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

Aishwarya Vaidya; Z620: Quantitative Biodiversity, Indiana University

19 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).

```
getwd()
```

```
## [1] "/cloud/project/QB2025_Vaidya/Week4-Beta"
```

```r
setwd("/cloud/project/QB2025_Vaidya/Week4-Beta")
require(vegan)
```

```
## Loading required package: vegan
```

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

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

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

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)
```

```
## Warning in data(doubs): data set 'doubs' not found
```

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. >A redundancy analysis i.e dbRDA, the best-fitting model included three key environmental factors: oxygen levels (oxy), biochemical oxygen demand (bdo), and nitrate concentration (nit). These variables together could explain about 53% ($R2 = 0.5303$) of the variation in fish communities. Oxygen had the strongest individual effect ($R2 = 0.2773$, $p = 0.002$), followed by biochemical oxygen demand ($R2 = 0.4009$, $p = 0.002$) and nitrate concentration ($R2 = 0.4981$, $p = 0.004$). A permutation test confirmed that the model as a whole was highly significant ($F = 10.262$, $p = 0.001$), meaning it's very unlikely these results happened by chance. When looking at how strongly each factor aligns with the main trends in fish distribution, all three variables were highly significant ($p = 0.001$), with oxygen and biochemical oxygen demand showing the strongest correlations.

```r
library(ade4)
data(doubs)
fish <- doubs$fish
fish <- fish[-8, ] # remove site 8 from data as has no observations
#Create "Factors" vector
quality <- c(rep("HQ", 13), rep("MQ", 5), rep("LQ", 6), rep("MQ", 5))
#Run PERMANOVA with adonis function
adonis2(fish ~ quality, method = "bray", permutations = 999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = fish ~ quality, permutations = 999, method = "bray")
```

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

```
library(indicspecies)

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   0.001 ***
## 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   0.002 **
## Cyca 0.866   0.001 ***
## Acce 0.866   0.001 ***
## Lele 0.863   0.003 **
## Titi 0.853   0.007 **
## Chto 0.829   0.002 **
## Rham 0.829   0.002 **
## Anan 0.829   0.002 **
## Eslu 0.827   0.019 *
## Pefl 0.806   0.012 *
## Blbj 0.791   0.002 **
## Scer 0.766   0.011 *
```

```
## Abbr 0.750   0.007 **
## Icme 0.661   0.027 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
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   0.001 ***
##
##  Group LQ  #sps.  2
##       stat p.value
## Alal 0.693   0.001 ***
## Ruru 0.473   0.031 *
##
##  Group MQ  #sps.  4
##       stat p.value
## Anan 0.571   0.009 **
## Spbi 0.557   0.006 **
## Chto 0.542   0.014 *
## Icme 0.475   0.045 *
##
##  Group LQ+MQ  #sps.  9
##       stat p.value
## Legi 0.658   0.003 **
## Baba 0.645   0.001 ***
## Rham 0.600   0.001 ***
## Acce 0.594   0.007 **
## Cyca 0.586   0.005 **
## Chna 0.571   0.010 **
## Blbj 0.571   0.013 *
## Gogo 0.523   0.012 *
## Abbr 0.499   0.038 *
## ---
## 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***: Despit R square of the PERMANOVA model explains only 45.76 % of the differences between species composition due to differences in habitat the differences are significant enough due to p value less than 0.001 suggesting the fish communities are significantly different as per habitat, based on whether they are upstream or downstream.IndVal suggest species that are strongly associated with a specific habitat differenciations like high , medium or low quality.We see that some fish species are strongly linked to specific habitat qualities—some thrive in high-quality habitats, while others seem totally fine in low or mixed-quality waters. Interestingly, the low + medium quality habitats had the highest number of associated species, suggesting that a lot of fish can tolerate suboptimal conditions. The phi coefficient analysis (r.g) backs this up, confirming which species are tied to which habitats. Now, if we match this up with the visualizations—like the Ward's cluster dendrogram, ordination plots (PCoA), and heatmaps, all tell the same story as far as conclusions are concerned. The clusters group together fish communities based on habitat quality, and the ordination migh show clear separation between habitat types. Thus we're seeing a clear pattern that some fish are picky and only thrive in high-quality environments, while others are adaptable and show up in multiple conditions.

