

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

Andrea Phillips; Z620: Quantitative Biodiversity, Indiana University

05 February, 2019

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. This 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 **Wednesday, February 6th, 2019 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, 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/andjr/Github2/QB2019_Phillips/2.Worksheets/8.BetaDiversity"
```

```
#setwd("/Users/andjr/Github2/QB2019_Phillips/2.Worksheets/8.BetaDiversity")
#jb <- read.table("JellyBeans.Source.txt", sep="\t", header = TRUE)
```

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

```
## This is vegan 2.5-4
```

```
##
```

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

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

```
##
```

```
##      lowess
```

```
## BiodiversityR 2.11-1: Use command BiodiversityRGUI() to launch the Graphical User Interface;
## to see changes use BiodiversityRGUI(changeLog=TRUE, backward.compatibility.messages=TRUE)
```

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, max.level = 1)
```

```
## List of 4
```

```
## $ env      : 'data.frame': 30 obs. of  11 variables:
```

```
## $ fish     : 'data.frame': 30 obs. of  27 variables:
```

```
## $ xy       : 'data.frame': 30 obs. of  2 variables:
```

```
## $ species : 'data.frame': 27 obs. of  4 variables:
```

```
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
```

```
head(doubs$fish)
```

```
##   Cogo Satr Phph Neba Thth Teso Chna Chto Lele Lece Baba Spbi Gogo Eslu
## 1   0   3   0   0   0   0   0   0   0   0   0   0   0   0
## 2   0   5   4   3   0   0   0   0   0   0   0   0   0   0
## 3   0   5   5   5   0   0   0   0   0   0   0   0   0   1
## 4   0   4   5   5   0   0   0   0   0   1   0   0   1   2
## 5   0   2   3   2   0   0   0   0   5   2   0   0   2   4
## 6   0   3   4   5   0   0   0   0   1   2   0   0   1   1
##   Pefl Rham Legi Scer Cyca Titi Abbr Icme Acce Ruru Blbj Alal Anan
## 1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2   0   0   0   0   0   0   0   0   0   0   0   0   0
## 3   0   0   0   0   0   0   0   0   0   0   0   0   0
## 4   2   0   0   0   0   1   0   0   0   0   0   0   0
## 5   4   0   0   2   0   3   0   0   0   5   0   0   0
## 6   1   0   0   0   0   2   0   0   0   1   0   0   0
```

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

- How many objects are in `doubs`?
- How many fish species are there in the `doubs` dataset?
- How many sites are in the `doubs` dataset?

Answer 1a: There are four lists in `doubs` (assuming that's what you mean by objects) **Answer**

1b: There are 27 fish species in the data set. **Answer 1c:** There are 30 sites in the data set.

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.

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

Answer 2a: The species richness is higher downstream, between 20-30 species, than upstream, which is between 0 and 11 species. The species richness between downstream and upstream varies, but most samples are similar to the downstream species richness, between 20-30. **An-**

swer 2b: Brown trout has little to no abundance downstream, showing up once. Upstream the

abundance is higher, up to 5. It's not until the upper half of the stream that Brown Trout shows up at all, ranging from 0 to 5 by sample. **Answer 2c:** While the fish richness shows downstream being very rich in species, it had no brown trout. Higher richness does not necessarily mean higher biodiversity. In this case upstream is probably a healthier ecosystem (or at least a healthier ecosystem for brown trout) than downstream, even though downstream has a high species richness.

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 = ""){
  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 <- 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)
  }

  #elseif
  # SbyS.pa <- decostand(site.by.species, method = "pa")
  # S <- ncol(SbyS.pa[,which(colSums(SbyS.pa) >0)])
  # a.bar <- mean(specnumber(SyS.pa))
  # b.w <- round(S/a.bar, 3)
  #return(b.w)
  #}

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

## [1] 0.5
```

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

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

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

- Describe how local richness (α) and turnover (β) contribute to regional (γ) fish diversity in the Doubs.
- Is the fish assemblage at site 1 more similar to the one at site 2 or site 10?
- 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: In general, beta diversity between sites depends heavily on the alpha diversity of those sites. As the alpha diversity of a fish site changes, that is going to in turn affect the beta diversity between sites. Looking at turnover, this is what is going to affect the gamma diversity of the whole landscape. As single sites change, we see changes between sites, which means changes on the ecological landscape level. This makes sense, because in this case gamma is going to be the product of alpha and beta diversity. **Answer 3b:** The fish assemblage at site 1 is more similar to site 10, with a `beta.w` of 0.714, than site 2 with a `beta.w` of 0.5. I believe this is interpreted as saying that site 1 and site 10 share 71.4% of their diversity, and site 1 and site 2 share 50% of their diversity. (Please correct me if I'm wrong, these indexes are way confusing to me :)) **Answer 3c:** This would change my interpretation to say that rather than meaning a proportion of (dis)similarity between sites, I would interpret beta to be an absolute value of diversity difference between the two sites, and requires knowledge of all of the sites to interpret.

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 metrics are heavily influenced by differences between sites/samples. In other words, if the sites are dissimilar because some species are not shared, the incidence-based metrics are going to be affected. However, if the rare species shows up in both sites, we may not notice the effect of rare species. Abundance-based metrics, however, consider each species individually. I believe this would make the abundance-based metrics more likely to be affected by rare species.

In the R code chunk below, do the following:

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

