, upper = TRUE, diag = TRUE)


fish.db
```

```
##             1          2          3          4          5          6          7
## 2  0.60000000                                                                  
## 3  0.68421053 0.14285714                                                       
## 4  0.75000000 0.33333333 0.18918919                                            
## 5  0.89189189 0.69565217 0.68000000 0.49090909                                 
## 6  0.75000000 0.39393939 0.29729730 0.19047619 0.41818182                      
## 7  0.68421053 0.14285714 0.12500000 0.24324324 0.64000000 0.24324324           
## 9  1.00000000 0.69230769 0.73333333 0.65714286 0.58333333 0.54285714 0.66666667
## 10 0.88235294 0.38461538 0.40000000 0.37142857 0.54166667 0.25714286 0.26666667
## 11 0.57142857 0.30434783 0.40740741 0.43750000 0.68888889 0.43750000 0.33333333
## 12 0.71428571 0.20000000 0.23529412 0.33333333 0.69230769 0.38461538 0.17647059
## 13 0.72727273 0.29032258 0.31428571 0.45000000 0.73584906 0.55000000 0.37142857
## 14 0.80645161 0.40000000 0.31818182 0.34693878 0.67741935 0.42857143 0.36363636
## 15 0.83333333 0.51111111 0.46938776 0.40740741 0.55223881 0.37037037 0.38775510
## 16 0.86046512 0.65384615 0.57142857 0.47540984 0.45945946 0.37704918 0.53571429
## 17 0.91489362 0.67857143 0.63333333 0.50769231 0.51282051 0.44615385 0.60000000
## 18 0.95555556 0.74074074 0.72413793 0.58730159 0.50000000 0.52380952 0.68965517
## 19 1.00000000 0.79310345 0.70967742 0.61194030 0.50000000 0.52238806 0.67741935
## 20 1.00000000 0.91176471 0.88888889 0.74025974 0.48888889 0.68831169 0.86111111
## 21 1.00000000 0.94594595 0.92307692 0.78313253 0.50000000 0.73493976 0.89743590
## 22 1.00000000 0.97619048 0.95454545 0.82795699 0.52830189 0.78494624 0.93181818
## 23 1.00000000 1.00000000 1.00000000 0.92000000 0.89473684 0.84000000 0.90000000
## 24 1.00000000 1.00000000 1.00000000 0.88888889 0.79591837 0.77777778 0.93548387
## 25 1.00000000 1.00000000 0.92592593 0.81250000 0.68888889 0.68750000 0.85185185
## 26 1.00000000 0.96363636 0.93220339 0.78125000 0.55844156 0.68750000 0.89830508
## 27 1.00000000 0.97333333 0.94936709 0.83333333 0.56701031 0.76190476 0.92405063
## 28 1.00000000 0.97560976 0.95348837 0.82417582 0.57692308 0.78021978 0.93023256
## 29 0.97777778 0.93939394 0.92233010 0.81481481 0.53719008 0.77777778 0.90291262
## 30 1.00000000 1.00000000 0.98095238 0.87272727 0.59349593 0.83636364 0.96190476
##             9         10         11         12         13         14         15
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10 0.57142857
## 11 0.76000000 0.44000000
## 12 0.68750000 0.37500000 0.24137931
## 13 0.81818182 0.57575758 0.33333333 0.18918919
## 14 0.76190476 0.47619048 0.43589744 0.21739130 0.19148936
## 15 0.65957447 0.40425532 0.50000000 0.33333333 0.38461538 0.24590164
## 16 0.70370370 0.51851852 0.64705882 0.55172414 0.59322034 0.44117647 0.26027397
## 17 0.68965517 0.51724138 0.63636364 0.58064516 0.61904762 0.50000000 0.40259740
## 18 0.64285714 0.57142857 0.69811321 0.66666667 0.70491803 0.60000000 0.46666667
## 19 0.66666667 0.63333333 0.82456140 0.75000000 0.81538462 0.67567568 0.56962025
## 20 0.68571429 0.77142857 0.91044776 0.89189189 0.92000000 0.83333333 0.70786517
## 21 0.76315789 0.81578947 0.91780822 0.92500000 0.95061728 0.86666667 0.76842105
## 22 0.76744186 0.86046512 0.95180723 0.95555556 