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%
## 0.113 0.152 0.180 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***: Mantel test basically tells us that fish communities are closely tied to stream conditions—when the environment changes, the fish species in the stream change too. The

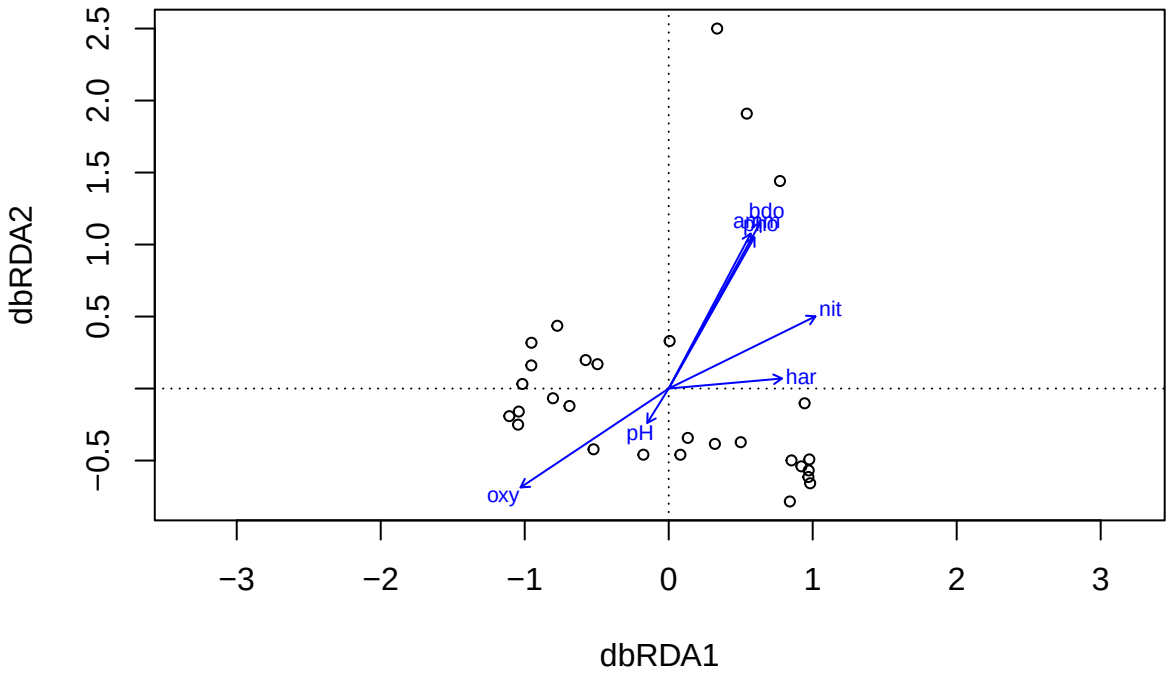
correlation is pretty strong at r = 0.60 with a super low p-value means this isn't just some random coincidence; it's a real pattern. This totally supports our idea that stream quality directly affects fish diversity. If we expected that healthier streams would have more diverse or distinct fish communities, these results prove the same. Besides, it also makes perfect sense as streams with similar water quality, flow, and habitat conditions tend to have similar fish species, because fish respond to their environment just like any other living thing.

ii. Constrained Ordination

In the R code chunk below, do the following:

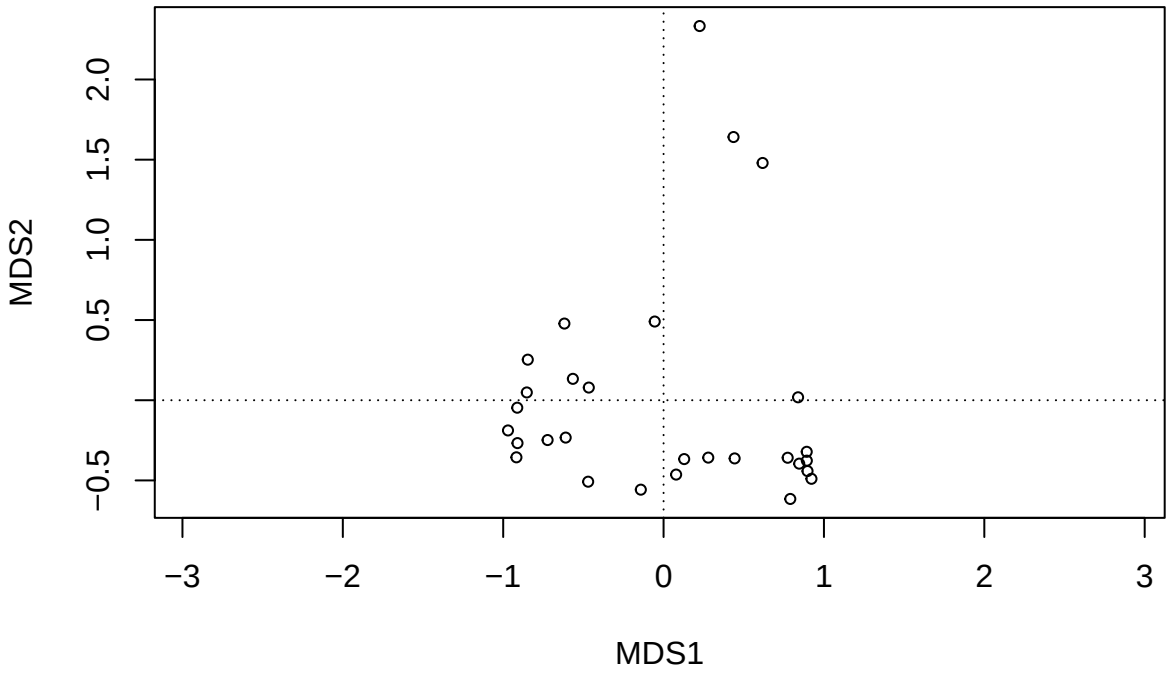
1. create an environmental matrix of the water chemistry data included in the `doubs` dataset using forward and reverse selection of variables,
2. conduct a redundancy analysis on the fish assemblages of the Doubs River,
3. use a permutation test to determine the significance of the constrained analysis,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

```
fish.db <- vegdist(fish, methood = "bray")
#Define environmental matrix
env.chem <- as.matrix(doubs$env[-8, 5:11])

#Perform dbRDA
doubs.dbrda <- dbrda(fish.db ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda)
```




```
#Fist, we will model only the intercept
doubs.dbrda.mod0 <- dbrda(fish.db ~ 1, as.data.frame(env.chem))

#Note there are ni vector here (we didn't constrain anything)
#Therefore, the axes suggest this is a simple MDS ( i.e , PCoA)
```

```
ordiplot(doubs.dbrda.mod0)
```




```
#Note we will model the full model with all explanatory variables
doubs.dbrda.mod1  <- dbrda(fish.db ~ .,as.data.frame(env.chem))

#Now we step through all combinations of explanatory variables in our model
# The function returns the model with the lowest AIC value
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
```

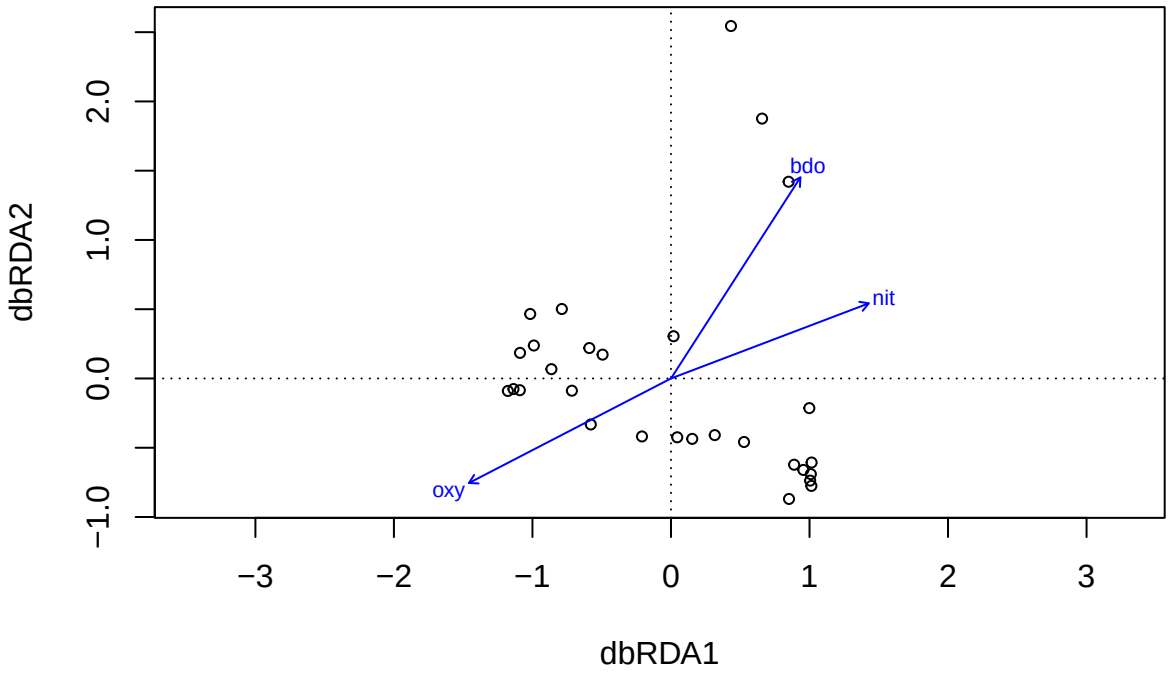
```
#Lets look at the model that was selected
doubs.dbrda$call
```

```
## 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)
```



