

# Assignment: Among Site (Beta) Diversity

Roy Z Moger-Reischer; Z620: Quantitative Biodiversity, Indiana University

28 January, 2017

## OVERVIEW

In this Assignment, we move beyond the investigation of within-site  $\alpha$ -diversity. We will explore  $\beta$ -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  $\beta$ -diversity
2. visualize  $\beta$ -diversity with heatmaps, cluster analysis, and ordination
3. test hypotheses about  $\beta$ -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 with 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] "C:/Users/rmoger/GitHub/QB2017_Moger-Reischer/Week3-Beta"
```

```
#setwd("/Users/rzmogerr/GitHub/QB2017_Moger-Reischer/Week3-Beta")
```

```
#I'll load all of the packages, following the style of the handout
```

```
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)
  }
}

## 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

## BiodiversityR 2.8-0: Use command BiodiversityRGUI() to launch the Graphical User Interface and to le

```

## 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)

## List of 4
## $ env      : 'data.frame': 30 obs. of  11 variables:
##  ..$ dfs: num [1:30] 3 22 102 185 215 324 268 491 705 990 ...
##  ..$ alt: num [1:30] 934 932 914 854 849 846 841 792 752 617 ...
##  ..$ slo: num [1:30] 6.18 3.43 3.64 3.5 3.18 ...
##  ..$ flo: num [1:30] 84 100 180 253 264 286 400 130 480 1000 ...
##  ..$ pH : num [1:30] 79 80 83 80 81 79 81 81 80 77 ...
##  ..$ har: num [1:30] 45 40 52 72 84 60 88 94 90 82 ...
##  ..$ pho: num [1:30] 1 2 5 10 38 20 7 20 30 6 ...
##  ..$ nit: num [1:30] 20 20 22 21 52 15 15 41 82 75 ...
##  ..$ amm: num [1:30] 0 10 5 0 20 0 0 12 12 1 ...
##  ..$ oxy: num [1:30] 122 103 105 110 80 102 111 70 72 100 ...
##  ..$ bdo: num [1:30] 27 19 35 13 62 53 22 81 52 43 ...
## $ fish     : 'data.frame': 30 obs. of  27 variables:
##  ..$ Cogo: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##  ..$ Satr: num [1:30] 3 5 5 4 2 3 5 0 0 1 ...
##  ..$ Phph: num [1:30] 0 4 5 5 3 4 4 0 1 4 ...
##  ..$ Neba: num [1:30] 0 3 5 5 2 5 5 0 3 4 ...
##  ..$ Thth: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
##  ..$ Teso: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...

```

```
## ..$ Chna: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Chto: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Lele: num [1:30] 0 0 0 0 5 1 1 0 0 2 ...
## ..$ Lece: num [1:30] 0 0 0 1 2 2 1 0 5 2 ...
## ..$ Baba: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Spbi: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Gogo: num [1:30] 0 0 0 1 2 1 0 0 0 1 ...
## ..$ Eslu: num [1:30] 0 0 1 2 4 1 0 0 0 0 ...
## ..$ Pefl: num [1:30] 0 0 0 2 4 1 0 0 0 0 ...
## ..$ Rham: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Legi: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Scer: num [1:30] 0 0 0 0 2 0 0 0 0 0 ...
## ..$ Cyca: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Titi: num [1:30] 0 0 0 1 3 2 0 0 1 0 ...
## ..$ Abbr: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Icme: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Acce: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Ruru: num [1:30] 0 0 0 0 5 1 0 0 4 0 ...
## ..$ Blbj: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Alal: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## ..$ Anan: num [1:30] 0 0 0 0 0 0 0 0 0 0 ...
## $ xy      : 'data.frame': 30 obs. of 2 variables:
## ..$ x: num [1:30] 88 94 102 100 106 112 114 110 136 168 ...
## ..$ y: num [1:30] 7 14 18 28 39 51 61 76 100 112 ...
## $ species: 'data.frame': 27 obs. of 4 variables:
## ..$ Scientific: chr [1:27] "Cottus gobio" "Salmo trutta fario" "Phoxinus phoxinus" "Nemacheilus ba
## ..$ French      : chr [1:27] "chabot" "truite fario" "vairon" "loche franche" ...
## ..$ English     : chr [1:27] "european bullhead" "brown trout" "minnow" "stone loach" ...
## ..$ code        : Factor w/ 27 levels "Abbr","Acce",...: 9 22 19 17 26 25 7 8 16 14 ...
```

```
str(doubs$env)
```

```
## 'data.frame': 30 obs. of 11 variables:
## $ dfs: num 3 22 102 185 215 324 268 491 705 990 ...
## $ alt: num 934 932 914 854 849 846 841 792 752 617 ...
## $ slo: num 6.18 3.43 3.64 3.5 3.18 ...
## $ flo: num 84 100 180 253 264 286 400 130 480 1000 ...
## $ pH : num 79 80 83 80 81 79 81 81 80 77 ...
## $ har: num 45 40 52 72 84 60 88 94 90 82 ...
## $ pho: num 1 2 5 10 38 20 7 20 30 6 ...
## $ nit: num 20 20 22 21 52 15 15 41 82 75 ...
## $ amm: num 0 10 5 0 20 0 0 12 12 1 ...
## $ oxy: num 122 103 105 110 80 102 111 70 72 100 ...
## $ bdo: num 27 19 35 13 62 53 22 81 52 43 ...
```

```
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
## 7	268	841	4.205	400	81	88	7	15	0	111	22
## 8	491	792	3.258	130	81	94	20	41	12	70	81
## 9	705	752	2.565	480	80	90	30	82	12	72	52
## 10	990	617	4.605	1000	77	82	6	75	1	100	43
## 11	1234	483	3.738	1990	81	96	30	160	0	115	27
## 12	1324	477	2.833	2000	79	86	4	50	0	122	30
## 13	1436	450	3.091	2110	81	98	6	52	0	124	24
## 14	1522	434	2.565	2120	83	98	27	123	0	123	38
## 15	1645	415	1.792	2300	86	86	40	100	0	117	21
## 16	1859	375	3.045	1610	80	88	20	200	5	103	27
## 17	1985	348	1.792	2430	80	92	20	250	20	102	46
## 18	2110	332	2.197	2500	80	90	50	220	20	103	28
## 19	2246	310	1.792	2590	81	84	60	220	15	106	33
## 20	2477	286	2.197	2680	80	86	30	300	30	103	28
## 21	2812	262	2.398	2720	79	85	20	220	10	90	41
## 22	2940	254	2.708	2790	81	88	20	162	7	91	48
## 23	3043	246	2.565	2880	81	97	260	350	115	63	164
## 24	3147	241	1.386	2976	80	99	140	250	60	52	123
## 25	3278	231	1.792	3870	79	100	422	620	180	41	167
## 26	3579	214	1.792	3910	79	94	143	300	30	62	89
## 27	3732	206	2.565	3960	81	90	58	300	26	72	63
## 28	3947	195	1.386	4320	83	100	74	400	30	81	45
## 29	4220	183	1.946	6770	78	110	45	162	10	90	42
## 30	4530	172	1.099	6900	82	109	65	160	10	82	44

##

## \$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
## 7	0	5	4	5	0	0	0	0	1	1	0	0	0	0
## 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## 9	0	0	1	3	0	0	0	0	0	5	0	0	0	0
## 10	0	1	4	4	0	0	0	0	2	2	0	0	1	0
## 11	1	3	4	1	1	0	0	0	0	1	0	0	0	0
## 12	2	5	4	4	2	0	0	0	0	1	0	0	0	0
## 13	2	5	5	2	3	2	0	0	0	0	0	0	0	0
## 14	3	5	5	4	4	3	0	0	0	1	1	0	1	1
## 15	3	4	4	5	2	4	0	0	3	3	2	0	2	0
## 16	2	3	3	5	0	5	0	4	5	2	2	1	2	1
## 17	1	2	4	4	1	2	1	4	3	2	3	4	1	1
## 18	1	1	3	3	1	1	1	3	2	3	3	3	2	1
## 19	0	0	3	5	0	1	2	3	2	1	2	2	4	1
## 20	0	0	1	2	0	0	2	2	2	3	4	3	4	2

## 21	0	0	1	1	0	0	2	2	2	2	4	2	5	3
## 22	0	0	0	1	0	0	3	2	3	4	5	1	5	3
## 23	0	0	0	0	0	0	0	0	0	1	0	0	0	0
## 24	0	0	0	0	0	0	1	0	0	2	0	0	1	0
## 25	0	0	0	0	0	0	0	0	1	1	0	0	2	1
## 26	0	0	0	1	0	0	1	0	1	2	2	1	3	2
## 27	0	0	0	1	0	0	1	1	2	3	4	1	4	4
## 28	0	0	0	1	0	0	1	1	2	4	3	1	4	3
## 29	0	1	1	1	1	1	2	2	3	4	5	3	5	5
## 30	0	0	0	0	0	0	1	2	3	3	3	5	5	4
##	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	
## 7	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 8	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 9	0	0	0	0	0	1	0	0	0	4	0	0	0	
## 10	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 11	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 12	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 13	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 14	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 15	0	0	0	0	0	1	0	0	0	0	0	0	0	
## 16	1	0	1	0	1	1	0	0	0	1	0	0	0	
## 17	2	1	1	0	1	1	0	0	0	2	0	2	1	
## 18	3	2	1	0	1	1	0	0	1	2	0	2	1	
## 19	1	2	1	1	1	2	1	0	1	5	1	3	1	
## 20	2	3	2	2	1	4	1	0	2	5	2	5	2	
## 21	3	3	2	2	2	4	3	1	3	5	3	5	2	
## 22	4	3	3	2	3	4	4	2	4	5	4	5	2	
## 23	0	0	0	0	0	0	0	0	0	1	0	2	0	
## 24	0	0	1	0	0	0	0	0	2	2	1	5	0	
## 25	0	0	0	1	0	0	0	0	1	1	0	3	0	
## 26	1	2	2	1	1	3	2	1	4	4	2	5	2	
## 27	1	3	3	1	2	5	3	2	5	5	4	5	3	
## 28	2	4	4	2	4	4	3	3	5	5	5	5	4	
## 29	4	5	5	2	3	3	4	4	5	5	4	5	4	
## 30	5	5	3	5	5	5	5	5	5	5	5	5	5	
##														
## \$xy														
##	x	y												
## 1	88	7												
## 2	94	14												
## 3	102	18												
## 4	100	28												
## 5	106	39												
## 6	112	51												
## 7	114	61												
## 8	110	76												
## 9	136	100												
## 10	168	112												