0.97802198 0.90000000 0.77142857
## 23 0.77777778 0.88888889 0.86666667 0.90909091 1.00000000 0.93750000 0.94594595
## 24 0.72413793 0.79310345 0.92307692 0.93939394 1.00000000 0.90697674 0.87500000
## 25 0.84000000 0.76000000 0.90909091 0.93103448 1.00000000 0.84615385 0.81818182
## 26 0.71929825 0.82456140 0.92592593 0.93442623 0.96774194 0.85915493 0.76315789
## 27 0.76623377 0.84415584 0.94594595 0.95061728 0.97560976 0.89010989 0.77083333
## 28 0.76190476 0.85714286 0.95061728 0.95454545 0.97752809 0.89795918 0.78640777
## 29 0.78217822 0.84158416 0.89795918 0.90476190 0.90566038 0.84347826 0.73333333
## 30 0.84466019 0.90291262 0.98000000 0.98130841 1.00000000 0.93162393 0.81967213
##            16         17         18         19         20         21         22
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17 0.26190476
## 18 0.34146341 0.13953488
## 19 0.39534884 0.31111111 0.25000000
## 20 0.58333333 0.42000000 0.32653061 0.23529412
## 21 0.62745098 0.49056604 0.40384615 0.29629630 0.10169492
## 22 0.66071429 0.55172414 0.47368421 0.38983051 0.18750000 0.10447761
## 23 0.90909091 0.83333333 0.82608696 0.84000000 0.86666667 0.87878788 0.89473684
## 24 0.81818182 0.69491525 0.64912281 0.63934426 0.57746479 0.61038961 0.65517241
## 25 0.76470588 0.74545455 0.66037736 0.61403509 0.67164179 0.69863014 0.73493976
## 26 0.63855422 0.54022989 0.45882353 0.32584270 0.21212121 0.20000000 0.25217391
## 27 0.66990291 0.57009346 0.48571429 0.37614679 0.19327731 0.13600000 0.12592593
## 28 0.69090909 0.57894737 0.50000000 0.41379310 0.22222222 0.16666667 0.12676056
## 29 0.65354331 0.51145038 0.44186047 0.41353383 0.24475524 0.18120805 0.11949686
## 30 0.72093023 0.57894737 0.52671756 0.48148148 0.29655172 0.23178808 0.18012422
##            23         24         25         26         27         28         29
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24 0.57894737
## 25 0.46666667 0.46153846
## 26 0.82978723 0.48275862 0.59259259
## 27 0.88059701 0.61538462 0.70270270 0.18867925
## 28 0.89189189 0.64705882 0.72839506 0.23893805 0.09774436
## 29 0.91208791 0.70588235 0.77551020 0.33846154 0.18666667 0.14649682
## 30 0.91397849 0.71153846 0.78000000 0.36363636 0.19736842 0.15723270 0.14772727
```

Question 5: Using the distance matrices from above, answer the following questions:

a. Does the resemblance matrix (`fish.db`) represent similarity or dissimilarity? What information in the resemblance matrix led you to arrive at your answer?
b. Compare the resemblance matrices (`fish.db` or `fish.ds`) you just created. How does the choice of the Sørensen or Bray-Curtis distance influence your interpretation of site (dis)similarity?

> ***Answer 5a***: The values in matrix represent dissimilarity, we see that there are no values for self paired sites. greater the value the more dissimilar the pairs are ***Answer 5b***:Bray-curtis takes abundance of species into account while Sorenson is an incidence based metric. Hence species eveness is not represented with the Sorenson metric and just have species idenities similar (irrespective of abundances) you are highly similar

4) VISUALIZING BETA-DIVERSITY

A. Heatmaps

In the R code chunk below, do the following:

1. define a color palette,
2. define the order of sites in the Doubs River, and
3. use the `levelplot()` function to create a heatmap of fish abundances in the Doubs River.

