Assignment: Among Site (Beta) Diversity

Erik Parker; Z620: Quantitative Biodiversity, Indiana University

31 January, 2017

OVERVIEW

In this Assignment, 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. Change "Student Name" on line 3 (above) with your name.
2. Complete as much of the exercise as possible during class; what you do not complete in class will need to be done on your own outside of class.
3. Use the Handout as a guide; it contains a more complete description of data sets along with the proper scripting needed to carry out the exercise.
4. Be sure to **answer the questions** in this exercise document; they also correspond to the Handout. Space for your answer is provided in this document and indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">".
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. When you are done wit the Assignment, **Knit** the text and code into a html file.
7. After Knitting, please submit the completed Assignment by creating a **pull request** via GitHub. Your pull request should include this file *beta_assignment.Rmd* and the html output of `Knitr` (*beta_assignment.html*).

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 "*/Week3-Beta*" folder, and
4. load the `vegan` R package (be sure to install if needed).

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

```
## [1] "/var/host/media/removable/USB Drive/GitHub/QB2017_Parker/Week3-Beta"
```

```
setwd("./")
require(vegan)
```

```
## Loading required package: vegan
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.4-2
```

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.

```
package.list <- c('vegan', 'ade4', 'viridis', 'gplots', 'BiodiversityR', 'indicspecies')
for (package in package.list) {
  if (!require(package, character.only = T, quietly = T)) {
    install.packages(package)
    library(package, character.only = T)
  }
}

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
```

```
str(doubs$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 ...
```

```
str(doubs$species)
```

```
## 'data.frame':    27 obs. of  4 variables:
##  $ Scientific: chr  "Cottus gobio" "Salmo trutta fario" "Phoxinus phoxinus" "Nemacheilus barbatulus"
##  $ French    : chr  "chabot" "truite fario" "vairon" "loche franche" ...
##  $ English   : chr  "european bullhead" "brown trout" "minnow" "stone loach" ...
##  $ code      : Factor w/ 27 levels "Abbr","Acce",..: 9 22 19 17 26 25 7 8 16 14 ...
```

```
fish <- as.data.frame(doubs$fish)
dim(fish)
```

```
## [1] 30 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:
The doubs list contains 4 different objects: env, fish, xy, and species.

Answer 1b:
There seems to be 27 fish species in the doubs dataset.

Answer 1c:
The doubs dataset contains 30 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:
From the figures included in the handout: it appears that fish richness is highest in the middle

bend of the river, and at the most downstream locations in the river. Richness is not very high comparitively at the upstream locations sampled.

Answer 2b:
Counter to the conclusion drawn from the fish richness data, it seems like the brown trout abundance is highest at the upstream sites, and at the sites immediately upstream from the middle bend of the river.

Answer 2c:
These patterns seem to point to the fact that richness is limited when examining patterns of biodiversity because it is an overall measure of the number of species present per site, but it doesn't tell you anything about what specific species are driving those patterns. Here in particular we see that the brown trout abundance values are in almost complete disagreement with the overal fish species richness, meaning that just looking at richness overall is only giving us part of the picture and in particular leaves out the arguably more interesting insight of who is driving the larger richness patterns seen.

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(site.by.species = fish, sitenum1 = 1, sitenum2 = 2, pairwise = TRUE)
```

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

```
beta.W(site.by.species = 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:
Based on the beta.W equation we are using above, it seems as though alpha and beta diversity contribute to gamma diversity multiplicatively. As both the species diversity within habitats (alpha), and the differentiation among habitats (beta) increases, so does the total species diversity in any given landscape (gamma).

Answer 3b:
Using the function constructed above, we see that the pairwise comparison between site 1 and 2 returns a lower value for Whitaker's species turnover than does the comparison between site 1 and 10, meaning that there is lower beta diversity between sites 1 and 2, and so that the sites are more similar to eachother.

Answer 3c:
This change in definition would lead to beta diversity being interpreted not as how many times more diverse a region (gamma) is than any local site (alpha), but how many more species are in the total region when compared to a local site.

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:
The incidence and abundance based metrics differ in their treatment of rare species in that the incidence based metrics seem to give more weight to rare species as their counts are explicitly assigned variables in the formulae. The abundance based measures don't give rare species this same treatment, and instead just perform a summation of the abundance of every species in the site, likely reducing the impact of rare species on the similarity between sites.

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 <- fish[-8,]

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

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

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

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

fish.db
```

```
##             1          2          3          4          5          6
## 1  0.00000000 0.60000000 0.68421053 0.75000000 0.89189189 0.75000000
## 2  0.60000000 0.00000000 0.14285714 0.33333333 0.69565217 0.39393939
## 3  0.68421053 0.14285714 0.00000000 0.18918919 0.68000000 0.29729730
## 4  0.75000000 0.33333333 0.18918919 0.00000000 0.49090909 0.19047619
## 5  0.89189189 0.69565217 0.68000000 0.49090909 0.00000000 0.41818182
## 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  0.68421053 1.00000000 0.88235294 0.57142857 0.71428571 0.72727273
## 2  0.14285714 0.69230769 0.38461538 0.30434783 0.20000000 0.29032258
## 3  0.12500000 0.73333333 0.40000000 0.40740741 0.23529412 0.31428571
## 4  0.24324324 0.65714286 0.37142857 0.43750000 0.33333333 0.45000000
## 5  0.64000000 0.58333333 0.54166667 0.68888889 0.69230769 0.73584906
## 6  0.24324324 0.54285714 0.25714286 0.43750000 0.38461538 0.55000000
## 7  0.00000000 0.66666667 0.26666667 0.33333333 0.17647059 0.37142857
## 9  0.66666667 0.00000000 0.57142857 0.76000000 0.68750000 0.81818182
## 10 0.26666667 0.57142857 0.00000000 0.44000000 0.37500000 0.57575758
## 11 0.33333333 0.76000000 0.44000000 0.00000000 0.24137931 0.33333333
## 12 0.17647059 0.68750000 0.37500000 0.24137931 0.00000000 0.18918919
## 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  0.80645161 0.83333333 0.86046512 0.91489362 0.95555556 1.00000000
## 2  0.40000000 0.51111111 0.65384615 0.67857143 0.74074074 0.79310345
## 3  0.31818182 0.46938776 0.57142857 0.63333333 0.72413793 0.70967742
## 4  0.34693878 0.40740741 0.47540984 0.50769231 0.58730159 0.61194030
## 5  0.67741935 0.55223881 0.45945946 0.51282051 0.50000000 0.50000000
## 6  0.42857143 0.37037037 0.37704918 0.44615385 0.52380952 0.52238806
## 7  0.36363636 0.38775510 0.53571429 0.60000000 0.68965517 0.67741935
## 9  0.76190476 0.65957447 0.70370370 0.68965517 0.64285714 0.66666667
## 10 0.47619048 0.40425532 0.51851852 0.51724138 0.57142857 0.63333333
## 11 0.43589744 0.50000000 0.64705882 0.63636364 0.69811321 0.82456140
## 12 0.21739130 0.33333333 0.55172414 0.58064516 0.66666667 0.75000000
## 13 0.19148936 0.38461538 0.59322034 0.61904762 0.70491803 0.81538462
## 14 0.00000000 0.24590164 0.44117647 0.50000000 0.60000000 0.67567568
## 15 0.24590164 0.00000000 0.26027397 0.40259740 0.46666667 0.56962025
## 16 0.44117647 0.26027397 0.00000000 0.26190476 0.34146341 0.39534884
## 17 0.50000000 0.40259740 0.26190476 0.00000000 0.13953488 0.31111111
## 18 0.60000000 0.46666667 0.34146341 0.13953488 0.00000000 0.25000000
## 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  1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
## 2  0.91176471 0.94594595 0.97619048 1.00000000 1.00000000 1.00000000
## 3  0.88888889 0.92307692 0.95454545 1.00000000 1.00000000 0.92592593
## 4  0.74025974 0.78313253 0.82795699 0.92000000 0.88888889 0.81250000
## 5  0.48888889 0.50000000 0.52830189 0.89473684 0.79591837 0.68888889
## 6  0.68831169 0.73493976 0.78494624 0.84000000 0.77777778 0.68750000
## 7  0.86111111 0.89743590 0.93181818 0.90000000 0.93548387 0.85185185
## 9  0.68571429 0.76315789 0.76744186 0.77777778 0.72413793 0.84000000
## 10 0.77142857 0.81578947 0.86046512 0.88888889 0.79310345 0.76000000
## 11 0.91044776 0.91780822 0.95180723 0.86666667 0.92307692 0.90909091
## 12 0.89189189 0.92500000 0.95555556 0.90909091 0.93939394 0.93103448
## 13 0.92000000 0.95061728 0.97802198 1.00000000 1.00000000 1.00000000
## 14 0.83333333 0.86666667 0.90000000 0.93750000 0.90697674 0.84615385
```

```
## 15 0.70786517 0.76842105 0.77142857 0.94594595 0.87500000 0.81818182
## 16 0.58333333 0.62745098 0.66071429 0.90909091 0.81818182 0.76470588
## 17 0.42000000 0.49056604 0.55172414 0.83333333 0.69491525 0.74545455
## 18 0.32653061 0.40384615 0.47368421 0.82608696 0.64912281 0.66037736
## 19 0.23529412 0.29629630 0.38983051 0.84000000 0.63934426 0.61403509
## 20 0.00000000 0.10169492 0.18750000 0.86666667 0.57746479 0.67164179
## 21 0.10169492 0.00000000 0.10447761 0.87878788 0.61038961 0.69863014
## 22 0.18750000 0.10447761 0.00000000 0.89473684 0.65517241 0.73493976
## 23 0.86666667 0.87878788 0.89473684 0.00000000 0.57894737 0.46666667
## 24 0.57746479 0.61038961 0.65517241 0.57894737 0.00000000 0.46153846
## 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  1.00000000 1.00000000 1.00000000 0.97777778 1.00000000
## 2  0.96363636 0.97333333 0.97560976 0.93939394 1.00000000
## 3  0.93220339 0.94936709 0.95348837 0.92233010 0.98095238
## 4  0.78125000 0.83333333 0.82417582 0.81481481 0.87272727
## 5  0.55844156 0.56701031 0.57692308 0.53719008 0.59349593
## 6  0.68750000 0.76190476 0.78021978 0.77777778 0.83636364
## 7  0.89830508 0.92405063 0.93023256 0.90291262 0.96190476
## 9  0.71929825 0.76623377 0.76190476 0.78217822 0.84466019
## 10 0.82456140 0.84415584 0.85714286 0.84158416 0.90291262
## 11 0.92592593 0.94594595 0.95061728 0.89795918 0.98000000
## 12 0.93442623 0.95061728 0.95454545 0.90476190 0.98130841
## 13 0.96774194 0.97560976 0.97752809 0.90566038 1.00000000
## 14 0.85915493 0.89010989 0.89795918 0.84347826 0.93162393
## 15 0.76315789 0.77083333 0.78640777 0.73333333 0.81967213
## 16 0.63855422 0.66990291 0.69090909 0.65354331 0.72093023
## 17 0.54022989 0.57009346 0.57894737 0.51145038 0.57894737
## 18 0.45882353 0.48571429 0.50000000 0.44186047 0.52671756
## 19 0.32584270 0.37614679 0.41379310 0.41353383 0.48148148
## 20 0.21212121 0.19327731 0.22222222 0.24475524 0.29655172
## 21 0.20000000 0.13600000 0.16666667 0.18120805 0.23178808
## 22 0.25217391 0.12592593 0.12676056 0.11949686 0.18012422
## 23 0.82978723 0.88059701 0.89189189 0.91208791 0.91397849
## 24 0.48275862 0.61538462 0.64705882 0.70588235 0.71153846
## 25 0.59259259 0.70270270 0.72839506 0.77551020 0.78000000
## 26 0.00000000 0.18867925 0.23893805 0.33846154 0.36363636
## 27 0.18867925 0.00000000 0.09774436 0.18666667 0.19736842
## 28 0.23893805 0.09774436 0.00000000 0.14649682 0.15723270
## 29 0.33846154 0.18666667 0.14649682 0.00000000 0.14772727
## 30 0.36363636 0.19736842 0.15723270 0.14772727 0.00000000
```

```
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
```

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) resemblance matrix represents dissimilarity because the diagonal values (eg. site 1 vs site 1) are all zero. This is saying that these sites are completely similar to each other, so they are not dissimilar at all. If this was a similarity matrix, these values would be 1.

