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

Ford Fishman; Z620: Quantitative Biodiversity, Indiana University

23 April, 2021

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, it is *imperative* that you **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 '8.BetaDiversity' folder.
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 (**8.BetaDiversity_2_Worksheet.Rmd**) with all code blocks filled out and questions answered) and the PDF output of `Knitr` (**8.BetaDiversity_2_Worksheet.pdf**).

The completed exercise is due on **Wednesday, April 23rd, 2021 before 09:00 AM**.

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 "*/8.BetaDiversity*" folder, and
4. load the `vegan` R package (be sure to install if needed).

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

```
## [1] "C:/Users/fordf/OneDrive/Documents/GitHub/QB2021_Fishman/2.Worksheets/8.BetaDiversity"
```

```
setwd("~/GitHub/QB2021_Fishman/2.Worksheets/8.BetaDiversity/")
library('vegan')
```

```
## Warning: package 'vegan' was built under R version 4.0.5
```

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

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

```
## This is vegan 2.5-7
```

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: package 'ade4' was built under R version 4.0.5
```

```
data(doubs)
str(doubs, max.level = 1)
```

```
## List of 4
##  $ env    :'data.frame': 30 obs. of  11 variables:
##  $ fish   :'data.frame': 30 obs. of  27 variables:
##  $ xy     :'data.frame': 30 obs. of  2 variables:
##  $ species:'data.frame': 27 obs. of  4 variables:
```

```
fish <- doubs$fish
fish <- 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))

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

```
##
## Call:
## adonis(formula = fish ~ quality, permutations = 999, method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
```

```
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## quality    2    3.0947 1.54733   10.97 0.45765  0.001 ***
## Residuals 26    3.6674 0.14105         0.54235
## Total     28    6.7621                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library("indicspecies")
```

```
## Warning: package 'indicspecies' was built under R version 4.0.5
```

```
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    0.02 *
##
##  Group HQ+MQ  #sps.  2
##       stat p.value
## Satr 0.860   0.006 **
## Phph 0.859   0.013 *
##
##  Group LQ+MQ  #sps.  20
##       stat p.value
## Alal 0.935   0.001 ***
## Gogo 0.933   0.001 ***
## Ruru 0.916   0.001 ***
## Legi 0.901   0.001 ***
## Baba 0.895   0.001 ***
## Chna 0.866   0.001 ***
## Spbi 0.866   0.001 ***
## Cyca 0.866   0.001 ***
## Acce 0.866   0.002 **
## Lele 0.863   0.003 **
## Titi 0.853   0.003 **
## Chto 0.829   0.001 ***
## Rham 0.829   0.003 **
## Anan 0.829   0.004 **
## Eslu 0.827   0.030 *
```

```
## Pefl 0.806   0.009 **
## Blbj 0.791   0.003 **
## Scer 0.766   0.010 **
## Abbr 0.750   0.011 *
## Icme 0.661   0.032 *
## ---
## 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   0.002 **
## Satr 0.650   0.001 ***
##
##  Group LQ  #sps.  2
##       stat p.value
## Alal 0.693   0.001 ***
## Ruru 0.473   0.030 *
##
##  Group MQ  #sps.  4
##       stat p.value
## Anan 0.571   0.009 **
## Spbi 0.557   0.004 **
## Chto 0.542   0.010 **
## Icme 0.475   0.040 *
##
##  Group LQ+MQ  #sps.  9
##       stat p.value
## Legi 0.658   0.004 **
## Baba 0.645   0.004 **
## Rham 0.600   0.004 **
## Acce 0.594   0.005 **
## Cyca 0.586   0.006 **
## Chna 0.571   0.006 **
## Blbj 0.571   0.010 **
## Gogo 0.523   0.010 **
## Abbr 0.499   0.033 *
```

```
## ---
## 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***: The PERMANOVA suggests that the fish species composition does vary with habitat quality. The IndVal and phi analyses did recover one cluster that was somewhat similar with both Phph and Satr, (though phi had an additional species). This cluster does appear distinct in the PCoA performed last week. Besides this, the clusters are somewhat different.

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.

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

env.dist <- vegdist(scale(doubs$env[-8,]), method = "euclid")

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

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = fish.dist, ydis = env.dist)
##
## Mantel statistic r: 0.604
##       Significance: 0.001
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.111 0.149 0.182 0.216
## 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***: The results suggest that there is a significant correlation between fish communities and environmental variables. If these conditions are correlated with stream quality, that would make sense. Perhaps these variables influence stream quality.

ii. Constrained Ordination

In the R code chunk below, do the following:

1. create an environmental matrix of the water chemistry data included in the `doubs` dataset using forward and reverse selection of variables,
2. conduct a redundancy analysis on the fish assemblages of the Doubs River,
3. use a permutation test to determine the significance of the constrained analysis,

4. use a permutation test to determine the correlation of each environmental factor on the constrained axes,
5. calculate the explained variation on the first and second constrained axes,
6. plot the constrained ordination results including labeled points for each site, and
7. add vectors that demonstrate the influence of each environmental factor the constrained ordination.

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

doubs.dbrda1 <- dbrda(fish.dist ~ ., as.data.frame(env.chem))
doubs.dbrda0 <- dbrda(fish.dist ~ 1, as.data.frame(env.chem))

doubs.dbrda <- ordiR2step(doubs.dbrda0, scope=formula(doubs.dbrda1), 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
```