```
fish <- doubs$fish
fish <- fish[-8, ]

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

fish.db <- vegdist(fish, method = "bray", diag = TRUE)
```

```
fish.ds <- vegdist(fish, method = "bray", binary = TRUE)
```

```
fish.ds
```

```
##           1           2           3           4           5           6
## 2  0.50000000
## 3  0.60000000 0.14285714
## 4  0.77777778 0.45454545 0.33333333
## 5  0.83333333 0.57142857 0.46666667 0.15789474
## 6  0.81818182 0.53846154 0.42857143 0.11111111 0.04761905
## 7  0.66666667 0.25000000 0.33333333 0.38461538 0.37500000 0.33333333
## 9  1.00000000 0.50000000 0.55555556 0.38461538 0.37500000 0.33333333
## 10 0.71428571 0.33333333 0.40000000 0.28571429 0.29411765 0.25000000
## 11 0.71428571 0.33333333 0.40000000 0.42857143 0.52941176 0.50000000
## 12 0.71428571 0.33333333 0.40000000 0.42857143 0.52941176 0.50000000
## 13 0.71428571 0.33333333 0.40000000 0.57142857 0.64705882 0.62500000
## 14 0.81818182 0.53846154 0.42857143 0.33333333 0.42857143 0.40000000
## 15 0.83333333 0.57142857 0.60000000 0.36842105 0.36363636 0.33333333
## 16 0.88888889 0.70000000 0.61904762 0.36000000 0.28571429 0.25925926
## 17 0.91304348 0.76000000 0.69230769 0.46666667 0.39393939 0.37500000
## 18 0.91666667 0.76923077 0.70370370 0.48387097 0.41176471 0.39393939
## 19 1.00000000 0.84615385 0.77777778 0.54838710 0.41176471 0.45454545
## 20 1.00000000 0.84000000 0.76923077 0.53333333 0.39393939 0.43750000
## 21 1.00000000 0.84615385 0.77777778 0.54838710 0.41176471 0.45454545
## 22 1.00000000 0.92000000 0.84615385 0.60000000 0.45454545 0.50000000
## 23 1.00000000 1.00000000 1.00000000 0.81818182 0.71428571 0.69230769
## 24 1.00000000 1.00000000 1.00000000 0.75000000 0.68421053 0.66666667
## 25 1.00000000 1.00000000 0.83333333 0.62500000 0.36842105 0.44444444
## 26 1.00000000 0.91666667 0.84000000 0.58620690 0.43750000 0.48387097
## 27 1.00000000 0.92000000 0.84615385 0.60000000 0.45454545 0.50000000
## 28 1.00000000 0.92000000 0.84615385 0.60000000 0.45454545 0.50000000
## 29 0.92592593 0.79310345 0.73333333 0.52941176 0.40540541 0.44444444
## 30 1.00000000 1.00000000 0.92000000 0.65517241 0.50000000 0.54838710
##           7           9          10          11          12          13
## 2
## 3
## 4
## 5
## 6
## 7
## 9  0.40000000
## 10 0.09090909 0.45454545
## 11 0.27272727 0.45454545 0.33333333
## 12 0.27272727 0.45454545 0.33333333 0.00000000
## 13 0.45454545 0.63636364 0.50000000 0.16666667 0.16666667
## 14 0.46666667 0.60000000 0.37500000 0.25000000 0.25000000 0.25000000
## 15 0.37500000 0.50000000 0.29411765 0.29411765 0.29411765 0.29411765
## 16 0.54545455 0.54545455 0.47826087 0.56521739 0.56521739 0.56521739
## 17 0.62962963 0.62962963 0.57142857 0.57142857 0.57142857 0.57142857
## 18 0.64285714 0.64285714 0.58620690 0.58620690 0.58620690 0.58620690
## 19 0.71428571 0.64285714 0.65517241 0.79310345 0.79310345 0.79310345
## 20 0.70370370 0.62962963 0.64285714 0.78571429 0.78571429 0.85714286
## 21 0.71428571 0.64285714 0.65517241 0.79310345 0.79310345 0.86206897
```

