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

Student Name; Z620: Quantitative Biodiversity, Indiana University

29 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] "/home/patgwall/Classwork/Current/QB/2.Worksheets/8.BetaDiversity"
```

```
setwd('~/Classwork/Current/QB/2.Worksheets/8.BetaDiversity/')
require(vegan)
```

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

```
## 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
require(ade4)
```

```
## Loading required package: ade4
```

```
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.

```
require(tidyr)
```

```
## Loading required package: tidyr
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
fish <- doubs[['fish']][-8,]
quality_vec <- c(rep('high', 13), rep('med', 5), rep('low', 6), rep('med', 5))
print(length(quality_vec) == nrow(fish))
```

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

```r
res <- adonis(fish ~ quality_vec)
print(res)
```

```
##
## Call:
## adonis(formula = fish ~ quality_vec)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##             Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## quality_vec  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 <- indicspecies::multipatt(fish, cluster = quality_vec, func = "IndVal.g", control = how(nperm =
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 med  #sps.  1
##       stat p.value
## Teso 0.686    0.02 *
##
##  Group high+med  #sps.  2
##       stat p.value
## Satr 0.860   0.004 **
## Phph 0.859   0.011 *
##
##  Group low+med  #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   0.002 **
## Acce 0.866   0.001 ***
## Lele 0.863   0.003 **
## Titi 0.853   0.003 **
## Chto 0.829   0.003 **
## Rham 0.829   0.002 **
## Anan 0.829   0.002 **
## Eslu 0.827   0.032 *
## Pefl 0.806   0.016 *
## Blbj 0.791   0.004 **
## Scer 0.766   0.008 **
## Abbr 0.750   0.009 **
## Icme 0.661   0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fish_rel <- vegan::decostand(fish, method = "total")
phi <- indicspecies::multipatt(fish_rel, cluster = quality_vec, 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 high  #sps.  3
##       stat p.value
## Phph 0.802   0.001 ***
## Neba 0.734   0.001 ***
## Satr 0.650   0.001 ***
##
##  Group low  #sps.  2
##       stat p.value
## Alal 0.693   0.001 ***
## Ruru 0.473   0.028 *
##
##  Group med  #sps.  4
##       stat p.value
## Anan 0.571   0.006 **
## Spbi 0.557   0.007 **
## Chto 0.542   0.008 **
## Icme 0.475   0.030 *
##
##  Group low+med  #sps.  9
##       stat p.value
## Legi 0.658   0.003 **
```

```
## Baba 0.645   0.005 **
## Rham 0.600   0.005 **
## Acce 0.594   0.006 **
## Cyca 0.586   0.006 **
## Chna 0.571   0.006 **
## Blbj 0.571   0.008 **
## Gogo 0.523   0.020 *
## Abbr 0.499   0.028 *
## ---
## 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***: These approaches do agree and they agree with the visualizations from the last worksheet. PERMANOVA identifies that the categorical predictors are significant and these roughly correspond to the 4 main groups that were aparent in the dendrogram and ordination. Previously I identified Satr and Alal as potential quality markers, based on their positions in the ordination. Both indicator value methods identify these species as significant predictors or correlates. Interestingly according to the above statistics these may not be the best predictors despoite their extreme values on the ordination.

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_bc <- vegan::vegdist(doubs$fish[-8,], method = "bray")
env_euc <- vegan::vegdist(doubs$env[-8, ], method = "euclid")
mantel(fish_bc, env_euc)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = fish_bc, ydis = env_euc)
##
## Mantel statistic r: 0.4677
##       Significance: 0.001
##
## Upper quantiles of permutations (null model):
##    90%    95%  97.5%    99%
## 0.0908 0.1234 0.1483 0.1866
## 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***: This shows us that fish community composition does correlate with environment type as we thought.

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.