```
# Order of Sites
library(gplots)
```

```
## Warning: package 'gplots' was built under R version 4.4.2
```

```
##
## Attaching package: 'gplots'
```

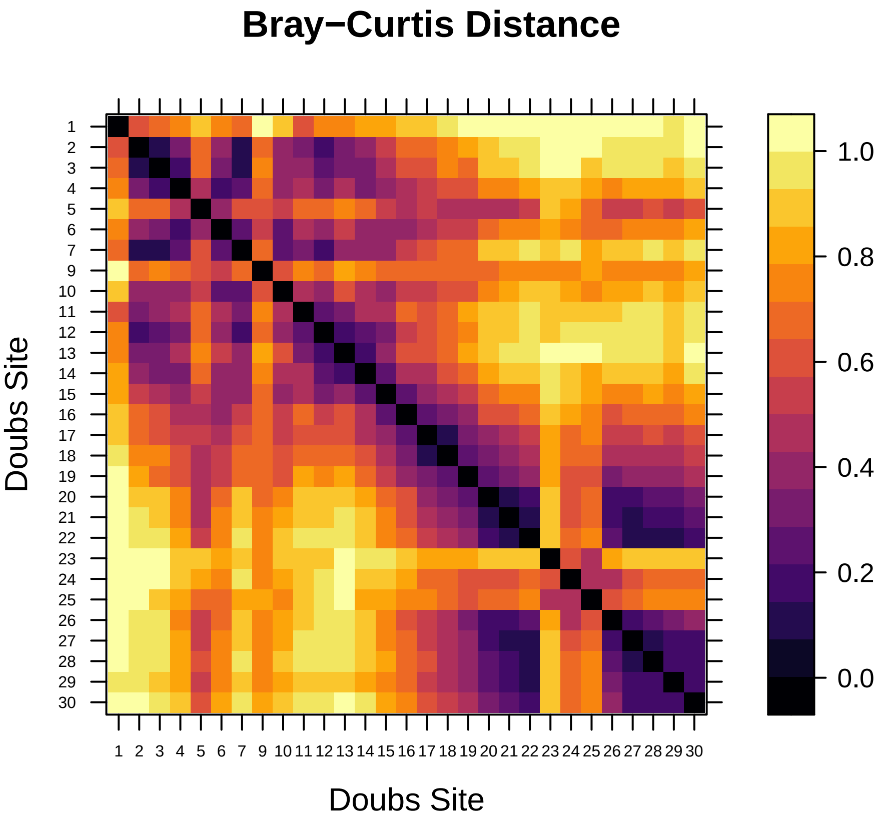
```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(viridis)
```

```
## Warning: package 'viridis' was built under R version 4.4.2
```

```
## Loading required package: viridisLite
```

```
order <- rev(attr(fish.db, "Labels"))

