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

Joshua Jones; Z620: Quantitative Biodiversity, Indiana University

16 April, 2021

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 today, it is *imperative* that you **push** this file to your GitHub repo, at whatever stage you are. Ths will enable you to pull your work onto your own computer.
6. When you have completed the worksheet, **Knit** the text and code into a single PDF file by pressing the `Knit` button in the RStudio scripting panel. This will save the PDF output in your '8.BetaDiversity' folder.
7. After Knitting, please submit the worksheet by making a **push** to your GitHub repo and then create a **pull request** via GitHub. Your pull request should include this file (**8.BetaDiversity_1_Worksheet.Rmd**) with all code blocks filled out and questions answered) and the PDF output of `Knitr` (**8.BetaDiversity_1_Worksheet.pdf**).

The completed exercise is due on **Friday, April 16th, 2021 before 09:00 AM**.

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, provide the code to:

1. clear your R environment,
2. print your current working directory,
3. set your working directory to your "*/8.BetaDiversity*" folder, and
4. load the `vegan` R package (be sure to install if needed).

```
rm(list = ls())
getwd()
```

```
## [1] "C:/Users/joshu/quantbio/QB2021_Jones/2.Worksheets/8.BetaDiversity"
```

```
#setwd("C:/Users/joshu/quantbio/QB2021_Jones/2.Worksheets/8.BetaDiversity")
library(vegan)
```

```
## Warning: package 'vegan' was built under R version 4.0.4
```

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

```
## Warning: package 'permute' was built under R version 4.0.4
```

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

```
## This is vegan 2.5-7
```

```
package.list <- c('vegan', 'ade4', 'viridis', 'gplots', 'BiodiversityR', 'indicspecies')
for (package in package.list){
  if (!require(package, character.only = TRUE, quietly = TRUE)) {
    install.packages(package)
    library(package,character.only = TRUE)
  }
}
```

```
## Warning: package 'ade4' was built under R version 4.0.5
```

```
## Warning: package 'viridisLite' was built under R version 4.0.5
```

```
## Warning: package 'gplots' was built under R version 4.0.4
```

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

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

```
## Warning: package 'BiodiversityR' was built under R version 4.0.5
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
## BiodiversityR 2.12-3: Use command BiodiversityRGUI() to launch the Graphical User Interface;
## to see changes use BiodiversityRGUI(changeLog=TRUE, backward.compatibility.messages=TRUE)
```

```
## Warning: package 'indicspecies' was built under R version 4.0.5
```

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, pleae 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 ...
```

```
head(doubs$env)
```

```
##   dfs alt   slo flo pH har pho nit amm oxy bdo
## 1   3 934 6.176  84 79  45   1  20   0 122  27
## 2  22 932 3.434 100 80  40   2  20  10 103  19
## 3 102 914 3.638 180 83  52   5  22   5 105  35
## 4 185 854 3.497 253 80  72  10  21   0 110  13
## 5 215 849 3.178 264 81  84  38  52  20  80  62
## 6 324 846 3.497 286 79  60  20  15   0 102  53
```

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 30 objects ***Answer 1b***: There are 27 fish species ***Answer 1c***: There are 30 sample sites

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***: There seems to be a relatively lower richness entering this portion of the river but halfway through the richess seems to get a lot higher, followed by a low diversity region, then another high diversity region leading to the downstream portion. ***Answer 2b***: There is high brown trout salmon abundance leading into this section of the river and right before the bend of the river but it seems to be relatively low or nonexistant along the rest of this portion of the river. ***Answer 2c***: Richness shows an overall measure of species presence but doesn't provide any information on specific species and their population levels within those locations.

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.