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

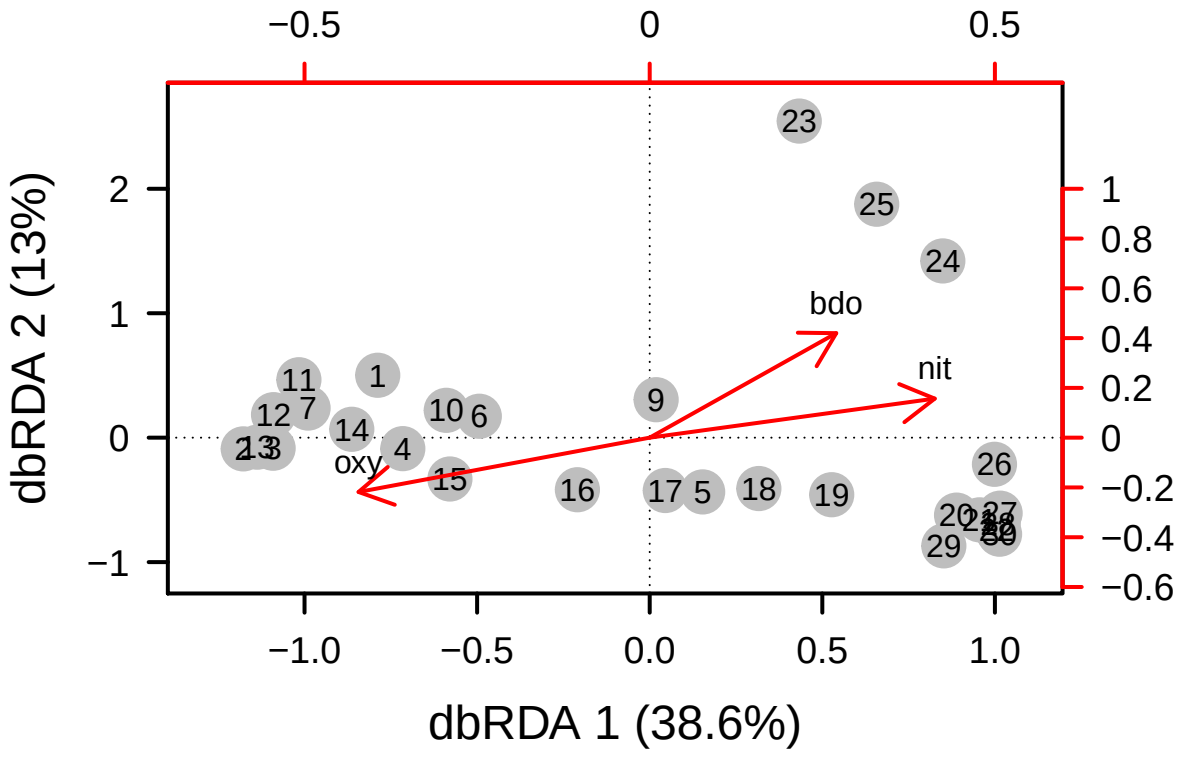
```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.dist ~ oxy + bdo + nit, data =
## as.data.frame(env.chem))
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     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)
```