# Heatmap
levelplot(as.matrix(fish.db)[,order], aspect = "iso", col.regions = inferno,
          xlab = "Doubs Site", ylab = "Doubs Site", scales = list(cex = 0.5),
          main = "Bray-Curtis Distance")
```

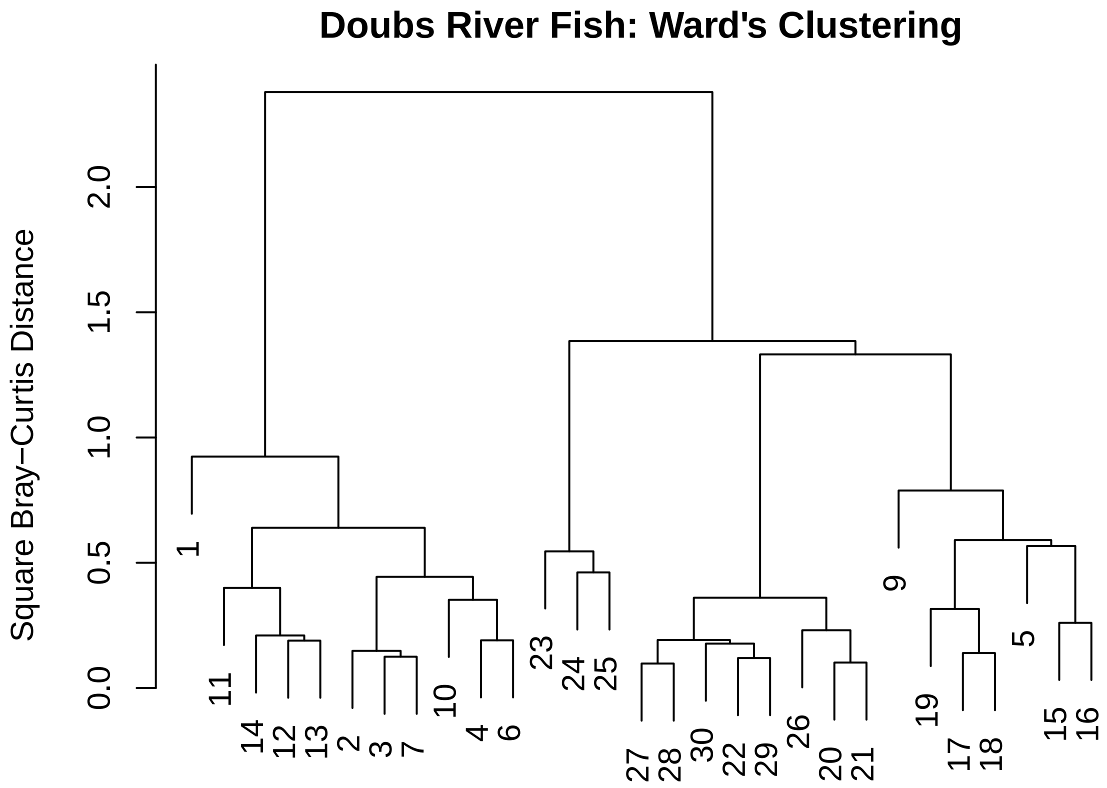


