8. Worksheet: Among Site (Beta) Diversity – Part 1

Richard Hull; Z620: Quantitative Biodiversity, Indiana University

10 April, 2021

OVERVIEW

In this worksheet, we move beyond the investigation of within-site α-diversity. We will explore β-diversity, which is defined as the diversity that occurs among sites. This requires that we examine the compositional similarity of assemblages that vary in space or time.

After completing this exercise you will know how to:

1. formally quantify β-diversity
2. visualize β-diversity with heatmaps, cluster analysis, and ordination
3. 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. Ths 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_1_Worksheet.Rmd**) with all code blocks filled out and questions answered) and the PDF output of `Knitr` (**8.BetaDiversity_1_Worksheet.pdf**).

The completed exercise is due on **Friday, April 16th, 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] "C:/Users/Rich Hull/GitHub/QB2021_Hull/2.Worksheets/8.BetaDiversity"
```

```
setwd("C:/Users/Rich Hull/GitHub/QB2021_Hull/2.Worksheets")
```

Install and load necessary packages

```
# Load packages
package.list <- c('vegan', 'ade4', 'viridis', 'gplots', 'BiodiversityR', 'indicspecies')
for (package in package.list){
  if (!require(package, character.only = TRUE, quietly = TRUE)) {
    install.packages(package)
    library(package, character.only = TRUE)
  }
}
```

```
## This is vegan 2.5-7

## Warning: package 'ade4' was built under R version 4.0.5

## Warning: package 'viridis' was built under R version 4.0.5

## Warning: package 'gplots' was built under R version 4.0.5

##
## Attaching package: 'gplots'

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

## Warning: package 'BiodiversityR' was built under R version 4.0.5

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

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

## Warning: package 'indicspecies' was built under R version 4.0.5
```

2) LOADING DATA

Load dataset

In the R code chunk below, do the following:

1. load the `doubs` dataset from the `ade4` package, and
2. explore the structure of the dataset.

```
# note, pleae do not print the dataset when submitting
# load data
data(doubs)
# explore structure
str(doubs, max.level = 1)
```

```
## List of 4
##  $ env    :'data.frame': 30 obs. of  11 variables:
##  $ fish   :'data.frame': 30 obs. of  27 variables:
##  $ xy     :'data.frame': 30 obs. of  2 variables:
##  $ species:'data.frame': 27 obs. of  4 variables:
```

```
head(doubs$env)
```

```
##   dfs alt   slo flo pH har pho nit amm oxy bdo
## 1   3 934 6.176  84 79  45   1  20   0 122  27
## 2  22 932 3.434 100 80  40   2  20  10 103  19
## 3 102 914 3.638 180 83  52   5  22   5 105  35
## 4 185 854 3.497 253 80  72  10  21   0 110  13
## 5 215 849 3.178 264 81  84  38  52  20  80  62
## 6 324 846 3.497 286 79  60  20  15   0 102  53
```

Question 1: Describe some of the attributes of the `doubs` dataset.

a. How many objects are in `doubs`?
b. How many fish species are there in the `doubs` dataset?
c. How many sites are in the `doubs` dataset?

Answer 1a: 4 items in list ***Answer 1b***: 27 species ***Answer 1c***: 30 sites

Visualizing the Doubs River Dataset

Question 2: Answer the following questions based on the spatial patterns of richness (i.e., α-diversity) and Brown Trout (*Salmo trutta*) abundance in the Doubs River.

a. How does fish richness vary along the sampled reach of the Doubs River?
b. How does Brown Trout (*Salmo trutta*) abundance vary along the sampled reach of the Doubs River?
c. What do these patterns say about the limitations of using richness when examining patterns of biodiversity?

Answer 2a: There are two large concentrations of high species richness, while the rest of the river has a low amount of species richness. ***Answer 2b***: Brown Trout abundance is highest in areas of the river that are low in species richness. ***Answer 2c***: Richness does not include abundance data, and therefore does not supply the entire picture. In this case, it appears there are sites that are high in species richness but generally have lower amounts of individuals per species, while certain other sites that have lower species richness are dominated by many individuals of the same few/one species.

3) QUANTIFYING BETA-DIVERSITY

In the R code chunk below, do the following:

1. write a function (`beta.w()`) to calculate Whittaker's β-diversity (i.e., β_w) that accepts a site-by-species matrix with optional arguments to specify pairwise turnover between two sites, and
2. use this function to analyze various aspects of β-diversity in the Doubs River.