```
## 11 186 130
## 12 205 145
## 13 222 167
## 14 228 182
## 15 252 190
## 16 266 209
## 17 245 203
## 18 225 200
## 19 206 194
## 20 189 193
## 21 187 201
## 22 192 212
## 23 192 228
## 24 179 233
## 25 145 217
## 26 91 187
## 27 65 174
## 28 49 164
## 29 27 151
## 30 8 133
##
## $species
##           Scientific           French           English code
## 1           Cottus gobio           chabot european bullhead Cogo
## 2           Salmo trutta fario       truite fario       brown trout Satr
## 3           Phoxinus phoxinus         vairon           minnow Phph
## 4           Nemacheilus barbatulus    loche franche       stone loach Neba
## 5           Thymallus thymallus        ombre           grayling Thth
## 6           Telestes soufia agassizi    blageon           blageon Teso
## 7           Chondrostoma nasus         hotu             nase Chna
## 8           Chondrostoma toxostoma      toxostome         toxostoma Chto
## 9           Leuciscus leuciscus         vandoise         common dace Lele
## 10          Leuciscus cephalus cephalus  chevaine          chub Lece
## 11           Barbus barbus barbeau fluviatile          barbel Baba
## 12           Spiralinus bipunctatus      spirlin          spirlin Spbi
## 13           Gobio gobio                goujon          gudgeon Gogo
## 14           Esox lucius                brochet          pike Eslu
## 15           Perca fluviatilis  perche fluviatile          perch Pefl
## 16           Rhodeus amarus             bouviere          bitterling Rham
## 17           Lepomis gibbosus  perche-soleil      pumpkinseed Legi
## 18          Scardinius erythropthalmus    rotengle          rudd Scer
## 19           Cyprinus carpio            carpe          carp Cyca
## 20           Tinca tinca                tanche          tench Titi
## 21           Abramis brama              breme  freshwater bream Abbr
## 22           Ictalurus melas  poisson chat    black bullhead Icme
## 23           Acerina cernua            gremille          ruffe Acce
## 24           Rutilus rutilus            gardon          roach Ruru
## 25           Blicca bjoerkna  breme bordeliere    silver bream Blbj
## 26           Alburnus alburnus          ablette          bleak Alal
## 27           Anguilla anguilla          anguille          eel Anan
```

```
print(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
```

## 2	0	5	4	3	0	0	0	0	0	0	0	0	0	0
## 3	0	5	5	5	0	0	0	0	0	0	0	0	0	1
## 4	0	4	5	5	0	0	0	0	0	1	0	0	1	2
## 5	0	2	3	2	0	0	0	0	5	2	0	0	2	4
## 6	0	3	4	5	0	0	0	0	1	2	0	0	1	1
##	Pefl	Rham	Legi	Scer	Cyca	Titi	Abbr	Icme	Acce	Ruru	Blbj	Alal	Anan	
## 1	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 2	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 3	0	0	0	0	0	0	0	0	0	0	0	0	0	
## 4	2	0	0	0	0	1	0	0	0	0	0	0	0	
## 5	4	0	0	2	0	3	0	0	0	5	0	0	0	
## 6	1	0	0	0	0	2	0	0	0	1	0	0	0	

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

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

**Answer 1a:**

It appears that `doubs` is a list of 4 substantial objects.

**Answer 1b:**

There are 27 columns in the fish dataframe. Thus, 27 spp.

**Answer 1c:**

There are 30 rows in that site-by-species matrix. Thus 30 sites.

## Visualizing the Doubs River Dataset

**Question 2:** Answer the following questions based on the spatial patterns of richness (i.e.,  $\alpha$ -diversity) and Brown Trout (*Salmo trutta*) abundance in the Doubs River.

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

**Answer 2a:**

It is perhaps a generalization, but sites tend to exhibit higher richness near mid-stream and downstream sites, compared to upstream.

**Answer 2b:**

Brown trout is more abundant at upstream sites (maybe brown trout is a weak competitor?)

**Answer 2c:**

Richness is a measure of numbers of different species. It doesn't say anything about the abundances of those species. Moreover, richness doesn't give any information about any of the individual species; rather, it gives information about the species as an aggregate. For example, even though brown trout are not present downstream (and therefore not contributing to richness), this could be a species that thrives when competitor species are absent.

## 3) QUANTIFYING BETA-DIVERSITY

In the R code chunk below, do the following:

1. write a function (`beta.w()`) to calculate Whittaker's  $\beta$ -diversity (i.e.,  $\beta_w$ ) 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  $\beta$ -diversity in the Doubs River.

```
beta.w <- function(site.by.species = "", sitenum1 = "", sitenum2 = "", pairwise = FALSE){
  # ONLY if we specify pairwise as TRUE, do this:
  if (pairwise == TRUE){
    # As a check, let's print an error if we do not provide needed arguments
    if (sitenum1 == "" | sitenum2 == "") {
      print("Error: please specify sites to compare")
      return(NA)}
    # If our function made it this far, let's calculate pairwise beta diversity
    site1 = site.by.species[sitenum1,] # Select site 1
    site2 = site.by.species[sitenum2,] # Select site 2
    site1 = subset(site1, select = site1 > 0) # Removes absences
    site2 = subset(site2, select = site2 > 0) # Removes absences
    gamma = union(colnames(site1), colnames(site2)) # Gamma species pool
    s = length(gamma) # Gamma richness
    a.bar = mean(c(specnumber(site1), specnumber(site2))) # Mean sample richness
    b.w = round(s/a.bar - 1, 3)
    return(b.w)
  }
  #if there is no pairwise specification
  else{
    SbyS.pa <- decostand(site.by.species, method = "pa") # convert to presence-absence
    S <- ncol(SbyS.pa[,which(colSums(SbyS.pa) > 0)]) # number of species in region
    a.bar <- mean(specnumber(SbyS.pa)) # average richness at each site
    b.w <- round(S/a.bar, 3)
    print(a.bar)
    return(c(b.w, a.bar, S))#also return alpha, and gamma (global richness)
  }
}

myab1<-beta.w(doubs$fish)

## [1] 12.5

my2<-beta.w(doubs$fish, 1, 2,pairwise=TRUE)
my10<-beta.w(doubs$fish, 1, 10,pairwise=TRUE)
```

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

- a. Describe how local richness ( $\alpha$ ) and turnover ( $\beta$ ) contribute to regional ( $\gamma$ ) 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  $\beta$  change if we instead defined beta additively (i.e.,  $\beta = \gamma - \alpha$ )?

**Answer 3a:**

We can use  $\gamma = (\alpha)(\beta) = (12.5)(2.16)$ , where alpha is the average richness across all the sites. Alpha is larger than beta, and we can say alpha contributes more to gamma.

```
myab1<-beta.w(doubs$fish)
```

```
myab1 [1] 2.16 12.50 27.00
```

**Answer 3b:**

```
my2<-beta.w(doubs$fish, 1, 2,pairwise=TRUE)
```



```
my10<-beta.w(doubs$fish, 1, 10,pairwise=TRUE)
```

```
my2
```

```
[1] 0.5
```

```
my10
```

```
[1] 0.714
```

This tells me that turnover is greater between site 1 and site 10 than it is between site 1 and site 2.

#### **Answer 3c:**

In this case, we would have  $\beta > \alpha$  ( $\alpha = 12.5$ ,  $\beta = 14.5$ ), and the interpretation could be that  $\alpha$  and  $\beta$  contribute about equally to  $\gamma$ .

## **The Resemblance Matrix**

In order to quantify  $\beta$ -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? > An incidence-based matrix naturally treats rare species equally—that is, where a species is has high or low richness, it still counts as a 1 (in contradistinction to a 0). For an abundance-based matrix there is a possibility that species with higher richness could be weighted differently than spp. represented by only a few individuals.

#### **Answer 4:**

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  $\tilde{S}_A$ , renson's Similarity ("fish.ds"), and
4. construct a resemblance matrix based on Bray-Curtis Distance ("fish.db").

```
fish <- doubs$fish#initialie the variable
fish <- fish[-8, ] # Remove site 8 from data bc it is empty

# Calculate Bray-Curtis
fish.db <- vegdist(fish, method = "bray")
# Calculate  $\tilde{S}_A$ , renson
fish.ds <- vegdist(fish, method = "bray", binary = TRUE)
fish.db
```

```
##           1           2           3           4           5           6
## 2  0.60000000
## 3  0.68421053 0.14285714
## 4  0.75000000 0.33333333 0.18918919
## 5  0.89189189 0.69565217 0.68000000 0.49090909
## 6  0.75000000 0.39393939 0.29729730 0.19047619 0.41818182
## 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
## 2
## 3
## 4
## 5
## 6
## 7
## 9 0.66666667
## 10 0.26666667 0.57142857
## 11 0.33333333 0.76000000 0.44000000
## 12 0.17647059 0.68750000 0.37500000 0.24137931
## 13 0.37142857 0.81818182 0.57575758 0.33333333 0.18918919
## 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
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10

```

```

## 11
## 12
## 13
## 14
## 15 0.24590164
## 16 0.44117647 0.26027397
## 17 0.50000000 0.40259740 0.26190476
## 18 0.60000000 0.46666667 0.34146341 0.13953488
## 19 0.67567568 0.56962025 0.39534884 0.31111111 0.25000000
## 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
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21 0.10169492
## 22 0.18750000 0.10447761
## 23 0.86666667 0.87878788 0.89473684
## 24 0.57746479 0.61038961 0.65517241 0.57894737
## 25 0.67164179 0.69863014 0.73493976 0.46666667 0.46153846
## 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
## 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 0.18867925
## 28 0.23893805 0.09774436
## 29 0.33846154 0.18666667 0.14649682
## 30 0.36363636 0.19736842 0.15723270 0.14772727

```

fish.ds

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

```

## 30 1.00000000 1.00000000 0.92000000 0.65517241 0.50000000 0.54838710
##          7          9          10          11          12          13
## 2
## 3
## 4
## 5
## 6
## 7
## 9 0.40000000
## 10 0.09090909 0.45454545
## 11 0.27272727 0.45454545 0.33333333
## 12 0.27272727 0.45454545 0.33333333 0.00000000
## 13 0.45454545 0.63636364 0.50000000 0.16666667 0.16666667
## 14 0.46666667 0.60000000 0.37500000 0.25000000 0.25000000 0.25000000
## 15 0.37500000 0.50000000 0.29411765 0.29411765 0.29411765 0.29411765
## 16 0.54545455 0.54545455 0.47826087 0.56521739 0.56521739 0.56521739
## 17 0.62962963 0.62962963 0.57142857 0.57142857 0.57142857 0.57142857
## 18 0.64285714 0.64285714 0.58620690 0.58620690 0.58620690 0.58620690
## 19 0.71428571 0.64285714 0.65517241 0.79310345 0.79310345 0.79310345
## 20 0.70370370 0.62962963 0.64285714 0.78571429 0.78571429 0.85714286
## 21 0.71428571 0.64285714 0.65517241 0.79310345 0.79310345 0.86206897
## 22 0.77777778 0.70370370 0.71428571 0.85714286 0.85714286 0.92857143
## 23 0.75000000 0.50000000 0.77777778 0.77777778 0.77777778 1.00000000
## 24 0.84615385 0.69230769 0.71428571 0.85714286 0.85714286 1.00000000
## 25 0.69230769 0.69230769 0.57142857 0.85714286 0.85714286 1.00000000
## 26 0.76923077 0.69230769 0.70370370 0.85185185 0.85185185 0.92592593
## 27 0.77777778 0.70370370 0.71428571 0.85714286 0.85714286 0.92857143
## 28 0.77777778 0.70370370 0.71428571 0.85714286 0.85714286 0.92857143
## 29 0.67741935 0.67741935 0.62500000 0.68750000 0.68750000 0.68750000
## 30 0.84615385 0.76923077 0.77777778 0.92592593 0.92592593 1.00000000
##          14          15          16          17          18          19
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15 0.14285714
## 16 0.33333333 0.28571429
## 17 0.37500000 0.33333333 0.12820513
## 18 0.39393939 0.35294118 0.15000000 0.02222222
## 19 0.57575758 0.52941176 0.25000000 0.15555556 0.13043478
## 20 0.62500000 0.57575758 0.28205128 0.18181818 0.15555556 0.02222222
## 21 0.63636364 0.58823529 0.30000000 0.20000000 0.17391304 0.04347826
## 22 0.68750000 0.63636364 0.33333333 0.22727273 0.20000000 0.06666667
## 23 0.84615385 0.85714286 0.80000000 0.76000000 0.76923077 0.76923077
## 24 0.77777778 0.78947368 0.68000000 0.60000000 0.54838710 0.48387097
## 25 0.66666667 0.68421053 0.60000000 0.60000000 0.54838710 0.48387097

```