B. Cluster Analysis

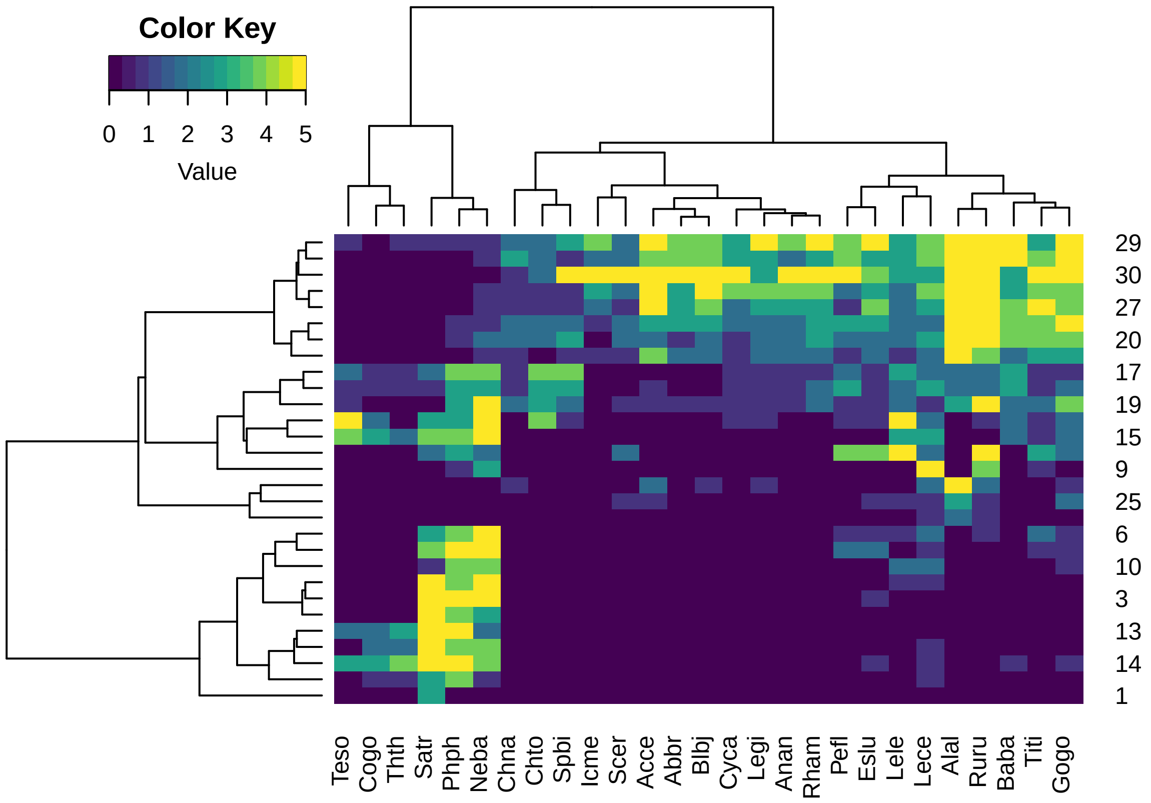
In the R code chunk below, do the following:

1. perform a cluster analysis using Ward's Clustering, and
2. plot your cluster analysis (use either `hclust` or `heatmap.2`).