```
{r}{r eval = FALSE, echo = FALSE} # Define Whittaker's beta diversity beta.w <- function(site.by.species
= ""){   SbyS.pa <- decostand(site.by.species, method = "pa")   S <- ncol(SbyS.pa[,which(colSums(SbyS.pa
> 0)])   a.bar <- mean(specnumber(SbyS.pa))   b.w <- round(S/a.bar, 3) } # Modify to
calculate Whittaker's beta diversity for turnover beta.w <- function(site.by.species =
"", sitenum1 = "", sitenum2 = "", pairwise = FALSE){   if (pairwise == TRUE){     if
(sitenum1 == "" | sitenum2 == ""){       print("Error: please specify sites to compare")
return(NA)}   site1 = site.by.species[sitenum1,]   site2 = site.by.species[sitenum2,]
site1 = subset(site1, select = site1 > 0)   site2 = subset(site2, select = site2 > 0)
gamma = union(colnames(site1), colnames(site2))   a = length(gamma)   a.bar = mean(c(specnumber(site1),
specnumber(site2)))   b.w = round(s/a.bar - 1, 3)   return(b.w)   }   else{     SbyS.pa
<- decostand(site.by.species, method = "pa")     S <- ncol(SbyS.pa[,which(colSums(SbyS.pa)
> 0)])     a.bar <- mean(specnumber(SbyS.pa))     b.w <- round(S/a.bar, 3)   } } # Calculate
Whittaker's beta diversity turnover in fish dataset wbturn <- beta.w(doubs$fish) # Calculate
Whittaker's beta diversity for site 1 and and 2, and then for 1 and 10 wbturn12 <- beta.w(doubs$fish,
sitenum1 = "1", sitenum2 = "2", pairwise = TRUE) wbturn110 <- beta.w(doubs$fish, sitenum1
= "1", sitenum2 = "10", pairwise = TRUE)
```

Question 3: Using your `beta.w()` function above, answer the following questions:

a. Describe how local richness (α) and turnover (β) contribute to regional (γ) fish diversity in the Doubs.
b. Is the fish assemblage at site 1 more similar to the one at site 2 or site 10?
c. Using your understanding of the equation $\beta_w = \gamma/\alpha$, how would your interpretation of β change if we instead defined beta additively (i.e., $\beta = \gamma - \alpha$)?

Answer 3a: Gamma diversity is how many times more diverse the regional species pool is than the average richness at each site. Therefore, the regional species pool is 2.16 times more diverse than the average species richness of each inidivual site. ***Answer 3b***: The fish assemblage at site 1 is more similar to the fish assemblage at site 2 than at site 10. ***Answer 3c***: Beta would no longer be proportional to either gamma or alpha diversity.

The Resemblance Matrix

In order to quantify β-diversity for more than two samples, we need to introduce a new primary ecological data structure: the **Resemblance Matrix**.

Question 4: How do incidence- and abundance-based metrics differ in their treatment of rare species?

Answer 4: Incidence based metrics place more emphasis on shared species and are therefore more likely to be influenced by rare species, which are less likely to be shared between sites. Abundance based metrics place more ephasis on abundant species, which are more likely to be shared between sites. Therefore, rare species are less likely to influence abundance based metrics than incidence based metrics.

In the R code chunk below, do the following:

1. make a new object, `fish`, containing the fish abundance data for the Doubs River,
2. remove any sites where no fish were observed (i.e., rows with sum of zero),
3. construct a resemblance matrix based on Sørensen's Similarity ("fish.ds"), and
4. construct a resemblance matrix based on Bray-Curtis Distance ("fish.db").

