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

Jocelyn Huang; Z620: Quantitative Biodiversity, Indiana University

11 February, 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).

```
remove(list = ls())
getwd()
```

```
## [1] "/cloud/project/QB2025_Huang/Week4-Beta"
```

```
setwd("/cloud/project/QB2025_Huang/Week4-Beta")
library(vegan)
```

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

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

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

```
library(ade4)
library(indicspecies)
```

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
data(doubs)
fish <- doubs$fish[-8,]
```

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.

```
quality <- c(rep("HQ", 13), rep("MQ", 5), rep("LQ", 6), rep("MQ", 5)) #Create "Factors" vector
adonis2(fish ~ quality, method = "bray", permutations = 999) #PERMANOVA
```

```
## 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
```

```
indval <- multipatt(fish, cluster = quality, func = "IndVal.g",
                    control = how(nperm = 999)) #Indicator Value
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   0.018 *
##
##  Group HQ+MQ  #sps.  2
##       stat p.value
## Satr 0.860   0.007 **
## 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   0.001 ***
## Cyca 0.866   0.001 ***
## Acce 0.866   0.002 **
## Lele 0.863   0.003 **
## Titi 0.853   0.004 **
## Chto 0.829   0.003 **
## Rham 0.829   0.002 **
## Anan 0.829   0.002 **
## Eslu 0.827   0.024 *
## Pefl 0.806   0.020 *
## Blbj 0.791   0.005 **
## Scer 0.766   0.015 *
## Abbr 0.750   0.008 **
## Icme 0.661   0.031 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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   0.002 **
##
##  Group LQ  #sps.  2
##       stat p.value
## Alal 0.693   0.001 ***
## Ruru 0.473   0.032 *
##
##  Group MQ  #sps.  4
##       stat p.value
## Anan 0.571   0.007 **
## Spbi 0.557   0.009 **
## Chto 0.542   0.011 *
## Icme 0.475   0.030 *
##
##  Group LQ+MQ  #sps.  9
##       stat p.value
## Legi 0.658   0.002 **
## Baba 0.645   0.003 **
## Rham 0.600   0.004 **
## Acce 0.594   0.005 **
## Cyca 0.586   0.003 **
## Chna 0.571   0.005 **
## Blbj 0.571   0.004 **
## Gogo 0.523   0.020 *
## Abbr 0.499   0.021 *
## ---
## 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***: Since the p-value is significant (p=0.001) with 999 permutation, we can conclude from the PERMANOVA test that fish species composition indeed varies with river quality. According to the Indicator Value of each species in each river quality group, blageon is a strong indicator species of "moderate quality" river with indicator score of 0.686. Brown trout (IndVal = 0.86) and minnow (0.859) are considered strong indicator species of river habitat that has at least moderate quality (high and moderate quality). On the other hand, multiple species are considered strong indicators species of "low and moderate quality" rivers, which means their presences indicate that the river does not have very high quality. Results of phi coefficient analysis show that species that prefer high quality rivers include minnow, brown trout, and stone loach, species that prefer low quality are common roach and bleak; while those prefer moderate quality are many: eel, spirlin,

toxostoma, and black bullhead. IndVal and phi coefficient analysis disagree with each other as IndVal results in blageon as a strong indicator species for moderate river quality, but phi coefficient results show that blageon is not strongly associated with moderate-quality river. However, phi coefficient is especially agreeing with the cluster dendogram made in the last worksheet.

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.

```
# Define 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)
```

```
##
## 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%
## 0.0979 0.1353 0.1650 0.1981
## Permutation: free
## Number of permutations: 999
```

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?

