

Geoprocessing in python (SS 2018)

Conservation Biogeography Lab & Earth Observation Lab

Geography Department, Humboldt-University Berlin

Matthias Baumann, Dirk Pflugmacher

Assignment VI – Protected area summary

Due date	Sunday, 2.6.2019 22:00
Submission form	1. csv-file 2. py-Script
Evaluation criteria	• Correctness of the asked results (49%). • Functionality of the py-script without external libraries (51%).

Goal of this assignment

The overarching goal of this assignment is to develop skills to successfully process vector data using python and OGR. In addition, we want to start making use of the advantage of scripts to efficiently process large files, which often turns out being problematic when using standard GIS software.

One requirement of this assignment is to not use python libraries through which you can perform the entire analysis using only a few lines of code. Instead, you are asked to directly interact with the shapefiles through ogr-commands (check above for the evaluation criteria)

Datasets

For this week, we have two shapefiles, which you can download under <https://tinyurl.com/y5bbgejt> or via the original homepages (see info below). Both shapefiles are large in their overall file size and their number of features. Specifically, the two datasets are:

1. The world database of protected areas. As the name suggests, the database contains polygons for each individual protected area on the globe, with different information such as the IUCN category. The original files are available under <https://protectedplanet.net/>
2. Country boundaries. The zonal shapefile, which is slightly modified compared to its original version (<https://gadm.org/data.html>) → the minimum unit is the country-level.

Your task

Your task is to provide a summary dataset, which indicates for each country in the world:

- (a) the number of designated terrestrial protected areas in it. In case a protected area is located in a border region of several countries, the protected area counts for each country individually.
- (b) the number of protected areas of each IUCN category. You can find out the attribute of the by examining the attribute table of the shapefile or by checking the meta-information online.
- (c) for each of the two summaries above (i.e., the number of PAs of the respective categories) calculate the mean PA size and the size of the largest PA in the country. In addition, extract the name of the largest PA and its establishment year. Compile the information into an output-file in form of a panel (see below).

Submission form

The required output is a csv-file in the order below, where each column is separated by a “,”.

Country ID	Country Name	PA Category	# PAs	Mean area of PAs	Area of largest PA	Name of largest PA	Year of establ. Of largest PA
1	Country01	ALL	1234	123.45	12134.56	NAME_of_PA	2005
1	Country01	1					
1	...	2					
1		...					
1		n					
2	Country2	ALL					
2	Country2	1					
2		2					
2		...					
2		n					
...		ALL					

Format of the assignment

This week's assignment can turn out to be rather complex and calculation intense, and there are (as usual) several ways of how to get to the finish line. Yet, the pathway you choose will determine to some extent the final processing time, so that carefully deciding on the path to run the data **before** starting the actual data processing is important. This week's assignment can therefore be done in groups of up to 4 people. During class time, I would like you to discuss and decide on a strategy to perform the analysis. After that, you are free to share the work or do the entire analysis by yourself. However, make sure that each of your group members submits the solutions in moodle. Likewise, the script that you will need to upload has to contain the names of ALL group members.