6. Worksheet: Among Site (Beta) Diversity – Part 2

Trang Nguyen; Z620: Quantitative Biodiversity, Indiana University

13 février, 2025

OVERVIEW

In this worksheet, we continue to explore concepts, statistics, and visualizations related to β-diversity. Now that you know how to formally quantify β-diversity, we will learn how to test hypotheses about β-diversity using multivariate statistics.

Directions:

1. In the Markdown version of this document in your cloned repo, change "Student Name" on line 3 (above) with your name.
2. Complete as much of the worksheet as possible during class.
3. Use the handout as a guide; it contains a more complete description of data sets along with examples of proper scripting needed to carry out the exercises.
4. Answer questions in the worksheet. Space for your answers is provided in this document and is indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">". You should notice that the answer is highlighted in green by RStudio (color may vary if you changed the editor theme).
5. Before you leave the classroom today, you should **push** this file to your GitHub repo, at whatever stage you are. This will enable you to pull your work onto your own computer.
6. When you have completed the worksheet, **Knit** the text and code into a single PDF file by pressing the `Knit` button in the RStudio scripting panel. This will save the PDF output in your Posit.cloud workspace: `/cloud/project/QB-2025/Week4-Beta/`
7. After Knitting, please submit the worksheet by making a **push** to your GitHub repo and then create a **pull request** via GitHub. Your pull request should include this file (**6.BetaDiversity_2_Worksheet.Rmd**) with all code blocks filled out and questions answered) and the PDF output of `Knitr` (**6.BetaDiversity_2_Worksheet.pdf**).

The completed exercise is due on **Wednesday, February 12th, 2025 before 12:00 PM (noon)**.

1) R SETUP

Typically, the first thing you will do in either an R script or an RMarkdown file is setup your environment. This includes things such as setting the working directory and loading any packages that you will need.

In the R code chunk below, provide the code to:

1. clear your R environment,
2. print your current working directory,
3. set your working directory to your `Week4-Beta/` folder.

4. load the **vegan** R package (be sure to install if needed).

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

```
## [1] "C:/Users/ttran/OneDrive - Indiana University/SP25 - Quantitative Biodiversity/QB2025_Nguyen/Weel
```

```
setwd(getwd())
```

2) LOADING DATA

Load dataset

In the R code chunk below, load the **doubs** dataset from the **ade4** package

```
# note, pleae do not print the dataset when submitting
library(ade4)
```

```
## Warning: le package 'ade4' a été compilé avec la version R 4.4.2
```

```
data(doubs)
```

3) 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.

```
library(vegan)
```

```
## Warning: le package 'vegan' a été compilé avec la version R 4.4.2
```

```
## Le chargement a nécessité le package : permute
```

```
## Warning: le package 'permute' a été compilé avec la version R 4.4.2
```

```
## Le chargement a nécessité le package : lattice
```

```
## This is vegan 2.6-8
```

```r
library(indicspecies)
```

```
## Warning: le package 'indicspecies' a été compilé avec la version R 4.4.2
```

```r
# Subset the fish data
fish = doubs$fish
fish = fish[-8, ]
# 1. doubs
quality = c(rep("HQ", 13),
    rep("MQ", 5), rep("LQ", 6),
    rep("MQ", 5))

# Run PERMANOVA
adonis2(fish ~ quality, method="bray", permutations=999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = fish ~ quality, permutations = 999, method = "bray")
##          Df SumOfSqs      R2     F Pr(>F)
## Model     2   3.0947 0.45765 10.97  0.001 ***
## Residual 26   3.6674 0.54235
## Total    28   6.7621 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
## IndVal matrix
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   0.029 *
##
```

```
##  Group HQ+MQ  #sps.  2
##       stat p.value
## Satr 0.860   0.004 **
## Phph 0.859   0.016 *
##
##  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   0.003 **
## Acce 0.866   0.002 **
## Lele 0.863   0.006 **
## Titi 0.853   0.004 **
## Chto 0.829   0.002 **
## Rham 0.829   0.003 **
## Anan 0.829   0.003 **
## Eslu 0.827   0.029 *
## Pefl 0.806   0.010 **
## Blbj 0.791   0.003 **
## Scer 0.766   0.007 **
## Abbr 0.750   0.006 **
## Icme 0.661   0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Phi coefficient analysis
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   0.003 **
## Satr 0.650   0.001 ***
##
##  Group LQ  #sps.  2
##       stat p.value
## Alal 0.693   0.001 ***
## Ruru 0.473   0.040 *
##
##  Group MQ  #sps.  4
##       stat p.value
## Anan 0.571   0.010 **
## Spbi 0.557   0.009 **
## Chto 0.542   0.016 *
## Icme 0.475   0.031 *
##
##  Group LQ+MQ  #sps.  9
##       stat p.value
## Legi 0.658   0.002 **
## Baba 0.645   0.001 ***
## Rham 0.600   0.002 **
## Acce 0.594   0.001 ***
## Cyca 0.586   0.006 **
## Chna 0.571   0.014 *
## Blbj 0.571   0.007 **
## Gogo 0.523   0.010 **
## Abbr 0.499   0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 1: Based on the PERMANOVA, IndVal, and phi coefficient analyses, what did you learn about the relationship between habitat quality and the fish species composition? Are the different analyses consistent with one another and do they agree with the visualizations (heat maps, cluster dendograms, ordinations) that you created?

> ***Answer 1***: From the last assignment, we saw that the fish species composition can be segregated into different clusters, where each cluster refers to a type of habitat quality (in this assignment). Here, we used PERMANOVA, IndVal and phi coefficients to see how well the habitat quality influences the fish species compositions. First, let's start with ANOVA. ANOVA is a statistical test that compares the amount of variation between groups to the amount of variation within each group, which allows us to study the effectiveness of segregating the variables into these groups. More precisely, here, we want to see how fish composition within each habitat group vary compared to how fish composition vary in different groups. The R2 of the model shows .45, which implies that the model explains 45% of the data, with a statistical signifiance. This means that habitat quality is a good measure to group the fish species composition. Next, using the IndVal test, we see that among the 27 fish species, there are 23 that are representative of two groups. For group with `MQ` habitat quality, `Teso` is the representative species. While for group `HQ+MQ`, there are 2 representative species, for group `LQ+MQ`, there are 20 representative species. All of them are significant. For phi coefficient test, we didn't see the same representative species as the IndVal test. The only group in phi coefficient test that is close enough to the IndVal test's `HQ+MQ` test is group `HQ` for two species Satr and Phph. I would say the phi coefficient test seems more reliable as it aligns better with the PCOA plot. The PCOA plot showed that the fish species Satr, Phph and Neba are more representative of a group (leftest on the PCOA1 axe),

while the species Alal, Ruru and Lece are more representative of another group (upper part on PCOA2). And these were captured by the phi coefficient test.

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 test the hypothesis that fish assemblages are correlated with stream environmental variables.

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

# Mantel test
mantel(fish.dist, env.dist, method="pearson", permutations = 1000)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = fish.dist, ydis = env.dist, method = "pearson",      permutations = 1000)
##
## Mantel statistic r: 0.604
##       Significance: 0.000999
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.113 0.149 0.175 0.192
## Permutation: free
## Number of permutations: 1000
```

Question 2: 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?

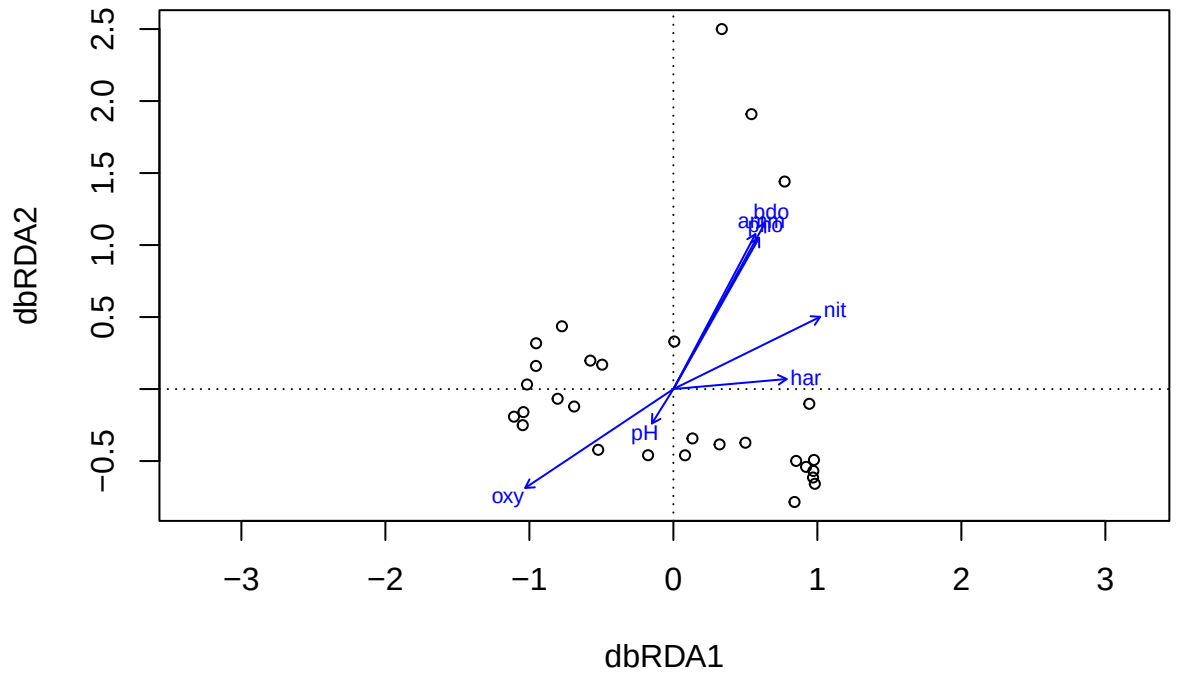
> ***Answer 2***:
> Based on what we learned from class, the Mantel class assessed the correlation between two dissimilarity matrices. Here, our first matrix is the dissimilarity matrix (Bray-Curtis distance) of the fishes composition between sites, and the second one is the dissimilarity matrix of environmental variables between sites. The test showed r=0.604, which means that there is relatively strong positive correlation. We can interpret it that as the environmental differences between sites increase, the dissimilarity in fish communities also tends to increase. Next, the signifiance of the test is 0.0009, meaning the p-val = 0.0001. This p-value indicates that there is only a 0.1% chance of obtaining a correlation of 0.604 or higher if there were actually no relationship between the fish and environmental distances. (Since the Mantel test is one tailed). This once again confirms our hypothesis that the stream quality influences the fish composition.