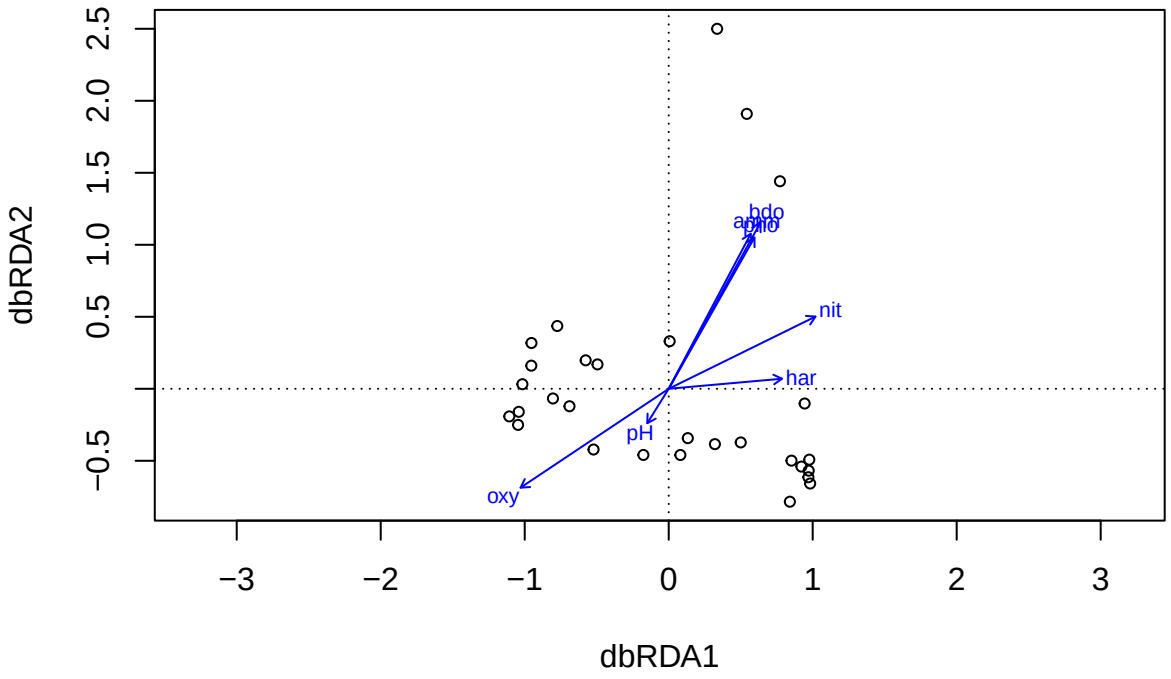
> ***Answer 2***:
> Result of the Mantel test suggests that fish diversity is significantly (p = 0.001) correlated with stream environmental conditions, and such correlation is moderately strong and positive (r = 0.604). Since the hypothesis states that fish community composition is more diverse in stream sections that have higher quality, Mantel test result proves this.

ii. Constrained Ordination

In the R code chunk below, do the following:

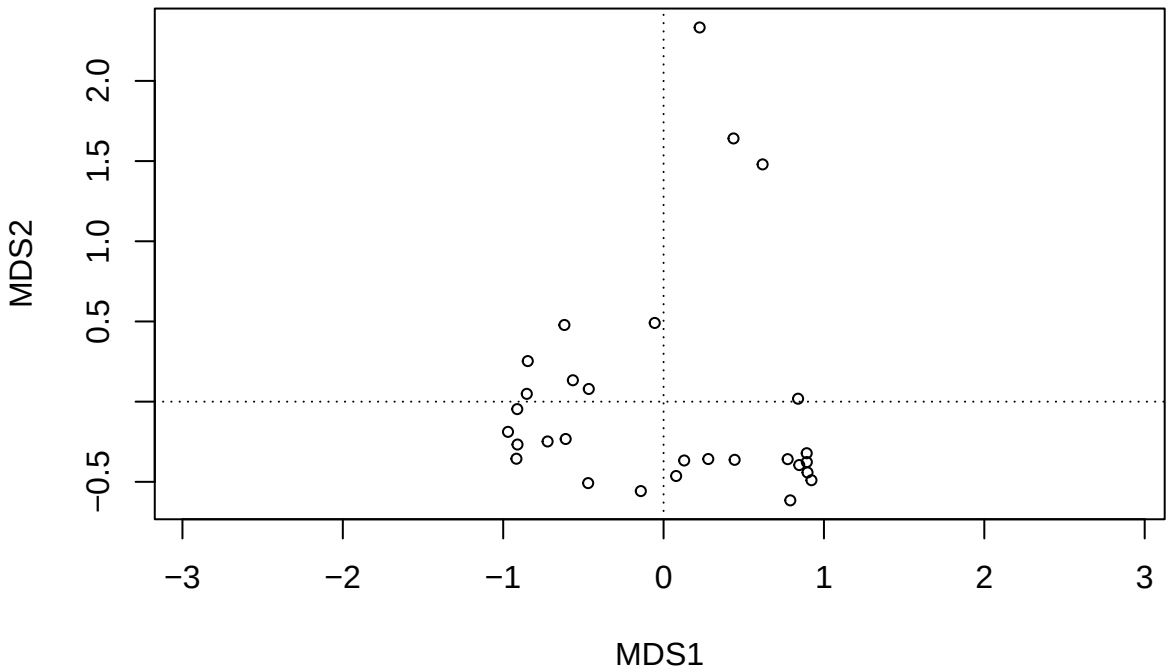
1. create an environmental matrix of the water chemistry data included in the `doubs` dataset using forward and reverse selection of variables,
2. conduct a redundancy analysis on the fish assemblages of the Doubs River,
3. use a permutation test to determine the significance of the constrained analysis,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

```
env.chem <- as.matrix(doubs$env[-8, 5:11]) # Environmental matrix
doubs.dbrda <- dbrda(fish.dist ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda) # test
```




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

doubs.dbrda.mod0 <- dbrda(fish.dist ~ 1, as.data.frame(env.chem)) # Model only the intercept
ordiplot (doubs.dbrda.mod0) # This is PCoA
```