```

## 22 0.77777778 0.70370370 0.71428571 0.85714286 0.85714286 0.92857143
## 23 0.75000000 0.50000000 0.77777778 0.77777778 0.77777778 1.00000000
## 24 0.84615385 0.69230769 0.71428571 0.85714286 0.85714286 1.00000000
## 25 0.69230769 0.69230769 0.57142857 0.85714286 0.85714286 1.00000000
## 26 0.76923077 0.69230769 0.70370370 0.85185185 0.85185185 0.92592593
## 27 0.77777778 0.70370370 0.71428571 0.85714286 0.85714286 0.92857143
## 28 0.77777778 0.70370370 0.71428571 0.85714286 0.85714286 0.92857143
## 29 0.67741935 0.67741935 0.62500000 0.68750000 0.68750000 0.68750000
## 30 0.84615385 0.76923077 0.77777778 0.92592593 0.92592593 1.00000000
##          14          15          16          17          18          19
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15 0.14285714
## 16 0.33333333 0.28571429
## 17 0.37500000 0.33333333 0.12820513
## 18 0.39393939 0.35294118 0.15000000 0.02222222
## 19 0.57575758 0.52941176 0.25000000 0.15555556 0.13043478
## 20 0.62500000 0.57575758 0.28205128 0.18181818 0.15555556 0.02222222
## 21 0.63636364 0.58823529 0.30000000 0.20000000 0.17391304 0.04347826
## 22 0.68750000 0.63636364 0.33333333 0.22727273 0.20000000 0.06666667
## 23 0.84615385 0.85714286 0.80000000 0.76000000 0.76923077 0.76923077
## 24 0.77777778 0.78947368 0.68000000 0.60000000 0.54838710 0.48387097
## 25 0.66666667 0.68421053 0.60000000 0.60000000 0.54838710 0.48387097
## 26 0.67741935 0.62500000 0.36842105 0.25581395 0.22727273 0.09090909
## 27 0.68750000 0.63636364 0.33333333 0.22727273 0.20000000 0.06666667
## 28 0.68750000 0.63636364 0.33333333 0.22727273 0.20000000 0.06666667
## 29 0.50000000 0.45945946 0.25581395 0.12500000 0.10204082 0.06122449
## 30 0.74193548 0.68750000 0.36842105 0.25581395 0.22727273 0.09090909
##          20          21          22          23          24          25
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17

```

```

## 18
## 19
## 20
## 21 0.02222222
## 22 0.04545455 0.02222222
## 23 0.76000000 0.76923077 0.76000000
## 24 0.46666667 0.48387097 0.46666667 0.45454545
## 25 0.46666667 0.48387097 0.46666667 0.45454545 0.37500000
## 26 0.06976744 0.04545455 0.02325581 0.75000000 0.44827586 0.44827586
## 27 0.04545455 0.02222222 0.00000000 0.76000000 0.46666667 0.46666667
## 28 0.04545455 0.02222222 0.00000000 0.76000000 0.46666667 0.46666667
## 29 0.08333333 0.06122449 0.08333333 0.79310345 0.52941176 0.52941176
## 30 0.06976744 0.04545455 0.02325581 0.75000000 0.44827586 0.44827586
##          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
## 25
## 26
## 27 0.02325581
## 28 0.02325581 0.00000000
## 29 0.10638298 0.08333333 0.08333333
## 30 0.04761905 0.02325581 0.02325581 0.10638298

```

fish.db

```

##          1          2          3          4          5          6
## 1 0.00000000
## 2 0.60000000 0.00000000
## 3 0.68421053 0.14285714 0.00000000
## 4 0.75000000 0.33333333 0.18918919 0.00000000
## 5 0.89189189 0.69565217 0.68000000 0.49090909 0.00000000
## 6 0.75000000 0.39393939 0.29729730 0.19047619 0.41818182 0.00000000
## 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

```



