Coding Test

Hude Hude
hh3024@columbia.edu

Basic Info

There are 5 parts in the test. I have completed task 1 - 4.
The zip file contains a collection of python programs, datasets in .tif format, output images, and a pdf file of coding note.

Breakdown

Task 1:

Get_AmazonBiome_new.py
Takes a brasil_coverage.tif and shapefiles of Amazon Biome as inputs. It generates amazon_coverage.tif as output.
Results: See the folder of coverage_amazon

Check_Shapefile.py
Takes shapefiles of Amazon Biome and visualize it for checking the validity of the shapefiles.
Results: Shapefile_Amazon_Biome.png in Outputs folder

Check_clip.py
Takes the output data obtained from Get_Amazon_new.py as input. It visualizes a part of the image (amazon_converage.tif) as well as prints out some random data points for checking.
Results: 1985_PartialSubset_Figure.png in Outputs folder

Task 2:

Get_Legacy.py
Takes the initial two Amazon Biome samples and generates legacy_coverage.tif, which represents the data of initial legacy forest.
Results: legacy_coverage.tif in coverage_amazon folder

Task 3:

Get_LegacyArea.py
Takes legacy_coverage.tif as input and compute the area of legacy forest in hectares.
Results:
Total Number of Pixels: 11892635340
Total Legacy Forest Area: 404506286.19 hectares.

Taks 4:

Get_Deforest_Nested.py
Takes amazon_coverage.tif in a range of year as inputs. Computes deforest rate for each year and update a dataset that tracking the changes to the legacy forest.
Results:

Total Deforest Area in 1987: 1911754.26 hectares
Deforest Rate in 1987: 0.4726%
Total Deforest Area in 1988: 3641055.30 hectares
Deforest Rate in 1988: 0.9001%
Total Deforest Area in 1989: 5053891.41 hectares
Deforest Rate in 1989: 1.2494%
Total Deforest Area in 1990: 6229313.46 hectares
Deforest Rate in 1990: 1.5400%
Total Deforest Area in 1991: 7460172.54 hectares
Deforest Rate in 1991: 1.8443%
Total Deforest Area in 1992: 9145510.20 hectares
Deforest Rate in 1992: 2.2609%
Total Deforest Area in 1993: 10772832.15 hectares
Deforest Rate in 1993: 2.6632%
Total Deforest Area in 1994: 12597069.42 hectares
Deforest Rate in 1994: 3.1142%

Get_Deforest_1987.py and Get_Deforest.py
Experiments for getting the generalized version.

None of the three successfully generate an output file that stores relevant information for task 5.

Task 5:
Have only some ideas but unable to start because of the issue mentioned above.

Coding:
I don't have experience dealing with tif data before. I massively consulted with online resources for understanding the data structure and methods that applies to it.

Memory Issue
I used a PC with 16GB RAM.

I have encountered memory error multiple times, especially when I try to store the entire dataset into arrays. I spent a great portion of time trying to find solutions given the memory capacity. I solved relevant memory issue for task 1 - 4. However, I stuck with Task 5 since it couldn't solve memory error at this step.

Running time is another issue, on average, each program takes a few minutes to finish. Get_Forest_Nested.py takes much longer (about an hour) to finish running.

More details are documented in the Coding_Note.pdf.