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

Caroline Edwards; 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] "/Users/carolineedwards/quant_bio/GitHub/QB2021_Edwards/2.Worksheets/8.BetaDiversity"
```

```
setwd("~/quant_bio/GitHub/QB2021_Edwards/2.Worksheets/8.BetaDiversity/")
package.list<-c('vegan','ade4','viridis','gplots','BiodiversityR','indicspecies','mobsim')
for (package in package.list){
  if (!require(package, character.only = TRUE, quietly =TRUE)){
    install.packages(package)
    library(package, character.only = TRUE)
  }
}
```

```
## This is vegan 2.5-6
```

```
## Warning: package 'ade4' was built under R version 3.6.2
```

```
## Warning: package 'gplots' was built under R version 3.6.2
```

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

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

```
## Warning: package 'BiodiversityR' was built under R version 3.6.2
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
## BiodiversityR 2.12-3: Use command BiodiversityRGUI() to launch the Graphical User Interface;
## to see changes use BiodiversityRGUI(changeLog=TRUE, backward.compatibility.messages=TRUE)
```

```
## Warning: package 'mobsim' was built under R version 3.6.2
```

2) LOADING DATA

Load dataset

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

```r
# note, pleae do not print the dataset when submitting
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.

```r
quality<- c(rep("HQ", 13), rep("MQ", 5), rep("LQ", 6), rep("MQ", 5))
fish<-doubs$fish
fish<-fish[-8,]
adonis(fish ~ quality, method = "bray", permutations = 999)
```

```
##
## Call:
## adonis(formula = fish ~ quality, permutations = 999, method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## quality    2    3.0947 1.54733   10.97 0.45765  0.001 ***
## Residuals 26    3.6674 0.14105         0.54235
## Total     28    6.7621                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
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   0.022 *
##
##  Group HQ+MQ  #sps.  2
##       stat p.value
## Satr 0.860   0.005 **
## Phph 0.859   0.011 *
##
##  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   0.001 ***
## Lele 0.863   0.002 **
## Titi 0.853   0.006 **
## Chto 0.829   0.001 ***
## Rham 0.829   0.001 ***
## Anan 0.829   0.001 ***
## Eslu 0.827   0.018 *
## Pefl 0.806   0.016 *
## Blbj 0.791   0.003 **
## Scer 0.766   0.008 **
## Abbr 0.750   0.008 **
## Icme 0.661   0.034 *
## ---
## 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   0.024 *
##
##  Group MQ  #sps.  4
##       stat p.value
## Anan 0.571   0.006 **
## Spbi 0.557   0.005 **
## Chto 0.542   0.011 *
## Icme 0.475   0.025 *
##
##  Group LQ+MQ  #sps.  9
##       stat p.value
## Legi 0.658   0.002 **
## Baba 0.645   0.002 **
## Rham 0.600   0.006 **
## Acce 0.594   0.004 **
## Cyca 0.586   0.006 **
## Chna 0.571   0.009 **
## Blbj 0.571   0.007 **
## Gogo 0.523   0.019 *
## 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***: The PERMANOVA shows that 45% of the variation can be explained by the quality level of the site. The IndVal shows that two species are associated with HQ+MQ, while the MQ+LQ habitats have 20 species associated with them. The The high quality habitats have only a few fish associated with them whereas low and medium quality habitats have many more fish associated with them.

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

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

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = fish.dist, ydis = env.dist)
##
## Mantel statistic r: 0.604
##       Significance: 0.001
##
## Upper quantiles of permutations (null model):
##    90%    95%  97.5%    99%
## 0.0903 0.1330 0.1662 0.1856
## 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?

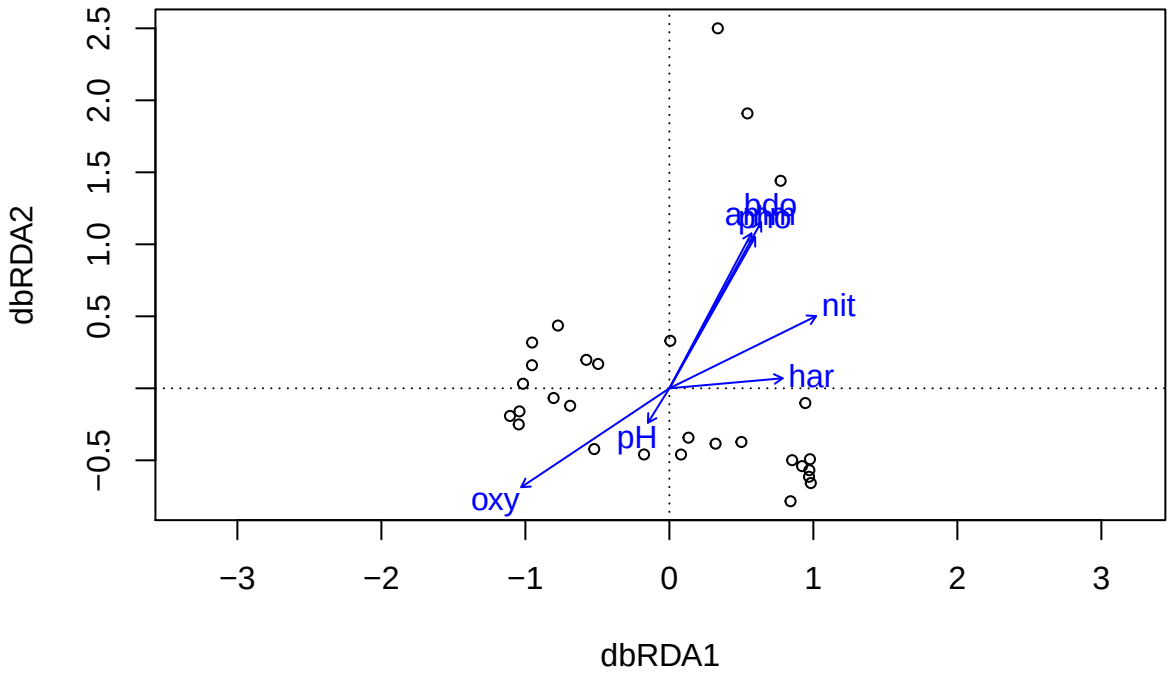
> ***Answer 2***: The mantel test shows that that the bray curtis distance between sites does increase with increased environmental difference, with a relatively strong positive correlation of 0.6 and a significance of 0.001. This means that fish diveristy is positively correlated with some environmental variables. From the previous test, it seems like in higher quality habitats there are only a few species present and at lower quality sites there are more species, so the mantel test suggest that increasing fish diversity could mean a less high quality habitat.

ii. Constrained Ordination

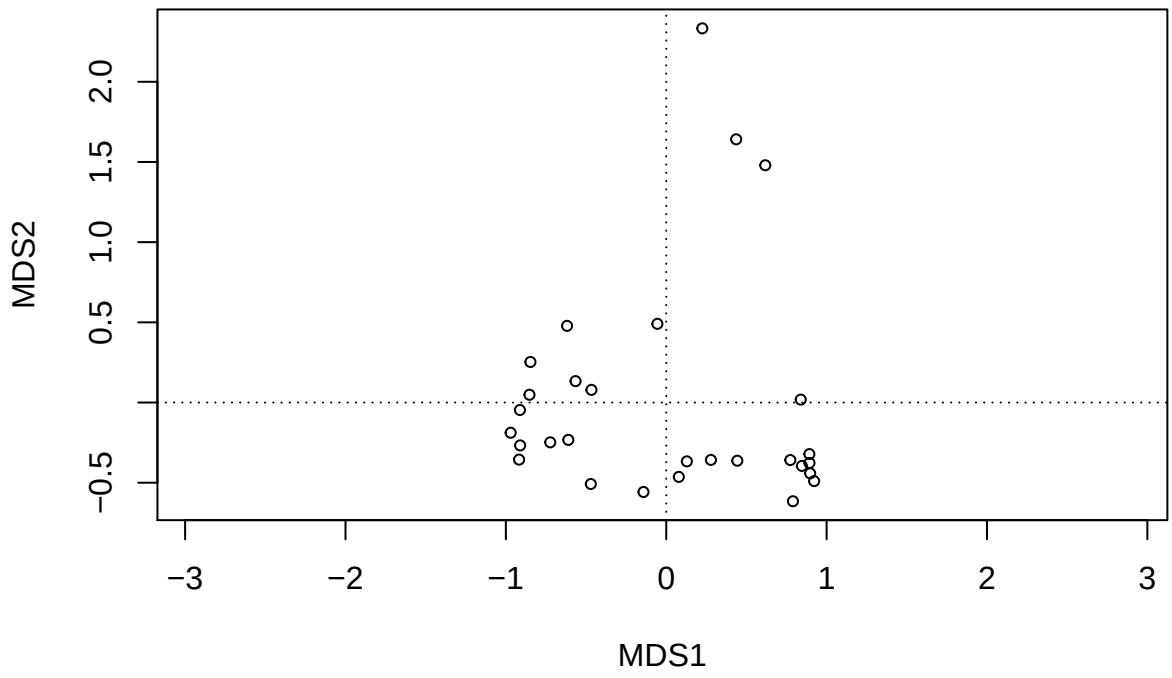
In the R code chunk below, do the following:

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])

fish.db<-vegdist(fish, method = "bray", diag = TRUE)
doubs.dbrda<-dbrda(fish.db ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda)
```




```
doubs.dbrda.mod0<-dbrda(fish.db~1, as.data.frame(env.chem))
ordiplot(doubs.dbrda.mod0)
```




```
doubs.dbrda.mod1<-dbrda(fish.db~., as.data.frame(env.chem))
doubs.dbrda<-ordiR2step(doubs.dbrda.mod0, doubs.dbrda.mod1, perm.max=200)
```

```
## Step: R2.adj= 0
## Call: fish.db ~ 1
##
##                 R2.adjusted
## <All variables>  0.53032584
## + oxy            0.27727176
## + nit            0.25755208
## + bdo            0.17477787
## + pho            0.14568614
## + har            0.14174915
## + amm            0.14142804
## <none>           0.00000000
## + pH            -0.01827054
##
##       Df    AIC      F Pr(>F)
## + oxy  1 47.939 11.742  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2772718
## Call: fish.db ~ oxy
##
```

```
##                 R2.adjusted
## <All variables>   0.5303258
## + bdo             0.4009000
## + amm             0.3474192
## + pho             0.3452702
## + har             0.3331357
## + nit             0.3316120
## <none>            0.2772718
## + pH              0.2586983
##
##       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
```

```
doubs.dbrda$call
```

