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

Herbert Sizek; Z620: Quantitative Biodiversity, Indiana University

27 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

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

```
## [1] "D:/GitHub/QB2021_Sizek/2.Worksheets/8.BetaDiversity"
```

```
setwd("D:/GitHub/QB2021_Sizek/2.Worksheets/8.BetaDiversity")
```

4. load the `vegan` R package (be sure to install if needed).

```
require(vegan)
```

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

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

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

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

```
require(ade4)
```

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

```
require(indicspecies)
```

```
## Loading required package: indicspecies
```

2) LOADING DATA

Load dataset

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

```
# note, pleae do not print the dataset when submitting
data(doubs)
```

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”.

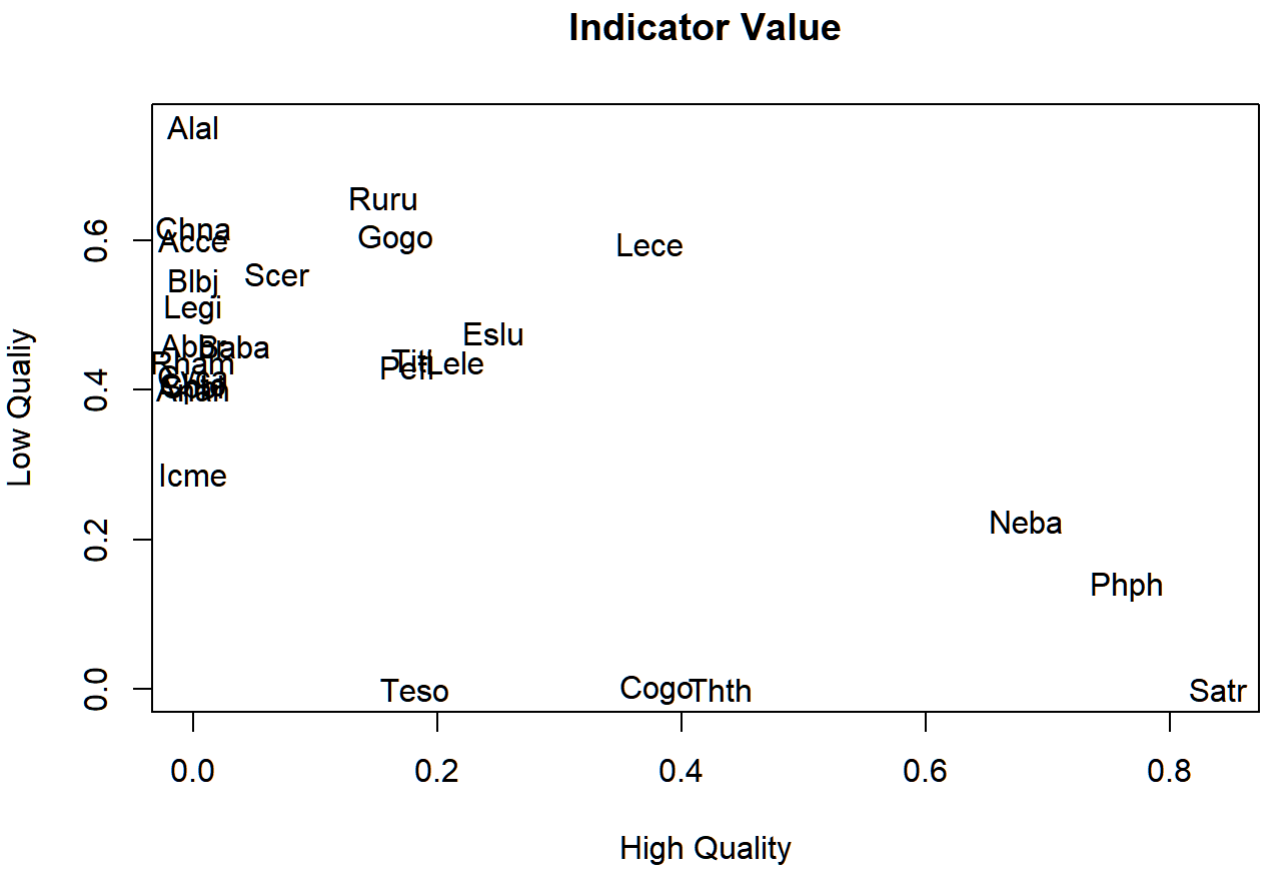
In the code chunk below, test the hypothesis that fish community composition varies with river quality.

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.

```
fish <- doubs$fish
fish<- fish[-which(rowSums(fish)==0),]
quality <- c(rep("HQ",13),rep("MQ",5),rep("LQ",6),rep("MQ",5))
#PERMANOVA
adonis(fish ~ quality, method ='bray',permutations =999)
```

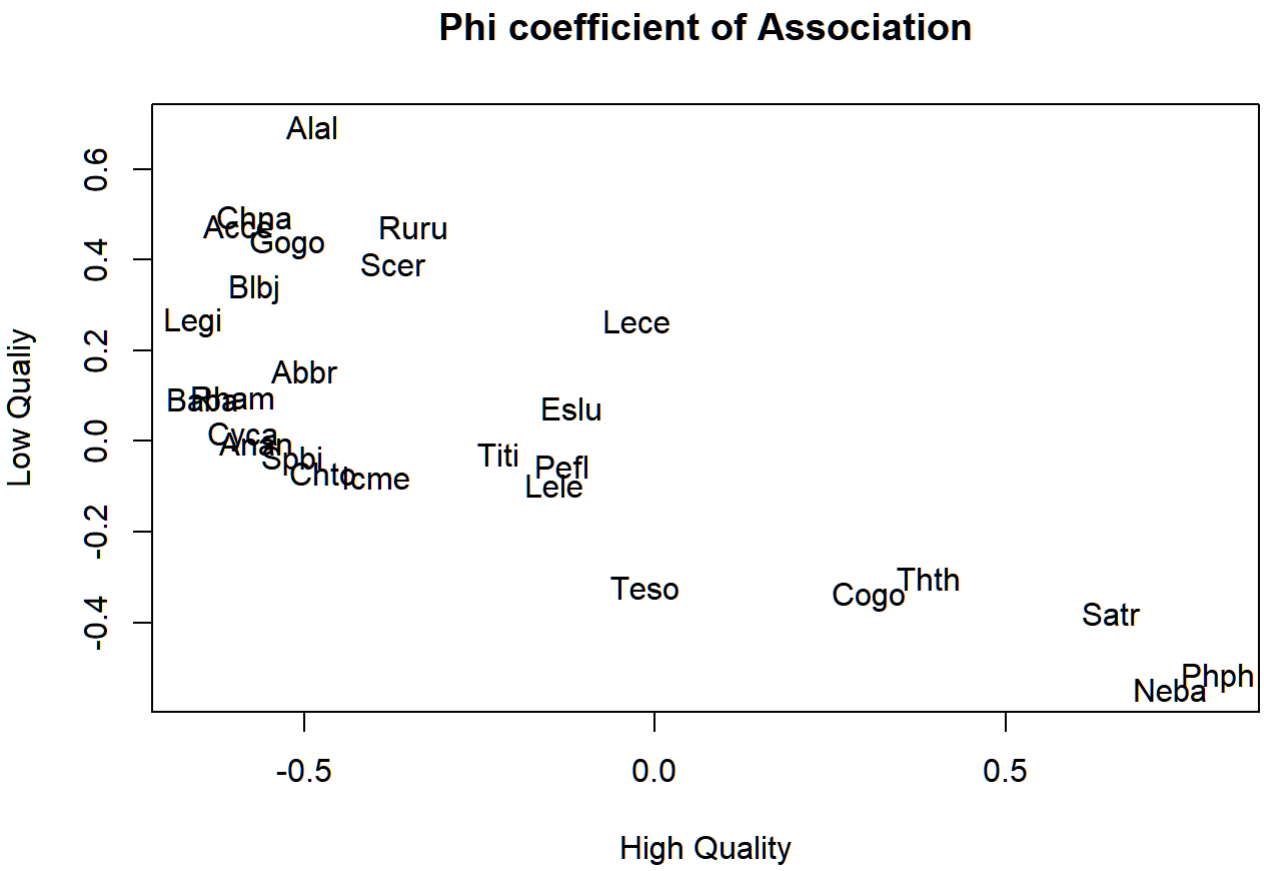
```
##
## Call:
## adonis(formula = fish ~ quality, permutations = 999, method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## quality    2    3.0947 1.54733   10.97 0.45765  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
```

```
# Indicator Species
indval <- multipatt(fish, cluster = quality, func = "IndVal.g",control=how(nperm = 999))
plot(indval$str[,1],indval$str[,2],xlab="High Quality",ylab="Low Qualiy",type="n",main="Indicato
r Value")
text(indval$str[,1],indval$str[,2],label=row.names(indval$str))
```




```
fish.rel <- decostand(fish, method ="total")
phi <- multipatt(fish.rel, cluster = quality, func= "r.g",control=how(nperm=999))
```

```
plot(phi$str[,1],phi$str[,2],xlab="High Quality",ylab="Low Qualiy",type="n", main="Phi coefficie
nt of Association")
text(phi$str[,1],phi$str[,2],label=row.names(phi$str))
```



