Assignment: Among Site (Beta) Diversity

Katie Beidler; Z620: Quantitative Biodiversity, Indiana University

30 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).

```
clr = function() {
  ENV = globalenv()
  ll = ls(envir = ENV)
  ll = ll[ll != "clr"]
  rm(list = ll, envir = ENV)
}
getwd()
```

```
## [1] "/Users/bhbeidler/GitHub/QB2017_Beidler/Week3-Beta"
```

```
setwd("/Users/bhbeidler/GitHub/QB2017_Beidler/Week3-Beta")

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

```
## This is vegan 2.4-2
```

```
##
## Attaching package: 'ade4'
```

```
## The following object is masked from 'package:vegan':
##
##     cca
```

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

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

```
# I could not get the BiodiversityR package to load- I tried loading it from the source...instead I jus
```

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.

```
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$fish)
```

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

```
# The fish object is the site by species matrix
```

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: Doubs is a list with 4 components including: an environmental dataframe, a fish dataframe, a xy dataframe and species dataframe ***Answer 1b***: There are 27 fish species in the `doubs` dataset. ***Answer 1c***: There are 30 sites in the `doubs` dataset.

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: Downstream richness values are higher compared to upstream values and there is another region of high richness towards the middle of the stream.

Answer 2b: Brown Trout (*Salmo trutta*) abundance generally increases as you move upstream.

Answer 2c: While richness tended to be higher downstream, the composition of the community likely differed between upstream and downstream regions of the stream as Brown Trout abundance was highest upstrea. Therefor there are limitations associated with richness when examining patterns of biodiversity across sites, because richness may be the same among sites but species composition may differ.

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.w1 = function(site.by.species = ""){
  SbyS.pa = decostand(site.by.species, method = "pa")
  S = ncol(SbyS.pa[,which(colSums(SbyS.pa) > 0)])
  a.bar = mean(specnumber(SbyS.pa))
  b.w = round(S/a.bar, 3)
  return(b.w)
}

# Beta diversity function that calcuates Whittaker's Beta diversity- the inpt is a sbys matrix with opt
beta.w2 = 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)
  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(c(b.w,a.bar,S))
}
}
# Calculating beta, average Richness at each site aka local richness, and gamma (S)
beta.w2(doubs$fish)
```

```
## [1]  2.16 12.50 27.00
```

```
# a.bar = average richness at each site = 12.5
# Whittakers turover is how many times the species composition changes completely among the two sites
# comparing site 1 to site 2 dissimilarity*
beta.w2(doubs$fish,1,2,pairwise = TRUE)
```

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

```
# comparing site 1 to site 10
beta.w2(doubs$fish,1,10,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***: beta = 2.16. This index quantifies how many more times more diverse the regional species pool (gamma = 27) is than the average richness at each site within the region (alpha = 12.5 in this case). For the River Doubs gamma diversity is around double alpha diversity. So gamma fish diversity is due to high richness and low turnover. ***Answer 3b***: Site 1 is more similar to site 2 than site 10.
>
> ***Answer 3c***: If you defined beta additively (i.e., $\beta = \gamma - \alpha$) then beta would be 14.5 and the interpretation would change as beta and alpha would be closer in value than before.

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 similarity metrics treat abundant and rare species equally, whereas abundance-based metrics do not. For example, the Morisita-Horn index is influenced by abundant species.

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
# Remove site 8 from data (no fish were observed)
fish = fish[-8, ]
# Calculate Jaccard
fish.dj = vegdist(fish, method = "jaccard", binary = TRUE)
# Calculate Sørensen
fish.ds = vegdist(fish, method = "bray", binary = TRUE)
## This should be equal to Sorensen index (binary Bray-Curtis in
## vegan)
# Calculate Bray-Curtis
fish.db = vegdist(fish, method = "bray")
fish.db = vegdist(fish, method = "bray", upper = TRUE, diag = TRUE)
fish.ds = vegdist(fish, method = "bray", binary = TRUE, upper = TRUE, diag = TRUE)
```

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 resemblance matrix (`fish.db`) represents disimilarity. The Bray–Curtis dissimilarity index ranges between 0 and 1, where 0 means that the sites share all the species and 1 means the two sites do not share any species. When you generate the square resemblence matrix you can see the resemblance of each site to itself and those values are equal to 0 meaning that Bray-Curtis represents disimilarity.

Answer 5b:Sørensen tends to be used when working with incidence data and Bray Curtis is used with Abundance data. The Sørensen and Bray-Curtis distance values vary slightly but both represent disimilarity.Your choice of Sørensen or Bray-Curtis distance can influence your interpretation of site, because Sørensen tends to put more emphasis on similarity of samples and values tend to be lower for Sørensen, you may interpret sites as being more similar.

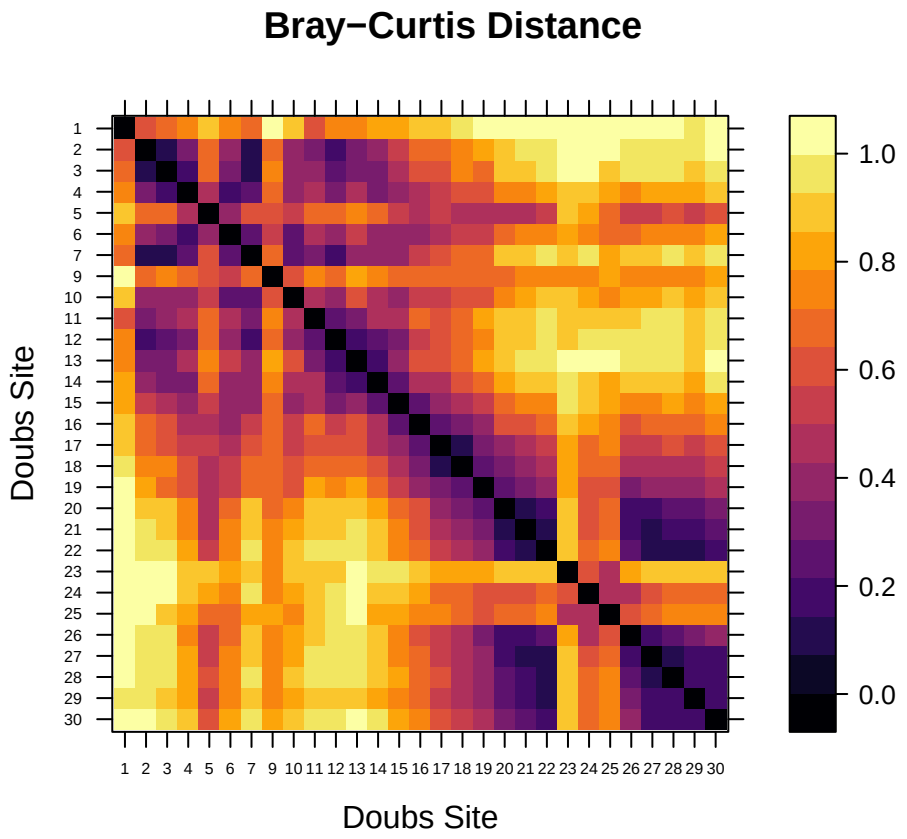
4) VISUALIZING BETA-DIVERSITY

A. Heatmaps

In the R code chunk below, do the following:

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

```
# Using the inferno color pallette
# Defining the Order of the sites
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")
```



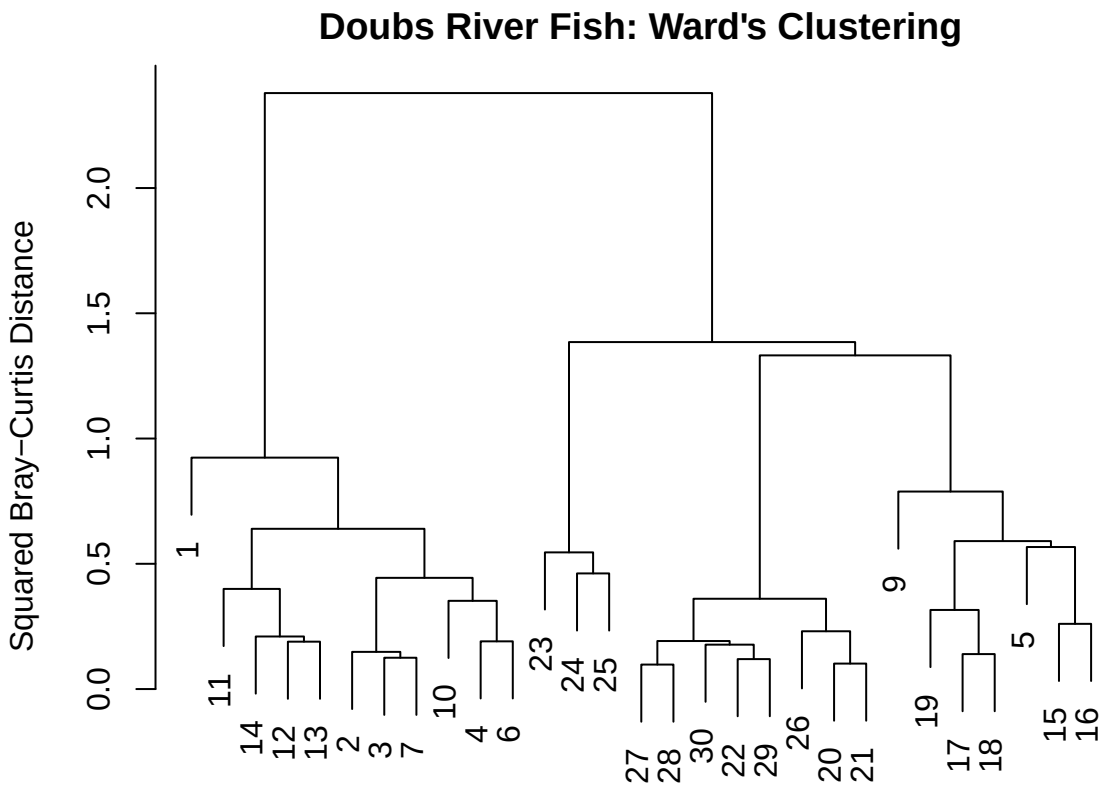

```
# It looks as though upstream sites within the river tend to be more similar with respect to richness c
```

B. Cluster Analysis

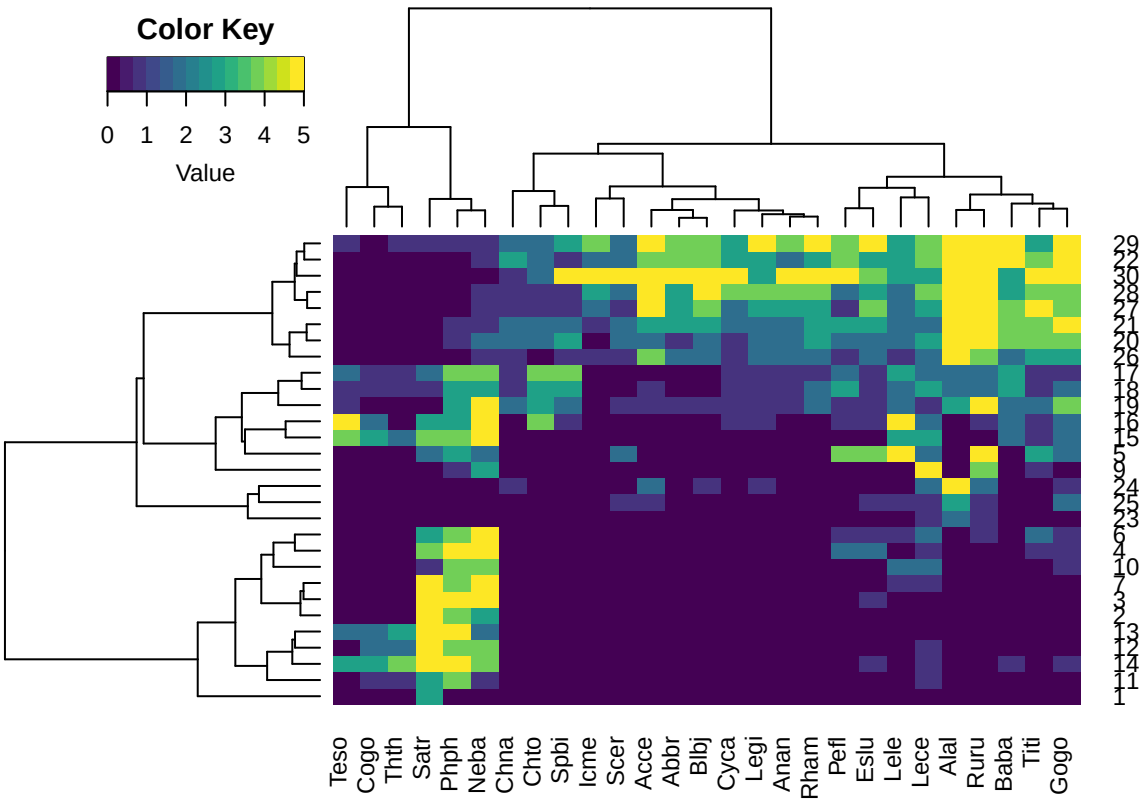
In the R code chunk below, do the following:

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




```
# Pairs are grouped according to similarity

# Heatmap that shows the abundance of different fish species at differnt sites
gplots::heatmap.2(as.matrix(fish), distfun = function(x) vegdist(x, method = "bray"),
hclustfun = function(x) hclust(x, method = "ward.D2"),
col = viridis, trace = "none", density.info = "none")
```



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

> ***Answer 6***:H1- Upstream sites have lower diversity and species abundances compared to mid-stream and downstream sites due to habitat quality.

C. Ordination

Principal Coordinates Analysis (PCoA)

Calling in the add species scores function because the BiodiversityR package would not load. The add.spec.scores() function in the BiodiversityR package produces species coordinates in ordination space with reflect the strength and direction each species has on ordination at the different sites.

