Bog_Data

Andrea Phillips and Emmi Mueller; Z620: Quantitative Biodiversity, Indiana University

01 March, 2019

```
#Clear workspace, set working directory and load packages
rm(list=ls())
getwd()
```

```
## [1] "C:/Users/emmim/GitHub/QB_Project"
```

```
#setwd("~/GitHub/QB_Project/")

require("tidyverse")
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------------------------
```

```
## v ggplot2 3.1.0     v purrr   0.3.0
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0
```

```
## -- Conflicts -----------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
require("vegan")
```

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

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

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

```
## This is vegan 2.5-4
```

```
require("ggplot2")
require("OTUtable")
```

```
## Loading required package: OTUtable
```

```
#Load bog otu data and print structure
data(otu_table)
otu_table <- as.data.frame(t(otu_table))
```

```
#Print rarefaction curve and rarefy otu_table to otu.rarefy
min.N <- min(rowSums(otu_table))
rarecurve(x = otu_table, step = 10, col = "blue", cex = 0.6, las = 1, main = "Bog rarefaction curve")
abline(0,1, col = 'red')
abline(v = min.N, col = 'green')
```

Bog rarefaction curve
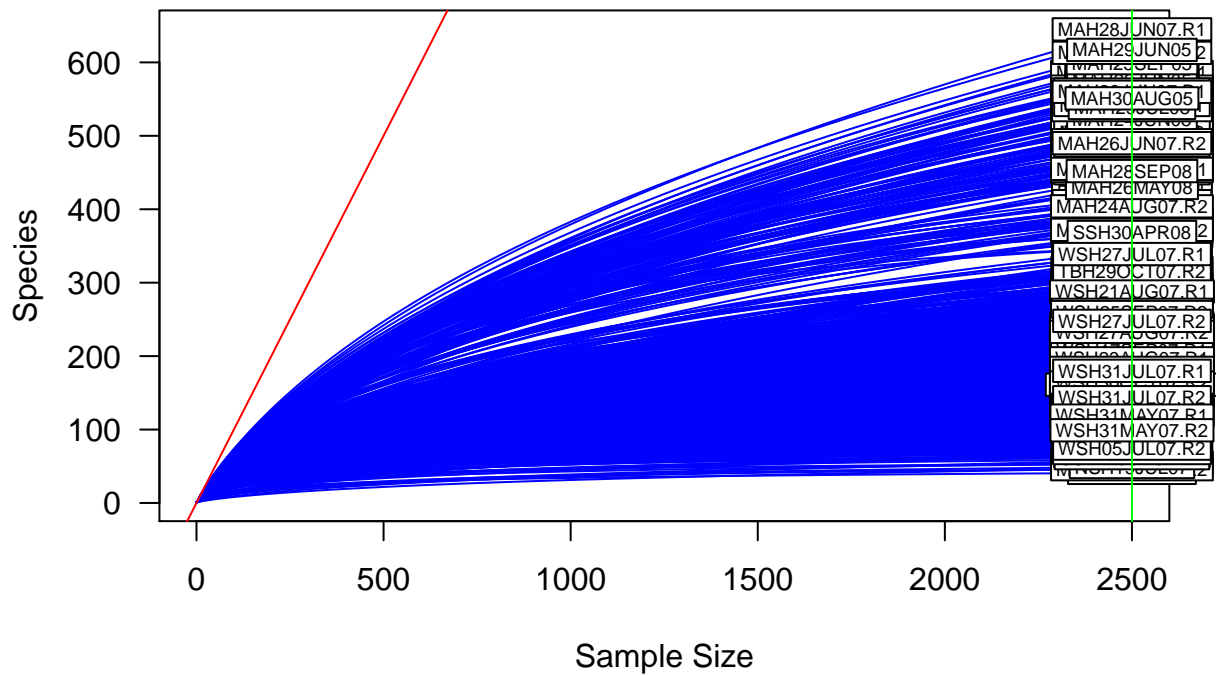

```r
#create otu_JUL07 of all bogs in first few days in JUL07
JUL07 <- c("CBE02JUL07.R1", "FBE02JUL07.R1", "HKE04JUL07.R1", "MAE04JUL07.R1", "NSE03JUL07.R1", "SSE03J
otu_JUL07 <- otu_table[grep(paste(JUL07, collapse = "|"), row.names(otu_table)),]
row.names(otu_JUL07) <- c("Crystal Bog", "Forestry Bog", "Hell's Kitchen", "Mary Lake", "North Sparkling

#looking at correlations between Crystal Bog and Mary Lake on the same date

#CBE01OCT07 and CBE27JUN07
(cor(otu_table[[24]], otu_table[[80]]))^2
```

```
## [1] 0.0005554524
```

```r
#CBE27JUN07 and MAE27JUN07
(cor(otu_table[[80]], otu_table[[444]]))^2
```

```
## [1] 0.0005536129
```

```r
#CBE01OCT07 and MAE01OCT07
(cor(otu_table[[24]], otu_table[[372]]))^2
```

```
## [1] 1.485061e-06
```

```r
#spatial diversity, create map
require("vegan")
#install.packages("sp")
require("sp")
```

```
## Loading required package: sp
```

```r
#install.packages("gstat")
require("gstat")
```

```
## Loading required package: gstat
```

```r
#install.packages("raster")
require("raster")
```

```
## Loading required package: raster
```

```
##
## Attaching package: 'raster'
```

```
## The following object is masked from 'package:OTUtable':
##
##     rotate
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```r
#install.packages("RgoogleMaps")
require("RgoogleMaps")
```

```
## Loading required package: RgoogleMaps
```

```r
#install.packages("maptools")
require("maptools")
```

```
## Loading required package: maptools
```

```
## Checking rgeos availability: TRUE
```

```r
#install.packages("rgdal")
require("rgdal")
```

```
## Loading required package: rgdal
```

```
## rgdal: version: 1.3-6, (SVN revision 773)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.2.3, released 2017/11/20
##  Path to GDAL shared files: C:/Users/emmim/R/win-library/3.5/rgdal/gdal
##  GDAL binary built with GEOS: TRUE
##  Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
##  Path to PROJ.4 shared files: C:/Users/emmim/R/win-library/3.5/rgdal/proj
##  Linking to sp version: 1.3-1
```

```r
#install.packages("simba")
require("simba")
```

```
## Loading required package: simba
```

```
## This is simba 0.3-5
```

```
##
## Attaching package: 'simba'
```