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 quality metric here is highly correlated with the Bray Curtis measure. There are also some speices that are indicators of “high quality” habitats which are inversely related to those of “low quality” habitats as seen by both the indicator value plot, and the phi coefficient of association. This generally agrees with previous analysis and observations, such as with Alal, Phph, and Neba.

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%
## 0.111 0.149 0.181 0.218
## 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 suggests that the distance between sampled sites correlates with the species correlations as determined by Bray-Curtis. This confirms the idea previously discussed, that the site quaility label correlates with the fish community diversity. We can not say that there is a causal relation with any of these tests as all of the tests we've preformed so far are corrlative not causitive, so I don't know about the stream quality influencing fish communities.

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,

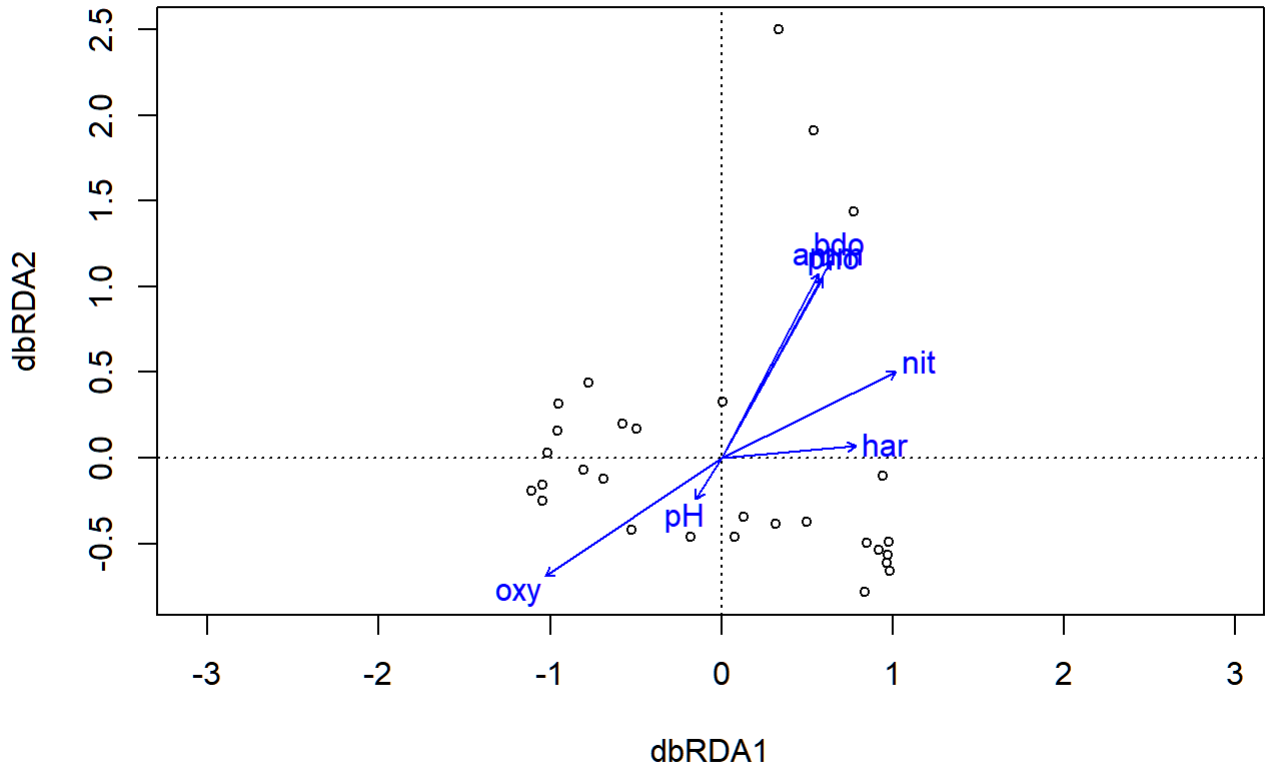
```
env.chem <- as.matrix(doubs$env[-8,5:11])
fish.db <- vegdist(fish,method="bray",upper=TRUE,diag=TRUE)
doubs.dbrda <- dbrda(fish.db ~ . , as.data.frame(env.chem))
var1 <- round(doubs.dbrda$CCA$eig[1]/sum(doubs.dbrda$CCA$eig),3)
var2 <- round(doubs.dbrda$CCA$eig[2]/sum(doubs.dbrda$CCA$eig),3)
var1
```

```
## dbRDA1
##  0.668
```

```
var2
```