```
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)
    site2 = subset(site2, select = site2 > 0)
    gamma = union(colnames(site1), colnames(site2))
    s     = length(gamma)
    a.bar = mean(c(specnumber(site1), specnumber(site2)))
    b.w   = round(s/a.bar - 1, 3)
    return(b.w)
  }
  else{
    SbyS.pa <- decostand(site.by.species, method = "pa")
    S <- ncol(SbyS.pa[,which(colSums(SbyS.pa) > 0)])
    a.bar <- mean(specnumber(SbyS.pa))
    b.w <- round(S/a.bar, 3)
    return(b.w)
  }
}

beta.w(doubs$fish)
```

```
## [1] 2.16
```

```
beta.w(doubs$fish,sitenum1 = 1, sitenum2 = 2, pairwise = TRUE)
```

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

```
beta.w(doubs$fish,sitenum1 = 1, sitenum2 = 10, pairwise = TRUE)
```

```
## [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: When measuring Whittaker's β-diversity I get a value of 2.16, and considering that that is equal to the ratio between gamma diversity to the average alpha diversity across the sites I would say that alpha diversity seems to be relatively low and that turnover must have a higher contribution to the overall gamma diversity. ***Answer 3b***: Comparing 1 to 2 gives a value of .5 while comparing 1 to 10 gives a value of .714, and since 0 is minimum and 10 is maximum beta diversity we can determine that site 2 is more similar in composition to site 1 than site 10 is. ***Answer 3c***: Instead of interpreting beta diversity as the ratio of gamma diversity to average alpha diversity, or the turnover across the community, I would interpret the beta diversity as the variation across sites, or an estimate over the differences in species present across sites in the study.

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: Incidence-based measures treat rare species as equal in weight as more common species while abundance-based weights them smaller based on the relative abundance.

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").

```
#Importing site-by-species matrix from list
fish <- doubs$fish

#Removing all rows with a total of 0 total observations
fish <-fish[rowSums(fish !=0) > 0,]

#Constructing resemblance matrix based on Sørensen's Similarity
fish.ds <- vegdist(fish, method = "bray", binary = TRUE)

#Constructing a resemblance matrix based on Bray-Curtis Distance
fish.db <- vegdist(fish, method = "bray")
```

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 Bray-Curtis Dissimilarity matrix is a measure of dissimilarity since the value given to the relationship between 1 and 2 is lower than the value between 1 and 10, and earlier we found that sites 1 and 2 had higher similarity to each other than was between 1 and 10. ***Answer 5b***: When considering the different measures the main difference in interpretation would be the amount of the community being considered in the measueres. By that I mean that since Sørensen

is considering only who is there and not what their abundancences across the communities are I would have to consider that less information about the total community is being used than in the Bray-Curtis and that that, depending on the question I'm asking, would affect how confident I would be in the result. For example, if I was considering how environmental conditions was affecting community composition I would prefer to utilize total abundance in my metric while if I was only considering how environmental conditions affected the ability of species to exist across sites I wouldn't care as much how many were there but moreso just if they are there.