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

```r
#install.packages("gplots")
require("gplots")
```

```
## Loading required package: gplots
```

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

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

```r
#install.packages("rgeos")
require("rgeos")
```

```
## Loading required package: rgeos
```

```
## rgeos version: 0.4-2, (SVN revision 581)
##  GEOS runtime version: 3.6.1-CAPI-1.10.1
##  Linking to sp version: 1.3-1
##  Polygon checking: TRUE
```

```r
require("viridis")
```

```
## Loading required package: viridis
```

```
## Loading required package: viridisLite
```

```r
require("gplots")
require("RColorBrewer")
```

```
## Loading required package: RColorBrewer
```

```r
boglocate <- read.table("NTL_MO_bogs_location.txt", header = TRUE, sep = "\t")
boglocate <- boglocate[1:8, 1:4]
otu.names <- names(otu_table)

#removed bogsbys because it just removed the first OTU from out sbys matrix

sites <- boglocate[,2]
lats <-as.numeric(boglocate[,3])
lons <-as.numeric(boglocate[,4])
bogDataPoints <- data.frame(sites, lats, lons)

palette <- brewer.pal(8, "Spectral")
bogDataPoints$colors <- palette


# #Add API key
# newmap <- GetMap(center = c(lat = 46.12,lon = -89.75), zoom = 11, destfile = "BogMap.png", maptype =
#
#
# PlotOnStaticMap(newmap, zoom = 11, cex = 2, col = 'blue')
# PlotOnStaticMap(newmap, lats, lons, cex = 3, pch = 20, col = bogDataPoints$colors, add = TRUE)
#
```

```
# tblleg <- unique(bogDataPoints[,c("sites","colors")])
# row.names(tblleg) <- NULL
#
# legend("topright", legend = tblleg$sites, fill = tblleg$colors,bg = "white")

#generate ssad for multiple randomly chosen OTU
ssad <-function(otu_table, otu){
  ad <-c(2,2)
  ad <- otu_table[, otu]
  ad = as.vector(t(x = ad))
  ad = ad[ad >0]
  return(ad)
}

par(mfrow = c(2,2))
ct <- 0
while (ct < 4){
  otu <- sample(1:length(otu_table), 1)
  ad <- ssad(otu_table, otu)
  if (length(ad) > 10 & sum(ad >100)){
    ct <- ct + 1
    plot(density(ad), col = "red", xlab = "Site abundance", ylab = "Probability Density", main = otu.na
  }
}
```
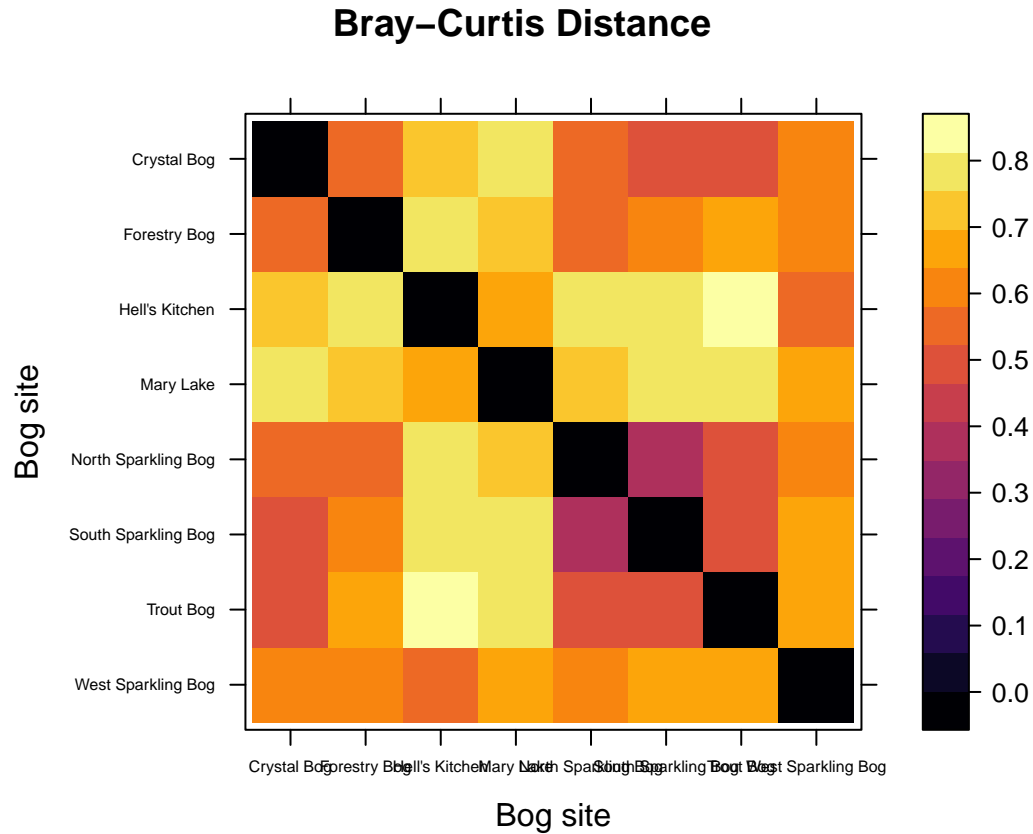