```
require(ggplot2)
```

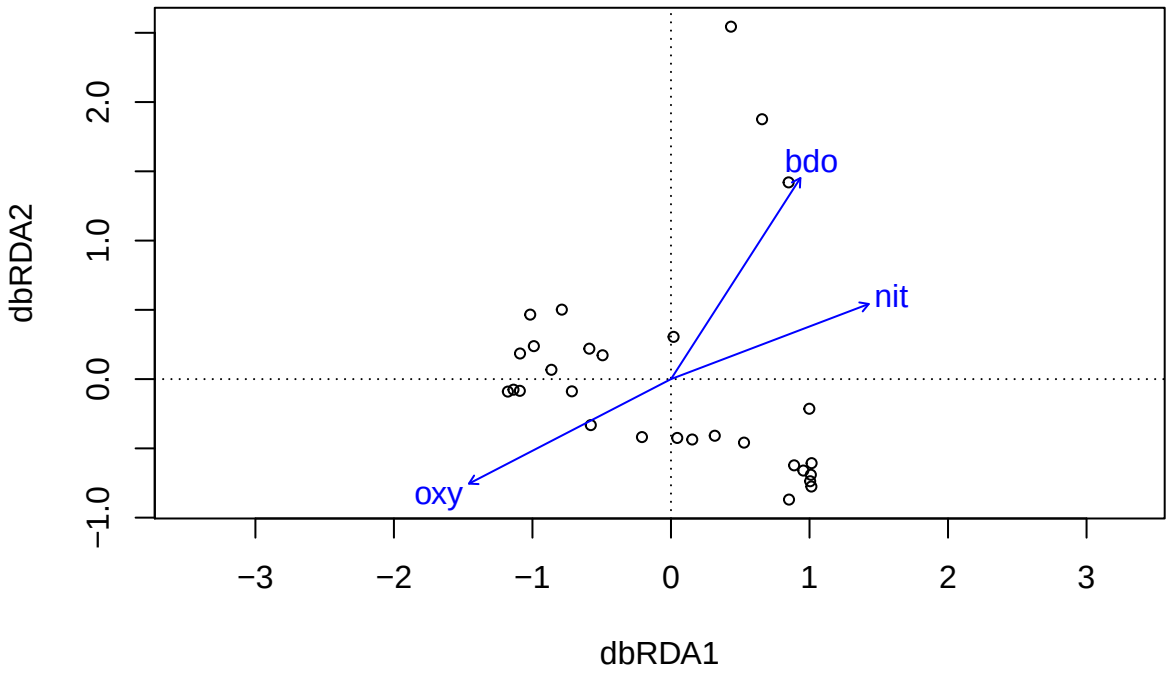
```
## Loading required package: ggplot2
```

```
# define environmental matrix
env_chem <- as.matrix(doubs$env[-8, 5:11])

# model selection
# no vars
doubs_dbrda_none <- vegan::dbrda(fish_bc ~ 1, as.data.frame(env_chem))
# all vars
doubs_dbrda_full <- vegan::dbrda(fish_bc ~ ., as.data.frame(env_chem))
# test fits
doubs_dbrda <- vegan::ordiR2step(doubs_dbrda_none, doubs_dbrda_full, perm_max = 200)
```

```
## Step: R2.adj= 0
## Call: fish_bc ~ 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_bc ~ 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  0.006 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.4009
## Call: fish_bc ~ 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_bc ~ oxy + bdo + nit
##
##                 R2.adjusted
## + amm             0.5415705
## <All variables>   0.5303258
## + pho             0.5277128
## + har             0.5218852
## <none>            0.4980793
## + pH              0.4843267
```

```
## check it out!
# doubs_dbrda$call
# doubs_dbrda$anova
vegan::ordiplot(doubs_dbrda)
```

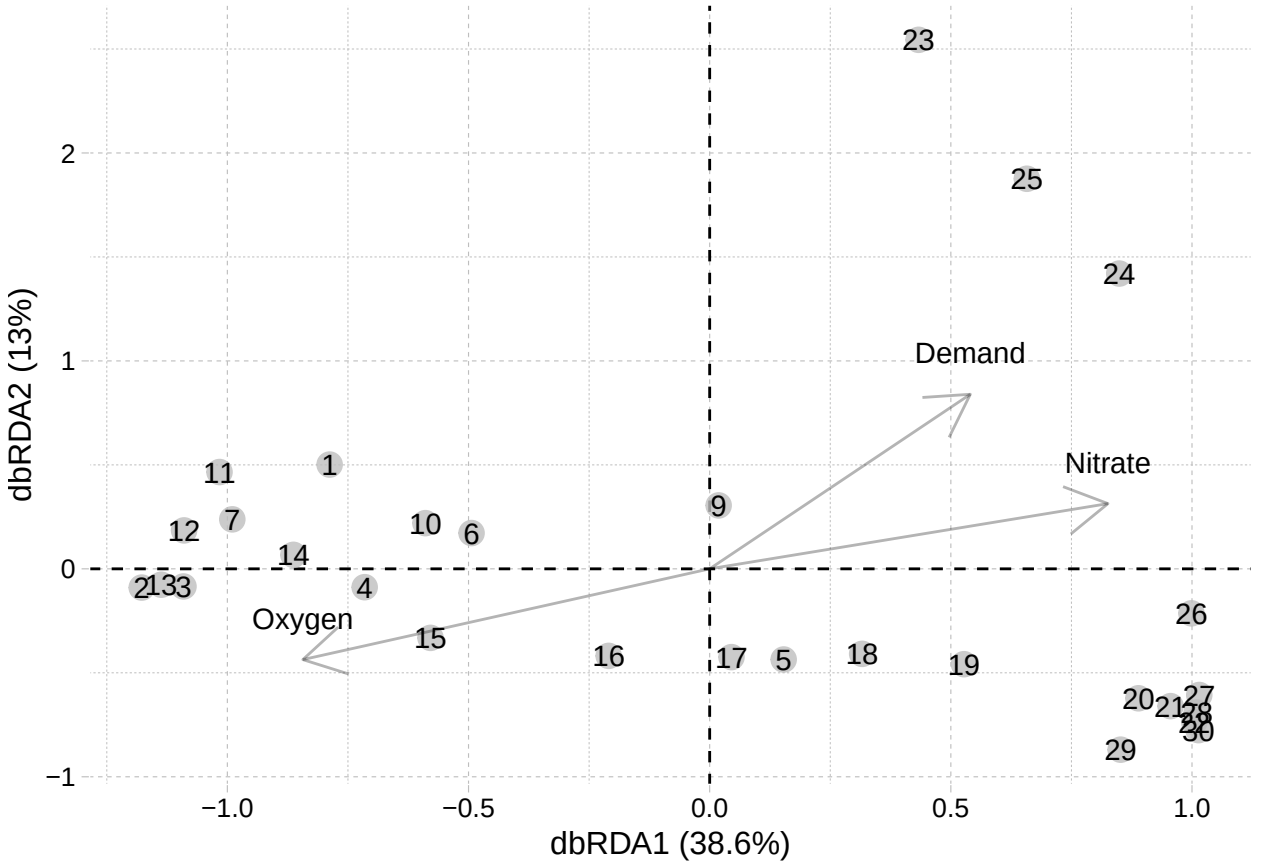



