# Geographical Ecology

Z620: Quantitative Biodiversity, Indiana University
? Feb. 2015

#### **OVERVIEW**

In this exercise, we will explore primary geographical aspects of biodiversity. You will also learn to: 1. Calculate univariate metrics of aggregation 2. Visualize geographical patterns 3. Test hypotheses of spatial correlation

## 1.) SETUP

Retrieve and Set Your Working Directory

```
rm(list=ls())
getwd()
setwd("~/GitHub/QuantitativeBiodiversity/Assignments/GeographicalEcology")
```

#### Load Basic Packages

As in previous exercises, we will use the **vegan** package to estimate aspects of diversity. We will also use the **BiodiversityR** package, which contains additional diversity estimators and related functions.

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

We will also use packages developed in R for geographical information systems (GIS). These packages will allow us to generate maps of study areas, project diversity data onto environmental layers, and conduct geographical analyses.

```
#install.packages("RgoogleMaps")
library(RgoogleMaps) # overlays of data onto a Google map.
```

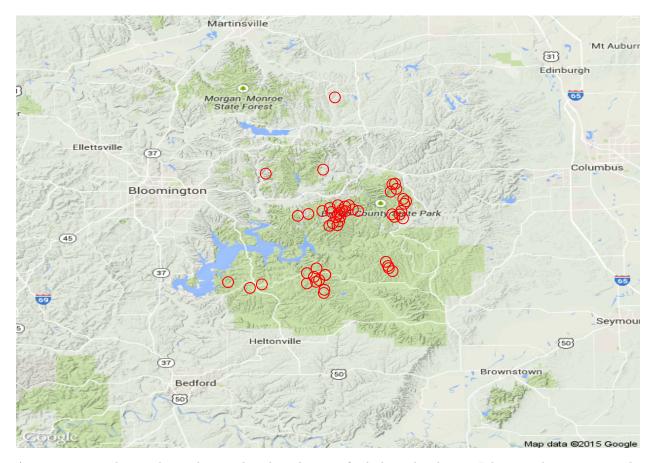
## 2.) MAKE A MAP

#### Generate a map using Google maps

Let's retrieve a map of Indiana using the GetMap function in the package RgoogleMaps. We will generate a map in the RStudio centered on Brown County, Indiana (39.1 degrees latitude, -86.3 degrees longitude)

```
newmap <- GetMap(center = c(39.1,-86.3), zoom = 10, destfile = "Brown_Co_map.png", maptype="terrain")
PlotOnStaticMap(newmap, zoom = 10, cex = 2, col='blue') # Now, plot the map in RStudio

Ponds <- read.table(file="BrownCoData/20130801_PondData.csv", head=TRUE, sep=",")
lats <- Ponds[, 3] # latitudes
lons <- Ponds[, 4] # longitudes
PlotOnStaticMap(newmap, lats, lons, cex=1.5, col='red', add = TRUE)</pre>
```



As you can see, the pond samples are distributed over a fairly large local area. Likewise, they are unevenly distributed in space, and over a complex topographic area. There are many possible ways to generate maps in R as well as to conduct geographical analyses. Some are offerred in the RGIS tutorial.

#### 3.) EXAMINE AGGREGATION

Aggregation can refer to the tendency for organisms to cluster within an environment, for events to occur within a particular period of time, and also refer to the tendency of a variable to take on a particular value, e.g., mean temperature, modal body size, average abundance.

In this component of the exercise, we will explore aspects of aggregation through graphical exploration and the calculation of univariate metrics.

#### Spatially implicit

Before examining how environmental variables, abundances, and aspects of diversity vary with distance or increasing area, we will examine how they vary within the landscape, regardless of their position in space.

#### Graphically examine environmental data

Before investigating correlations or conducting statistical tests, we can ask whether our data appear to be normally distributed, highly skewed, multi-modal, or even spurious. A basic graphical examination can be a powerful way to visualize many underlying statistical properties at once.

Let's begin with a simple graphical exploration of our environmental data. We will use kernel density curves, which are similar in use to histograms but avoid arbitrarily binning continuous data. Kernel density curves attempt to account for uncertainty and reveal the 'probability density' of a variable taking values within a particular range.