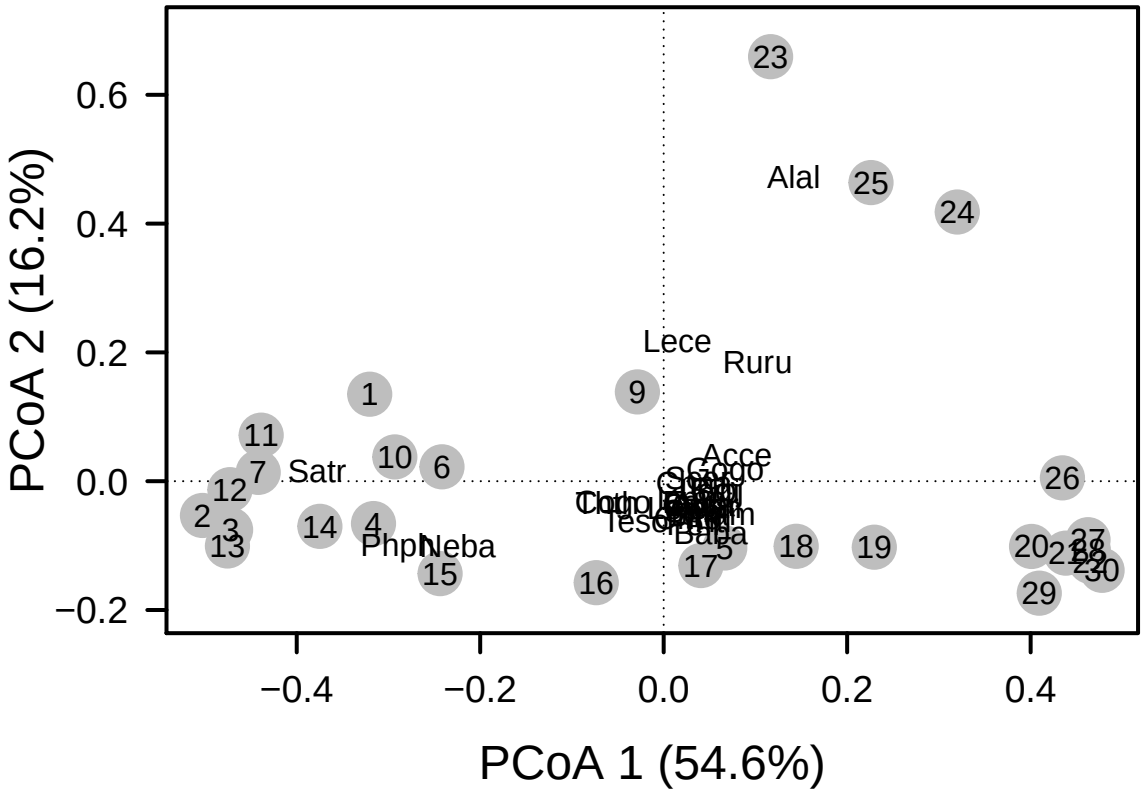
```
`add.spec.scores` <-function(ordi,comm,method="cor.scores",multi=1,Rscale=F,scaling="1") {
    ordiscores <- scores(ordi,display="sites")
    n <- ncol(comm)
    p <- ncol(ordiscores)
    specscores <- array(NA,dim=c(n,p))
    rownames(specscores) <- colnames(comm)
    colnames(specscores) <- colnames(ordiscores)
    if (method == "cor.scores") {
      for (i in 1:n) {
        for (j in 1:p) {specscores[i,j] <- cor(comm[,i],ordiscores[,j],method="pearson")}
      }
    }
    if (method == "wa.scores") {specscores <- wascores(ordiscores,comm)}
```

```
  if (method == "pcoa.scores") {
    rownames(ordiscores) <- rownames(comm)
    eigenv <- ordi$eig
    accounted <- sum(eigenv)
    tot <- 2*(accounted/ordi$GOF[2])-(accounted/ordi$GOF[1])
    eigen.var <- eigenv/(nrow(comm)-1)
    neg <- length(eigenv[eigenv<0])
    pos <- length(eigenv[eigenv>0])
    tot <- tot/(nrow(comm)-1)
    eigen.percen <- 100*eigen.var/tot
    eigen.cumpercen <- cumsum(eigen.percen)
    constant <- ((nrow(comm)-1)*tot)^0.25
    ordiscores <- ordiscores * (nrow(comm)-1)^-0.5 * tot^-0.5 * constant
    p1 <- min(p, pos)
    for (i in 1:n) {
      for (j in 1:p1) {
        specscores[i,j] <- cor(comm[,i],ordiscores[,j])*sd(comm[,i])/sd(ordiscores[,j])
        if(is.na(specscores[i,j])) {specscores[i,j]<-0}
      }
    }
    if (Rscale==T && scaling=="2") {
      percen <- eigen.var/tot
      percen <- percen^0.5
      ordiscores <- sweep(ordiscores,2,percen,"/")
      specscores <- sweep(specscores,2,percen,"*")
    }
    if (Rscale==F) {
      specscores <- specscores / constant
      ordiscores <- ordi$points
    }
    ordi$points <- ordiscores
    ordi$eig <- eigen.var
    ordi$eig.percen <- eigen.percen
    ordi$eig.cumpercen <- eigen.cumpercen
    ordi$eigen.total <- tot
    ordi$R.constant <- constant
    ordi$Rscale <- Rscale
    ordi$scaling <- scaling
  }
  specscores <- specscores * multi
  ordi$cproj <- specscores
  return(ordi)
}
```

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.

```r
# Perform a Principal Coordinates Analysis to visualize beta-diversity
fish.pcoa = cmdscale(fish.db, eig = TRUE, k = 3)

# Calculate the variation explained by the first three axes in your ordination
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)


# plot the PCoA ordination
# Define Plot Parameters
par(mar = c(5, 5, 1, 2) + 0.1)
# Initiate Plot
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)
# Add Axes
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)

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

# Identify influential species

# First we calculate the relative abundances of each species at each site
fishREL = fish
for(i in 1:nrow(fish)){
fishREL[i, ] = fish[i, ] / sum(fish[i, ])
}
# Now, we use this information to calculate and add species scores
fish.pcoa = add.spec.scores(fish.pcoa,fishREL,method = "pcoa.scores")
text(fish.pcoa$cproj[ ,1], fish.pcoa$cproj[ ,2],
labels = row.names(fish.pcoa$cproj), col = "black")
```



In the R code chunk below, do the following:

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

```
#identify influential species and add species coordinates to PCoA plot.
spe.corr = add.spec.scores(fish.pcoa, fishREL, method = "cor.scores")$cproj
corrcut = 0.7 # user defined cutoff
imp.spp = spe.corr[abs(spe.corr[, 1]) >= corrcut | abs(spe.corr[, 2]) >= corrcut, ]
# Permutation Test for Species Abundances Across Axes
fit = envfit(fish.pcoa, fishREL, perm = 999)
fit
```

```
##
## ***VECTORS
##
##          Dim1     Dim2     r2 Pr(>r)
## Cogo -0.83884 -0.54438 0.2982  0.015 *
## Satr -0.99904  0.04371 0.4326  0.004 **
## Phph -0.94110 -0.33813 0.7814  0.001 ***
## Neba -0.91413 -0.40543 0.6234  0.001 ***
## Thth -0.87692 -0.48063 0.2634  0.022 *
## Teso -0.44704 -0.89452 0.1700  0.090 .
## Chna  0.99707 -0.07644 0.4612  0.001 ***
## Chto  0.42032 -0.90738 0.2579  0.021 *
## Lele  0.33041 -0.94384 0.0495  0.551
## Lece  0.06856  0.99765 0.3399  0.008 **
## Baba  0.54118 -0.84091 0.6752  0.001 ***
```

```
## Spbi  0.57341 -0.81927 0.4138  0.004 **
## Gogo  0.97507  0.22188 0.3753  0.003 **
## Eslu  0.72044 -0.69352 0.1673  0.105
## Pefl  0.43762 -0.89916 0.3048  0.012 *
## 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
```

```
spc = doubs$species
```