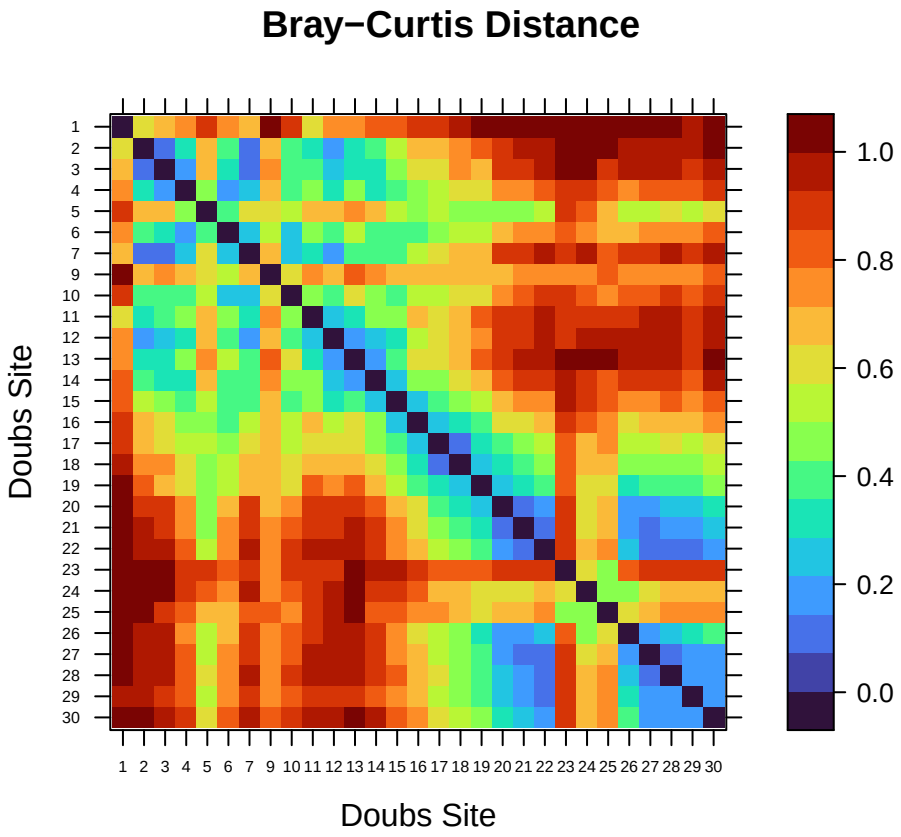
4) VISUALIZING BETA-DIVERSITY

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.

```
#Define Order of sites
order <- rev(attr(fish.db, "Labels"))

#Plot Heatmap
levelplot(as.matrix(fish.db)[,order], aspect = "iso", col.regions = turbo, xlab = "Doubs Site", ylab =
```