Otu0450

Otu0322

Otu0081

Otu1708
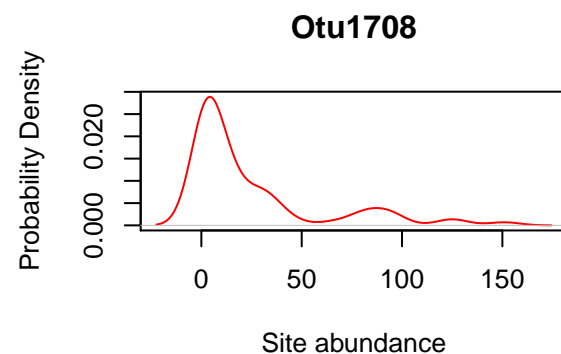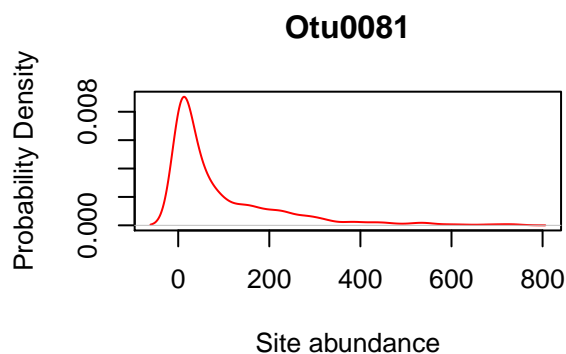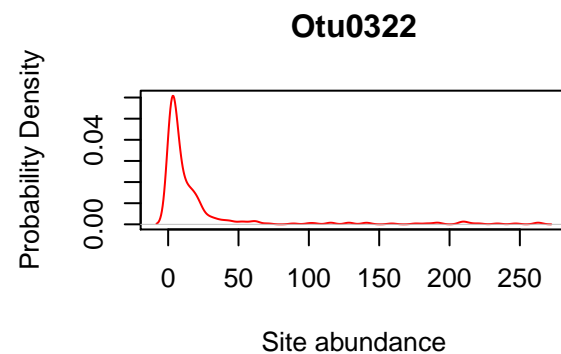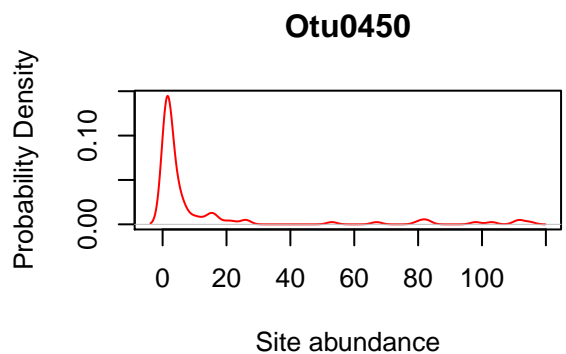
```
#beta diversity of bog communities
otu_JUL07.db <- vegdist(otu_JUL07, method = "bray", upper = TRUE, diag = TRUE)
order <- rev(attr(otu_JUL07.db, "Labels"))
levelplot(as.matrix(otu_JUL07.db)[, order], aspect = "iso", col.regions = inferno, xlab = "Bog site", yl
```
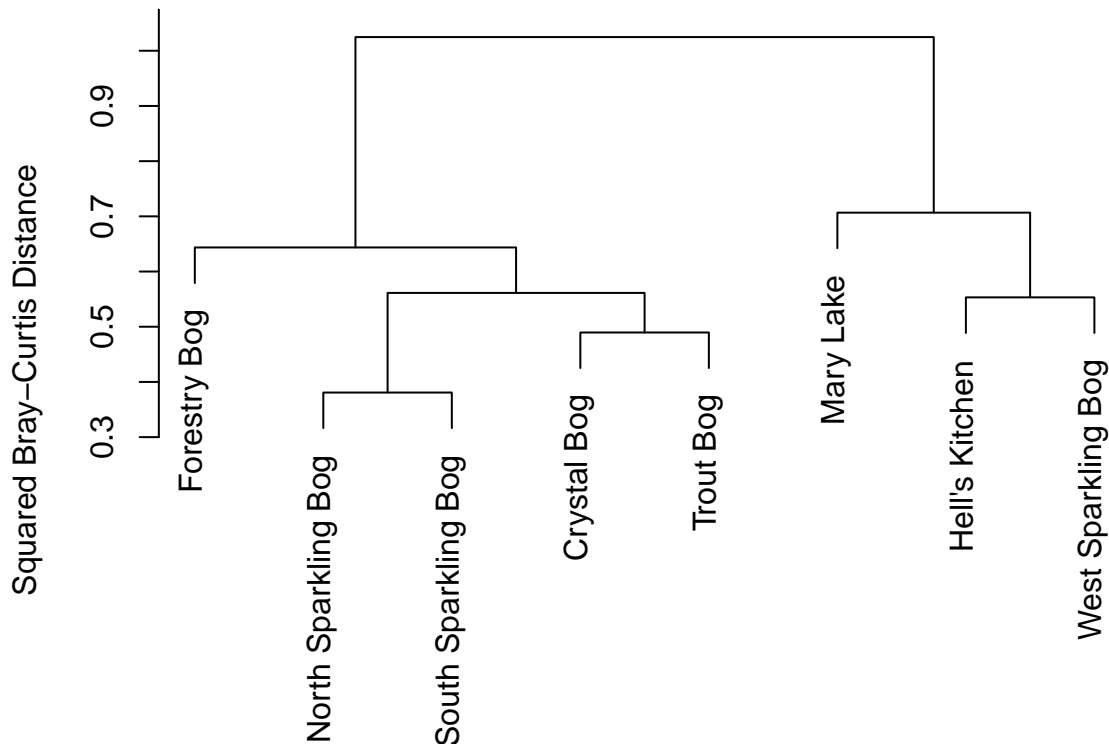