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: Sites are grouped along the PCoA axes with respect to the variation in the abundance of different fish species which likely corresponds to their location either upstream, midstream or downstream in the River Doubs. Thus, upstream sites have more similar abundances of fish species and similar species composition compared to other regions. PCoA1 likely corresponds to an environmental variable associated with different stream regions (streamflow, temperature, water quality etc.)

Answer 7b:The following species can be used as indicator species for river quality: *Phoxinus phoxinus*, *Nemacheilus barbatulus*, *Rhodeus amarus*, *Lepomis gibbosus*, *Cyprinus carpio*, *Abramis brama*, *Acerina cernua*, *Blicca bjoerkna*, *Alburnus alburnus*, and *Anguilla anguilla*.

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.

```r
# Earlier work done in the Doubs River suggests that the river has 4 distinct regions of habitat qualit

# Create "Factors" vector for categorizing habitat quality in the Doubs River
quality = c(rep("HQ", 13), rep("MQ", 5), rep("LQ", 6), rep("MQ", 5))
# Run PERMANOVA with adonis function
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
```

```r
# Calculating indicator values to for each species in each habitat group- how many individual species r
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   0.016 *
##
##  Group HQ+MQ  #sps.  2
##       stat p.value
## Satr 0.860   0.003 **
## Phph 0.859   0.009 **
##
##  Group LQ+MQ  #sps.  20
##       stat p.value
```

```
## Alal 0.935   0.001 ***
## Gogo 0.933   0.001 ***
## Ruru 0.916   0.001 ***
## Legi 0.901   0.001 ***
## Baba 0.895   0.001 ***
## Chna 0.866   0.002 **
## Spbi 0.866   0.001 ***
## Cyca 0.866   0.002 **
## Acce 0.866   0.002 **
## Lele 0.863   0.002 **
## Titi 0.853   0.001 ***
## Chto 0.829   0.002 **
## Rham 0.829   0.002 **
## Anan 0.829   0.002 **
## Eslu 0.827   0.021 *
## Pefl 0.806   0.007 **
## Blbj 0.791   0.003 **
## Scer 0.766   0.006 **
## Abbr 0.750   0.006 **
## Icme 0.661   0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Habitat preferences of each species. In this case, we can calculate the phi coefficient of associatio
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   0.027 *
##
##  Group MQ  #sps.  4
```

```
##       stat p.value
## Anan 0.571   0.012 *
## Spbi 0.557   0.009 **
## Chto 0.542   0.007 **
## Icme 0.475   0.032 *
##
##  Group LQ+MQ  #sps.  9
##       stat p.value
## Legi 0.658   0.003 **
## Baba 0.645   0.001 ***
## Rham 0.600   0.004 **
## Acce 0.594   0.006 **
## Cyca 0.586   0.006 **
## Chna 0.571   0.009 **
## Blbj 0.571   0.010 **
## Gogo 0.523   0.015 *
## Abbr 0.499   0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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***: Fish community composition varies significantly with river quality (R2= 0.45, P = 0.001). Certain species have preferences for certain quality levels. For example, *Phoxinus phoxinus*, *Nemacheilus barbatulus*, and *Salmo trutta fario* prefer high quality sites. Whereas, *Alburnus alburnus* and *Rutilus rutilus* prefer low quality sites. The following species can be used as indicator species for high-medium quality sites, *Phoxinus phoxinus* and *Salmo trutta fario*. *Telestes soufia agassizi* is an indicator species for medium quality sites. The remaining species are indicators for low quality sites.

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.

```
# Distance matrices for both fish communites and environmental factors
fish.dist = vegdist(doubs$fish[-8, ], method = "bray")
env.dist = vegdist(scale(doubs$env[-8,]),method = "euclid")