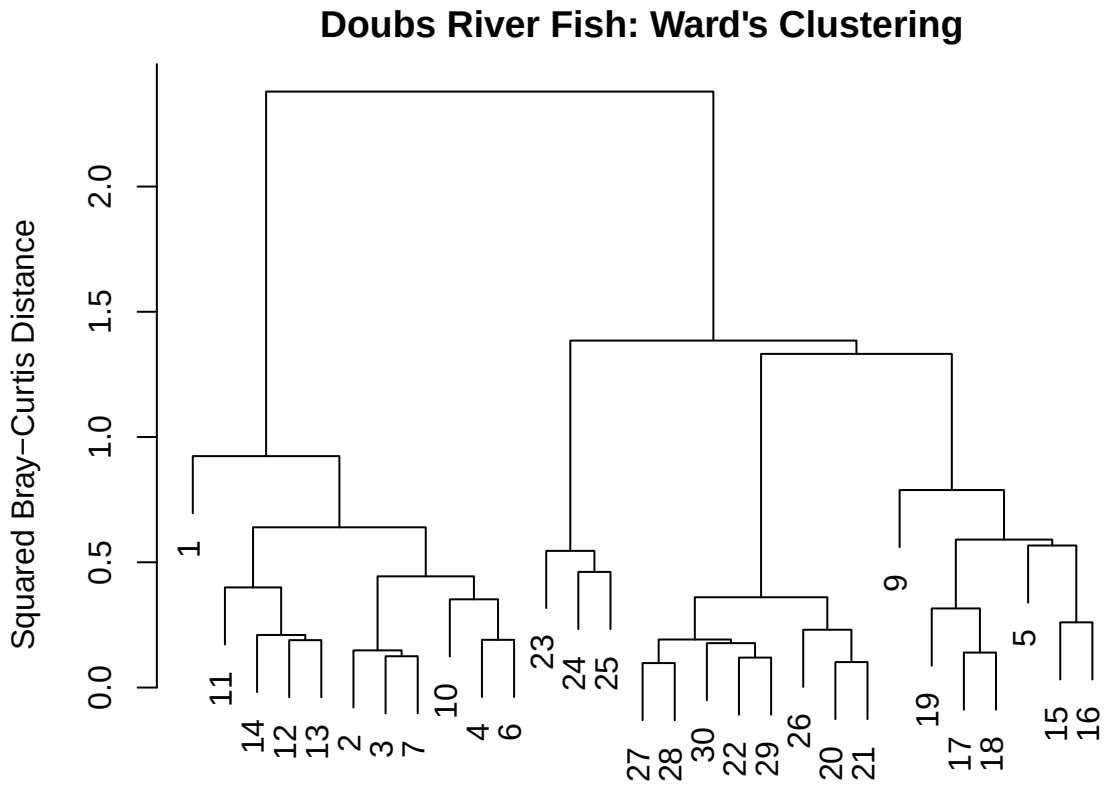


B. Cluster Analysis

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`).

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

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



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:

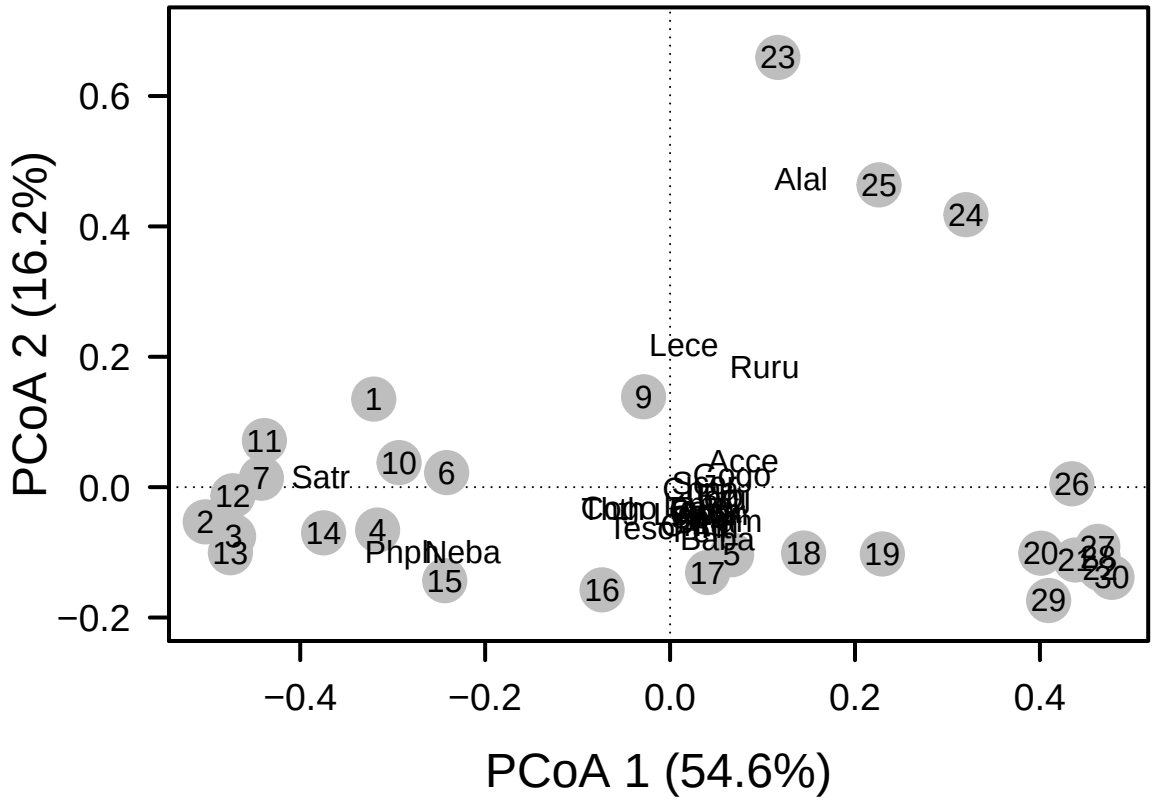
C. Ordination

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

#Define Plot Parameters
par(mar = c(5,5,1,2), .1)

#Initiate Plot
plot(fish.pcoa$points[,1], fish.pcoa$points[,2], ylim = c(-.2, .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)

