Assignment: Spatial Diversity

Michelle Benavidez; Z620: Quantitative Biodiversity, Indiana University

11 February, 2017

OVERVIEW

This assignment will emphasize primary concepts and patterns associated with spatial diversity, while using R as a Geographic Information Systems (GIS) environment. Complete the assignment by refering to examples in the handout.

After completing this assignment you will be able to:
1. Begin using R as a geographical information systems (GIS) environment.
2. Identify primary concepts and patterns of spatial diversity.
3. Examine effects of geographic distance on community similarity.
4. Generate simulated spatial data.

Directions:

1. Change "Student Name" on line 3 (above) with your name.
2. Complete as much of the assignment as possible during class; what you do not complete in class will need to be done on your own outside of class.
3. Use the handout as a guide; it contains a more complete description of data sets along with the proper scripting needed to carry out the assignment.
4. Be sure to **answer the questions** in this assignment document. Space for your answer is provided in this document and indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">".
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. When you are done wit the assignment, **Knit** the text and code into an html file.
7. After Knitting, please submit the completed assignment by creating a **pull request** via GitHub. Your pull request should include this file *spatial_assignment.Rmd* and the html output of `Knitr` (*spatial_assignment.html*).

1) R SETUP

In the R code chunk below, provide the code to:

1. Clear your R environment
2. Print your current working directory,
3. Set your working directory to your "*/Week4-Spatial*" folder, and

```
rm(list=ls())
getwd()
setwd("C:/users/Michelle/GitHub/QB2017_Benavidez/Week4-Spatial")
```

2) LOADING R PACKAGES

In the R code chunk below, do the following:

1. Install and/or load the following packages: `vegan`, `sp`, `gstat`, `raster`, `RgoogleMaps`, `maptools`, `rgdal`, `simba`, `gplots`, `rgeos`

```r
require("vegan")
require("sp")
require("gstat")
require("raster")
require("RgoogleMaps")
require("maptools")
require("rgdal")
require("simba")
require("gplots")
require("rgeos")
```

Question 1: What are the packages `simba`, `sp`, and `rgdal` used for?

> ***Answer 1***: The 'simba' is used for calculating and processing similarity data (for binary data) and reshaping data into matrices. The 'sp' package is for analying spatial data, such as map plotting, coordinate retrieving etc... The 'rgdal' package is used for spatial mapping. It provides bindings to the Frank Warmerdam's Geospatial Data Abstraction Library (software for reading raster and otehr geospatial files) and the PROJ.4 library (UNIX based library tthat converts between lat/long and cartesian coordinates).

3) LOADING DATA

In the R code chunk below, use the example in the handout to do the following:

1. Load the Site-by-Species matrix for the Indiana ponds datasets: BrownCoData/SiteBySpecies.csv
2. Load the Environmental data matrix: BrownCoData/20130801_PondDataMod.csv
3. Assign the operational taxonomic units (OTUs) to a variable 'otu.names'
4. Remove the first column (i.e., site names) from the OTU matrix.

```r
Ponds=read.table(file="BrownCoData/20130801_PondDataMod.csv",head=T,sep=",")
OTUs=read.csv(file="BrownCoData/SiteBySpecies.csv",head=T,sep=",")
otu.names=names(OTUs)
OTUs=as.data.frame(OTUs[-1])
```

Question 2a: How many sites and OTUs are in the SiteBySpecies matrix?

> ***Answer 2a***: 51 sites and 16383 OTUs

Question 2b: What is the greatest species richness found among sites?

> ***Answer 2b***: 16383

4) GENERATE MAPS

In the R code chunk below, do the following:

1. Using the example in the handout, visualize the spatial distribution of our samples with a basic map in RStudio using the `GetMap` function in the package `RgoogleMaps`. This map will be centered on Brown County, Indiana (39.1 latitude, -86.3 longitude).

```r
lats=as.numeric(Ponds[,3])
lons=as.numeric(Ponds[,4])

newmap=GetMap(center=c(39.1,-86.3),zoom=10)
PlotOnStaticMap(newmap,zoom=10,cex=2,col='blue')
PlotOnStaticMap(newmap,lats,lons,cex=1,pch=20,col='red',add=T)
```
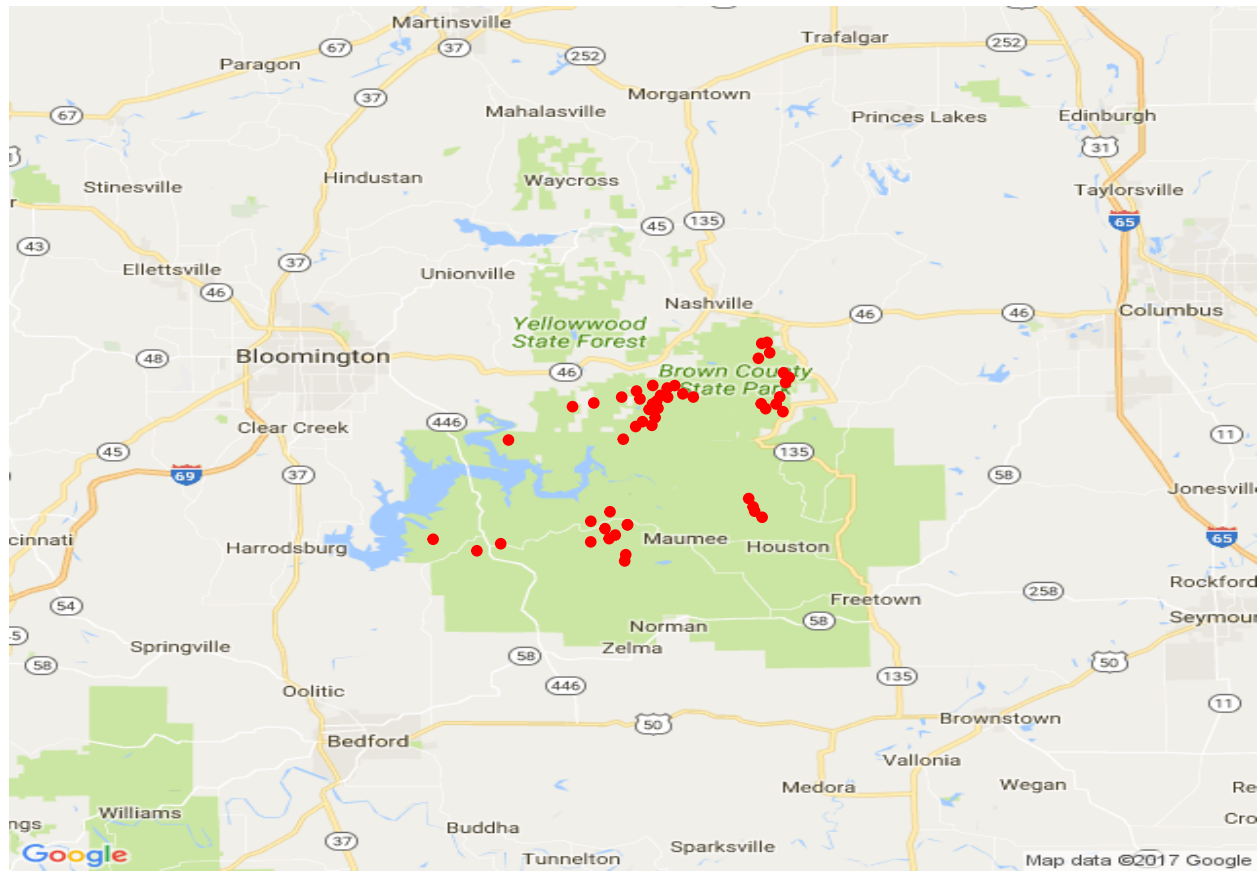
Question 3: Briefly describe the geographical layout of our sites.

    Answer 3: Sites are all located witin Brown County State Park. They are patterned in four distinct clusters with a few outliers around Monroe Lake. Two of the groups are located closer to the edge of the park while the other two are located closer to the park's center.

In the R code chunk below, do the following:

1. Using the example in the handout, build a map by combining lat-long data from our ponds with land cover data and data on the locations and shapes of surrounding water bodies.

```r
# 1. Import TreeCover.tif as a raster file.
Tree.Cover=raster("TreeCover/TreeCover.tif")

# 2. Plot the % tree cover data
plot(Tree.Cover,xlab="Longitude",ylab=Lattitude",
    main='Map of geospatial data for % tree cover,\nwater bodies, and sample sites')

# 3. Import water bodies as a shapefile.
Water.Bodies=readShapeSpatial("water/water.shp")

# 4. Plot the water bodies around our study area, i.e., Monroe County.
plot(Water.Bodies,border='cyan',axes=T,add=T)

# 5. Convert lat-long data for ponds to georeferenced points.
Refuge.Ponds=SpatialPoints(cbind(lons,lats))

# 6. Plot the refuge pond locations
```
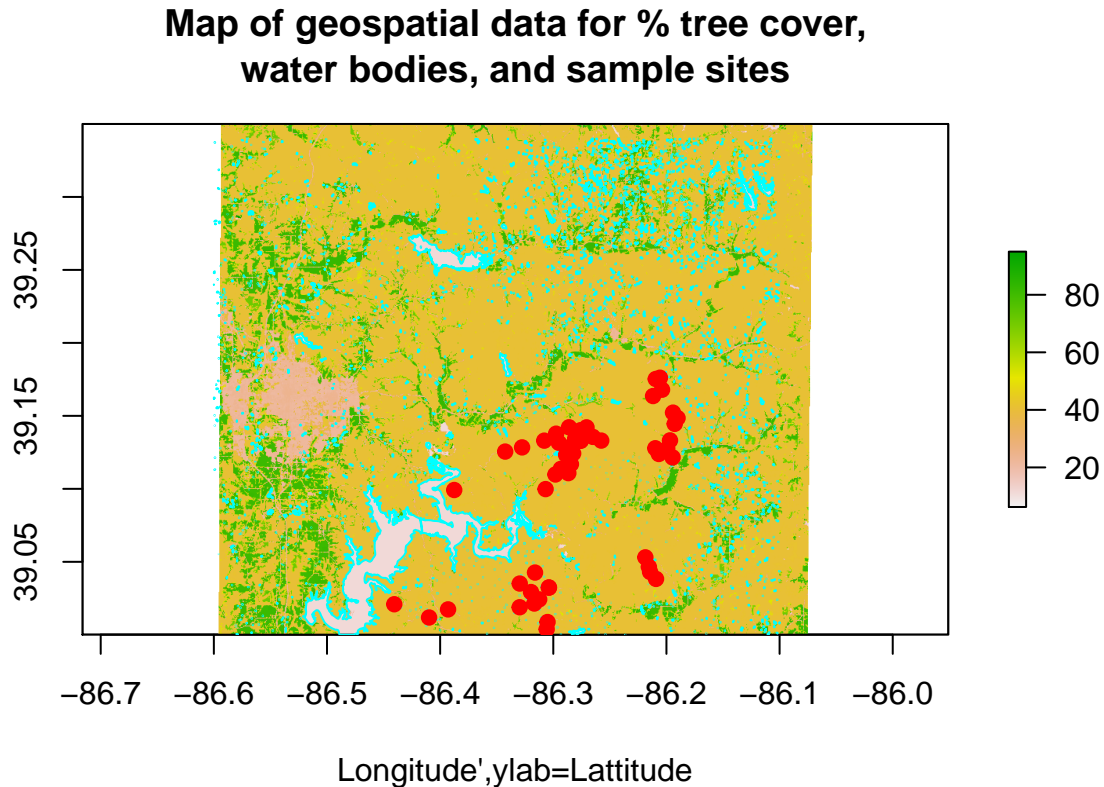
```
plot(Refuge.Ponds,line='r',col='red',pch=20,cex=1.5,add=T)
```

Map of geospatial data for % tree cover, water bodies, and sample sites

Longitude',ylab=Lattitude

Question 4a: What are datums and projections?

> ***Answer 4a***: A datum is basically a point of refernce on the globe that the projection uses to plot the reference point on a 2D map.

5) UNDERSTANDING SPATIAL AUTOCORRELATION

Question 5: In your own words, explain the concept of spatial autocorrelation.

> ***Answer 5***: The internal correlation of a variable through space. In other words, it is a measure of how much a viable correlates with itself over space.

6) EXAMINING DISTANCE-DECAY

Question 6: In your own words, explain what a distance decay pattern is and what it reveals.

> ***Answer 6***: The distance-decay pattern describes the diminishing interactions as two species/places/things of interest become more spatially distant.

In the R code chunk below, do the following:

1. Generate the distance decay relationship for bacterial communities of our refuge ponds and for some of the environmental variables that were measured. Note: You will need to use some of the data transformations within the *semivariogram* section of the handout.