```
# make object fish and remove sites with 0
fish <- doubs$fish
fish <- fish[-8, ]
# calculate Jaccard
fish.dj <- vegdist(fish, method = "jaccard", binary = TRUE)
# calculate Bray-Curtis
fish.db <- vegdist(fish, method = "bray")
# calculate Sorensen
fish.ds <- vegdist(fish, method = "bray", binary = TRUE)
```

Question 5: Using the distance matrices from above, answer the following questions:

a. Does the resemblance matrix (`fish.db`) represent similarity or dissimilarity? What information in the resemblance matrix led you to arrive at your answer?
b. Compare the resemblance matrices (`fish.db` or `fish.ds`) you just created. How does the choice of the Sørensen or Bray-Curtis distance influence your interpretation of site (dis)similarity?

Answer 5a: Dissimilarity. Sites 1 and 9 have a value of 1, but share no common species.
Answer 5b: Sorensen's method is more conservative and calculates lower dissimilarity values than the Bray-Curtis method.

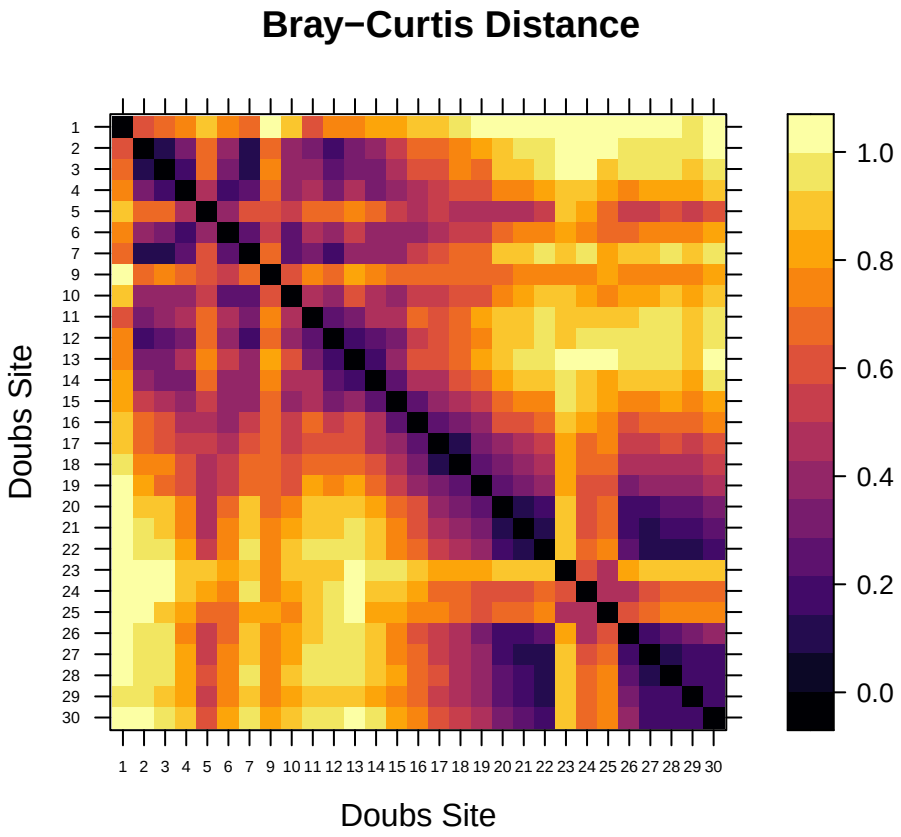
4) VISUALIZING BETA-DIVERSITY

A. Heatmaps

In the R code chunk below, do the following:

1. define a color palette,
2. define the order of sites in the Doubs River, and
3. use the `levelplot()` function to create a heatmap of fish abundances in the Doubs River.

```
# define order of sites
order <- rev(attr(fish.db, "Labels"))
# plot heatmap
levelplot(as.matrix(fish.db)[, order], aspect = "iso", col.regions = inferno, xlab = "Doubs Site", ylab
```

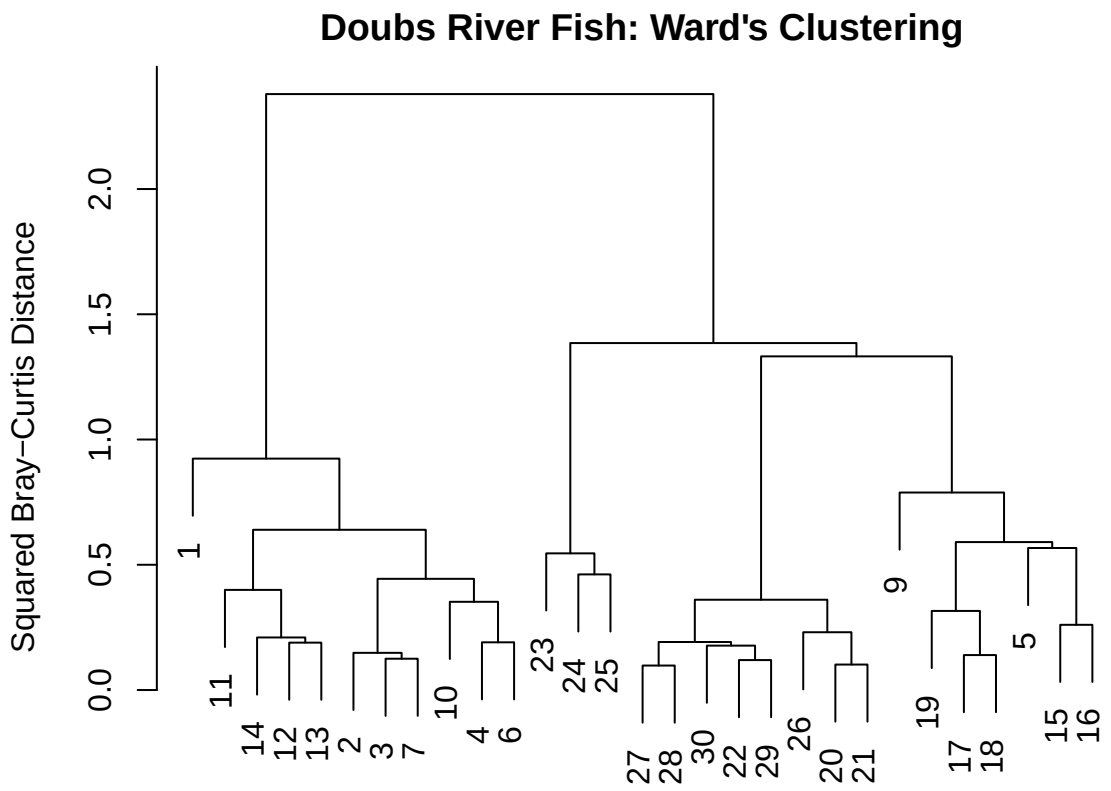


B. Cluster Analysis

In the R code chunk below, do the following:

1. perform a cluster analysis using Ward's Clustering, and
2. plot your cluster analysis (use either `hclust` or `heatmap.2`).