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

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

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

vals <- scores(doubs.dbrda, display="wa")
plot(vals, xlim=c(-1.3, 1.1), ylim=c(-1.1, 2.7),
            xlab=paste0("dbRDA 1 (", dbrda.explainvar1, "%)"),
            ylab=paste0("dbRDA 2 (", dbrda.explainvar2, "%)"),
            pch=16, cex=2.0, type="n", cex.lab=1.5, cex.axis=1.2, axes=F
            )
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(vals, pch=19, cex=3, bg="grey", col="grey")
text(vals, labels = row.names(vals))

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 primary variables contributing to the variation are oxygen, nitrogen, and `bdo` (as well as variables correlated with `bdo`).

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.

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

# 2
rs <- rowSums(fish)/sum(fish)

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

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  0.008 **
## ---
## 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  0.008 **
## ---
## 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  0.018 *
## ---
```

```
## 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   0.01 **
## ---
## 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  0.054 .
## ---
## 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]

# 3
doubs.total.env <- dbrda(fish.dist ~ env.mod)
doubs.total.space <- dbrda(fish.dist ~ space.mod)
```

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

#4
permutest(doubs.env.cond.space, permutations = 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, permutations = 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, permutations = 999)
```

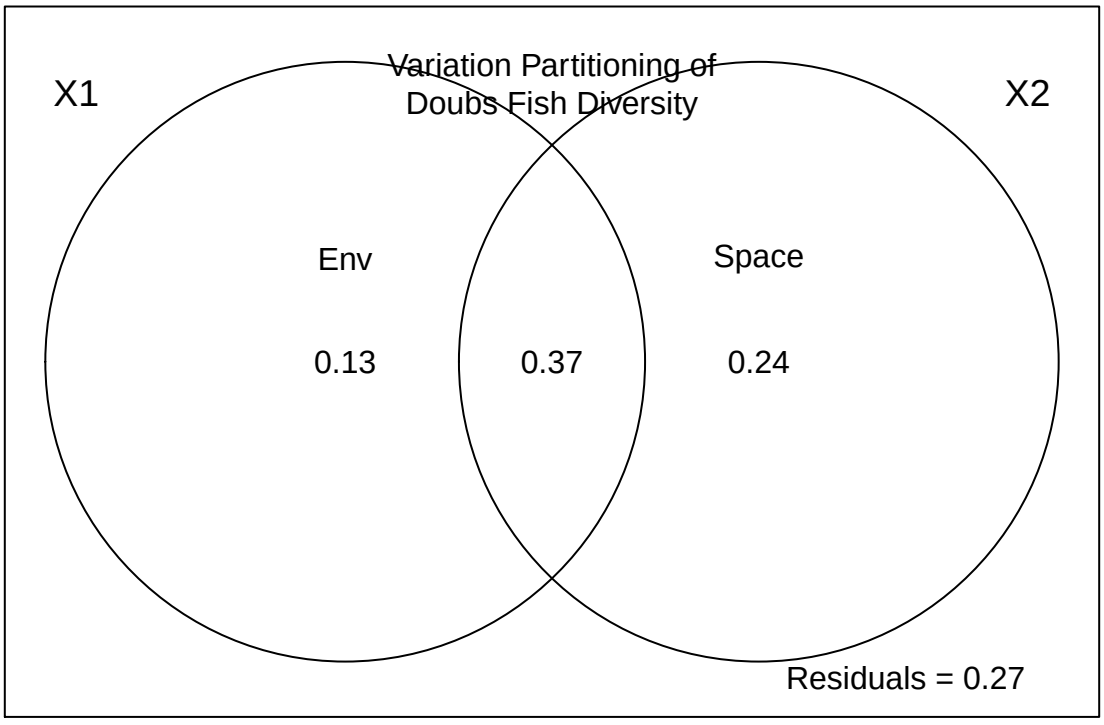
```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.dist ~ env.mod)
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     3  3.7317 10.262  0.001 ***
## Residual 25  3.0304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
permutest(doubs.total.space, permutations = 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
```

```r
#5
doubs.varpart <- varpart(fish.dist, env.mod, space.mod)

#6
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 largest proportion of variation is due to the spatially structured environment, followed by the residuals, space, and finally, the environment.

SYNTHESIS

As in the previous worksheet, use the **mobsim** package from the DataWrangling module to simulate two local communities each containing 1000 individuals (N) and 25 species (S), but with one having a random spatial distribution and the other having a patchy spatial distribution. Take ten (10) subsamples from each site using the quadrat function and answer the following questions:

1) Perform a PERMANOVA to test whether or not the spatial distribution of species affects species composition.