Bray−Curtis Distance

```
JUL07.ward <- hclust(otu_JUL07.db, method = "ward.D2")

par(mar = c(1,5,2,2) + 0.1)
plot(JUL07.ward, main = "Bog Sites: Ward's Clustering", ylab = "Squared Bray-Curtis Distance")
```

Bog Sites: Ward's Clustering


```
bog.pcoa <- cmdscale(otu_JUL07.db, eig = TRUE, k= 3)
explainvar1 <- round(bog.pcoa$eig[1] / sum(bog.pcoa$eig), 3) * 100
explainvar2 <- round(bog.pcoa$eig[2] / sum(bog.pcoa$eig), 3) * 100
explainvar3 <- round(bog.pcoa$eig[3] / sum(bog.pcoa$eig), 3) * 100

par(mar = c(5,5,1,2) + 0.1)

plot(bog.pcoa$points[,1], bog.pcoa$points[,2],
     xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
     ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
     pch = 16, cex = 3.0, type = "n", cex.lab = 1.2, cex.axis = 1.2, axes = FALSE, xlim = c(-0.4, 0.5))

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

points(bog.pcoa$points[,1], bog.pcoa$points[,2], pch = 19, cex = 3, bg = "gray", col = "gray")

text(bog.pcoa$points[,1], bog.pcoa$points[,2], labels = row.names(bog.pcoa$points))
```
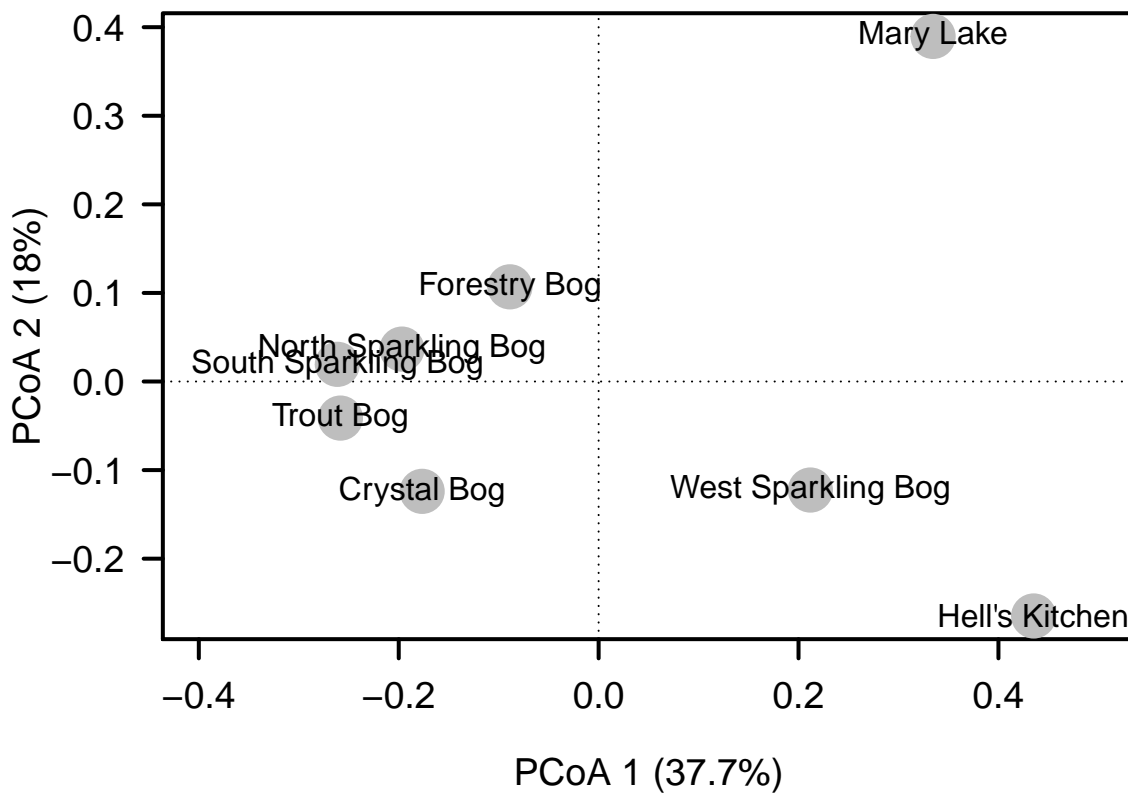
```
#Set up metadata for JUL07 sites
data(metadata)
rows <- c(4971,4977,5004,5025,4982,4988,4962,4998)
metadata_JUL07 <- metadata[rows, -1]
row.names(metadata_JUL07) <- c("Crystal Bog", "Forestry Bog","Hell's Kitchen","Mary Lake","North Sparkli
num = 1
rows <- c(4976,4981,5024,5048,4987,4997,4970,5003)
while(num - 1 < nrow(metadata_JUL07)){
  depth_current <- metadata$Depth[rows[num]]
  metadata_JUL07$Depth[num] <- depth_current
  num <- num + 1
}
metadata_JUL07 <- metadata_JUL07[,c("Depth", "DO", "Temperature")]
metadata_JUL07 <- cbind(metadata_JUL07, lats)
metadata_JUL07 <- cbind(metadata_JUL07, lons)
bogcca <- cca(otu_JUL07, metadata_JUL07)

#Mantel Test
jul07.dist <- vegdist(otu_JUL07, method = "bray", upper = TRUE, diag = TRUE)
jul07.env.dist <- vegdist(scale(metadata_JUL07), method = "euclid")

mantel(jul07.dist, jul07.env.dist)

##
## Mantel statistic based on Pearson's product-moment correlation
##
```

```
## Call:
## mantel(xdis = jul07.dist, ydis = jul07.env.dist)
##
## Mantel statistic r: 0.6039
##       Significance: 0.029
##
## Upper quantiles of permutations (null model):
##    90%   95% 97.5%    99%
## 0.359 0.512 0.612 0.686
## Permutation: free
## Number of permutations: 999
```

```r
print(jul07.dist)
```

```
##                    Crystal Bog Forestry Bog Hell's Kitchen Mary Lake
## Crystal Bog             0.0000       0.5304         0.7348    0.7736
## Forestry Bog            0.5304       0.0000         0.7880    0.7216
## Hell's Kitchen          0.7348       0.7880         0.0000    0.6952
## Mary Lake               0.7736       0.7216         0.6952    0.0000
## North Sparkling Bog     0.5316       0.5684         0.7664    0.7252
## South Sparkling Bog     0.4884       0.5892         0.8012    0.7604
## Trout Bog               0.4896       0.6588         0.8136    0.7828
## West Sparkling Bog      0.5972       