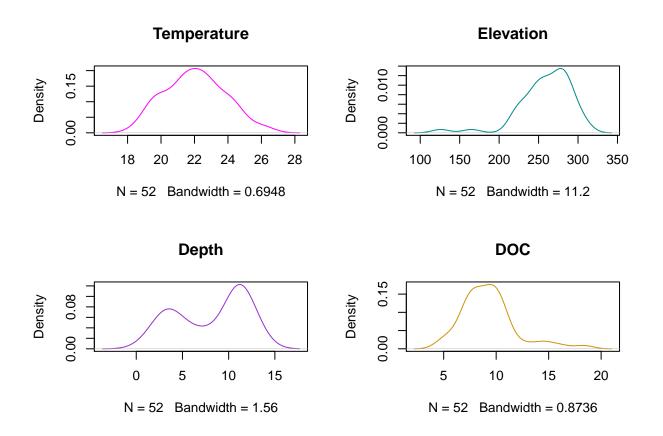
```
EnvData <- Ponds[5:21]
par(mfrow=c(2, 2))

data <- as.numeric(EnvData[, "Temp"])
plot(density(data), col = 'magenta', main = "Temperature")

data <- as.numeric(EnvData[, "Elevation"])
plot(density(data), col = 'DarkCyan', main = "Elevation")

data <- as.numeric(EnvData[, "Depth"])
plot(density(data), col = 'darkorchid', main = "Depth")

data <- as.numeric(EnvData[, "DOC"])
plot(density(data), col = 'darkgoldenrod3', main = "DOC")</pre>
```



**Graphical examination:** While distributions of temperature, elevation, and dissolved organic carbon appear to aggregated around a single modal value, depth appears to have two distinct modes. While temperature appears more or less normally distributed around a mean of 22 degrees celsius, elevation is clearly left skewed while DOC is clearly right skewed.

We have already begun to learn something about our environmental data without accounting for the locations of our samples. Likewise, we can do the same for the distribution of abundance and diversity among taxa.

#### Graphically examine taxonomic data

Load the site-by-OTU matrix for samples taken from ponds in and around Brown County, Indiana.

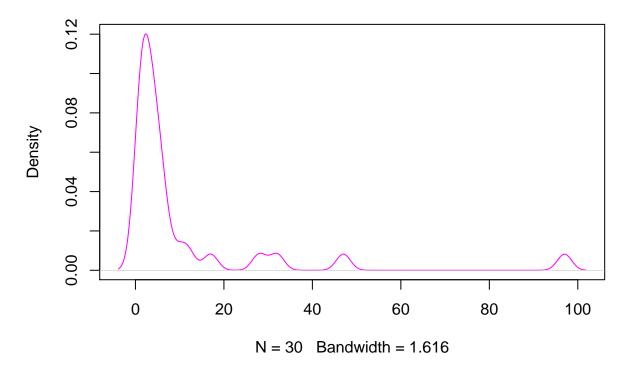
```
# Site by OTU matrix
OTUs <- read.csv(file="BrownCoData/SiteBySpecies.csv", head=TRUE, sep=",")
otu.names <- names(OTUs)
OTUs <- as.data.frame(OTUs[-1]) # remove first column (site names)
Site.N <- as.vector(rowSums(OTUs)) # no. reads</pre>
```

Construct kernel-density curves for the distribution of abundance among individual OTU's across the Brown County Pond dataset.

```
ad <- c(0,0)
otu <- 1
while (length(ad) <= 10){
   otu <- sample(1:length(OTUs), 1)

   ad <- OTUs[, otu]
   ad = as.vector(t(x = ad))
   ad = ad[ad > 0]
}
plot(density(ad), col = 'magenta', main = otu.names[otu])
```

## Otu000481



As you can see, sampled abundance for a given OTU is often aggregated, revealing many sites where the OTU is relatively rare and many where it is relatively more common. In fact, this uneven spatial distribution of abundance is common occurrence in ecological systems. This spatially implicit distribution of abundance is often referred to as the species spatial abundance distribution (SSAD).

**Question 1**: In the site-by-species matrix, each row represents a site and each column represents a species. If the SSAD is generated by considering all rows for a single column, then what do we obtain when we consider all columns for a given row? Have examined this sort of data structure before? If so, elaborate?