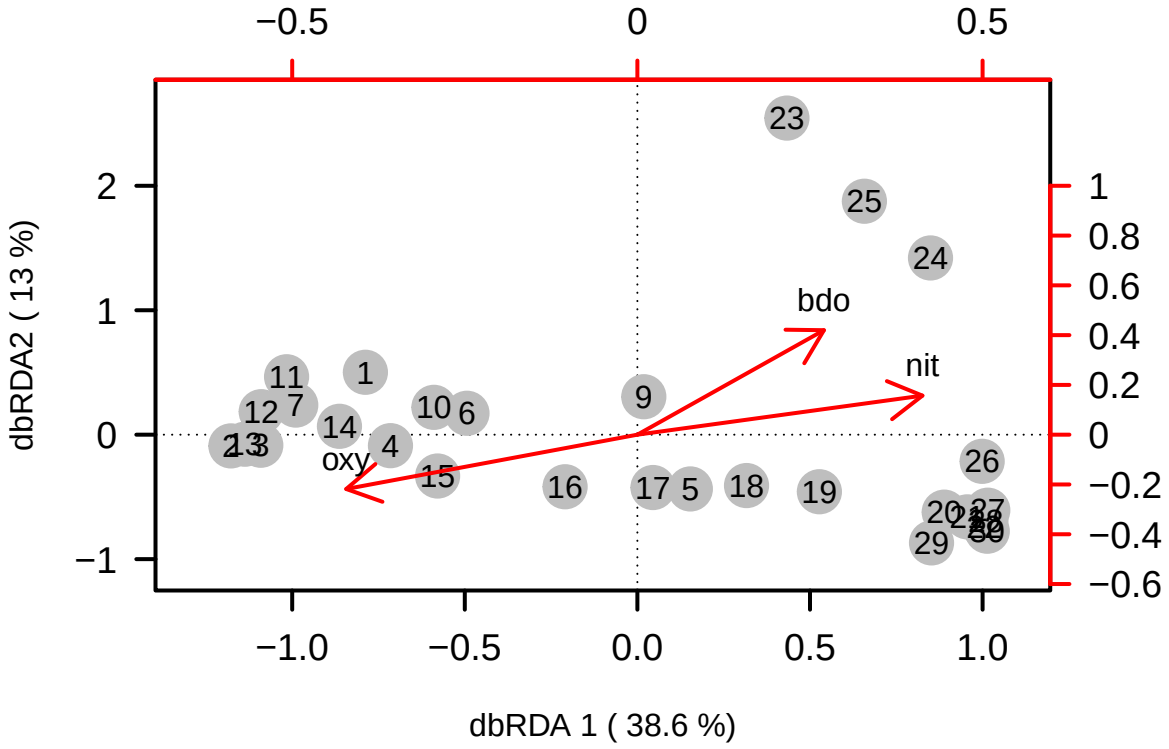

```
#Permutation tests to evaluate significance
permutest(doubs.dbrda, permutations = 999)
```

```
##
## 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
```

```r
#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 Parametes
par(mar = c(5,5,4,4) + 0.1)
#Initiate Plot
plot(scores(doubs.dbrda, display = "wa"),  xlim = c(-1.3,1.1),
     ylim = c(-1.1,2.7), xlab = paste("dbRDA 1 (", dbrda.explainvar1, "%)",
     sep = " "), ylab =paste("dbRDA2 (", dbrda.explainvar2, "%)", sep = " "),
     cex.axis = 1.2, axes = FALSE)

#Add Axes
axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las =1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las =1)
abline(h = 0, v = 0, lty = 3)
box(lwd =2)

#Add Points and Labels
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")))

#Add Environmental Vectors
vectors <- scores(doubs.dbrda, display = "bp")
#row.names(vectors) <- rownames(vectors)
arrows(0,0,vectors[, 1], vectors[ , 2],
       lwd =2, lty =1, length = 0.2, col = "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***: Looks like fish communities are pretty picky about their water quality! The analysis shows that oxygen levels, pollution (BOD), and nitrate concentrations are the biggest factors shaping which fish live where. Streams with similar levels of these things tend to have similar fish populations. BOD (a measure of organic pollution) has the strongest impact, meaning that dirtier water with more organic waste is likely messing with fish diversity. Oxygen and nitrate also play big roles, which makes sense as fish need oxygen to survive, and too many nutrients like nitrate could throw ecosystems off balance. This could suggest better water quality could mean healthier, more diverse fish communities.

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
```