```
# looks like oxygen (oxy), dissolved nitrate (nit), and the oxygen demand (bdo)
# are the useful uncorrelated environmental features

# permutest(doubs_dbrda, permutations = 999)
# envfit(doubs_dbrda, env_chem[,c(4,6,7)], perm = 999)

# Let's make a fancy plot
expvar <- round(sapply(doubs_dbrda$CCA$eig, function(x) x /
                       sum(c(doubs_dbrda$CCA$eig, doubs_dbrda$CA$eig))), 3) * 100

# convert the plots to a dataframe
dbrda_tib <- tibble::as_tibble(scores(doubs_dbrda)$sites) %>%
  mutate(name = c(1:7, 9:(nrow(scores(doubs_dbrda)$sites) + 1)))
# env data
env_vecs <- tibble::as_tibble(scores(doubs_dbrda, display = "bp"))
env_vecs <- mutate(env_vecs, name = c('Oxygen', 'Demand', 'Nitrate'))

print(env_vecs)
```

```
## # A tibble: 3 x 3
##   dbRDA1 dbRDA2 name
##    <dbl>  <dbl> <chr>
## 1 -0.844 -0.436 Oxygen
## 2  0.540  0.840 Demand
## 3  0.826  0.314 Nitrate
```

```
g <- ggplot() +
  geom_hline(yintercept=0, linetype=2) +
  geom_vline(xintercept = 0, linetype=2) +
  geom_point(data=dbrda_tib, aes(x = dbRDA1, y = dbRDA2), size = 4, alpha = 0.8, color = 'grey') +
  geom_text(data=dbrda_tib, aes(x=dbRDA1, y = dbRDA2, label = name))+
  geom_segment(data=env_vecs, aes(x = 0, y = 0, xend = dbRDA1, yend = dbRDA2), arrow = arrow(), alpha =
  geom_text(data=env_vecs, aes(x = dbRDA1, y = dbRDA2 + 0.2, label = name)) +
```

```
  xlab(paste('dbRDA1 (', expvar[1], '%)', sep = '')) +
  ylab(paste('dbRDA2 (', expvar[2], '%)', sep = '')) +
  ggthemes::theme_pander()
g
```



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***: We see dissolved oxygen concentration, dissolved nitrate concentration, and the oxygen demand of river life are the significant relationships.

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.

```
# environmental matrix
env_mod <- model.matrix(~ oxy + bdo + nit, as.data.frame(env_chem))[, -1]

# get space decomposition
```

```r
rs <- rowSums(fish) / sum(fish)
doubs_pcnmw <- vegan::pcnm(dist(doubs$xy[-8,]), w = rs, dist.ret = TRUE)
doubs_space <- as.data.frame(scores(doubs_pcnmw))
doubs_pcnm_none <- dbrda(fish_bc ~ 1, doubs_space)
doubs_pcnm_full <- dbrda(fish_bc ~ ., doubs_space)
step_pcnm <- vegan::ordiR2step(doubs_pcnm_none, doubs_pcnm_full, perm.max = 200)
```

```
## Step: R2.adj= 0
## Call: fish_bc ~ 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_bc ~ PCNM2
##
##                  R2.adjusted
## <All variables>   0.6260113
## + PCNM3           0.3429270
## + PCNM5           0.3057368
## + PCNM1           0.2885396
## + PCNM16          0.2786746
## + PCNM14          0.2744520
## + PCNM15          0.2692809
## + PCNM6           0.2659866
## + PCNM13          0.2636194
## + PCNM9           0.2517847
## + PCNM8           0.2496240
## + PCNM10          0.2434688
## + PCNM7           0.2431476
## + PCNM17          0.2404343
```

