

GEOG 4092/5092: Assignment 1

Data preparation and vector geoprocessing

Due next class—September 3 (10 points)

Goal: Learn to work with multiple data files in an automated framework. Gain familiarity with the Pandas and Geopandas module to conduct vector geoprocessing and relational joins.

Data: Download the file *lab1.zip* from Canvas and unzip it into your home directory. The geopackage (*lab1.gpkg*) contains a feature class of two watersheds in northern Colorado and spatial and tabular data from the USDA Soil Survey Geographic Database (SSURGO). There are 9 feature classes that cover the area of the two watersheds. The feature classes delineate survey polygons used to record soil properties in SSURGO. The file geodatabase also contains 9 tables with attribute data that correspond to the feature classes. Column names in the tables correspond to the following: musym (map unit symbol, which is a classification scheme for mapped areas); aws025wta (available water storage 0-25cm); aws0150wta (available water storage 0-150cm); and drclassdcd (drainage class, indicating the dominant condition).

Task: You will process the SSURGO data by joining each of the tables to the corresponding feature class and then you will merge the 9 feature classes. Next, you will intersect the SSURGO data with the watershed boundaries and finally count the number of map unit polygons within each watershed.

Part I:

- 1) Join each of the tables to the corresponding feature class. Find a common field to join on.
- 2) Add a new field named *mapunit* to each joined feature class and calculate the values for all features as the map unit ID of the feature class. The map unit ID is the suffix of the layer name (e.g., co641).
- 3) Merge the 9 feature classes into a new feature class.
- 4) Intersect the merged feature class with the watershed boundaries.

Part II:

Find the number of features in the resulting intersected feature class that correspond to each watershed.

Report these two numbers in a print statement at the end of your script. Hint: Look at the **groupby** function from pandas, which column should you group by?

Take the time to investigate. It might be easier to first work on one feature class for Part I steps 1 and 2, and develop a prototype for your workflow. Then, loop through the 9 shapefiles and repeat the single tasks. Go through the single steps very carefully and make use of the Geopandas and Pandas documentation to find out more about usage and the required function and syntax.

By the start of next class upload your script (*lastname_lab1.ipynb*) to Canvas.

Grading: You will be evaluated on the following: your notebook runs without errors (2); Part I: joins tables to feature classes (2), adds and calculates *mapunit* field (1), merges the 9 feature class and intersects the result with the watershed boundaries (1); Part II: prints the number of intersected features in each watershed boundary (2); and the clarity, logic and efficiency of your code (2).