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

#First, we will weight each site by its relative abundance
rs <- rowSums(fish)/sum(fish)
#Next, we will perform PCNM
doubs.pcnmw <- pcnm(dist(doubs$xy[-8,]), w = rs, dist.ret = T)

#Eigenvectors associated with the positive eigenvalues are meaningful
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)
```

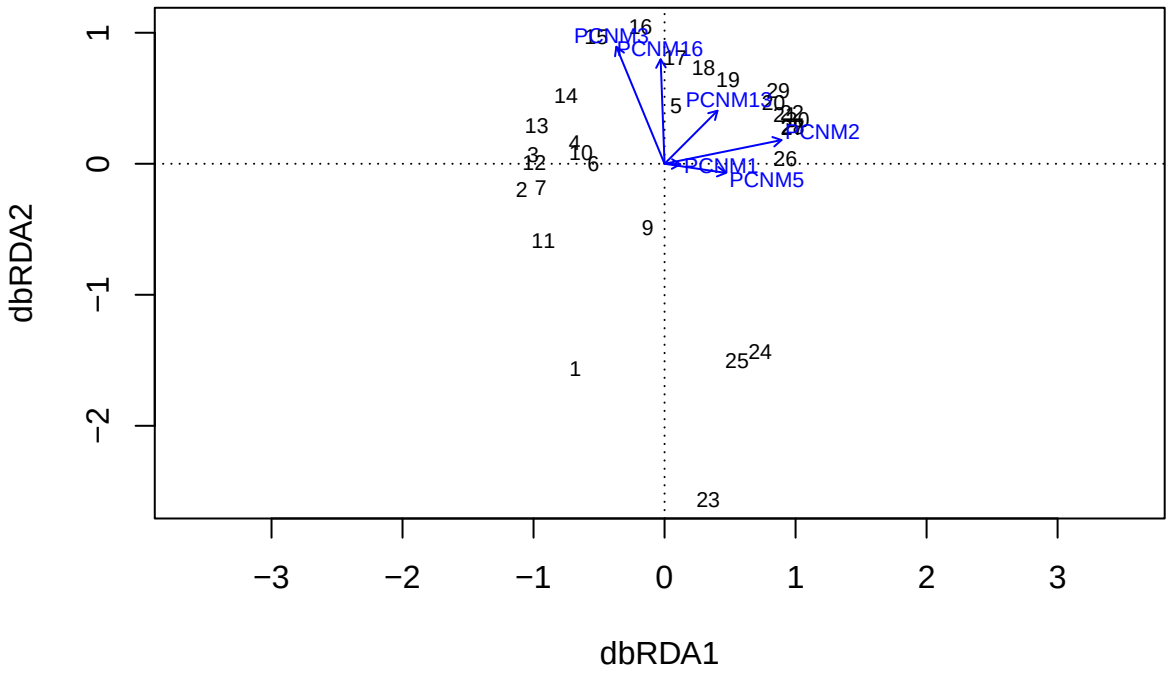
```
## Step: R2.adj= 0
## Call: fish.db ~ 1
##
##                   R2.adjusted
## <All variables>  0.626011301
## + PCNM2          0.235370423
## + PCNM3          0.078394885
## + PCNM13         0.065305668
## + PCNM5          0.046185074
## + PCNM6          0.032809156
## + PCNM16         0.030486700
## + 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  0.006 **
## ---
## 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  0.022 *
## ---
## 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  0.016 *
## ---
## 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  0.058 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Because this is another dbRDA, we could visualise the biplot
# showing how each vector explains variation across sites
plot(step.pcnm)
```