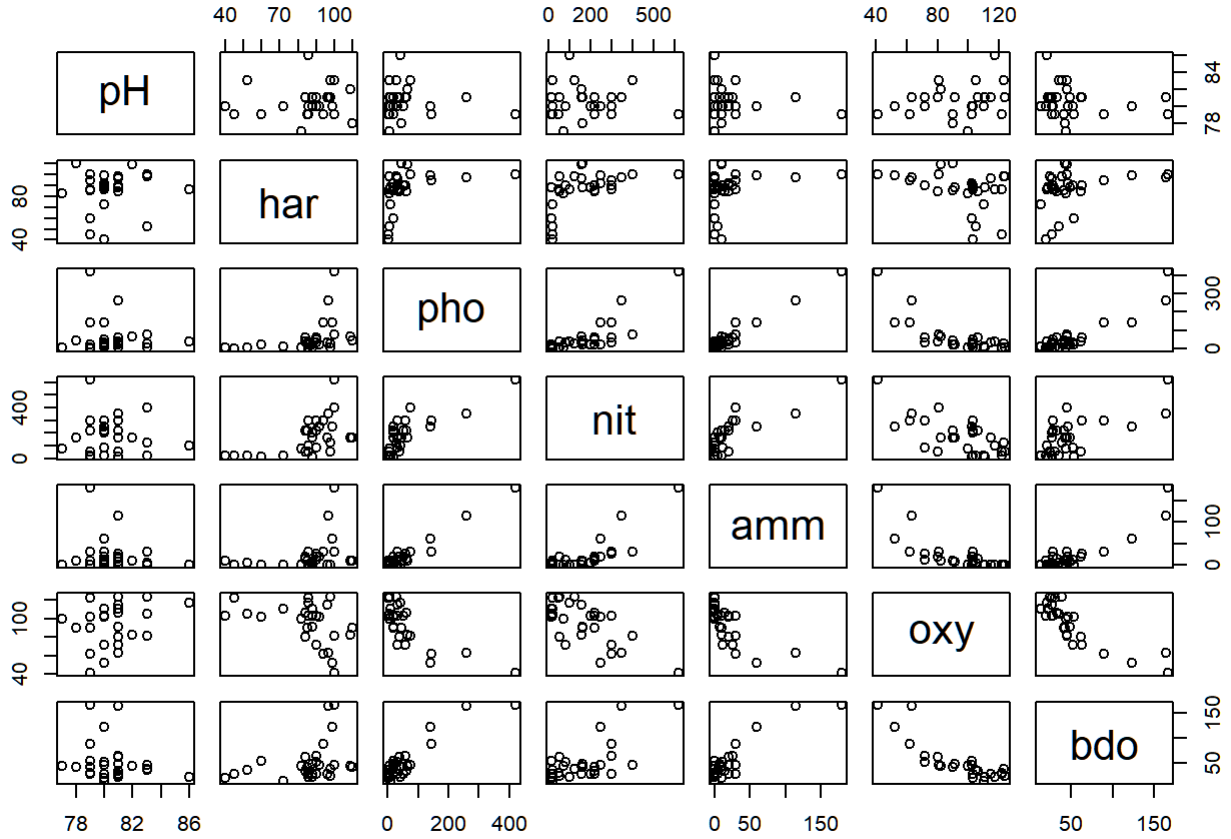
```
## dbRDA2
##  0.211
```

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



This first round shows that the eigenvector for multiple environmental variables are aligned with each other, this probably indicates that they are highly correlated.

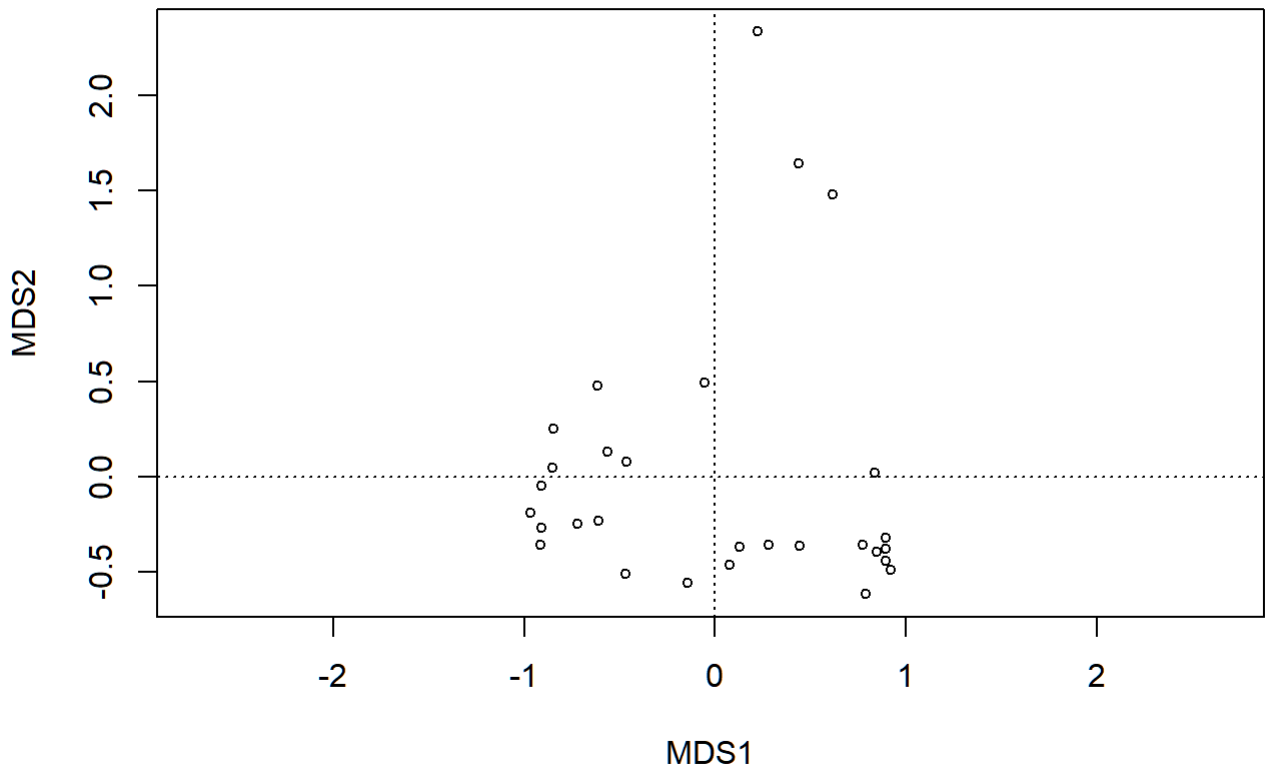
```
pairs(env.chem)
```



from a brief glance betwen pho amm and bdo this does appear to be the case.