```
# perform cluster analysis
fish.ward <- hclust(fish.db, method = "ward.D2")
# plot cluster
par(mar = c(1,5,2,2) + 0.1)
plot(fish.ward, main = "Doubs River Fish: Ward's Clustering", ylab = "Squared Bray-Curtis Distance")
```



Question 6: Based on cluster analyses and the introductory plots that we generated after loading the data, develop an ecological hypothesis for fish diversity the `doubs` data set?

> ***Answer 6***: There are two main sets of sites which both contain high species richness but differ in their species composition. Other sites are predominatey comprised of a few dominant species

C. Ordination

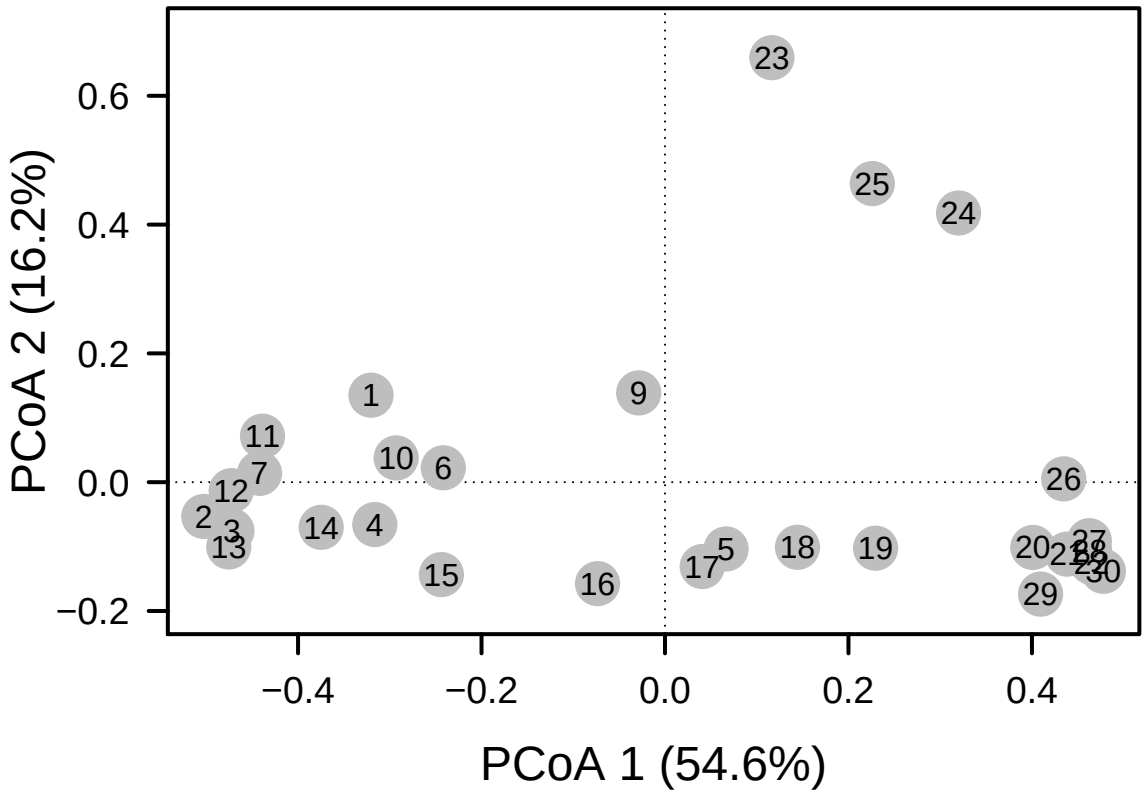
Principal Coordinates Analysis (PCoA)

In the R code chunk below, do the following:

1. perform a Principal Coordinates Analysis to visualize beta-diversity
2. calculate the variation explained by the first three axes in your ordination
3. plot the PCoA ordination,
4. label the sites as points using the Doubs River site number, and
5. identify influential species and add species coordinates to PCoA plot.