# Mantel Test
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%
## 0.0922 0.1282 0.1532 0.1895
## Permutation: free
## Number of permutations: 999
```

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?

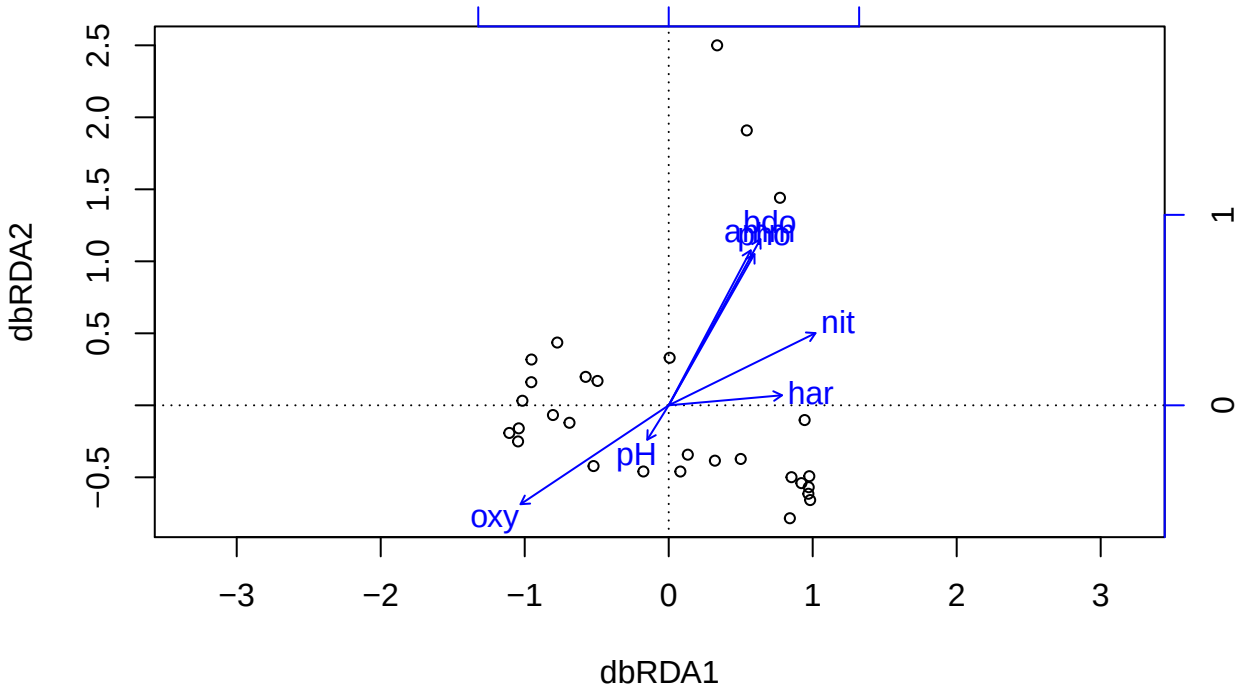
> ***Answer 9***: Yes, fish assemblages are correlated with stream environmental variables (r = 0.60, P = 0.001). The results from the Mantel test suggest that fish diversity is influenced by stream environmental conditions. These results support my previous hypothesis about stream quality influencing fish species diversity and abundances.

ii. Constrained Ordination

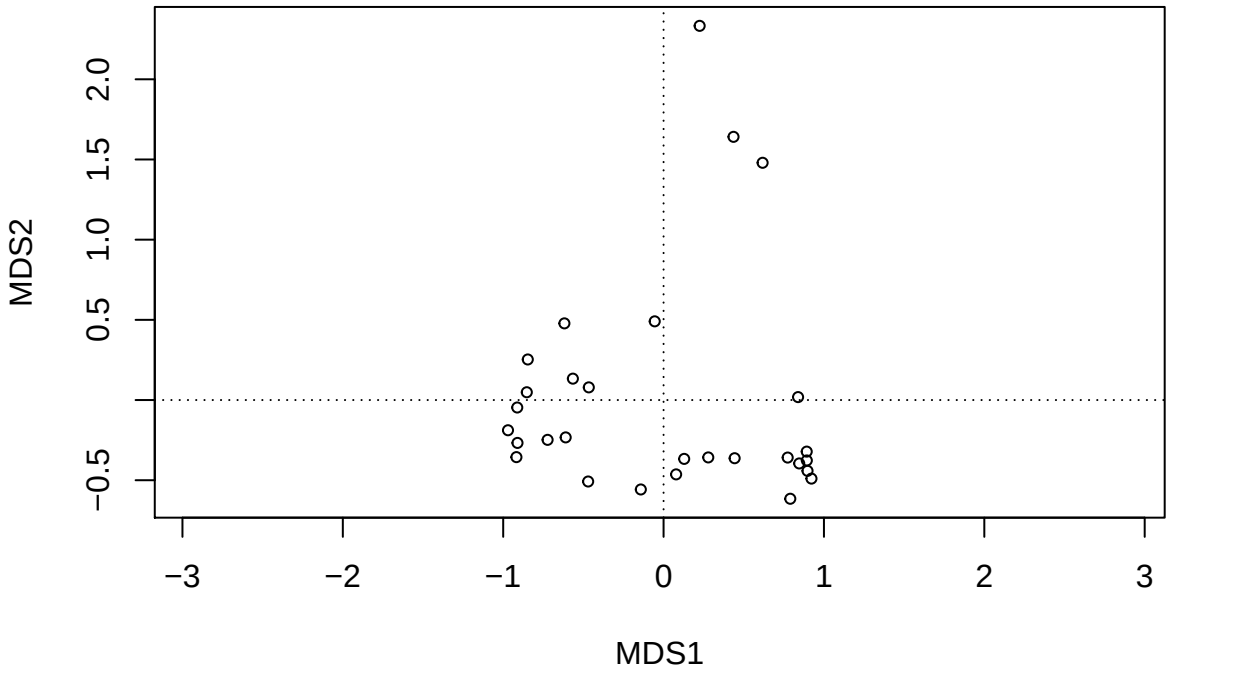
In the R code chunk below, do the following:

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.

```
# Define environmental matrix
env.chem = as.matrix(doubs$env[-8 , 5:11])
# Perform dbRDA (Redundancy analysis)
doubs.dbrda = dbrda(fish.db ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda)
```




```
# First, we will model only the intercept
doubs.dbrda.mod0 = dbrda(fish.db ~ 1, as.data.frame(env.chem))
# Note there are no vectors here (we didn't constrain anything)
# Therefore, the axes suggest this is a simple MDS (i.e., PCoA)
ordiplot(doubs.dbrda.mod0)
```




```
# Next, we will model the full model, with all explanatory variables
doubs.dbrda.mod1 = dbrda(fish.db ~ ., as.data.frame(env.chem))
# Now we step through all combinations of explanatory variables in our model
```

```r
# The function returns the model with the lowest AIC value
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
```

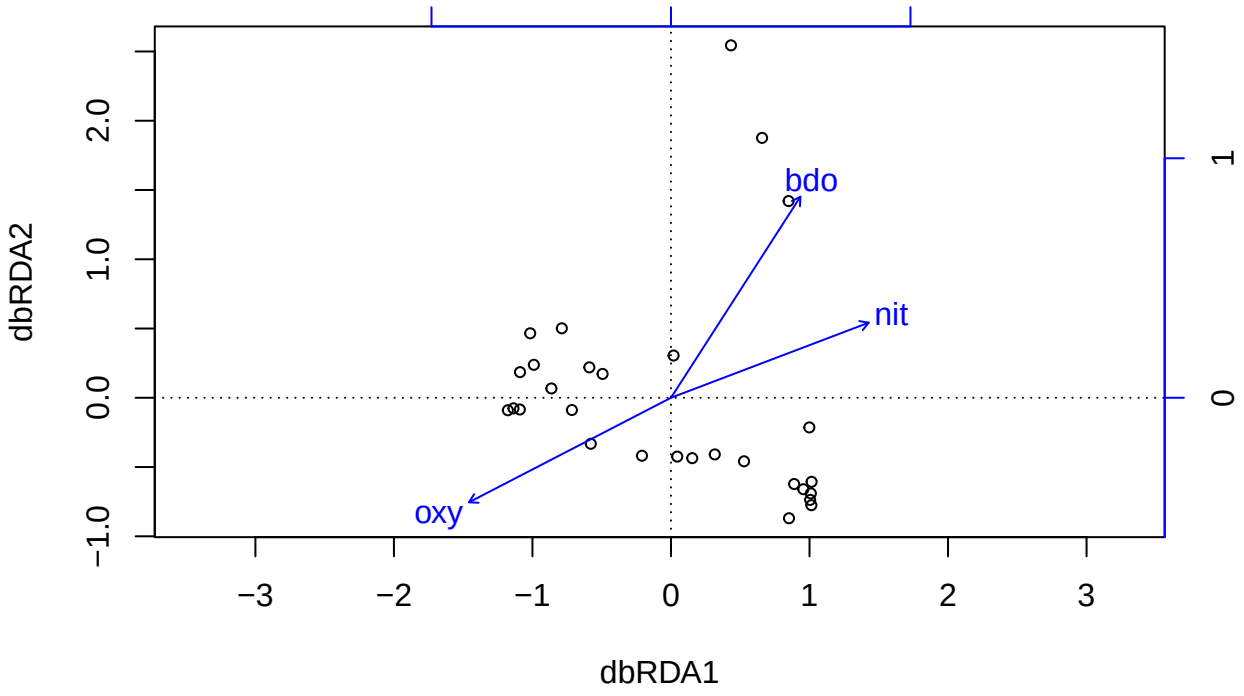
```
# Lets look at the model that was selected
doubs.dbrda$call
```

