# Among Site (Beta) Diversity

*Student Name; Z620: Quantitative Biodiversity, Indiana University*

*22 January, 2017*

## OVERVIEW

In this assignment, we move beyond the investigation of within-site $\alpha$-diversity. We will explore $\beta$-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 $\beta$-diversity
2. visualize $\beta$-diversity with heatmaps, cluster analysis, and ordination
3. test hypotheses about $\beta$-diversity using multivariate statistics

## Directions:

1. Change "Student Name" on line 3 (above) with your name.
2. Complete as much of the exercise 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 exercise.
4. Be sure to **answer the questions** in this exercise document; they also correspond to the handout. 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, **Knit** the text and code into a html file.
7. After Knitting, please submit the completed assignment by creating a **pull request** via GitHub. Your pull request should include this file *beta_assignment.Rmd* and the html output of `Knitr` (*beta_assignment.html*).

## 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 "*/Week3-Beta*" folder, and
4. load the `vegan` R package (be sure to install if needed).

## 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.

***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***:
      ***Answer 1b***:
      ***Answer 1c***:

**Visualizing the Doubs River Dataset**

***Question 2***: Answer the following questions based on the spatial patterns of richness (i.e., $\alpha$-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***:
      ***Answer 2b***:
      ***Answer 2c***:

# 3) QUANTIFYING BETA-DIVERSITY

In the R code chunk below, do the following:

1. write a function (`beta.w()`) to calculate Whittaker's $\beta$-diversity (i.e., $\beta_w$) 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 $\beta$-diversity in the Doubs River.

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

    a. Describe how local richness ($\alpha$) and turnover ($\beta$) contribute to regional ($\gamma$) 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 $\beta$ change if we instead defined beta addititvely (i.e., $\beta = \gamma - \alpha$)?

      ***Answer 3a***:
      ***Answer 3b***:
      ***Answer 3c***:

**The Resemblance Matrix**

In order to quantify $\beta$-diversity for more than two samples, we need to introduce our second 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***:

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").

**Question 5**: Does the resemblance matrix (`fish.db`) represent similarity or dissimilarity? What information in the resemblance matrix led you to arrive at your answer?

> ***Answer 5***:

# 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.

## 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`).

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

> ***Answer 6***:

## 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.

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.

**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***:
> ***Answer 7b***:

# 5) 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.

***Question 8***: Based on the PERMANOVA, IndVal, and phi coefficient analyses, what did you learn about the relationship between habitat quality and the fish species composition?

> ***Answer 8***:

## B. Mutivariate 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 determine if these matrices are correlated, and test the hypothesis that fish assemblages are correlated with stream environmental variables.

***Question 9***: 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 9***:

### ii. Constrained Ordination

In the R code chunk below, do the following:

1. create an environmental matrix of the the water chemistry data included in the `doubs` dataset using forward and reverse selection of variables,
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,
4. use a permutation test to determine the correlation of each environmental factor on the constrained axes,
5. calculate the explained variation on the first and second constrained axes,
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 the constrained ordination.

***Question 10***: 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 10***:

### iii. Variation Partitioning

In the code chunk below,

1. Create a matrix model of the selected environmental variables,
2. Create a matrix model of the selected PCNM axes,
3. Perform constrained and partial constrained ordinations using the spatial and environmental models you just created,
4. Test the significance of each of your constrained ordinations using permutation tests,
5. Partition the variation among sites into the relative importance of space, environment, spatially structured environment, and residuals,
6. Plot the variation partitioning output to visualize it.

***Question 11***: Interpret the variation partitioning results.

    ***Answer 11***:

## SYNTHESIS

1. Using what you have learned from the visualizations and statistical analyses above, make an overall conclusion about spatial patterns and key predictors of fish $\beta$-diversity in the Doubs River.

2. Load the dataset you are using for your project. Perform an ordination to visualize your dataset. Using this ordination, develop some hypotheses relevant to $\beta$-diversity and identify the appropriate tools you would use to test them.