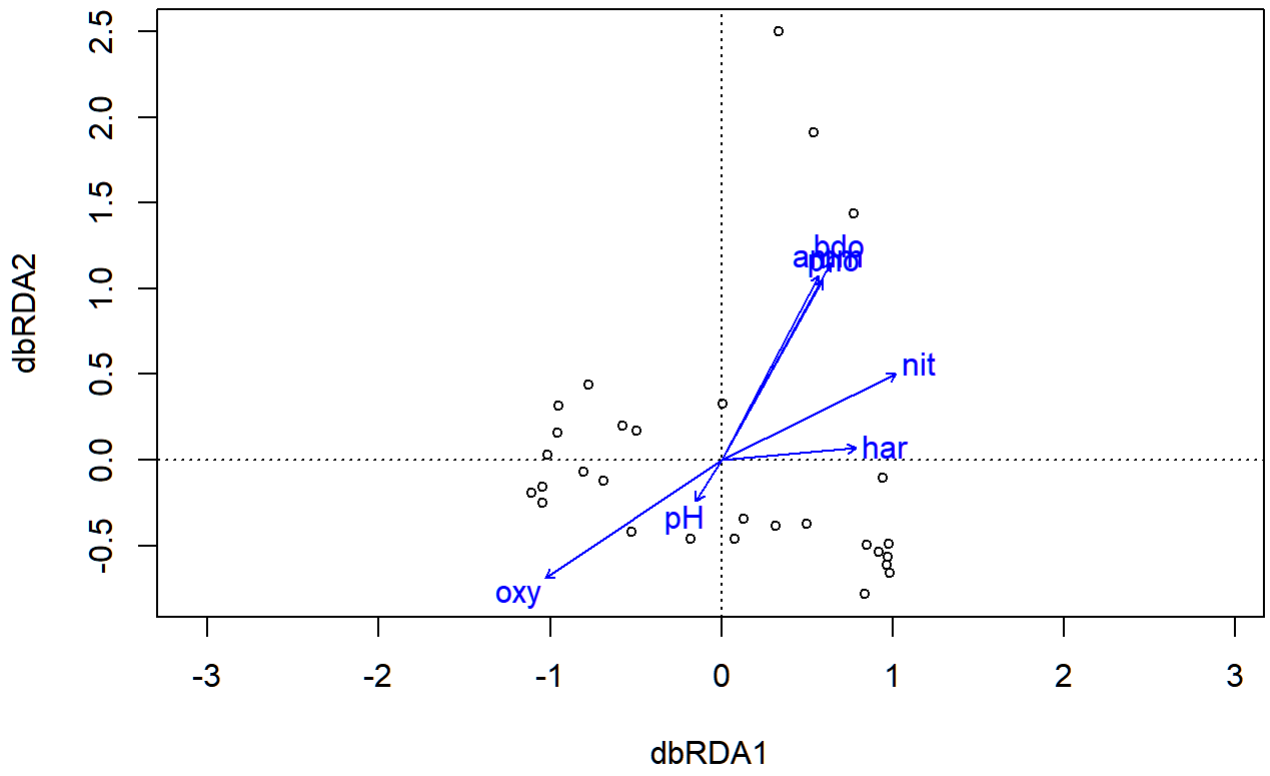
forward case (no vectors):

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



Reverse case with all variables

```
doubs.dbrda.mod1 <- dbrda(fish.db ~ ., as.data.frame(env.chem))
ordiplot(doubs.dbrda.mod1)
```



now do the different possible combinations of variables and accounts

```
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  0.004 **
## ---
## 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
```

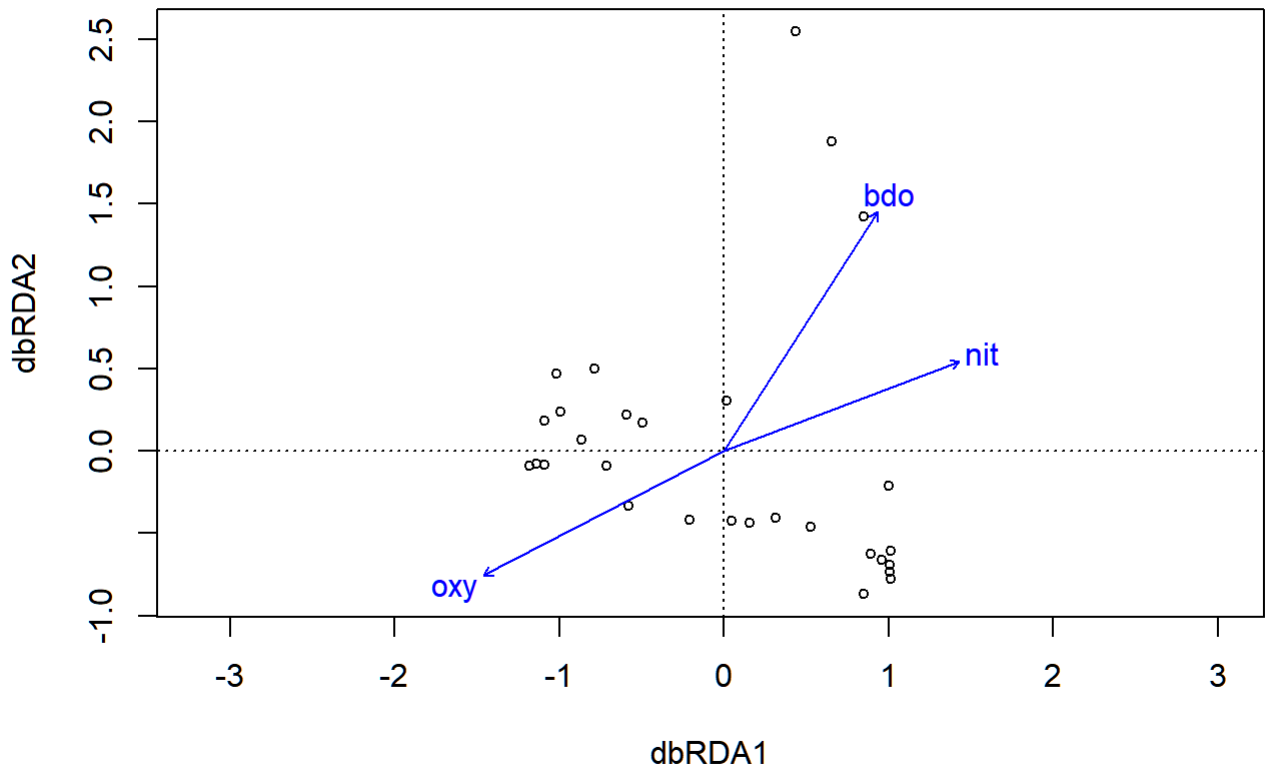
```
## dbrda(formula = fish.db ~ oxy + bdo + nit, data = as.data.frame(env.chem))
```

```
doubs.dbrda$anova
```

	R2.adj <dbl>	Df <dbl>	AIC <dbl>	F <dbl>	Pr(>F) <dbl>
+ oxy	0.2772718	1	47.93933	11.742086	0.002
+ bdo	0.4009000	1	43.40432	6.571630	0.002
+ nit	0.4980793	1	39.13432	6.033983	0.004
<All variables>	0.5303258	*NA*	*NA*	*NA*	*NA*

4 rows

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



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,

```
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
```

4. use a permutation test to determine the correlation of each environmental factor on the constrained axes,

```
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
```

5. calculate the explained variation on the first and second constrained axes,

```
var1 <- round(doubs.dbrda$CCA$eig[1]/sum(doubs.dbrda$CCA$eig),3)
var2 <- round(doubs.dbrda$CCA$eig[2]/sum(doubs.dbrda$CCA$eig),3)
var1
```