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.

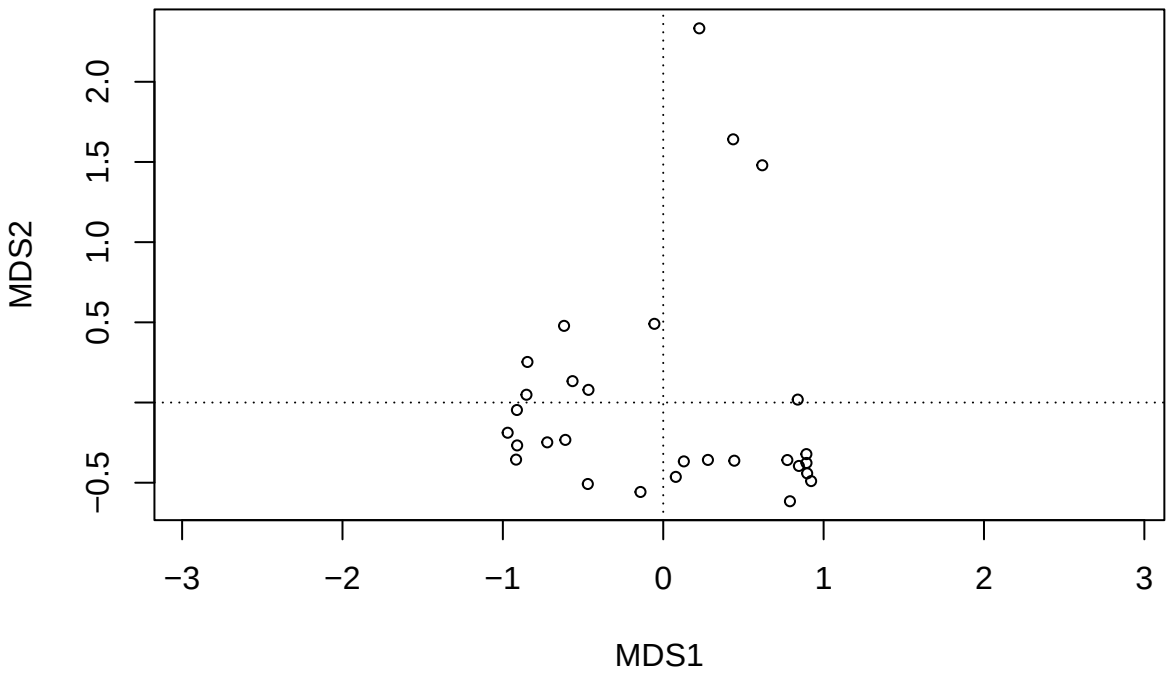
```
# Bray Curtis distance for fish
fish.db = vegdist(fish, method="bray")

# Creat env matrix
env.chem = as.matrix(doubs$env[-8, 5:11])

# dbRDA
# ?dbrda test formula, data using euclidean distance
# Here, we test fish.db data using all variables in env.chem
doubs.dbrda.modfull = dbrda(fish.db ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda.modfull)
```



```r
# Check the pairwise correlation
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
```

```r
# wee see that pho and nit are very correlated 0.8, similar to oxy and pho -.76
# Highly correlated variables => may be overfit


# ================================================
# Try model with constant explained variables
doubs.dbrda.mod0 = dbrda(
    fish.db ~ 1,
    as.data.frame(env.chem)
    )
ordiplot(doubs.dbrda.mod0) # no vector as we don't use any variable in env.chem
```




```
# Next, we will test all combination from 0 explanatory variable to full
# This functions returns the one that has lowest AIC
doubs.dbrda = ordiR2step(doubs.dbrda.mod0,
   doubs.dbrda.modfull, 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
##
##       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
##
##       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
```

```
# Summary of selected model
doubs.dbrda$call # formula = fish.db ~ oxy + bdo + nit
```

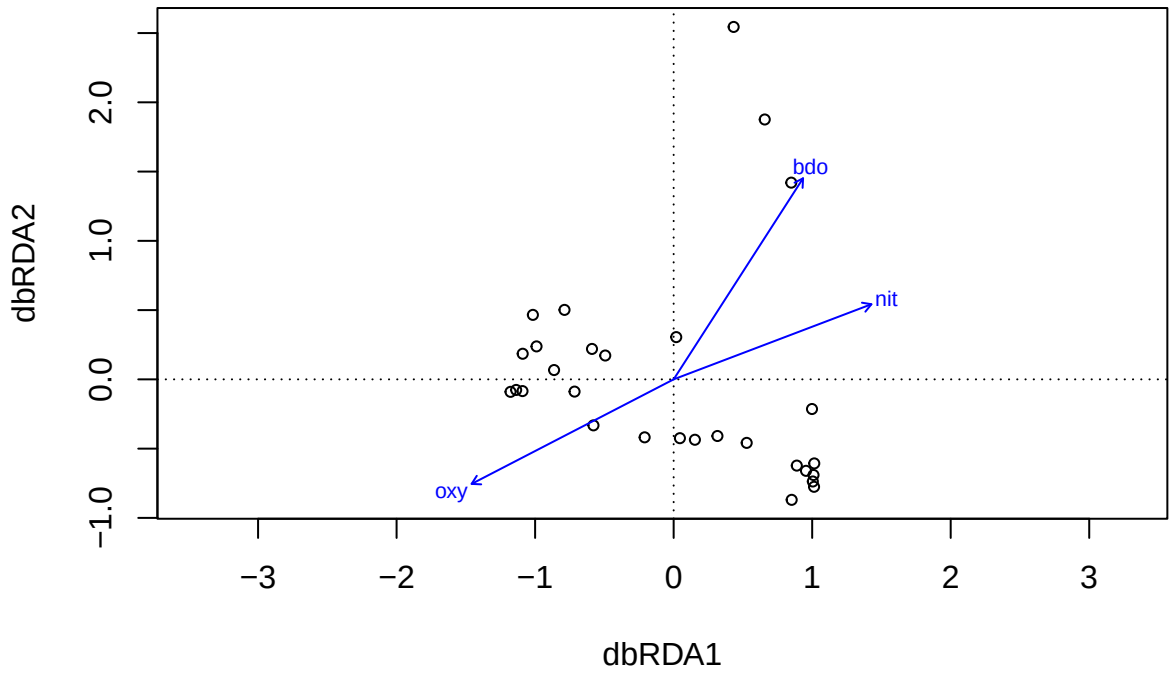
```
## dbrda(formula = fish.db ~ oxy + bdo + nit, data = as.data.frame(env.chem))
```

```
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)
```



Question 3: 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 3***: The variables that contributed to the variation in fish community structure are : `oxy, bdo, nit`. All these variables have a significant p value. The model explains 0.53 of the varaition in the fish community.

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.

```
# Create a matrix model for the envr dataq
env.mod = model.matrix( ~ oxy + bdo + nit, as.data.frame(env.chem))[,-1]
env.mod
```

```
##    oxy bdo nit
## 1  122  27  20
## 2  103  19  20
## 3  105  35  22
## 4  110  13  21
## 5   80  62  52
## 6  102  53  15
## 7  111  22  15
## 9   72  52  82
## 10 100  43  75
## 11 115  27 160
## 12 122  30  50
## 13 124  24  52
## 14 123  38 123
## 15 117  21 100
## 16 103  27 200
## 17 102  46 250
## 18 103  28 220
## 19 106  33 220
## 20 103  28 300
## 21  90  41 220
## 22  91  48 162
## 23  63 164 350
## 24  52 123 250
## 25  41 167 620
## 26  62  89 300
## 27  72  63 300
## 28  81  45 400
## 29  90  42 162
## 30  82  44 160
```

```
#cCreate spatial model
# first, weight each site by its relative abundance
rs = rowSums(fish) / sum(fish)

# Perform PCNM
doubs.pcnmw = pcnm(dist(doubs$xy[-8,]), w=rs, dist.re=T)
# doubs.pcnmw
doubs.pcnmw$values > 0 # Extract only eigenvectors associated with positive eigenvalues
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [13]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE
```

```r
# Perform model selection of spatial data
doubs.space = as.data.frame(scores(doubs.pcnmw))
doubs.pcnm.mod0 = dbrda(fish.db ~ 1,doubs.space) # no var
doubs.pcnm.mod1 = dbrda(fish.db ~ .,doubs.space) # all var
# Stepwise model selection
step.pcnm = ordiR2step(
    doubs.pcnm.mod0,
    doubs.pcnm.mod1,
    perm.max=200)
```

