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

Lana Bolin; 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. 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 **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] "/Users/lana/GitHub/QB2019_Bolin/2.Worksheets/8.BetaDiversity"
```

```
setwd("~/GitHub/QB2019_Bolin/2.Worksheets/8.BetaDiversity/")
require("vegan")
```

```
## Loading required package: vegan
```

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

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

```
## This is vegan 2.5-3
```

```
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)
  }
}
```

```
##
## 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)
```

```
## 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 ...
```

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: 4
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: There is higher richess in the downstream region, as well as in the little crook midriver.
Answer 2b: Brown Trout abundance decreases as you head downstream.
Answer 2c: Richness tells us little about specific taxa that we may be interested in. In this example, our focal species is most abundant in areas with overall low 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, 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)
  }
}

# pairwise between sites 1 and 2
beta.w(doubs$fish, 1, 2, TRUE)
```

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

```
# pairwise between sites 1 and 10
beta.w(doubs$fish, 1, 10, 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: (α) and (β) diversity contribute to (γ) diversity in a multiplicative way - as either increases, so does (γ) diversity.
Answer 3b: The fish assemblage at site 1 is more similar to the one at site 2 because there is lower pairwise beta diversity.
Answer 3c: Instead of being a measure of how many times more diverse the regional pool is relative to the mean local diversity, it would instead be a measure of how many more species are found in the regional pool than in the average local pool.

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 treat rare species as equivalent to common species, while abundance-based metrics do not.

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, ]

fish.ds <- vegdist(fish, method = "bray", binary = TRUE)
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: Dissimilarity, because sites further away from each other tend to have values at or closer to 1, while sites closer to each other tend to hvae values loser to 0.
Answer 5b: Qualitatively the matrices look pretty similar - there doesn't seem to be a constitutive bias between them (i.e. sometimes dissimilarity between two species is higher using Bray-Curtis, and sometimes it's lower). Sørensen distance is based on incidence data, so it relies on comparing shared and unique species between sites. Bray-Curtis distance, on the other hand, uses abundance data, so it takes into account the relative abundances of different species. It's important to keep these differences in mind when interpreting data; for example, two sites may have high dissimilarity using Bray-Curtis distance but low dissimilarity using Sørensen if the sites have the same species, but they occur in very different abundances.

4) VISUALIZING BETA-DIVERSITY

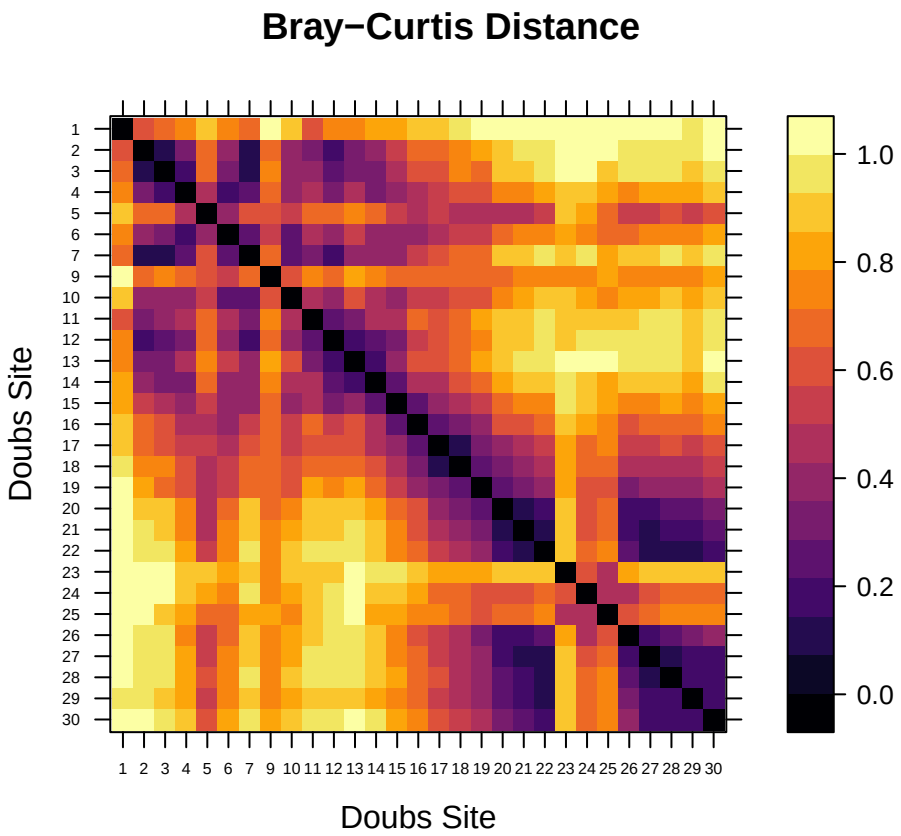
A. Heatmaps

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 <- 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")
```