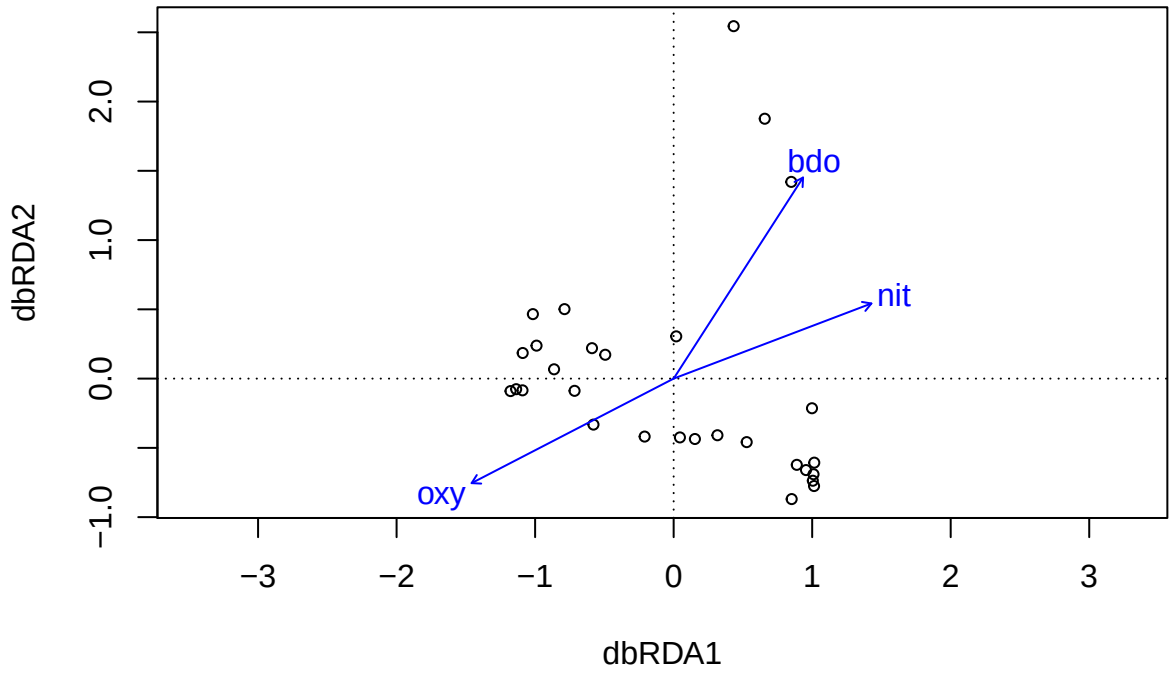
```
## dbrda(formula = fish.db ~ oxy + bdo + nit, data = as.data.frame(env.chem))
```

```
doubs.dbrda$anova
```

```
##                  R2.adj Df    AIC       F Pr(>F)
## + oxy           0.27727  1 47.939 11.7421  0.002 **
## + bdo           0.40090  1 43.404  6.5716  0.002 **
## + nit           0.49808  1 39.134  6.0340  0.002 **
```

```
## <All variables> 0.53033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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



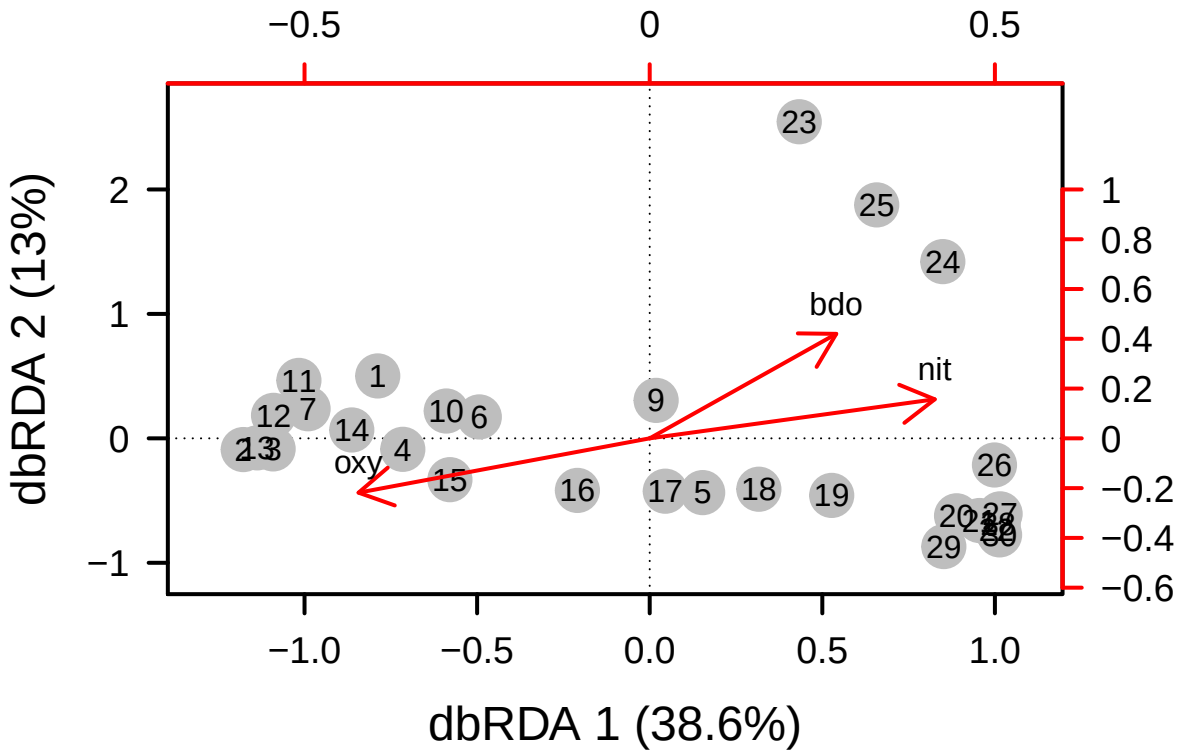