```
doubs.dbrda.mod1 <- dbrda(fish.dist ~ ., as.data.frame (env.chem)) # Full model
# Function returns the model with lowest AIC value
```

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

```
## Step: R2.adj= 0
## Call: fish.dist ~ 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.dist ~ 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.dist ~ 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.dist ~ oxy + bdo + nit
##
##                 R2.adjusted
## + amm             0.5415705
## <All variables>   0.5303258
## + pho             0.5277128
## + har             0.5218852
## <none>            0.4980793
## + pH              0.4843267
```

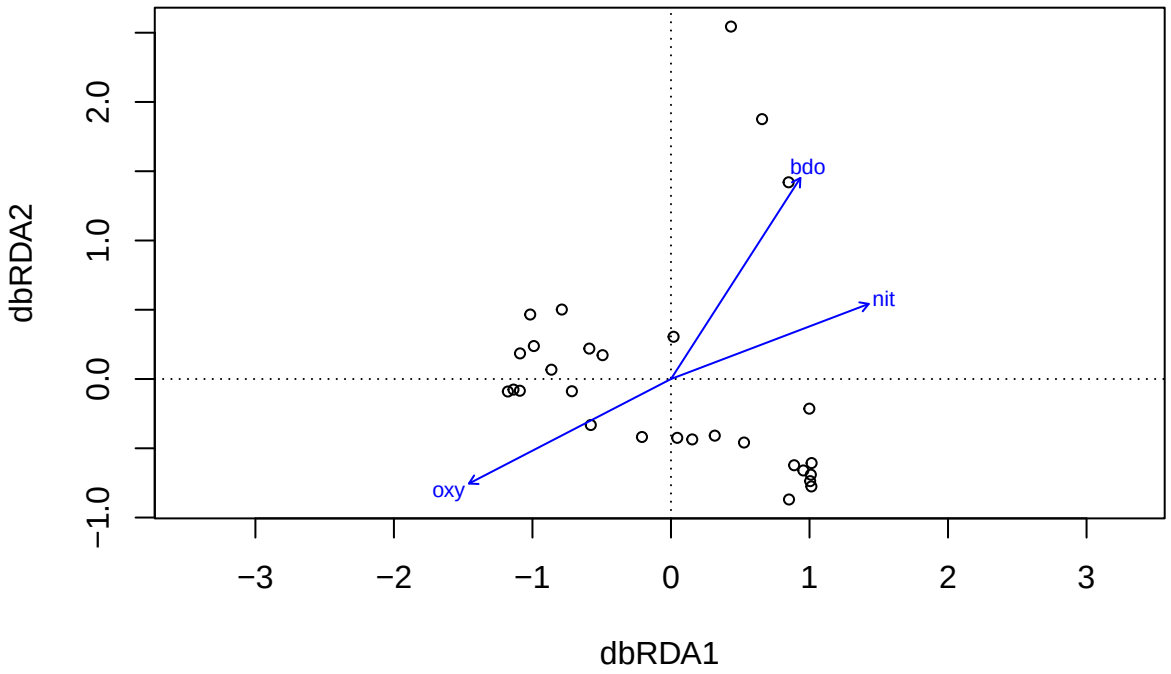
```
# What is the model:
doubs.dbrda$call
```

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

```
doubs.dbrda$anova
```

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

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




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

```
##
## Permutation test for dbrda under reduced model
```

```
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.dist ~ oxy + bdo + nit, data =
## as.data.frame(env.chem))
## 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
```

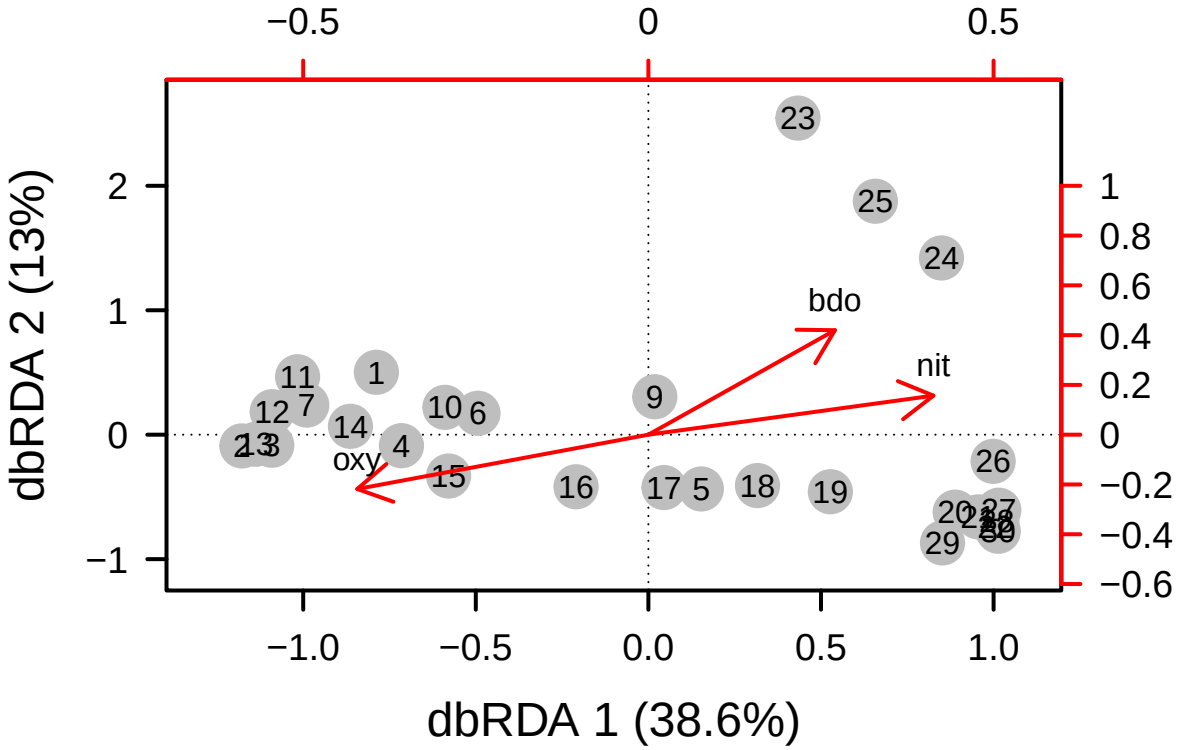
```
envfit(doubs.dbrda, env.chem[, c(4,6,7)], perm = 999) # correlation
```