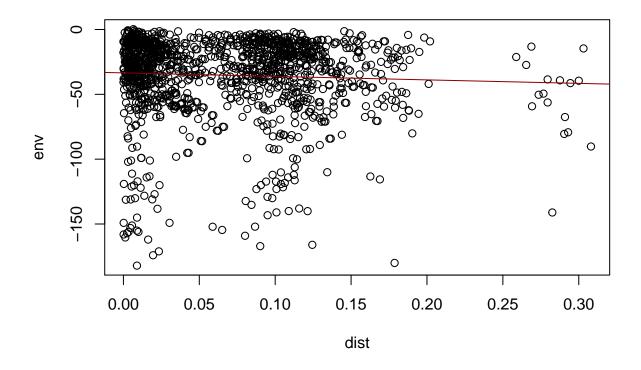
Answer 1:

## 4.) EXAMINE DISTANCE-DECAY

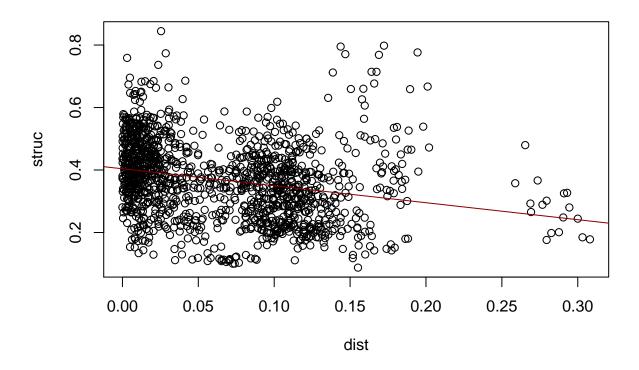
First law of geography...spatial autocorrelation...decrease in similarity with distance...

```
#install.packages("simba")
library(simba)
```

```
# calculate the similarity (Bray-Curtis) between the plots
struc.dist <- 1 - vegdist(OTUs)</pre>
# calculate geographical distance between plots
coord.dist <- dist(as.matrix(lats,lons))</pre>
# transform environmental data to numeric types
temp <- as.numeric(EnvData[, "Temp"])</pre>
elev <- as.numeric(EnvData[, "Elevation"])</pre>
depth <- as.numeric(EnvData[, "Depth"])</pre>
doc <- as.numeric(EnvData[, "DOC"])</pre>
# calculate the distance (Euclidean) between the plots regarding environmental variables
env.dist <- 1 - vegdist(cbind(temp, elev, depth, doc), "euclidean")</pre>
# transform all distance matrices into list format:
struc.dist.ls <- liste(struc.dist, entry="struc")</pre>
env.dist.ls <- liste(env.dist, entry="env")</pre>
coord.dist.ls <- liste(coord.dist, entry="dist")</pre>
# create a data frame containg plot information, geographical distance,
# similarity of environment, and similarity of community
df <- data.frame(coord.dist.ls, env.dist.ls[,3], struc.dist.ls[,3])</pre>
names(df)[4:5] <- c("env", "struc")</pre>
attach(df)
#df <- subset(df, struc != 0)
# plot Distance-Decay relationships with regression lines in red
par(mfrow=c(1, 1))
plot(dist, env)
abline(lm(env ~ dist), col="red4")
```



```
par(mfrow=c(1, 1))
plot(dist, struc)
abline(lm(struc ~ dist), col="red4")
```



```
# Is the slope significantly different?
res <- diffslope(dist, env, dist, struc)
res</pre>
```

```
##
## Is difference in slope significant?
## Significance is based on 1000 permutations
##
## Call:
## diffslope(x1 = dist, y1 = env, x2 = dist, y2 = struc)
##
## Difference in Slope: -26.44
## Significance: 0.038
##
## Empirical upper confidence limits of r:
     90%
           95% 97.5%
                       99%
##
##
    19.0
          24.3 29.4
                      34.2
```

## 5.) EXAMINE SPECIES-AREA RELATIONSHIP, a.k.a, taxa-area relationship

The species-area relationship (SAR) describes the rate at which we discover or accumulate species with increasing area. The general relationship of the SAR ... and is of the form  $S=c^*A^z$  ... which was first predicted by Arrhenius (1921)... wow, that's old.

# Random accmulation of sites

Accumulation of sites by proximity

```
OTU.obs <- function(x = ""){
  rowSums(x > 0) * 1
}
```