```
## 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)
```



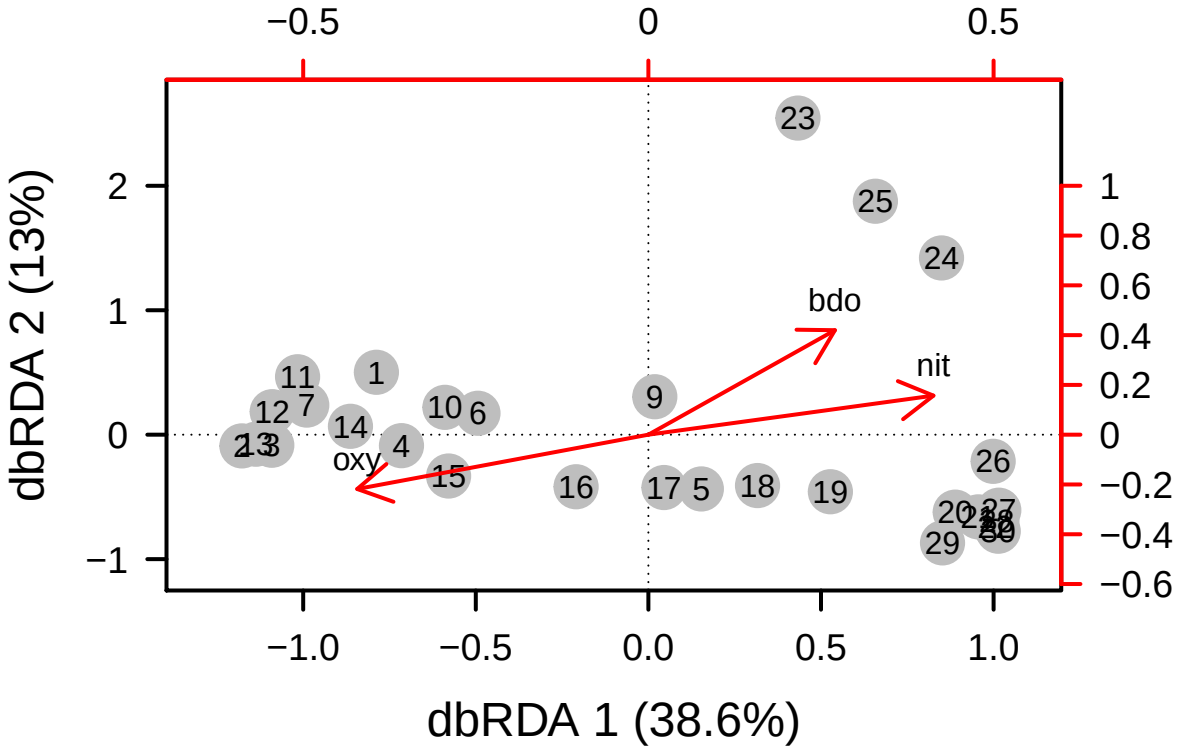

```
# Permutation tests to evaluate significance
permutest(doubs.dbrda, permutations = 999)
```

```
##
## 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
```

```r
# Calculate Explained Variation
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

# Now, let's plot the ordination for the selected model.
# Define Plot Parameters
par(mar = c(5, 5, 4, 4) + 0.1)
# Initiate Plot
plot(scores(doubs.dbrda, display = "wa"), xlim = c(-1.3, 1.1), ylim = c(-1.1, 2.7),
xlab = paste("dbRDA 1 (", dbrda.explainvar1, "%)", sep = ""),
ylab = paste("dbRDA 2 (", dbrda.explainvar2, "%)", sep = ""),
pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE)
# Add Axes
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)
# Add Points & Labels
points(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")))
# Add Environmental Vectors
vectors = scores(doubs.dbrda, display = "bp")
#row.names(vectors) = rownames(vectors)
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***: The following environmental variables are contributing to variation in fish community structure include: Dissolved oxygen [mgL-1], Biological oxygen demand [mgL-1], and Nitrate concentration [mgL-1].

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.

```
# Remember, our environmental model uses oxy, bdo, and nit and has R2 of 0.53
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
```

```r
# Let's create a matrix model for our environmental data
env.mod = model.matrix(~ oxy + bdo + nit, as.data.frame(env.chem))[,-1]
# First, we will weight each site by its relative abundance
rs = rowSums(fish)/sum(fish)
# Next, we will perform PCNM
doubs.pcnmw = pcnm(dist(doubs$xy[-8,]), w = rs, dist.ret = T)
# PCNM can return negative eigenvalues, but only the
# eigenvectors associated with the positive eigenvalues are meaningful
doubs.pcnmw$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
```

```r
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
```

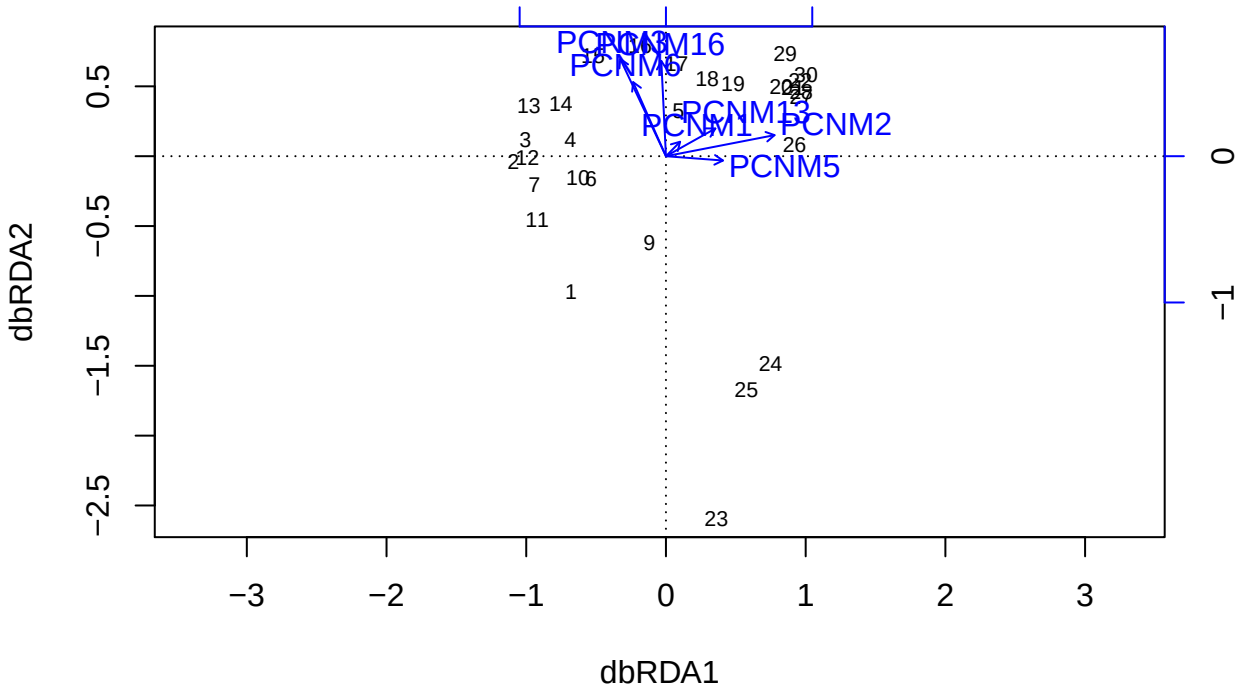
```
## <All variables>   0.6260113
## + PCNM3           0.3429270
## + PCNM5           0.3057368
## + PCNM1           0.2885396
## + PCNM16          0.2786746
## + PCNM14          0.2744520
## + PCNM15          0.2692809
## + PCNM6           0.2659866
## + PCNM13          0.2636194
## + PCNM9           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   0.02 *
## ---
## 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  0.012 *
## ---
## 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  0.014 *
## ---
## 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  0.042 *
## ---
## 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   0.08 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Because this is another dbRDA, we could visualize the biplot
# showing how each vector explains variation across sites
plot(step.pcnm)
```




```
# The object `step.pcnm` now contains the selected model.
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  0.020 *
## + PCNM1         0.47215  1 41.411 4.0570  0.012 *
## + PCNM13        0.52124  1 39.346 3.4612  0.010 **
## + PCNM16        0.57680  1 36.480 4.0192  0.014 *
## + PCNM6         0.60431  1 35.182 2.5296  0.042 *
## <All variables> 0.62601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# We can now construct a spatial model using only the selected PCNM axes.
space.mod = model.matrix(~ PCNM2 + PCNM3 + PCNM5 + PCNM1 +
PCNM13 + PCNM16 + PCNM6, doubs.space)[,-1]
# First conduct constrained ordinations
doubs.total.env = dbrda(fish.db ~ env.mod)
doubs.total.space = dbrda(fish.db ~ space.mod)
# Next construct partial constrained ordinations
doubs.env.cond.space = dbrda(fish.db ~ env.mod + Condition(space.mod))
doubs.space.cond.env = dbrda(fish.db ~ space.mod + Condition(env.mod))
# Next test for significance of the dbRDA fractions.
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
```

```
dev.off
```