Answer 5b: The similarity between sites seems to be much higher when calculated using Sorensen's method when compared to the similarity given using Bray-Curtis (1-the dissimilarity value). Based on the variables for each method, I think this might be due the larger impact afforded by rare species in Sorensen's method. So, it seems like a higher number of rare species present in a site will lead to more disagreement between the two methods.

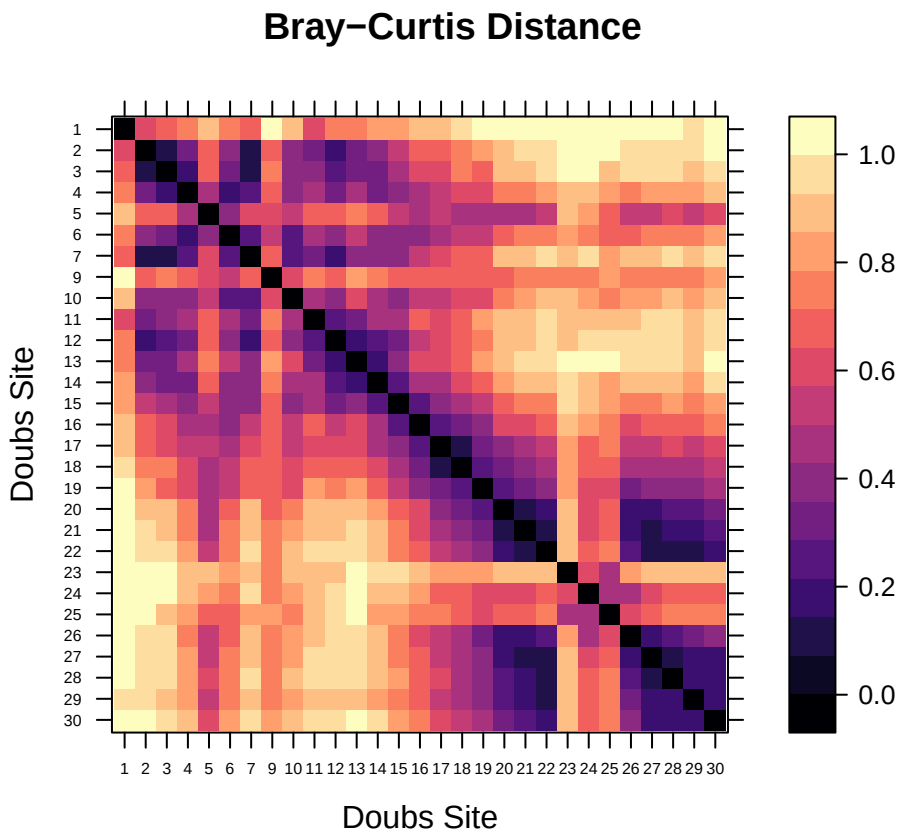
4) VISUALIZING BETA-DIVERSITY

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 = magma, xlab = "Doubs Site", ylab =
```

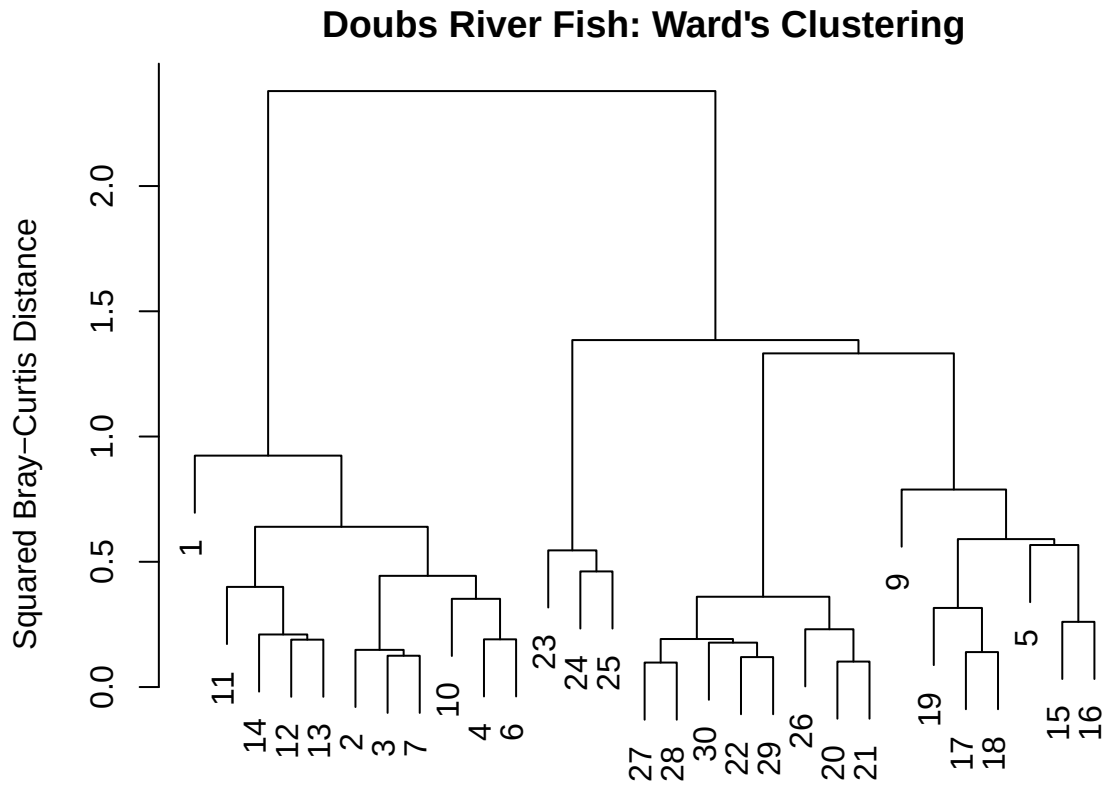


B. Cluster Analysis

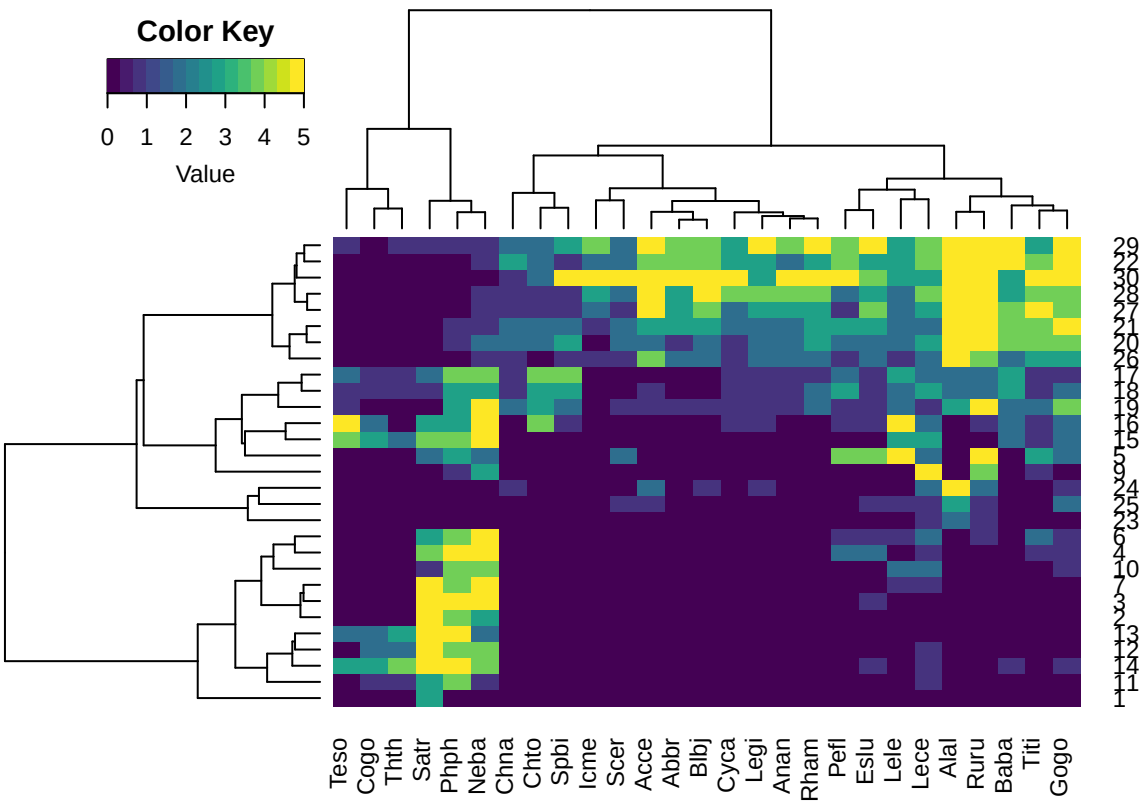
In the R code chunk below, do the following:

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