```

## 10 0.88235294 0.38461538 0.40000000 0.37142857 0.54166667 0.25714286
## 11 0.57142857 0.30434783 0.40740741 0.43750000 0.68888889 0.43750000
## 12 0.71428571 0.20000000 0.23529412 0.33333333 0.69230769 0.38461538
## 13 0.72727273 0.29032258 0.31428571 0.45000000 0.73584906 0.55000000
## 14 0.80645161 0.40000000 0.31818182 0.34693878 0.67741935 0.42857143
## 15 0.83333333 0.51111111 0.46938776 0.40740741 0.55223881 0.37037037
## 16 0.86046512 0.65384615 0.57142857 0.47540984 0.45945946 0.37704918
## 17 0.91489362 0.67857143 0.63333333 0.50769231 0.51282051 0.44615385
## 18 0.95555556 0.74074074 0.72413793 0.58730159 0.50000000 0.52380952
## 19 1.00000000 0.79310345 0.70967742 0.61194030 0.50000000 0.52238806
## 20 1.00000000 0.91176471 0.88888889 0.74025974 0.48888889 0.68831169
## 21 1.00000000 0.94594595 0.92307692 0.78313253 0.50000000 0.73493976
## 22 1.00000000 0.97619048 0.95454545 0.82795699 0.52830189 0.78494624
## 23 1.00000000 1.00000000 1.00000000 0.92000000 0.89473684 0.84000000
## 24 1.00000000 1.00000000 1.00000000 0.88888889 0.79591837 0.77777778
## 25 1.00000000 1.00000000 0.92592593 0.81250000 0.68888889 0.68750000
## 26 1.00000000 0.96363636 0.93220339 0.78125000 0.55844156 0.68750000
## 27 1.00000000 0.97333333 0.94936709 0.83333333 0.56701031 0.76190476
## 28 1.00000000 0.97560976 0.95348837 0.82417582 0.57692308 0.78021978
## 29 0.97777778 0.93939394 0.92233010 0.81481481 0.53719008 0.77777778
## 30 1.00000000 1.00000000 0.98095238 0.87272727 0.59349593 0.83636364
##          7          9          10          11          12          13
## 1
## 2
## 3
## 4
## 5
## 6
## 7 0.00000000
## 9 0.66666667 0.00000000
## 10 0.26666667 0.57142857 0.00000000
## 11 0.33333333 0.76000000 0.44000000 0.00000000
## 12 0.17647059 0.68750000 0.37500000 0.24137931 0.00000000
## 13 0.37142857 0.81818182 0.57575758 0.33333333 0.18918919 0.00000000
## 14 0.36363636 0.76190476 0.47619048 0.43589744 0.21739130 0.19148936
## 15 0.38775510 0.65957447 0.40425532 0.50000000 0.33333333 0.38461538
## 16 0.53571429 0.70370370 0.51851852 0.64705882 0.55172414 0.59322034
## 17 0.60000000 0.68965517 0.51724138 0.63636364 0.58064516 0.61904762
## 18 0.68965517 0.64285714 0.57142857 0.69811321 0.66666667 0.70491803
## 19 0.67741935 0.66666667 0.63333333 0.82456140 0.75000000 0.81538462
## 20 0.86111111 0.68571429 0.77142857 0.91044776 0.89189189 0.92000000
## 21 0.89743590 0.76315789 0.81578947 0.91780822 0.92500000 0.95061728
## 22 0.93181818 0.76744186 0.86046512 0.95180723 0.95555556 0.97802198
## 23 0.90000000 0.77777778 0.88888889 0.86666667 0.90909091 1.00000000
## 24 0.93548387 0.72413793 0.79310345 0.92307692 0.93939394 1.00000000
## 25 0.85185185 0.84000000 0.76000000 0.90909091 0.93103448 1.00000000
## 26 0.89830508 0.71929825 0.82456140 0.92592593 0.93442623 0.96774194
## 27 0.92405063 0.76623377 0.84415584 0.94594595 0.95061728 0.97560976
## 28 0.93023256 0.76190476 0.85714286 0.95061728 0.95454545 0.97752809
## 29 0.90291262 0.78217822 0.84158416 0.89795918 0.90476190 0.90566038
## 30 0.96190476 0.84466019 0.90291262 0.98000000 0.98130841 1.00000000
##          14          15          16          17          18          19
## 1
## 2

```

```

## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14 0.00000000
## 15 0.24590164 0.00000000
## 16 0.44117647 0.26027397 0.00000000
## 17 0.50000000 0.40259740 0.26190476 0.00000000
## 18 0.60000000 0.46666667 0.34146341 0.13953488 0.00000000
## 19 0.67567568 0.56962025 0.39534884 0.31111111 0.25000000 0.00000000
## 20 0.83333333 0.70786517 0.58333333 0.42000000 0.32653061 0.23529412
## 21 0.86666667 0.76842105 0.62745098 0.49056604 0.40384615 0.29629630
## 22 0.90000000 0.77142857 0.66071429 0.55172414 0.47368421 0.38983051
## 23 0.93750000 0.94594595 0.90909091 0.83333333 0.82608696 0.84000000
## 24 0.90697674 0.87500000 0.81818182 0.69491525 0.64912281 0.63934426
## 25 0.84615385 0.81818182 0.76470588 0.74545455 0.66037736 0.61403509
## 26 0.85915493 0.76315789 0.63855422 0.54022989 0.45882353 0.32584270
## 27 0.89010989 0.77083333 0.66990291 0.57009346 0.48571429 0.37614679
## 28 0.89795918 0.78640777 0.69090909 0.57894737 0.50000000 0.41379310
## 29 0.84347826 0.73333333 0.65354331 0.51145038 0.44186047 0.41353383
## 30 0.93162393 0.81967213 0.72093023 0.57894737 0.52671756 0.48148148
##          20          21          22          23          24          25
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20 0.00000000
## 21 0.10169492 0.00000000
## 22 0.18750000 0.10447761 0.00000000
## 23 0.86666667 0.87878788 0.89473684 0.00000000
## 24 0.57746479 0.61038961 0.65517241 0.57894737 0.00000000
## 25 0.67164179 0.69863014 0.73493976 0.46666667 0.46153846 0.00000000
## 26 0.21212121 0.20000000 0.25217391 0.82978723 0.48275862 0.59259259
## 27 0.19327731 0.13600000 0.12592593 0.88059701 0.61538462 0.70270270

```

