Assignment 1 – GGS590 GeoAI - Problem Formulation for GeoAI Workflows

Due February 4th 2026

This assignment aims to reinforces the core ideas from this week's class, including:

- Loading and inspecting geospatial data
- Managing coordinate reference systems (CRS)
- Translating geographic questions into testable hypotheses
- Operationalizing those hypotheses as computable workflows
- Reflecting on where human judgment is required in GeoAI

The assignment has five sections to cover, which include the following:

1. Loading and inspecting your data (20%)

Take one of the datasets covered in class, then load in and inspect key attributes of the data. If you prefer, you can use other datasets (e.g., depending on your interests). Make sure you report:

- The number of features
- The geometry types
- The CRS
- Three useful attribute columns which you could use in an analysis (and briefly state why/how)

Now create two processed data layers from your dataset where you manipulate the geometries. You have the flexibility to choose. For example, you might convert linestrings to points every *n* meters, or convert polygons to points/linestrings. Be creative.

2. Research question and hypothesis (20%)

Formulate a geographic research question to be investigated using the datasets at your disposal.

For example, "are country centroids closer to borders in Europe than in Africa?" (please do not copy).

Write one testable hypothesis and state how the hypothesis could be falsified.

3. GeoAI-ready workflow specification (20%)

Create a problem specification that then includes the following, to enable you to test your hypothesis:

- Input layers
- CRS and justification
- Spatial operations
- Outputs
- Assumptions and constraints
- Validation checks

4. Hypothesis-driven spatial analysis (25%)

Implement your workflow in Python, e.g., using the packages we covered in this week's class (GeoPandas, Pandas, NumPy etc.):

- Reproject data to an appropriate projected CRS.

- Perform spatial operations required to test your hypothesis.
- Compute a quantitative result.
- Produce at least one map visualizing the result.

5. Reflection on judgment and limitations (15%)

Provide a description of:

- What parts of this analysis required human judgment rather than coding skill?
- What is one important assumption or proxy in your analysis?
- What is one way your result could be misleading?
- What would you change if you had better data?

Submission format

Please write your analysis in a Google Colab notebook and then use the print function to save it to a .pdf file for submission on Canvas. Without submitting your files like this, you will receive a 50-point penalty to your overall grade (as you need to provide both the code and the answers).

Grading rubric

Component	Points
Part A: Data loading & inspection	20
Part B: Research question & hypothesis	20
Part C: Workflow specification	20
Part D: Spatial analysis & visualization	25

The Mason Honor code applies. Please also remember to follow Mason AI use policy.