```
## + PCNM11          0.2366833
## <none>            0.2353704
## + PCNM12          0.2288789
## + PCNM4           0.2189522
##
##         Df    AIC      F Pr(>F)
## + PCNM3  1 46.083 5.4196  0.004 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.342927
## Call: fish_bc ~ PCNM2 + PCNM3
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM5           0.4076020
## + PCNM1           0.3970300
## + PCNM16          0.3853210
## + PCNM15          0.3828748
## + PCNM14          0.3781827
## + PCNM13          0.3770376
## + PCNM6           0.3595644
## + PCNM8           0.3556885
## + PCNM7           0.3541631
## + PCNM10          0.3526775
## + PCNM17          0.3513683
## + PCNM9           0.3433672
## <none>            0.3429270
## + PCNM11          0.3416399
## + PCNM12          0.3396547
## + PCNM4           0.3311509
##
##         Df    AIC      F Pr(>F)
## + PCNM5  1 43.941 3.8385  0.006 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.407602
## Call: fish_bc ~ PCNM2 + PCNM3 + PCNM5
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM1           0.4721469
## + PCNM16          0.4631976
## + PCNM15          0.4589111
## + PCNM14          0.4535248
## + PCNM13          0.4511582
## + PCNM6           0.4305640
## + PCNM7           0.4261965
## + PCNM8           0.4224505
## + PCNM17          0.4181666
## + PCNM10          0.4154485
## + PCNM11          0.4112178
## + PCNM9           0.4111995
```

```
## + PCNM12          0.4087602
## <none>            0.4076020
## + PCNM4           0.3976526
##
##         Df    AIC     F Pr(>F)
## + PCNM1  1 41.411 4.057  0.004 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.4721469
## Call: fish_bc ~ PCNM2 + PCNM3 + PCNM5 + PCNM1
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM13          0.5212427
## + PCNM16          0.5208668
## + PCNM15          0.5161770
## + PCNM14          0.5147355
## + PCNM6           0.4999020
## + PCNM7           0.4936559
## + PCNM8           0.4904113
## + PCNM17          0.4856884
## + PCNM10          0.4835952
## + PCNM11          0.4760087
## + PCNM9           0.4751424
## + PCNM12          0.4747221
## <none>            0.4721469
## + PCNM4           0.4651218
##
##          Df    AIC      F Pr(>F)
## + PCNM13  1 39.346 3.4612  0.024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.5212427
## Call: fish_bc ~ 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)
```

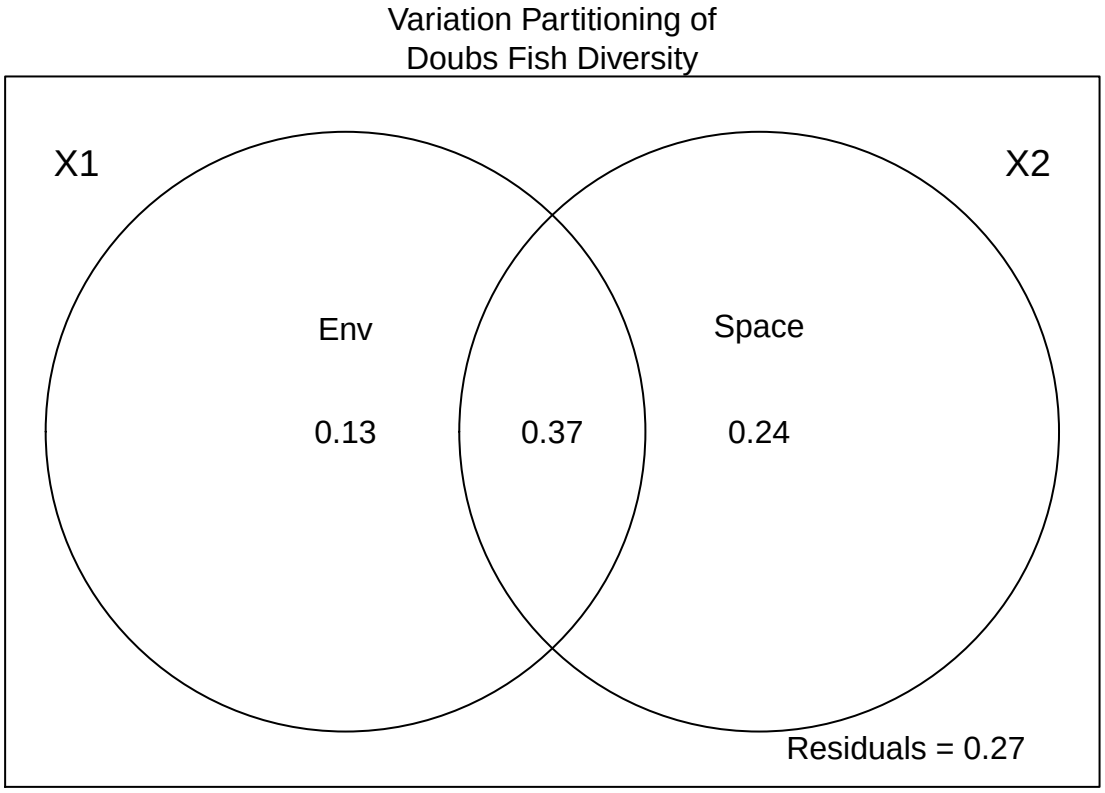
```
## + PCNM16  1 36.48 4.0192  0.022 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.5767968
## Call: fish_bc ~ PCNM2 + PCNM3 + PCNM5 + PCNM1 + PCNM13 + PCNM16
##
##                 R2.adjusted
## <All variables>   0.6260113
## + PCNM6           0.6043089
## + PCNM8           0.5970286
## + PCNM12          0.5946888
## + PCNM7           0.5946475
## + PCNM9           0.5883735
## + PCNM10          0.5851333
## + PCNM15          0.5846468
## <none>            0.5767968
## + PCNM17          0.5748533
## + PCNM4           0.5733749
## + PCNM11          0.5711176
## + PCNM14          0.5652509
##
##         Df    AIC      F Pr(>F)
## + PCNM6  1 35.182 2.5296   0.05 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.6043089
## Call: fish_bc ~ 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  0.102
```

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

# build the full models
doubs_total_env <- vegan::dbrda(fish_bc ~ env_mod)
doubs_total_space <- vegan::dbrda(fish_bc ~ space_mod)
```

```r
doubs_env_cond_space <- vegan::dbrda(fish_bc ~ env_mod + Condition(space_mod))
doubs_space_cond_env <- vegan::dbrda(fish_bc ~ space_mod + Condition(env_mod))

## test for significance
# vegan::permutest(doubs_total_env)
# vegan::permutest(doubs_total_space)
# vegan::permutest(doubs_env_cond_space)
# vegan::permutest(doubs_space_cond_env)

# using the built-in varpart() function
doubs_varpart <- vegan::varpart(fish_bc, env_mod, space_mod)
doubs_varpart
```

```
##
## Partition of squared Bray distance in dbRDA
##
## Call: vegan::varpart(Y = fish_bc, 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
```

```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")
```