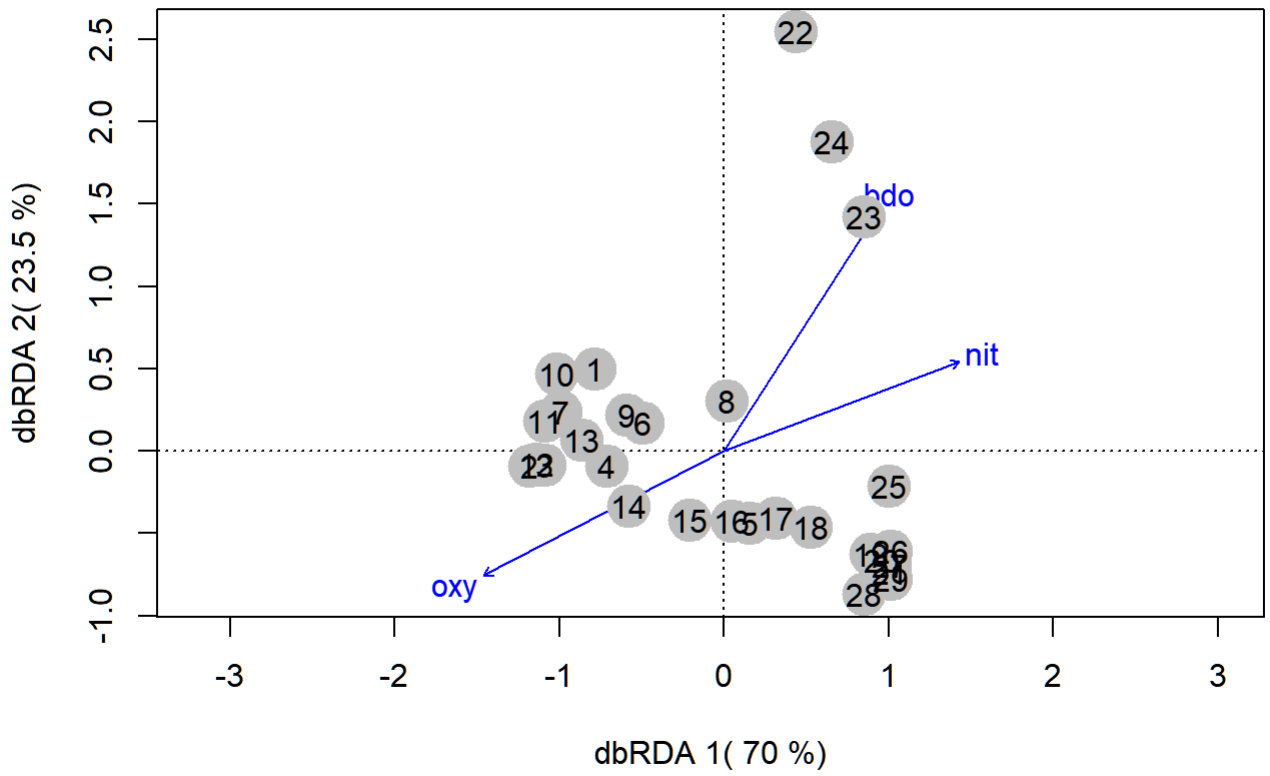
```
## dbRDA1
##    0.7
```

```
var2
```

```
## dbRDA2
##  0.235
```

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.

```
ordiplot(doubs.dbrda,
         xlab = paste("dbRDA 1(",var1*100,"%)"),
         ylab = paste("dbRDA 2(",var2*100,"%)"))
points(scores(doubs.dbrda,display = "wa"), bg="gray", col="gray", pch =19,cex=3)
text(scores(doubs.dbrda, display = "wa"))
```



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***: oxygen, nitrogen, and the set of bdo, amm, and pho

iii. Variation Partitioning

In the code chunk below,
> Some notes on what is occuring here for my records. We are interested in how much (un)certainty environment and spatial variables are giving through a correlation matrix. The eigenvectors of the matrix are a function of the certainty of the matrix. In the case of PCNM, the eigenvectors explain a specific amount of certainty of the system. From that we can calculate the conditional PCNM, the certainty conditioned on a maximization, to give us the mutual information between the different conditions, the environment and the space. The joint entropy expresses the residual, is how much uncertainty is left in the system by knowing the environment and the space. The language used is a direct copy from information theory and mutual information, but I'm less sure if the math would works it all the way through.

1. Create a matrix model of the selected environmental variables,

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

2. Create a matrix model of the selected PCNM axes,

```
rs <- rowSums(fish)/sum(fish)
doubs.pcnmw <- pcnm(dist(doubs$x[-8,]),w=rs,dist.ret=TRUE)
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)
```

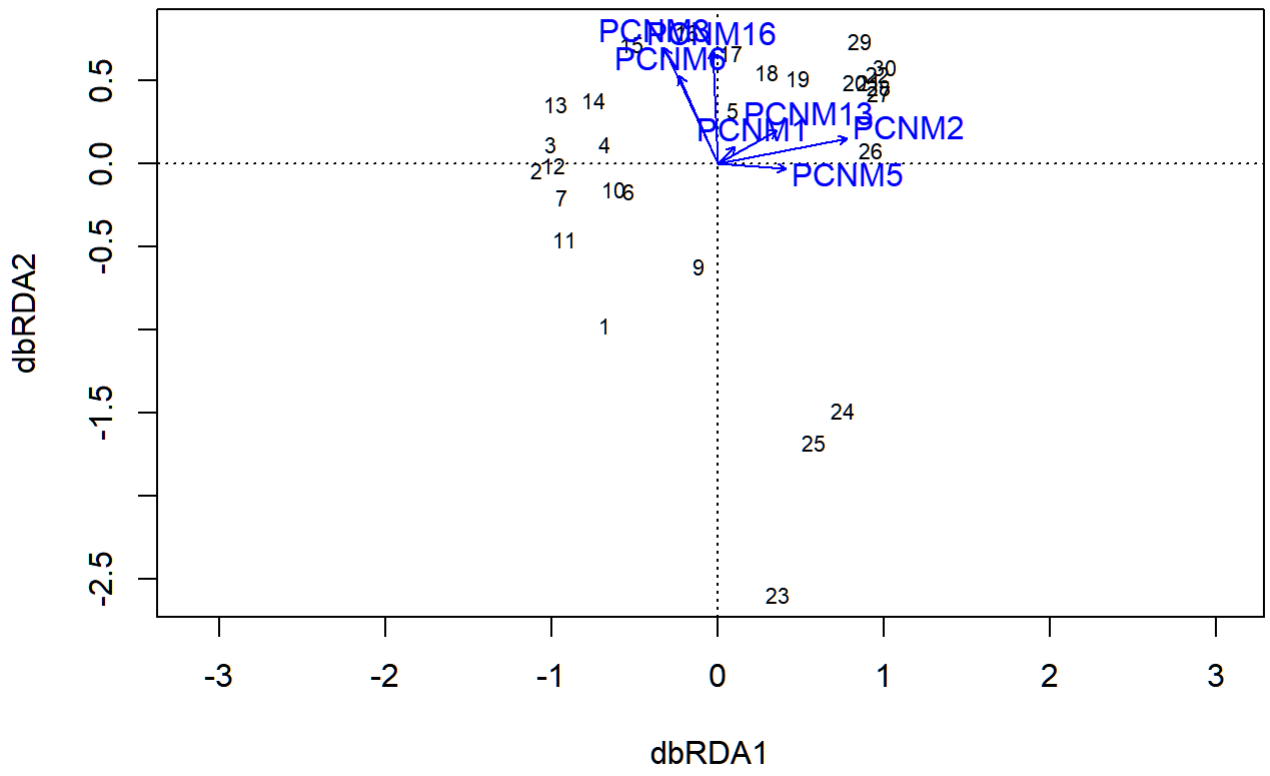
```
## Step: R2.adj= 0
## Call: fish.db ~ 1
##
##                  R2.adjusted
## <All variables>  0.626011301
## + PCNM2          0.235370423
## + PCNM3          0.078394885
## + PCNM13         0.065305668
## + PCNM5          0.046185074
## + PCNM6          0.032809156
## + PCNM16         0.030486700
## + 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  0.004 **
## ---
## 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  0.004 **
## ---
## 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  0.018 *
## ---
## 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  0.014 *
```

```
## ---
## 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  0.014 *
## ---
## 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  0.038 *
## ---
## 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  0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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