```
#The object 'step.pcnm' now contains the selected model.
step.pcnm$anova
```

```
##                  R2.adj Df    AIC      F Pr(>F)
## + PCNM2         0.23537  1 49.574 9.6190  0.002 **
## + PCNM3         0.34293  1 46.083 5.4196  0.002 **
## + PCNM5         0.40760  1 43.941 3.8385  0.014 *
## + PCNM1         0.47215  1 41.411 4.0570  0.006 **
## + PCNM13        0.52124  1 39.346 3.4612  0.022 *
## + PCNM16        0.57680  1 36.480 4.0192  0.016 *
## <All variables> 0.62601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#We can now construct a spatial model using only the selected PCNM axes.
space.mod <- model.matrix (~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13 + PCNM16 + PCNM6, doubs.space)[, -1]

#First conduct constrained ordinations
doubs.total.env <- dbrda(fish.db ~ env.mod)
doubs.total.space <- dbrda(fish.db ~ space.mod)

#Next construct partial constrained ordinations
doubs.env.cond.space <- dbrda(fish.db ~ env.mod + Condition(space.mod))
doubs.space.cond.env <- dbrda(fish.db ~ space.mod + Condition(env.mod))

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

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
```

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

```
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
```

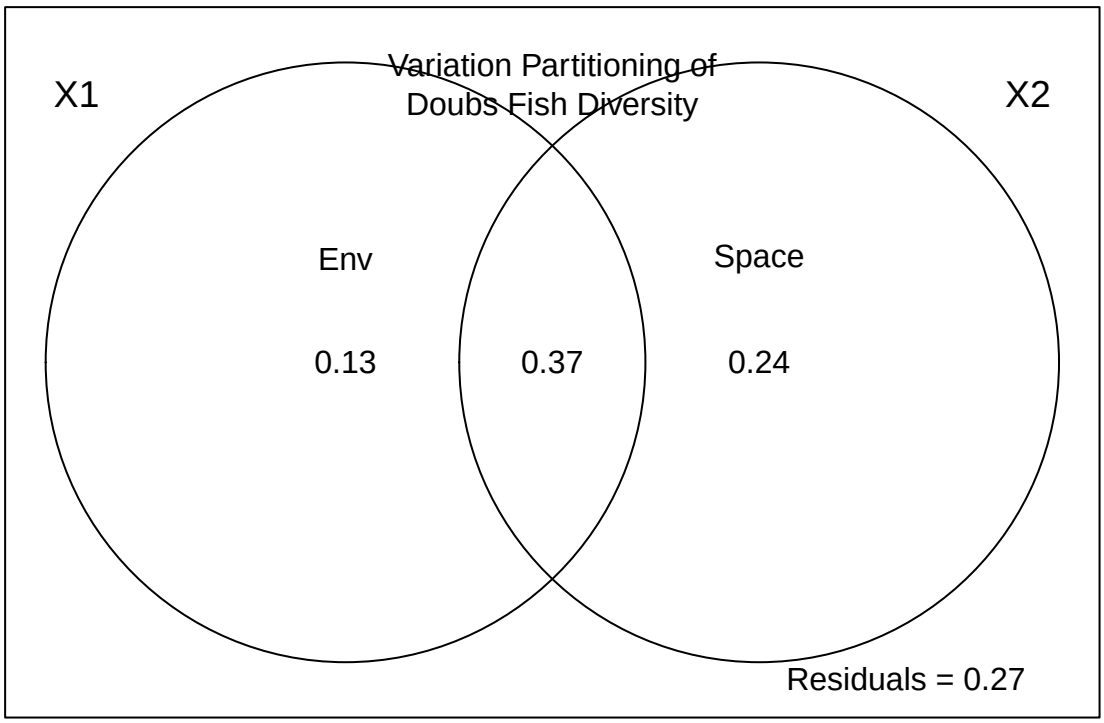
```
permutest(doubs.total.space, permutations = 999)
```

```
##
## 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
```

```r
#Using the built in varpart () function
doubs.varpart <- varpart (fish.db, env.mod, space.mod)
doubs.varpart
```

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