```
library(mobsim)
```

```
## Warning: package 'mobsim' was built under R version 4.0.4
```

```
com.ran <- sim_poisson_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))
comm_mat.ran <- sample_quadrats(com.ran, n_quadrats = 10, quadrat_area = 0.1,
               method = "random", avoid_overlap = T,plot = F)
```

```
## Warning in sample_quadrats(com.ran, n_quadrats = 10, quadrat_area = 0.1, : Cannot find a sampling lay
##                                                 Install the package spatstat for an improved method for r
##                                                 Use less quadrats or smaller quadrat area, or set avoid_
```

```
# obtain site by species
sbs.ran <- comm_mat.ran$spec_dat

com.clust <- sim_thomas_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))

comm_mat.clust <- sample_quadrats(com.clust, n_quadrats = 10, quadrat_area = 0.1,
               method = "random", avoid_overlap = T,plot = F)
```

```
## Warning in sample_quadrats(com.clust, n_quadrats = 10, quadrat_area = 0.1, : Cannot find a sampling l
##                                                 Install the package spatstat for an improved method for r
##                                                 Use less quadrats or smaller quadrat area, or set avoid_
```

```
# obtain site by species
sbs.clust <- comm_mat.clust$spec_dat

distrib <- c(rep("rand",10), rep("patch",10))
df.total <- rbind(sbs.ran, sbs.clust)
adonis(df.total~distrib, method="bray", permutations=999)
```

```
##
## Call:
## adonis(formula = df.total ~ distrib, permutations = 999, method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## distrib    1   0.93655 0.93655  13.788 0.43374  0.001 ***
## Residuals 18   1.22268 0.06793         0.56626
## Total     19   2.15923                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

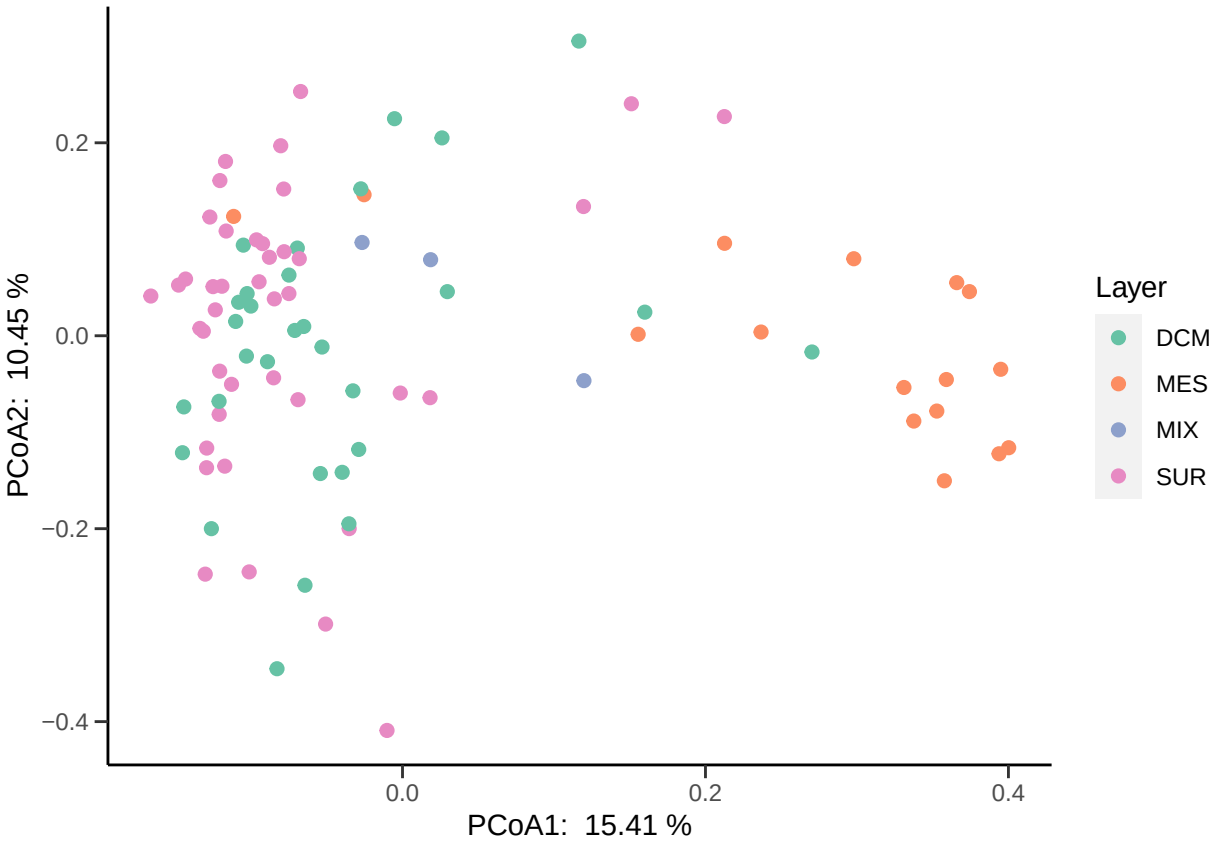
2) Load the dataset you are using for your Team Project. Perform an ordination to visualize your dataset. Using this ordination, develop some hypotheses relevant to β-diversity. Use a statistic to test one of these hypotheses. Succinctly explain the finding and its relevance to your system.