```
step.pcnm$anova
```

	R2.adj <dbl>	Df <dbl>	AIC <dbl>	F <dbl>	Pr(>F) <dbl>
+ PCNM2	0.2353704	1	49.57372	9.619039	0.004
+ PCNM3	0.3429270	1	46.08295	5.419644	0.004
+ PCNM5	0.4076020	1	43.94068	3.838544	0.018
+ PCNM1	0.4721469	1	41.41138	4.056958	0.014
+ PCNM13	0.5212427	1	39.34605	3.461157	0.020
+ PCNM16	0.5767968	1	36.48005	4.019224	0.014
+ PCNM6	0.6043089	1	35.18163	2.529645	0.038
<All variables>	0.6260113	*NA*	*NA*	*NA*	*NA*

8 rows

3. Perform constrained and partial constrained ordinations using the spatial and environmental models you just created,

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

4. Test the significance of each of your constrained ordinations using permutation tests,

```
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
```

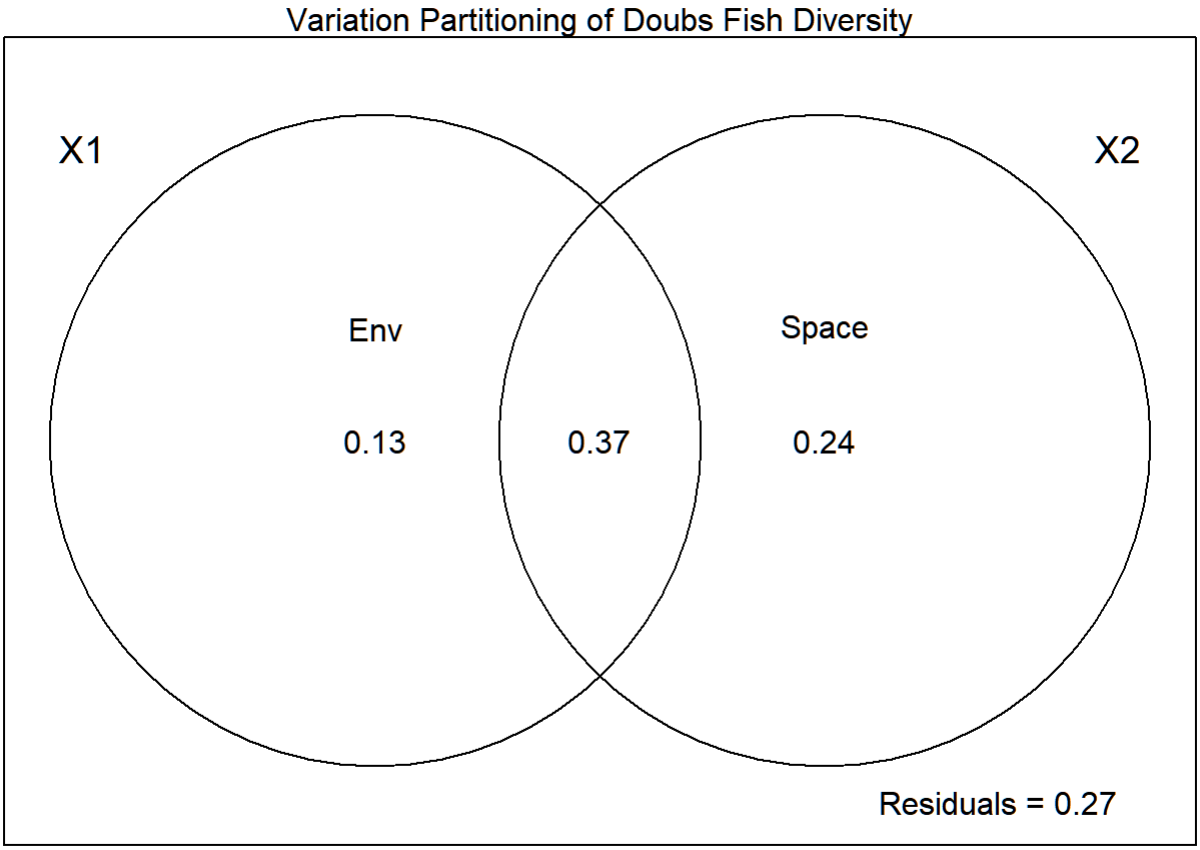
5. Partition the variation among sites into the relative importance of space, environment, spatially structured environment, and residuals,

```
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
```

6. Plot the variation partitioning output to visualize it.

```
par(mar=c(2,2,2,2))
plot(doubs.varpart)
text(1,0.25,"Space")
text(0,0.25,"Env")
mtext("Variation Partitioning of Doubs Fish Diversity",side =3)
```



Question 4: Interpret the variation partitioning results.

> ***Answer 4***: I don't know how to explain this succinctly. From what I understand, we are comparing two eigenvectors. They are pointing in different directions, so the shared component, the information provided by the environment and space, resolves 0.37 of the variation of species diversity. The two eigenvectors independent components resolve another 0.13 and 0.24 of the variation in species diversity, while their is an additional 0.27 not accounted for by these two eigenvectors, the residuals. Because this is not causal analysis, we can say that the environment and space information correlates to the species diversity more so than just the space or the environment (with environment being the lowest).

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.