0.6120         0.5532    0.6472
##                    North Sparkling Bog South Sparkling Bog Trout Bog
## Crystal Bog                     0.5316              0.4884    0.4896
## Forestry Bog                    0.5684              0.5892    0.6588
## Hell's Kitchen                  0.7664              0.8012    0.8136
## Mary Lake                       0.7252              0.7604    0.7828
## North Sparkling Bog             0.0000              0.3808    0.5020
## South Sparkling Bog             0.3808              0.0000    0.4916
## Trout Bog                       0.5020              0.4916    0.0000
## West Sparkling Bog              0.6088              0.6584    0.6796
##                    West Sparkling Bog
## Crystal Bog                    0.5972
## Forestry Bog                   0.6120
## Hell's Kitchen                 0.5532
## Mary Lake                      0.6472
## North Sparkling Bog            0.6088
## South Sparkling Bog            0.6584
## Trout Bog                      0.6796
## West Sparkling Bog             0.0000
```

```r
#dbRDA from metadata
metadata_JUL07 <- as.matrix(metadata_JUL07)

#dbRDA from metadata only lats + lons
bog.dbrda <- dbrda(otu_JUL07~ lats + lons + Temperature, as.data.frame(metadata_JUL07), dist = "bray")

#bog.dbrda <- dbrda(otu_JUL07~ ., as.data.frame(metadata_JUL07), dist = "bray")

require("psych")
```

```
## Loading required package: psych
```

```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:simba':
##
##     sim

## The following object is masked from 'package:OTUtable':
##
##     shannon

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```r
psych::corr.test(metadata_JUL07)
```

```
## Call:psych::corr.test(x = metadata_JUL07)
## Correlation matrix
##             Depth    DO Temperature  lats  lons
## Depth        1.00  0.24        0.94  0.94 -0.77
## DO           0.24  1.00        0.32  0.30 -0.04
## Temperature  0.94  0.32        1.00  0.95 -0.78
## lats         0.94  0.30        0.95  1.00 -0.74
## lons        -0.77 -0.04       -0.78 -0.74  1.00
## Sample Size
## [1] 8
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##             Depth   DO Temperature lats lons
## Depth        0.00 1.00        0.00 0.00 0.15
## DO           0.57 0.00        1.00 1.00 1.00
## Temperature  0.00 0.45        0.00 0.00 0.15
## lats         0.00 0.46        0.00 0.00 0.17
## lons         0.02 0.93        0.02 0.03 0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```r
# bog.dbrda.mod0 <- dbrda(jul07.dist ~ 1, as.data.frame(metadata_JUL07))
# bog.dbrda.mod1 <- dbrda(jul07.dist ~ ., as.data.frame(metadata_JUL07))
# bog.dbrda <- ordiR2step(bog.dbrda.mod0, bog.dbrda.mod1, perm.max = 200)

bog.dbrda$call
```
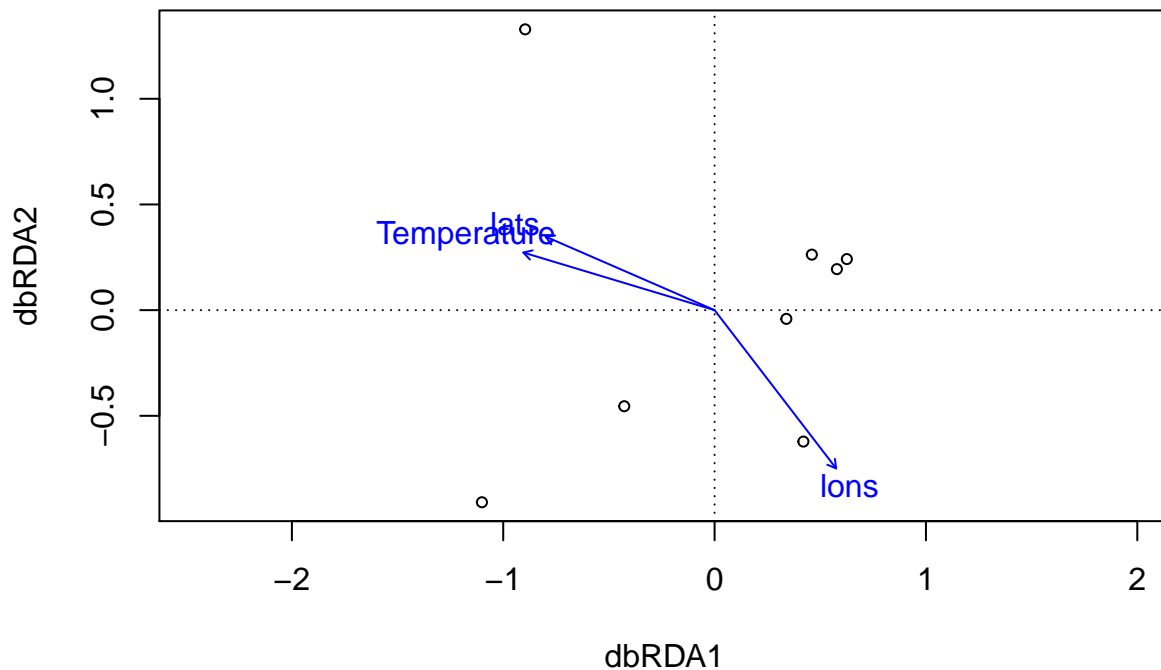
```
## dbrda(formula = otu_JUL07 ~ lats + lons + Temperature, data = as.data.frame(metadata_JUL07),
##     distance = "bray")
```

```r
bog.dbrda$anova
```

```
## NULL
```

```r
ordiplot(bog.dbrda)
```