```

## 26 0.67741935 0.62500000 0.36842105 0.25581395 0.22727273 0.09090909
## 27 0.68750000 0.63636364 0.33333333 0.22727273 0.20000000 0.06666667
## 28 0.68750000 0.63636364 0.33333333 0.22727273 0.20000000 0.06666667
## 29 0.50000000 0.45945946 0.25581395 0.12500000 0.10204082 0.06122449
## 30 0.74193548 0.68750000 0.36842105 0.25581395 0.22727273 0.09090909
##          20          21          22          23          24          25
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21 0.02222222
## 22 0.04545455 0.02222222
## 23 0.76000000 0.76923077 0.76000000
## 24 0.46666667 0.48387097 0.46666667 0.45454545
## 25 0.46666667 0.48387097 0.46666667 0.45454545 0.37500000
## 26 0.06976744 0.04545455 0.02325581 0.75000000 0.44827586 0.44827586
## 27 0.04545455 0.02222222 0.00000000 0.76000000 0.46666667 0.46666667
## 28 0.04545455 0.02222222 0.00000000 0.76000000 0.46666667 0.46666667
## 29 0.08333333 0.06122449 0.08333333 0.79310345 0.52941176 0.52941176
## 30 0.06976744 0.04545455 0.02325581 0.75000000 0.44827586 0.44827586
##          26          27          28          29
## 2
## 3
## 4
## 5
## 6
## 7
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21

```

```
## 22
## 23
## 24
## 25
## 26
## 27 0.02325581
## 28 0.02325581 0.00000000
## 29 0.10638298 0.08333333 0.08333333
## 30 0.04761905 0.02325581 0.02325581 0.10638298

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

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

- Does the resemblance matrix (`fish.db`) represent similarity or dissimilarity? What information in the resemblance matrix led you to arrive at your answer?
- Compare the resemblance matrices (`fish.db` or `fish.ds`) you just created. How does the choice of the Sørensen or Bray-Curtis distance influence your interpretation of site (dis)similarity?

**Answer 5a:**

I wanted to use the numbers on the 1:1 diagonal to indicate which type of resemblance matrix. That is, if the resemblance of each site to itself is 1, it is a similarity matrix; if it is 0, the matrix measure dissimilarity. However, R didn't show that diagonal without specifying `diag = TRUE`. I looked at the handout, where it is indicated that Bray-Curtis is a Dissimilarity metric. This makes sense, because the numerator is a difference. Therefore, if for a given site, variable in the numerator is subtracted from itself, the difference will equal 0, and the sum of the 0s will give a dissimilarity of 0.

**Answer 5b:** It was not easy to intuitively compare the matrices without using some sort of visualization. I couldn't detect major differences just by looking at the numerical outputs of those matrices. The two methods seemed to output comparable results.

## 4) VISUALIZING BETA-DIVERSITY

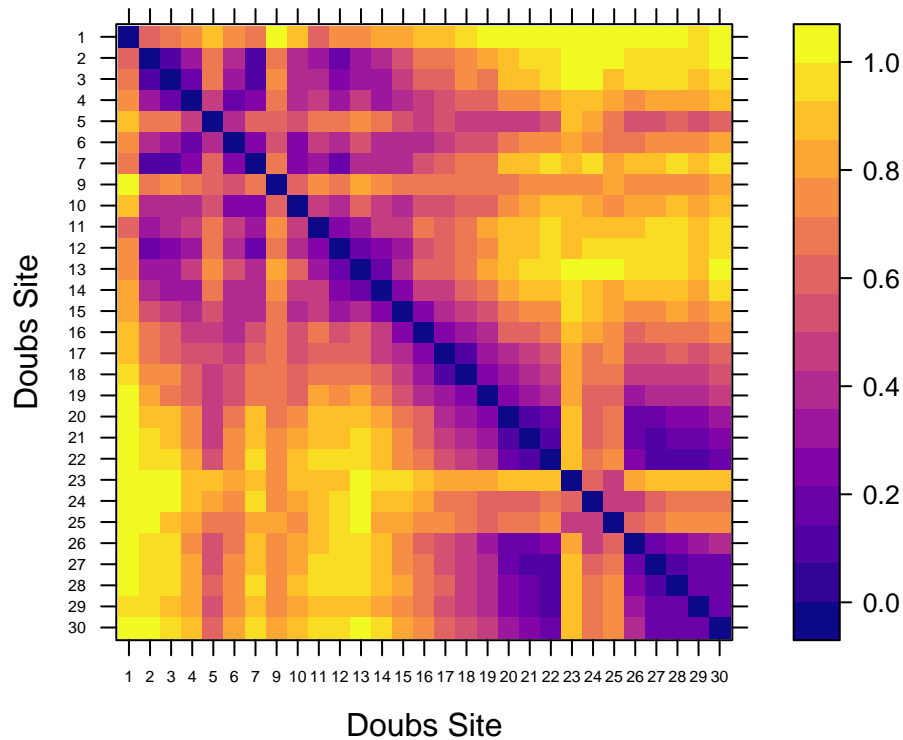
### A. Heatmaps

In the R code chunk below, do the following:

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

```
library(viridis)
order <- rev(attr(fish.db, "Labels")) #define a reversed order
levelplot(as.matrix(fish.db)[, order], aspect = "iso", col.regions = plasma, xlab = "Doubs Site", ylab = "Fish Abundance")
```

## Bray–Curtis Distance



### B. Cluster Analysis

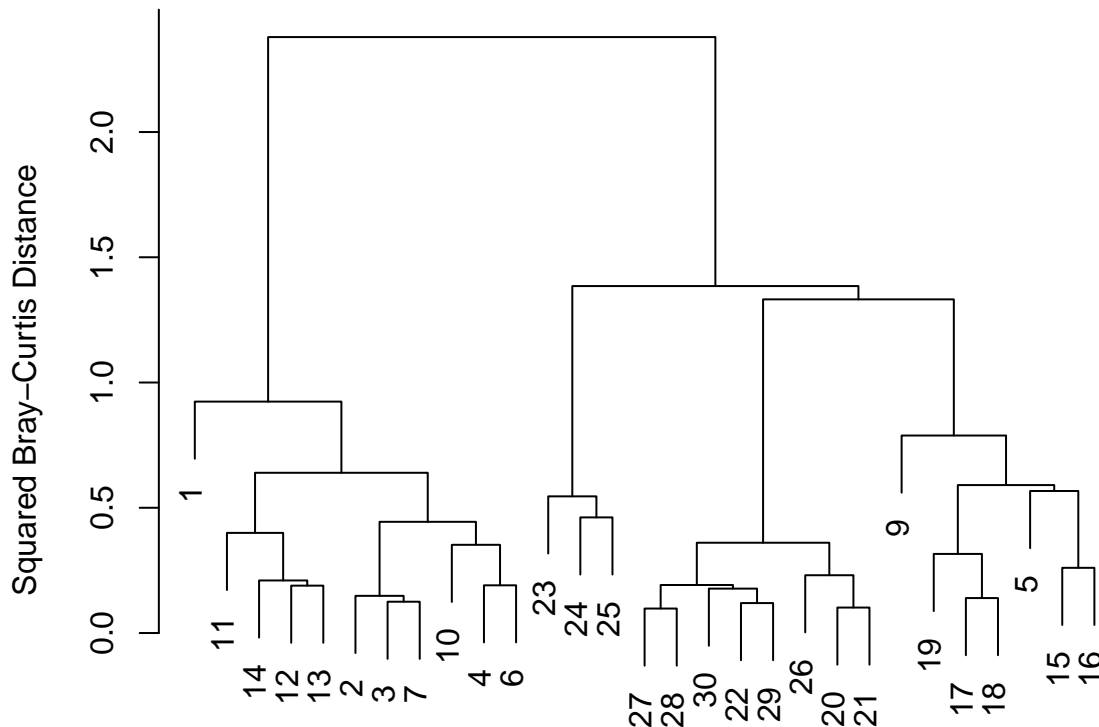
In the R code chunk below, do the following:

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

```
fish.ward <- hclust(fish.db, method = "ward.D2") #now visualize phylogenetically
par(mar = c(1, 5, 2, 2) + 0.1) #set up display settings
plot(fish.ward, main = "Doubs River Fish: Ward's Clustering", ylab = "Squared Bray-Curtis Distance") #plo
```



## Doubs River Fish: Ward's Clustering



*#I like that this tree is not ultrametric*

**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:** The plot of sites in x-y space indicated that there were spatial clusters in the sampling of sites: Downstream; Midstream; and Upstream. There were a few sites that fell between the Upstream and the oxbow portion of Midstream. On the Ward's Clustering cluster analysis, there are 3 large closely clustered clades, and 1 smaller clade. I could hypothesize that these clades correspond to geographic locations along the stream from upstream to downstream. That is, fish communities tend to more closely resemble other fish communities which are physically closely together in terms of how far the stream flows between the communities' locations.

This hypothesis could be (in)validated by plotting the site identifier numbers on the x-y space map.

If the site numbers correspond 1 through 30 to sites most Upstream through most Downstream (and based on section 5A, I think this is true), then my hypothesis is somewhat supported: The cluster on the left is mostly Upstream sties; there is a small Midstream/Downstream clade of 3; on the right is a Midstream clade; in the middle-right is a Downstream clade.

The clades are less visible on the heatmap.