```
# conduct PCoA
fish.pcoa <- cmdscale(fish.db, eig = TRUE, k = 3)
# calculate variation explained by first three axes
explainvar1 <- round(fish.pcoa$eig[1] / sum(fish.pcoa$eig), 3) * 100
explainvar2 <- round(fish.pcoa$eig[2] / sum(fish.pcoa$eig), 3) * 100
explainvar3 <- round(fish.pcoa$eig[3] / sum(fish.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)
```

```r
# plot PCoA ordination
par(mar = c(5, 5, 1, 2) + 0.1)
plot(fish.pcoa$points[ ,1], fish.pcoa$points[ ,2], ylim = c(-0.2, 0.7), xlab = paste("PCoA 1 (", explai
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(fish.pcoa$points[ ,1], fish.pcoa$points[ ,2],
       pch = 19, cex = 3, bg = "gray", col = "gray")
text(fish.pcoa$points[ ,1], fish.pcoa$points[ ,2], labels = row.names(fish.pcoa$points))
```



In the R code chunk below, do the following:

1. identify influential species based on correlations along each PCoA axis (use a cutoff of 0.70), and
2. use a permutation test (999 permutations) to test the correlations of each species along each axis.

```r
# calculate relative abundance of each sp at each site
fishREL <- fish
for(i in 1:nrow(fish)){
  fishREL[i, ] = fish [i, ] / sum(fish[i, ])
}
# calculate and add species scores
fish.pcoa <- add.spec.scores(fish.pcoa, fishREL, method = "pcoa.scores")
# cutoff
spe.corr <- add.spec.scores(fish.pcoa, fishREL, method = "cor.scores")$cproj
```