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
df <- read.table("data/env-geo-div-class.txt", sep = "\t", header = TRUE,row.names = 1)
vcs.o <-read.csv("data/vcs.csv", row.names = 1)
vcs <- as.data.frame(t(vcs.o))
vcs <- subset(vcs, ! rownames(vcs) %in% c("MSP114"))
df1 <- subset(df, ! rownames(df) %in% c("MSP114"))


# numerical features
num_cols <- unlist(lapply(df, is.numeric))

# numerical df
df.num <- df1[, num_cols]
df.sparse <- subset(df.num, select=-c(Richness,Evenness,ShannonH))


#convert to relative abundance table
vcs.rel <- vcs/rowSums(vcs)
vcs.db <- vegdist(vcs.rel, method="bray")

vcs.pcoa <- cmdscale(vcs.db, eig=T, k=3)

pcoa2_eig <- (vcs.pcoa$eig)[1:2]/sum(vcs.pcoa$eig)
site.pcoa <- data.frame(vcs.pcoa$points)[1:2]
site.pcoa <- cbind(site.pcoa, df1$Layer)
colnames(site.pcoa) <- c("PCoA1","PCoA2","Layer")

ggplot(site.pcoa, aes(PCoA1, PCoA2))+
  geom_point(aes(color = Layer), size = 2)+
  scale_color_brewer(palette = "Set2")+
  labs(x=paste("PCoA1: ", round(100*pcoa2_eig[1],2),"%"), y = paste("PCoA2: ", round(100*pcoa2_eig[2],2
      theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"),
        axis.ticks.length = unit(5,"pt"),)
```



>: The variation along the first axis is primarily associated with these depth categories, specifically the mesopelagic zone (`MES`). I predict that the rest of the variation is driven by nutrient differences from samples to sample.

```
env.data <- read.csv("data/metadata-imputed.csv", row.names = 1)
dbrda1 <- dbrda(vcs.db ~ ., as.data.frame(env.data))
dbrda0 <- dbrda(vcs.db ~ 1, as.data.frame(env.data))

dbrda <- ordiR2step(dbrda0, scope=formula(dbrda1), perm.max=200)
```

```
## Step: R2.adj= 0
## Call: vcs.db ~ 1
##
##                 R2.adjusted
## <All variables> 0.239543078
## + Depth         0.086962869
## + Nitrate       0.086688128
## + Temperature   0.083388743
## + Phosphate     0.082227200
## + Oxygen        0.052284391
## + Latitude_N    0.045583124
## + Longitude_E   0.035090310
## + Chlorophyll_a 0.023288542
## + Nitrite       0.008446059
## <none>          0.000000000
##
##         Df    AIC      F Pr(>F)
```