## 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.

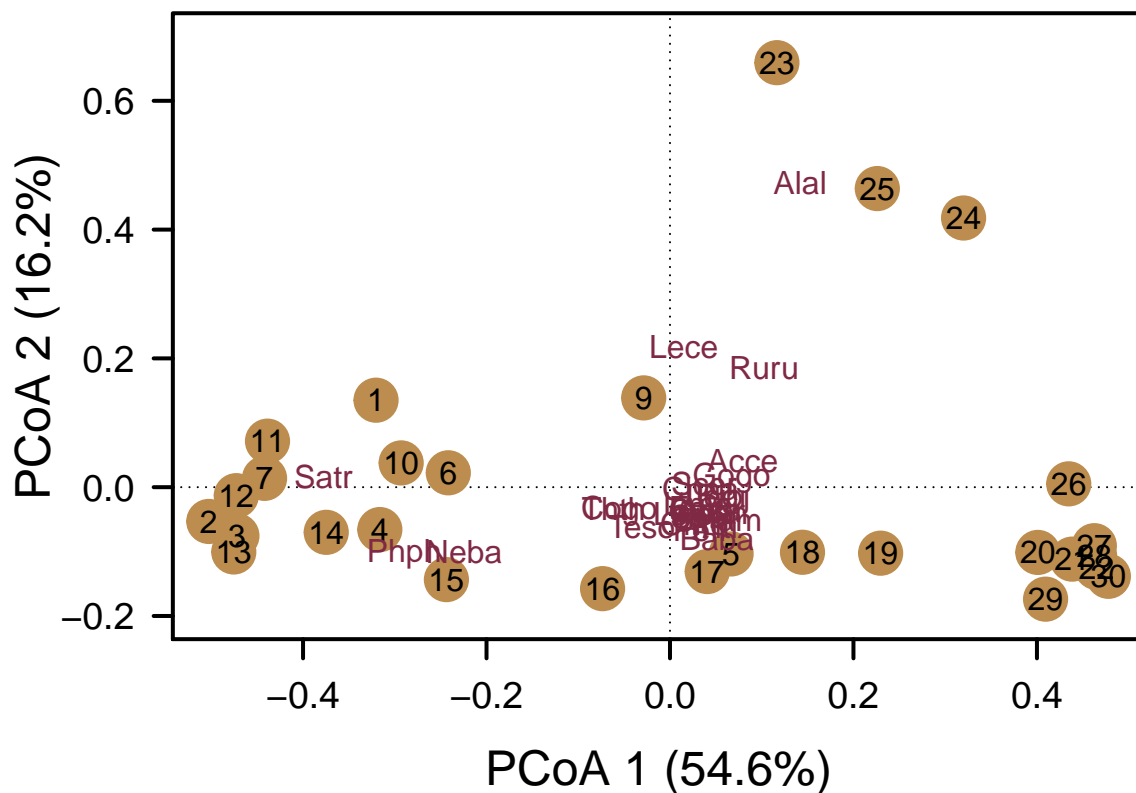
```
fish.pcoa <- cmdscale(fish.db, eig = TRUE, k = 3)#do the PCoA
str(fish.pcoa)

## List of 5
## $ points: num [1:29, 1:3] -0.3204 -0.5027 -0.4721 -0.3163 0.0666 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:29] "1" "2" "3" "4" ...
## .. ..$ : NULL
## $ eig : num [1:29] 3.695 1.098 0.71 0.415 0.305 ...
## $ x : NULL
## $ ac : num 0
## $ GOF : num [1:2] 0.748 0.78

explainvar1 <- round(fish.pcoa$eig[1] / sum(fish.pcoa$eig), 3) * 100#for each of the first three eigenv
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)#total amt explained in these 3 dimensions

par(mar = c(5, 5, 1, 2) + 0.1)#structure the figure output
# Initiate Plot
plot(fish.pcoa$points[,1], fish.pcoa$points[,2], ylim = c(-0.2, 0.7), xlab = paste("PCoA 1 (", explain
# 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 = "#bc8d4f", col = "#bc8d4f")
text(fish.pcoa$points[,1], fish.pcoa$points[,2],
labels = row.names(fish.pcoa$points))
#Now IDfy the most influential spp
fishREL <- fish
for(i in 1:nrow(fish)){
  fishREL[i, ] = fish[i, ] / sum(fish[i, ])
}

fish.pcoa <- add.spec.scores(fish.pcoa,fishREL,method = "pcoa.scores")
text(fish.pcoa$cproj[,1], fish.pcoa$cproj[,2],labels = row.names(fish.pcoa$cproj), col = "#802944")
```



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 # user defined cutoff
imp.spp <- spe.corr[abs(spe.corr[, 1]) >= corrcut | abs(spe.corr[, 2]) >= corrcut, ] #now get those spp
print(imp.spp) #and print the latter list
```

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

```
#do a permut'n test for species abundance across axes
fit <- envfit(fish.pcoa, fishREL, perm = 999)
```

**Question 7:** Address the following questions about the ordination results of the Doubs data set:

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

### **Answer 7a:**

I think the clustering is based on the community composition of the 30 sites, and therefore the question confuses me somewhat.

The sites cluster similarly in the PCoA as they did in the Ward's Clustering.

One could posit a similar hypothesis that the clustering is related to geographic upstream/downstream distance. If the site numbers correspond 1 through 30 to sites most Upstream through most Downstream (and based on section 5A, I think this is true), this hypothesis would be somewhat supported.

### **Answer 7b:**

Perhaps it would make sense to focus on the most-influential species. The correlation coefficient test indicated that Phph, Neba, Rham, Legi, Cyca, Abbr, Acce, Blbj, Alal, and Anan spp exhibited a correlation coefficient  $> 0.7$  along at least PCoA dimension. Whether or not these species are particularly sensitive to changes in water quality, I think that even if these spp. are equally sensitive to water quality as other fish spp., if there is a response by the stream organisms, it would be more easily detectable in these species.

## **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)) #site 8 doesnt exist here
adonis(fish ~ quality, method = "bray", permutations = 999) #now look for differences in fish community

##
## 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

#p <= 0.001 < 0.05
indval <- multipatt(fish, cluster = quality, func = "IndVal.g", control = how(nperm=999))
summary(indval) #are there spp which are especially representative of a particular level of the quality

##
```

```

## 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    0.032 *
##
## Group HQ+MQ #sps. 2
##      stat p.value
## Satr 0.860    0.006 **
## Phph 0.859    0.013 *
##
## 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    0.001 ***
## Spbi 0.866    0.002 **
## Cyca 0.866    0.001 ***
## Acce 0.866    0.003 **
## Lele 0.863    0.004 **
## Titi 0.853    0.006 **
## Chto 0.829    0.003 **
## Rham 0.829    0.005 **
## Anan 0.829    0.004 **
## Eslu 0.827    0.021 *
## Pefl 0.806    0.019 *
## Blbj 0.791    0.006 **
## Scer 0.766    0.007 **
## Abbr 0.750    0.010 **
## Icme 0.661    0.020 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#All spp overlap with medium quality

#or calculate affinity of particular species for particular quality factor levels
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  0.003 **
##
## Group LQ #sps. 2
##      stat p.value
## Alal 0.693  0.001 ***
## Ruru 0.473  0.033 *
##
## Group MQ #sps. 4
##      stat p.value
## Anan 0.571  0.010 **
## Spbi 0.557  0.013 *
## Chto 0.542  0.017 *
## Icme 0.475  0.041 *
##
## Group LQ+MQ #sps. 9
##      stat p.value
## Legi 0.658  0.002 **
## Baba 0.645  0.003 **
## Rham 0.600  0.005 **
## Acce 0.594  0.002 **
## Cyca 0.586  0.008 **
## Chna 0.571  0.002 **
## Blbj 0.571  0.008 **
## Gogo 0.523  0.022 *
## Abbr 0.499  0.026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

*#all spp have significant affinities for one, or a combination of two, factor levels.*

**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:** perMANOVA indicates that spp. are not distributed among sites of varying quality under the distribution predicted by a null model.

IndVal indicates that there are not particular species which are very indicative of low or high quality waters. 23 species were significant indicators, but all of those indications included medium quality.

phi coefficient analysis indicates that 1/3 of the species have an affinity for one particular level of water quality, and 2/3 of the species have affinity for  $\leq 2$  levels of water quality.

## 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.

```
#make matrices
fish.dist <- vegdist(doubs$fish[-8, ], method = "bray")
env.dist <- vegdist(scale(doubs$env[-8,]),method = "euclid")

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%
## 0.100 0.139 0.169 0.192
## 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:**

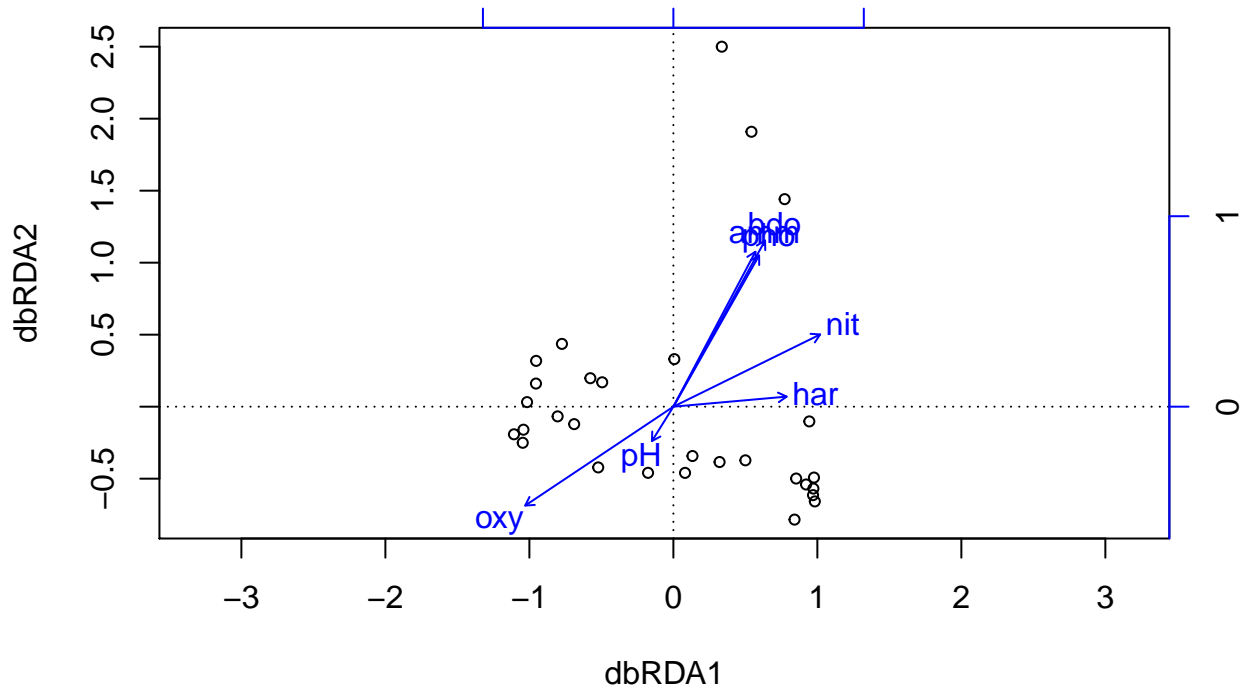
Fish diversity is correlated with environmental variables ( $r = 0.604$ ). This supports the results of the categorical analyses above, which in turn supports an hypothesis in which some species do have particular affinities for differing levels of water quality.