```
##
## ***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
```

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

# Define plot parameters
par(mar = c(5,5,4,4) + 0.1)
plot(scores(doubs.dbrda, display = "wa"), xlim = c(-1.3, 1.1),
     ylim = c(-1.1, 2.7), xlab = paste("dbRDA 1 (", dbrda.explainvar1, "%)",
     sep = ""), ylab = paste("dbRDA 2 (", dbrda.explainvar2, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5,
     cex.axis = 1.2, axes = FALSE)
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)
points(scores(doubs.dbrda, display = "wa"), pch = 19, cex = 3, bg = "gray",
       col = "gray")
text(scores(doubs.dbrda, display = "wa"),
     labels = row.names(scores(doubs.dbrda, display = "wa")))
vectors <- scores(doubs.dbrda, display = "bp")
arrows(0, 0, vectors[, 1], vectors [, 2],
       lwd = 2, lty = 1, length = 0.2, col = "red")
text(vectors[, 1], vectors[, 2], pos = 3,
     labels = row.names(vectors))
axis(side = 3, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red",
     lwd = 2.2, at = pretty(range(vectors[, 1])) * 2,
```

```
     labels = pretty(range(vectors[, 1])))
axis(side = 4, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red",
     lwd = 2.2, at = pretty(range(vectors[, 2])) * 2,
     labels = pretty(range(vectors[, 2])))
```



Question 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 selected model with lowest AIC includes environmental variables oxygen level ("oxy"), "bdo", and nitrogen level ("nit"). Since the anova results show that these three parameters all have significant p-value of 0.002, all three environmental factors seem to contribute significantly to the overall fish community structure. According to the distance-based ordination plot, oxygen level seems to contribute a lot to sites 1, 14, 4, 10, 6, 15, and 16, while effect of bdo is more pronounced in site 9, and that of nitrogen level in sites 17, 5, 18, and 19. On the other hand, oxygen level (being the longest vector) has the strongest influence on the fish community composition among the selected three environmental conditions, and bdo has the weakest. While the sites that are strongly influenced by oxygen level are also points that are closer to each other on the ordination plot, this could also mean that the oxygen levels in these sites are very similar and which lead to similar composition.

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]
rs <- rowSums(fish)/sum(fish) # relative abundance of each site
doubs.pcnmw <- pcnm(dist(doubs$xy[-8,]), w = rs, dist.ret = T)
doubs.pcnmw$values > 0 # filter for the non-negatice eigenvalues as these are the only that are meaning
```