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

```
##
## Permutation test for dbrda under reduced model
##
## Permutation: free
## Number of permutations: 999
##
## Model: dbrda(formula = otu_JUL07 ~ lats + lons + Temperature, data
## = as.data.frame(metadata_JUL07), distance = "bray")
## Permutation test for all constrained eigenvalues
##          Df Inertia       F Pr(>F)
## Model     3 0.89129 2.0022  0.005 **
## Residual  4 0.59354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
envfit(bog.dbrda, metadata_JUL07, perm = 999)
```

```
##
## ***VECTORS
##
##                 dbRDA1    dbRDA2     r2 Pr(>r)
## Depth         -0.90615   0.42296 0.7616  0.035 *
## DO            -0.91788  -0.39686 0.1096  0.736
## Temperature   -0.94725   0.32049 0.9467  0.001 ***
## lats          -0.90444   0.42659 0.8123  0.034 *
```

```
## lons           0.60744 -0.79437 0.9697   0.004 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 999
```

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

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

plot(scores(bog.dbrda, display = "wa"), xlim = c(-1.3, 1.5), ylim = c(-1.5, 2.7), xlab = paste("dbRDA 1

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

points(scores(bog.dbrda, display = "wa"), pch = 19, cex = 3, bg = "gray", col = "gray")
text(scores(bog.dbrda, display = "wa"), labels = row.names(scores(bog.dbrda, display = "wa")))

vectors <- scores(bog.dbrda, display = "bp")