```
corrcut <- 0.7
imp.spp <- spe.corr[abs(spe.corr[, 1]) >= corrcut | abs(spe.corr[, 2]) >= corrcut, ]
# permutate
fit <- envfit(fish.pcoa, fishREL, perm = 999)
```

Question 7: Address the following questions about the ordination results of the `doubs` data set:

a. Describe the grouping of sites in the Doubs River based on fish community composition.
b. Generate a hypothesis about which fish species are potential indicators of river quality.

> ***Answer 7a***: Many sites have a diverse species richness, while a number of sites are dominated by Satr, Php, Neba, Alal, Lece, and Ruru ***Answer 7b***: Satr, Php, Neba, Alal, Lece, and Ruru are all species that are indicators of low river quality because most species cannot naturally withstand low water quality, leading to a number of sites with low water quality being dominated by many individuals of the same few species and sites in higher water quality maintaining a high species diversity with low numbers of each species.

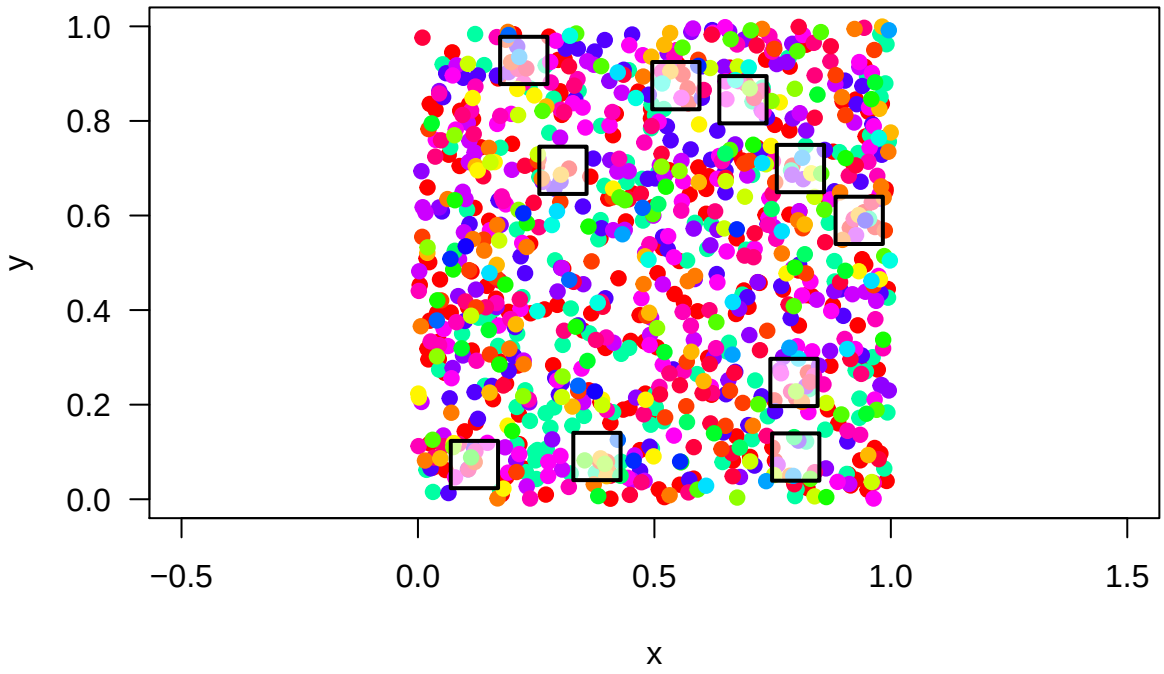
SYNTHESIS

Using the `mobsim` package from the DataWrangling module last week, 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:

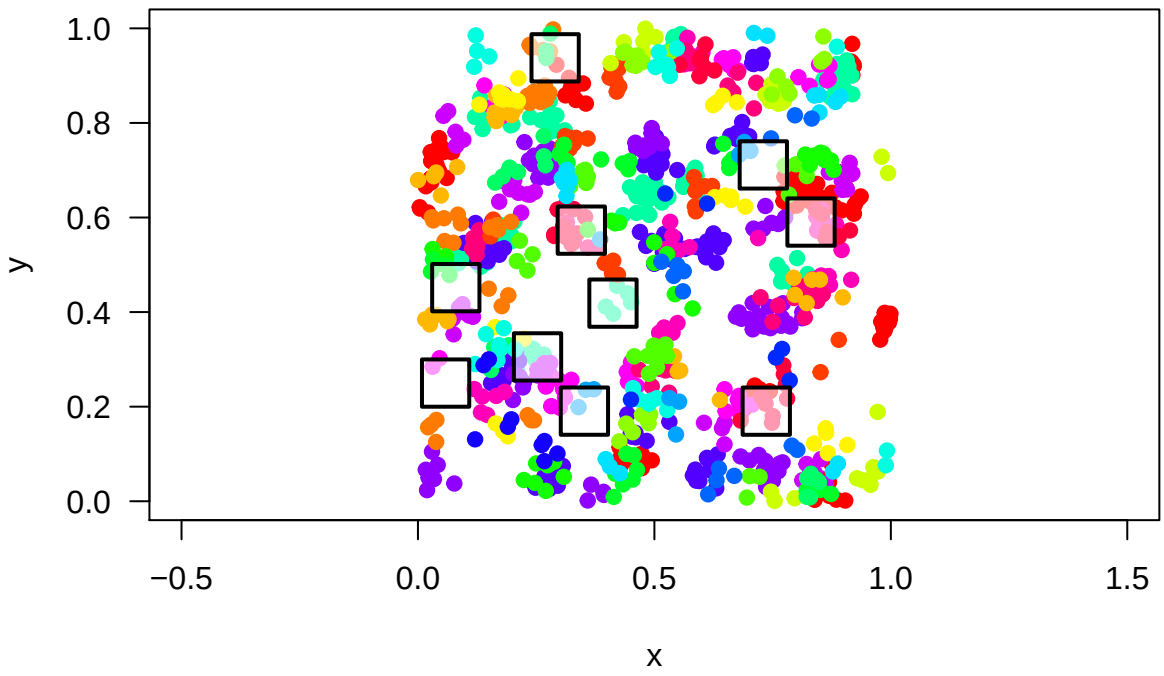
```
# simulate random spatial distribution comm
require(mobsim)
```

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