### 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.

```
#make envtl matrix
env.chem <- as.matrix(doubs$env[-8 , 5:11])
#dbRDA, redundancy analysis
doubs.dbrda <- dbrda(fish.db ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda)
```



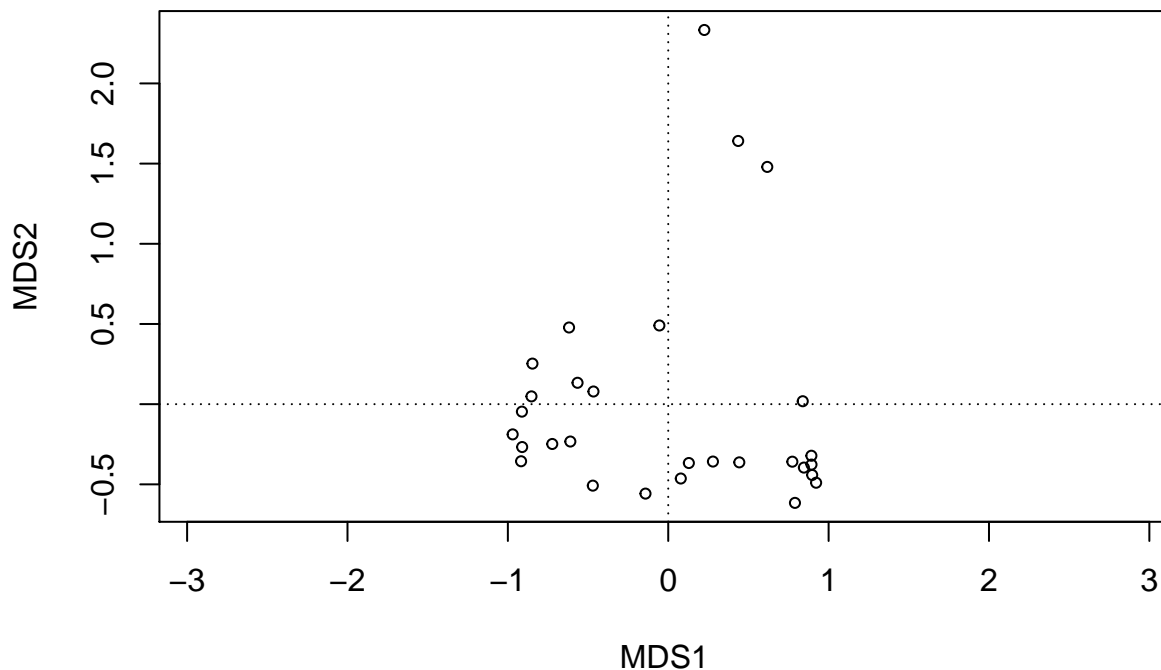
```
psych::corr.test(env.chem)
```

```
## Call:psych::corr.test(x = env.chem)
## Correlation matrix
##      pH    har    pho    nit    amm    oxy    bdo
## pH   1.00  0.08 -0.08 -0.04 -0.12  0.19 -0.16
## har  0.08  1.00  0.37  0.53  0.30 -0.37  0.34
## pho -0.08  0.37  1.00  0.80  0.97 -0.76  0.91
## nit -0.04  0.53  0.80  1.00  0.80 -0.69  0.68
## amm -0.12  0.30  0.97  0.80  1.00 -0.75  0.90
## oxy  0.19 -0.37 -0.76 -0.69 -0.75  1.00 -0.84
## bdo -0.16  0.34  0.91  0.68  0.90 -0.84  1.00
## Sample Size
## [1] 29
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##      pH    har    pho    nit    amm    oxy    bdo
## pH   0.00  1.00  1.00  1.00  1.00  1.00  1.00
## har  0.66  0.00  0.46  0.03  0.83  0.46  0.59
## pho  0.68  0.05  0.00  0.00  0.00  0.00  0.00
## nit  0.83  0.00  0.00  0.00  0.00  0.00  0.00
```



```
## amm 0.53 0.12 0.00 0.00 0.00 0.00 0.00
## oxy 0.32 0.05 0.00 0.00 0.00 0.00 0.00
## bdo 0.40 0.07 0.00 0.00 0.00 0.00 0.00
##
## To see confidence intervals of the correlations, print with the short=FALSE option
#nit, amm, oxy, bdo are all significantly correlated with one another.
#let's be more parsimonious and use fewer variables---avoid model overfitting
# First, we will model only the intercept
doubts.dbrda.mod0 <- dbrda(fish.db ~ 1, as.data.frame(env.chem))

ordiplot(doubts.dbrda.mod0)
```



```
#the full model:
doubts.dbrda.mod1 <- dbrda(fish.db ~ ., as.data.frame(env.chem))
# Now step through all combinations of explanatory variables
doubts.dbrda <- ordiR2step(doubts.dbrda.mod0, doubts.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
```

*#Then lowest AIC model is returned; add trace argument bc I want to see more models, but I'm not sure i*  
doubts.dbrda <- `ordiR2step`(doubts.dbrda.mod0, doubts.dbrda.mod1, `perm.max` = 200, `trace` = TRUE)

```
## 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
```

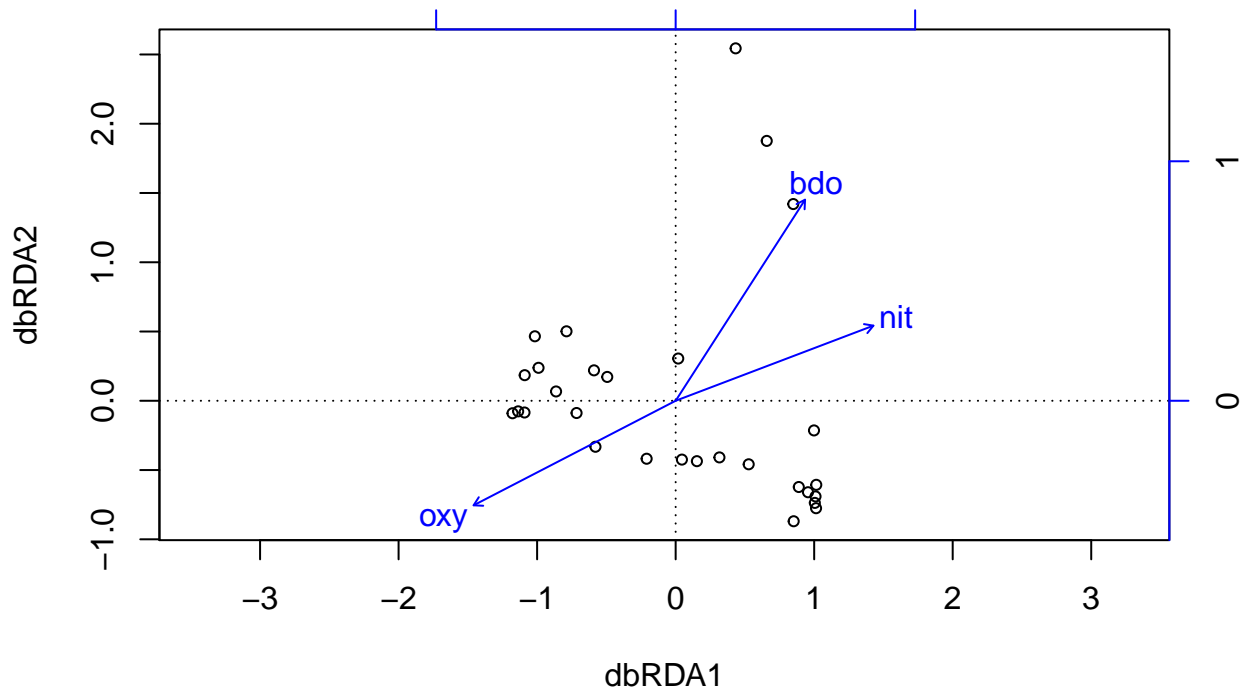
```

## + 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
# examine model that was selected
doubts.dbrda$call

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

```
#P <= 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
# Calc explained variation
dbrda.explainvar1 <- round(doubs.dbrda$CCA$eig[1] /sum(c(doubs.dbrda$CCA$eig, doubs.dbrda$CA$eig)), 3)
dbrda.explainvar2 <- round(doubs.dbrda$CCA$eig[2] /sum(c(doubs.dbrda$CCA$eig, doubs.dbrda$CA$eig)), 3)
print(dbrda.explainvar1+dbrda.explainvar2)

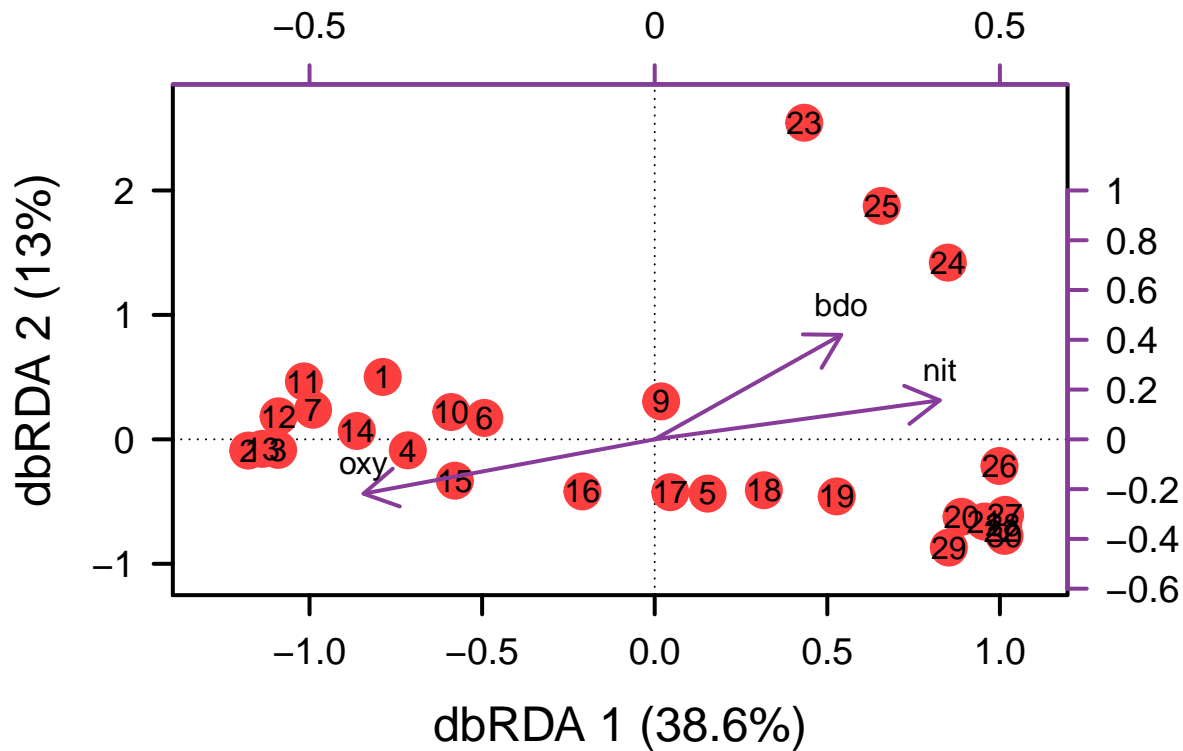
## dbRDA1
## 51.6

#first two axes explain 51.6% of variance

# 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 = ""))
# 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)
# points & labels
points(scores(doubs.dbrda, display = "wa"), pch = 19, cex = 2.5, bg = "gray", col = "#fc3e3f")
text(scores(doubs.dbrda, display = "wa"), labels = row.names(scores(doubs.dbrda, display = "wa")))
# envtl 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 = "#883c9a")
text(vectors[,1], vectors[, 2], pos = 3, labels = row.names(vectors))
axis(side = 3, lwd.ticks=2, cex.axis=1.2, las = 1, col = "#883c9a", lwd = 2.2, at = pretty(range(vectors[,1])))
axis(side = 4, lwd.ticks=2, cex.axis=1.2, las = 1, col = "#883c9a", lwd = 2.2, at = pretty(range(vectors[,2])))

```



#okay, it worked...

**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 constrained ordination was able to explain 51.6% of the variance using just the nit, oxy, and bdo variables. All these variables were highly significant ( $P < 0.005$ ) in the lowest AIC constrained ordination model.

### 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.

