HONR 39900 – Midterm Project Guidelines

Justin A. Gould
gould29@purdue.edu

April 9, 2021

DUE DATE: 2021/12/13 23:59 EDT

Final Submission Instructions
To receive credit for the project, do the following:

1. Submit Postgres database (following proper database design principles): Documentation of the database you created including descriptions of your database's tables, the columns within the tables and the relationships between the tables. An ERD and/or screen captures may be effective ways to convey this information.

2. Submit all data used, map(s) you generate, demo videos of application(s) you develop, etc. to Brightspace before the due date.

3. Submit Short (2-3-pages) write-up of justifying your distribution plan and a demonstration of how your database solves the problem, and explaining your analysis approach (e.g., data used, preprocessing, feature engineering, PostGIS queries, etc.).

All details *w.r.t.* project description, expectations, etc. will be outlined on the following pages.

Project Topic and Description: *Topic:*

- Apply Geospatial Analytics and GIS Skills in a Domain of your Choice

Description:
We have covered a lot of content in our course. From map database design principles and basic GIS data plotting to map matching and route optimization, we now have the skills to work on a wide variety of projects in industry and academic research. The intent of this class is to be applied, in the sense that I have taught you things that you will use in industry, should you work on geospatial-related projects. In the spirit of being an applied class, I wanted to give all of you the opportunity to show off your skills and work in a space about which you're excited. It could be related to your major, or even a passion! I am excited to see what you come up with. Please see below for specific guidelines, requirements, and suggestions.

Project Requirements:
While there is no "hard requirement" in terms of what you can do for your project, there are a few requirements:

- You are free to select any topic, domain, scope for your project.
- You are free to use any data, techniques, etc. for your project.
- While you are not required to use PostGIS, you must store your data in a Postgres DB and present the ERD at the time of submission.
- Similarly, you must use at least 2 data sources that you can `JOIN`. For example, if your project is determining where to build ethanol manufacturing facilities, you could use: a shapefile, population data, locations of corn fields, locations of pipelines, E85 consumption, etc.
- The project must be written in Python.
- Your project needs to be novel (e.g., determining where to build a new car dealership considering OEM sales, population, competitor market share, etc.), or provide a novel approach to an existing problem (e.g., how to equitably redistrict congressional district to improve gerrymandering and access to representation).

Deliverables:

1. Postgres database (following proper database design principles): Documentation of the database you created including descriptions of your database's tables, the columns within the tables and the relationships between the tables. An ERD and/or screen captures may be effective ways to convey this information.
2. All data used, map(s) you generate, demo videos of application(s) you develop, etc.
3. Short (2-3-pages) write-up of justifying your distribution plan and a demonstration of how your database solves the problem, and explaining your analysis approach (e.g., data used, preprocessing, feature engineering, PostGIS queries, etc.).
4. 4-5-minute presentation to be completed in class: outlining data you used, approach, map(s), etc.

Available Datasets:
You can use any data you would like, but these data are available for you on our GitHub, to possibly help you get started. If you have trouble finding data, reach out to your instructor.

1. U.S. county geospatial data (Shapefile): `UScounties`
2. Demographic data: `population_data.csv`
3. Shapefile of all U.S. congressional districts of the 114th U.S. Congress: `districts114.zip`
4. US Presidential Election results in 2012 and 2016, by county: `election.zip`

Also visit the following for ideas of datasets:

- https://geodacenter.github.io/data-and-lab/: University of Chicago listing of GIS data
- https://hub.arcgis.com/search: Free GIS datasets from ESRI
- https://sedac.ciesin.columbia.edu/: Socioeconomic data from NASA/Columbia
- https://neo.sci.gsfc.nasa.gov/: Earth-related data from NASA (e.g., atmospheric, environmental, climate, etc.)
- https://datasetsearch.research.google.com/: General dataset search from Google
- https://www.data.gov/: U.S. government's dataset search engine
- https://data.cdc.gov/: Public health data from the U.S. Centers for Disease Control

Grading Scale:
The final project is worth a total of 450 points, or 45% of your final grade. You will earn points based on the following:

- Project Content: 410 points...
  - Postgres DB: 150 points (Acceptable ERD, based on principles discussed in class, and write-up containing requested information)
  - Map(s): 85 points (Quality of map(s): following the map design principles discussed in class)
  - Originality of Project: 50 points
  - Depth and Techniques Used : 100 points (do you use advanced techniques taught throughout the class?)
  - Write-up: 25 points (Quality of write-up, in terms of containing requested information, grammar, spelling, and flow)
- In-class Presentation: 40 points...
  1. Knowledge of Subject Matter (Maximum of 5 points, as outlined in rubric)
  2. Communication Skills/Clarity (Maximum of 5 points, as outlined in rubric)
  3. Poise/Confidence (Maximum of 5 points, as outlined in rubric)
  4. Method of Presentation (Maximum of 5 points, as outlined in rubric)
  5. Voice (Maximum of 5 points, as outlined in rubric)
  6. Visual Contact (Maximum of 5 points, as outlined in rubric)
  7. Evidence of Preparation (Maximum of 5 points, as outlined in rubric)
  8. Orderly Sequence (Maximum of 5 points, as outlined in rubric)