Coding Test

You have 3 days from the moment you receive this test to return your answer to me. This is meant to assess both your current coding skills and implementation capabilities (eg your current work computer). Don't worry if you can't make much progress or finish the entire test. Document your progress and answers in your prepared output document, which is what you'll send to me. I recommend using R or Python for the exercise.

Download the .tif files in the folder

https://www.dropbox.com/scl/fo/wnai0zowvkz5rpvjtxgve/h?rlkey=5a79b1w32za7dmgrjik0oowpm&dl=0

The folder contains annual satellite imagery of land cover for Brazil. It's basically a giant picture where each pixel maps a 30 x 30-meter satellite shot of the same land every year. Don't try to open the file on your PDF reader (eg Preview, Adobe) as it will crash your computer.

Each pixel in this picture has a unique color from a specified color pallet. Each color represents one type of land cover (for example water body, soybeans, tropical forest) in that area and is represented by a number. Each pixel is georeferenced, so that by tracking the same pixel over time one can learn what happened to the land cover in that area. For instance, a pixel once occupied by forest may be deforested another year, and eventually become used by livestock or become a built structure.

1) The files contain satellite data for the entire country of Brazil. We'll only work with the Amazon biome for this exercise. Your first task is to subset the .tif file (also called raster) to only include pixels contained in the Amazon biome boundary. You can easily find Amazon biome shapefiles online for download.

2) Always working with the raster file only for the Amazon biome, start with the files for years 1985 and 1986. For each pixel, if the color code is contained in (1, 3, 4, 5, 6, 49, 10, 11, 12, 32, 29, 50, 13), assign that pixel as "forest". Assign it as "not forest" otherwise. If the same pixel is classified as "forest" both in 1985 and 1986, this will be our initial forest stock, or legacy forest. You can think of this as the "original forest" (at least that's as far back we can measure with this satellite data).

3) Given that each polygon spans 900 meters of actual land, calculate the total legacy forest area in hectares (convert it by multiplying by 0.09).

4) For each year from 1987 to 2020, we'll now calculate deforestation rates. We define a pixel as "deforested" if it started as legacy forest and then changes to human cover. Human cover color codes are (14, 15, 18, 19, 39, 20, 40, 62, 41, 36, 46, 47, 35, 48, 9, 21, 24, 30). Once a pixel becomes "deforested" (human cover-coded), it stays deforested forever. Be clever about how you implement this step and leverage the fact that you'll have increasingly fewer pixels in your relevant sample year after year (since you can discard "deforested" pixels). Calculate annual deforestation rates for as many years as you can, starting in 1987 and advancing in time (<u>hint: deforestation rate can be calculated as # of deforested pixels/# of legacy forest pixels</u>).

5) Generate one plot of the Amazon biome color-coded in green as “legacy forest” and gray as “other” (the “not legacy forest” pixels) for the initial forest stock. Put this plot next to another plot of the latest year you processed in the previous step – say 2020 – where you again plot in green the “legacy forest”, in gray “other”, and in yellow “deforested”. It will be hard to plot all pixels (there will be approximately 9 billion pixels). There are plotting tools that allow you to shrink the number of plotted pixels or simplify rendering. For reference, it takes me about 30-45 seconds to generate both plots.

Document all your answers and code written in your output. If you’re not able to calculate everything or run your code for all years, run it as much as you can and explain in the output your approach. Feel free to provide your answers in tables or free-form text; think about what’s the best way to convey each piece of data.