```
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

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

#spatial model:
rs <- rowSums(fish)/sum(fish)
doubts.pcnmw <- pcnm(dist(doubts$xy[-8,]), w = rs, dist.ret = T)
doubts.pcnmw$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

#pare down the model
doubts.space <- as.data.frame(scores(doubts.pcnmw))
doubts.pcnm.mod0 <- dbrda(fish.db ~ 1, doubts.space)
doubts.pcnm.mod1 <- dbrda(fish.db ~ ., doubts.space)
step.pcnm <- ordiR2step(doubts.pcnm.mod0, doubts.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 0.006 **
## ---
## 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 0.008 **
## ---
## 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 0.02 *

```

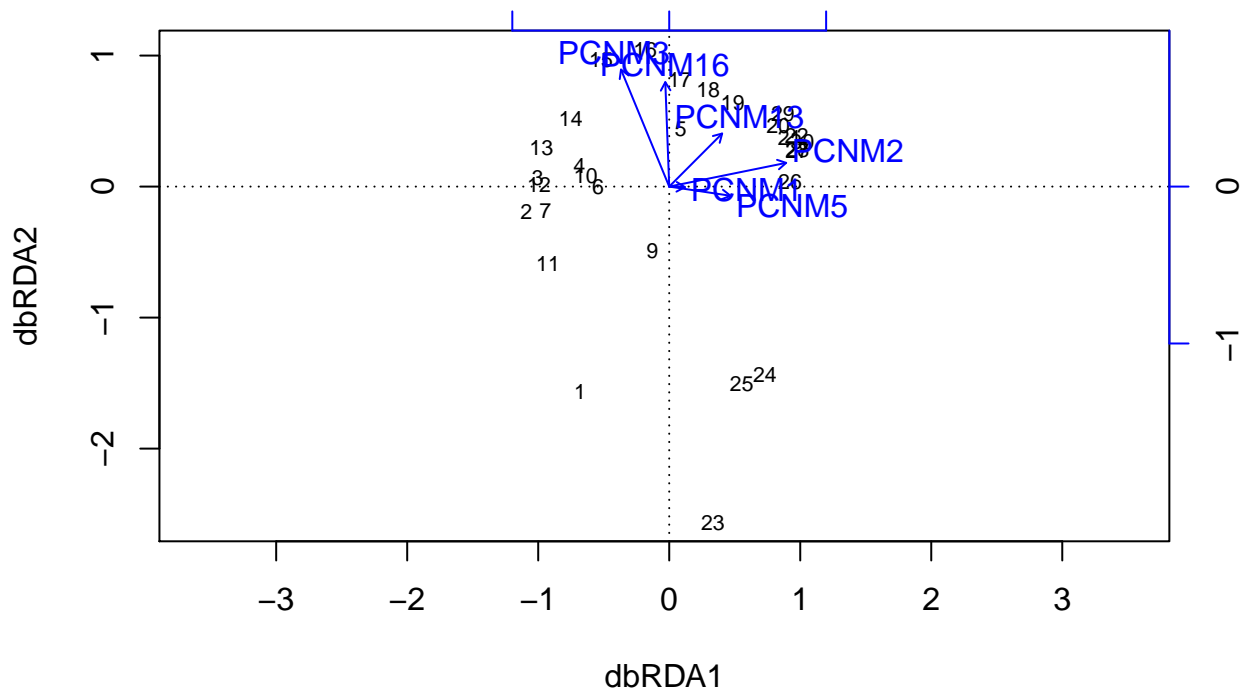
```

## ---
## 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 0.052 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#biplot
plot(step.pcnm)#it's convoluted
```



```
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 0.006 **
## + PCNM1    0.47215 1 41.411 4.0570 0.008 **
## + PCNM13   0.52124 1 39.346 3.4612 0.020 *
## + PCNM16   0.57680 1 36.480 4.0192 0.012 *
## <All variables> 0.62601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#spatial model with only the PCNMs indicated in the lowest AIC model
```

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

```
#constrained ordinations
```

```
doubts.total.env <- dbrda(fish.db ~ env.mod)
```

```

doubts.total.space <- dbrda(fish.db ~ space.mod)
#partial constrained ordination stt one varb is controlled for
doubts.env.cond.space <- dbrda(fish.db ~ env.mod + Condition(space.mod))
doubts.space.cond.env <- dbrda(fish.db ~ space.mod + Condition(env.mod))
#simulations for p-value testing
permutest(doubts.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(doubts.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(doubts.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(doubts.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

```

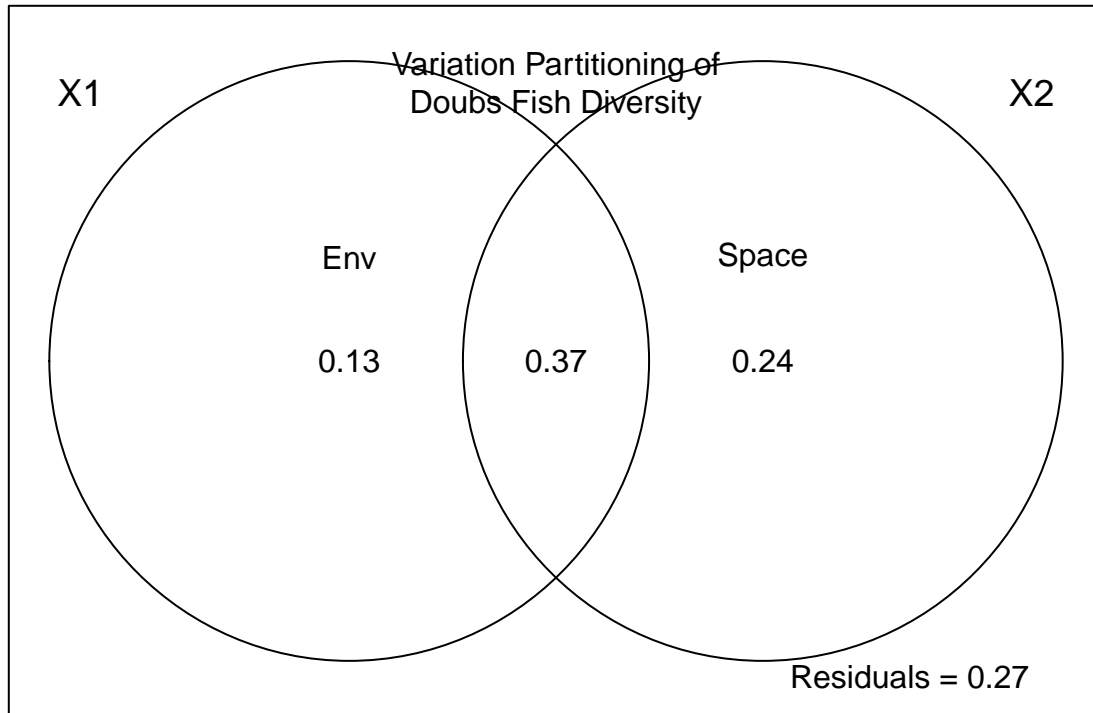
```

#all are significant, P <= 0.001
#visualize proportions of variance explained by each type of variables
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:** The majority (73-74%) of the variance can be explained by environment and/or space. The largest portion (37%) is explained by spatially structured environment factors. Space alone explained more variation (24%) than environment alone did (13%).

This indicates that the environmental variables are correlated with geographic location. When this is taken into account, the model can make better predictions about community compositions.

## 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  $\beta$ -diversity and identify the appropriate tools you would use to test them.

```
plant<-read.csv("C:\\Users\\rmoge\\GitHub\\QB2017_DivPro\\Data\\HF_plants.csv")
#print(dim(plant))
#print(names(plant))

#install.packages("dplyr")
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
plants.sortedT<-arrange(plant, treatment)
plants.sortedY<-arrange(plants.sortedT,year)
plants.sortedP<-arrange(plants.sortedY,plot)
my2009plants.sortedY<-filter(plants.sortedY, year==2009)
dim(my2009plants.sortedY)

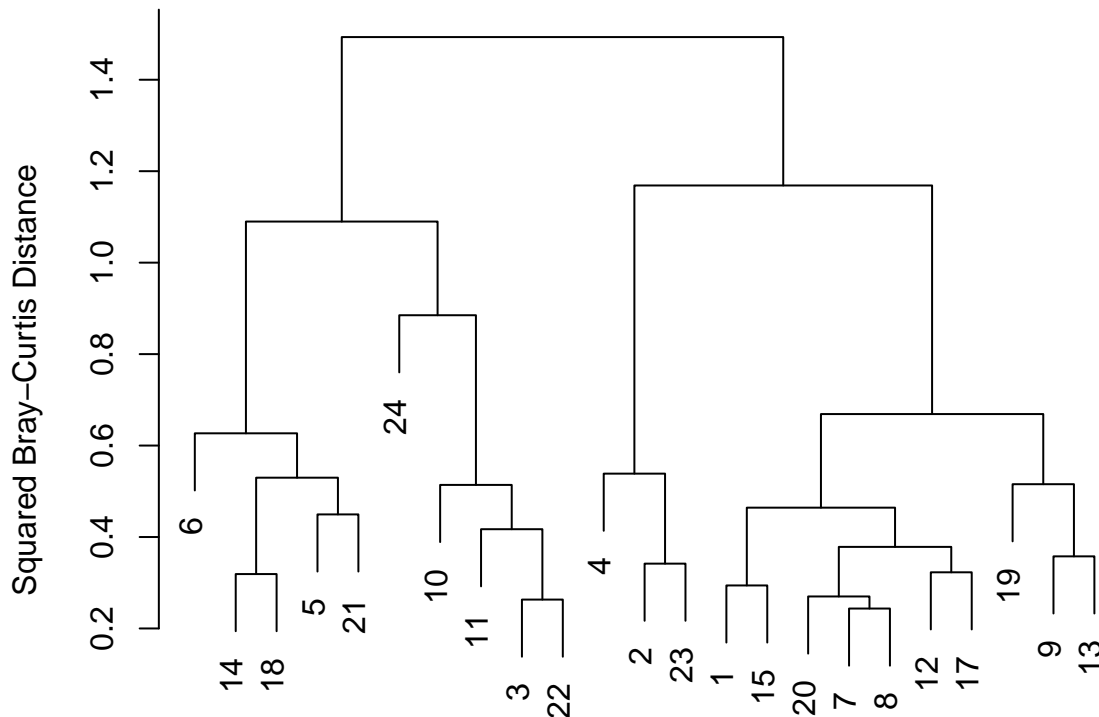
## [1] 24 43
sbs2009plants.sortedYt<-my2009plants.sortedY[4:43]
sbs2009plants.sortedYtt<-sbs2009plants.sortedYt[-16,]

soilenv<-read.csv("C:\\Users\\rmoge\\GitHub\\QB2017_DivPro\\Data\\HF_soilresp.csv")
my2009soilenv<-filter(soilenv,year==2009,day==14,month==1)
temporary_soil<-my2009soilenv[-9]
sbs2009soilenv<-as.matrix(temporary_soil[8:11])

# Calculate Bray-Curtis
#plant.db <- vegdist(sbs2009plants.sortedYt, method = "bray")
#plant.db
plant.db <- vegdist(as.matrix(sbs2009plants.sortedYtt), method = "bray", upper = TRUE, diag = TRUE)
plant.ward <- hclust(plant.db, method = "ward.D2")#now visualize phylogenetically
par(mar = c(1, 5, 2, 2) + 0.1)#set up display settings
plot(plant.ward, main = "HF sites 2009: Ward's Clustering",ylab = "Squared Bray-Curtis Distance")#plot
```



