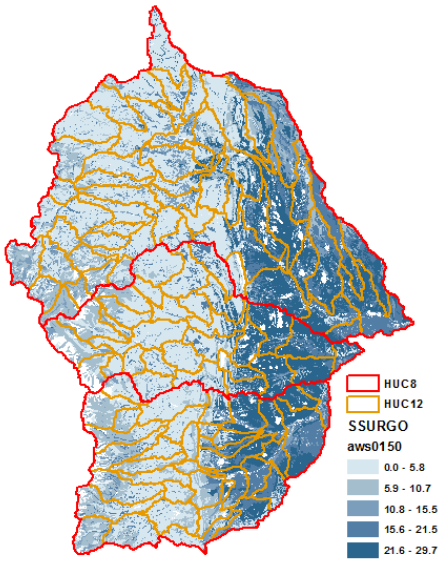
GEOG 4092/5092: Assignment 3
Random Spatial Sampling

Due October 1 (30 points)

Objective: Learn about spatial sampling, particularly random sampling regimes. Broaden understanding of vector geometry structures to encode and deconstruct objects. Learn to effectively implement functions in your code.

Data: Copy the *lab3.zip* file from Canvas into your working directory. This lab will use the SSURGO data from lab1, but for larger extent, and watersheds for the same area at two different levels of detail (HUC8 and HUC12). The HUC classification is a hierarchical system, where level 12 watersheds are nested within larger level 8 watersheds. We will use the available water storage 0-150cm (*aws0150)* variable from the SSURGO data in this lab.



Goals: Write a generic framework that implements a stratified random sampling of points within a polygon feature class. Generate two stratified random samples based on point density within the HUC8 and HUC12 watersheds. Then summarize the *aws0150* variable based on each sample. Compare the mean values of *aws0150* within each HUC8 watershed based on the two sampling routines conducted within the HUC8 versus the HUC12 boundaries.

Part I: Create a stratified random sampling approach for HUC8 and HUC12 watersheds

1. Calculate the number of points (*n*) to sample in each polygon (i.e., watershed):
    a. Get the extent of the polygon.
    b. Get the area of the polygon, and convert it to square kilometers.
    c. Calculate the number of points using a point density of **0.05 points/sqkm**. Round to the nearest integer.
2. Create a stratified random sample of *n* points within each polygon.
    a. Randomly generate points within the extent of the polygon using **random seed 0**, i.e., **random.seed(0)**.
    b. Test if each point is within the polygon boundary.
    c. If so, encode the point and indicate the HUC8 watershed that contains the point. NOTE: the first 8 digits of the HUC12 code are the HUC8 code.

Remember: You will create two samples, one based on the HUC8 watershed boundaries and one based on HUC12 boundaries. In Part II, you will summarize *aws0150* in the SSURGO data using these two sets of points, but you will do this summary at the HUC8 level for both samples. So, you need to store the HUC8 code for each point whether you are sampling within HUC8 or HUC12 boundaries. Remember that the HUC12 watersheds are nested within the HUC8 watersheds, and the HUC12 codes are the HUC8 codes plus four more digits at the end.

Part II: Summarize the SSURGO attributes by level of sampling

1. Calculate the mean *aws0150* for each HUC8 watershed from the HUC8- and HUC12-based point samples, i.e., you will have three mean values for each point sample. Hint: look at the pandas group by method.
2. In a final print statement, report the mean values for each HUC8 watershed from the two samples and your conclusions regarding the different results of sampling within HUC8 versus HUC12 watersheds. You should print out six total mean values (three for the HUC8-based point sample and three for the HUC12-based sample.

3. Implement at least one meaningfully used function. This means that your function effectively creates a useful, reusable operation. Functions should be put at the top of your script and should be well documented.

Suggestions:

Start by developing your script as a simple procedural workflow. Then implement generic structures using loops. In the second week you should implement a function (or functions) as reusability mechanisms for the steps you created in the first week. Think about which parts can be considered "autonomous entities" of execution, which you might use again in another context—these reusable pieces should be implemented as functions. Remember to write your script in a generic way; think about how you can develop your script to accommodate any number of watersheds and watershed levels.

By the start of next class in two weeks upload your script (*lastname_lab3.py*) to Canvas.

Grading: You will be evaluated on the following: your script runs without errors (4); Part I - completes step 1 (6) and step 2 (10); Part II - calculate means (2), prints correct values and conclusions (1), and uses a function effectively (2); and the clarity, logic and efficiency of your script (5).

Optional Challenge: Conduct the same analysis except by using a regular grid over the study area. Part of this challenge is to figure out how to create a regular grid that will have approximately the same point density (0.05 points per sqkm).