4

```r
xy=data.frame(env = Ponds$TDS, pond.name = Ponds$Sample_ID, lats = Ponds$lat, lons = Ponds$long)

coordinates(xy)=~lats+lons

proj4string(xy)=CRS("+proj=longlat +datum=NAD83")

UTM=spTransform(xy, CRS("+proj=utm +zone=51 +ellps=WGS84"))

UTM=as.data.frame(UTM)
xy$lats_utm <- UTM[,2]
xy$lons_utm <- UTM[,3]



# 1) Calculate Bray-Curtis similarity between plots using the `vegdist()` function
comm.dist=1 - vegdist(OTUs)

# 2) Assign UTM lattitude and longitude data to 'lats' and 'lons' variables
lats=as.numeric(xy$lats_utm)
lons=as.numeric(xy$lons_utm)

# 3) Calculate geographic distance between plots and assign to the variable 'coord.dist'
coord.dist=dist(as.matrix(lats, lons))

# 4) Transform environmental data to numeric type, and assign to variable 'x1'
x1=as.numeric(Ponds$"SpC")

# 5) Using the `vegdist()` function in `simba`, calculate the Euclidean distance between the plots for
env.dist=vegdist(x1, "euclidean")

# 6) Transform all distance matrices into database format using the `liste()` function in `simba`:
comm.dist.ls=liste(comm.dist, entry="comm")
env.dist.ls=liste(env.dist, entry="env")
coord.dist.ls=liste(coord.dist, entry="dist")

# 7) Create a data frame containing similarity of the environment and similarity of community.
df=data.frame(coord.dist.ls, env.dist.ls[,3], comm.dist.ls[,3])

# 8) Attach the columns labels 'env' and 'struc' to the dataframe you just made.
names(df)[4:5]=c("env","struc")
attach(df)

# 9) After setting the plot parameters, plot the distance-decay relationships, with regression lines in
par(mfrow=c(1, 2), pty="s")
plot(env, struc, xlab="Environmental Distance", ylab="1 - Bray-Curtis",
     main = "Environment", col='SteelBlue')

OLS=lm(struc~env)
OLS
abline(OLS,col="red4")

 plot(dist, struc, xlab="Geographic Distance", ylab="1 - Bray-Curtis", main="Community\nComposition", co
```
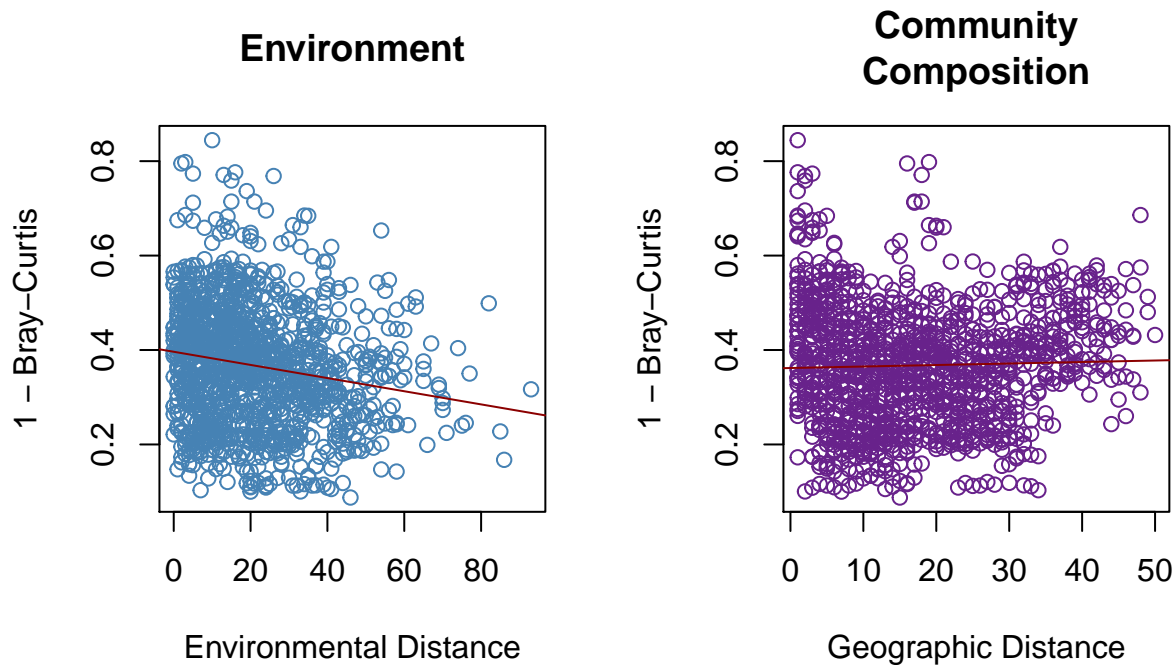
```
OLS=lm(struc ~ dist)
OLS
abline(OLS,col="red4")
```

Environment

Community Composition


```
# 10) Use `simba` to calculates the difference in slope or intercept of two regression lines
diffslope(env, struc, dist, struc)
```

Question 7: What can you conclude about community similarity with regars to environmental distance and geographic distance?

> *Answer 7*: There is a slight negative relationship between community similarity and environmental distance, but the community stay relatively stable over geographic distance. Accoutning for both simultaneouly, there is a significant slope (p < 0.001) that shows for every unit of change, species composition decreases by 0.17%

7) EXAMINING SPECIES SPATIAL ABUNDANCE DISTRIBUTIONS

Question 8: In your own words, explain the species spatial abundance distribution and what it reveals.

> *Answer 8*: The species spatial abundance distrbution describes how individual organisms are arranged in a given landscape. It can inform us about the rate and abundances at which you might encounter organisms in that particular space.