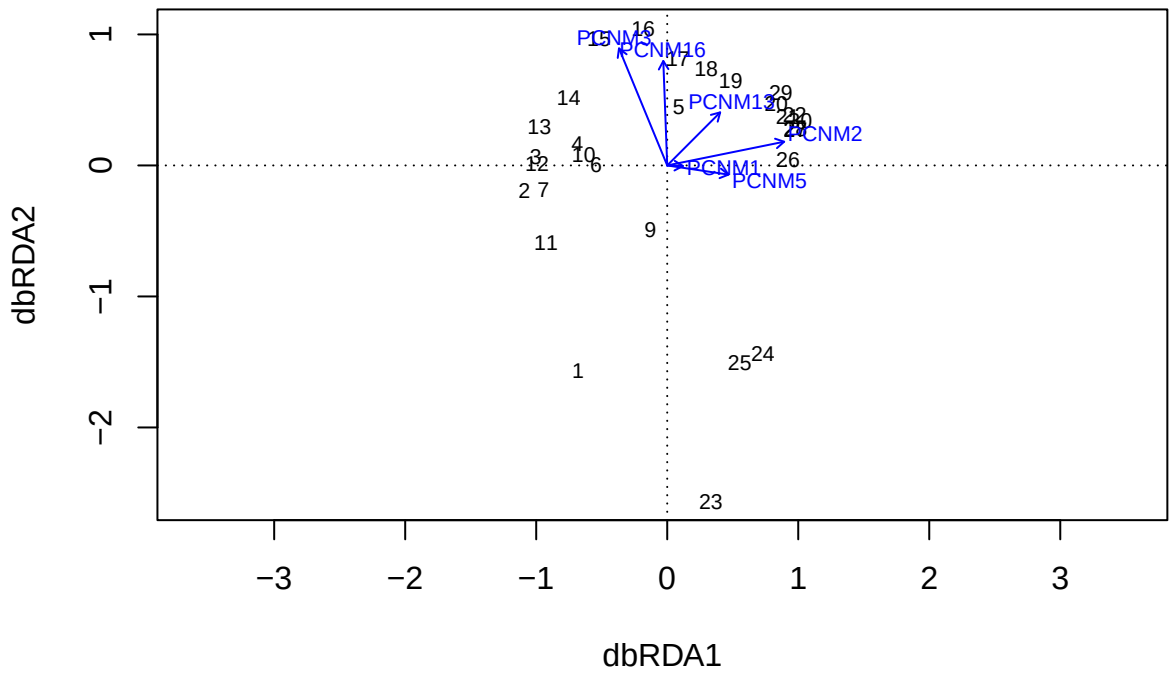
```
## Step: R2.adj= 0
## Call: fish.db ~ 1
##
##                   R2.adjusted
## <All variables>  0.626011301
## + PCNM2          0.235370423
## + PCNM3          0.078394885
## + PCNM13         0.065305668
## + PCNM5          0.046185074
## + PCNM6          0.032809156
## + PCNM16         0.030486700
## + 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
##
##         Df    AIC      F Pr(>F)
## + PCNM3  1 46.083 5.4196  0.004 **
## ---
## 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>   0.6260113
## + PCNM5           0.4076020
## + PCNM1           0.3970300
## + PCNM16          0.3853210
## + PCNM15          0.3828748
## + PCNM14          0.3781827
## + PCNM13          0.3770376
## + PCNM6           0.3595644
## + PCNM8           0.3556885
## + PCNM7           0.3541631
## + PCNM10          0.3526775
## + PCNM17          0.3513683
## + PCNM9           0.3433672
## <none>            0.3429270
## + PCNM11          0.3416399
## + PCNM12          0.3396547
## + PCNM4           0.3311509
##
##         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
##
##         Df    AIC     F Pr(>F)
## + PCNM1  1 41.411 4.057  0.004 **
## ---
## 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
##
##          Df    AIC      F Pr(>F)
## + PCNM13  1 39.346 3.4612  0.024 *
## ---
## 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
##
##          Df   AIC      F Pr(>F)
## + PCNM16  1 36.48 4.0192  0.024 *
## ---
## 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
##
##         Df    AIC      F Pr(>F)
## + PCNM6  1 35.182 2.5296  0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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




```
# Check the portion of explained variation of the fish composition using the spatial 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  0.004 **
## + PCNM5         0.40760  1 43.941 3.8385  0.006 **
## + PCNM1         0.47215  1 41.411 4.0570  0.004 **
## + PCNM13        0.52124  1 39.346 3.4612  0.024 *
## + PCNM16        0.57680  1 36.480 4.0192  0.024 *
## <All variables> 0.62601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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


# ========================================
# Compare env and spatial model
doubs.total.env = dbrda(fish.db ~ env.mod)
doubs.total.space = dbrda(fish.db ~ space.mod)
```

```r
#  PArtial constrained ordination
# ?Condition --> control the second explanatory matrix
doubs.env.cond.space = dbrda(fish.db ~ env.mod + Condition(space.mod))
doubs.space.cond.env = dbrda(fish.db ~ space.mod
+ Condition(env.mod))

# Test for significance of the dbRDA fractions
permutest(doubs.env.cond.space,permutations=999)
```

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.db ~ env.mod + Condition(space.mod))
## Permutation test for all constrained eigenvalues
##          Df Inertia     F Pr(>F)
## Model     3 0.85158 4.423  0.001 ***
## Residual 18 1.15519
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.db ~ space.mod + Condition(env.mod))
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     7  1.8752 4.1741  0.001 ***
## Residual 18  1.1552
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.db ~ env.mod)
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     3  3.7317 10.262  0.001 ***
## Residual 25  3.0304
```

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

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

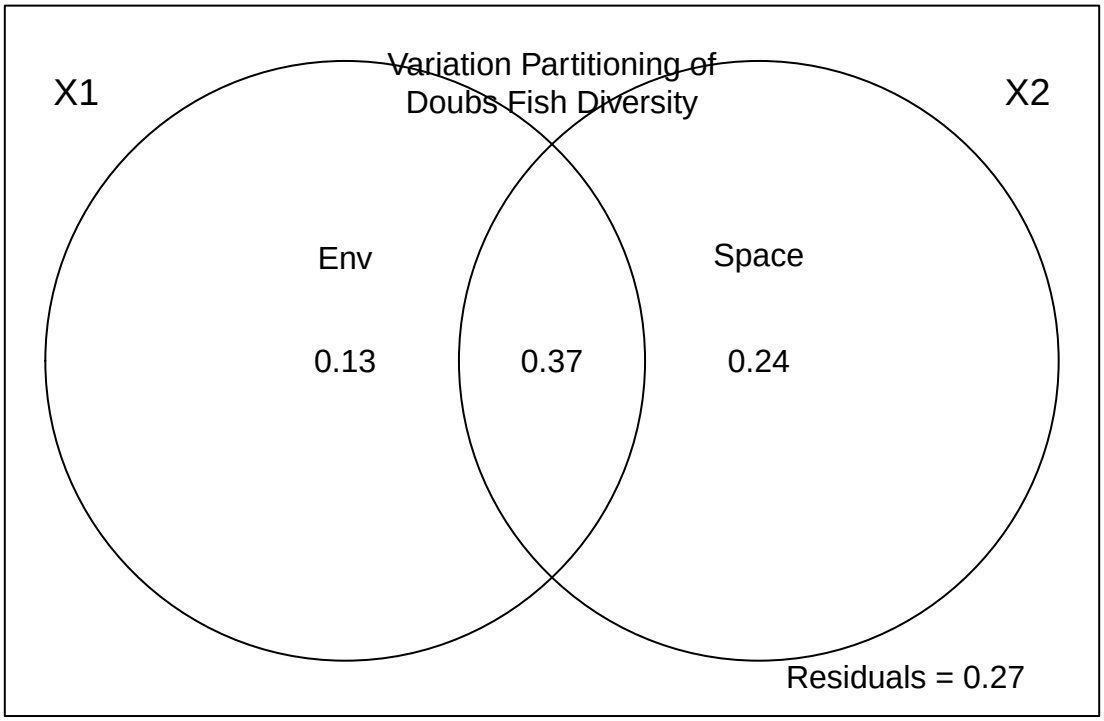
```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.db ~ space.mod)
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     7  4.7553 7.1089  0.001 ***
## Residual 21  2.0068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Calculate the fraction of variation explained by space alone, by env alone and by both, and by neithe
 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+c] = 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] = X2|X1            7                 0.23618     TRUE
## [c]                    0                 0.36813    FALSE
## [d] = Residuals                          0.26574    FALSE
## ---
## Use function 'dbrda' 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 4: Interpret the variation partitioning results.

> ***Answer 4***: For the last partition table, we see that the total sum of squre in the dissimilarity data is 6.67. Two sets of explanatory variables in consideration are: env.mod and space.mod (X2 - the spatial variables derived from PCNM). First, when using the environmental variables (plus any variation they share with spatial factors), the adjusted R^2 is 0.498. This means that 49.8% of the variation in the fish community is explained. Secondly, when using the spatial variables (plus their shared effects with environmental variables), the model explains 60.4% of the variation. When both sets of variables are used together, the model explains about 73% of the total variation in the fish community. If we consider only environmental effect, the model explains only 13% of the variation. If we consider only pure spatial effect, the model explains about 23.6% of the variation. The variation explained by both environment and space is about 36% of the variation. The environmental variables on their own explain only a small portion (13%) of the variation, while the spatial variables independently explain a larger fraction (23.6%). A substantial amount of variation (36%) is explained by the overlap between the two, and together they account for 73% of the total variation in the fish community composition.

SYNTHESIS

Load the dataset from that you and your partner are using for the team project. Use one of the hypothesis-testing tools introduced in the beta diversity module. Interpret the findings of your data with respect to principles of biodiversity.

Answer: After our previous analysis of the zoobenthos data, we discovered that over the past 10 years there were only three distinct sites where we had environmental data over time (at least 2+ years). To work around this, we treated each site's data for each year as a separate sample. When we examined the heatmap of zoobenthos abundance, we saw that the samples clustered into three main groups, which makes sense given that there are three sites. However, we also noticed that sites 0 and 2 tend to be similar to each other, while site 1 maintains a more consistent zoobenthos composition over the years. Furthermore, Ward's Clustering showed quite an interesting pattern: sites 0 and 2 are sometimes more similar to each other than to themselves in certain years (for example, the cluster containing 0_1995 and 0_2002 at the third rightmost cluster in the dendrogram).

Now, we will now use PERMANOVA to determine if the zoobenthos composition differs significantly between sites and across years.