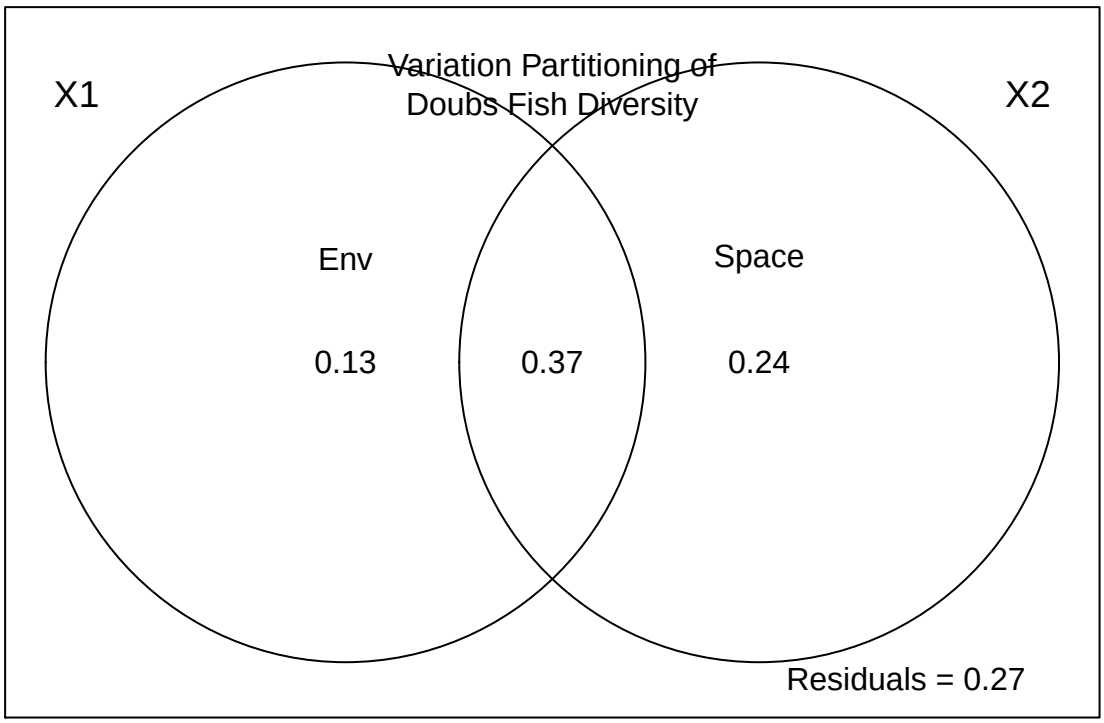
```
## function (which = dev.cur())
## {
```

```
##     if (which == 1)
##         stop("cannot shut down device 1 (the null device)")
##     .External(C_devoff, as.integer(which))
##     dev.cur()
## }
## <bytecode: 0x7fdb4ca034b0>
## <environment: namespace:grDevices>
```

```
# Using the built-in varpart() function
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***: Around 13% of the variation in fish diversity in the Doubs River can be attributed to environmental variables, most notably, biological oxygen demand and dissolved oxygen & nitrate concentrations. Almost a quarter of the variation in fish diversity (24%) can be attributed to changes in spatial positionalong the river. Spatial & environmental factors explain about 37% of the variation in fish species diversity. Around 27% of the variaiton is unexplained. Spatial position explains ~2x more than environmental variables when assessed independently, however, they explain the most variation (37%) together. It seems as though spatial and environmental variation together are driving fish community strucutre in the River Doubs.

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.