In the R code chunk below, do the following:

1. Define a function that will generate the SSAD for a given OTU.

2. Draw six OTUs at random from the IN ponds dataset and and plot their SSADs as kernel density curves. Use **while loops** and **if** statements to accomplish this.

```r
siteN=rowSums(OTUs)
# 1. Define an SSAD function
ssad=function(x){
  ad=c(2, 2)
  ad=OTUs[, otu]
  ad = as.vector(t(x = ad))
  ad = ad[ad > 0]
}

# 2. Set plot parameters
par(mfrow=c(2, 3))

# 3. Declare a counter variable
ct=0

# 4. Write a while loop to plot the SSADs of six species chosen at random
while(ct<6){
  otu=sample(1:length(OTUs),1)
  ad=ssad(otu)
  if(length(ad)>10&sum(ad>100)){
    ct=ct+1
    plot(density(ad),col='red',xlab='Site Abundance',
         ylab='Probability Density',main=otu.names[otu])
  }
}
```

Otu001243

Otu000501

Otu000076

Otu000136

Otu000512

Otu000232
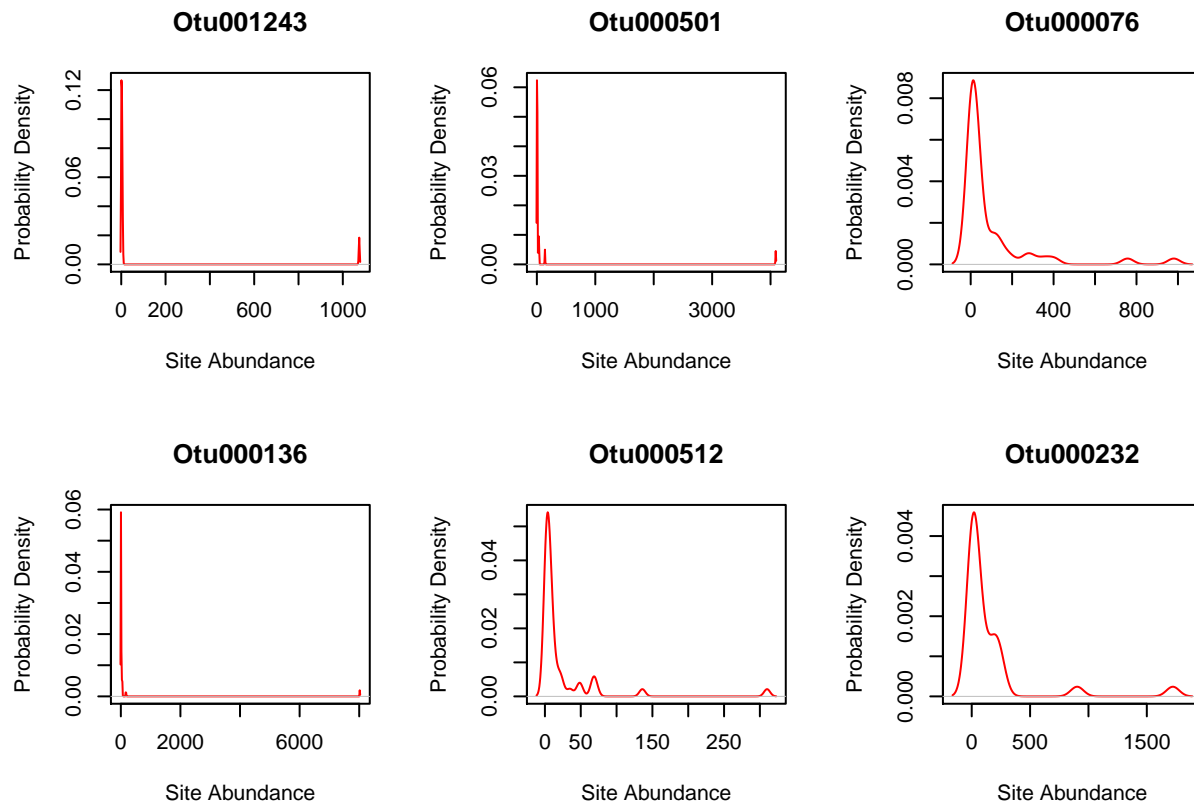
8) UNDERSTANDING SPATIAL SCALE

Many patterns of biodiversity relate to spatial scale.

Question 9: List, describe, and give examples of the two main aspects of spatial scale

> *Answer 9*: Two main aspects: Extent, Grain. Extent describes the greatest distance in a given study area. Grain, on the other hand, is the smallest distance in a given study area. For example if you are measuring biodiversity in a 20ha plot the extent is 20ha. If you were measuring diversity per hectare in that 20ha plot, then 1ha would be the grain of that study. As another example, if you were measuring diversity along 100km interstate and had that study site partitioned into 0.25km plots, then the extent is 100km and the grain is 0.25km.

9) CONSTRUCTING THE SPECIES-AREA RELATIONSHIP

Question 10: In your own words, describe the species-area relationship.