Question 4: Interpret the variation partitioning results.

> ***Answer 4***: The environment and spatial considerations separately and together explain most of the variation except for a remaining unexplained fraction in the residuals. Space alone is more informative than the environment alone.

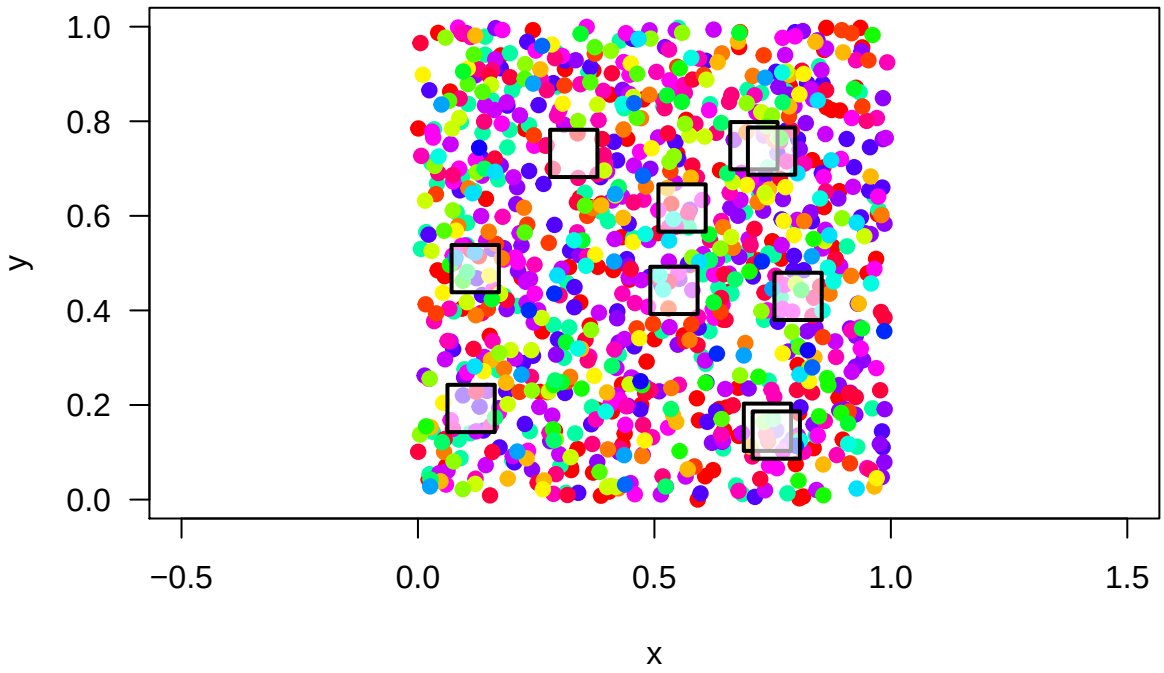
SYNTHESIS

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.