```
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")
```




```
gplots::heatmap.2(as.matrix(fish), distfun = function(x) vegdist(x, method = "bray"), hclustfun = functi
```



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: Upstream sites are less diverse due to a lack of resources and the subsequent dominance of a few well adapted species, while downstream and middle fork sites show greater diversity due to an abundance of resources allowing for greater niche partitioning facilitating coexistence between species.

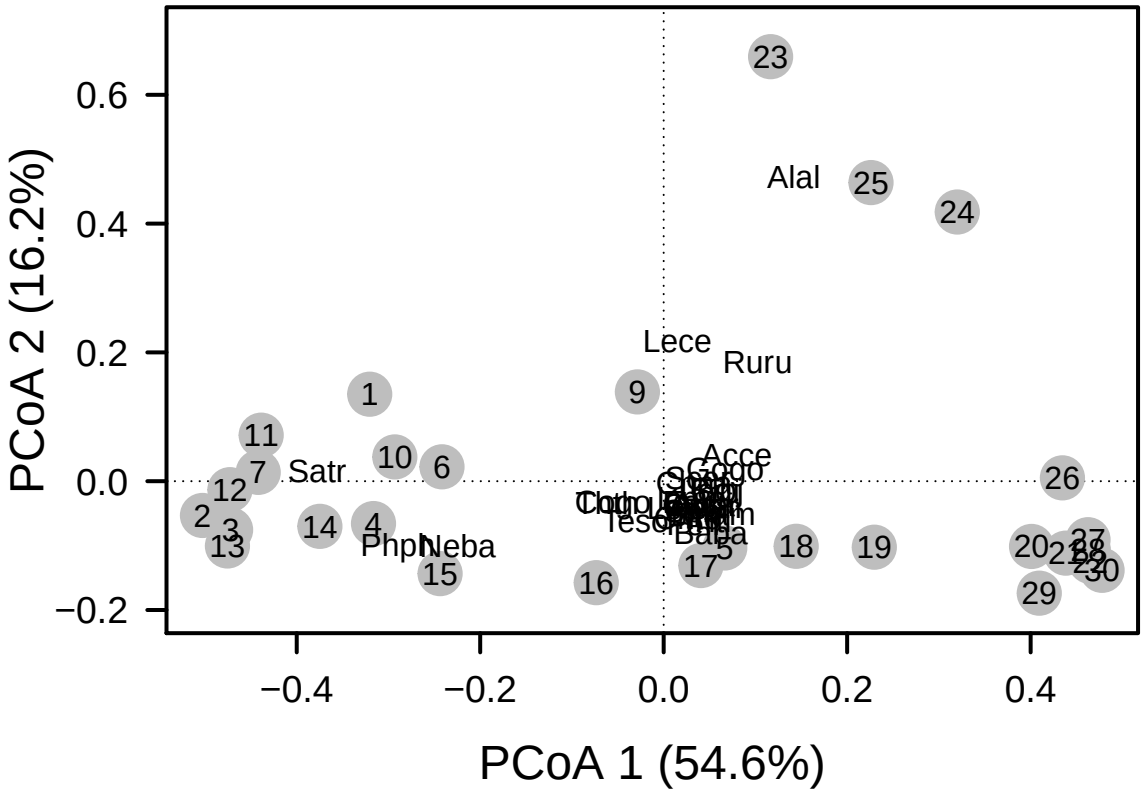
C. Ordination

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) # defining plot parameters

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)
# Plotting data

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)
# Adding axes, lines through origins

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))
# Adding and labeling points

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

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")
```




```
#Use relative abundance info to calculate and add species scores to plot
```

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 <- 0.7
imp.spp <- spe.corr[abs(spe.corr[,1]) >= corrcut | abs(spe.corr[,2]) >= corrcut,]
# Identification of important species based on PCoA axis correlation with cutoff of 0.7.

fit <- envfit(fish.pcoa, fishREL, perm = 999)
#Permutation test for species abundance based on correlation level defined above

fit
```

```
##
## ***VECTORS
##
##          Dim1     Dim2     r2 Pr(>r)
## Cogo -0.83884 -0.54438 0.2982  0.021 *
## 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  0.035 *
## Teso -0.44704 -0.89452 0.1700  0.098 .
## Chna  0.99707 -0.07644 0.4612  0.001 ***
## Chto  0.42032 -0.90738 0.2579  0.024 *
## Lele  0.33041 -0.94384 0.0495  0.501
```

```
## Lece  0.06856  0.99765 0.3399  0.014 *
## 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  0.100 .
## Pefl  0.43762 -0.89916 0.3048  0.008 **
## 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  0.010 **
## 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:
The grouping of sites based on fish community composition seems to very closely follow the geographic relationship illustrated in the fish richness plot on page 2 of the handout. Upstream sites, with low diversity, seem to largely cluster together on the left side of the PCoA plot, showing large seperation on the major axis of determination from the other more species rich sites. As expected based on this, the most species rich downstream and middle bend sites cluster together on the far right side of the plot, far from the species poor upstream sites. Sites with intermediate richness are interspersed between these two main clusters along the first axis, with little movement along axis 2 (which explains significantly less variation.)

Answer 7b: Based on all of the information I have so far, it would seem that the upstream sites are by and large the most challenging habitats in the stream, as they show the lowest species richness and so presumably require special adaptations to properly exploit. This means that the presence of spcies found primarily in species poor sites (such as Satr, Phph, and Neba) is a good indicator of habitat quality generally, and an indication of poor habitat quality specifically.

5) HYPOTHESIS TESTING

A. Multivariate Procedures for Categorical Designs

Earlier work done in the Doubs River suggested that the river has four distinct regions of habitat quality: the first region (sites 1-14) of "high quality"; the second (sites 15 - 19) and fourth (sites 26 - 30) of "moderate quality"; and the third (sites 20 - 25) of "low quality".

In the code chunk below, test the hypothesis that fish community composition varies with river quality.

1. create a factor vector that categorizes habitat quality in the Doubs River,

2. use the multivariate analyses for categorical predictors to describe how fish community structure relates to habitat quality.

```
quality <- c(rep("HQ", 13), rep("MQ", 5), rep("LQ", 6), rep("MQ", 5))

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

```
##
## Call:
## adonis(formula = fish ~ quality, permutations = 999, method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## quality    2    3.0947 1.54733   10.97 0.45765  0.001 ***
## Residuals 26    3.6674 0.14105         0.54235
## Total     28    6.7621                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
indval <- multipatt(fish, cluster = quality, func = "IndVal.g", control = how(nperm = 999))
summary(indval)
```

```
##
##  Multilevel pattern analysis
##  ---------------------------
##
##  Association function: IndVal.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 27
##  Selected number of species: 23
##  Number of species associated to 1 group: 1
##  Number of species associated to 2 groups: 22
##
##  List of species associated to each combination:
##
##  Group MQ  #sps.  1
##       stat p.value
## Teso 0.686   0.027 *
##
##  Group HQ+MQ  #sps.  2
##       stat p.value
## Satr 0.860   0.003 **
## Phph 0.859   0.017 *
##
##  Group LQ+MQ  #sps.  20
##       stat p.value
## Alal 0.935   0.001 ***
## Gogo 0.933   0.001 ***
## Ruru 0.916   0.002 **
## Legi 0.901   0.001 ***
```

```
## Baba 0.895   0.001 ***
## Chna 0.866   0.001 ***
## Spbi 0.866   0.002 **
## Cyca 0.866   0.002 **
## Acce 0.866   0.001 ***
## Lele 0.863   0.004 **
## Titi 0.853   0.006 **
## Chto 0.829   0.001 ***
## Rham 0.829   0.002 **
## Anan 0.829   0.002 **
## Eslu 0.827   0.018 *
## Pefl 0.806   0.014 *
## Blbj 0.791   0.003 **
## Scer 0.766   0.012 *
## Abbr 0.750   0.005 **
## Icme 0.661   0.023 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Calculating an indicator value for each species in each habitat group. Significant scores close to 1 s

fish.rel <- decostand(fish, method = "total")
phi <- multipatt(fish.rel, cluster = quality, func = "r.g", control = how(nperm = 999))
summary(phi)
```

```
##
##  Multilevel pattern analysis
##  ---------------------------
##
##  Association function: r.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 27
##  Selected number of species: 18
##  Number of species associated to 1 group: 9
##  Number of species associated to 2 groups: 9
##
##  List of species associated to each combination:
##
##  Group HQ  #sps.  3
##       stat p.value
## Phph 0.802   0.001 ***
## Neba 0.734   0.001 ***
## Satr 0.650   0.001 ***
##
##  Group LQ  #sps.  2
##       stat p.value
## Alal 0.693   0.001 ***
## Ruru 0.473   0.037 *
##
##  Group MQ  #sps.  4
##       stat p.value
## Anan 0.571   0.008 **
## Spbi 0.557   0.010 **
## Chto 0.542   0.011 *
```

```
## Icme 0.475   0.033 *
##
##  Group LQ+MQ  #sps.  9
##       stat p.value
## Legi 0.658   0.002 **
## Baba 0.645   0.001 ***
## Rham 0.600   0.005 **
## Acce 0.594   0.007 **
## Cyca 0.586   0.005 **
## Chna 0.571   0.005 **
## Blbj 0.571   0.004 **
## Gogo 0.523   0.016 *
## Abbr 0.499   0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Habitat preference of each species. Range from -1 (strong avoidance of a group) to 0 (no preference) t
```

Question 8: Based on the PERMANOVA, IndVal, and phi coefficient analyses, what did you learn about the relationship between habitat quality and the fish species composition?

> ***Answer 8***: I'm still not totally sure if I buy the earlier classification of habitat quality in this environment (it seems so counter intuitive to me that the high quality sites have so few species compared to the low quality ones. Must be that the few species in high quality do such a good job outcompeting the other species? But these low quality sites can still support 20+ species, seems like a lot!), but given they are correct: these analyses are telling us that the high quality sites are dominated by a very select group of species (Satr, Phph, and Neba) which serve as strong indicators of high quality habitat, and also have strong preferences for those habitat types. The remaining species seem to be most strongly associated with low quality sites, and medium quality too, but not really medium quality sites alone. So, if a species can't find a place in the very exclusive high quality areas, they seem to specialize on low quality habitats primarily.

B. Multivariate Procedures for Continuous Designs

i. Mantel Test

In the R code chunk below, do the following:

1. create distance matrices for both fish communities and environmental factors, and
2. use a Mantel test to determine if these matrices are correlated, and test the hypothesis that fish assemblages are correlated with stream environmental variables.

```
fish.dist <- vegdist(doubs$fish[-8,], method = "bray")
env.dist <- vegdist(scale(doubs$env[-8,]), method = "euclid")
#Defining our distance matrices for a mantel test. Using a bray-curtis distance matrix of site-species

mantel(fish.dist,env.dist)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = fish.dist, ydis = env.dist)
##
## Mantel statistic r: 0.604
##       Significance: 0.001
```

```
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.105 0.137 0.165 0.193
## Permutation: free
## Number of permutations: 999
```