#Add axis
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)

#Add Points & Labels
points(fish.pcoa$points[,1], fish.pcoa$points[,2],
       pch = 19, cex = 3, bg = "gray", col = "gray")
text(fish.pcoa$points[,1], fish.pcoa$points[,2],
     labels = row.names(fish.pcoa$points))

#Calculating relative abundance of species at each site
fishREL <- fish
for (i in 1:nrow(fish)){
  fishREL[i,] = fish[i,] / sum(fish[i,])
}

# Calculating and adding species scores
fish.pcoa <- add.spec.scores(fish.pcoa,fishREL, method = "pcoa.scores")
text(fish.pcoa$cproj[,1], fish.pcoa$cproj[,2],
     labels = row.names(fish.pcoa$cproj), col = "black")
```



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(fish.pcoa, fishREL, method = "cor.scores")$cproj
corrcut <- .7
imp.spp <- spe.corr[abs(spe.corr[,1]) >= corrcut | abs(spe.corr[,2]) >= corrcut, ]

fit <- envfit(fish.pcoa, fishREL, perm = 999)
```

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***: There seems to be 4 main goupings, the largest one driven buy high populations of Satr., Phph., and Neba, a smaller one driven by a higher population of Alal, and two others, one with a middling community composition, and another perhapse with a lack of those species that differentiate the first group from all the others. ***Answer 7b***: I would hypothesize that Alal. is a species that could be an indication of negative river quality and that Satr., Phph., and Neba. would be indicators of good river quality.

SYNTHESIS

Using the `mobsim` package from the DataWrangling module last week, simulate two local communities each containing 1000 individuals (N) and 25 species (S), but with one having a random spatial distribution and the other having a patchy spatial distribution. Take ten (10) subsamples from each site using the quadrat function and answer the following questions:

```
library(mobsim)
```

```
## Warning: package 'mobsim' was built under R version 4.0.4
```

```
#Make plot of whole community
com <- sim_poisson_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm", sad_coef = list("meanlog" =
comm_mat <- sample_quadrats(com, n_quadrats = 1, quadrat_area = .9, method = "random", avoid_overlap = '
```

```
# simulate community with `mobsim`
com1 <- sim_poisson_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))

# Lay down sampling quadrats on the community
comm_mat1 <- sample_quadrats(com1, n_quadrats = 10, quadrat_area = 0.01,
             method = "random", avoid_overlap = T)
```

```
# Rename sampled areas as quadrats
quads <- c("site1", "site2", "site3", "site4", "site5", "site6", "site7",
           "site8", "site9", "site10")
row.names(comm_mat1$xy_dat) <- quads
row.names(comm_mat1$spec_dat) <- quads

# simulate community with `mobsim`
com2 <- sim_thomas_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))

# Lay down sampling quadrats on the community
comm_mat2 <- sample_quadrats(com2, n_quadrats = 10, quadrat_area = 0.01,
             method = "random", avoid_overlap = T)
```

```
# Rename sampled areas as quadrats
quads <- c("site1", "site2", "site3", "site4", "site5", "site6", "site7",
           "site8", "site9", "site10")
row.names(comm_mat2$xy_dat) <- quads
row.names(comm_mat2$spec_dat) <- quads
```

