

Advanced Geoscripting - Final project

Hannah Weiser - 3405946 - h.weiser@stud.uni-heidelberg.de

10.11.2020

1 Before the assignment - Objectives

My intention for the final assignment was to conduct an exploratory data analysis in a Jupyter Notebook mainly using Pandas and Geopandas. I wanted to work with ecological data and investigate the distribution of certain insect species. As a second step, I wanted to write a python program with a final workflow. In my program, I used two APIs: The (GBIF API) and the Eurostat (GISCO API).

My main objectives for the project were

- using git and committing code regularly.
- using one or multiple APIs.
- organizing the code in functions.
- documenting which dependencies are needed
- documenting how to run the entire program
- documenting what the individual functions take as arguments and what they return.
- integrating error handling.
- practicing writing tests.

2 After the assignment - Reflection

Overall, I learned a lot during the final project by using APIs which I was not familiar with before. I found that Jupyter Notebooks are very useful for exploring the data and experimenting with different analysis and I will keep using this approach in future projects. After this interactive exploratory data analysis, it was easy to create a nice workflow in a python program and to decide for which steps I need to write a function and how the function should look like.

I wrote the final python script in PyCharm which helps to maintain code style by analysing the code, noticing errors and giving warnings and suggestions.

I managed to regularly use git during development. It was very helpful to develop the habit of committing code regularly. This also improved my overall organization in the project, because it made me work on smaller tasks and components at a time. I still sometimes "forgot" to commit and had to commit

multiple or larger changes at once. So there is still the potential to strengthen this habit.

After an initial working implementation, I added docstrings for the functions and took care of handling errors (e.g. missing entries in the configuration file, undefined coordinate reference system in GeoDataFrames, etc.). Actively trying to find exceptions and edge cases which might throw errors was beneficial for making the functions and the overall program more flexible. This is beneficial for running the analysis with different input data or reusing parts of the code for other projects.

I also documented the overall program separately in the github README. Especially noting down the required dependencies or, even better, providing an anaconda environment file is a practice that I will make a new habit. This facilitates sharing the code or collaborating on it, because others can quickly set up the environment and run it without hassle.

Finally, I practiced unit testing. I wrote two tests for one function to make sure it works in different cases. This is a useful practice to find bugs in the first place and to make sure the code is still running correctly even after making more changes on the respective function, class or method.

While I think my program is well organized overall, it could be further improved when it comes to readability. Nonetheless, I think it is documented well enough so that I still understand what it does and how to use it when I look at it again in two years, which is the most important thing for me at this time.