```
##  [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
```

```
# Model selection that drops the redundant eigenvalues:
doubs.space <- as.data.frame(scores(doubs.pcnmw))
doubs.pcnm.mod0 <- dbrda(fish.dist ~ 1, doubs.space)
doubs.pcnm.mod1 <- dbrda(fish.dist ~ ., doubs.space)
step.pcnm <- ordiR2step(doubs.pcnm.mod0, doubs.pcnm.mod1, perm.max = 200)
```

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

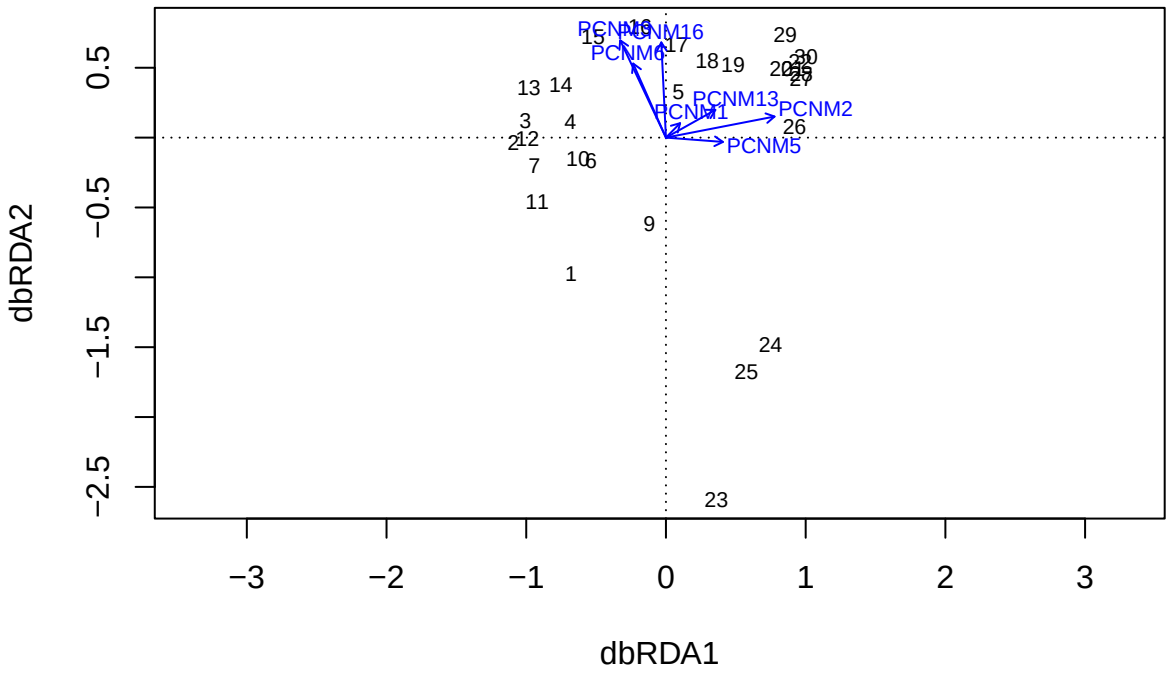
```
## Step: R2.adj= 0.2353704
## Call: fish.dist ~ PCNM2
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM3           0.3429270
## + PCNM5           0.3057368
## + PCNM1           0.2885396
## + PCNM16          0.2786746
## + PCNM14          0.2744520
## + PCNM15          0.2692809
## + 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  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.342927
## Call: fish.dist ~ 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.dist ~ 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  0.006 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.4721469
## Call: fish.dist ~ 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.dist ~ 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  0.004 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.5767968
## Call: fish.dist ~ 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   0.05 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.6043089
## Call: fish.dist ~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13 + PCNM16 +      PCNM6
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM8           0.6248697
## + PCNM12          0.6208788
## + PCNM10          0.6170988
## + PCNM7           0.6142419
```

```
## + PCNM15          0.6140369
## + PCNM9           0.6107110
## <none>            0.6043089
## + PCNM17          0.6037430
## + PCNM11          0.5978305
## + PCNM4           0.5963667
## + PCNM14          0.5932113
##
##         Df    AIC     F Pr(>F)
## + PCNM8  1 34.219 2.151  0.094 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(step.pcnm) # Visualize the biplot for spatial factors' influence on sites composition
```




```
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  0.006 **
## + PCNM13         0.52124  1 39.346 3.4612  0.024 *
## + PCNM16         0.57680  1 36.480 4.0192  0.004 **
## + PCNM6          0.60431  1 35.182 2.5296  0.050 *
## <All variables> 0.62601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

# Use partial constrained ordination, which requires two explanatory matrix instead of one (2 layers of
doubs.total.env <- dbrda(fish.dist ~ env.mod)
```

```r
doubs.total.space <- dbrda(fish.dist ~ space.mod)
doubs.env.cond.space <- dbrda(fish.dist ~ env.mod + Condition(space.mod))
doubs.space.cond.env <- dbrda(fish.dist ~ space.mod + Condition(env.mod))

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

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.dist ~ 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, permutation = 999)
```

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.dist ~ 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, permutation = 999)
```

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.dist ~ 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
```