```
com1 <- sim_poisson_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))
# simulate patchy spatial distribution
com2 <- sim_thomas_community(s_pool = 25, n_sim = 1000, sad_type = "lnorm",
        sad_coef = list("meanlog" = 2, "sdlog" = 1))
# divide both communities into 10 quadrats of the same size
# Lay down sampling quadrats on the community
comm_mat1 <- sample_quadrats(com1, n_quadrats = 10, quadrat_area = 0.01,
             method = "random", avoid_overlap = T)
```




```
# Rename sampled areas as quadrats
quads <- c("quad1", "quad2", "quad3", "quad4", "quad5", "quad6", "quad7",
           "quad8", "quad9", "quad10")
row.names(comm_mat1$xy_dat) <- quads
row.names(comm_mat1$spec_dat) <- quads
# Lay down sampling quadrats on the community
comm_mat2 <- sample_quadrats(com2, n_quadrats = 10, quadrat_area = 0.01,
             method = "random", avoid_overlap = T)
```



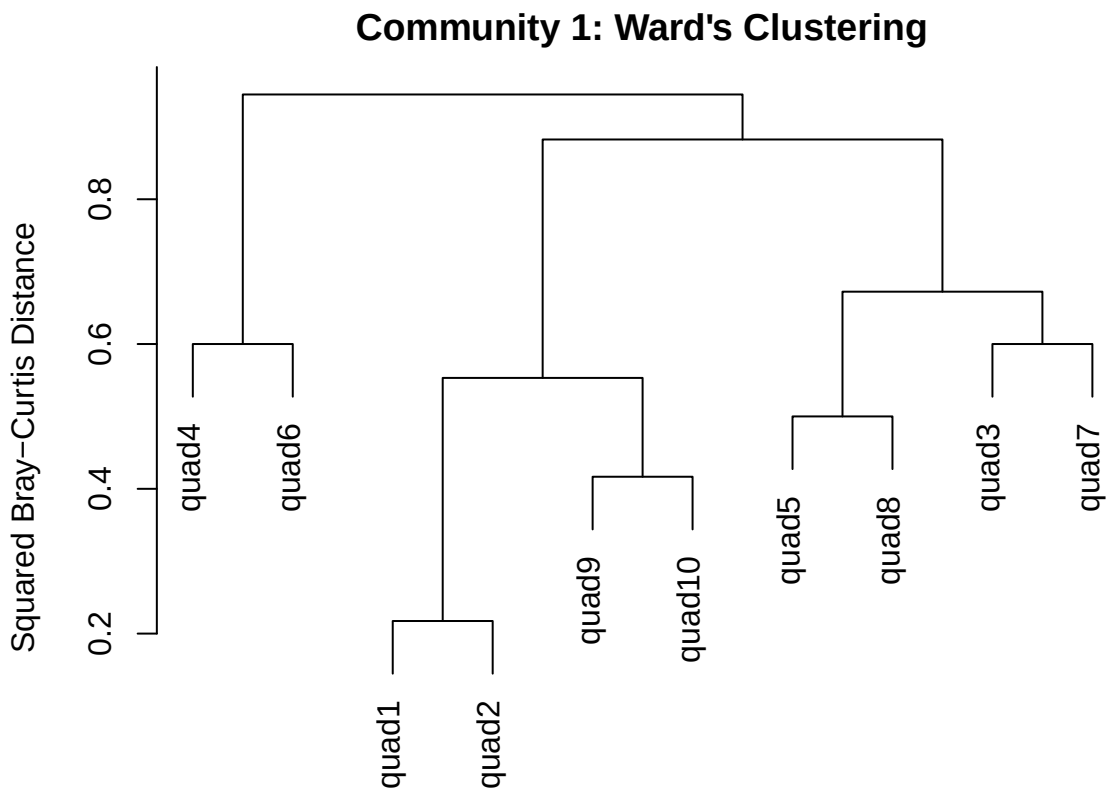