```
permutest(doubs.dbrda, permutations = 999)
```

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = fish.db ~ 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)*100
dbrda.explainvar2<-round(doubs.dbrda$CCA$eig[2]/
                         sum(c(doubs.dbrda$CCA$eig, doubs.dbrda$CA$eig)),3)*100

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 nitrate and dissolved oxygen were the two environmental variables that seemed to be contributing most to driving variation in fish community structure.

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)
doubs.pcnmw<-pcnm(dist(doubs$xy[-8,]), w=rs, dist.ret = T)
doubs.pcnmw$values>0
```

```
##  [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
```

```
doubs.space<-as.data.frame(scores(doubs.pcnmw))
doubs.pcnm.mod0<-dbrda(fish.db~1, doubs.space)
doubs.pcnm.mod1<-dbrda(fish.db~., doubs.space)
step.pcnm<-ordiR2step(doubs.pcnm.mod0, doubs.pcnm.mod1, perm.max=200)
```

```
## Step: R2.adj= 0
## Call: fish.db ~ 1
##
##                  R2.adjusted
## <All variables>  0.626011301
## + PCNM2          0.235370423
## + PCNM3          0.078394885
## + PCNM13         0.065305668
## + PCNM5          0.046185074
## + PCNM6          0.032809156
## + PCNM16         0.030486700
## + PCNM14         0.029680999
## + PCNM9          0.020357410
## + PCNM15         0.013632610
## + PCNM8          0.009411968
## + PCNM1          0.003986221
## + PCNM17         0.002415012
## + PCNM10         0.001326442
## <none>           0.000000000
## + PCNM7         -0.001861430
## + PCNM11        -0.006841522
## + PCNM4         -0.007089863
## + PCNM12        -0.014396973
##
##         Df    AIC     F Pr(>F)
## + PCNM2  1 49.574 9.619  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.2353704
## Call: fish.db ~ PCNM2
##
##                  R2.adjusted
## <All variables>   0.6260113
```

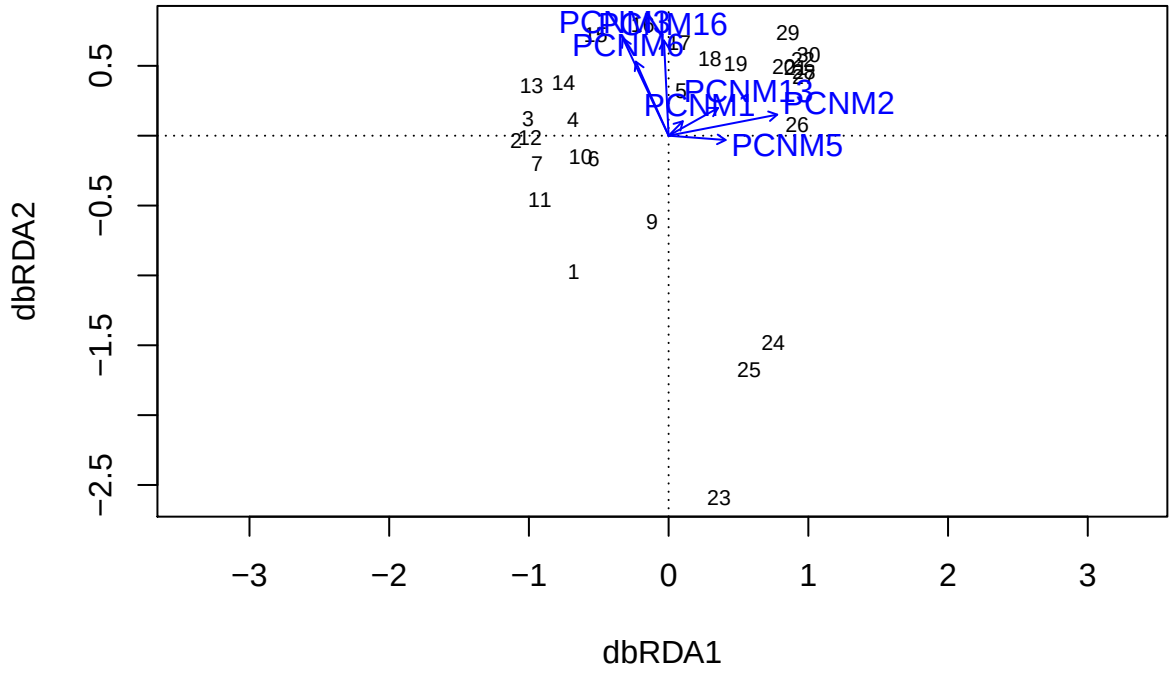
```
## + PCNM3           0.3429270
## + PCNM5           0.3057368
## + PCNM1           0.2885396
## + PCNM16          0.2786746
## + PCNM14          0.2744520
## + PCNM15          0.2692809
## + PCNM6           0.2659866
## + PCNM13          0.2636194
## + PCNM9           0.2517847
## + PCNM8           0.2496240
## + PCNM10          0.2434688
## + PCNM7           0.2431476
## + PCNM17          0.2404343
## + PCNM11          0.2366833
## <none>            0.2353704
## + PCNM12          0.2288789
## + PCNM4           0.2189522
##
##         Df    AIC      F Pr(>F)
## + PCNM3  1 46.083 5.4196  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.342927
## Call: fish.db ~ PCNM2 + PCNM3
##
##                 R2.adjusted
## <All variables>   0.6260113
## + 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  0.014 *
## ---
## 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  0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.4721469
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5 + PCNM1
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM13          0.5212427
## + PCNM16          0.5208668
## + PCNM15          0.5161770
## + PCNM14          0.5147355
## + PCNM6           0.4999020
## + PCNM7           0.4936559
## + PCNM8           0.4904113
## + PCNM17          0.4856884
## + PCNM10          0.4835952
## + PCNM11          0.4760087
## + PCNM9           0.4751424
## + PCNM12          0.4747221
## <none>            0.4721469
## + PCNM4           0.4651218
##
##          Df    AIC      F Pr(>F)
## + PCNM13  1 39.346 3.4612   0.02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.5212427
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM16          0.5767968
## + PCNM15          0.5715331
```

```
## + PCNM14          0.5698343
## + PCNM6           0.5475140
## + PCNM7           0.5392074
## + PCNM8           0.5379134
## + PCNM11          0.5281106
## + PCNM9           0.5267003
## + PCNM10          0.5265029
## + PCNM12          0.5255581
## <none>            0.5212427
## + PCNM17          0.5171800
## + PCNM4           0.5152311
##
##          Df   AIC      F Pr(>F)
## + PCNM16  1 36.48 4.0192  0.018 *
## ---
## 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  0.044 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.6043089
## Call: fish.db ~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13 + PCNM16 + PCNM6
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM8           0.6248697
## + PCNM12          0.6208788
## + PCNM10          0.6170988
## + PCNM7           0.6142419
## + PCNM15          0.6140369
## + PCNM9           0.6107110
## <none>            0.6043089
## + PCNM17          0.6037430
## + PCNM11          0.5978305
```

```
## + PCNM4           0.5963667
## + PCNM14          0.5932113
##
##         Df    AIC     F Pr(>F)
## + PCNM8  1 34.219 2.151  0.082 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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