```
#Looks for monotonic correlations (either always + or always -) between the two matrices supplied. Give
#Here testing hypothesis that fish assemblages are correlated with stream environmental variables. Sign
```

Question 9: What do the results from our Mantel test suggest about fish diversity and stream environmental conditions? How does this relate to your hypothesis about stream quality influencing fish communities?

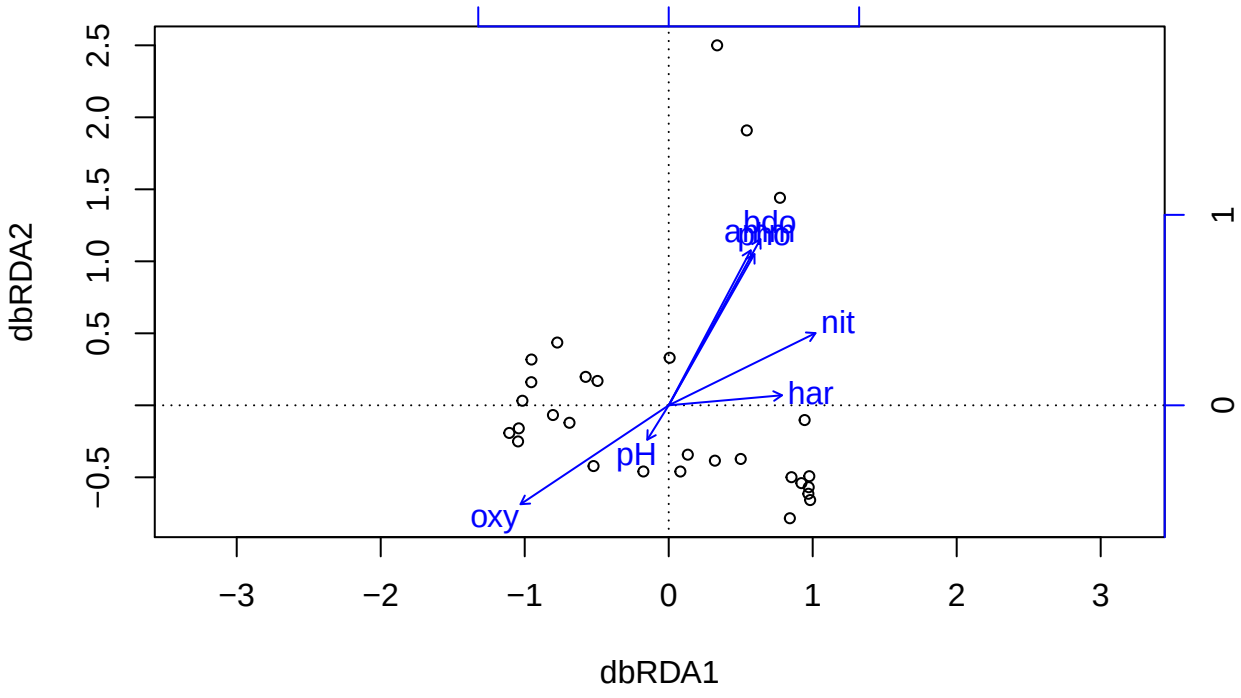
> ***Answer 9***:
> The Mantel test performed above returned a significant, positive r value, suggesting that there is a relatively high positive correlation between fish diversity and stream environmental variables. I believe that this means that as the measures of stream environmental conditions, contained in the site by environment matrix, increase so too does fish diversity. If I am intrepreting these results right, this Mantel test seems to support my earlier hypothesis, yet also refute the suggestion of the earlier work described above which classified the most species rich sites as low or medium quality, and the species poor sites as high quality. It seems to me that if that were the case, we would have seen a negative value for r returned from this Mantel test.

ii. Constrained Ordination

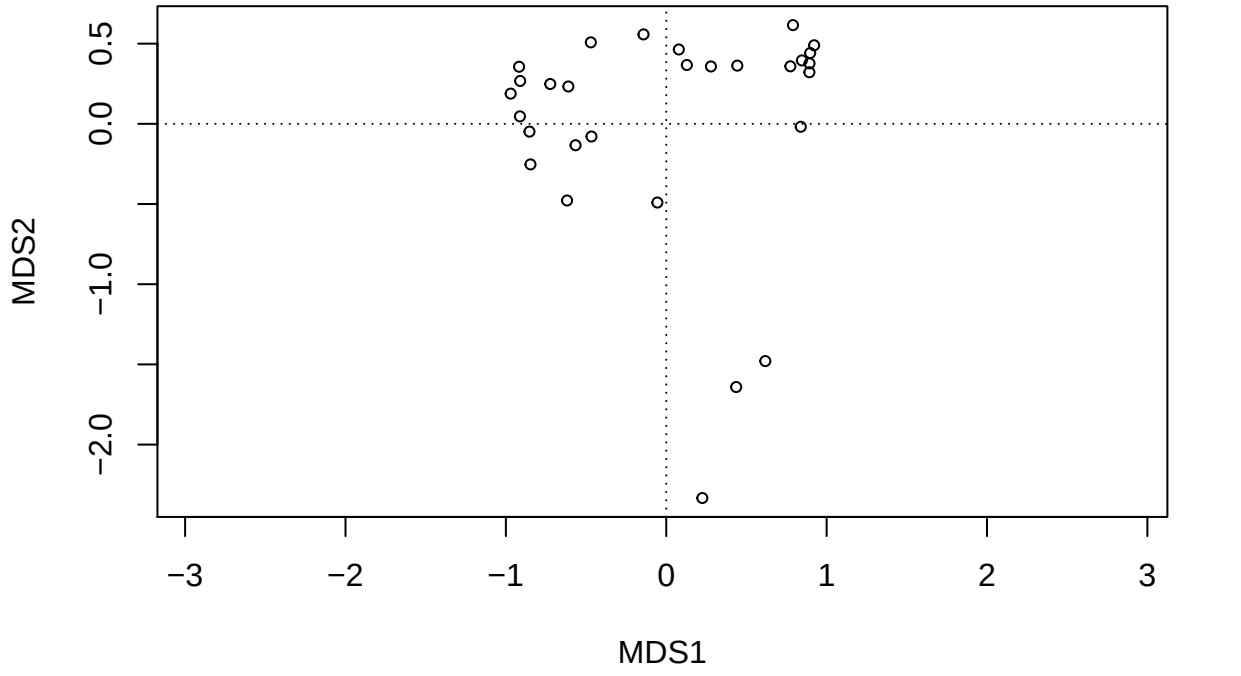
In the R code chunk below, do the following:

1. create an environmental matrix of the water chemistry data included in the `doubs` dataset using forward and reverse selection of variables,
2. conduct a redundancy analysis on the fish assemblages of the Doubs River,
3. use a permutation test to determine the significance of the constrained analysis,
4. use a permutation test to determine the correlation of each environmental factor on the constrained axes,
5. calculate the explained variation on the first and second constrained axes,
6. plot the constrained ordination results including labeled points for each site, and
7. add vectors that demonstrate the influence of each environmental factor the constrained ordination.

```
env.chem <- as.matrix(doubs$env[-8, 5:11])

doubs.dbrda <- dbrda(fish.db ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda)
```




```
doubs.dbrda.mod0 <- dbrda(fish.db ~ 1, as.data.frame(env.chem))

ordiplot(doubs.dbrda.mod0)
```




```
doubs.dbrda.mod1 <- dbrda(fish.db ~ ., as.data.frame(env.chem))

doubs.dbrda <- ordiR2step(doubs.dbrda.mod0, doubs.dbrda.mod1, perm.max = 200)
```

```
## Step: R2.adj= 0
## Call: fish.db ~ 1
##
```

```
##                 R2.adjusted
## <All variables>  0.53032584
## + oxy            0.27727176
## + nit            0.25755208
## + bdo            0.17477787
## + pho            0.14568614
## + har            0.14174915
## + amm            0.14142804
## <none>           0.00000000
## + pH            -0.01827054
##
##       Df    AIC      F Pr(>F)
## + oxy  1 47.939 11.742  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2772718
## Call: fish.db ~ oxy
##
##                 R2.adjusted
## <All variables>   0.5303258
## + bdo             0.4009000
## + amm             0.3474192
## + pho             0.3452702
## + har             0.3331357
## + nit             0.3316120
## <none>            0.2772718
## + pH              0.2586983
## - oxy             0.0000000
##
##       Df    AIC      F Pr(>F)
## + bdo  1 43.404 6.5716  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.4009
## Call: fish.db ~ oxy + bdo
##
##                 R2.adjusted
## <All variables>   0.5303258
## + nit             0.4980793
## + har             0.4695121
## <none>            0.4009000
## + pho             0.3938042
## + amm             0.3869134
## + pH              0.3865240
## - bdo             0.2772718
## - oxy             0.1747779
##
##       Df    AIC     F Pr(>F)
## + nit  1 39.134 6.034  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Step: R2.adj= 0.4980793
## Call: fish.db ~ oxy + bdo + nit
##
##                   R2.adjusted
## + amm               0.5415705
## <All variables>     0.5303258
## + pho               0.5277128
## + har               0.5218852
## <none>              0.4980793
## + pH                0.4843267
## - nit               0.4009000
## - oxy               0.3420426
## - bdo               0.3316120
```

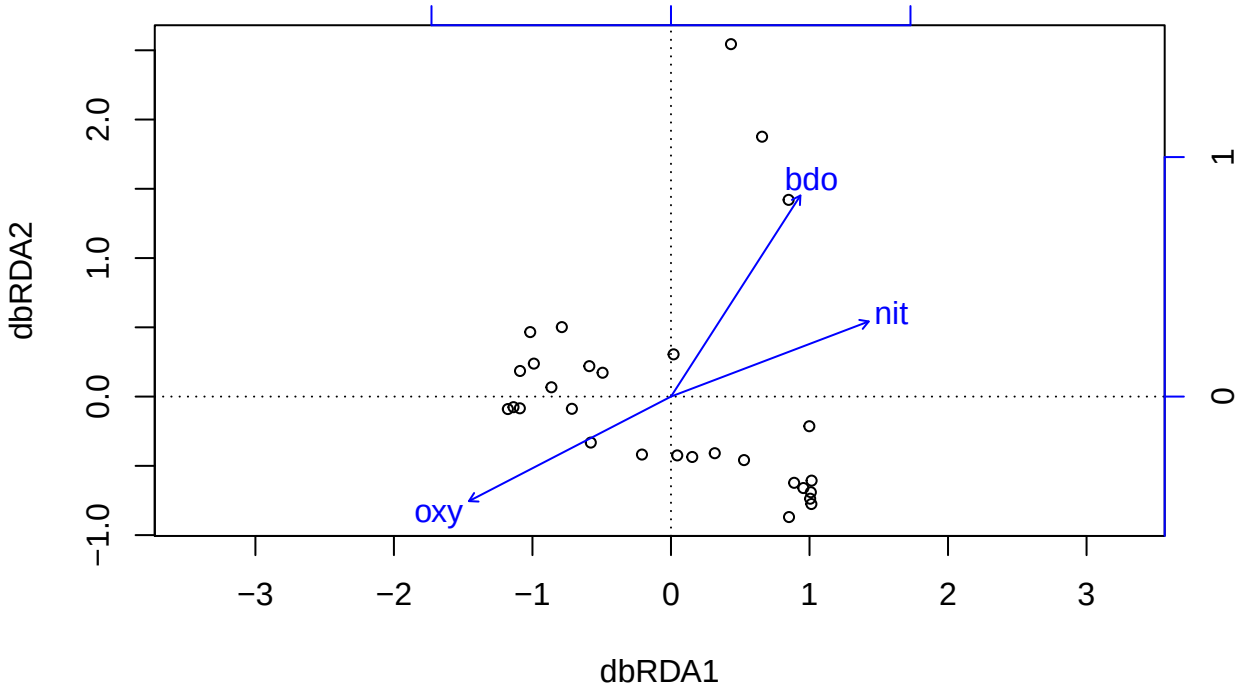
```
doubs.dbrda$call
```

```
## dbrda(formula = fish.db ~ oxy + bdo + nit, data = as.data.frame(env.chem))
```

```
doubs.dbrda$anova
```

```
##                   R2.adj Df    AIC       F Pr(>F)
## + oxy            0.27727  1 47.939 11.7421  0.002 **
## + bdo            0.40090  1 43.404  6.5716  0.002 **
## + nit            0.49808  1 39.134  6.0340  0.002 **
## <All variables> 0.53033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ordiplot(doubs.dbrda)
```