```
library(vegan)
library(indicspecies)


# Load data
# zoobenthos = read.csv(paste0(getwd(),"/Week4-Beta/data/SbS_full.csv"), row.names=1)
zoobenthos = read.csv("data/SbS_full.csv", row.names=1)
# We create a factor vector that categorizes the sites based on the dendogram
groups = c(
    rep("GR2", 1), # 0_1994
    rep("GR3", 1), # 0_1995
    rep("GR2", 5), # 0_1997 - 0_2001
    rep("GR3", 1), # 0_2002
    rep("GR2", 2), # 0_2003 - 0_2004
    rep("GR1", 7), # 1_1994 - 1_2004
    rep("GR3", 9)) # 2_1994 - 2_2004

# Run PERMANOVA
adonis2(zoobenthos ~ groups, method="bray", permutations=999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = zoobenthos ~ groups, permutations = 999, method = "bray")
##          Df SumOfSqs      R2      F Pr(>F)
## Model     2   4.0111 0.42778 8.5972  0.001 ***
## Residual 23   5.3654 0.57222
## Total    25   9.3765 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## IndVal matrix
indval = multipatt(zoobenthos, cluster=groups, func = "IndVal.g", control = how(nperm=100))
summary(indval)
```

```
##
##  Multilevel pattern analysis
```

```
##  ---------------------------
##
##  Association function: IndVal.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 103
##  Selected number of species: 33
##  Number of species associated to 1 group: 31
##  Number of species associated to 2 groups: 2
##
##  List of species associated to each combination:
##
##  Group GR1  #sps.  15
##                         stat p.value
## Pygospio.elegans       1.000  0.0099 **
## Mytilus.edulis         0.998  0.0099 **
## Macoma.balthica        0.990  0.0099 **
## Mya.arenaria           0.986  0.0099 **
## Gammarus.salinus       0.926  0.0099 **
## Gammarus.zaddachi      0.926  0.0099 **
## Jaera.albifrons        0.926  0.0099 **
## Saduria.entomon        0.926  0.0099 **
## Crangon.crangon        0.895  0.0099 **
## Gammarus.oceanicus     0.845  0.0099 **
## Alcyonidium.polyoum    0.796  0.0099 **
## Tubifex.costatus       0.756  0.0198 *
## Bathyporeia.pilosa     0.655  0.0198 *
## Hydrobia.ulvae         0.652  0.0099 **
## Tubificoides.benedeni 0.535  0.0495 *
##
##  Group GR2  #sps.  16
##                                  stat p.value
## Abra.alba                       0.999  0.0099 **
## Heteromastus.filiformis         0.978  0.0099 **
## Lagis.koreni                    0.935  0.0099 **
## Trochochaeta.multisetosa        0.895  0.0099 **
## Corbula.gibba                   0.887  0.0099 **
## Euchone.papillosa               0.866  0.0099 **
## Nephtys.hombergii               0.852  0.0198 *
## Phyllodoce..Anaitides..mucosa 0.812  0.0297 *
## Pherusa.plumosa                 0.791  0.0099 **
## Pholoe.assimilis                0.791  0.0099 **
## Terebellides.stroemii           0.790  0.0099 **
## Nephtys.ciliata                 0.773  0.0099 **
## Scalibregma.inflatum            0.707  0.0099 **
## Nemertina                       0.675  0.0198 *
## Ophiura.albida                  0.612  0.0396 *
## Parvicardium.ovale              0.612  0.0495 *
##
##  Group GR2+GR3  #sps.  2
##                                     stat p.value
## Arctica.islandica                  0.946  0.0099 **
## Polydora..Polydora..quadrilobata 0.827  0.0495 *
## ---
```

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

```r
# Phi coefficient analysis
zoobenthos.rel = decostand(zoobenthos, method='total')
phi = multipatt(zoobenthos.rel, cluster=groups, func="r.g", control=how(nperm=100))
summary(phi)
```

```
##
##  Multilevel pattern analysis
##  ---------------------------
##
##  Association function: r.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 103
##  Selected number of species: 27
##  Number of species associated to 1 group: 26
##  Number of species associated to 2 groups: 1
##
##  List of species associated to each combination:
##
##  Group GR1  #sps.  12
##                      stat p.value
## Pygospio.elegans   0.950  0.0099 **
## Macoma.balthica    0.832  0.0099 **
## Jaera.albifrons    0.825  0.0099 **
## Mya.arenaria       0.763  0.0099 **
## Mytilus.edulis     0.762  0.0099 **
## Gammarus.zaddachi  0.595  0.0099 **
## Bathyporeia.pilosa 0.563  0.0396 *
## Tubifex.costatus   0.545  0.0099 **
## Gammarus.oceanicus 0.513  0.0099 **
## Saduria.entomon    0.471  0.0099 **
## Gammarus.salinus   0.468  0.0099 **
## Hydrobia.ulvae     0.460  0.0099 **
##
##  Group GR2  #sps.  11
##                         stat p.value
## Abra.alba              0.832  0.0099 **
## Terebellides.stroemii 0.597  0.0099 **
## Pholoe.assimilis      0.582  0.0099 **
## Nemertina             0.553  0.0198 *
## Lagis.koreni          0.522  0.0099 **
## Parvicardium.ovale    0.517  0.0396 *
## Pherusa.plumosa       0.487  0.0198 *
## Euchone.papillosa     0.465  0.0099 **
## Ampharete.baltica     0.408  0.0396 *
## Ophiura.albida        0.367  0.0297 *
## Scalibregma.inflatum  0.349  0.0099 **
##
##  Group GR3  #sps.  3
##                                  stat p.value
## Scoloplos.armiger               0.521  0.0297 *
## Pontoporeia.femorata            0.516  0.0297 *
```

```
## Polydora..Polydora..quadrilobata 0.515  0.0297 *
##
##  Group GR2+GR3  #sps.  1
##                    stat p.value
## Diastylis.rathkei 0.535  0.0198 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comments about PERMANOVA, IndVal and phi coefficient analyses: Here we see that the PERMANOVA test supports that there are significant differences in zoobenthos composition between the 3 groups that differentiate the sites and some sites in specific years. For IndVal analysis, we saw that there are 34 species that are significantly associated with the groups. For example, species such as Pygospio elegans, Mytilus edulis, and Macoma balthica are strongly associated with GR1. There are 16 species that are significantly linked to GR2, and 3 species show an association with the combination of groups GR2 and GR3. This is interesting as we saw that sites 0 and 2 are sometimes more similar to each other than to themselves in certain years. The phi coefficient analysis showed that the species Pygospio elegans, Mytilus edulis, and Macoma balthica are strongly associated with GR1, which is consistent with the IndVal analysis. Here we see a similar grouping as IndVal.

```r
# Load environmental data

# zoobenthos.env = read.csv(paste0(getwd(),"/Week4-Beta/data/env_data.csv"), row.names=1)
zoobenthos.env = read.csv("data/env_data.csv", header = TRUE, row.names=1)

# Distance matrices of the zoobenthos data
zoobenthos.dist = vegdist(zoobenthos, method="bray")
zoobenthos.env.dist = vegdist(scale(zoobenthos.env), method="euclid")


# Mantel test
mantel(zoobenthos.dist, zoobenthos.env.dist, method="pearson", permutations = 1000)
```