```
# Loading required packages
require("dplyr")
```

```
## Loading required package: dplyr
```

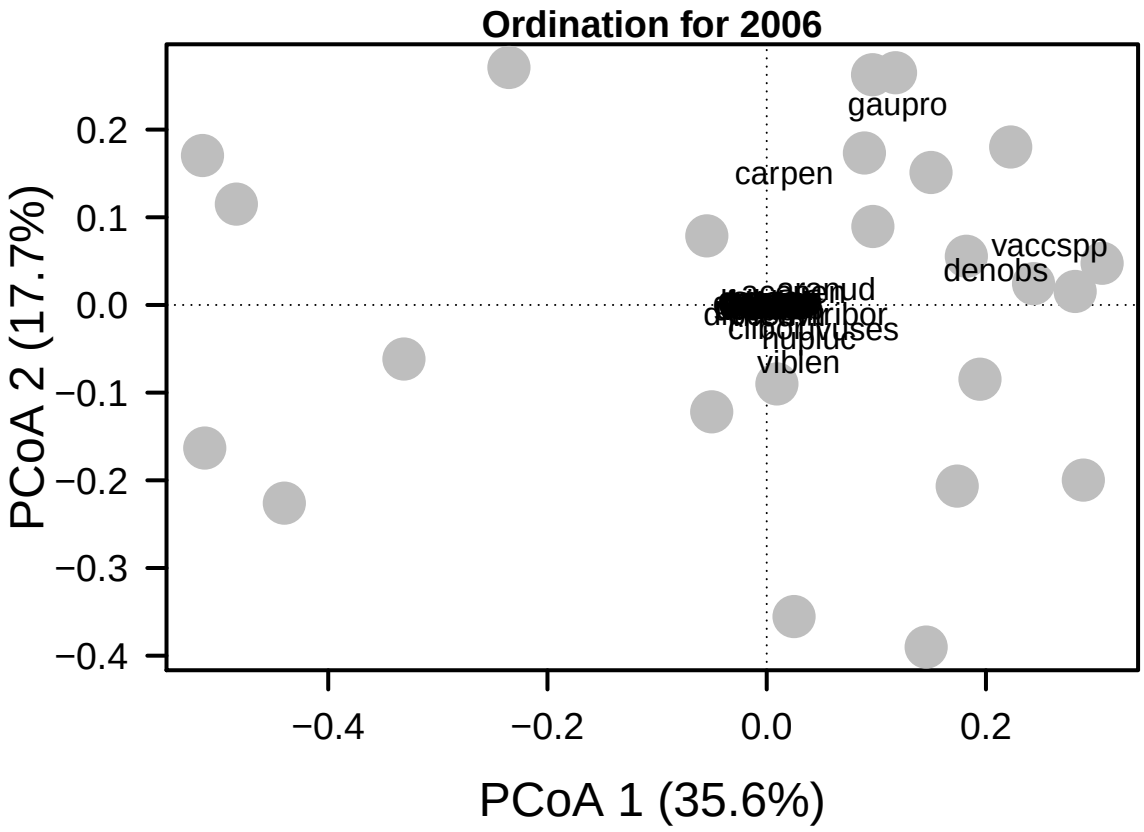
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Reading in tree diversity data
plant = read.csv("/Users/bhbeidler/GitHub/QB2017_DivPro/Data/HF_plants.csv")

# Subsetting the data into the different years
plant_06 = filter(plant, year == 2006)
plant_07 = filter(plant, year == 2007)
plant_08 = filter(plant, year == 2008)
plant_09 = filter(plant, year == 2009)

# Separating out the treatments from the site by species matricies
plant_06_sbys = plant_06[ ,4:43]
plant_07_sbys = plant_07[ ,4:43]
plant_08_sbys = plant_08[ ,4:43]
plant_09_sbys = plant_09[ ,4:43]

# Ordination

# construct a resemblance matrix based on Bray-Curtis Distance ("plant_2006.db") for 2006
plant_06.db = vegdist(plant_06_sbys, method = "bray", upper = TRUE, diag = TRUE)
# Perform a Principal Coordinates Analysis to visualize beta-diversity
plant_06.pcoa = cmdscale(plant_06.db, eig = TRUE, k = 3)

# Calculate the variation explained by the first three axes in your ordination
explainvar1 = round(plant_06.pcoa$eig[1] / sum(plant_06.pcoa$eig), 3) * 100
explainvar2 = round(plant_06.pcoa$eig[2] / sum(plant_06.pcoa$eig), 3) * 100
explainvar3 = round(plant_06.pcoa$eig[3] / sum(plant_06.pcoa$eig), 3) * 100
sum.eig = sum(explainvar1, explainvar2, explainvar3)


# plot the PCoA ordination
# Define Plot Parameters
par(mar = c(5, 5, 1, 2) + 0.1)
# Initiate Plot
plot(plant_06.pcoa$points[ ,1], plant_06.pcoa$points[ ,2],
xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
main = "Ordination for 2006",
pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE)
# Add Axes
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)

# Add Points & Labels
points(plant_06.pcoa$points[ ,1], plant_06.pcoa$points[ ,2],
pch = 16, cex = 3, bg = "gray", col = "gray")
text(plant_06.pcoa$points[ ,1], plant_06.pcoa$points[ ,2],
labels = row.names(plant_06.pcoa$points))
```

```
# Identify influential species
# First we calculate the relative abundances of each species at each site
plant_06REL = plant_06_sbys
for(i in 1:nrow(plant_06_sbys )){
plant_06REL[i, ] = plant_06_sbys [i, ] / sum(plant_06_sbys [i, ])
}
# Now, we use this information to calculate and add species scores
plant_06.pcoa= add.spec.scores(plant_06.pcoa,plant_06REL,method = "pcoa.scores")
text(plant_06.pcoa$cproj[ ,1], plant_06.pcoa$cproj[ ,2],
labels = row.names(plant_06.pcoa$cproj), col = "black")
```




```
# Hypothesis Testing
# H1: Soil treatment (either warming or N fertilization) influenced beta diversity during 2006.

# Multivariate Procedures for Categorical designs

# There are 40 tree species. For each year and each treatment there are 6 sites. Treatments include: 1=
# Create "Factors" vector for categorizing soil treatments
soil_treat = as.factor(plant_06[ ,3])

# Run PERMANOVA with adonis function
PERMA.06 = adonis(plant_06_sbys ~ soil_treat, method = "bray", permutations = 999)
PERMA.07 = adonis(plant_07_sbys ~ soil_treat, method = "bray", permutations = 999)
PERMA.08 = adonis(plant_08_sbys ~ soil_treat, method = "bray", permutations = 999)
PERMA.09 = adonis(plant_09_sbys ~ soil_treat, method = "bray", permutations = 999)

PERMA.06
```

```
##
## Call:
## adonis(formula = plant_06_sbys ~ soil_treat, permutations = 999,      method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##            Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## soil_treat  3    0.2292 0.076403 0.34078 0.04863  0.997
## Residuals  20    4.4840 0.224198         0.95137
## Total      23    4.7132                 1.00000
```

```
PERMA.07
```

```
##
## Call:
## adonis(formula = plant_07_sbys ~ soil_treat, permutations = 999,      method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##            Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## soil_treat  3    0.2948 0.098282 0.40013 0.05662  0.997
## Residuals  20    4.9125 0.245626         0.94338
## Total      23    5.2074                 1.00000
```

```
PERMA.08
```

```
##
## Call:
## adonis(formula = plant_08_sbys ~ soil_treat, permutations = 999,      method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##            Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## soil_treat  3    0.4008 0.13361 0.59238 0.08161  0.942
## Residuals  20    4.5110 0.22555         0.91839
## Total      23    4.9118                 1.00000
```

```
PERMA.09
```

```
##
## Call:
## adonis(formula = plant_09_sbys ~ soil_treat, permutations = 999,      method = "bray")
##
## Permutation: free
```

```
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##            Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## soil_treat  3    0.5218 0.17393 0.85599 0.11379  0.655
## Residuals  20    4.0637 0.20319         0.88621
## Total      23    4.5855                 1.00000
```

Answer Synthesis: Testing to see if the warming or N treatments influence beta diversity. I used a PERMANOVA to test the hypothesis that tree community composition varies with soil treatment. There was no siginificant effect of soil treatment on beta diversity for any of the years. To perform the PERMANOVA I had to subset the data into the different years. To test for the effects of both time and soil treatments on tree community composition we could perform a RDA-partitioning the variance between treatments and time. The study may have been too short to capture any changes in diversity. There is also a companion dataset for this study that looks at ant diversity in the same plots over the same time period. The next step will be to look at ant diversity in addition to tree diversity.

‘‘‘