```
# Cluster Analysis
fish.ward <- hclust(fish.db, method = "ward.D2")


#Plot for Cluster
par(mar = c(1,5,2,2) + 0.1)
plot(fish.ward, main = "Doubs River Fish: Ward's Clustering", ylab = "Square Bray-Curtis Distance")
```



```r
gplots::heatmap.2(as.matrix(fish),
                  distfun = function(x) vegdist(x, method = "bray"),
                  hclustfun = function(x) hclust(x, method = "ward.D2"),
                  col = viridis, trace = "none", density.info = "none")
```



Question 6: Based on cluster analyses and the introductory plots that we generated after loading the data, develop an ecological hypothesis for fish diversity the `doubs` data set?

> ***Answer 6***: From this we can assume that most sites from number 1 - 14 are very similar in their species composition and are likely to be in a similar habitat which leads to a similar community structure. The same logic can be applied for the other sets of plots. It also looks like in 1-14 sites most species appear to be low relative abudance, while in plots 20- 29 some species are dominant (high relative abundance)

C. Ordination

Principal Coordinates Analysis (PCoA)

In the R code chunk below, do the following:

1. perform a Principal Coordinates Analysis to visualize beta-diversity
2. calculate the variation explained by the first three axes in your ordination
3. plot the PCoA ordination,
4. label the sites as points using the Doubs River site number, and
5. identify influential species and add species coordinates to PCoA plot.

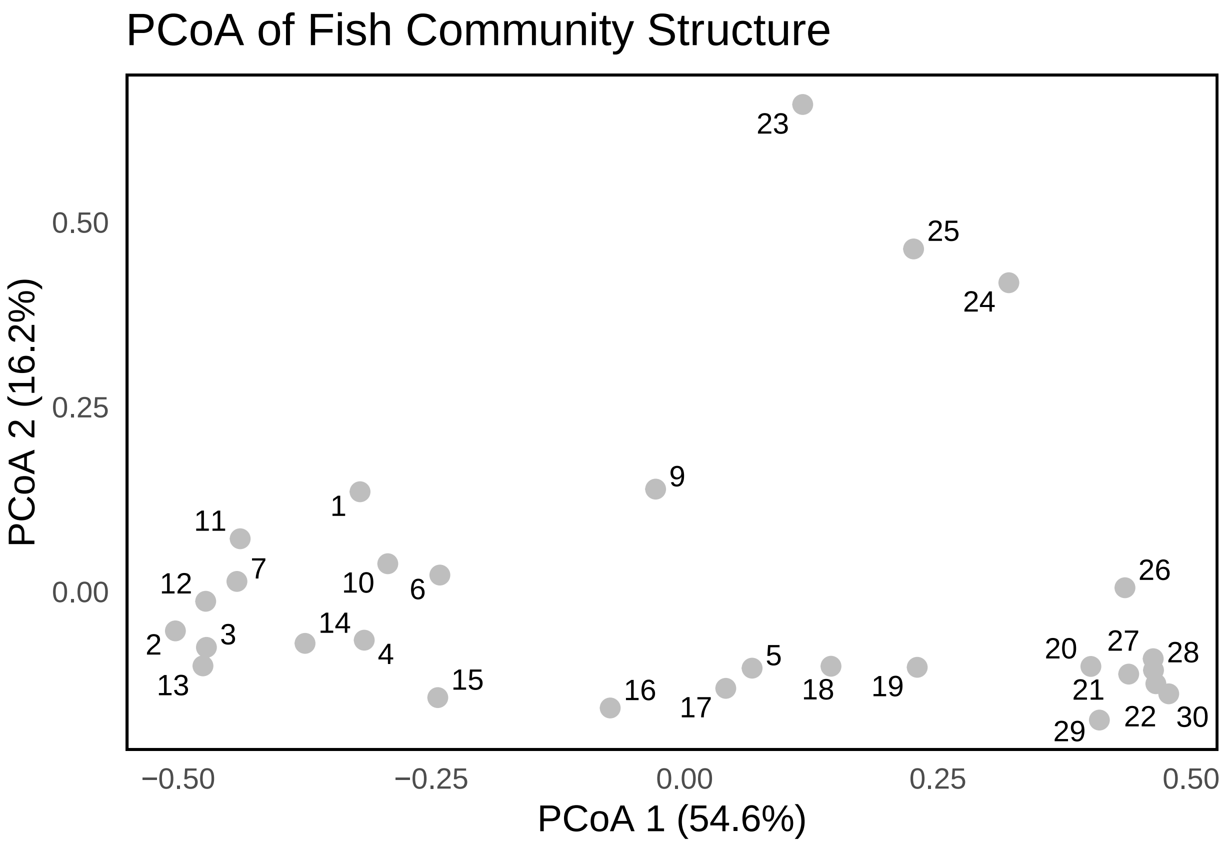
```
fish.pcoa <- cmdscale(fish.db, eig = TRUE, k = 3)