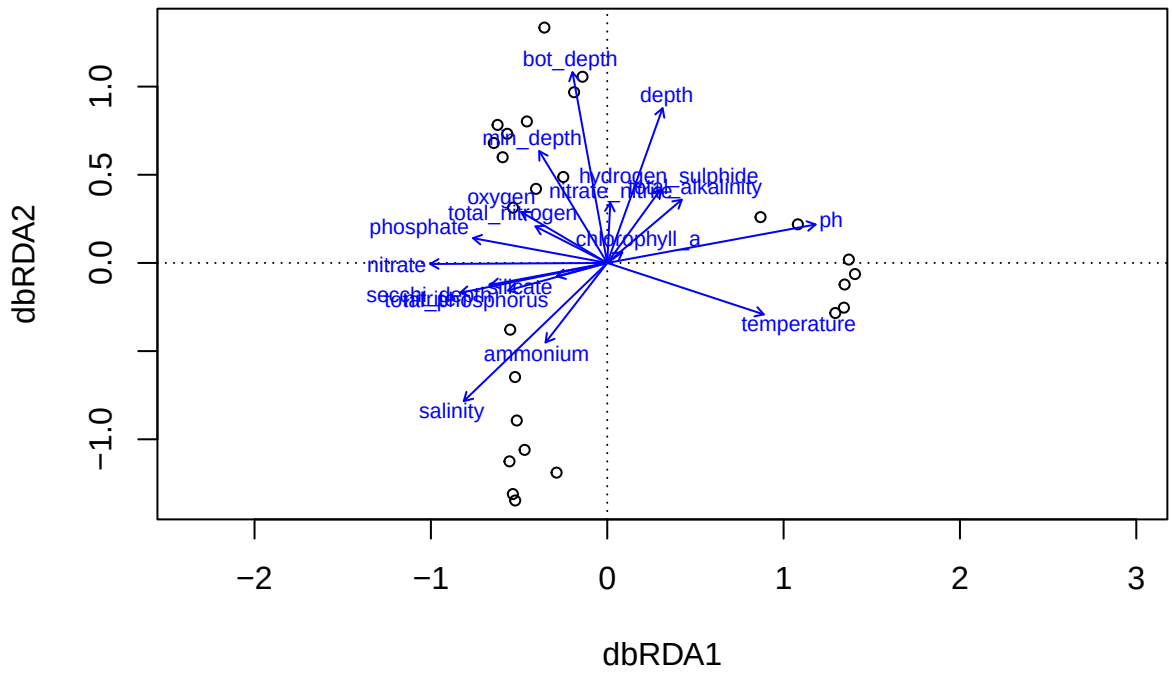
```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = zoobenthos.dist, ydis = zoobenthos.env.dist, method = "pearson",      permutations = 1
##
## Mantel statistic r: 0.451
##       Significance: 0.000999
##
## Upper quantiles of permutations (null model):
##    90%    95%  97.5%    99%
## 0.0759 0.0998 0.1272 0.1572
## Permutation: free
## Number of permutations: 1000
```

Comments about Mantel test: In the Mantel test, we compare the dissimilarities in zoobenthos community composition with the differences in environmental conditions (using Euclidean distances on scaled data). First, we see that the statistic r is 0.451, which is a moderate positive correlation. This means that sites that are more different in their environmental conditions tend to also be more different in their zoobenthos communities. Here, the p-value is 0.0001, which is very small. This indicates that the correlation is statistically significant.

```
# Bray Curtis distance for zoobenthos
zoobenthos.db = vegdist(zoobenthos, method="bray")

# Creat env matrix
env.vars = as.matrix(zoobenthos.env)

# dbRDA
# png("Outputs/dbrda_plot_all.png", width = 4000, height = 3000, res = 500)
zoobenthos.dbrda.modfull = dbrda(zoobenthos.db ~ ., as.data.frame(env.vars))
```

```
##
## Some constraints or conditions were aliased because they were redundant. This
## can happen if terms are linearly dependent (collinear): 'max_depth'
```

```
ordiplot(zoobenthos.dbrda.modfull)
```




```
# dev.off()

# Check the pairwise correlation
psych::corr.test(env.vars)
```

```
## Call:psych::corr.test(x = env.vars)
## Correlation matrix
##                  min_depth max_depth bot_depth secchi_depth depth temperature
```