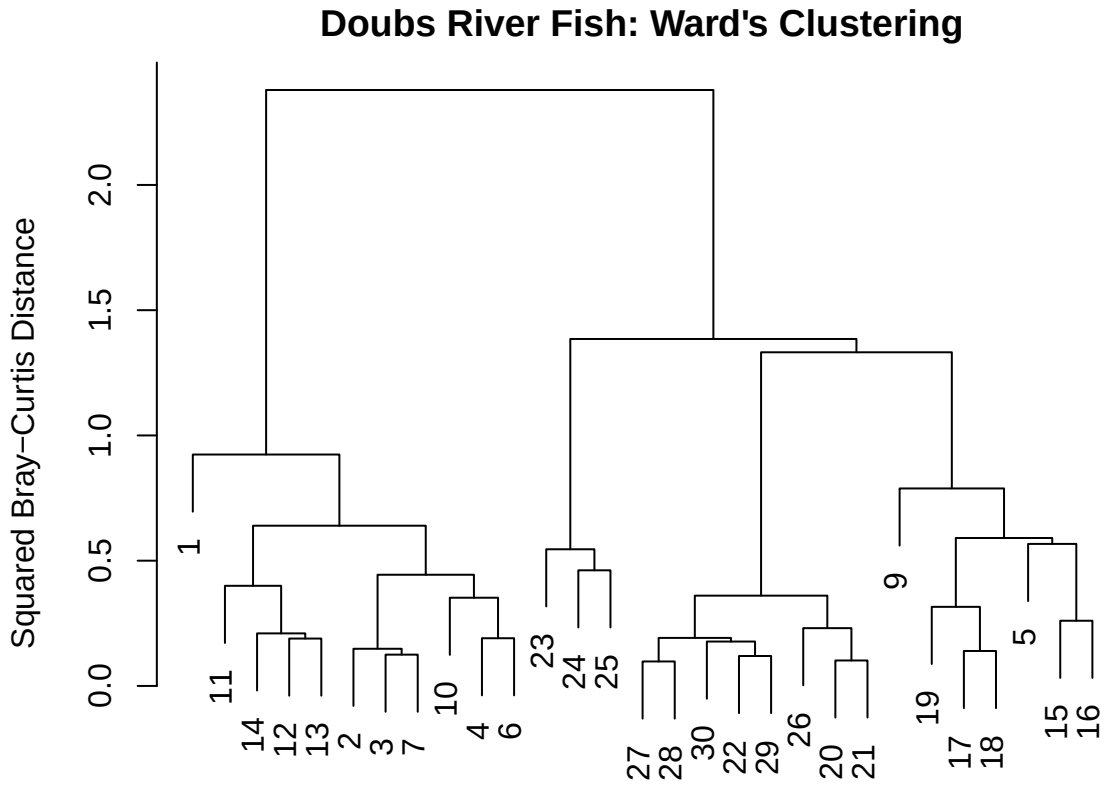
B. Cluster Analysis

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`).

```r
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")
```



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***: We saw earlier that fish richness appears to negatively correlate with brown trout abundance, with the upstream tending to have low fish richness and high brown trout abuncance, and the downstream tending to have high fish richness and low brown trout abuncance. Our cluster analysis shows that upstream sites generally cluster together, as do downstream sites. One hypothesis is that these patterns are seen because brown trout is a keystone species whose presence/absence has large effects on the community composition.

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)