```
permutest(doubs.dbrda, permutations = 999)
```

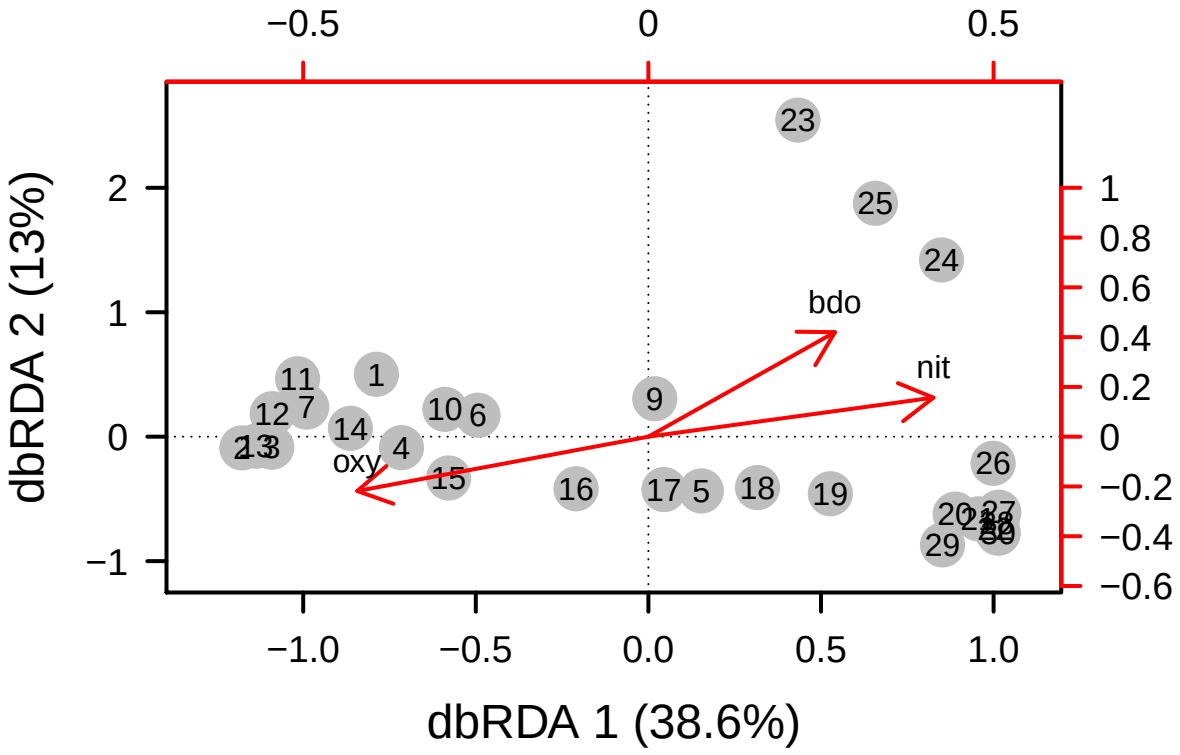
```
##
## Permutation test for dbrda
##
```

```
## Permutation: free
## Number of permutations: 999
##
## Call: dbrda(formula = fish.db ~ oxy + bdo + nit, data =
## as.data.frame(env.chem))
## Permutation test for all constrained eigenvalues
## Pseudo-F:     10.2619 (with 3, 25 Degrees of Freedom)
## Significance:     0.001
```

```
envfit(doubs.dbrda, env.chem[,c(4,6,7)], perm = 999)
```

```
##
## ***VECTORS
##
##       dbRDA1   dbRDA2     r2 Pr(>r)
## nit  0.87724  0.48005 0.6431  0.001 ***
## oxy -0.82864 -0.55979 0.7656  0.001 ***
## bdo  0.55603  0.83116 0.8939  0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 999
```

```
dbrda.explainvar1 <- round(doubs.dbrda$CCA$eig[1] /
                           sum(c(doubs.dbrda$CCA$eig, doubs.dbrda$CA$eig)), 3) *100
dbrda.explainvar2 <- round(doubs.dbrda$CCA$eig[2] /
                           sum(c(doubs.dbrda$CCA$eig, doubs.dbrda$CA$eig)), 3) *100

par(mar = c(5,5,4,4) + 0.1)

plot(scores(doubs.dbrda, display = "wa"), xlim = c(-1.3, 1.1), ylim = c(-1.1, 2.7), xlab = paste("dbRDA
ylab = paste("dbRDA 2 (", dbrda.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(scores(doubs.dbrda, display = "wa"),
       pch = 19, cex = 3, bg = "gray", col = "gray")
text(scores(doubs.dbrda, display = "wa"),
     labels = row.names(scores(doubs.dbrda, display = "wa")))

vectors <- scores(doubs.dbrda, display = "bp")
arrows(0, 0, vectors[,1], vectors[,2],
       lwd = 2, lty = 1, length = 0.2, col = "red")
text(vectors[,1], vectors[,2], pos = 3,
     labels = row.names(vectors))
axis(side = 3, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red", lwd = 2.2,
     at = pretty(range(vectors[,1])) * 2, labels = pretty(range(vectors[,1])))
axis(side = 4, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red", lwd = 2.2,
     at = pretty(range(vectors[,2])) * 2, labels = pretty(range(vectors[,2])))
```



Question 10: Based on the constrained ordination, what are the environmental variables (or groups of correlated variables) that seem to be contributing to variation in fish community structure?

> ***Answer 10***: Based on the results of the constrained ordination above, it seems that the environmental variables which most contribute to fish community structure are Nitrogen, Oxygen, and bdo (bound diffused oxygen?). It seems that of these three, which all contribute, oxygen and nitrogen are the strongest drivers of differentiation between divergent fish community groups.

iii. Variation Partitioning

In the code chunk below,

1. Create a matrix model of the selected environmental variables,
2. Create a matrix model of the selected PCNM axes,
3. Perform constrained and partial constrained ordinations using the spatial and environmental models you just created,
4. Test the significance of each of your constrained ordinations using permutation tests,
5. Partition the variation among sites into the relative importance of space, environment, spatially structured environment, and residuals,
6. Plot the variation partitioning output to visualize it.

```
env.mod <- model.matrix(~ oxy + bdo + nit, as.data.frame(env.chem))[,-1]

rs <- rowSums(fish)/sum(fish)

doubs.pcnmw <- pcnm(dist(doubs$xy[-8,]), w = rs, dist.ret = T)

doubs.pcnmw$values > 0
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [12]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
```

```
## [23] FALSE FALSE FALSE FALSE
```

```
doubs.space <- as.data.frame(scores(doubs.pcnmw))
doubs.pcnm.mod0 <- dbrda(fish.db ~ 1, doubs.space)
doubs.pcnm.mod1 <- dbrda(fish.db ~ ., doubs.space)
step.pcnm <- ordiR2step(doubs.pcnm.mod0, doubs.pcnm.mod1, perm.max = 200)
```

```
## Step: R2.adj= 0
## Call: fish.db ~ 1
##
##                   R2.adjusted
## <All variables>  0.626011301
## + PCNM2          0.235370423
## + PCNM3          0.078394885
## + PCNM13         0.065305668
## + PCNM5          0.046185074
## + PCNM6          0.032809156
## + PCNM16         0.030486700
## + PCNM14         0.029680999
## + PCNM9          0.020357410
## + PCNM15         0.013632610
## + PCNM8          0.009411968
## + PCNM1          0.003986221
## + PCNM17         0.002415012
## + PCNM10         0.001326442
## <none>           0.000000000
## + PCNM7         -0.001861430
## + PCNM11        -0.006841522
## + PCNM4         -0.007089863
## + PCNM12        -0.014396973
##
##         Df    AIC     F Pr(>F)
## + PCNM2  1 49.574 9.619  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2353704
## Call: fish.db ~ PCNM2
##
##                  R2.adjusted
## <All variables>   0.6260113
## + PCNM3           0.3429270
## + PCNM5           0.3057368
## + PCNM1           0.2885396
## + PCNM16          0.2786746
## + PCNM14          0.2744520
## + PCNM15          0.2692809
## + PCNM6           0.2659866
## + PCNM13          0.2636194
## + PCNM9           0.2517847
## + PCNM8           0.2496240
## + PCNM10          0.2434688
## + PCNM7           0.2431476
## + PCNM17          0.2404343
## + PCNM11          0.2366833
```

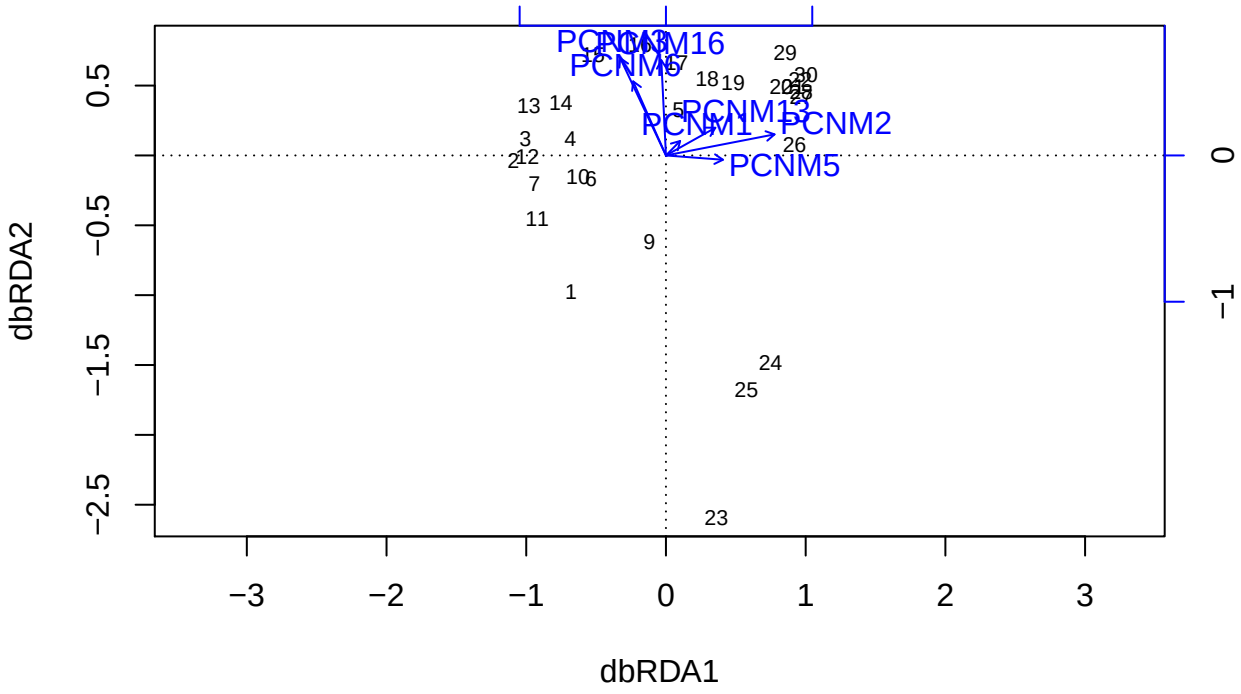
```
## <none>            0.2353704
## + PCNM12          0.2288789
## + PCNM4           0.2189522
## - PCNM2           0.0000000
##
##         Df    AIC      F Pr(>F)
## + PCNM3  1 46.083 5.4196  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.342927
## Call: fish.db ~ PCNM2 + PCNM3
##
##                 R2.adjusted
## <All variables>  0.62601130
## + PCNM5          0.40760197
## + PCNM1          0.39703000
## + PCNM16         0.38532100
## + PCNM15         0.38287481
## + PCNM14         0.37818268
## + PCNM13         0.37703761
## + PCNM6          0.35956442
## + PCNM8          0.35568849
## + PCNM7          0.35416308
## + PCNM10         0.35267745
## + PCNM17         0.35136832
## + PCNM9          0.34336720
## <none>           0.34292704
## + PCNM11         0.34163988
## + PCNM12         0.33965471
## + PCNM4          0.33115086
## - PCNM3          0.23537042
## - PCNM2          0.07839489
##
##         Df    AIC      F Pr(>F)
## + PCNM5  1 43.941 3.8385  0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.407602
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM1           0.4721469
## + PCNM16          0.4631976
## + PCNM15          0.4589111
## + PCNM14          0.4535248
## + PCNM13          0.4511582
## + PCNM6           0.4305640
## + PCNM7           0.4261965
## + PCNM8           0.4224505
## + PCNM17          0.4181666
## + PCNM10          0.4154485
```