```r
permutest(doubs.total.space, permutation = 999)
```

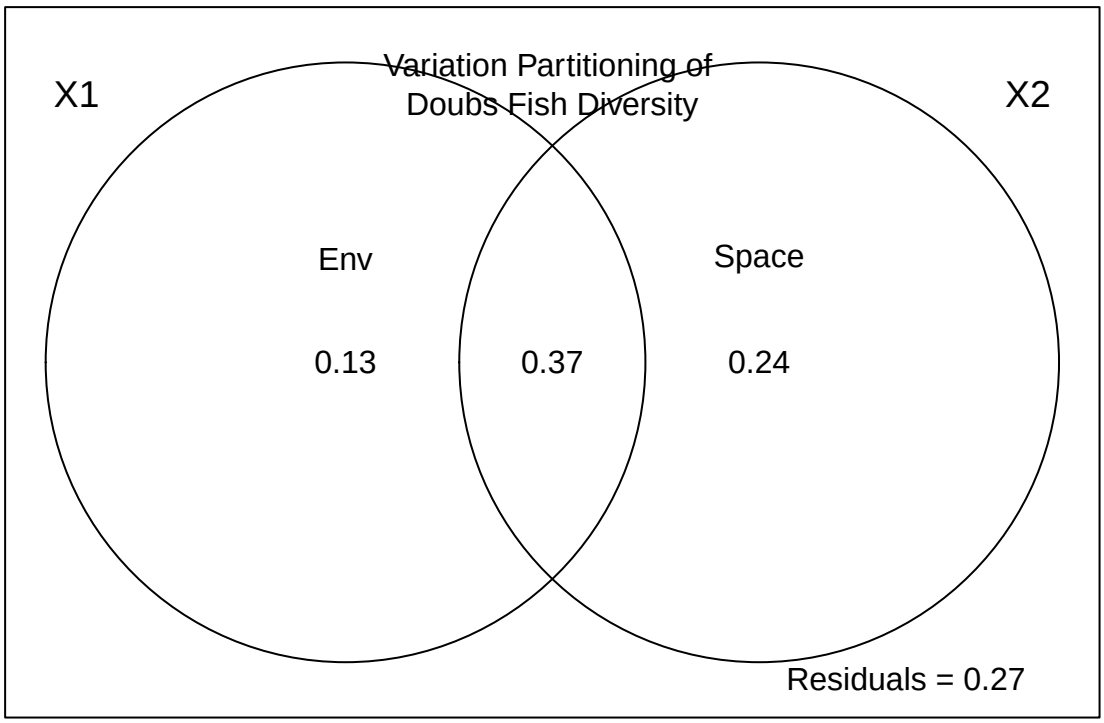
```
##
## Permutation test for dbrda under reduced model
```

```
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.dist ~ 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
```

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

```
##
## Partition of squared Bray distance in dbRDA
##
## Call: varpart(Y = fish.dist, 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***: The variation partitioning plot shows that 13% of the total variation in the fish community composition among sites are explained by the environmental condition alone, while 24% of the variation can be explained by space alone. 37% of the variation in community structure is likely driven by spcially structured environmental variation (the shared effect of environment and space).

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.

```
### Load data: ----
# Load species-by-site matrix:
bac <- read.table("/cloud/project/QB2025_Huang/Group-project/bacteria_div.txt", header = TRUE, sep = "\t
# drop till 1 site by location:
bac <- bac[!grepl("_3|_2|_1", rownames(bac)), ]
# Load environmental data
bac.env <- read.table("/cloud/project/QB2025_Huang/Group-project/bac_env.txt", header = TRUE, sep = "\t"
# Load spatial data
bac.xy <- read.table("/cloud/project/QB2025_Huang/Group-project/bac_xy.txt", header = TRUE, sep = "\t",

# Make distance matrices:
# Bray-Curtis:
bac.bc <- vegdist(bac, method = "bray", upper = TRUE, diag = TRUE)

### PERMANOVA: ----
land_type <- bac.env$Landscape
habitat <- bac.env$Habitat
adonis2(bac ~ land_type, method = "bray", permutation = 999)
```