# Percen variation explained by the first three axes
sum.eig
```

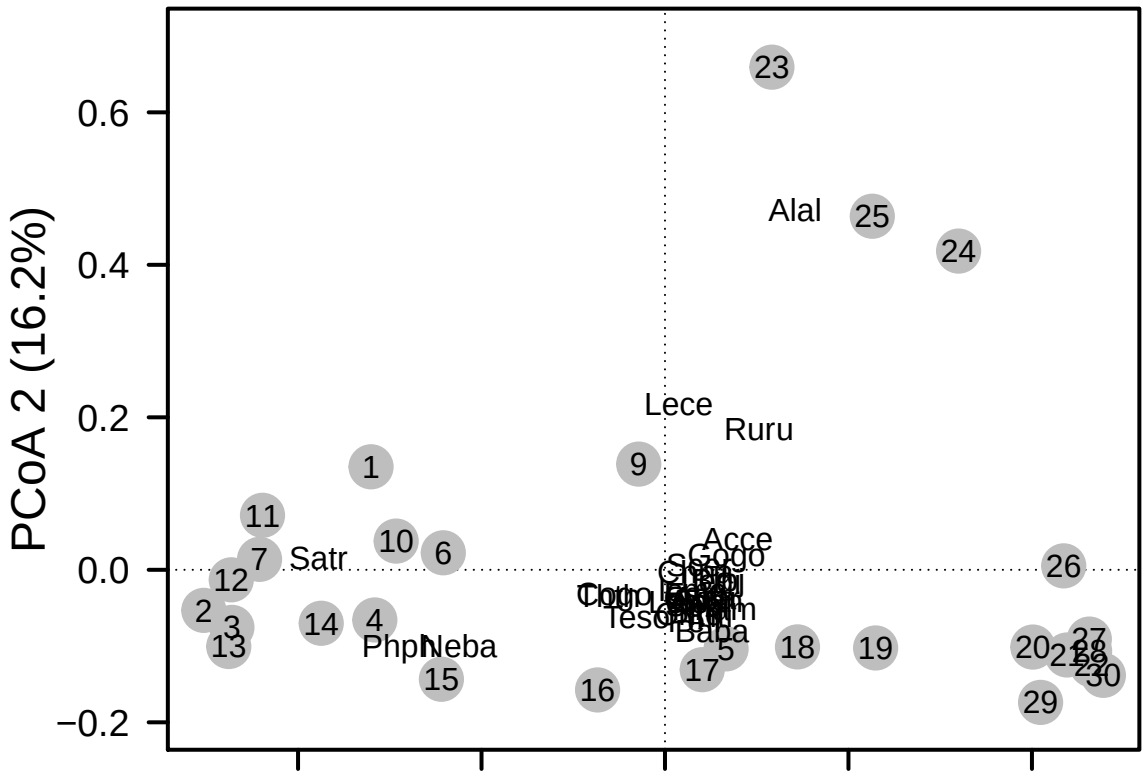
```
## [1] 81.3
```

```
# Plot of PCoA ordination
par(mar = c(1, 5, 2, 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, 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))

### I'm not sure why my x-axis label isn't showing up...

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

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.

```
# Influencial species
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, ]
imp.spp
```

```
##            Dim1       Dim2        Dim3
## Phph -0.8674640 -0.1699316 -0.12463098
## Neba -0.7674114 -0.1855678 -0.36963830
## Rham  0.8088751 -0.4192567  0.14136301
## Legi  0.8201759 -0.1701803  0.12423941
## Cyca  0.7595122 -0.4442926  0.17313658
## Abbr  0.7704744 -0.3452714  0.29277803
## Acce  0.7635195  0.2155765  0.10288179
## Blbj  0.8118483 -0.1324698  0.25581178
## Alal  0.4471283  0.8119843 -0.05167131
## Anan  0.7974122 -0.3918972  0.20944968
```

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

```
##
```

```
## ***VECTORS
##
##          Dim1     Dim2     r2 Pr(>r)
## Cogo -0.83884 -0.54438 0.2982  0.007 **
## Satr -0.99904  0.04371 0.4326  0.005 **
## 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  0.021 *
## Teso -0.44704 -0.89452 0.1700  0.084 .
## Chna  0.99707 -0.07644 0.4612  0.002 **
## Chto  0.42032 -0.90738 0.2579  0.020 *
## Lele  0.33041 -0.94384 0.0495  0.530
## Lece  0.06856  0.99765 0.3399  0.010 **
## Baba  0.54118 -0.84091 0.6752  0.001 ***
## Spbi  0.57341 -0.81927 0.4138  0.002 **
## Gogo  0.97507  0.22188 0.3753  0.002 **
## Eslu  0.72044 -0.69352 0.1673  0.096 .
## Pefl  0.43762 -0.89916 0.3048  0.009 **
## 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  0.008 **
## 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  0.001 ***
## 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
```

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***: Alal seems to influence community composition of sites 23, 24, and 25. Phph and Neva seem to influence community composition of the group of sites lower on PCoA 1. ***Answer 7b***: Because Phph and Neva are associated with the upstream sites that have low species diversity, they may be indicators of river 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 pairswise 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?

```r
jelly <- read.table("JellyBeans.txt", sep = "\t", header = TRUE)
jelly.source <- read.table("JellyBeans.Source.txt", sep = "\t", header = TRUE)

# Make a df for each population sampled
jelly.A <- jelly[jelly$Group == "A", ]
jelly.A <- jelly.A[, -c(1,2)]      # remove Site and Group columns

jelly.B <- jelly[jelly$Group == "B", ]
jelly.B <- jelly.B[, -c(1,2)]      # remove Site and Group columns

# Mean pairwise difference: group A
bray.A <- vegdist(jelly.A, method = "bray")
mean(bray.A)
```

```
## [1] 0.2714244
```

```r
# Mean pairwise difference: group B
bray.B <- vegdist(jelly.B, method = "bray")
mean(bray.B)
```

```
## [1] 0.3302977
```

```r
# T-test for A vs B comparison
t.test(bray.A, bray.B)
```

```
##
##  Welch Two Sample t-test
##
## data:  bray.A and bray.B
## t = -2.3151, df = 7.7248, p-value = 0.05041
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1178815510  0.0001348894
## sample estimates:
## mean of x mean of y
## 0.2714244 0.3302977
```

```r
# T-tests for A vs Source and for B vs Source
t.test(bray.A, jelly.source$Count)
```

```
##
##  Welch Two Sample t-test
##
## data:  bray.A and jelly.source$Count
## t = -4.6131, df = 25, p-value = 0.0001016
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -53.40434 -20.43743
## sample estimates:
##  mean of x  mean of y
##  0.2714244 37.1923077
```

```r
t.test(bray.B, jelly.source$Count)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  bray.B and jelly.source$Count
## t = -4.6057, df = 25, p-value = 0.0001035
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -53.34551 -20.37851
## sample estimates:
##  mean of x  mean of y
##  0.3302977 37.1923077
```