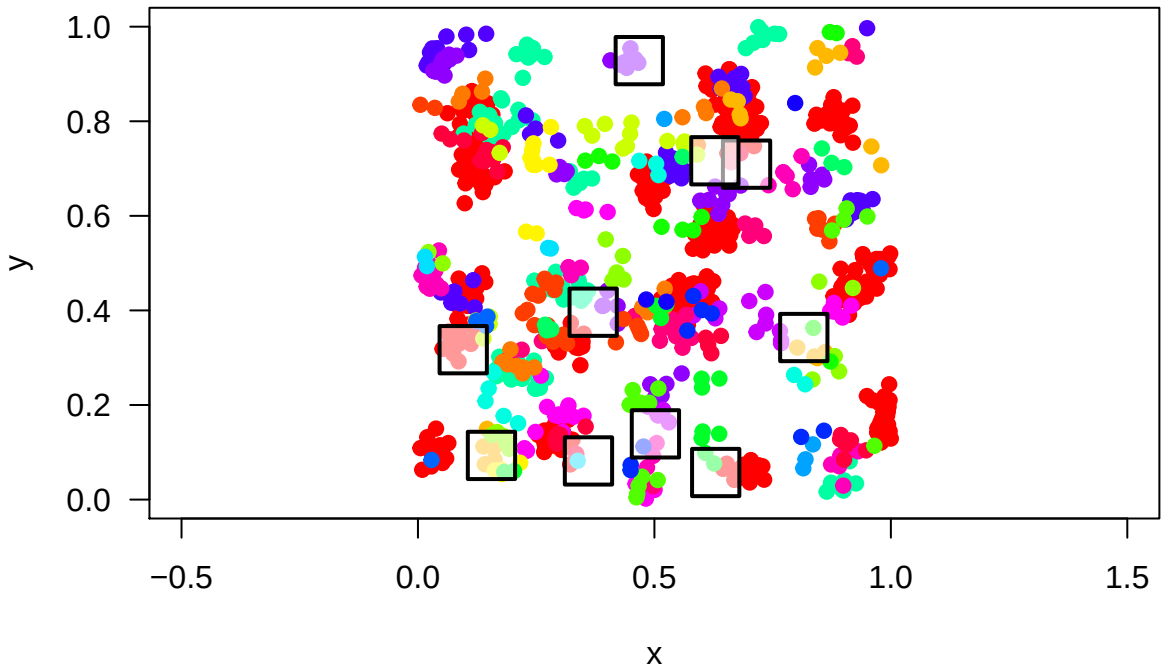
```
# poisson community
pois_comm <- mobsim::sim_poisson_community(25, 1000)
pois_quad <- mobsim::sample_quadrats(pois_comm, 10)
```

```
## Warning in mobsim::sample_quadrats(pois_comm, 10): There are overlapping
## sampling squares in the design
```




```
# thomas community
thom_comm <- mobsim::sim_thomas_community(25, 1000)
thom_quad <- mobsim::sample_quadrats(thom_comm, 10)
```

```
## Warning in mobsim::sample_quadrats(thom_comm, 10): There are overlapping
## sampling squares in the design
```




```
# factor vector
sources <- c(rep("poisson", 10), rep("thomas", 10))

# full SbyS
all_sites <- bind_rows(pois_quad$spec_dat, thom_quad$spec_dat)
```

```
res <- vegan::adonis(all_sites ~ sources)
print(res)
```

```
##
## Call:
## vegan::adonis(formula = all_sites ~ sources)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## sources    1    1.1913 1.19131  4.2549 0.19119  0.001 ***
## Residuals 18    5.0397 0.27998         0.80881
## Total     19    6.2310                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a significant relationship between source type and species composition!

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(readr)
```

```
## Loading required package: readr
```

```
require(dplyr)
require(tidyr)
require(vegan)
require(purrr)
```

```
## Loading required package: purrr
```

```
require(ggplot2)

# abundance data
df <- read_csv('~/projects/QB-2021-project/data/taxapercent.csv')
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   X1 = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```
taxa <- df %>%
  mutate(layer = X1) %>%
  select(-"pop%", -X1)
names(taxa)
```

```
##  [1] "actinopterygii%"    "anura%"             "chelonia%"
##  [4] "crocodylidae%"      "squamata%"          "aves%"
```

```
##  [7] "insectivora%"       "chiroptera%"         "primates%"
## [10] "rodentia%"          "lagomorpha%"         "canidae%"
## [13] "mustelidae%"        "viverridae%"         "hyaenidae%"
## [16] "felidae%"           "proboscidea%"        "hyracoidea%"
## [19] "equidae%"           "chaclicotheriidae%" "rhinocerotidae%"
## [22] "suidae%"            "hippopotamidae%"     "giraffidae%"
## [25] "bovidae%"           "layer"
```

```
labels <- as_vector(taxa['layer'])

abundance_bc <- vegan::vegdist(taxa[, 1:(ncol(taxa)-1)])

# define environmental matrix
environ <- read_csv("~/projects/QB-2021-project/data/totalenv.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   age_ma = col_double(),
##   carb_c13 = col_double(),
##   carb_o18 = col_double()
## )
```

```
times <- read_csv("~/projects/QB-2021-project/data/unitages.csv")
```