```
require(mobsim)
```

```
## Loading required package: mobsim
```

```
## Warning: package 'mobsim' was built under R version 3.6.3
```

```
comA <- sim_poisson_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))

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

Take ten (10) subsamples from each site using the quadrat function and answer the following questions:

```
comm_matA <- sample_quadrats(comA, n_quadrats = 10, quadrat_area = 0.03,
             avoid_overlap = T, plot=F)
```

```
## Warning in sample_quadrats(comA, n_quadrats = 10, quadrat_area = 0.03, avoid_overlap = T, : C
annot find a sampling layout with no overlap.
##                                             Install the package spatstat for an improved meth
od for non-overlapping squares,
##                                             Use less quadrats or smaller quadrat area, or set
avoid_overlap to FALSE.
```

```
row.names(comm_matA$spec_dat)= paste("R",row.names(comm_matA$spec_dat),sep="")
comm_matB <- sample_quadrats(comB, n_quadrats = 10, quadrat_area = 0.03,
             avoid_overlap = T, plot=F)
```

```
## Warning in sample_quadrats(comB, n_quadrats = 10, quadrat_area = 0.03, avoid_overlap = T, : C
annot find a sampling layout with no overlap.
##                                             Install the package spatstat for an improved meth
od for non-overlapping squares,
##                                             Use less quadrats or smaller quadrat area, or set
avoid_overlap to FALSE.
```

```
row.names(comm_matB$spec_dat)= paste("C",row.names(comm_matB$spec_dat),sep="")
```

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

> This question is poorly worded. Because of the data structure, we can not determine causation, only correlation. Also, there is no indication of what the treatments/replicates are. Here is a random go at a potential way that this could be done, but it isn't that good because of the different underlying distributions.

```
require(plyr)
```

```
## Loading required package: plyr
```

```
## Warning: package 'plyr' was built under R version 3.6.3
```

```
genstructure <- c(rep("Random",10),rep("Clustered",10))
comm_matAB <- rbind.fill(comm_matA$spec_dat,comm_matB$spec_dat)
comm_matAB[is.na(comm_matAB)] <-0
adonis(comm_matAB ~ genstructure, method="bray", permutations=999)
```

```
##
## Call:
## adonis(formula = comm_matAB ~ genstructure, permutations = 999,      method = "bray")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## genstructure  1    1.0675 1.06748    5.29 0.22714  0.001 ***
## Residuals    18    3.6322 0.20179         0.77286
## Total        19    4.6997                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

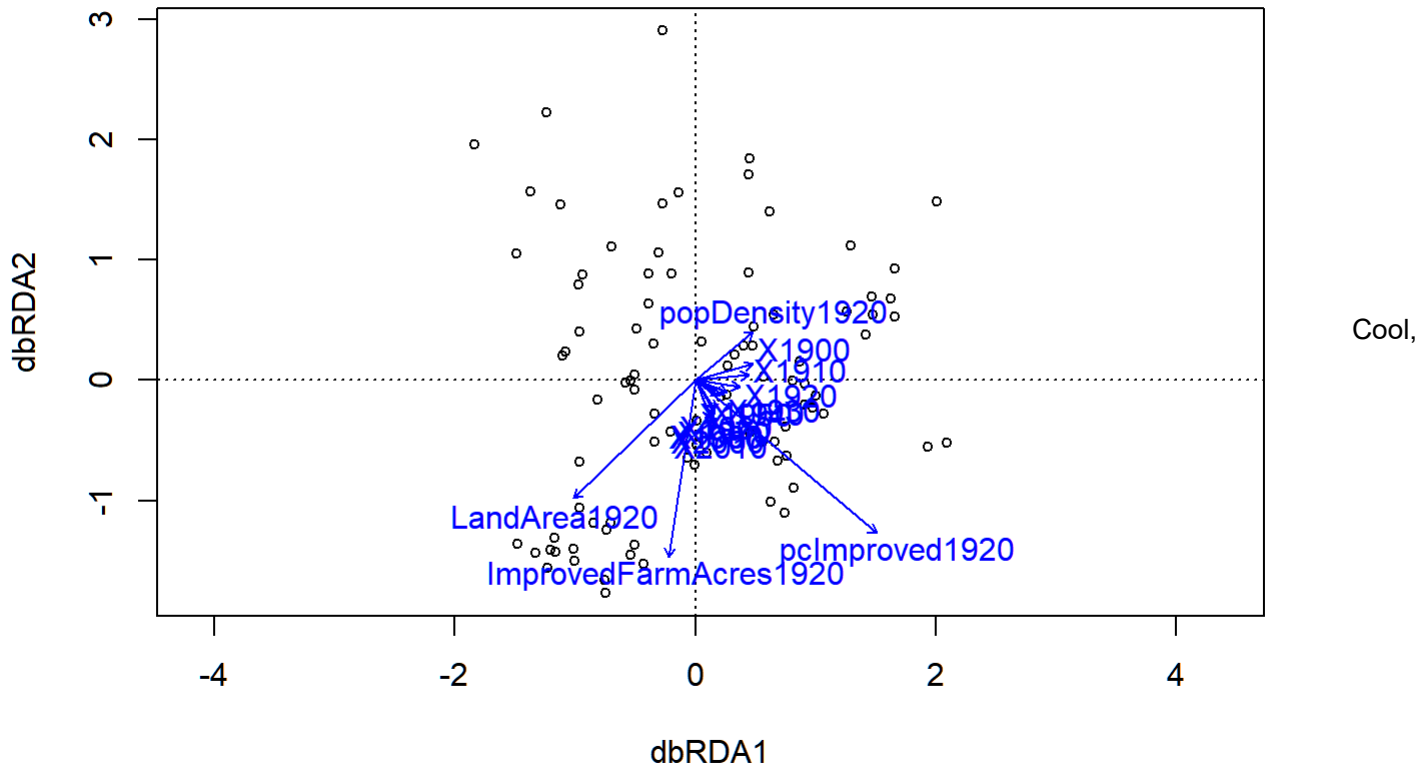
2. Load the dataset you are using for your Team Project. Perform an ordination to visualize your dataset. Using this ordination, develop some hypotheses relevant to β-diversity. Use a statistic to test one of these hypotheses. Succinctly explain the finding and its relevance to your system.

```
# Woody trees data
woodytreedeamssm3 <- as.data.frame(read.csv("data/woodytrees.csv",header=TRUE))
rownames(woodytreedeamssm3)<- woodytreedeamssm3[,1]
woodytreedeamssm3<- woodytreedeamssm3[,-1]
# US population Census 1920
UScensus <- as.data.frame(read.csv("data/IndianaCensus.csv",header=TRUE))
rownames(UScensus) <- UScensus[,1]
UScensus <- UScensus[,-1]
# Census of Agriculture of 1920
AgCensus <- as.data.frame(read.csv("data/IndianaAgCensus1920.csv",header=TRUE))
rownames(AgCensus) <- AgCensus[,1]
AgCensus <- AgCensus[,-1]
```

```
popData <- cbind(USCensus[-1,],AgCensus[-1,])
popData$pcImproved1920 <- popData$ImprovedFarmAcres1920/popData$LandArea1920
popData$popDensity1920 <- popData$X1920/popData$LandArea1920


woodytreedeamssm3.db <- vegdist(woodytreedeamssm3,method='bray',binary = TRUE,upper=TRUE,diag=TR
UE)

woodytreedeamssm3.dbrda <- dbrda(woodytreedeamssm3.db ~.,as.data.frame(popData))
ordiplot(woodytreedeamssm3.dbrda)
```



Cool,

from this it is clear that the census population levels are minimally correlated, while the land area of counties and improved acres are more correlated.