```r
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 partition diagram breaks down whatit is that influences fish diversity in the Doubs River. Oxygen (oxy) seems to have the strongest influence, explaining a significant portion of the variation (R2 adjusted = 0.277, F = 11.74, p = 0.002). On the spatial side, PCNM2, PCNM3, and PCNM5 were being the most important. The final model, explains about 62.6% of the variation. Permutation tests confirm that both environmental and spatial effects are highly significant ($p < 0.001$). Even when accounting for space, environmental effects remain strong (F = 4.423, p = 0.001), and vice versa for spatial effects (F = 4.174, p = 0.001). About 13% of the variation in fish diversity is due to the environment alone, while 24% due to spatial factors. The biggest chunk, 37%, is influenced by both, meaning that where the fish are found is tied to both their surroundings and how they move. However, 27% of the variation remains unexplained(from the residuals at the bottom), so other factors or random chance might also be at play.

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. >Habitat explains 5.9% of the variation while land type explains 6.4% of the variation in microbacterial community based on the adjusted R2 values but the p-value being more than 0.05 suggests that this influence is not certain and might have occured by chance.dbRDA analysis suggests that environmental factors like Mean Annual Precipitation (MAP), Mean Annual Temperature (MAT), and average temperature variation significantly influence species composition.

```
load("/cloud/project/QB2025_Vaidya/longdataBac_objects2_datadryad.rda")
Bacteria <- longdataBac_datadryad #rename
rm(longdataBac_datadryad)
load("/cloud/project/QB2025_Vaidya/Bac_wide_plot_final2_datadryad (1).rda")
Bac_env <- Bac_wide_plot #rename
rm(Bac_wide_plot)
bac <- read.table("/cloud/project/QB2025_Vaidya/bacteria_div.txt",
                        header = TRUE, sep = "\t", stringsAsFactors = FALSE, row.names = 1)
```

```r
## Data cleaning ----
#Matrix based on diff habitat type
#bac_by_site <- with(Bacteria, tapply(Counts, list(PlotID, Sender), sum, default = 0))
#write.table(bac_by_site, file = "bacteria_div.txt", sep = "\t", row.names = TRUE, col.names = NA, quot

#drop some rows
bac.reduced <- bac[!grepl("_3|_2|_1", rownames(bac)), ]
env.reduced <- Bac_env[!grepl("_3|_2|_1", Bac_env$PlotID), ]
rownames(env.reduced) <- env.reduced$PlotID
env.reduced <- env.reduced[, -c(1, 2, 3, 9)]
#write.table(env.reduced, file = "bac_env.txt", sep = "\t", row.names = TRUE, col.names = NA, quote = F
Bacteria.reduced <- Bacteria[!grepl("_3|_2|_1", Bacteria$PlotID), ]
xy <- aggregate(cbind(POINT_X, POINT_Y) ~ PlotID, data = Bacteria.reduced, FUN = mean) #make the spatia

## Distance matrix
# Bray-Curtis:
bac.bc <- vegdist(bac.reduced, method = "bray", upper = TRUE, diag = TRUE)
# Jaccard
bac.jac <- vegdist(bac.reduced, method = "jaccard", upper = TRUE, diag = TRUE)

## Try hypothesis testing ----

### PERMANOVA ----
# None of these are significant, no matter using jaccard or bray
land_type <- env.reduced$Landscape
habitat <- env.reduced$Habitat
adonis2(bac.reduced ~ land_type, method = "jaccard", permutation = 999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = bac.reduced ~ land_type, permutations = 999, method = "jaccard")
##          Df SumOfSqs      R2      F Pr(>F)
## Model     2   0.7378 0.06402 1.2995  0.125
## Residual 38  10.7868 0.93598
## Total    40  11.5246 1.00000
```

```r
adonis2(bac.reduced ~ land_type, method = "bray", permutation = 999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = bac.reduced ~ land_type, permutations = 999, method = "bray")
##          Df SumOfSqs      R2      F Pr(>F)
## Model     2   0.6006 0.07295 1.4952  0.094 .
## Residual 38   7.6313 0.92705
## Total    40   8.2319 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
adonis2(bac.reduced ~ habitat, method = "jaccard", permutation = 999)
```