explainvar1 <- round(fish.pcoa$eig[1]/sum(fish.pcoa$eig), 3) * 100
explainvar2 <- round(fish.pcoa$eig[2]/sum(fish.pcoa$eig), 3) * 100
```

```r
explainvar3 <- round(fish.pcoa$eig[3]/sum(fish.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)

#----------Define Plot Parameters------
library(ggplot2)
library(ggrepel)  # For better text label positioning
```

```
## Warning: package ’ggrepel’ was built under R version 4.4.2
```

```r
# Create a data frame from fish.pcoa
fish_pcoa_df <- data.frame(
  PCoA1 = fish.pcoa$points[,1],
  PCoA2 = fish.pcoa$points[,2],
  Site = row.names(fish.pcoa$points)  # Site labels
)

# Create the plot
ggplot(fish_pcoa_df, aes(x = PCoA1, y = PCoA2, label = Site)) +
  geom_point(size = 3, color = "gray") +  # Plot points
  geom_text_repel(size = 4, max.overlaps = Inf) +  # Fix overlap issue
  labs(
    x = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
    y = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
    title = "PCoA of Fish Community Structure"
  ) +
  theme_minimal(base_size = 14) +  # Clean minimal theme
  theme(
    panel.grid.major = element_blank(),  # Remove major grid
    panel.grid.minor = element_blank(),  # Remove minor grid
    panel.border = element_rect(color = "black", fill = NA, size = 1)  # Add box
  )
```

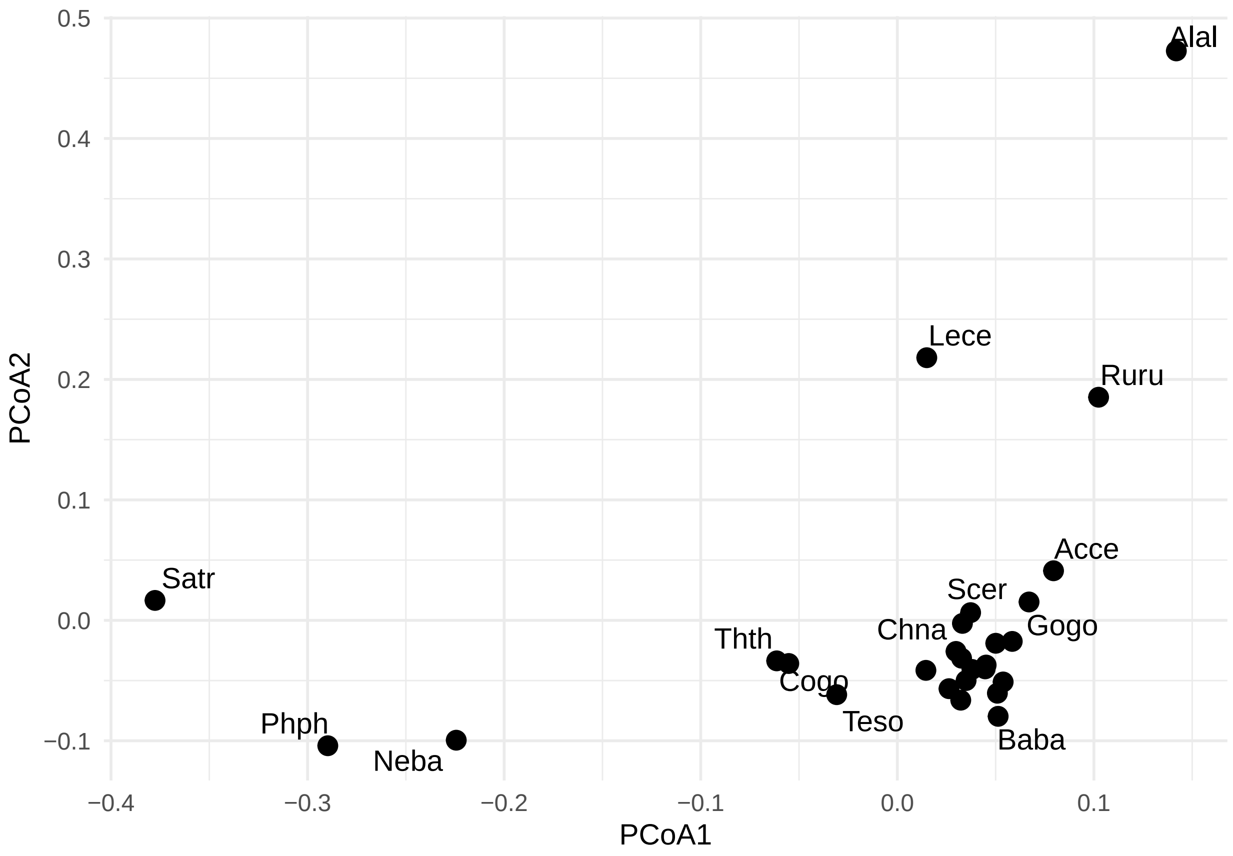
```
## Warning: The ‘size‘ argument of ‘element_rect()‘ is deprecated as of ggplot2 3.4.0.
## i Please use the ‘linewidth‘ argument instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



```
#First we calculate the relative abundances of each species at each site
fishREL <- fish
for(i in 1:nrow(fish)){
  fishREL[i, ] =fish[i, ] / sum(fish[i, ])
}

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
    }
    specscores <- specscores * multi
    ordi$cproj <- specscores
    return(ordi)
  }


#species scores based ont he relative abundance of different species


library(vegan)

fish.pcoa <- add.spec.scores.class(fish.pcoa,fishREL, method = "pcoa.scores")

# Convert species projections into a data frame
species_proj <- data.frame(
  PCoA1 = fish.pcoa$cproj[,1],
  PCoA2 = fish.pcoa$cproj[,2],
  Label = row.names(fish.pcoa$cproj)
)
```

```
# Plot species projections
ggplot(species_proj, aes(x = PCoA1, y = PCoA2)) +
  geom_point(color = "black", size = 3) +
  geom_text_repel(aes(label = Label), size = 4) +
  theme_minimal()
```

```
## Warning: ggrepel: 13 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



In the R code chunk below, do the following:

1. identify influential species based on correlations along each PCoA axis (use a cutoff of 0.70), and
2. use a permutation test (999 permutations) to test the correlations of each species along each axis.