```
## 28 0.22222222 0.16666667 0.12676056 0.89189189 0.64705882 0.72839506
## 29 0.24475524 0.18120805 0.11949686 0.91208791 0.70588235 0.77551020
## 30 0.29655172 0.23178808 0.18012422 0.91397849 0.71153846 0.78000000
##      26      27      28      29      30
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26 0.00000000
## 27 0.18867925 0.00000000
## 28 0.23893805 0.09774436 0.00000000
## 29 0.33846154 0.18666667 0.14649682 0.00000000
## 30 0.36363636 0.19736842 0.15723270 0.14772727 0.00000000
```

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

- Does the resemblance matrix (`fish.db`) represent similarity or dissimilarity? What information in the resemblance matrix led you to arrive at your answer?
- 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: It represents dissimilarity because the diagonal is 0. This means that when a site is compared to itself, it is 100% similar or 0% dissimilar.

Answer 5b: Sørensen places greater emphasis on unshared taxa, so the site may be shown as more dissimilar based on unshared taxa than Bray-Curtis shows.

4) VISUALIZING BETA-DIVERSITY

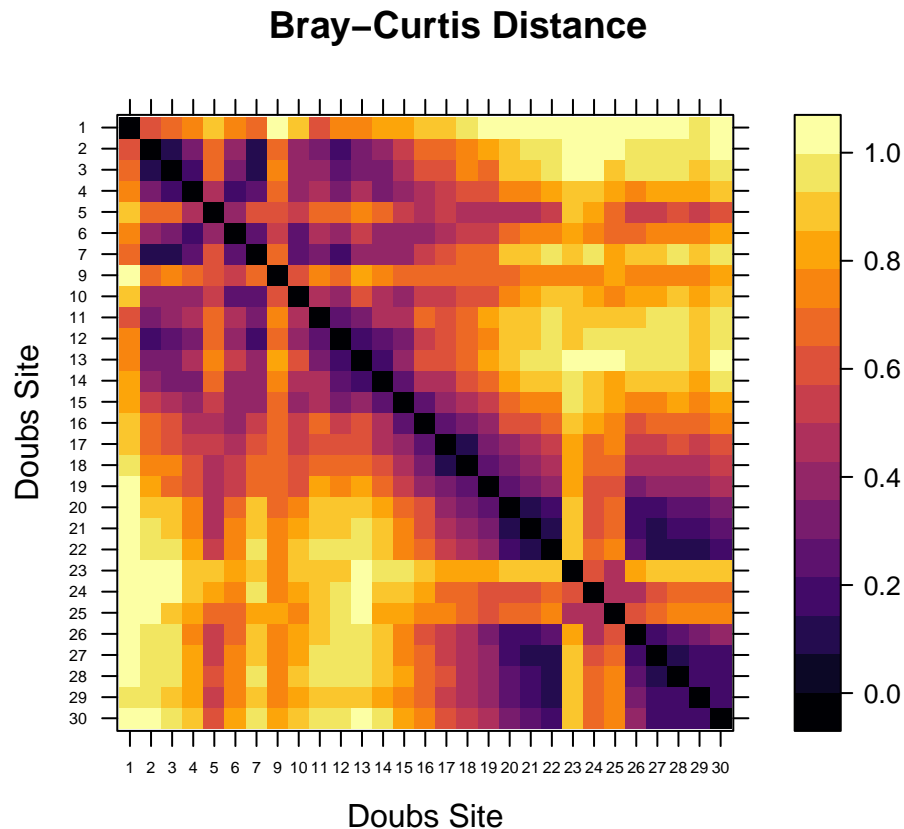
A. Heatmaps

In the R code chunk below, do the following:

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

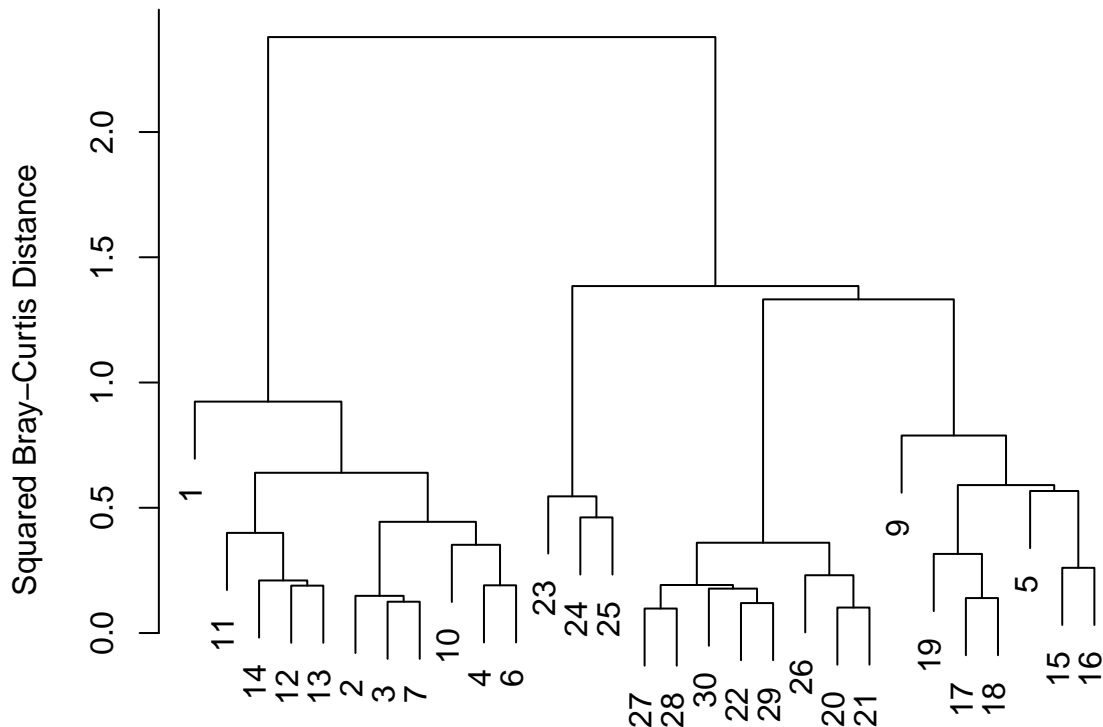
In the R code chunk below, do the following:

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

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

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

Doubs River Fish: Ward's Clustering



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: I believe from this data we can conclude that there are a few pockets of environmentally liveable areas for fish species. These sites seem to have different species compositions, but not entirely different species compositions, so it is possible that there is little gene flow between the different liveable sites.

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)
```

```

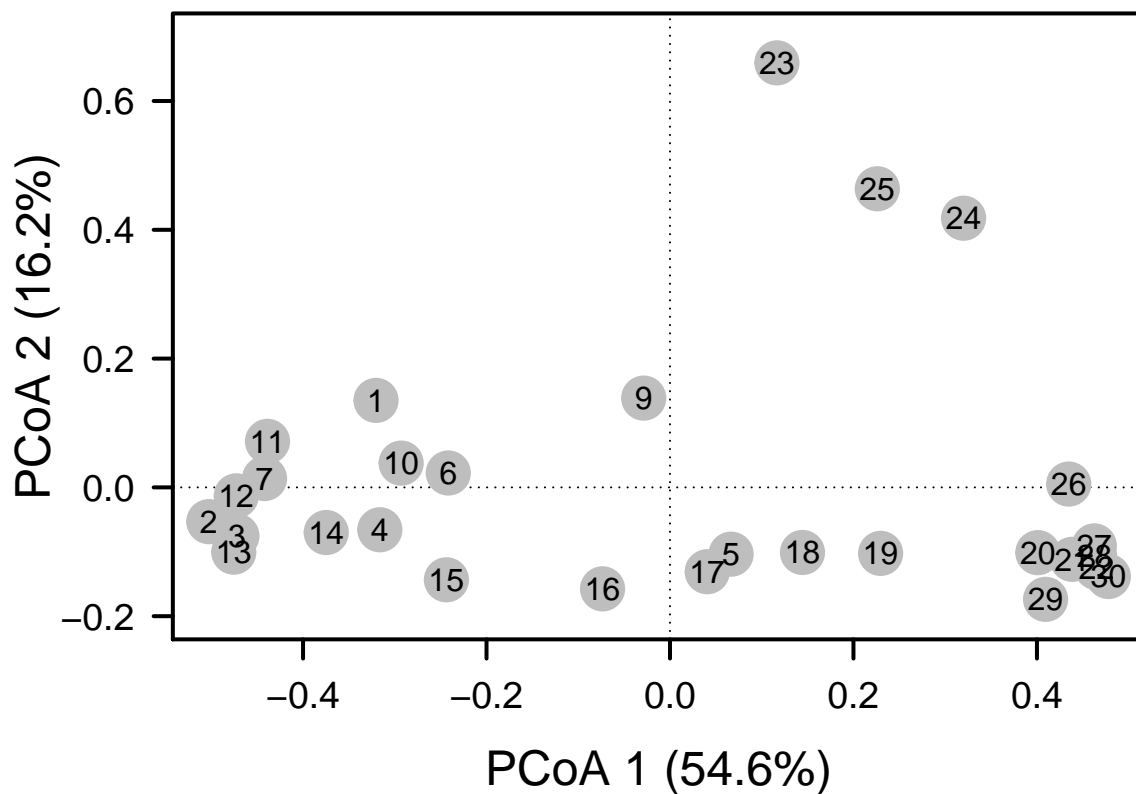
par(mar = c(5, 5, 1, 2) + 0.1)

plot(fish.pcoa$points[,1], fish.pcoa$points[,2], ylim = c(-0.2, 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(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))

