Geoprocessing in python (SS 2019)

Geography Department, Humboldt-University Berlin
Matthias Baumann, Dirk Pflugmacher

Assignment I – Basic python scripting

Due date	Sunday, 21.4.2019 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

The overarching goal of this assignment is to get familiar with the scripting environment (python, pycharm) and perform basic processing tasks. Important components are *variable assignments*, *loops*, *string-searches* (*regular expressions*), and a low number of *easy calculations*. There will be two overall tasks, which are individually described. This assignment is a good opportunity to develop and create a scripting structure, which you can use in subsequent assignments. An example of how such a script may look like is provided in moodle, and you can use this to tailor your own script that comforts you best. Of course, you can also create an entirely new script.

Exercise I – sanity check

Suppose, you (or your colleague) has downloaded a large number of Landsat scenes for you, which you will need in your analysis. The data come from the USGS and contain surface reflectance values as well as sensor quality bands and meta-information. You can find the data in moodle (Part01_Landsat.rar), in which the RAW-archives were already extracted (*caution: for the purpose of data storage, these are only dummy files*). For this assignment please check several things:

1. The Landsat sensor family has now overall four satellites (L4, L5, L7, L8), and for each of the four we have a different number of scenes available. Please assess for each footprint, how many scenes from each individual sensor are located in the folders and submit the numbers into moodle under question 1.
2. Although generally very reliable, in some rare cases not all files are being transferred by the USGS or got lost during the extraction process. As a result, some scenes may be incomplete, which can have implications in later steps of a longer processing chain. This makes a pre-check of each individual scene necessary. The task here is to (a) count the number of scenes that do not have the "correct" number of files in them (<u>caution: the number of files may vary between the different sensors!</u>); and (b) generate a text-file, in which each corrupt scene (*i.e., the entire file path*) is written as an individual line. Submit the number of erroneous scenes into question 2 in moodle and upload a txt-file under question 3.

Exercise II – data overview

Assume you are receiving several geospatial datasets from a collaborator, who is an economist and not a GIS expert and therefore does not have an overview of the usability of the data. As it turns out, your collaborator also was not the producer of the datasets, but got the data from someone else, which is another reason for us to thoroughly check the datasets, to highlight corrupt datasets, and to generally count the number of layers that we can work with later on. For this exercise, please check:

1. The number of SHP-files and the number of raster-files that are in the folder (*again: for the purpose of data storage, these are only dummy files)*. Please make sure that you only count the number of layers and not the number of files (e.g., a SHP-file is composed of more than one file, the same may be true for raster-files). Once you have your results, please upload them for both layer types (vector and raster) into moodle under question 4.
2. Some layers are incomplete, which is particularly true for the vector layers (i.e., the SHP-files). Some of them are missing a *dbf*-file, while for others the *projection information* is missing. The task is to identify these layers. Please submit into moodle under question 5 the number of incomplete vector and raster layers. Under question 6 then provide a list of the layer names.