```
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  0.014 *
## + PCNM1         0.47215  1 41.411 4.0570  0.012 *
## + PCNM13        0.52124  1 39.346 3.4612  0.020 *
## + PCNM16        0.57680  1 36.480 4.0192  0.018 *
## + PCNM6         0.60431  1 35.182 2.5296  0.044 *
## <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]

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

```
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
```

```
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
```

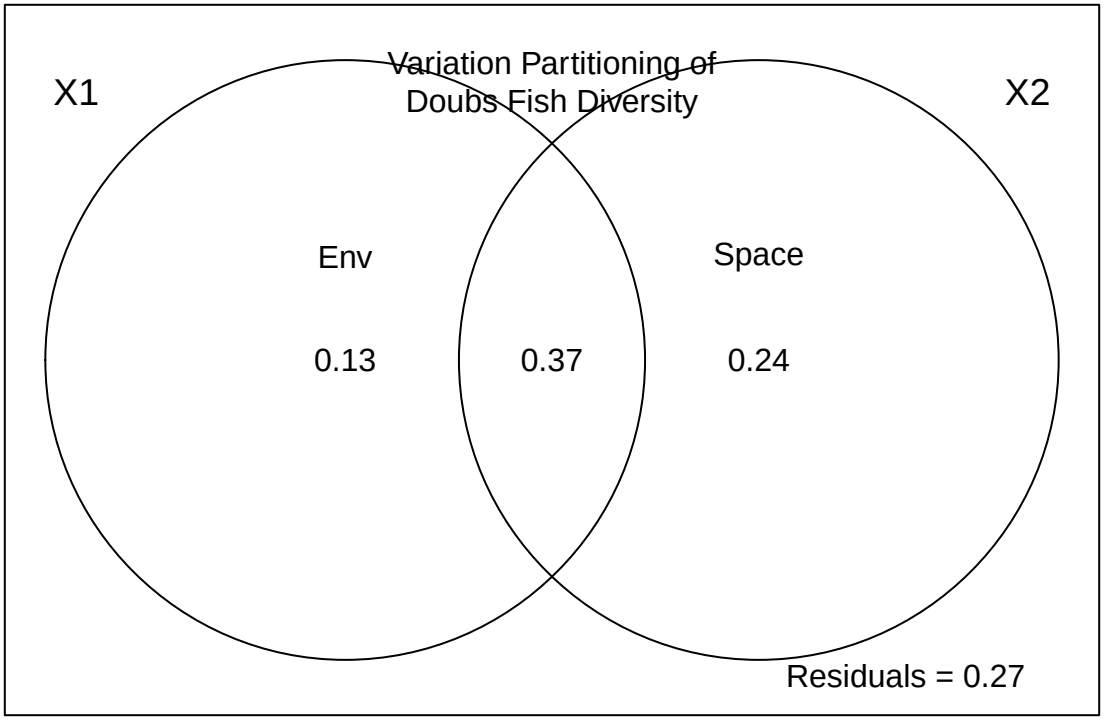
```
permutest(doubs.total.space, permutations=999)
```

```
##
## 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
```

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

```
##
## Partition of squared Bray distance in dbRDA
##
## Call: varpart(Y = fish.db, X = env.mod, space.mod)
##
## Explanatory tables:
## X1:  env.mod
## X2:  space.mod
##
## No. of explanatory tables: 2
## Total variation (SS): 6.7621
## No. of observations: 29
##
## Partition table:
##                       Df R.squared Adj.R.squared Testable
## [a+b] = X1             3   0.55186       0.49808     TRUE
## [b+c] = X2             7   0.70323       0.60431     TRUE
## [a+b+c] = X1+X2       10   0.82917       0.73426     TRUE
## Individual fractions
## [a] = X1|X2            3                 0.12995     TRUE
## [b]                    0                 0.36813    FALSE
## [c] = X2|X1            7                 0.23618     TRUE
## [d] = Residuals                          0.26574    FALSE
## ---
## Use function '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)
```