1) Compare the average pairwise similarity among subsamples in site 1 (random spatial distribution) to the average pairwise similarity among subsamples in site 2 (patchy spatial distribution). Use a t-test to determine whether compositional similarity was affected by the spatial distribution. Finally, compare the compositional similarity of site 1 and site 2 to the source community?

```
#Distance Matrix for random distributions
#Importing site-by-species matrix from list
randdots <- comm_mat1$spec_dat
```

```r
#Removing all Columns with a total of 0 total observations
randdots <- randdots[,colSums(randdots !=0) > 0]

#Constructing a resemblance matrix based on Bray-Curtis Distance
randdots <- vegdist(randdots, method = "bray")
print(randdots)
```

```
##            site1     site2     site3     site4     site5     site6     site7
## site2  0.6842105
## site3  0.7333333 0.6666667
## site4  0.7647059 0.8000000 0.6250000
## site5  0.8888889 0.7142857 0.7647059 0.7894737
## site6  0.6000000 0.6521739 0.6842105 0.6190476 0.5454545
## site7  0.6842105 0.4545455 0.7777778 0.8000000 0.6190476 0.6521739
## site8  0.8888889 0.7142857 0.7647059 0.7894737 0.4000000 0.5454545 0.5238095
## site9  0.7894737 0.6363636 0.5555556 0.7000000 0.5238095 0.3913043 0.5454545
## site10 0.7272727 0.7600000 0.8095238 0.6521739 0.5833333 0.4615385 0.6000000
##            site8     site9
## site2
## site3
## site4
## site5
## site6
## site7
## site8
## site9  0.3333333
## site10 0.5000000 0.5200000
```

```r
#Distance Matrix for patchy distributions
patchdots <- comm_mat2$spec_dat

#Removing all rows with a total of 0 total observations
patchdots <- patchdots[,colSums(patchdots !=0) > 0]

#Constructing a resemblance matrix based on Bray-Curtis Distance
patchdots <- vegdist(patchdots, method = "bray")
```

```
## Warning in vegdist(patchdots, method = "bray"): you have empty rows: their
## dissimilarities may be meaningless in method "bray"
```

```
## Warning in vegdist(patchdots, method = "bray"): missing values in results
```

```r
print(patchdots)
```

```
##            site1     site2     site3     site4     site5     site6     site7
## site2  1.0000000
## site3  1.0000000 0.8750000
## site4  1.0000000 1.0000000 1.0000000
## site5  1.0000000 1.0000000 0.8787879 1.0000000
## site6  1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## site7  1.0000000 1.0000000 1.0000000       NaN 1.0000000 1.0000000
```

```
## site8  1.0000000 1.0000000 1.0000000 1.0000000 0.8666667 1.0000000 1.0000000
## site9  1.0000000 1.0000000 1.0000000       NaN 1.0000000 1.0000000       NaN
## site10 0.6666667 0.9166667 0.9285714 1.0000000 1.0000000 1.0000000 1.0000000
##            site8     site9
## site2
## site3
## site4
## site5
## site6
## site7
## site8
## site9  1.0000000
## site10 1.0000000 1.0000000
```

```
##############################################################################################
#This section is from me not knowing that t.test could take both seperate matrices as inputs, please ac
##############################################################################################
#Fusing matrices into dataframe
#distance <- as.matrix(patchdots)
#distance <- as.data.frame(as.table(distance))
#distance2 <- as.matrix(randdots)
#distance2 <- as.data.frame(as.table(distance2))

#for (i in 1:nrow(distance)){
#  distance[i,4] <- "Patchy"
#  distance2[i,4] <- "Random"
#}

#names(distance) <- c("First Site", "Second Site", "Beta", "Distribution")
#names(distance2) <- c("First Site", "Second Site", "Beta", "Distribution")

#distances <- rbind(distance, distance2)

#t <- t.test(Distribution ~ Beta, data = distances)
#with(distances, t.test(Beta[Distribution == Patchy], Beta[Distribution == Random]))
```

```
t.test(patchdots,randdots)
```

```
##
##  Welch Two Sample t-test
##
## data:  patchdots and randdots
## t = 15.104, df = 62.984, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2860455 0.3732753
## sample estimates:
## mean of x mean of y
## 0.9793419 0.6496815
```