```
## + PCNM11          0.4112178
## + PCNM9           0.4111995
## + PCNM12          0.4087602
## <none>            0.4076020
## + PCNM4           0.3976526
## - PCNM5           0.3429270
## - PCNM3           0.3057368
## - PCNM2           0.1195237
##
##         Df    AIC     F Pr(>F)
## + PCNM1  1 41.411 4.057  0.014 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.4721469
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5 + PCNM1
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM13          0.5212427
## + PCNM16          0.5208668
## + PCNM15          0.5161770
## + PCNM14          0.5147355
## + PCNM6           0.4999020
## + PCNM7           0.4936559
## + PCNM8           0.4904113
## + PCNM17          0.4856884
## + PCNM10          0.4835952
## + PCNM11          0.4760087
## + PCNM9           0.4751424
## + PCNM12          0.4747221
## <none>            0.4721469
## + PCNM4           0.4651218
## - PCNM1           0.4076020
## - PCNM5           0.3970300
## - PCNM3           0.3691841
## - PCNM2           0.1269210
##
##          Df    AIC      F Pr(>F)
## + PCNM13  1 39.346 3.4612   0.01 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.5212427
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM16          0.5767968
## + PCNM15          0.5715331
## + PCNM14          0.5698343
## + PCNM6           0.5475140
## + PCNM7           0.5392074
## + PCNM8           0.5379134
```

```
## + PCNM11          0.5281106
## + PCNM9           0.5267003
## + PCNM10          0.5265029
## + PCNM12          0.5255581
## <none>            0.5212427
## + PCNM17          0.5171800
## + PCNM4           0.5152311
## - PCNM13          0.4721469
## - PCNM1           0.4511582
## - PCNM5           0.4350790
## - PCNM3           0.4111185
## - PCNM2           0.2307026
##
##          Df   AIC      F Pr(>F)
## + PCNM16  1 36.48 4.0192  0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.5767968
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13 + PCNM16
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM6           0.6043089
## + PCNM8           0.5970286
## + PCNM12          0.5946888
## + PCNM7           0.5946475
## + PCNM9           0.5883735
## + PCNM10          0.5851333
## + PCNM15          0.5846468
## <none>            0.5767968
## + PCNM17          0.5748533
## + PCNM4           0.5733749
## + PCNM11          0.5711176
## + PCNM14          0.5652509
## - PCNM16          0.5212427
## - PCNM13          0.5208668
## - PCNM1           0.5136241
## - PCNM5           0.4764463
## - PCNM3           0.4676690
## - PCNM2           0.2646853
##
##         Df    AIC      F Pr(>F)
## + PCNM6  1 35.182 2.5296  0.034 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.6043089
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13 + PCNM16 + PCNM6
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM8           0.6248697
## + PCNM12          0.6208788
```

```
## + PCNM10          0.6170988
## + PCNM7           0.6142419
## + PCNM15          0.6140369
## + PCNM9           0.6107110
## <none>            0.6043089
## + PCNM17          0.6037430
## + PCNM11          0.5978305
## + PCNM4           0.5963667
## + PCNM14          0.5932113
## - PCNM6           0.5767968
## - PCNM13          0.5495723
## - PCNM16          0.5475140
## - PCNM1           0.5362957
## - PCNM3           0.5092539
## - PCNM5           0.4965968
## - PCNM2           0.2824470
##
##         Df    AIC     F Pr(>F)
## + PCNM8  1 34.219 2.151  0.102
```

```
plot(step.pcnm)
```




```
step.pcnm$anova
```

```
##                  R2.adj Df    AIC      F Pr(>F)
## + PCNM2         0.23537  1 49.574 9.6190  0.002 **
## + PCNM3         0.34293  1 46.083 5.4196  0.002 **
## + PCNM5         0.40760  1 43.941 3.8385  0.012 *
## + PCNM1         0.47215  1 41.411 4.0570  0.014 *
## + PCNM13        0.52124  1 39.346 3.4612  0.010 **
## + PCNM16        0.57680  1 36.480 4.0192  0.012 *
## + PCNM6         0.60431  1 35.182 2.5296  0.034 *
## <All variables> 0.62601
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
space.mod <- model.matrix(~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13 + PCNM16 + PCNM6, doubs.space)[,-1]

doubs.total.env <- dbrda(fish.db ~ env.mod)
doubs.total.space <- dbrda(fish.db ~ space.mod)

doubs.env.cond.space <- dbrda(fish.db ~ env.mod + Condition(space.mod))
doubs.space.cond.env <- dbrda(fish.db ~ space.mod + Condition(env.mod))

permutest(doubs.env.cond.space, permutations = 999)
```

```
##
## Permutation test for dbrda
##
## Permutation: free
## Number of permutations: 999
##
## Call: dbrda(formula = fish.db ~ env.mod + Condition(space.mod))
## Permutation test for all constrained eigenvalues
## Pseudo-F:     4.423025 (with 3, 18 Degrees of Freedom)
## Significance:     0.001
```

```
permutest(doubs.space.cond.env, permutations = 999)
```

```
##
## Permutation test for dbrda
##
## Permutation: free
## Number of permutations: 999
##
## Call: dbrda(formula = fish.db ~ space.mod + Condition(env.mod))
## Permutation test for all constrained eigenvalues
## Pseudo-F:     4.174109 (with 7, 18 Degrees of Freedom)
## Significance:     0.001
```

```
permutest(doubs.total.env, permutations = 999)
```

```
##
## Permutation test for dbrda
##
## Permutation: free
## Number of permutations: 999
##
## Call: dbrda(formula = fish.db ~ env.mod)
## Permutation test for all constrained eigenvalues
## Pseudo-F:     10.2619 (with 3, 25 Degrees of Freedom)
## Significance:     0.001
```

```
permutest(doubs.total.space, permutations = 999)
```

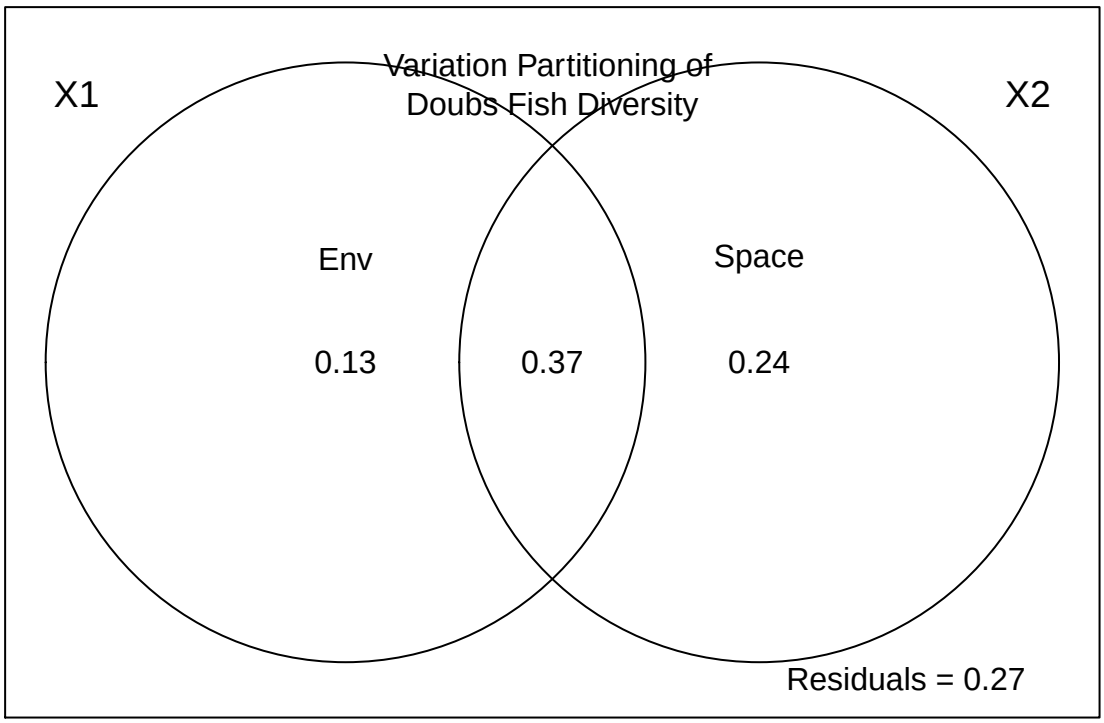
```
##
## Permutation test for dbrda
##
## Permutation: free
## Number of permutations: 999
```

```
##
## Call: dbrda(formula = fish.db ~ space.mod)
## Permutation test for all constrained eigenvalues
## Pseudo-F:     7.108896 (with 7, 21 Degrees of Freedom)
## Significance:     0.001
```

```
doubs.varpart <- varpart(fish.db, env.mod, space.mod)
doubs.varpart
```

```
##
## Partition of squared Bray distance in dbRDA
##
## Call: varpart(Y = fish.db, X = env.mod, space.mod)
##
## Explanatory tables:
## X1:  env.mod
## X2:  space.mod
##
## No. of explanatory tables: 2
## Total variation (SS): 6.7621
## No. of observations: 29
##
## Partition table:
##                       Df R.squared Adj.R.squared Testable
## [a+b] = X1             3   0.55186       0.49808     TRUE
## [b+c] = X2             7   0.70323       0.60431     TRUE
## [a+b+c] = X1+X2       10   0.82917       0.73426     TRUE
## Individual fractions
## [a] = X1|X2            3                 0.12995     TRUE
## [b]                    0                 0.36813    FALSE
## [c] = X2|X1            7                 0.23618     TRUE
## [d] = Residuals                          0.26574    FALSE
## ---
## Use function 'capscale' to test significance of fractions of interest
```