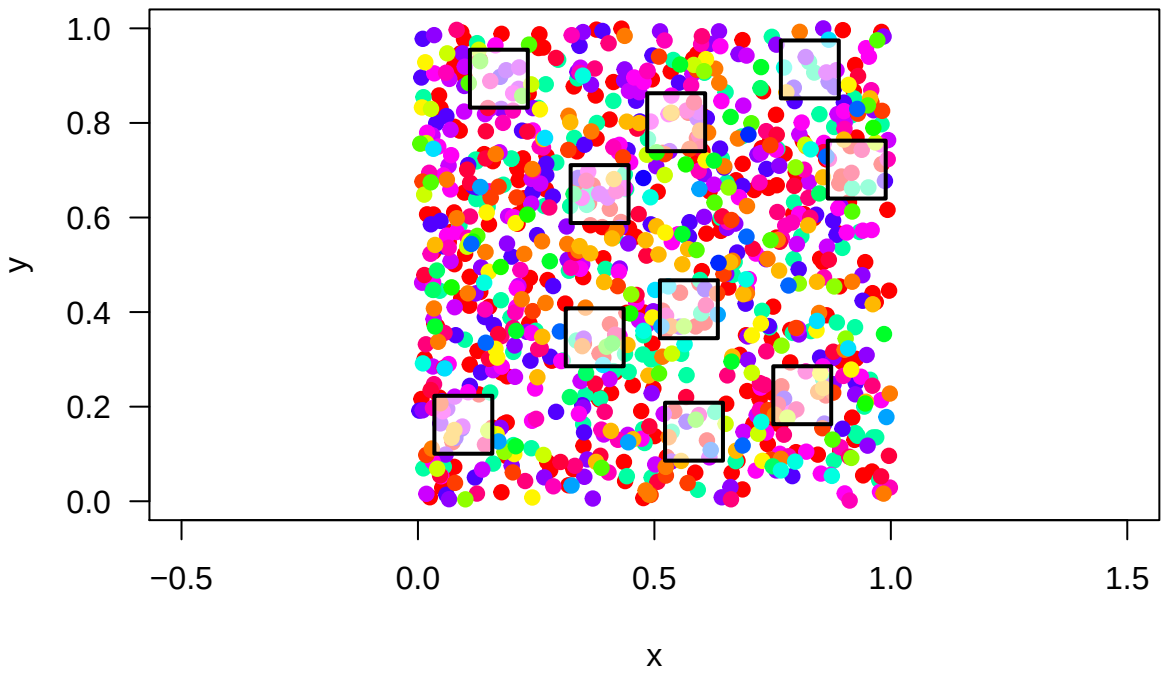
Question 4: Interpret the variation partitioning results.

> ***Answer 4***: The variation partitioning was testing whether fish community changes were driven by environmental factors or spatial factors or spatially structured environmental variation. This analysis shows that 0.13 and of the variation was driven by environmental factors alone and 0.24 of the variation was driven by spatial location. Spatially structured environmental variation explained 0.37 of the variation, so both spacial and environmental components are important for driving fish community differences.

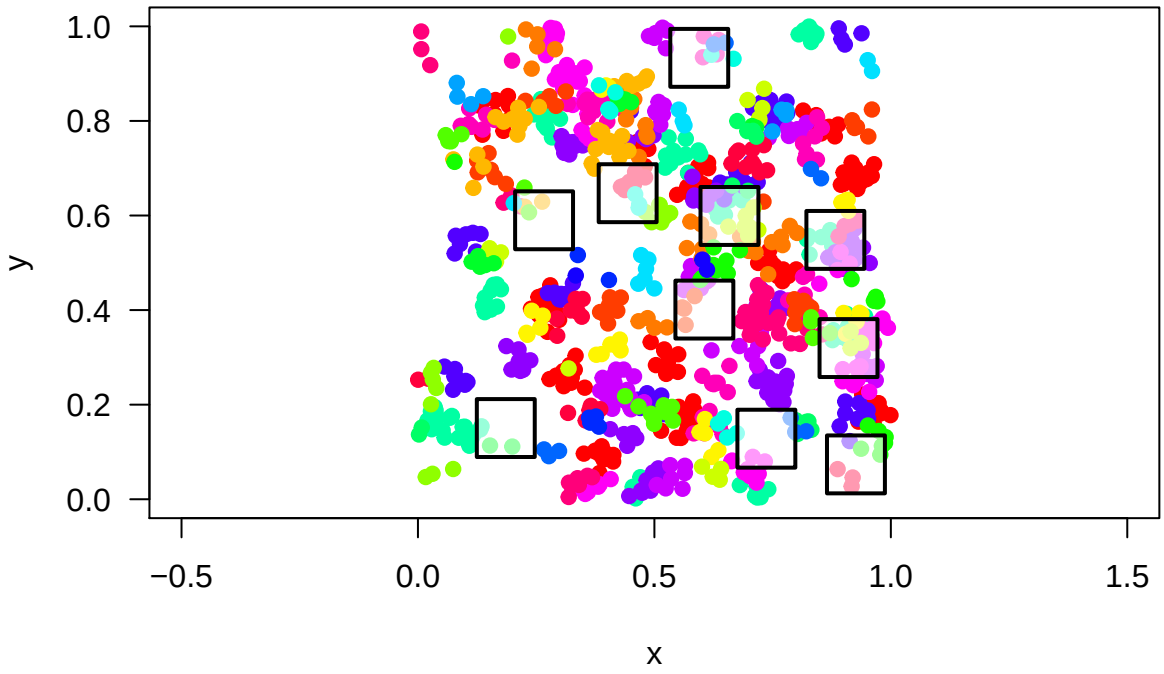
SYNTHESIS

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:

```
com_random <- sim_poisson_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))
#plot(com_random)
com_random_quads <- sample_quadrats(com_random, n_quadrats = 10, quadrat_area = 0.015,
             method = "random", avoid_overlap = T)
```




```
com_patchy <- sim_thomas_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))
#plot(com_patchy)
com_patchy_quads <- sample_quadrats(com_patchy, n_quadrats = 10, quadrat_area = 0.015,
             method = "random", avoid_overlap = T)
```