```
woodytreedeamssm3.dbrda.mod0 <-dbrda(woodytreedeamssm3.db~1,as.data.frame(popData))
# ordiplot(woodytreedeamssm3.dbrda.mod0)
```

```
woodytreedeamssm3.dbrda.mod1 <- dbrda(woodytreedeamssm3.db~., as.data.frame(popData))
# ordiplot(woodytreedeamssm3.dbrda.mod1)
```

```
woodytreedeamssm3.dbrda <- ordiR2step(woodytreedeamssm3.dbrda.mod0,woodytreedeamssm3.dbrda.mod1,
perm.max=200)
```

```
## Step: R2.adj= 0
## Call: woodytreedeamssm3.db ~ 1
##
##                           R2.adjusted
## <All variables>          6.106465e-02
## + pcImproved1920         2.455358e-02
## + LandArea1920           1.344601e-02
## + ImprovedFarmAcres1920  1.107929e-02
## + X1910                  1.469237e-04
## + X1900                  7.384027e-05
## <none>                   0.000000e+00
## + popDensity1920        -4.247067e-04
## + X1920                 -5.000502e-04
## + X1930                 -1.276819e-03
## + X1940                 -1.385791e-03
## + X1950                 -1.706708e-03
## + X1980                 -1.773088e-03
## + X1990                 -1.799462e-03
## + X2000                 -1.809320e-03
## + X1970                 -1.820925e-03
## + X2010                 -1.923783e-03
## + X1960                 -2.040614e-03
##
##                  Df    AIC      F Pr(>F)
## + pcImproved1920  1 260.31 3.2906  0.002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.02455358
## Call: woodytreedeamssm3.db ~ pcImproved1920
##
##                          R2.adjusted
## <All variables>          0.06106465
## + ImprovedFarmAcres1920  0.03958766
## + LandArea1920           0.03861442
## + X1910                  0.02490417
## + popDensity1920         0.02464179
## + X1920                  0.02459080
## <none>                   0.02455358
## + X1900                  0.02448953
## + X1930                  0.02402390
## + X1940                  0.02393225
## + X1950                  0.02362122
## + X1970                  0.02358689
## + X1980                  0.02356842
## + X1990                  0.02350552
## + X2000                  0.02345176
## + X1960                  0.02335954
## + X2010                  0.02330089
##
##                          Df    AIC      F Pr(>F)
## + ImprovedFarmAcres1920  1 259.86 2.4088  0.004 **
## ---
```

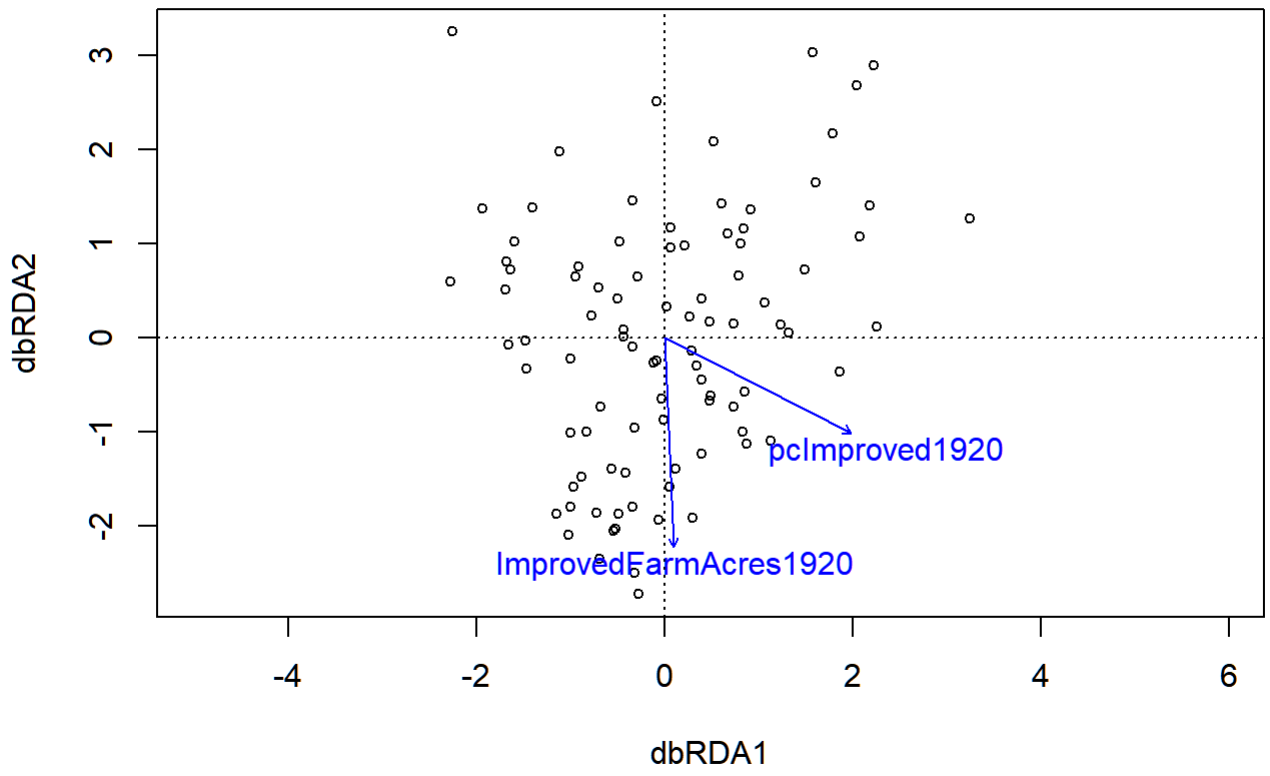
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step: R2.adj= 0.03958766
## Call: woodytreedeamssm3.db ~ pcImproved1920 + ImprovedFarmAcres1920
##
##                  R2.adjusted
## <All variables>   0.06106465
## + X1910           0.04162100
## + X1900           0.04157352
## + X1920           0.04065366
## + popDensity1920  0.03981553
## + X1930           0.03975224
## <none>            0.03958766
## + X1940           0.03947776
## + X1950           0.03905245
## + X1970           0.03884227
## + X1980           0.03867984
## + X1990           0.03864197
## + X2000           0.03860370
## + X1960           0.03860028
## + X2010           0.03846327
## + LandArea1920    0.03465277
##
##         Df    AIC      F Pr(>F)
## + X1910  1 260.62 1.1888  0.288
```

```
woodytreedeamssm3.dbrda$anova
```

	R2.adj	Df	AIC	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
+ pcImproved1920	0.02455358	1	260.3143	3.290619	0.002
+ ImprovedFarmAcres1920	0.03958766	1	259.8574	2.408840	0.004
<All variables>	0.06106465	*NA*	*NA*	*NA*	*NA*

3 rows

```
ordiplot(woodytreedeamssm3.dbrda)
```



OOF only 4% of variation explained by these the agricultural information we inputted, which probably isn't that bad given that we are talking about species diversity across an entire state, but also not good because we don't have a set of variables that explain the variance found. There is high likelyhood that this variation could just be due to somewhat random effects.