```
par(mar = c(2,2,2,2))
plot(doubs.varpart)
text(1, 0.25, "Space")
text(0, 0.25, "Env")
mtext("Variation Partitioning of \nDoubs Fish Diversity", side = 3, line = -3)
```



Question 11: Interpret the variation partitioning results.

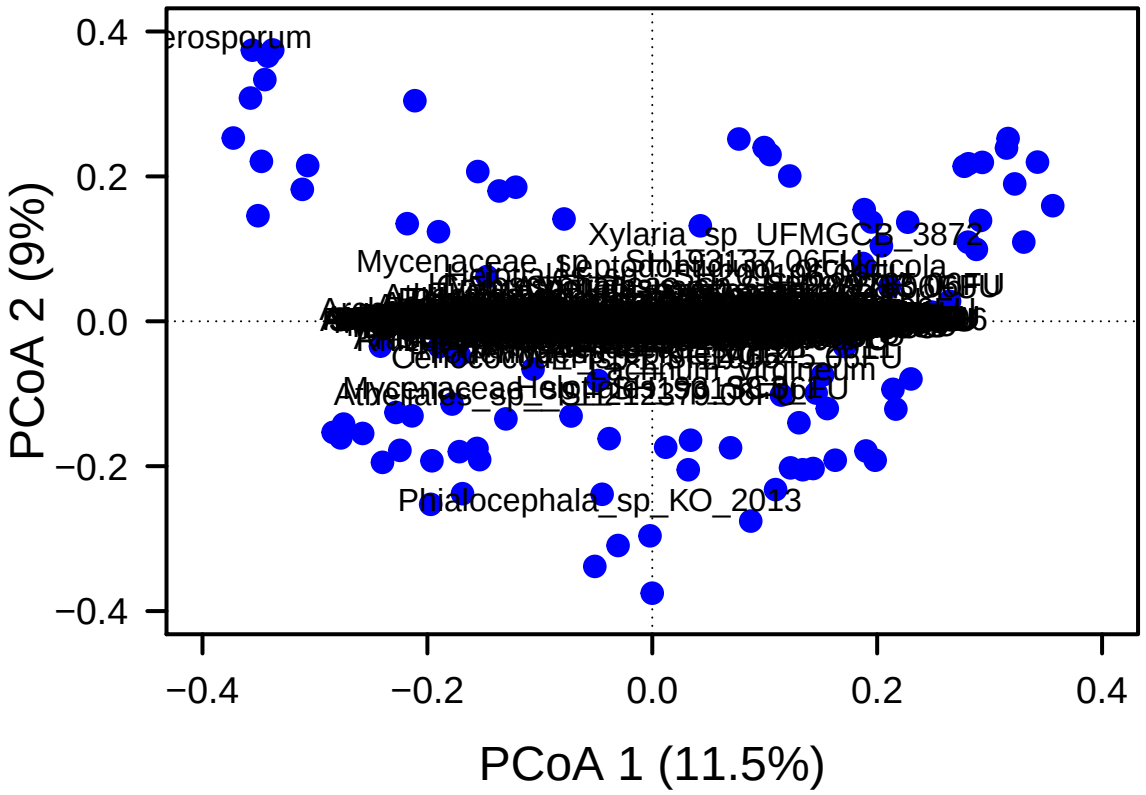
> ***Answer 11***: It seems as though most variation in fish community structure in the Doubs River can be explained due to an interaction between environmental variables and spatial position of sites. The interaction between those two factors explains 37% of the variation in fish diversity seen, which doesn't seem to be a huge amount, but is good enough to be the largest single contribution of the variables examined here. Space and environment alone both also explain a good amount of the variation by themselves, but behind their interaction the second largest explanatory category is neither space nor environment. This suggests to me that either 1) there is something else going on here that is so far unaccounted for or 2) a good portion of the variation in fish community structure seen between sites is due to something like random chance or sampling error.

SYNTHESIS

Load the dataset you are using for your project. Perform an ordination to visualize your dataset. Using this ordination, develop some hypotheses relevant to β-diversity and identify the appropriate tools you would use to test them.

```
endophyte.data <- read.csv("/media/removable/USB Drive/Z620.jay.non.GitHub/Endophyte communities and tr
endophyte.data <- endophyte.data[,-1]

endophyte.bray <- vegdist(endophyte.data, method = "bray")
endophyte.sorensen <- vegdist(endophyte.data, method = "bray", binary = TRUE)


endophyte.pcoa.bray <- cmdscale(endophyte.bray, eig = TRUE, k = 3)
endophyte.pcoa.sorensen <- cmdscale(endophyte.sorensen, eig = TRUE, k = 3)

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

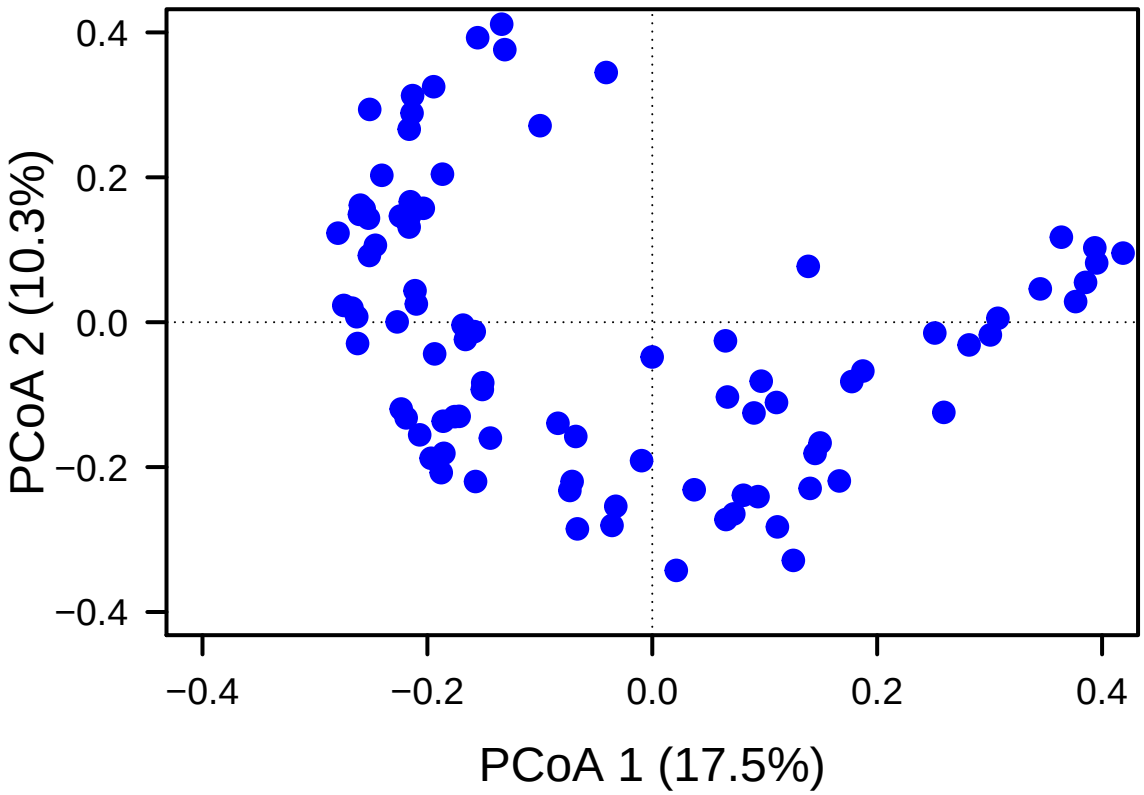
```
sum.eig.endo.bray <- sum(explainvar1.bray, explainvar2.bray, explainvar3.bray)

#PCoA ordination plot of bray endophyte data below

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

plot(endophyte.pcoa.bray$points[ ,1], endophyte.pcoa.bray$points[ ,2],
     xlab = paste("PCoA 1 (", explainvar1.bray, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2.bray, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE, ylim = c(-.4,.4), xli

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(endophyte.pcoa.bray$points[,1], endophyte.pcoa.bray$points[,2],
       pch = 19, cex = 1.5, bg = "gray", col = "blue")
text(endophyte.pcoa.bray$points[,1], endophyte.pcoa.bray$points[,2],
     labels = row.names(endophyte.pcoa.bray$points))


endophyte.pcoa.bray <- add.spec.scores(endophyte.pcoa.bray,endophyte.data,method = "pcoa.scores")
text(endophyte.pcoa.bray$cproj[ ,1], endophyte.pcoa.bray$cproj[, 2],
     labels = row.names(endophyte.pcoa.bray$cproj), col = "black")
```



Well. That Bray-Curtis dissimilarity index version of the ordination sure doesn't look like much. Most of the points are clustered around the center, and the ones that aren't don't seem to be

segregating into any real groupings of note. Maybe those points in the top left of the plot count for something though, and there is a particular influential fungal species up there (which in this view is cut off, but I found it to be Piloderma sphaerosporum). I'm not sure if it is totally appropriate to choose a different resemblance matrix index here just because the plot doesn't look great, and the major axes don't explain much of the variation, but I decided to give it a try anyways below with Sorensen's. I'm just nervous about this as Bray-Curtis seems the most appropriate for abundance-based data such as ours, which is given in relative abundances.