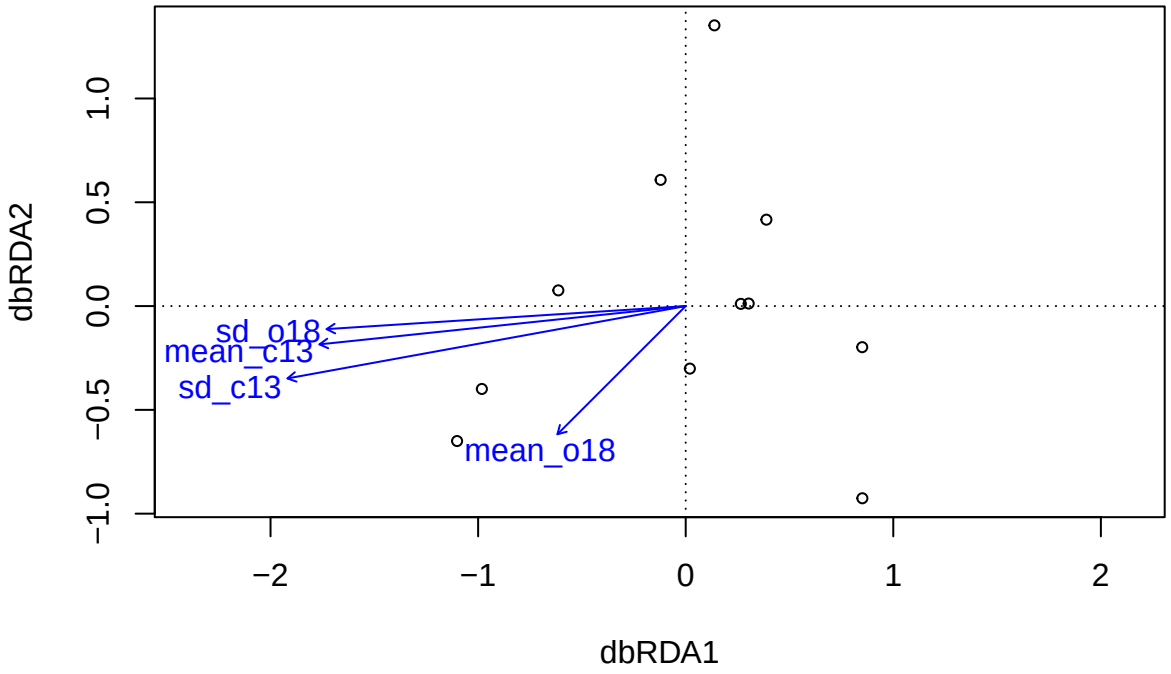
```
##
## -- Column specification --------------------------------------------------------
## cols(
##   layer = col_character(),
##   age_ma = col_double()
## )
```

```
# lower bed1
window_size = 0.4
df_list <- vector("list", 10)
for (i in 1:11) {
  row <- environ %>%
  filter(age_ma > times$age_ma[i] & age_ma < times$age_ma[i] + window_size) %>%
  mutate(layer = labels[i]) %>%
  summarize(layer = layer, mean_c13 = mean(carb_c13), sd_c13 = sd(carb_c13),
            mean_o18 = mean(carb_o18), sd_o18 =  sd(carb_o18))
  df_list[[i]] <- row[1,]
}
env <- bind_rows(df_list)
# model selection
# no vars
taxa_dbrda_none <- vegan::dbrda(abundance_bc ~ 1, env[,2:ncol(env)])
# all vars
taxa_dbrda_full <- vegan::dbrda(abundance_bc ~ ., env[,2:ncol(env)])
# test fits
taxa_dbrda <- vegan::ordiR2step(taxa_dbrda_none, taxa_dbrda_full, perm_max = 200)
```

```
## Step: R2.adj= 0
## Call: abundance_bc ~ 1
##
##                 R2.adjusted
```

```
## + sd_c13          0.17845003
## + mean_c13        0.14062734
## + sd_o18          0.13926644
## <All variables>   0.09558481
## <none>            0.00000000
## + mean_o18       -0.02161078
```

```
## check it out!
# taxa_dbrda$call
# taxa_dbrda$anova

# nmone of the env vectors matter but lets plot the whole thing anyway
vegan::ordiplot(taxa_dbrda_full)
```