arrows(0, 0, vectors[,1], vectors[,2], lwd = 2, lty = 1, length = 0.2, col = "red")
text(vectors[,1], vectors[,2], pos = 3, labels = row.names(vectors))
axis(side=3, lwd.ticks = 2, cex.axis=1.2, las = 1, col = "red", lwd = 2.2, at = pretty(range(vectors[,1]
axis(side=4, lwd.ticks = 2, cex.axis=1.2, las = 1, col = "red", lwd = 2.2, at = pretty(range(vectors[,2]
```
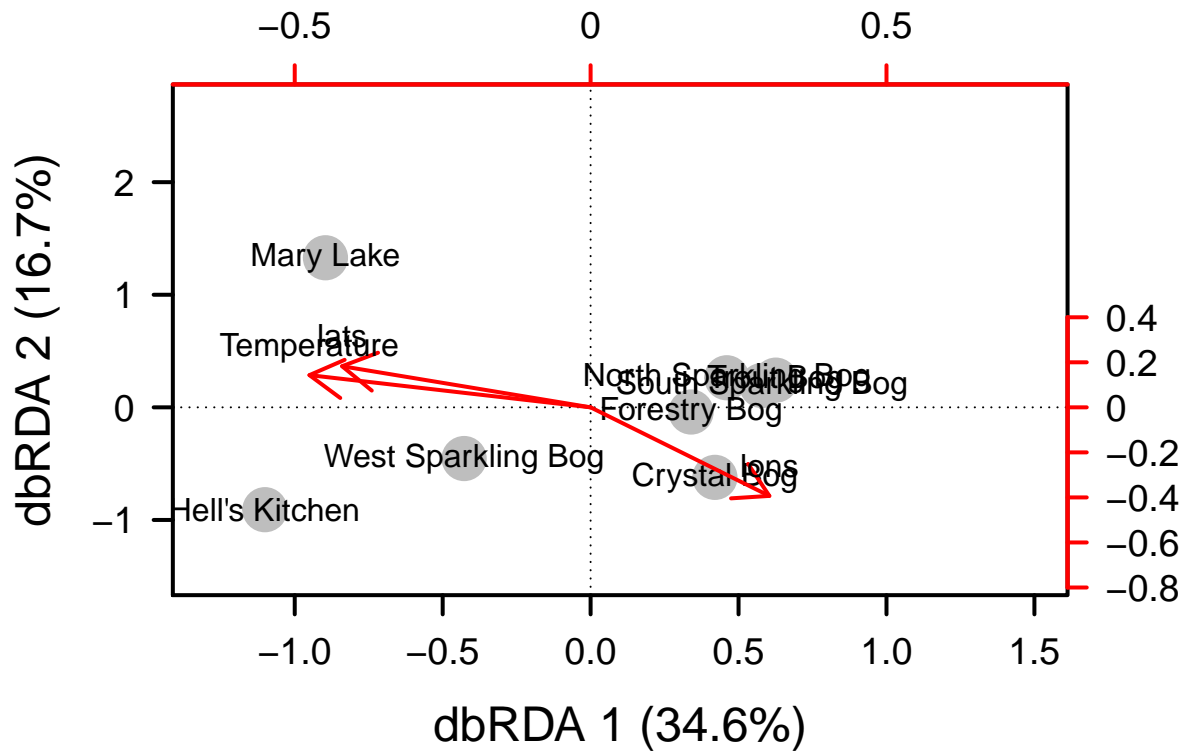
```
#Relativized PCoA
bog.pcoa <- cmdscale(jul07.dist, eig = TRUE, k= 3)
explainvar1 <- round(bog.pcoa$eig[1] / sum(bog.pcoa$eig), 3) * 100
explainvar2 <- round(bog.pcoa$eig[2] / sum(bog.pcoa$eig), 3) * 100
explainvar3 <- round(bog.pcoa$eig[3] / sum(bog.pcoa$eig), 3) * 100

par(mar = c(5,5,1,2) + 0.1)

plot(bog.pcoa$points[,1], bog.pcoa$points[,2], ylim = c(-0.2, 0.7),
    xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
    ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
    pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes = FALSE)

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

points(bog.pcoa$points[,1], bog.pcoa$points[,2], pch = 19, cex = 3, bg = "gray", col = "gray")

text(bog.pcoa$points[,1], bog.pcoa$points[,2], labels = row.names(bog.pcoa$points))
```
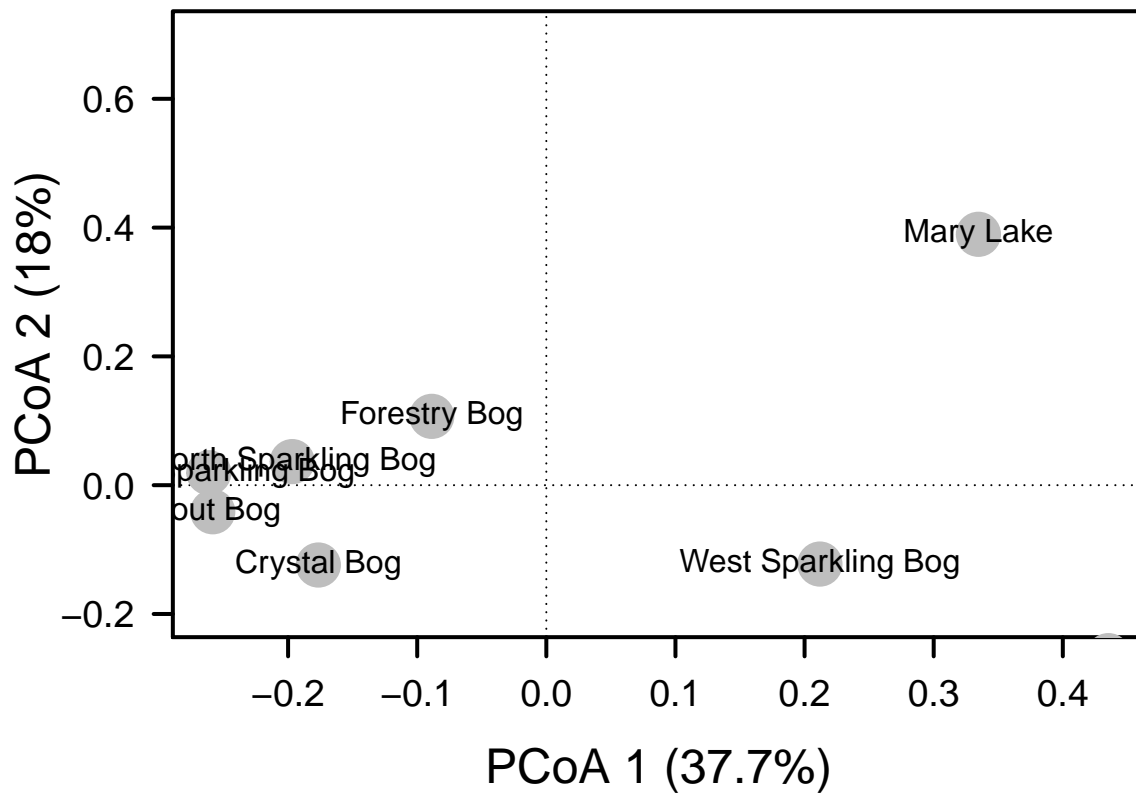
```
bogREL <- otu_JUL07
for (i in 1:nrow(otu_JUL07)){
  bogREL[i,] = otu_JUL07[i,]/sum(otu_JUL07[i,])
}
```