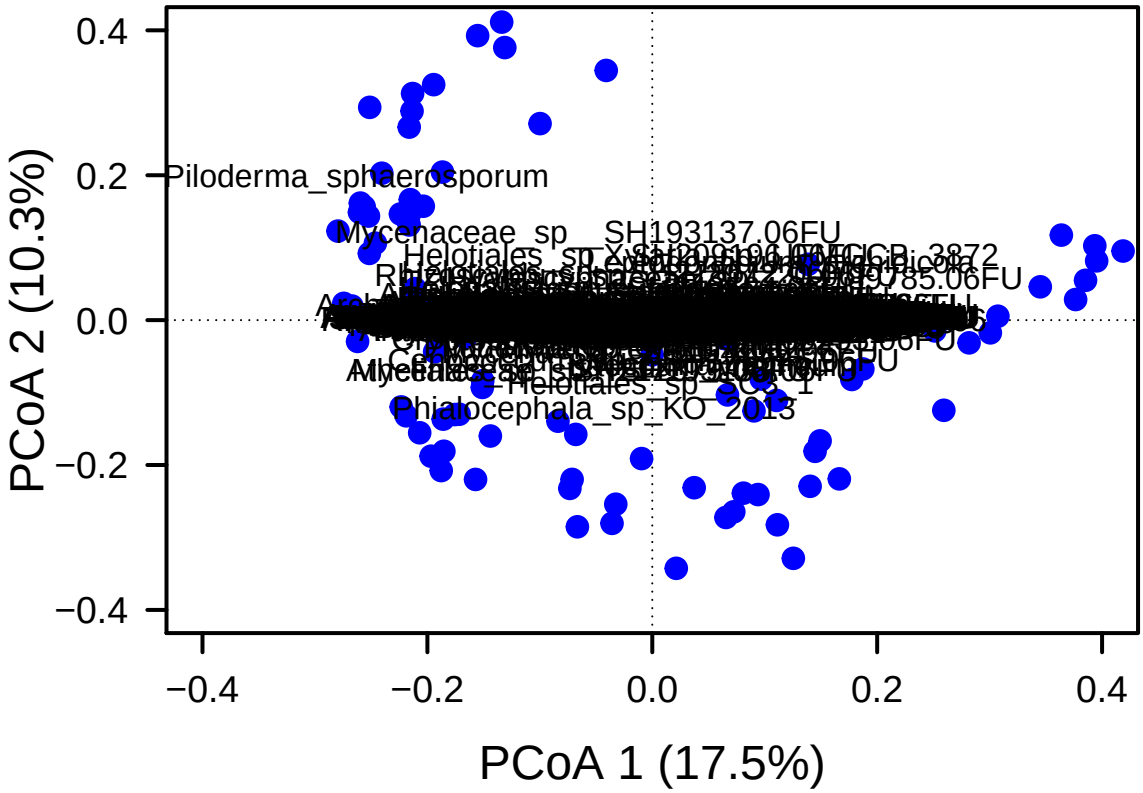
```
endophyte.pcoa.sorensen <- cmdscale(endophyte.sorensen, eig = TRUE, k = 3)

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

sum.eig.endo.sorensen <- sum(explainvar1.sorensen, explainvar2.sorensen, explainvar3.sorensen)

#PCoA ordination plot of bray endophyte data below

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

plot(endophyte.pcoa.sorensen$points[ ,1], endophyte.pcoa.sorensen$points[ ,2],
     xlab = paste("PCoA 1 (", explainvar1.sorensen, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2.sorensen, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE, ylim = c(-.4,.4), xli

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(endophyte.pcoa.sorensen$points[,1], endophyte.pcoa.sorensen$points[,2],
       pch = 19, cex = 1.5, bg = "gray", col = "blue")
text(endophyte.pcoa.sorensen$points[,1], endophyte.pcoa.sorensen$points[,2],
     labels = row.names(endophyte.pcoa.sorensen$points))
```




```
#endophyte.pcoa.sorensen <- add.spec.scores(endophyte.pcoa.sorensen,endophyte.data,method = #"pcoa.scor
#text(endophyte.pcoa.sorensen$cproj[ ,1], endophyte.pcoa.sorensen$cproj[, 2],
 #    labels = row.names(endophyte.pcoa.sorensen$cproj), col = "black")
```

```
endophyte.pcoa.sorensen <- cmdscale(endophyte.sorensen, eig = TRUE, k = 3)

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

sum.eig.endo.sorensen <- sum(explainvar1.sorensen, explainvar2.sorensen, explainvar3.sorensen)

#PCoA ordination plot of bray endophyte data below

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

plot(endophyte.pcoa.sorensen$points[ ,1], endophyte.pcoa.sorensen$points[ ,2],
     xlab = paste("PCoA 1 (", explainvar1.sorensen, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2.sorensen, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE, ylim = c(-.4,.4), xl

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(endophyte.pcoa.sorensen$points[,1], endophyte.pcoa.sorensen$points[,2],
       pch = 19, cex = 1.5, bg = "gray", col = "blue")
text(endophyte.pcoa.sorensen$points[,1], endophyte.pcoa.sorensen$points[,2],
```

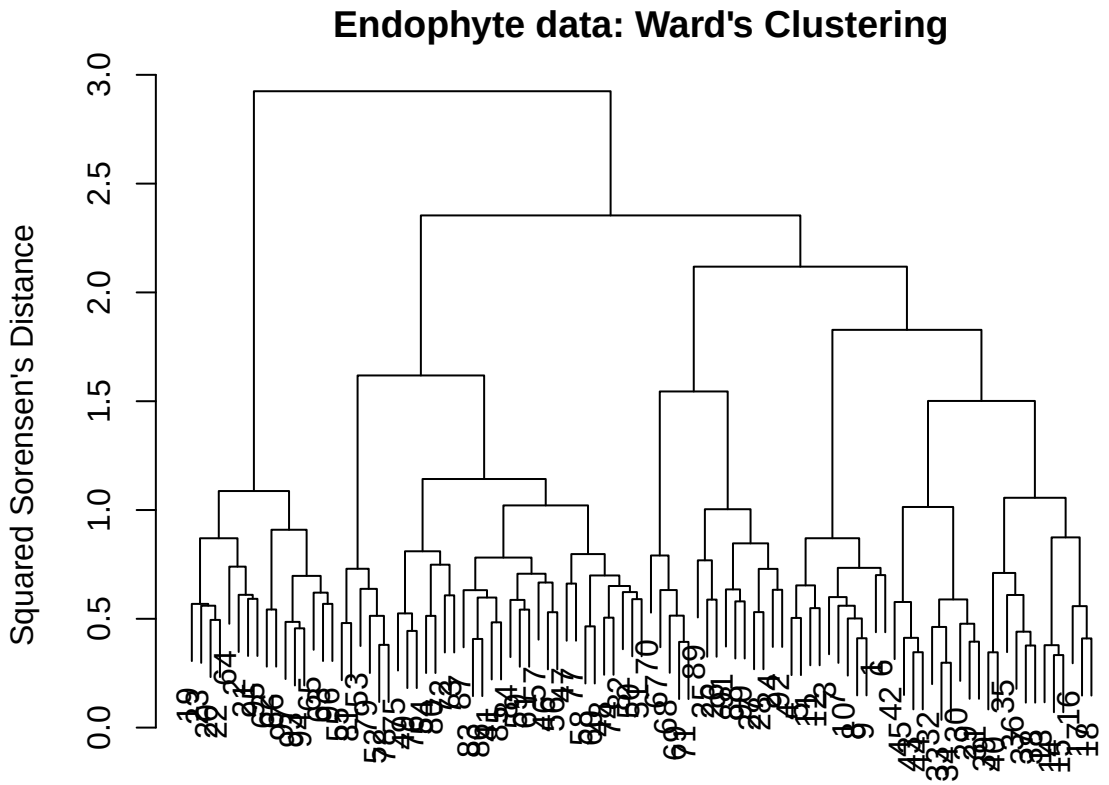
```
     labels = row.names(endophyte.pcoa.sorensen$points))


endophyte.pcoa.sorensen <- add.spec.scores(endophyte.pcoa.sorensen,endophyte.data,method = "pcoa.scores
text(endophyte.pcoa.sorensen$cproj[ ,1], endophyte.pcoa.sorensen$cproj[, 2],
     labels = row.names(endophyte.pcoa.sorensen$cproj), col = "black")
```



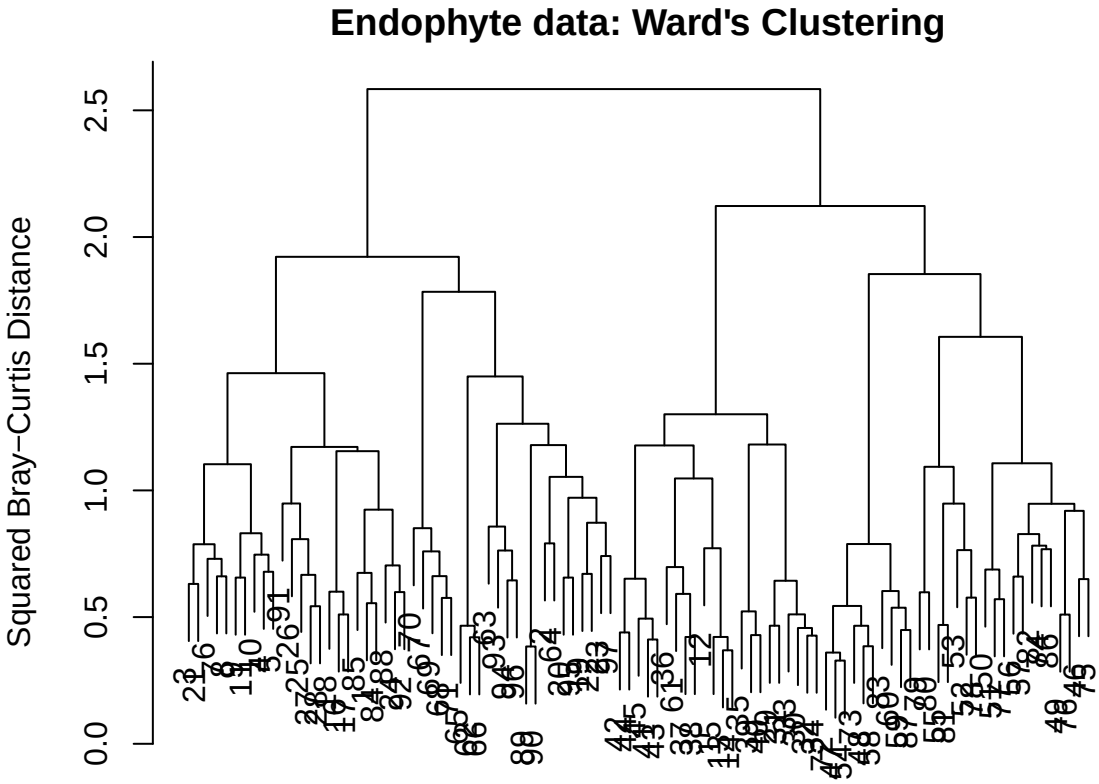
Again, a mess around the middle, especially with the fungal species names overlayed on the plot, but it does look a bit better than the Bray plot above. Still have the same influential fungal species in the top left grouping of points, and see more seperation overall. Also good is the fact that the two major axes explain much more variation in this plot than in the previous one. It's really unfortunate that for some reason the site numbers don't want to display on the points here, as that would be really helpful in distinguising groupings (are groups all one tree species? One region?), but for the life of me I can't get them to work out right.

Overall, it seems from these ordinations that Piloderma sphaerosporum is far and away the most influential species present in this dataset and it seems like its presence in a few sites is leading to their increased differentiation when compared to the rest of the sites. Based only on this, I hypothesize that this species' presence is an indicator of challenging habitats and that it is an influential species in these top left sites not due to its actions excluding other species, but due to its ability to survive where other fungal species cannot. Furthermore, I predict that the most highly differentiated sites according to this ordination correspond not to invasive Pinus species in our dataset, but to the two outgroups, the Nothofagus and Sylvestris. I hypothesize that beta diversity is highest between tree sites of different species (Contorta vs Sylvestris) as opposed to between tree sites of different location (European vs USA Contorta) To test these hypotheses, I need to better incorporate site information into my analyses (a given!), and work on gathering some environmental matrics (lat/long data, temperature, soil conditions, precipitation) not currently in the dataset. The addition of this environmental data would allow for a Mantel test, or other multivariate procedures, to be done.

```
endophyte.ward <- hclust(endophyte.sorensen, method = "ward.D2")

par(mar = c(1,5,2,2) + 0.1)
plot(endophyte.ward, main = "Endophyte data: Ward's Clustering", ylab = "Squared Sorensen's Distance")
```




```
endophyte.ward.bray <- hclust(endophyte.bray, method = "ward.D2")

par(mar = c(1,5,2,2) + 0.1)
plot(endophyte.ward.bray, main = "Endophyte data: Ward's Clustering", ylab = "Squared Bray-Curtis Distar
```



I was hoping these cluster plots might provide a little more insight into what's going on by site, but they really don't give too much carity. The better one, like with the ordination above, is the Sorensen's plot, but it doesn't tell much still. The best thing I can take from it is that there are a few consecutive sites on the far left and right (19,20,22,23 on the left / 13-18 on the right) that cluster together. These sites correspond to the Southern Hemisphere P. Contorta, and USA P. Contorta sites in the dataset. So, it seems as though location may be playing a role in community structure differentation for at least some sites?