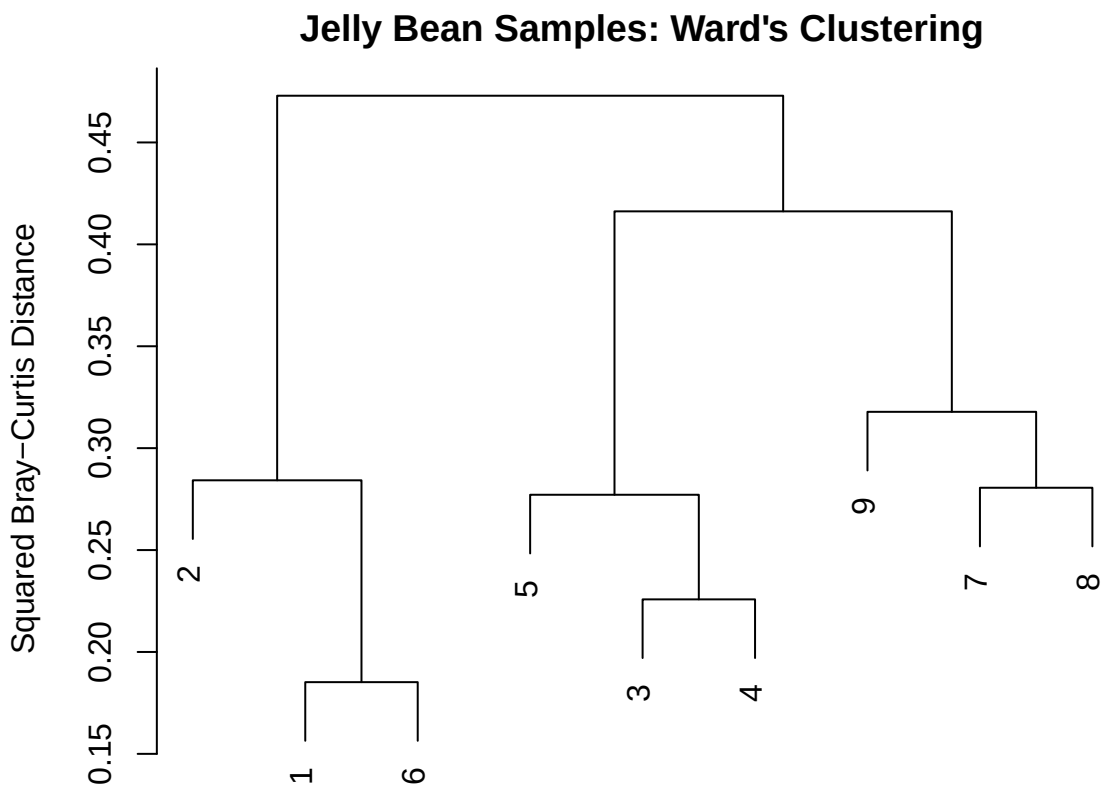
Dissimilarity was lower in group A (0.27) than group B (0.33), so subsamples were more similar to each other in group A than in group B. The groups were marginally different from each other ($t = -2.32$, $p = 0.0504$).
Both subcommunities were compositionally different from the source community ($P < 0.001$ for both).

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?

```r
jelly.db <- vegdist(jelly[, -c(1:2)], method = "bray")

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

par(mar = c(1, 5, 2, 2) + 0.1)
plot(jelly.full.ward,
     main = "Jelly Bean Samples: Ward's Clustering",
     ylab = "Squared Bray-Curtis Distance",
     col = jelly$Group)
```



I couldn't figure out how to color my cluster diagram according to Group, but Group A is sites 1, 2, 3, 6, and 8. Three A sites - 1, 2, and 6 - cluster together, suggesting a difference between A and B. Howver, site 3 clusters with two B sites, as does site 8. So while there's a little bit of evidence for differences in composition between groups A and B, it's not overpowering.