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

```
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = bac.reduced ~ habitat, permutations = 999, method = "jaccard")
##          Df SumOfSqs      R2      F Pr(>F)
## Model     3   0.6876 0.05967 0.7826  0.875
## Residual 37  10.8369 0.94033
## Total    40  11.5246 1.00000
```

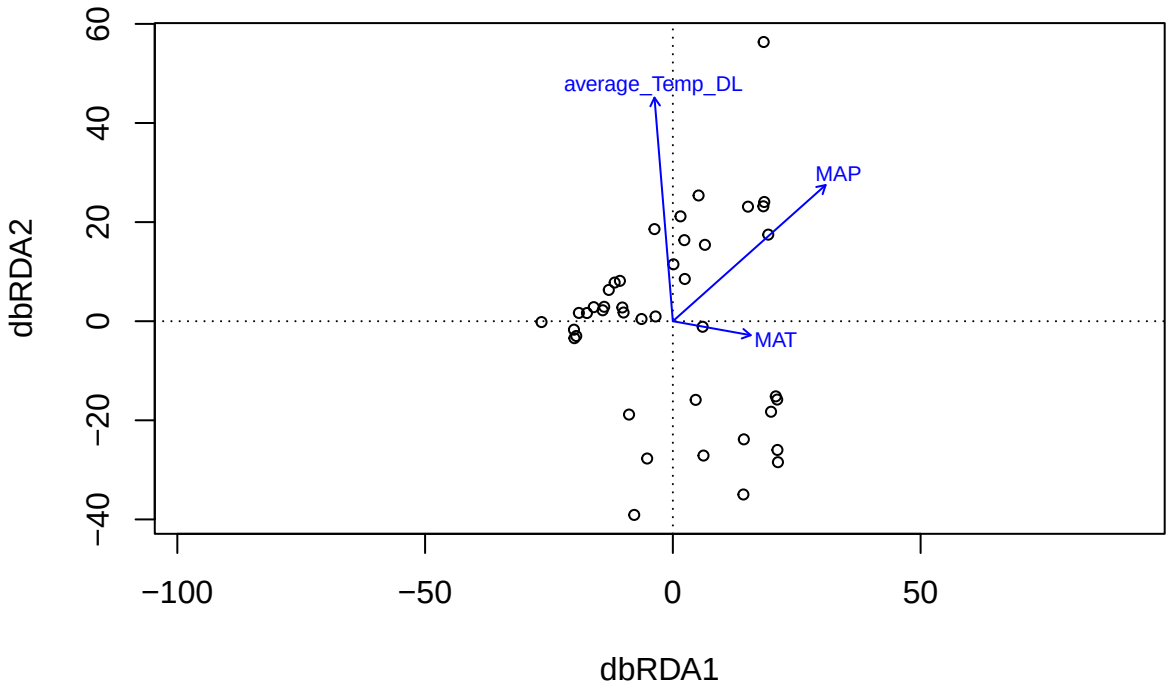
```
adonis2(bac.reduced ~ habitat, method = "bray", permutation = 999)
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = bac.reduced ~ habitat, permutations = 999, method = "bray")
##          Df SumOfSqs      R2      F Pr(>F)
## Model     3   0.4394 0.05338 0.6955  0.875
## Residual 37   7.7925 0.94662
## Total    40   8.2319 1.00000
```

```
### Mantel test ----
env.bc <- vegdist(scale(env.reduced[3:5]), method = "euclid") # env matrix
mantel(bac.bc, env.bc)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = bac.bc, ydis = env.bc)
##
## Mantel statistic r: 0.09388
##       Significance: 0.15
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.117 0.162 0.197 0.238
## Permutation: free
## Number of permutations: 999
```

```
### Constrained Ordination ----
env <- as.matrix(env.reduced[3:5]) #Continous env conditions
env_clean <- na.omit(env)
bac_clean <- bac[rownames(env_clean), , drop = FALSE] # Ensure matching rows
env[is.na(env)] <- mean(env, na.rm = TRUE)  # Replace with column mean
bac <- bac[rownames(env), , drop = FALSE]
bac.dbrda <- dbrda(bac ~ ., as.data.frame(env_clean)) # Using abundance-based distance
bac.dbrda_j <- dbrda(bac.jac ~ ., as.data.frame(env)) # using incidence based distance
ordiplot(bac.dbrda)
```




```
ordiplot(bac.dbrda_j) #This does not have ANOVA result
```