```



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.

```

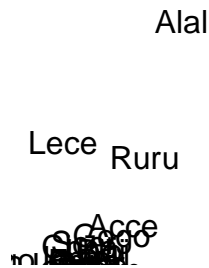
fishREL <- fish
for(i in 1:nrow(fish)){
  fishREL[i, ] = fish[i, ] / sum(fish[i, ])
}

```

```

}
plot.new()
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")

```



```

spe.corr <- add.spec.scores(fish.pcoa, fishREL, method = "cor.scores")$cproj
corrcut <- 0.7
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:

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

Answer 7a: We essentially have two main groupings of sites, one of high quality and one of low quality. It also appears that is a sort of gradient of change in quality between sites on a continuum between the two main groupings. **Answer 7b:** It appears that `Satr` and `Alal` are likely quality indicators, and to a lesser extent `Phpn`, `Neba`, `Lece`, and `Ruru`. The presence of `Satr`, to provide an example, demarcates one of the groupings, and the other grouping is lacking `Satr`, which can be interpreted to mean that higher quality sites are more likely to contain `Satr` (assuming the presence of a species means higher quality and not lower quality.)

SYNTHESIS

Using the jelly bean data from class (i.e., JellyBeans.Source.txt and JellyBeans.txt):

- 1) Compare the average pairwise similarity among subsamples in group A to the average pairwise similarity among subsamples in group B. Use a t-test to determine whether compositional similarity was affected by the “vicariance” event. Finally, compare the compositional similarity of jelly beans in group A and group B to the source community?
- 2) Create a cluster diagram or ordination using the jelly bean data. Are there any visual trends that would suggest a difference in composition between group A and group B?

```
getwd()
```

```
## [1] "C:/Users/andjr/Github2/QB2019_Phillips/2.Worksheets/8.BetaDiversity"
```

```
jellybean <- read.table("JellyBeans.txt", sep = "\t", header = TRUE)
jellybeansource <- read.table("JellyBeans.Source.txt", sep = "\t", header = TRUE)

#Below I added sample 8 to group B so the vectors
#are the same length. Arbitrary and not idea, but I
#couldn't figure out how to make it work with vectors
#of different lengths :)

jellybeanA <- c(1, 2, 3, 6, 8)
jellybeanB <- c(4, 5, 7, 8, 9)
jellybeanA <- jellybean[jellybeanA, 3:30]
jellybeanB <- jellybean[jellybeanB, 3:30]

beta.jb <- 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.jb = round(s/a.bar - 1, 3)
    return(b.jb)
  }

  #I don't know how to compare all the subsamples
#pairwise without writing a million lines of code.
#We haven't gone over doing t-tests in R, and there's
#no explanation of what the vicariance event is anywhere...

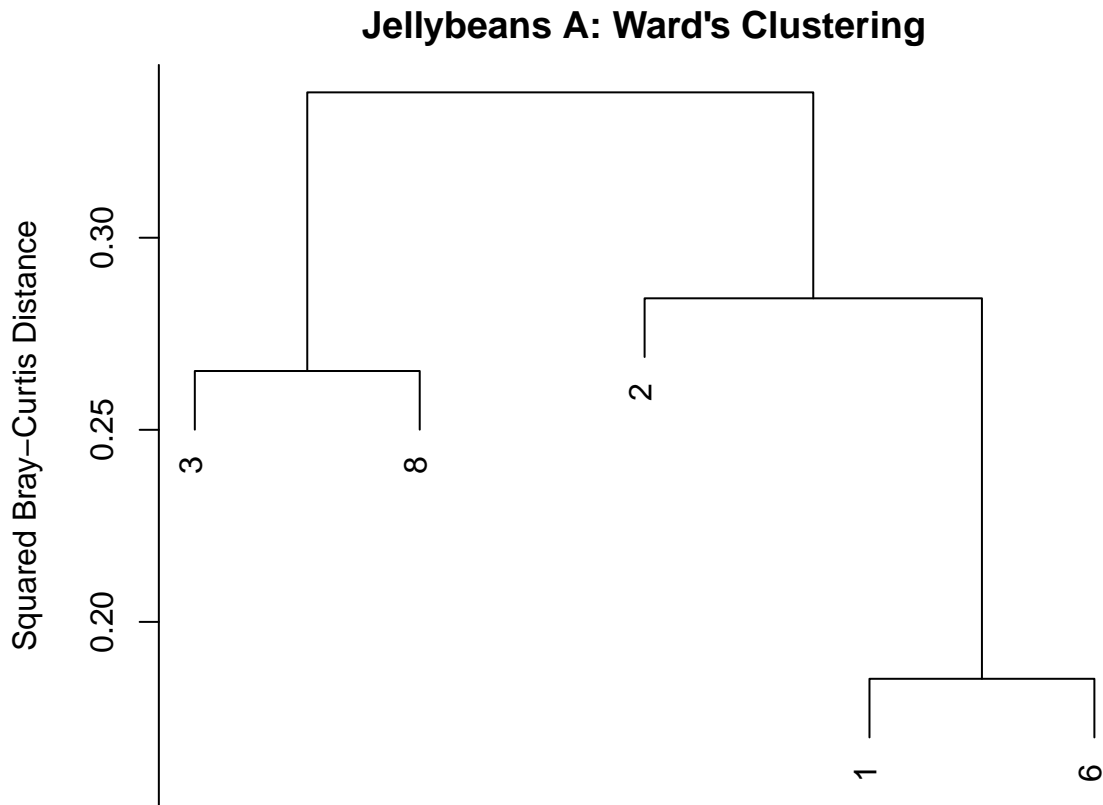
jellybean = jellybean[1:9, 3:30]