## HF sites 2009: Ward's Clustering



*#I will try to make sense of the clustering based on known treatments that were applied.  
 #Unfortunately 19;9;13 are all different treatments.  
 #4;2;23 are none controls, but each is a different experimental treatment.  
 #6;5;14;18;21 run the gamut of treatments.  
 #10;11;3;22 are all non-controls, but 24 is a control plot  
 #12;17 are not the same treatment.  
 #20;7;8 are not the same treatment.  
 #1;15 are not the same treatment.  
 #I am not able to detect any categorical patterns in the clustering.*

```
#Next do PCoA
plant.pcoa <- cmdscale(plant.db, eig = TRUE, k = 3)#do the PCoA
str(plant.pcoa)
```

```
## List of 5
## $ points: num [1:23, 1:3] -0.00742 -0.43847 0.28123 -0.36578 0.24496 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:23] "1" "2" "3" "4" ...
## .. ..$ : NULL
## $ eig : num [1:23] 1.356 0.818 0.647 0.526 0.333 ...
## $ x : NULL
## $ ac : num 0
## $ GOF : num [1:2] 0.57 0.598
```

```
explainvar1 <- round(plant.pcoa$eig[1] / sum(plant.pcoa$eig), 3) * 100#for each of the first three eigenvalues
explainvar2 <- round(plant.pcoa$eig[2] / sum(plant.pcoa$eig), 3) * 100
```

```

explainvar3 <- round(plant.pcoa$eig[3] / sum(plant.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3) #total amt explained in these 3 dimensions

par(mar = c(5, 5, 1, 2) + 0.1) #structure the figure output
# Initiate Plot
plot(plant.pcoa$points[,1], plant.pcoa$points[,2], ylim = c(-0.2, 0.7), xlab = paste("PCoA 1 (", explainvar1, "%)", sep=""),
# 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.pcoa$points[,1], plant.pcoa$points[,2],
pch = 19, cex = 3, bg = "#fc066e", col = "#fc066e")
text(plant.pcoa$points[,1], plant.pcoa$points[,2],
labels = row.names(plant.pcoa$points))

#Now IDfy the most influential spp
plantREL <- sbs2009plants.sortedYtt
for(i in 1:nrow(sbs2009plants.sortedYtt)){
  plantREL[i, ] = sbs2009plants.sortedYtt[i, ] / sum(sbs2009plants.sortedYtt[i, ])
}

plant.pcoa <- add.spec.scores(plant.pcoa, plantREL, method = "pcoa.scores")

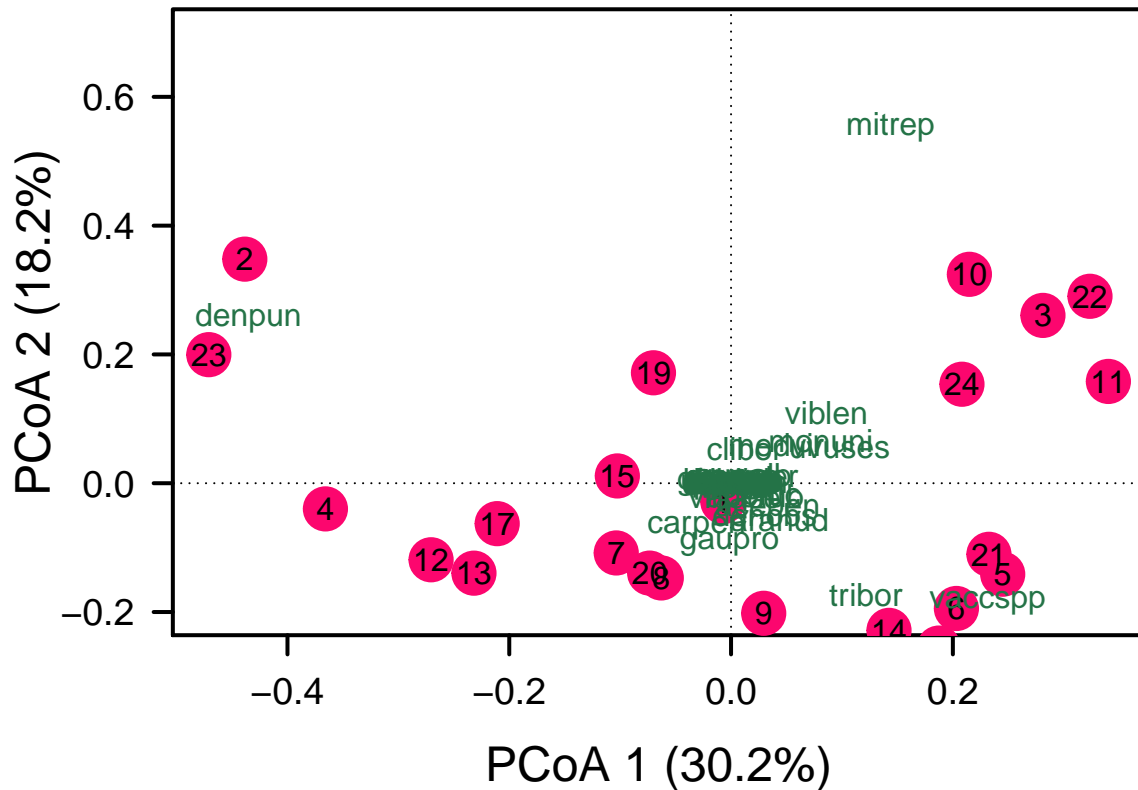
```

[illegible]

```
## Warning in cor(comm[, i], ordiscores[, j]): the standard deviation is zero
```

```
## Warning in cor(comm[, i], ordiscores[, j]): the standard deviation is zero
```

```
text(plant.pcoa$proj[,1], plant.pcoa$proj[,2], labels = row.names(plant.pcoa$proj), col = "#247347")
```



```
#i used exactly the same commands as for ordination for fish communities. However, this time I sometime
#24 errors indicating 'stdev is zero' and the site identification numbers are not plotted on the ordina
#I don't know what is going wrong.
```

```
#When the figure does plot correctly, I see the same patterns as in the "phylogenetic clustering": the
#in the ordination are not coherent with respect to the climate treatment that was applied to the plots
```

```
#
```

```
#Tribor; vacc; denpun; viblen species appear to be the most influential spp. I will examine these taxa
```

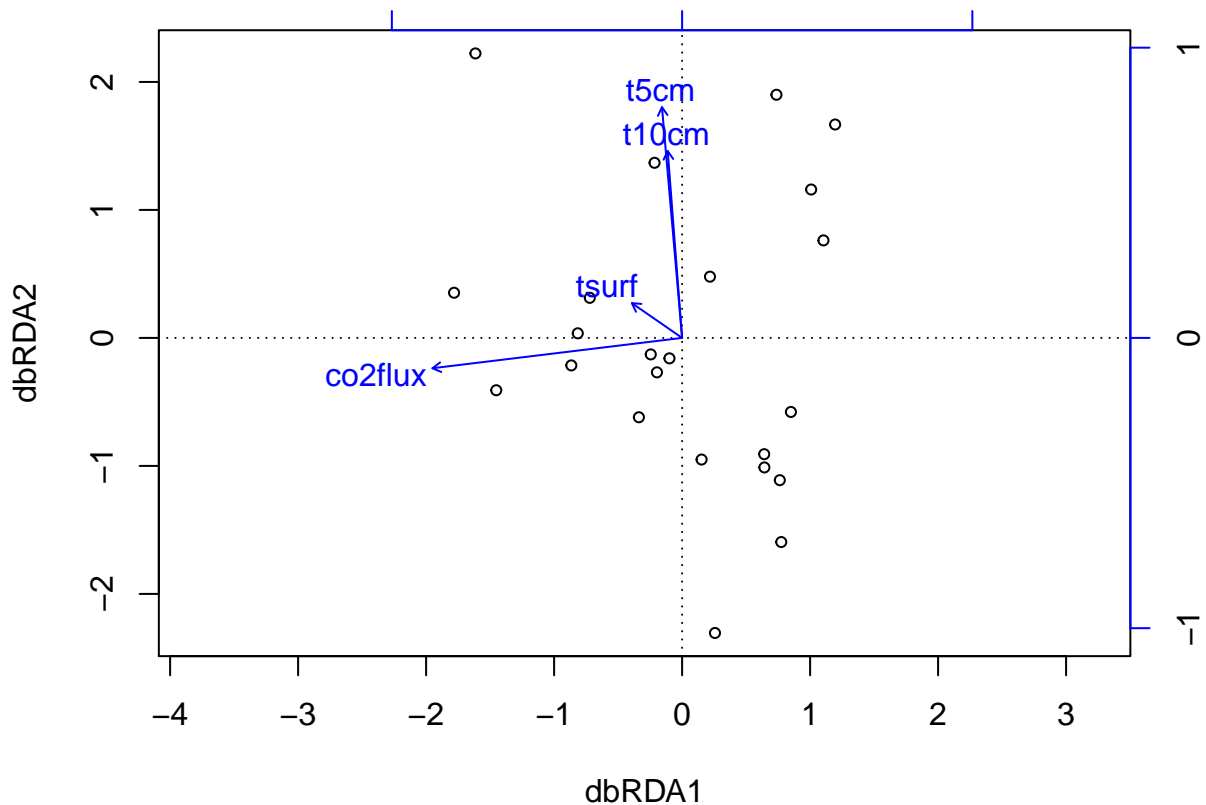
```
#
```

```
#Constrained ordination to examine soil abiotic properties as drivers of tree species diversity
#That is, these are soil environmental variables which were measured observationally, unlike the climat
#modification treatments I have referred to heretofore.
```

```
#dbRDA, redundancy analysis
```

```
HF.dbrda <- dbrda(plant.db ~ ., as.data.frame(sbs2009soilenv))
```