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

```
random_sites<-(com_random_quads$spec_dat)
rownames(random_sites)<-c("R1","R2","R3","R4","R5","R6","R7","R8","R9","R10")
patchy_sites<-(com_patchy_quads$spec_dat)
rownames(patchy_sites)<-c("P1","P2","P3","P4","P5","P6","P7","P8","P9","P10")
all_sites<-rbind(random_sites,patchy_sites)
site_type<-c("R","R","R","R","R","R","R","R","R","R","P","P","P","P","P","P","P","P","P","P")
adonis(all_sites~site_type, method="bray", permutations = 999)
```

```
##
## Call:
## adonis(formula = all_sites ~ site_type, permutations = 999, method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##            Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## site_type   1    0.9811 0.98105  3.3168 0.15559  0.001 ***
## Residuals 18    5.3241 0.29579         0.84441
## Total     19    6.3052                 1.00000
## ---
```

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

The spatial distribution of species does affects the species composition, however with a correlation of only 0.14, the amount of variation left unexplained is high.

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.

The ordination shows that N and C are pretty correlated and might be driving some of the variation and P might as well, but it's kind of hard from this visualization since the amount of variation isn't quanitfied. I used a mantel test to test for a correlation between plant and fungi communities, and the result was that the plant dissimilarity matrix has a 0.3412 correlation with the fungi dissimilarity matrix. I tested whether there might be indicator species for the plant and fungi communities for both the ridge and snowbed and the result was that there was one species of plant for both the ridge and snowbed that were strongly (>.90) associated with that site type.

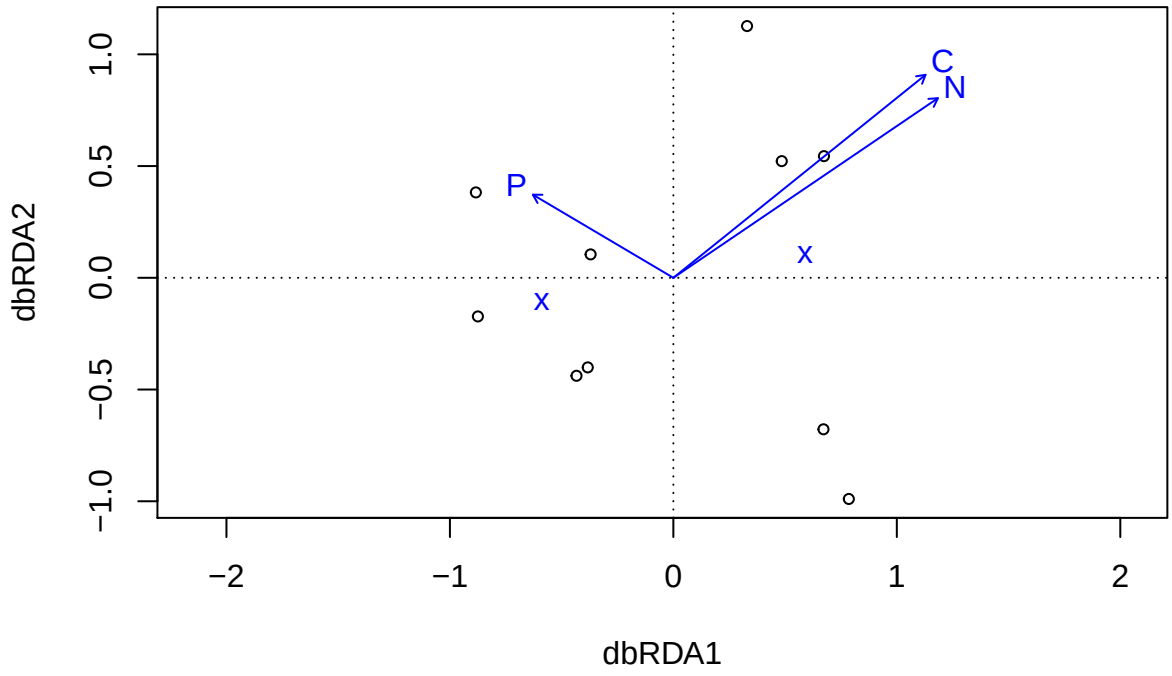
```
#import data and format
veg<-read.csv("~/quant_bio/GitHub/DiversityProject/alpine_ridge_data/veg_.csv")
row.names(veg)<-veg[,1]
veg<-t(veg[,-c(1,48,49)])
otu_sxs<-read.csv("~/quant_bio/GitHub/DiversityProject/alpine_ridge_data/OTU_table.csv")
otu_sxs<-t(otu_sxs)
env_var<-read.delim("~/quant_bio/GitHub/DiversityProject/alpine_ridge_data/variables.txt")
env_var$site<- c(rep("S1", 5), rep("S2", 5), rep("S3", 5), rep("S4", 5), rep("S5", 5), rep("R1", 5), rep

#take mean of all samples of environmental data at each site, so there is only one value per site
env<-tapply(env_var$P, env_var$site, mean)
env<-as.data.frame(env)
env$P<-tapply(env_var$P, env_var$site, mean)
env$N<-tapply(env_var$N, env_var$site, mean)
env$C<-tapply(env_var$C, env_var$site, mean)
env$V<-c(rep("R", 5), rep("S", 5))
env<-env[,-1]

#rarefaction of otu samples
site_species.r <- rrarefy(otu_sxs, 1000)
```

```
## Warning in rrarefy(otu_sxs, 1000): some row sums < 'sample' and are not rarefied
```

```
richness <- rowSums((site_species.r > 0) * 1)
minimum.r <- min(rowSums(site_species.r))
rarefy <- rarefy(x = site_species.r, sample = minimum.r, se = TRUE)

#remove samples containing less than 1000 reads (R1.14, R1.55.2, R2.25, S2.78.2)
df.site_species.r <- as.data.frame(site_species.r)
rarefied_site_species <- data.frame()
for (i in 1:nrow(df.site_species.r)){
  if (rowSums(df.site_species.r[i,]) >= 1000){
    rarefied_site_species <- rbind(rarefied_site_species, df.site_species.r[i,])
  }
```

```
}
otu<-as.data.frame(rarefied_site_species)
otu_site<-c(rep("S1", 5), rep("S2", 5), rep("S3", 5), rep("S4", 5), rep("S5", 5), rep("R1", 4), rep("R2

#add up all samples of fungi data at each site, so there is one total value per site
otu_total<-matrix(nrow=10,ncol=806)
for (i in 1:(ncol(otu)-1)){
  otu_total[,i]<-tapply(otu[,i],otu_site,sum)
}
```

```
#ordination plot of plant communities with environmental variables C,N,P
veg.dist<- vegdist(veg, method = "bray")
veg.dbrda.env<-dbrda(veg.dist~., env)
ordiplot(veg.dbrda.env)
```




```
#PERMANOVA test the amount of variation in the plant communities explained by the site type (ridge or s
site_var<-c("R1","R2","R3","R4","R5","S1","S2","S3","4S","S5")
adonis(veg~env$P+env$C+env$N+env$V+site_var,method="bray",permutations=999)
```

```
##
## Call:
## adonis(formula = veg ~ env$P + env$C + env$N + env$V + site_var,      permutations = 999, method = "
##
## Permutation: free
```

```
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## env$P      1   0.18482       0       0 0.12711      1
## env$C      1   0.33514       0       0 0.23049      1
## env$N      1   0.20660       0       0 0.14209      1
## env$V      1   0.21789       0       0 0.14985      1
## site_var   5   0.50959       0       0 0.35046      1
## Residuals  0   0.00000     Inf         0.00000
## Total      9   1.45405                 1.00000
```

```
#mantel test to test the correlation of the fungi and plant bray-curtis distances and the plant with th
otu.dist<- vegdist(otu_total, method="bray")
env.dist<- vegdist(scale(env[,-4]), method="euclid")
veg.dist<- vegdist(veg, method = "bray")

mantel(veg.dist,otu.dist)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = veg.dist, ydis = otu.dist)
##
## Mantel statistic r: 0.3591
##       Significance: 0.018
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.218 0.276 0.330 0.390
## Permutation: free
## Number of permutations: 999
```