jelly.db <- vegdist(jellybeanA, method = "bray", upper = TRUE, diag = TRUE)
```



```
jelly.ward <- hclust(jelly.db, method = "ward.D2")

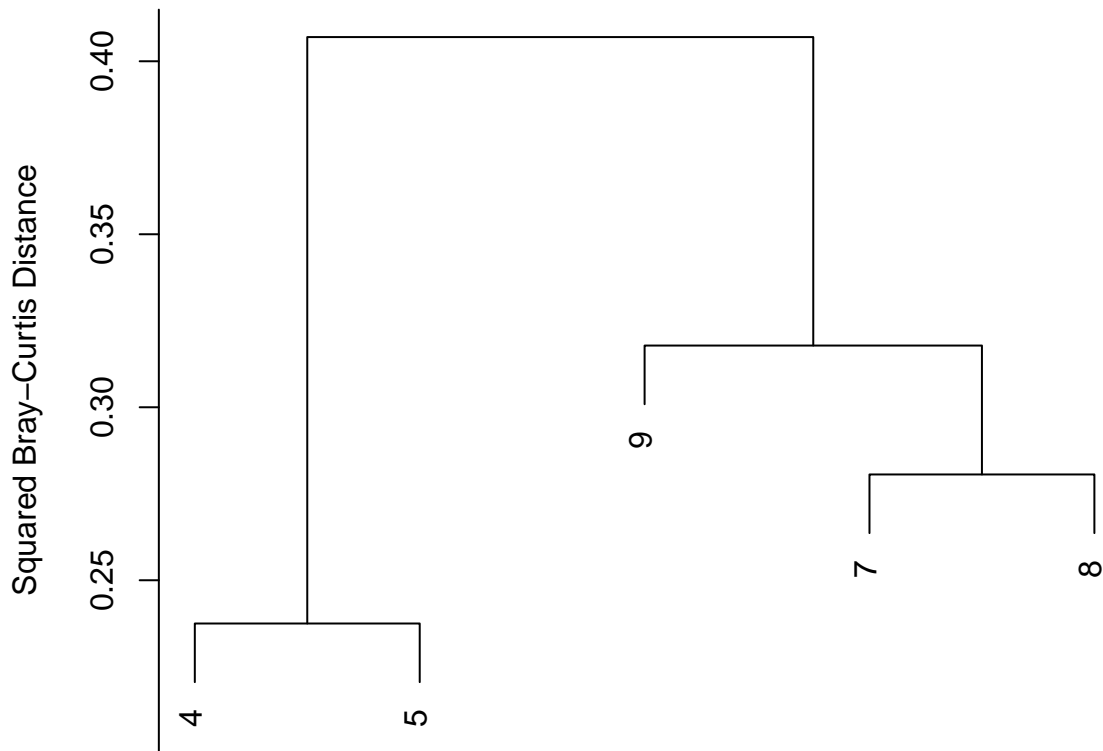
par(mar = c(1, 5, 2, 2) + 0.1)
plot(jelly.ward, main = "Jellybeans A: Ward's Clustering",
     ylab = "Squared Bray-Curtis Distance")
```



```
jelly.db <- vegdist(jellybeanB, method = "bray", upper = TRUE, diag = TRUE)
jelly.ward <- hclust(jelly.db, method = "ward.D2")

par(mar = c(1, 5, 2, 2) + 0.1)
plot(jelly.ward, main = "Jellybeans B: Ward's Clustering",
     ylab = "Squared Bray-Curtis Distance")
```

Jellybeans B: Ward's Clustering



*#Yes, it appears groups A and B vary in composition qualitatively--
#the Bray-Curtis distance shows different patterns for A and B.
#In group A, the groups are closer together in general, note the change
#in the axis. In group B, samples 7, 8, and 9 are fairly similar, varying
#a bit from 4 and 5. In sample A, 1 and 6 are a bit different from group 2,
#which collectively vary from samples 3 and 8, which are very similar to each other.*