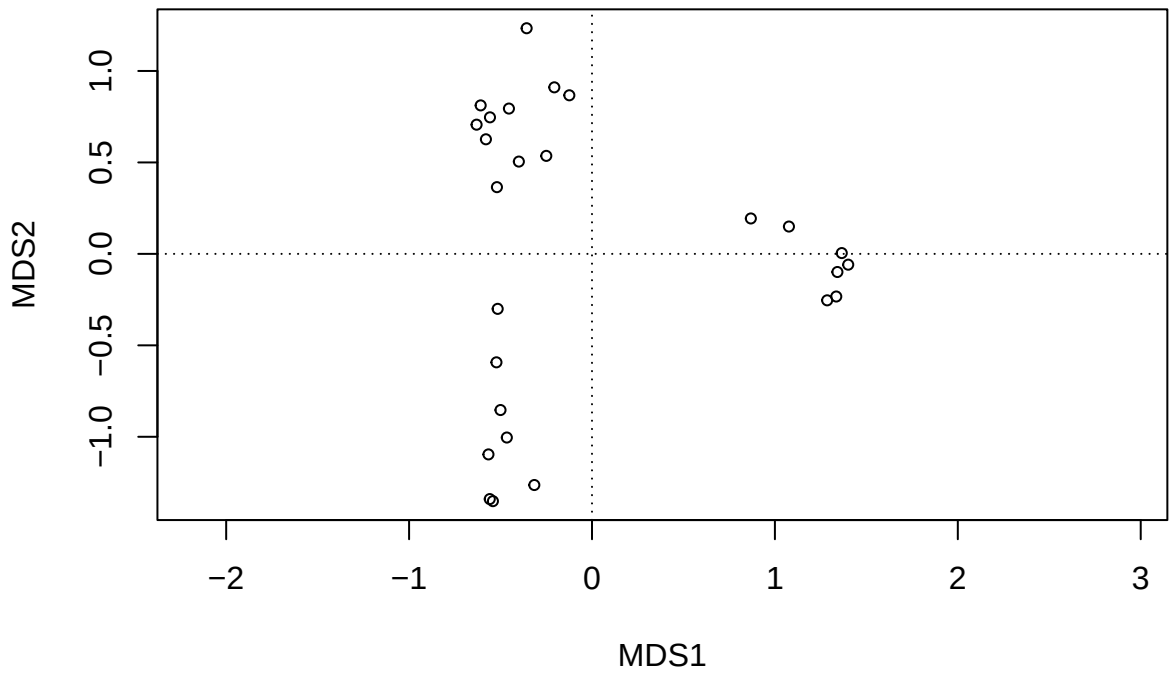
```
## min_depth              1.00      1.00      0.78         0.21  0.52       -0.35
## max_depth              1.00      1.00      0.78         0.21  0.52       -0.35
## bot_depth              0.78      0.78      1.00         0.03  0.77       -0.30
## secchi_depth           0.21      0.21      0.03         1.00  0.04       -0.42
## depth                  0.52      0.52      0.77         0.04  1.00        0.07
## temperature           -0.35     -0.35     -0.30        -0.42  0.07        1.00
## salinity              -0.34     -0.34     -0.63         0.50 -0.66       -0.27
## oxygen                 0.31      0.31      0.34         0.03 -0.16       -0.71
## phosphate              0.19      0.19      0.04         0.36 -0.20       -0.27
## total_phosphorus      -0.26     -0.26     -0.32        -0.02 -0.42       -0.20
## silicate              -0.11     -0.11     -0.29         0.06 -0.26        0.22
## nitrate_nitrite        0.30      0.30      0.32        -0.16  0.27       -0.06
## nitrate                0.06      0.06     -0.06         0.38 -0.32       -0.50
## nitrite               -0.03     -0.03     -0.21         0.27 -0.48       -0.43
## ammonium              -0.21     -0.21     -0.50         0.14 -0.48        0.07
## total_nitrogen         0.36      0.36      0.21         0.27 -0.07       -0.30
## hydrogen_sulphide      0.12      0.12      0.26        -0.40  0.22       -0.04
## ph                    -0.10     -0.10      0.08        -0.60  0.38        0.58
## total_alkalinity       0.11      0.11      0.25        -0.38  0.11        0.15
## chlorophyll_a         -0.09     -0.09     -0.06        -0.41 -0.36       -0.15
##                    salinity oxygen phosphate total_phosphorus silicate
## min_depth             -0.34   0.31      0.19            -0.26    -0.11
## max_depth             -0.34   0.31      0.19            -0.26    -0.11
## bot_depth             -0.63   0.34      0.04            -0.32    -0.29
## secchi_depth           0.50   0.03      0.36            -0.02     0.06
## depth                 -0.66  -0.16     -0.20            -0.42    -0.26
## temperature           -0.27  -0.71     -0.27            -0.20     0.22
## salinity               1.00  -0.11      0.48             0.57     0.45
## oxygen                -0.11   1.00     -0.12            -0.02    -0.54
## phosphate              0.48  -0.12      1.00             0.34     0.69
## total_phosphorus       0.57  -0.02      0.34             1.00     0.38
## silicate               0.45  -0.54      0.69             0.38     1.00
## nitrate_nitrite       -0.25   0.05     -0.04            -0.06    -0.23
## nitrate                0.69   0.14      0.59             0.65     0.39
## nitrite                0.65   0.13      0.55             0.55     0.47
## ammonium               0.60  -0.22      0.51             0.31     0.55
## total_nitrogen         0.17   0.12      0.61             0.14     0.28
## hydrogen_sulphide     -0.44   0.13     -0.33            -0.27    -0.17
## ph                    -0.75  -0.26     -0.59            -0.48    -0.30
## total_alkalinity      -0.51   0.22     -0.36            -0.31    -0.35
## chlorophyll_a         -0.07   0.21      0.19             0.05     0.17
##                    nitrate_nitrite nitrate nitrite ammonium total_nitrogen
## min_depth                     0.30    0.06   -0.03    -0.21           0.36
## max_depth                     0.30    0.06   -0.03    -0.21           0.36
## bot_depth                     0.32   -0.06   -0.21    -0.50           0.21
## secchi_depth                 -0.16    0.38    0.27     0.14           0.27
## depth                         0.27   -0.32   -0.48    -0.48          -0.07
## temperature                  -0.06   -0.50   -0.43     0.07          -0.30
## salinity                     -0.25    0.69    0.65     0.60           0.17
## oxygen                        0.05    0.14    0.13    -0.22           0.12
## phosphate                    -0.04    0.59    0.55     0.51           0.61
## total_phosphorus             -0.06    0.65    0.55     0.31           0.14
## silicate                     -0.23    0.39    0.47     0.55           0.28
## nitrate_nitrite               1.00   -0.03   -0.11    -0.04          -0.15
```

```
## nitrate                    -0.03    1.00    0.82     0.34           0.46
## nitrite                    -0.11    0.82    1.00     0.40           0.32
## ammonium                   -0.04    0.34    0.40     1.00           0.09
## total_nitrogen             -0.15    0.46    0.32     0.09           1.00
## hydrogen_sulphide           0.07   -0.30   -0.08    -0.32          -0.24
## ph                          0.25   -0.77   -0.65    -0.39          -0.34
## total_alkalinity            0.14   -0.25   -0.15    -0.40           0.04
## chlorophyll_a              -0.02    0.08    0.05    -0.12           0.36
##                   hydrogen_sulphide    ph total_alkalinity chlorophyll_a
## min_depth                      0.12 -0.10             0.11         -0.09
## max_depth                      0.12 -0.10             0.11         -0.09
## bot_depth                      0.26  0.08             0.25         -0.06
## secchi_depth                  -0.40 -0.60            -0.38         -0.41
## depth                          0.22  0.38             0.11         -0.36
## temperature                   -0.04  0.58             0.15         -0.15
## salinity                      -0.44 -0.75            -0.51         -0.07
## oxygen                         0.13 -0.26             0.22          0.21
## phosphate                     -0.33 -0.59            -0.36          0.19
## total_phosphorus              -0.27 -0.48            -0.31          0.05
## silicate                      -0.17 -0.30            -0.35          0.17
## nitrate_nitrite                0.07  0.25             0.14         -0.02
## nitrate                       -0.30 -0.77            -0.25          0.08
## nitrite                       -0.08 -0.65            -0.15          0.05
## ammonium                      -0.32 -0.39            -0.40         -0.12
## total_nitrogen                -0.24 -0.34             0.04          0.36
## hydrogen_sulphide              1.00  0.31             0.48          0.05
## ph                             0.31  1.00             0.36          0.11
## total_alkalinity               0.48  0.36             1.00          0.04
## chlorophyll_a                  0.05  0.11             0.04          1.00
## Sample Size
## [1] 26
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                   min_depth max_depth bot_depth secchi_depth depth temperature
## min_depth              0.00      0.00      0.00         1.00  1.00        1.00
## max_depth              0.00      0.00      0.00         1.00  1.00        1.00
## bot_depth              0.00      0.00      0.00         1.00  0.00        1.00
## secchi_depth           0.30      0.30      0.90         0.00  1.00        1.00
## depth                  0.01      0.01      0.00         0.86  0.00        1.00
## temperature            0.08      0.08      0.13         0.03  0.73        0.00
## salinity               0.09      0.09      0.00         0.01  0.00        0.19
## oxygen                 0.13      0.13      0.09         0.89  0.43        0.00
## phosphate              0.34      0.34      0.86         0.07  0.32        0.18
## total_phosphorus       0.19      0.19      0.11         0.94  0.03        0.34
## silicate               0.59      0.59      0.16         0.77  0.21        0.29
## nitrate_nitrite        0.13      0.13      0.11         0.43  0.19        0.76
## nitrate                0.78      0.78      0.77         0.06  0.11        0.01
## nitrite                0.87      0.87      0.30         0.19  0.01        0.03
## ammonium               0.29      0.29      0.01         0.50  0.01        0.72
## total_nitrogen         0.07      0.07      0.31         0.19  0.72        0.13
## hydrogen_sulphide      0.55      0.55      0.20         0.05  0.29        0.85
## ph                     0.61      0.61      0.70         0.00  0.06        0.00
## total_alkalinity       0.59      0.59      0.22         0.06  0.61        0.47
## chlorophyll_a          0.67      0.67      0.78         0.04  0.07        0.47
##                   salinity oxygen phosphate total_phosphorus silicate
```

```
## min_depth             1.00   1.00      1.00             1.00     1.00
## max_depth             1.00   1.00      1.00             1.00     1.00
## bot_depth             0.11   1.00      1.00             1.00     1.00
## secchi_depth          1.00   1.00      1.00             1.00     1.00
## depth                 0.04   1.00      1.00             1.00     1.00
## temperature           1.00   0.01      1.00             1.00     1.00
## salinity              0.00   1.00      1.00             0.37     1.00
## oxygen                0.59   0.00      1.00             1.00     0.77
## phosphate             0.01   0.57      0.00             1.00     0.02
## total_phosphorus      0.00   0.91      0.09             0.00     1.00
## silicate              0.02   0.00      0.00             0.06     0.00
## nitrate_nitrite       0.22   0.82      0.83             0.75     0.25
## nitrate               0.00   0.50      0.00             0.00     0.05
## nitrite               0.00   0.53      0.00             0.00     0.01
## ammonium              0.00   0.28      0.01             0.13     0.00
## total_nitrogen        0.40   0.56      0.00             0.49     0.16
## hydrogen_sulphide     0.02   0.53      0.10             0.18     0.40
## ph                    0.00   0.19      0.00             0.01     0.13
## total_alkalinity      0.01   0.29      0.07             0.12     0.08
## chlorophyll_a         0.74   0.30      0.34             0.81     0.41
##                   nitrate_nitrite nitrate nitrite ammonium total_nitrogen
## min_depth                    1.00    1.00    1.00     1.00           1.00
## max_depth                    1.00    1.00    1.00     1.00           1.00
## bot_depth                    1.00    1.00    1.00     1.00           1.00
## secchi_depth                 1.00    1.00    1.00     1.00           1.00
## depth                        1.00    1.00    1.00     1.00           1.00
## temperature                  1.00    1.00    1.00     1.00           1.00
## salinity                     1.00    0.02    0.06     0.21           1.00
## oxygen                       1.00    1.00    1.00     1.00           1.00
## phosphate                    1.00    0.24    0.62     1.00           0.18
## total_phosphorus             1.00    0.06    0.56     1.00           1.00
## silicate                     1.00    1.00    1.00     0.57           1.00
## nitrate_nitrite              0.00    1.00    1.00     1.00           1.00
## nitrate                      0.87    0.00    0.00     1.00           1.00
## nitrite                      0.59    0.00    0.00     1.00           1.00
## ammonium                     0.83    0.09    0.04     0.00           1.00
## total_nitrogen               0.47    0.02    0.11     0.68           0.00
## hydrogen_sulphide            0.73    0.14    0.69     0.11           0.24
## ph                           0.21    0.00    0.00     0.05           0.09
## total_alkalinity             0.49    0.22    0.46     0.04           0.85
## chlorophyll_a                0.91    0.72    0.81     0.57           0.07
##                   hydrogen_sulphide   ph total_alkalinity chlorophyll_a
## min_depth                      1.00 1.00            1.00             1
## max_depth                      1.00 1.00            1.00             1
## bot_depth                      1.00 1.00            1.00             1
## secchi_depth                   1.00 0.19            1.00             1
## depth                          1.00 1.00            1.00             1
## temperature                    1.00 0.34            1.00             1
## salinity                       1.00 0.00            1.00             1
## oxygen                         1.00 1.00            1.00             1
## phosphate                      1.00 0.27            1.00             1
## total_phosphorus               1.00 1.00            1.00             1
## silicate                       1.00 1.00            1.00             1
## nitrate_nitrite                1.00 1.00            1.00             1
```

```
## nitrate                    1.00 0.00          1.00          1
## nitrite                    1.00 0.06          1.00          1
## ammonium                   1.00 1.00          1.00          1
## total_nitrogen             1.00 1.00          1.00          1
## hydrogen_sulphide          0.00 1.00          1.00          1
## ph                         0.13 0.00          1.00          1
## total_alkalinity           0.01 0.07          0.00          1
## chlorophyll_a              0.82 0.60          0.86          0
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```r
# ================================================
# Try model with constant explained variables
zoobenthos.dbrda.mod0 = dbrda(zoobenthos.db  ~ 1, as.data.frame(env.vars))
ordiplot(zoobenthos.dbrda.mod0) # no vector as we don't use any variable in env.chem
```




```r
# Next, we will test all combination from 0 explanatory variable to full
# This functions returns the one that has lowest AIC
zoobenthos.dbrda = ordiR2step(zoobenthos.dbrda.mod0, zoobenthos.dbrda.modfull, perm.max=200)
```

```
##
## Some constraints or conditions were aliased because they were redundant. This
## can happen if terms are linearly dependent (collinear): 'max_depth'

## Step: R2.adj= 0
```

```
## Call: zoobenthos.db ~ 1
##
##                      R2.adjusted
## <All variables>       0.40984583
## + ph                  0.21640369
## + nitrate             0.17149814
## + salinity            0.15818299
## + nitrite             0.11229503
## + temperature         0.10999716
## + bot_depth           0.10598531
## + phosphate           0.08620862
## + depth               0.07291467
## + min_depth           0.06382860
## + max_depth           0.06382860
## + secchi_depth        0.06367906
## + total_phosphorus    0.06094895
## + ammonium            0.03746229
## + hydrogen_sulphide   0.02788227
## + total_nitrogen      0.02560828
## + oxygen              0.02245370
## + total_alkalinity    0.01248279
## + silicate            0.01100400
## <none>                0.00000000
## + nitrate_nitrite    -0.00909813
## + chlorophyll_a      -0.02682847
##
##      Df    AIC      F Pr(>F)
## + ph  1 53.772 7.9042  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2164037
## Call: zoobenthos.db ~ ph
##
##                      R2.adjusted
## <All variables>        0.4098458
## + bot_depth            0.3405184
## + salinity             0.3054451
## + min_depth            0.2837509
## + max_depth            0.2837509
## + depth                0.2815544
## + nitrate              0.2515670
## + ammonium             0.2385972
## + hydrogen_sulphide    0.2338244
## + total_phosphorus     0.2323457
## + temperature          0.2298373
## + nitrate_nitrite      0.2276313
## + total_nitrogen       0.2261635
## + silicate             0.2227789
## + nitrite              0.2216704
## + phosphate            0.2212616
## + oxygen               0.2190926
## <none>                 0.2164037
## + secchi_depth         0.2137524
```

```
## + total_alkalinity    0.2054901
## + chlorophyll_a       0.1967794
##
##             Df    AIC      F Pr(>F)
## + bot_depth  1 50.182 5.5168  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.3405184
## Call: zoobenthos.db ~ ph + bot_depth
##
##                     R2.adjusted
## <All variables>       0.4098458
## + nitrate             0.3829257
## + min_depth           0.3617994
## + max_depth           0.3617994
## + total_phosphorus    0.3616251
## + total_nitrogen      0.3552496
## + hydrogen_sulphide   0.3546443
## + silicate            0.3517829
## + nitrite             0.3508278
## + salinity            0.3499296
## + phosphate           0.3478617
## + secchi_depth        0.3448757
## + nitrate_nitrite     0.3435697
## + ammonium            0.3423172
## <none>                0.3405184
## + total_alkalinity    0.3312191
## + chlorophyll_a       0.3261141
## + depth               0.3256556
## + oxygen              0.3239526
## + temperature         0.3227967
##
##           Df    AIC      F Pr(>F)
## + nitrate  1 49.298 2.5806  0.008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.3829257
## Call: zoobenthos.db ~ ph + bot_depth + nitrate
##
##                     R2.adjusted
## <All variables>       0.4098458
## + min_depth           0.4060148
## + max_depth           0.4060148
## + hydrogen_sulphide   0.4007186
## + secchi_depth        0.3931849
## + total_phosphorus    0.3928446
## + salinity            0.3895423
## + total_nitrogen      0.3891182
## + nitrate_nitrite     0.3864817
## + nitrite             0.3860522
## + silicate            0.3848200
## + ammonium            0.3844651
```

```
## <none>                0.3829257
## + phosphate           0.3798900
## + total_alkalinity    0.3742776
## + depth               0.3689548
## + temperature         0.3659681
## + oxygen              0.3658756
## + chlorophyll_a       0.3643132
##
##             Df    AIC      F Pr(>F)
## + min_depth  1 49.097 1.8552  0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

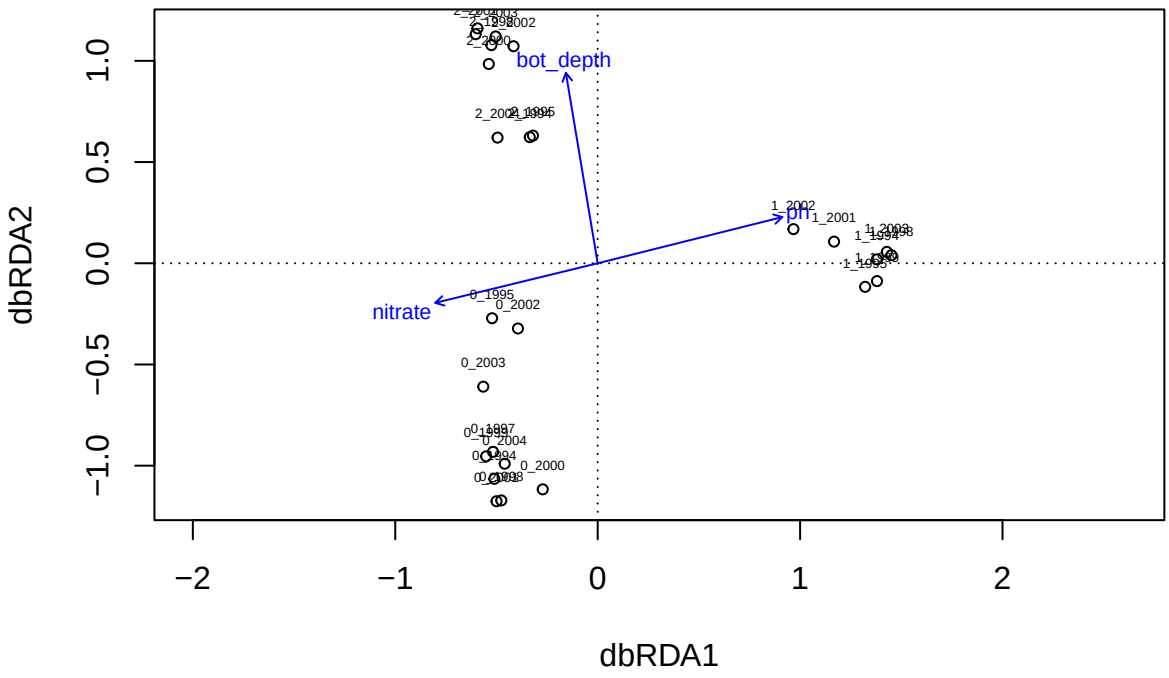
```
# Summary of selected model
zoobenthos.dbrda$call   # zoobenthos.db ~ ph + bot_depth + nitrate + min_depth
```

