Geospatial Data Science Exam Project Instructions, Spring 2023

The submission is a written project report about the application of geospatial data science *either* to **answer a research question** or to **create a prototype of a digital product**. It may range from a technical workflow proof of concept to research data exploration. The project should explore or solve a problem with a geospatial dimension and may focus on any aspect of spatial data collection, visualization, analysis, or statistical evaluation. The submission has two parts: (a) commented code deposited on Github (or similar code repository) and (b) the associated report that describes the project and links to the code repository.

If anything is unclear after reading this document, please ask in the forum or in class.

Project groups

The group projects must consist of 1 to 3 people. It is thus allowed to work alone, but we strongly recommend you to work in groups if at all possible – it usually results in better projects!

Project approval

Before starting to work on your project, you must submit a very short project proposal at the latest **by March 31st** in the Google Sheet linked at LearnIT. The reason is to avoid unrealistic plans or substantial overlaps between groups. All proposals need to be approved by the course manager by April 11th. You will then have at least 6 weeks to work on the project until the submission deadline on **May 26th**.

Example data sets and projects

Groups are welcome to collect their own data sets. However, it might be more straightforward to use an existing, publicly available data set. Go here for example data, research, projects, and libraries: KSGEDAS1KU-20231: Exam Resources (itu.dk)

There is no restriction on libraries: Feel free to use whatever works best, as long as you document it.

Project hand-in

You must hand in:

- report.pdf: A project report.
- gitlog.txt: Your repo's git log, e.g. by running: `git log > gitlog.txt`
- If you hand in data analysis or research: code.zip: One zip file containing Jupyter notebook(s) (.ipynb) of your commented code that runs fully without errors, assuming that all used data files are in place, and reproduces your findings. Do not include the data files here.
- If you hand in a prototype of a digital product, do not submit your code, but provide the link to your code repository in the report. Also provide the URL or a way where your prototype can be tried out.

Using a layout like in the provided LaTeX/Overleaf template, not including references and appendices, project reports should be:

- 6-10 pages for groups of size 1
- 8-12 pages for groups of size 2
- 10-14 pages for groups of size 3

Your report should contain *at least* one or more figures that visualize geospatial data.
It is up to you whether you want to use the provided LaTeX template or use another one – but using LaTeX or similar is strongly encouraged.

The report should include the motivation and background for your project, a description of the data and the data processing, your results, a critical discussion of both process and outcome, as well as conclusions and future work (see below for more details).

You are expected to make use of and refer to some of the geospatial concepts and methods we have worked with in the course throughout the report.

You are free to organize your report as you see fit, but the suggested report structure is as follows:

Title or project name, authors

1.Introduction and Goal

Here you provide the context and motivation for the problem. What are your research questions, or what does your prototype intend to do/solve? Explain the particular *spatial* problem you are tackling and its relevance.

2. Problems and Background

Provide context (historical, cultural, technical) for your project and introduce the background and related work in literature (cite or list relevant literature on the research problem; list other scripts and software in this area etc.)

Give an overview of the (cultural, historical, social, technical, etc.) problems to be solved by the project and the role of the different digital tools in achieving the aims of the project

Specify your approach and why you have chosen it.

3. Data acquisition and processing

List and cite all sources of data used in this paper, comment on their fitness for purpose (format, quality, and provenance). Focus on the spatial component of your data, its origin, and precision and reliability. Provide details of data manipulation and transformation. Link to processing scripts where relevant.

4. Results

Provide and explain the results of your investigation, illustrated with figures where essential and relevant. Relate to lessons learnt, counts, statistics, maps or other outcomes.

Briefly comment on 1) the main elements of your digital workflow, highlighting challenges and decision-making bottlenecks (e.g. how did you transform point data to make it into a continuous surface?) 2) functions/tricks you found useful and wish to promote or credit. Remember that the technical tasks should not clutter/interfere with your overall narrative and data analysis (unless your project is about developing a technical pipeline)

For 'technical pipeline' projects: provide a guided tour of your pipeline to facilitate its reproducibility, explaining your choices, clarifying dependencies, and referring to the scripts/tools you compiled in GitHub.

5. Discussion

Evaluate the results in light of the data sources and research premises/assumptions. How representative, reliable, complete and precise are your results? How transferable or generalizable? Give an account on the major short-coming(s) of your methodology / data / prototype. Briefly evaluate the results in light of digital tools, the learning process, time on task, vis-à-vis the final product.

For 'technical pipeline' projects: Provide a comparison with other state-of-the-art data or software if any exists for the same task (kindly cite relevant work, scripts, etc.)

6. Conclusions and future work

Here you provide a short summary of the results of the project, the achieved (or missed) goals and highlight the most important lessons learnt while working on the project. Indicate how

the methods, data, analysis, or prototype functionalities could be improved or extended in future work.

References

At least 5, both domain-based literature and references to digital tutorials or internet resources consulted.

Appendix A – Metadata tables (required)

Table 1 – Software metadata

	Software metadata description	***Please fill in this column***
S1	Current software version	*Example: gds:6.1, Python 3.9*
S2	Permanent link to your code in your Github repository	*Example: https://github.com/mszell/bikenwgrowth/releases/tag/1.0.0*
S3	Legal Software License	*Example: Creative Commons 4.0*
S4	Computing platform / Operating System	*Example: Linux 18.04, OS X, Microsoft Windows 10,...*
S5	Installation requirements & dependencies for software not used in class	
S6	If available Link to software documentation for special software	*Example: http://mozart.github.io/documentation/*
S6	Support email for questions	

Table 2 – Data metadata (use the template below or create your own metadata table)

	Data metadata description	***Please fill in this column (you can link to license and metadata descriptions online; where relevant remember to articulate data provenance and quality)***
D1	Data License	*List the license for your own data, and communicate the license of other used datasets*
D2	Dataset name / main properties	*Example: Geochem_dk.grid. SpatialRaster layer, resolution 30x30m, attributes: "ID" "Elev" "pH" generated by KU Dept of Geochemistry, 2020 interpolated from 5x5km grid sampling, source: www.xxxx.dk*

Appendix B - Contribution statement (mandatory for group projects)

For individualized grading of group projects, here you must clarify who of your group is responsible for the different parts of the project submission. Please state:

For at least 2 sections of your report, who was primary contributor (and optionally also second and/or third contributor)

Who was primary (optionally also second and/or third) contributor for any additional major tasks such as data collection, data preparation, analysis, programming, or visualization.