> *Answer 10*: Typically, as you look over larger and larger areas of land/habitat you will discover more species. This relationship can be visualized as the species-area curve, which quantifies this phenomenon.

In the R code chunk below, provide the code to:

1. Simulate the spatial distribution of a community with 100 species, letting each species have between 1 and 1,000 individuals.

```r
par(mfrow=c(2,2))

# 1. Declare variables to hold simulated community and species information
sd=0.75
mn=1.5

# 2. Populate the simulated landscape
x1 <- rnorm(10000, mean = mn, sd =sd)
y1 <- rnorm(10000, mean = mn, sd =sd)
x2 <- rnorm(10000, mean = -mn, sd =sd)
y2 <- rnorm(10000, mean = -mn, sd =sd)
x3 <- rnorm(10000, mean = -mn, sd =sd)
y3 <- rnorm(10000, mean = mn, sd =sd)
x4 <- rnorm(10000, mean = mn, sd =sd)
y4 <- rnorm(10000, mean = -mn, sd =sd)

plot(x1,y1, xlim=c(-4, 4), ylim=c(-6, 6), pch=".", cex=2, col='Steelblue')
plot(x2,y2, xlim=c(-2, 2), ylim=c(-2, 2), pch=".", cex=2,col='Steelblue')
plot(x3,y3, xlim=c(-1, 1), ylim=c(-1, 1), pch=".", cex=2,col='Steelblue')
plot(x4,y4, xlim=c(-0.5, 0.5), ylim=c(-0.5, 0.5), pch=".", cex=2, col='Steelblue')
```
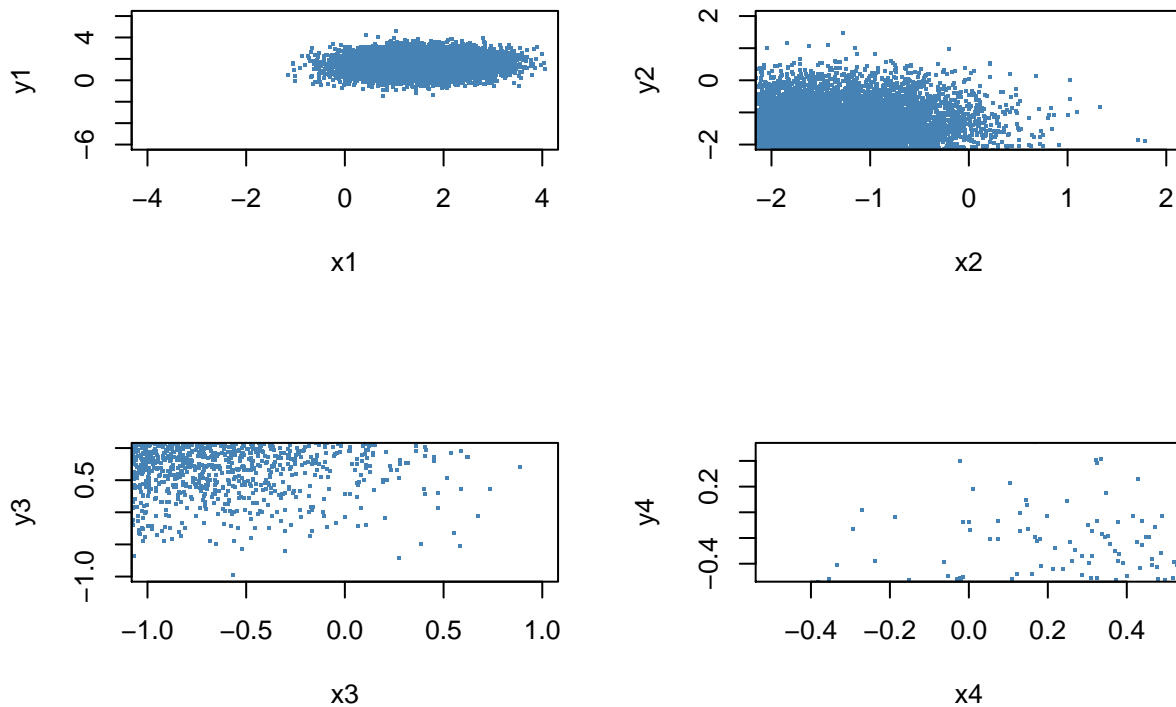

While consult the handout for assistance, in the R chunk below, provide the code to:

1. Use a nested design to examine the SAR of our simulated community.
2. Plot the SAR and regression line.

```r
par(mfrow=c(1, 2), pty="s")
community=c()
species=c()

plot(0, 0, col='white', xlim = c(0, 100), ylim = c(0, 100),
     xlab='x coordinate', ylab='y coordinate',
main='A simulated landscape occupied by 100
species, having 1 to 1000 individuals each.')
while (length(community) < 100){
  std <- runif(1, 1, 10)
  ab <- sample(1000, 1)
  x <- rnorm(ab, mean = runif(1, 0, 100), sd = std)
  y <- rnorm(ab, mean = runif(1, 0, 100), sd = std)
  color <- c(rgb(runif(1),runif(1),runif(1)))

  points(x, y, pch=".", col=color)
  species <- list(x, y, color)
  community[[length(community)+1]] <- species
}

# 1. Declare the spatial extent and lists to hold species richness and area data

lim=10
S.list=c()
A.list=c()


# 2. Construct a 'while' loop and 'for' loop combination to quantify the numbers of species for progres

while (lim <= 100){
  S <- 0
  for (sp in community){
    xs <- sp[[1]]
    ys <- sp[[2]]
    sp.name <- sp[[3]]
    xy.coords <- cbind(xs, ys)
    for (xy in xy.coords){
      if (max(xy) <= lim){
        S <- S + 1
        break
      }
    }
  }
  S.list <- c(S.list, log10(S))
  A.list <- c(A.list, log10(lim^2))
  lim <- lim * 2
}

# 3. Be sure to log10-transform the richness and area data
# This portion was done in the last few lines of the while loop above.
```
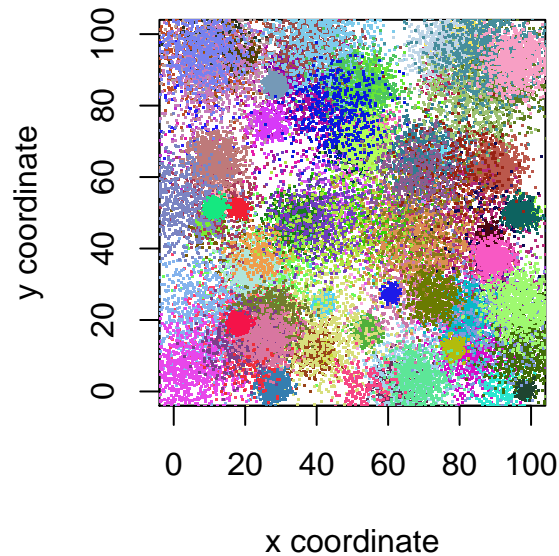
A simulated landscape occupied by
**species, having 1 to 1000 individuals**

In the R code chunk below, provide the code to:

1. Plot the richness and area data as a scatter plot.
2. Calculate and plot the regression line
3. Add a legend for the z-value (i.e., slope of the SAR)

```r
results <- lm(S.list ~ A.list)
plot(A.list, S.list, col="dark red", pch=20, cex=2,
     main="Species-area relationship",
     xlab='ln(Area)', ylab='ln(Richness)')
abline(results, col="red", lwd=2)
int <- round(results[[1]][[1]],2)
z <- round(results[[1]][[2]],2)
legend(x=2, y=2, paste(c('slope = ', z), collapse = " "), cex=0.8,
       box.lty=0)
```

Species–area relationship
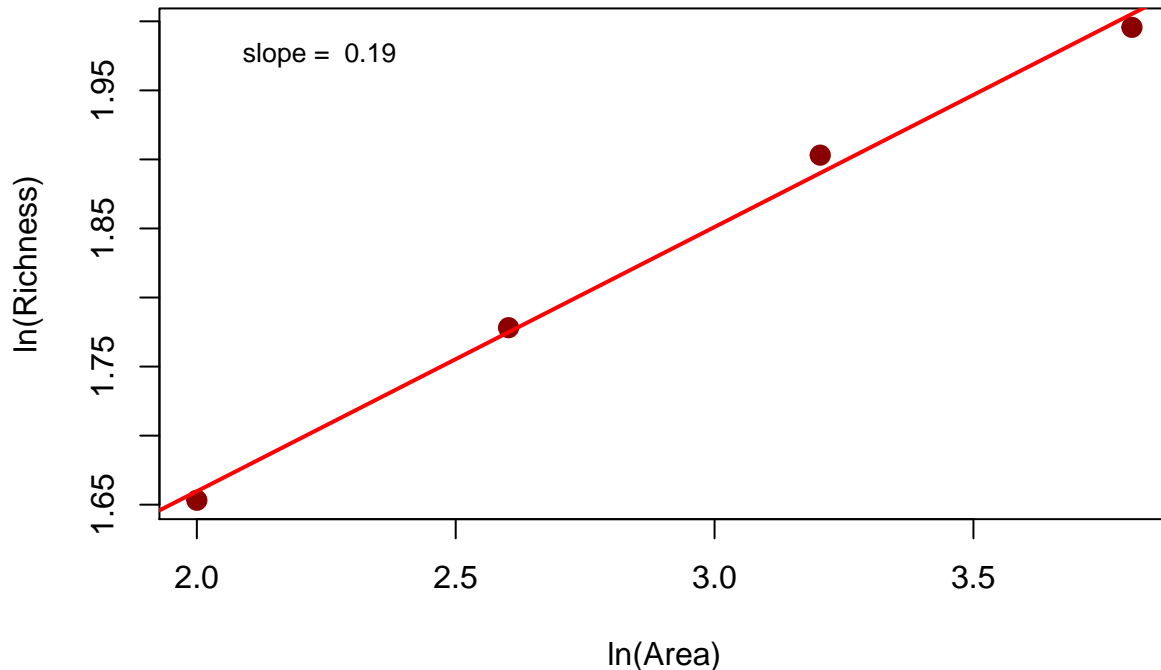
Question 10a: Describe how richness relates to area in our simulated data by interpreting the slope of the SAR.