```
## + Depth  1 240.66 9.4769  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.08696287
## Call: vcs.db ~ Depth
##
##                 R2.adjusted
## <All variables>  0.23954308
## + Latitude_N     0.13114530
## + Temperature    0.12756423
## + Oxygen         0.12360158
## + Phosphate      0.12240354
## + Longitude_E    0.12233552
## + Nitrate        0.11791244
## + Nitrite        0.09895100
## + Chlorophyll_a  0.09853169
## <none>           0.08696287
##
##              Df    AIC      F Pr(>F)
## + Latitude_N  1 237.17 5.4749  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.1311453
## Call: vcs.db ~ Depth + Latitude_N
##
##                 R2.adjusted
## <All variables>   0.2395431
## + Phosphate       0.1655662
## + Temperature     0.1649808
## + Oxygen          0.1608059
## + Nitrate         0.1585886
## + Longitude_E     0.1564760
## + Nitrite         0.1435822
## + Chlorophyll_a   0.1408352
## <none>            0.1311453
##
##             Df    AIC      F Pr(>F)
## + Phosphate  1 234.49 4.5888  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.1655662
## Call: vcs.db ~ Depth + Latitude_N + Phosphate
##
##                 R2.adjusted
## <All variables>   0.2395431
## + Oxygen          0.1902595
## + Temperature     0.1874882
## + Longitude_E     0.1852672
## + Chlorophyll_a   0.1757525
## + Nitrite         0.1695995
## + Nitrate         0.1695021
```