```
#Relationship between environmental distance, geographic distance and community distance
#distance decay relationship
#install.packages("simba")
require("simba")

bog_loc <- read.table("NTL_MO_bogs_location.txt", header = TRUE, sep = "\t")
otu_JUL07.db <- vegdist(otu_JUL07, method = "bray", upper = TRUE, diag = TRUE)
#sites <- boglocate[,2]


lats <-as.numeric(boglocate[,3])
lons <-as.numeric(boglocate[,4])
#bogDataPoints <- data.frame(sites, lats, lons)
coord.dist <- dist(as.matrix(lats, lons))
x1 <- as.numeric(metadata_JUL07[,3])
env.dist <- vegdist(x1, "euclidean")
comm.dist.ls <- liste(otu_JUL07.db, entry = "comm")
env.dist.ls <- liste(env.dist, entry = "env")
coord.dist.ls <- liste(coord.dist, entry = "dist")
df <- data.frame(coord.dist.ls, env.dist.ls[,3], comm.dist.ls[,3])
names(df)[4:5] <- c("env", "struc")
attach(df)
```

```
par(mfrow=c(1, 3), pty = "s")
plot(env, struc, xlab = "Environmental Distance", ylab = "Bray-Curtis Dissimilarity",
     main = "Environmental Distance", col = 'SteelBlue')

OLS <- lm(struc ~ env)
OLS
```

```
##
## Call:
## lm(formula = struc ~ env)
##
## Coefficients:
## (Intercept)          env
##      0.54039       0.07119
```

```
abline(OLS, col = "red4")

plot(dist, struc, xlab = "Geographic Distance", ylab = "Bray-Curtis Dissimilarity",
     main = "Community\nComposition\n", col = 'darkorchid4')

OLS <- lm(struc ~ dist)
OLS
```

```
##
## Call:
## lm(formula = struc ~ dist)
##
## Coefficients:
## (Intercept)         dist
##       0.5535       0.8739
```
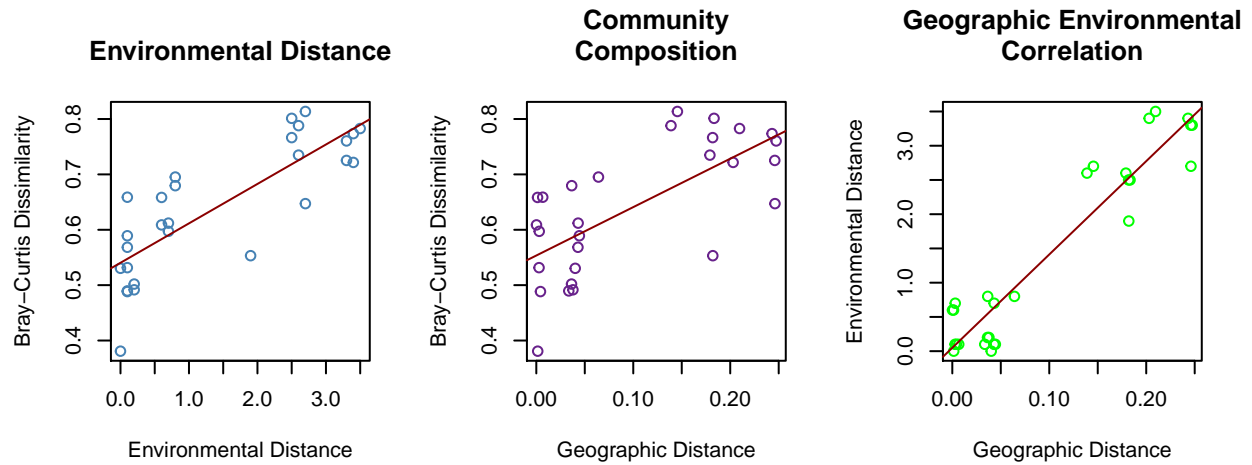
```
abline(OLS, col = "red4")

plot(dist, env, xlab = "Geographic Distance", ylab = "Environmental Distance", main = "Geographic Enviro

OLS <- lm(env ~ dist)
OLS
```

```
##
## Call:
## lm(formula = env ~ dist)
##
## Coefficients:
## (Intercept)         dist
##       0.0498      13.6236
```

```
abline(OLS, col = "red4")
```

Environmental Distance

Community Composition

Geographic Environmental Correlation


```
diffslope(env, struc, dist, struc)
```

```
##
## Is difference in slope significant?
## Significance is based on 1000 permutations
##
## Call:
## diffslope(x1 = env, y1 = struc, x2 = dist, y2 = struc)
##
## Difference in Slope: -0.8027
## Significance: 0.001
##
## Empirical upper confidence limits of r:
##     90%     95%   97.5%     99%
## 0.0208  0.0265  0.0313  0.0377
```