```
spe.corr <- add.spec.scores.class(fish.pcoa, fishREL, method = "cor.scores")$cproj
corrcut <- 0.7 #user defined cutoff
imp.spp <- spe.corr[abs(spe.corr[,1]) >= corrcut | abs(spe.corr[,2]) >= corrcut, ]

#Permutation test for Spcies Abundances Acros Axes
fit <- envfit(fish.pcoa, fishREL, perm = 999)
fit
```

```
##
## ***VECTORS
```

```
##
##          Dim1     Dim2     r2 Pr(>r)
## Cogo -0.83884 -0.54438 0.2982  0.015 *
## Satr -0.99904  0.04371 0.4326  0.004 **
## Phph -0.94110 -0.33813 0.7814  0.001 ***
## Neba -0.91413 -0.40543 0.6234  0.001 ***
## Thth -0.87692 -0.48063 0.2634  0.020 *
## Teso -0.44704 -0.89452 0.1700  0.075 .
## Chna  0.99707 -0.07644 0.4612  0.005 **
## Chto  0.42032 -0.90738 0.2579  0.019 *
## Lele  0.33041 -0.94384 0.0495  0.547
## Lece  0.06856  0.99765 0.3399  0.006 **
## Baba  0.54118 -0.84091 0.6752  0.001 ***
## Spbi  0.57341 -0.81927 0.4138  0.001 ***
## Gogo  0.97507  0.22188 0.3753  0.004 **
## Eslu  0.72044 -0.69352 0.1673  0.090 .
## Pefl  0.43762 -0.89916 0.3048  0.010 **
## Rham  0.72476 -0.68901 0.8301  0.001 ***
## Legi  0.93461 -0.35568 0.7016  0.001 ***
## Scer  0.98569  0.16858 0.3533  0.013 *
## Cyca  0.68181 -0.73153 0.7743  0.001 ***
## Titi  0.64378 -0.76521 0.4586  0.001 ***
## Abbr  0.77254 -0.63497 0.7128  0.001 ***
## Icme  0.75626 -0.65427 0.5270  0.003 **
## Acce  0.88799  0.45986 0.6294  0.001 ***
## Ruru  0.48379  0.87518 0.5177  0.001 ***
## Blbj  0.95802 -0.28671 0.6766  0.001 ***
## Alal  0.28755  0.95777 0.8592  0.001 ***
## Anan  0.74277 -0.66954 0.7894  0.001 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## Permutation: free
## Number of permutations: 999
```

```r
#Create "Factors" vector

quality <- c(rep("HQ", 13), rep("MQ", 5), rep("LQ", 6), rep("MQ", 5))
#Run PERMANOVA with adonis function

adonis2(fish ~ quality, method = "bray", permutations = 999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = fish ~ quality, permutations = 999, method = "bray")
##          Df SumOfSqs      R2     F Pr(>F)
## Model     2   3.0947 0.45765 10.97  0.001 ***
## Residual 26   3.6674 0.54235
## Total    28   6.7621 1.00000
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

Question 7: Address the following questions about the ordination results of the `doubs` data set:

a. Describe the grouping of sites in the Doubs River based on fish community composition.
b. Generate a hypothesis about which fish species are potential indicators of river quality.

> ***Answer 7a***: As hypothesized before we see that there is clustering from 1-14 and the rest cluster similarly separately. This suggests that these group are formed of sites that are similar with each other
> ***Answer 7b***: Anan, Acce and Alal appear to be the most sensitive since since they have the highest correlation coeeficient values.

SYNTHESIS

Load the dataset from that you and your partner are using for the team project. Use one of the tools introduced in the beta diversity module to visualize your data. Describe any interesting patterns and identify a hypothesis is relevant to the principles of biodiversity.