Answer 10a: There is a positive relationship between species richness and area, so as the area of the land increases species richness increases as well.

Question 10b: What does the y-intercept of the SAR represent?

Answer 10b: The Y-intercept represents the log transformed values of species richness in a given area of the dataset.

SYNTHESIS

Load the dataset you are using for your project. Plot and discuss either the geogrpahic Distance-Decay relationship, the SSADs for at least four species, or any variant of the SAR (e.g., random accumulation of plots or areas, accumulation of contiguous plots or areas, nested design).

For this portion of the assignment I have looked at SSADs for six speices during one year (2005) of our time series data. I had to adjust the 'if' part of the loop compared to our homework activity as there were not many species in the dataset that met that criteria. From the adjusted numbers is would seem as if there might be a small collection of abundant species, but most are rare. There is definitely some differences in density probabilites between the species. Figuremage below.
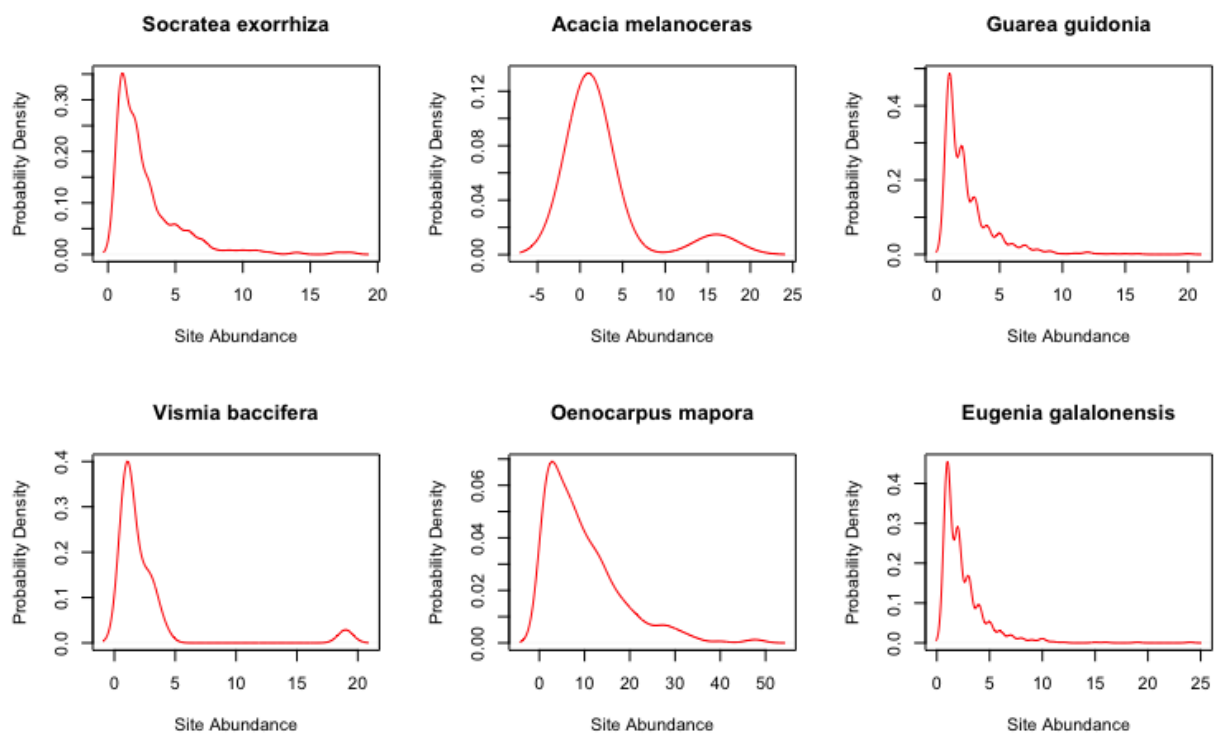
Figure 1: SSADs for six random tree species from a 50 hectare plot in BCI during 2005