```
## <none>            0.1655662
##
##          Df    AIC      F Pr(>F)
## + Oxygen  1 232.73 3.6226  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.1902595
## Call: vcs.db ~ Depth + Latitude_N + Phosphate + Oxygen
##
##                 R2.adjusted
## <All variables>   0.2395431
## + Longitude_E     0.2105869
## + Temperature     0.2039998
## + Chlorophyll_a   0.2007219
## + Nitrite         0.1928422
## <none>            0.1902595
## + Nitrate         0.1893234
##
##               Df    AIC      F Pr(>F)
## + Longitude_E  1 231.38 3.1887  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2105869
## Call: vcs.db ~ Depth + Latitude_N + Phosphate + Oxygen + Longitude_E
##
##                 R2.adjusted
## <All variables>   0.2395431
## + Temperature     0.2259622
## + Chlorophyll_a   0.2217673
## + Nitrite         0.2124659
## <none>            0.2105869
## + Nitrate         0.2094238
##
##               Df    AIC      F Pr(>F)
## + Temperature  1 230.53 2.6686  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2259622
## Call: vcs.db ~ Depth + Latitude_N + Phosphate + Oxygen + Longitude_E +      Temperature
##
##                 R2.adjusted
## <All variables>   0.2395431
## + Chlorophyll_a   0.2372994
## + Nitrite         0.2275638
## <none>            0.2259622
## + Nitrate         0.2258622
##
##                 Df    AIC      F Pr(>F)
## + Chlorophyll_a  1 230.11 2.2338  0.018 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Step: R2.adj= 0.2372994
## Call: vcs.db ~ Depth + Latitude_N + Phosphate + Oxygen + Longitude_E +      Temperature + Chlorophyll
##
##                 R2.adjusted
## <All variables>   0.2395431
## + Nitrate         0.2389549
## + Nitrite         0.2379572
## <none>            0.2372994
##
##           Df    AIC      F Pr(>F)
## + Nitrate  1 230.81 1.1784  0.226
```

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

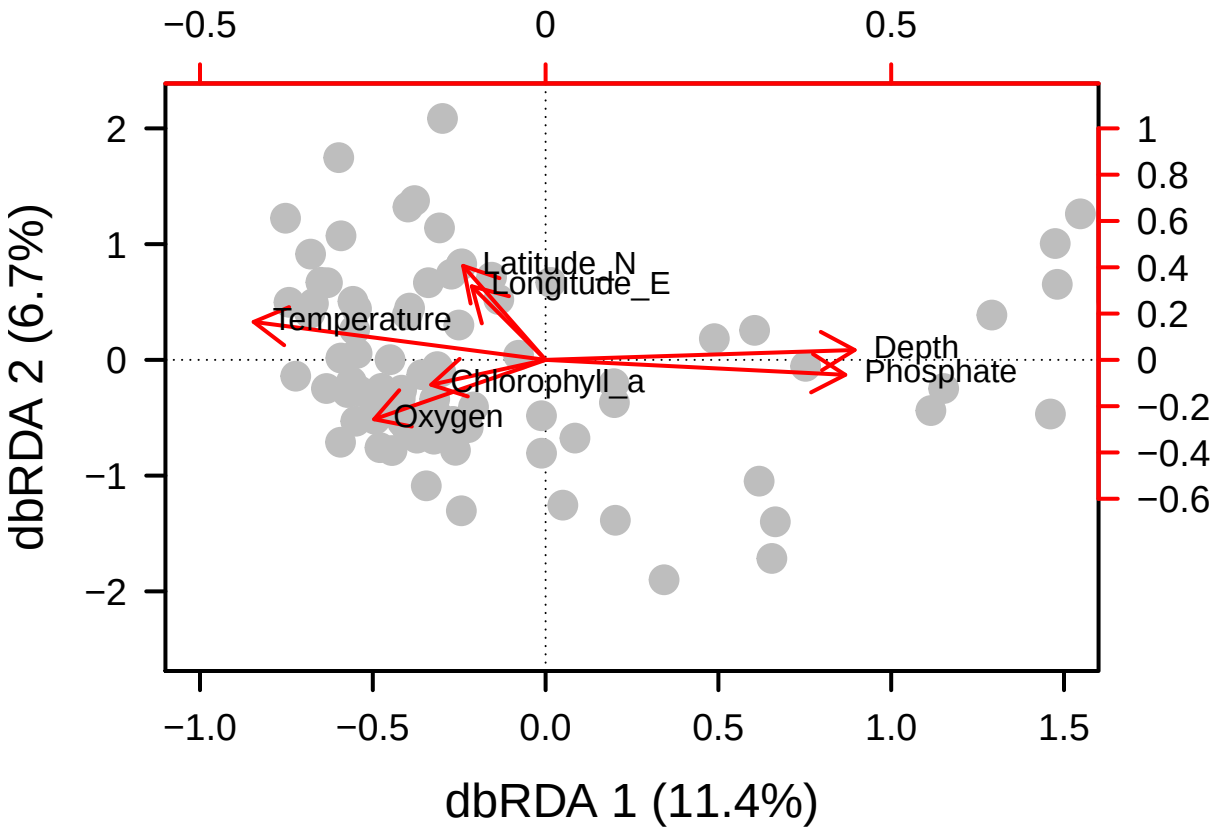
```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = vcs.db ~ Depth + Latitude_N + Phosphate + Oxygen
## + Longitude_E + Temperature + Chlorophyll_a, data =
## as.data.frame(env.data))
## Permutation test for all constrained eigenvalues
##          Df Inertia      F Pr(>F)
## Model     7  4.6179 4.9558  0.001 ***
## Residual 82 10.9157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## ***VECTORS
##
##                dbRDA1   dbRDA2     r2 Pr(>r)
## Temperature -0.92990  0.36781 0.6329  0.001 ***
## Nitrite      0.81090  0.58519 0.0267  0.304
## Phosphate    0.98504 -0.17232 0.5881  0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 999
```

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


vals <- scores(dbrda, display="wa")
par(mar=c(5,4,2,4) + 0.1)
plot(vals,
     xlim=c(-1, 1.5), ylim=c(-2.5, 2.2),
     xlab=paste0("dbRDA 1 (", dbrda.explainvar1, "%)"),
     ylab=paste0("dbRDA 2 (", dbrda.explainvar2, "%)"),
     pch=16, cex=2.0, type="n", cex.lab=1.5, cex.axis=1.2, axes=F
```

```
)
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(vals, pch=19, cex=2, bg="grey", col="grey")

vectors <- scores(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=4, 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])))
```



>: Variation on the primary axis is mostly driven by depth, phosphate, and temperature, while the biogeographical location explains much of the remaining variation. Therefore, spatial factors likely are more important than environmental factors.