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

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = veg.dist, ydis = env.dist)
##
## Mantel statistic r: 0.3351
##       Significance: 0.023
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.209 0.277 0.327 0.372
## Permutation: free
## Number of permutations: 999
```

```r
#look at indicator species for the two site types in plants and fungi
indval_veg<-multipatt(veg, cluster=c(rep("R", 5), rep("S", 5)), func = "r.g", control=how(nperm=999))
summary(indval_veg)
```

```
##
##  Multilevel pattern analysis
##  ---------------------------
##
##  Association function: r.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 49
##  Selected number of species: 4
##  Number of species associated to 1 group: 4
##
##  List of species associated to each combination:
##
##  Group R  #sps.  2
##                      stat p.value
## Cladonia.arbuscula 0.946   0.011 *
## Empetrum.nigrum    0.691   0.038 *
##
##  Group S  #sps.  2
##                   stat p.value
## Salix.herbaceae 0.908   0.011 *
## Polytrichum.sp. 0.770   0.028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
indval_otu<-multipatt(otu_total, cluster=c(rep("R", 5), rep("S", 5)), func = "r.g", control=how(nperm=99
summary(indval_otu)
```

```
##
##  Multilevel pattern analysis
##  ---------------------------
##
##  Association function: r.g
##  Significance level (alpha): 0.05
##
##  Total number of species: 806
##  Selected number of species: 24
##  Number of species associated to 1 group: 24
##
##  List of species associated to each combination:
##
##  Group R  #sps.  8
##      stat p.value
## 578 0.870   0.006 **
## 493 0.820   0.006 **
## 635 0.816   0.044 *
## 718 0.781   0.006 **
## 342 0.745   0.043 *
```

```
## 575 0.692   0.043 *
## 571 0.596   0.050 *
## 717 0.557   0.006 **
##
##  Group S  #sps.  16
##      stat p.value
## 629 0.800   0.006 **
## 500 0.721   0.043 *
## 465 0.696   0.014 *
## 676 0.692   0.043 *
## 743 0.686   0.006 **
## 354 0.667   0.026 *
## 286 0.655   0.043 *
## 198 0.641   0.043 *
## 520 0.636   0.043 *
## 89  0.636   0.014 *
## 475 0.605   0.043 *
## 711 0.595   0.043 *
## 96  0.562   0.044 *
## 202 0.541   0.043 *
## 103 0.524   0.006 **
## 144 0.453   0.043 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```