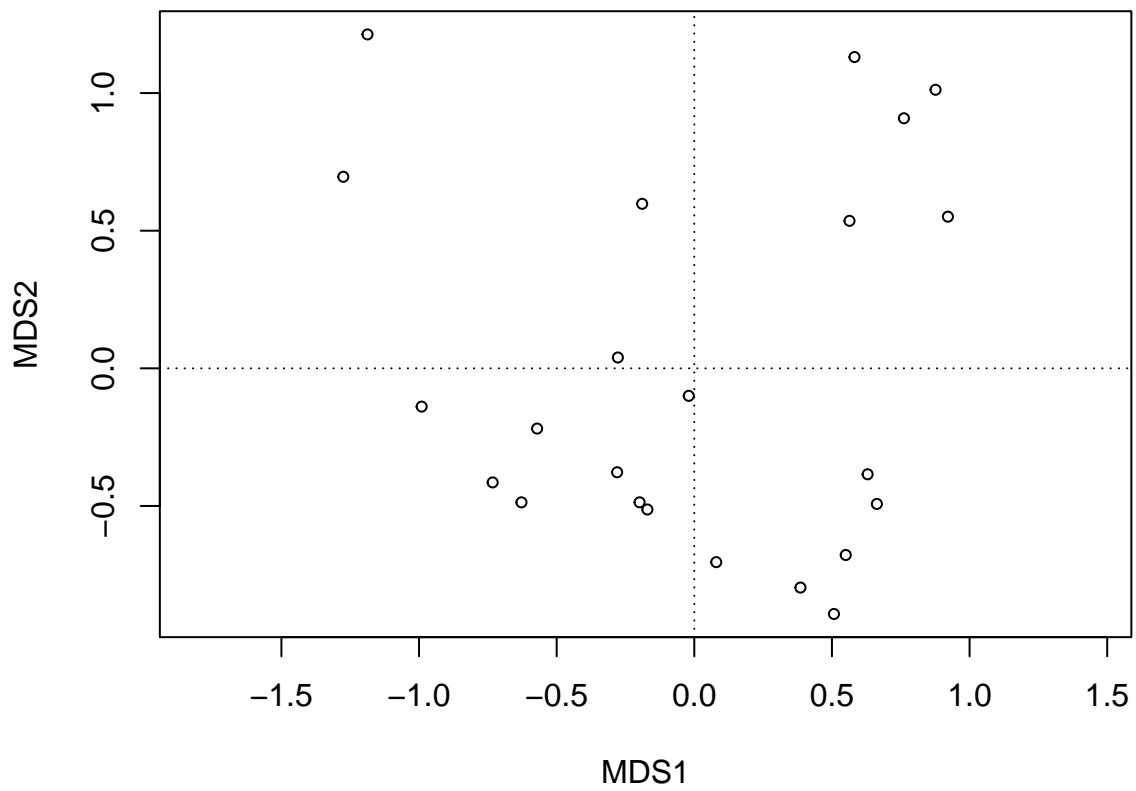
```
ordiplot(HF.dbrda)
```



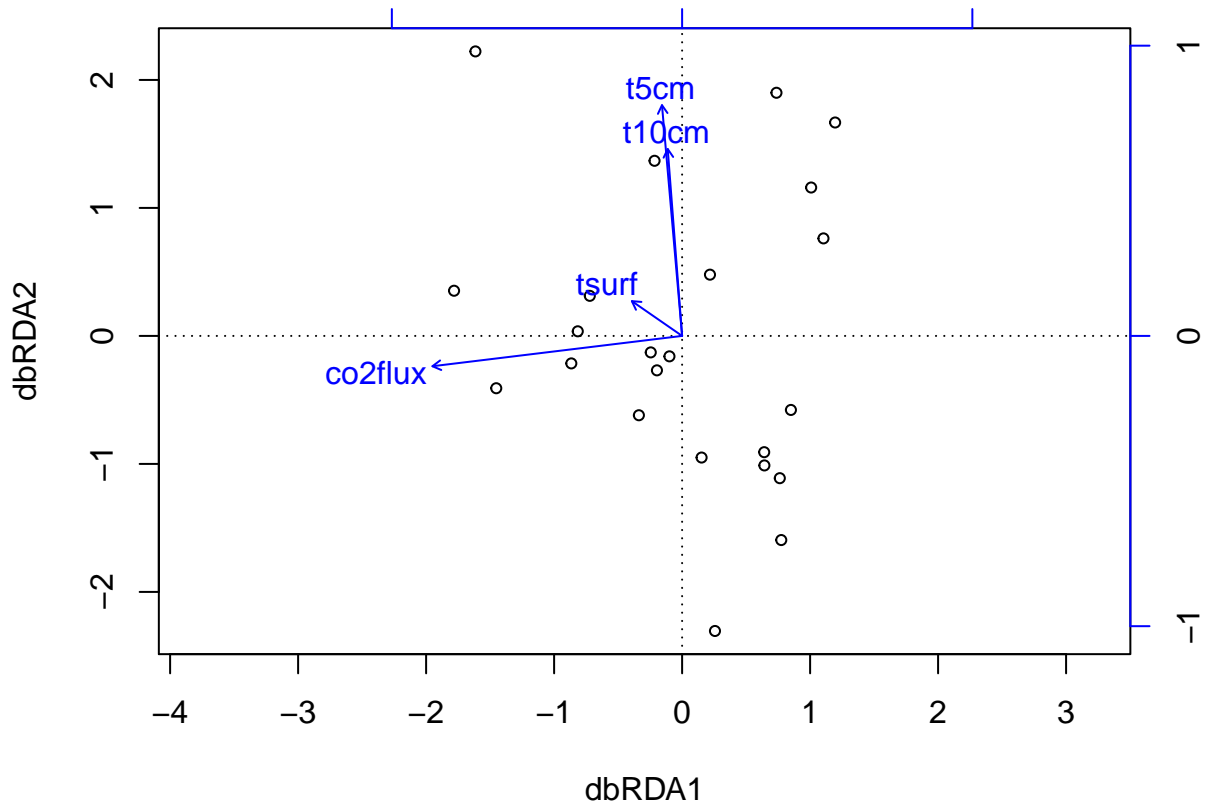
```
psych::corr.test(sbs2009soilenv)
```

```
## Call:psych::corr.test(x = sbs2009soilenv)
## Correlation matrix
##      co2flux tsurf  t5cm t10cm
## co2flux   1.00 -0.28 -0.14 -0.19
## tsurf     -0.28  1.00  0.02  0.02
## t5cm      -0.14  0.02  1.00  0.97
## t10cm     -0.19  0.02  0.97  1.00
## Sample Size
## [1] 23
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##      co2flux tsurf  t5cm t10cm
## co2flux   0.00  0.94   1    1
## tsurf     0.19  0.00   1    1
## t5cm      0.53  0.91   0    0
## t10cm     0.39  0.93   0    0
##
## To see confidence intervals of the correlations, print with the short=FALSE option
#none of the variables are all significantly correlated with one another.
# First, we will model only the intercept
HF.dbrda.mod0 <- dbrda(plant.db ~ 1, as.data.frame(sbs2009soilenv))

ordiplot(HF.dbrda.mod0)
```



```
#the full model:  
HF.dbrda.mod1 <- dbrda(plant.db ~ ., as.data.frame(sbs2009soilenv))  
ordiplot(HF.dbrda.mod1)
```



```
# Permutation tests to evaluate significance
permutest(HF.dbrda.mod1, permutations = 999)
```

```
##
## Permutation test for dbrda
##
## Permutation: free
## Number of permutations: 999
##
## Call: dbrda(formula = plant.db ~ co2flux + tsurf + t5cm + t10cm,
## data = as.data.frame(sbs2009soilenv))
## Permutation test for all constrained eigenvalues
## Pseudo-F:      1.393738 (with 4, 18 Degrees of Freedom)
## Significance:    0.116
```

```
#P = 0.116
envfit(HF.dbrda.mod1, sbs2009soilenv, perm = 999)
```

```
##
## ***VECTORS
##
##          dbRDA1  dbRDA2    r2 Pr(>r)
## co2flux -0.98639 -0.16440 0.4488 0.003 **
## tsurf   -0.94535  0.32605 0.0204 0.776
## t5cm      0.01372  0.99991 0.2207 0.083 .
## t10cm     0.03095  0.99952 0.1448 0.209
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 999
#only co2 flux is a significant explanatory variable

# Calc explained variation
dbrda.explainvar1 <- round(HF.dbrda.mod1$CCA$eig[1] /sum(c(HF.dbrda.mod1$CCA$eig, HF.dbrda.mod1$CA$eig))
dbrda.explainvar2 <- round(HF.dbrda.mod1$CCA$eig[2] /sum(c(HF.dbrda.mod1$CCA$eig, HF.dbrda.mod1$CA$eig))
print(dbrda.explainvar1+dbrda.explainvar2)

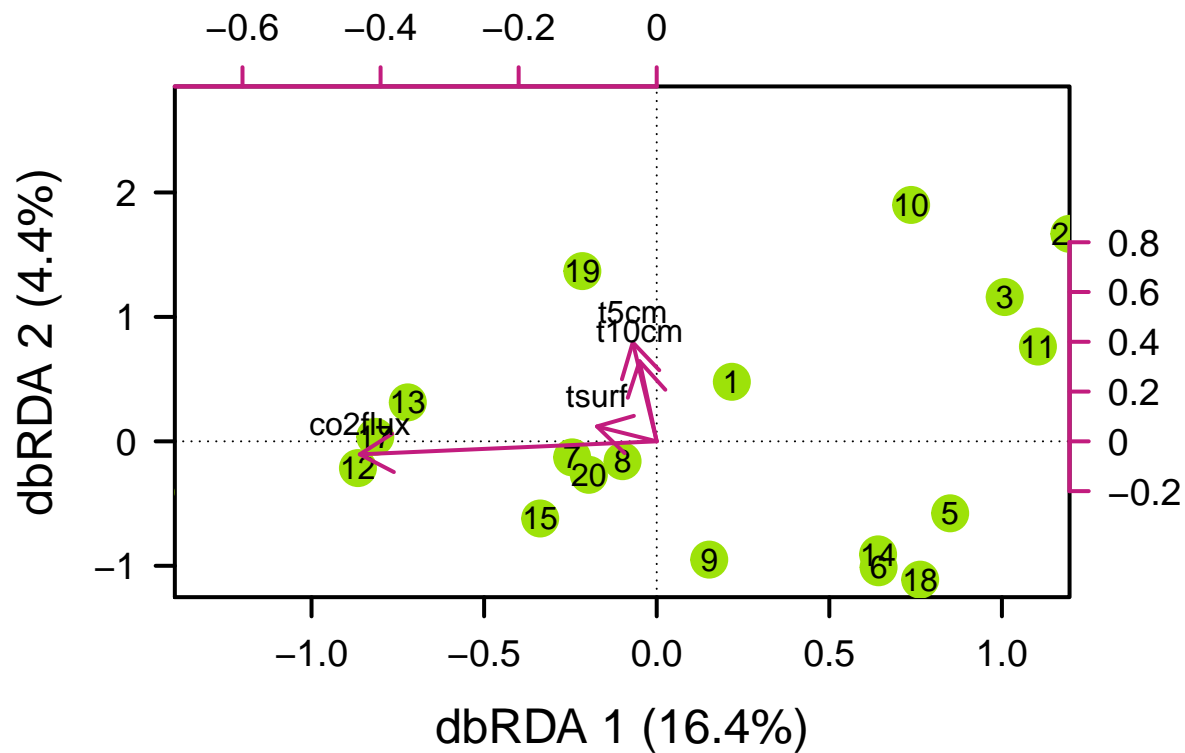
## dbRDA1
##    20.8

#first two axes explain 20.8% of variance

# Define Plot Parameters
par(mar = c(5, 5, 4, 4) + 0.1)
# Initiate Plot
plot(scores(HF.dbrda.mod1, 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 = ""))
# 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)
# points & labels
points(scores(HF.dbrda.mod1, display = "wa"), pch = 19, cex = 2.5, bg = "#f3e634", col = "#9de209")
text(scores(HF.dbrda.mod1, display = "wa"), labels = row.names(scores(HF.dbrda.mod1, display = "wa")))
# envtl Vectors
vectors <- scores(HF.dbrda.mod1, display = "bp")
#row.names(vectors) <- rownames(vectors)

arrows(0, 0, vectors[,1], vectors[, 2],
lwd = 2, lty = 1, length = 0.2, col = "#c21c7e")
text(vectors[,1], vectors[, 2], pos = 3, labels = row.names(vectors))
axis(side = 3, lwd.ticks=2, cex.axis=1.2, las = 1, col = "#c21c7e", lwd = 2.2, at = pretty(range(vectors[,1])))
axis(side = 4, lwd.ticks=2, cex.axis=1.2, las = 1, col = "#c21c7e", lwd = 2.2, at = pretty(range(vectors[,2])))

```



#this plot makes it even more clear that co2 is the best predictor of tree community composition  
 #In the plot it also appears that t5cm and t10cm (temperature at 5cm and 10cm soil depths) are very similar  
 #I tried to match up site identifiers with treatment (control, N addition, warming, or both warming + N)  
 #Perhaps this is not surprising--the permutations test was not significant.