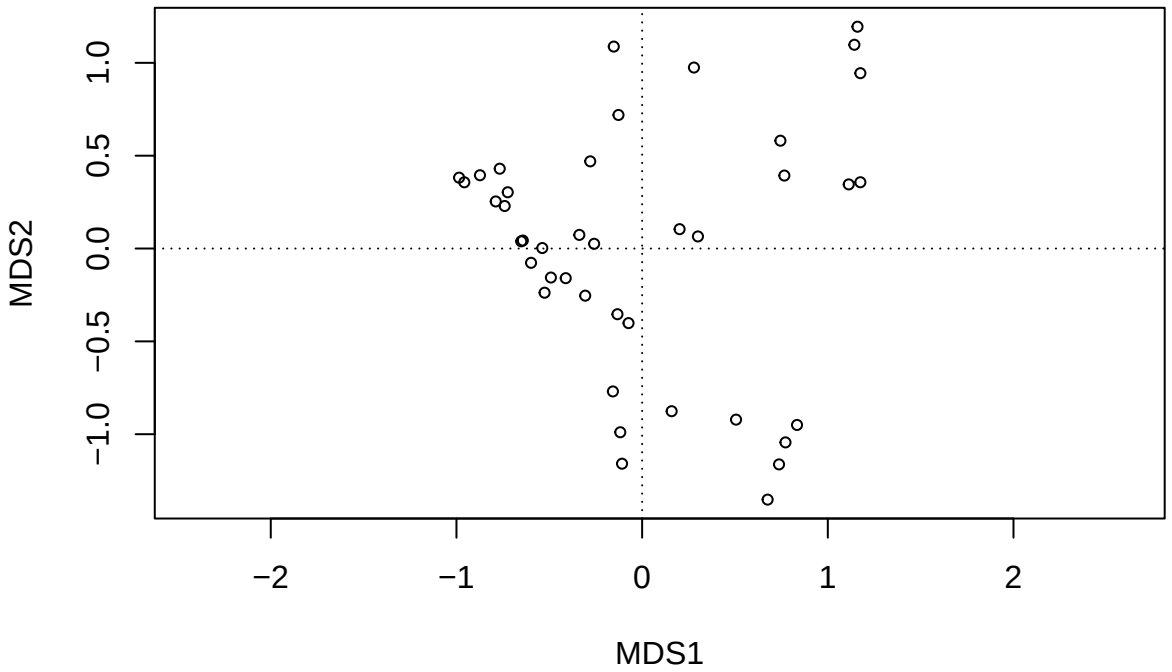
```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = bac ~ land_type, permutations = 999, method = "bray")
##          Df SumOfSqs      R2      F Pr(>F)
## Model     2   0.6006 0.07295 1.4952  0.091 .
## Residual 38   7.6313 0.92705
## Total    40   8.2319 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
adonis2(bac ~ habitat, method = "bray", permutation = 999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = bac ~ habitat, permutations = 999, method = "bray")
##          Df SumOfSqs      R2      F Pr(>F)
## Model     3   0.4394 0.05338 0.6955  0.887
## Residual 37   7.7925 0.94662
## Total    40   8.2319 1.00000
```

```
### Constrained Ordination ----
bac.envcon <- as.matrix(bac.env[3:5]) #Continous env conditions
bac.dbrda <- dbrda(bac.bc ~ ., as.data.frame(bac.envcon))

bac.dbrda.mod0 <- dbrda(bac.bc ~ 1, as.data.frame(bac.envcon)) # Model only the intercept
ordiplot (bac.dbrda.mod0) # This is PCoA
```




```
bac.dbrda.mod1 <- dbrda(bac.bc ~ ., as.data.frame (bac.envcon)) # Full model
# Function returns the model with lowest AIC value
bac.dbrda <- ordiR2step(bac.dbrda.mod0, bac.dbrda.mod1, perm.max = 200)
```

```
## Step: R2.adj= 0
## Call: bac.bc ~ 1
##
##                     R2.adjusted
## <All variables>    0.087967024
## + MAP              0.033955737
## <none>             0.000000000
## + MAT             -0.002921586
## + average_Temp_DL -0.003899852
##
##       Df    AIC     F Pr(>F)
## + MAP  1 86.962 2.406  0.036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.03395574
## Call: bac.bc ~ MAP
##
##                   R2.adjusted
## <All variables>    0.08796702
## + MAT              0.06038480
## <none>             0.03395574
## + average_Temp_DL  0.03007787
##
##       Df    AIC     F Pr(>F)
## + MAT  1 86.759 2.097  0.042 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.0603848
## Call: bac.bc ~ MAP + MAT
##
##                   R2.adjusted
## <All variables>    0.08796702
## + average_Temp_DL  0.08796702
## <none>             0.06038480
##
##                   Df    AIC      F Pr(>F)
## + average_Temp_DL  1 86.445 2.1492  0.026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.08796702
## Call: bac.bc ~ MAP + MAT + average_Temp_DL
##
```

```
bac.dbrda$call
```