```
# Rename sampled areas as quadrats
quads <- c("quad1", "quad2", "quad3", "quad4", "quad5", "quad6", "quad7",
           "quad8", "quad9", "quad10")
row.names(comm_mat2$xy_dat) <- quads
row.names(comm_mat2$spec_dat) <- quads
```

1) Compare the average pairwise similarity among subsamples in site 1 (random spatial distribution) to the average pairswise similarity among subsamples in site 2 (patchy spatial distribution). Use a t-test to determine whether compositional similarity was affected by the spatial distribution. Finally, compare the compositional similarity of site 1 and site 2 to the source community?

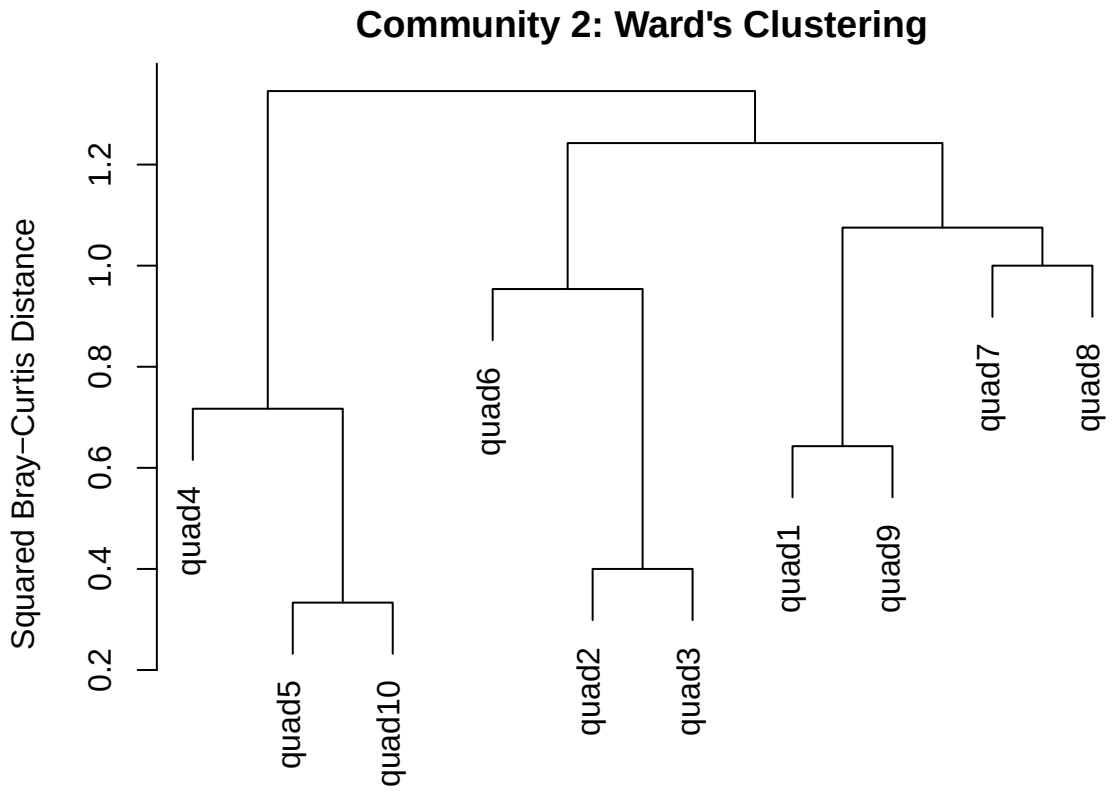
```
# calculate similarities
# calculate Jaccard
comm1j <- vegdist(comm_mat1$spec_dat, method = "jaccard", binary = TRUE)
comm2j <- vegdist(comm_mat2$spec_dat, method = "jaccard", binary = TRUE)
# calculate average
comm1avg <- mean(comm1j)
comm2avg <- mean(comm2j)
# perform t-test
comm12 <- t.test(comm1j, comm2j)
# Site 1 and 2 have a significant difference in similary in composition of sites. Site 1 has a much hig
```

2) Create a cluster diagram or ordination using your simulated data. Are there any visual trends that would suggest a difference in composition between site 1 and site 2? Describe.

```r
# calculate Bray-Curtis for comm1 and comm2
comm1.db <- vegdist(comm_mat1$spec_dat, method = "bray")
comm2.db <- vegdist(comm_mat2$spec_dat, method = "bray")
# perform cluster analysis of site 1
comm1cluster <- hclust(comm1.db, method = "ward.D2")
# plot cluster site 1
par(mar = c(1,5,2,2) + 0.1)
plot(comm1cluster, main = "Community 1: Ward's Clustering", ylab = "Squared Bray-Curtis Distance")
```




```r
# perform cluster analysis of site 2
comm2cluster <- hclust(comm2.db, method = "ward.D2")
# plot cluster site 2
par(mar = c(1,5,2,2) + 0.1)
plot(comm2cluster, main = "Community 2: Ward's Clustering", ylab = "Squared Bray-Curtis Distance")
```




```
# Community one has less uniformity in similarity between quadrats, whereas community two has a very ev
```