```
## dbrda(formula = zoobenthos.db ~ ph + bot_depth + nitrate, data = as.data.frame(env.vars))
```

```
zoobenthos.dbrda$anova
```

```
##                  R2.adj Df    AIC      F Pr(>F)
## + ph            0.21640  1 53.772 7.9042  0.002 **
## + bot_depth     0.34052  1 50.182 5.5168  0.002 **
## + nitrate       0.38293  1 49.298 2.5806  0.008 **
## <All variables> 0.40985
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# png("Outputs/dbrda_plot_best.png", width = 4000, height = 3000, res = 500)
ordiplot(zoobenthos.dbrda)
text(zoobenthos.dbrda, display = "sites", cex = 0.4, pos = 3)
```




```
# dev.off()
```

Comments about the constrained ordination: To keep this short, we see that 4 environmental variables pH, bottom depth, and nitrate are significant predictors of zoobenthos community composition. Together, they explain about 41% of the variation. The stepwise model selection confirms that each of these variables contributes significantly, and the ordination plot provides a visual representation of these environmental gradients and their effect on community structure. We also note that these variables are not highly correlated with each other, which is good for the model.

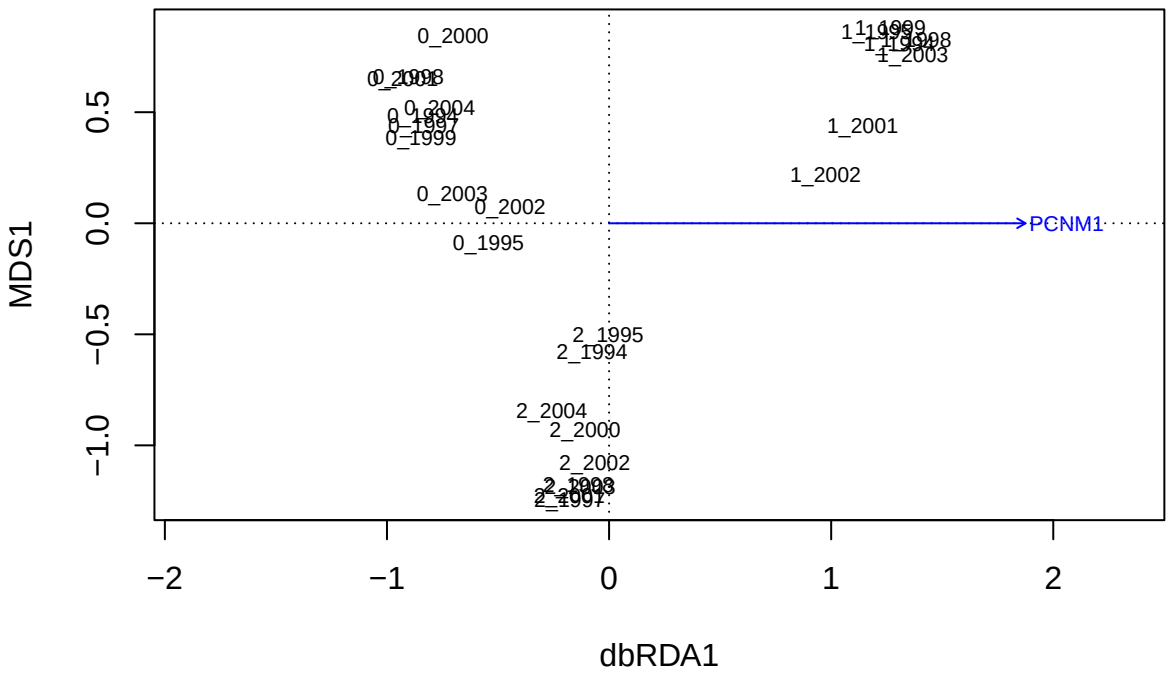
```
# Create a matrix model for the envr data
env.mod = model.matrix( ~ ph + bot_depth + min_depth + nitrate, as.data.frame(env.vars))[,-1]
# env.mod

