Geoprocessing in python (SS 2019)

Conservation Biogeography Lab, Geography Department,
Humboldt-University Berlin
Matthias Baumann, Dirk Pflugmacher

Assignment II – Geopandas

Due date	Sunday, 5.5.2019 10pm
Submission form	1. Questions in relation to the scripting task within a moodle-quiz. 2. py-Script
Evaluation criteria	• Correctness of the asked results (50%). • Functionality of the py-script (50%).

Goal of this assignment

- Become familiar with pandas
  - Read and write data
  - Indexing and selection
  - Merge and join
  - Aggregation and grouping, pivot tables
- Geopandas
  - Add projection
  - Add geometry
  - Save to shapefile
- Shapely
  - Point and Polygon class
- Strengthen you knowledge of loops, list-comprehensions, iterator functions

Exercise I

The European Land Use/Cover Area frame Survey (LUCAS) provides a unique source of publicly available land cover survey data (https://ec.europa.eu/eurostat/de/web/lucas/overview). LUCAS collects information on land cover and land use, agro-environmental, and soil data every three years in the European Union since 2006 and at a systematic sampling grid with points spaced 2 km apart. Assume, you downloaded two tables from the LUCAS webpage: 1) the primary table with the land cover and land use data recorded at each sampling point (DE_2015_20180724.csv), 2) the grid table containing the sampling locations (GRID_CSVEXP_20171113.csv). Your task is to filter out unsuitable observations and then create a shapefile with the points locations GPS_LONG and GPS_LAT from the grid table, but also including the records from the primary table. The coordinate system is: https://www.spatialreference.org/ref/epsg/4326/

Filter criteria for LUCAS samples:

OBSERVED	1= point is observed; 2= Point is not visible; 3= point in marine sea; 4= point out of national territory
OBS_TYPE	1= field survey - point visible at a distance less than 100 m; 2= field survey - point visible at a distance more than 100m; 3= photo interpreted in the field since point not visible; 4= point not surveyed
OBS_DIRECT	1= on the point; 2= look to north rule applied; 3= look to east rule applied; 8= not relevant
OBS_DIST	Expressed in meters; Distance from the point (based on GPS coordinates)
OBS_RADIUS	1= 1.5m; 2= 20m
AREA_SIZE	1= less than 0.5ha; 2= between 0.5ha and 1ha; 3= between 1ha and 10ha; 4= more than 10ha; 8= N.R.
LC1_PERCENT	1= less than 5%; 2= between 5% and 10%; 3= between 10% and 25%; 4= between 25% and 50%; 5= between 50% and 75%; 6= more than 75%; 8= N.R.
FEATURE_WIDTH	1= less than 20m (not important) 2= more than 20m 8= N.R.

1. Filter out potentially unsuitable observations. Save the observations that meet the criteria below into a new data frame and answer the following questions: 1) How many observations are left?, 2) How many land cover classes (LC1) were recorded?, 3) What is the average observer distance (meters)?, 4) What is the average observer distance of class A21?, 5) How many samples are in the land cover class (LC1) with the most samples?

   Filter criteria:
   - OBS_TYPE = 1
   - OBS_DIRECT = 1
   - OBS_RADIUS <= 2
   - AREA_SIZE >= 2
   - FEATURE_WIDTH > 1
   - LC1_PCT >= 5

2. Create a GeoDataFrame of the filtered LUCAS samples and save it as an ESRI shapefile.

Exercise II

Assume you used EarthExplorer (https://earthexplorer.usgs.gov) to select all Landsat-8 Collection 1 scenes acquired over Germany and downloaded their metadata (LANDSAT_8_C1_313804.csv). Your task is to explore and filter the metadata of the available Landsat scenes and to build a shapefile from selected scene boundaries.

1. Filter out potential scenes using the filter criteria below. Save the scenes that meet the criteria below into a new data frame and answer the following questions: 1) What is the average geometric accuracy in the X and Y direction: 'Geometric RMSE Model X' and 'Geometric RMSE Model Y', 2) What is the average land cloud cover, 3) How many Landsat footprints (unique Path/Row combination) are there?

   Filter criteria:
   - Day/Night Indicator = DAY
   - Data Type Level-1 = OLI_TIRS_L1TP
   - Land Cloud Cover < 70

2. Create a GeoDataFrame of the filtered Landsat scenes and save it as an ESRI shapefile (Hint: “Use UR Corner Lat dec”, etc.)