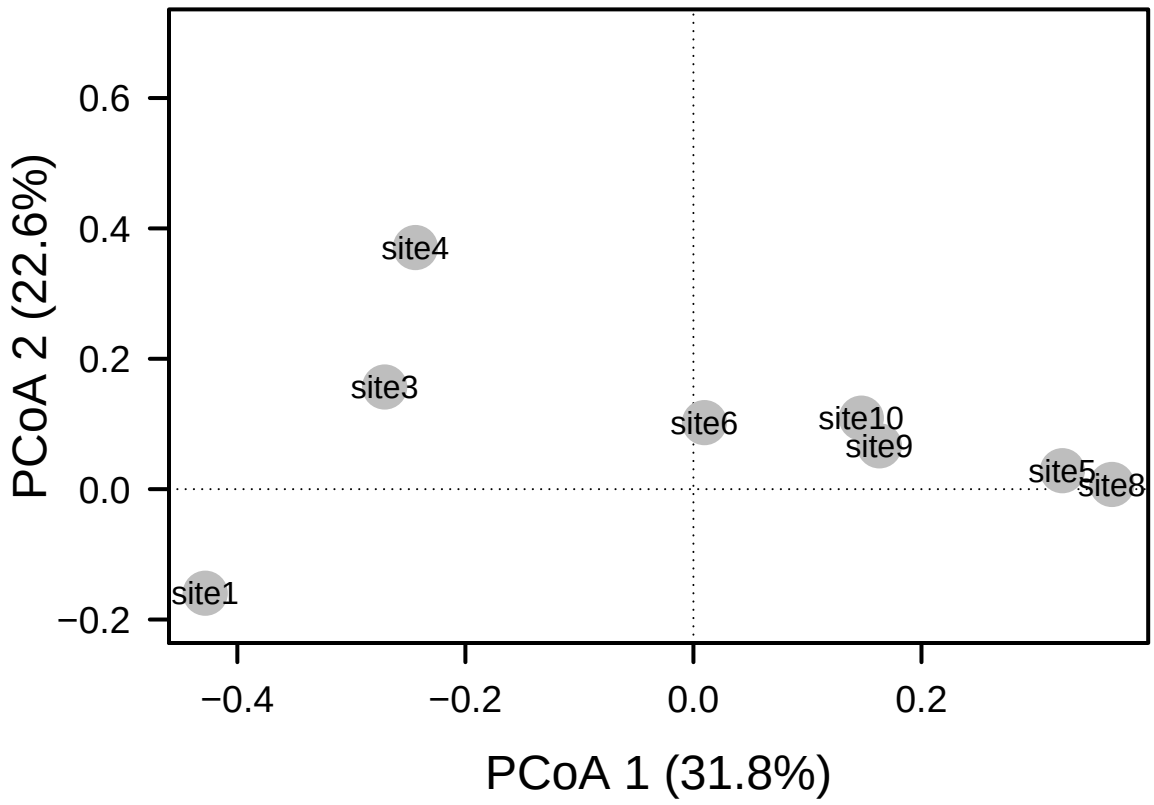
```
whole <- comm_mat$spec_dat
randdots2 <- comm_mat1$spec_dat
```

```r
#move add site 1 and 2 to dataframe with the whole area samples (for some reason site one is site 11 in
whole <- rbind(whole, randdots2[c(1,2),])

#bray-curtis measurements across these three sites
whole <- vegdist(whole, method = "bray")
print(whole)
```

```
##            site1    site11
## site11 0.9822813
## site2  0.9757174 0.6842105
```

2) Create a cluster diagram or ordination using your simulated data. Are there any visual trends that would suggest a difference in composition between site 1 and site 2? Describe.

```r
randdots.pcoa <- cmdscale(randdots, eig = TRUE, k = 3)

explainvar1 <- round(randdots.pcoa$eig[1]/sum(randdots.pcoa$eig), 3) * 100
explainvar2 <- round(randdots.pcoa$eig[2]/sum(randdots.pcoa$eig), 3) * 100
explainvar3 <- round(randdots.pcoa$eig[3]/sum(randdots.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)

#Define Plot Parameters
par(mar = c(5,5,1,2), .1)

#Initiate Plot
plot(randdots.pcoa$points[,1], randdots.pcoa$points[,2], ylim = c(-.2, .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)

#Add axis
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)

#Add Points & Labels
points(randdots.pcoa$points[,1], randdots.pcoa$points[,2],
       pch = 19, cex = 3, bg = "gray", col = "gray")
text(randdots.pcoa$points[,1], randdots.pcoa$points[,2],
     labels = row.names(randdots.pcoa$points))
```



Are there any visual trends that would suggest a difference in composition between site 1 and site 2? No, they seem relatively close together in PCoA space, but since they're random each time the script is run, they are sometimes very far apart.