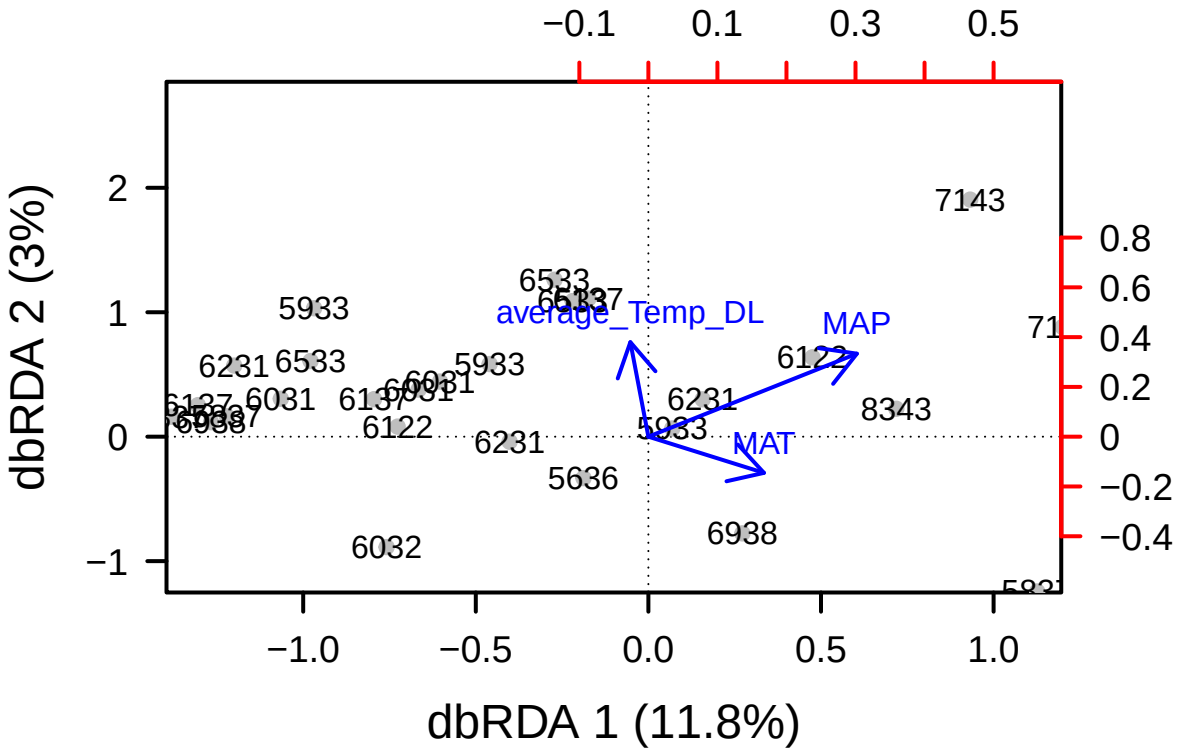
```
## dbrda(formula = bac.bc ~ MAP + MAT + average_Temp_DL, data = as.data.frame(bac.envcon))
```

```
bac.dbrda$anova
```

```
##                     R2.adj Df    AIC      F Pr(>F)
## + MAP             0.033956  1 86.962 2.4060  0.036 *
## + MAT             0.060385  1 86.759 2.0970  0.042 *
## + average_Temp_DL 0.087967  1 86.445 2.1492  0.026 *
```

```
## <All variables>   0.087967
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
bc.explainvar1 <-  round(bac.dbrda$CCA$eig[1] /
                           sum(c(bac.dbrda$CCA$eig, bac.dbrda$CA$eig)), 3) * 100
bc.explainvar2 <- round(bac.dbrda$CCA$eig[2] /
                          sum(c(bac.dbrda$CCA$eig, bac.dbrda$CA$eig)), 3) * 100
# Plot the ordination plot:
par(mar = c(5,5,4,4) + 0.1)
plot(scores(bac.dbrda, display = "wa"), xlim = c(-1.3, 1.1), ylim = c(-1.1, 2.7),
     xlab = paste("dbRDA 1 (", bc.explainvar1, "%)",sep = ""),
     ylab = paste("dbRDA 2 (", bc.explainvar2, "%)", sep = ""),
     pch = 16, cex = 2.0, type = "n", cex.lab = 1.5,
     cex.axis = 1.2, axes = FALSE)
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)
points(scores(bac.dbrda, display = "wa"), pch = 19, cex = 1, bg = "gray",
       col = "gray")
text(scores(bac.dbrda, display = "wa"),
     labels = substr(row.names(scores(bac.dbrda, display = "wa")), 1, 4))
bc.vectors <- scores(bac.dbrda, display = "bp")
arrows(0, 0, bc.vectors[, 1], bc.vectors[, 2],
       lwd = 2, lty = 1, length = 0.2, col = "blue")
text(bc.vectors[, 1], bc.vectors[, 2], pos = 3,
     labels = row.names(bc.vectors), col = "blue")
axis(side = 3, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red",
     lwd = 2.2, at = pretty(range(bc.vectors[, 1])) * 2,
     labels = pretty(range(bc.vectors[, 1])))
axis(side = 4, lwd.ticks = 2, cex.axis = 1.2, las = 1, col = "red",
     lwd = 2.2, at = pretty(range(bc.vectors[, 2])) * 2,
     labels = pretty(range(bc.vectors[, 2])))
```



Answer: We performed constrained ordination analysis on the effect of three environmental variables (average local temperature, mean annual precipitation, mean annual temperature) on the bacterial community structure. The ANOVA results for the distanced-based redundancy analysis show that all three environmental variables contribute significantly to the bacterial community structure. Since the vector of Mean annual precipitation (MAP) is the longest, this means that the influence of precipitation to the community composition of bacteria is the strongest among the three environmental factors. Mean annual precipitation vector draws across sites 5933, 6231, and 6122, which means the composition of these three sites are strongly associated with the amount of mean precipitation of the sites. We also performed a permutation ANOVA (PERMANOVA) for the categorical data (habitat type and land use) of the sites, and found that neither is significantly influencing the bacterial community composition. However, with a p-value of 0.106, land use type (whether the sites is used for agriculture, urban, or near-natural) has a tendency of influencing the bacteria diversity.