#cCreate spatial model
# first, weight each site by its relative abundance
rs = rowSums(zoobenthos) / sum(zoobenthos)

# Load spatial data
# zoobenthos.coords = read.csv(paste0(getwd(),"/Week4-Beta/data/site_coords.csv"), header = TRUE, row.n
zoobenthos.coords = read.csv("data/site_coords.csv", header = TRUE, row.names=1)
zoobenthos.pcnmw = pcnm(dist(zoobenthos.coords), w=rs, dist.re=T) # Perform PCNM on the coordinates
zoobenthos.pcnmw$values > 0 # Extract only eigenvectors associated with positive eigenvalues
```

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

```r
# Perform model selection of spatial data
zoobenthos.space = as.data.frame(scores(zoobenthos.pcnmw))
zoobenthos.pcnmw.mod0 = dbrda(zoobenthos.db ~ 1, zoobenthos.space) # no var
zoobenthos.pcnmw.mod1 = dbrda(zoobenthos.db ~ ., zoobenthos.space) # all var
# Stepwise model selection
step.pcnm = ordiR2step(zoobenthos.pcnmw.mod0, zoobenthos.pcnmw.mod1, perm.max=200)
```

```
## Step: R2.adj= 0
## Call: zoobenthos.db ~ 1
##
##                 R2.adjusted
## <All variables>   0.2015856
## + PCNM1           0.2015856
## <none>            0.0000000
##
##         Df    AIC      F Pr(>F)
## + PCNM1  1 54.259 7.3121  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2015856
## Call: zoobenthos.db ~ PCNM1
##
```

```r
# png("Outputs/step_pcnm.png", width = 4000, height = 3000, res = 500)
plot(step.pcnm)
```




```
# dev.off()
step.pcnm$call
```

```
## dbrda(formula = zoobenthos.db ~ PCNM1, data = zoobenthos.space)
```

```
# Check the portion of explained variation of the fish composition using the spatial model
step.pcnm$anova
```

```
##                  R2.adj Df    AIC      F Pr(>F)
## + PCNM1         0.20159  1 54.259 7.3121  0.002 **
## <All variables> 0.20159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Final spatial model
space.mod = model.matrix(~ PCNM1, zoobenthos.space)[,-1]


# ============================================
# Compare env and spatial model
zoobenthos.total.env = dbrda(zoobenthos.db ~ env.mod)
zoobenthos.total.space = dbrda(zoobenthos.db ~ space.mod)

#  PArtial constrained ordination
```

```r
# ?Condition --> control the second explanatory matrix
zoobenthos.env.cond.space = dbrda(zoobenthos.db ~ env.mod + Condition(space.mod))
zoobenthos.space.cond.env = dbrda(zoobenthos.db ~ space.mod + Condition(env.mod))

# Test for significance of the dbRDA fractions
permutest(zoobenthos.env.cond.space,permutations=999)
```

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = zoobenthos.db ~ env.mod + Condition(space.mod))
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     4  2.8123 3.2143  0.001 ***
## Residual 20  4.3746
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = zoobenthos.db ~ space.mod + Condition(env.mod))
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     1  0.3038 1.3889  0.162
## Residual 20  4.3746
```

```r
permutest(zoobenthos.total.env,permutations=999)
```

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = zoobenthos.db ~ env.mod)
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     4  4.6981 5.2721  0.001 ***
## Residual 21  4.6784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

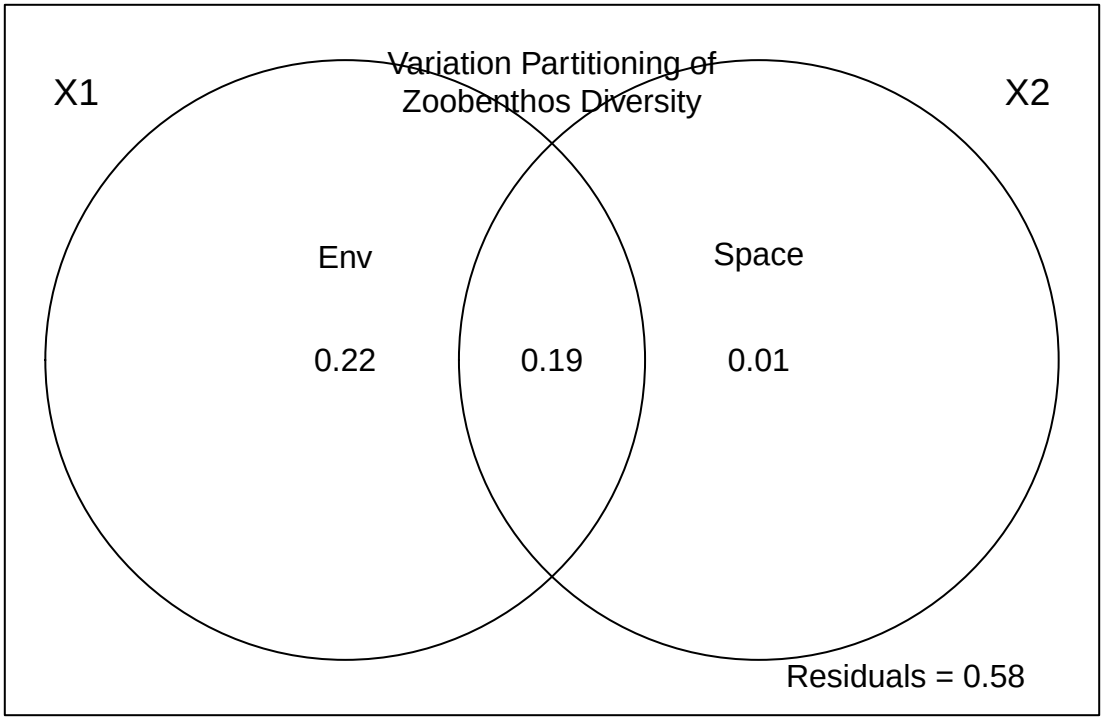
```r
permutest(zoobenthos.total.space,permutations=999)
```

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = zoobenthos.db ~ space.mod)
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     1  2.1896 7.3121  0.001 ***
## Residual 24  7.1869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Calculate the fraction of variation explained by space alone, by env alone and by both, and by neithe
zoobenthos.varpart = varpart(zoobenthos.db, env.mod, space.mod)
zoobenthos.varpart
```

```
##
## Partition of squared Bray distance in dbRDA
##
## Call: varpart(Y = zoobenthos.db, X = env.mod, space.mod)
##
## Explanatory tables:
## X1:  env.mod
## X2:  space.mod
##
## No. of explanatory tables: 2
## Total variation (SS): 9.3765
## No. of observations: 26
##
## Partition table:
##                       Df R.squared Adj.R.squared Testable
## [a+c] = X1             4   0.50105       0.40601     TRUE
## [b+c] = X2             1   0.23352       0.20159     TRUE
## [a+b+c] = X1+X2        5   0.53345       0.41681     TRUE
## Individual fractions
## [a] = X1|X2            4                 0.21523     TRUE
## [b] = X2|X1            1                 0.01080     TRUE
## [c]                    0                 0.19079    FALSE
## [d] = Residuals                          0.58319    FALSE
## ---
## Use function 'dbrda' to test significance of fractions of interest
```

```r
# png("Outputs/zoobenthos_varpart.png", width = 4000, height = 3000, res = 500)
par(mar = c(2,2,2,2))
plot(zoobenthos.varpart)
text(1, 0.25, "Space")
text(0, 0.25, "Env")
mtext("Variation Partitioning of\nZoobenthos Diversity", side = 3, line =-3)
```




```
# dev.off()
```

Comments about the variation partitioning including spatial data:

For the spatial model PCNM We identified that there is only a single eigenvector, PCNM1, best captured the spatial structure. This spatial predictor (PCNM1) explains about 20% of the variation in zoobenthos community composition (adjusted R^2= 0.202).

For the model using spatial variables controlling for environment variables, We see that the model is significant (adjusted R-squared 20%). This means that that the environmental variables (pH, bottom depth, nitrate and minimal depth) significantly explain community variation even after accounting for spatial structure.

In contrast, for the model using environemental variables controlling for spatial factors, we see that the model has less explainatory power (R^2 adjusted = 20%). This suggests that the spatial variation is not a significant driver of the zoobenthos biodiversities between sites.

If we take each model excluding the other set of explainatory variables: For the model using only environemental variables, interestingly, the adjusted R squared value decreased by half (21%) For the model using only spatial variable, expectedly, the adjusted R squared value is very low (1%). We also see that when using both environment and space together, the model explains 40.7% of the variation (R&2 adjusted). This shows that we need to take in account additional factors that influence the variation of biodiversities between sites of the zoobenthos.