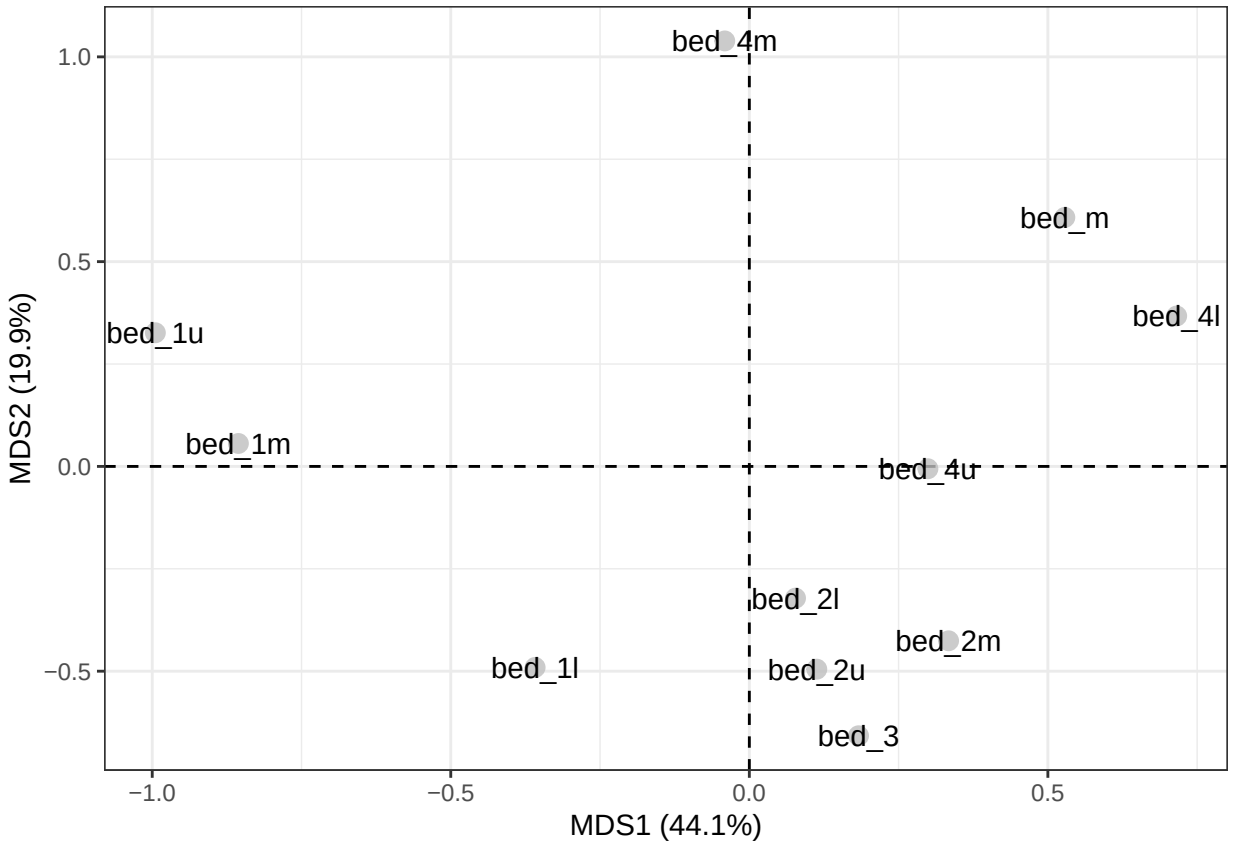



```
# Let's make a fancy plot
expvar <- round(sapply(taxa_dbrda$CA$eig, function(x) x /
                       sum(c(taxa_dbrda$CA$eig))), 3) * 100

# convert the plots to a dataframe
dbrda_tib <- tibble::as_tibble(scores(taxa_dbrda)) %>% mutate(layer = labels)

env_vecs <- tibble::as_tibble(scores(taxa_dbrda, display = "bp"))
env_vecs <- mutate(env_vecs, name = "vec") # c('Oxygen', 'Demand', 'Nitrate'))

print(env_vecs)
```

```
## # A tibble: 0 x 1
## # ... with 1 variable: name <chr>
```

```
g <- ggplot() +
  geom_hline(yintercept=0, linetype=2) +
  geom_vline(xintercept = 0, linetype=2) +
  geom_point(data=dbrda_tib, aes(x = MDS1, y = MDS2), size = 3, alpha = 0.8, color = 'grey') +
  geom_text(data=dbrda_tib, aes(x = MDS1, y = MDS2, label = layer)) +
```

```
  #geom_segment(data=env_vecs, aes(x = 0, y = 0, xend = dbRDA1, yend = dbRDA2), arrow = arrow(), alpha =
  #geom_text(data=env_vecs, aes(x = dbRDA1, y = dbRDA2 + 0.2, label = name)) +
  xlab(paste('MDS1 (', expvar[1], '%)', sep = '')) +
  ylab(paste('MDS2 (', expvar[2], '%)', sep = '')) +
  theme_bw()
g
```



So none of the environmental vectors matter according to R2. The fancy ggplot shows PCoA, while the `vegan`-generated `ordiplot` shows the full model. Just to double check let's see if PERMANOVA finds a relationship between the mean isotope values and the species composition.

```
mean_carbon <- as_vector(env[,'mean_c13'])
mean_oxygen <- as_vector(env[,'mean_o18'])
print(adonis(taxa[, 1:(ncol(taxa)-1)] ~ mean_oxygen), method = 'bray', permutations=999)
```

```
##
## Call:
## adonis(formula = taxa[, 1:(ncol(taxa) - 1)] ~ mean_oxygen)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##             Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## mean_oxygen  1   0.06759 0.067587 0.78846 0.08055  0.572
## Residuals    9   0.77148 0.085720         0.91945
```

```
## Total       10   0.83907                  1.00000
```

```
print(adonis(taxa[, 1:(ncol(taxa)-1)] ~ mean_carbon), method = 'bray', permutations=999)
```

```
##
## Call:
## adonis(formula = taxa[, 1:(ncol(taxa) - 1)] ~ mean_carbon)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##             Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## mean_carbon  1   0.19010 0.190104  2.6364 0.22656  0.036 *
## Residuals    9   0.64897 0.072107         0.77344
## Total       10   0.83907                  1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The PERMANOVA shows us that there is a significant relationship between Carbon-13 and species composition but oxygen seems to play no role.