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19042)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] tcltk     stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] mobsim_0.2.0         indicspecies_1.7.9   BiodiversityR_2.12-3
```

```
##  [4] gplots_3.1.1         viridis_0.6.0        viridisLite_0.4.0
##  [7] ade4_1.7-16          vegan_2.5-7          lattice_0.20-41
## [10] permute_0.9-5
##
## loaded via a namespace (and not attached):
##   [1] nlme_3.1-149        bitops_1.0-6        insight_0.13.2
##   [4] RColorBrewer_1.1-2  progress_1.2.2      tools_4.0.3
##   [7] backports_1.2.1     utf8_1.2.1          R6_2.5.0
##  [10] rpart_4.1-15        KernSmooth_2.23-17  nortest_1.0-4
##  [13] Hmisc_4.5-0         DBI_1.1.1           mgcv_1.8-33
##  [16] colorspace_2.0-0    nnet_7.3-14         tidyselect_1.1.0
##  [19] gridExtra_2.3       prettyunits_1.1.1   curl_4.3
##  [22] compiler_4.0.3      htmlTable_2.1.0     sandwich_3.0-0
##  [25] tcltk2_1.2-11       effects_4.2-0       caTools_1.18.2
##  [28] scales_1.1.1        checkmate_2.0.0     proxy_0.4-25
##  [31] RcmdrMisc_2.7-1     stringr_1.4.0       digest_0.6.27
##  [34] relimp_1.0-5        minqa_1.2.4         foreign_0.8-80
##  [37] rmarkdown_2.6       rio_0.5.26          base64enc_0.1-3
##  [40] jpeg_0.1-8.1        pkgconfig_2.0.3     htmltools_0.5.1.1
##  [43] lme4_1.1-26         Rcmdr_2.7-1         highr_0.8
##  [46] htmlwidgets_1.5.3   rlang_0.4.10        readxl_1.3.1
##  [49] rstudioapi_0.13     generics_0.1.0      zoo_1.8-9
##  [52] gtools_3.8.2        dplyr_1.0.5         zip_2.1.1
##  [55] car_3.0-10          magrittr_2.0.1      Formula_1.2-4
##  [58] Matrix_1.2-18       Rcpp_1.0.6          munsell_0.5.0
##  [61] fansi_0.4.2         abind_1.4-5         lifecycle_1.0.0
##  [64] stringi_1.5.3       yaml_2.2.1          carData_3.0-4
##  [67] MASS_7.3-53         grid_4.0.3          parallel_4.0.3
##  [70] forcats_0.5.1       crayon_1.4.1        haven_2.3.1
##  [73] splines_4.0.3       hms_1.0.0           knitr_1.31
##  [76] pillar_1.5.1        boot_1.3-25         glue_1.4.2
##  [79] evaluate_0.14       mitools_2.4         latticeExtra_0.6-29
##  [82] data.table_1.14.0   nloptr_1.2.2.2      png_0.1-7
##  [85] vctrs_0.3.6         cellranger_1.1.0    gtable_0.3.0
##  [88] purrr_0.3.4         assertthat_0.2.1    ggplot2_3.3.3
##  [91] xfun_0.20           openxlsx_4.2.3      survey_4.0
##  [94] e1071_1.7-6         class_7.3-17        survival_3.2-7
##  [97] tibble_3.1.0        cluster_2.1.0       statmod_1.4.35
## [100] ellipsis_0.3.1
```