GEP 664 Final Project - Build a Spatial Database

Spring 2019

The final project is worth 20 points.

Due Dates

Presentation Due - Tue May 14th 5:30pm (submit to Box so I can collate before class)

Final Project Due - Tue May 21st 9:30pm (can submit to Box file size permitting)

1 Introduction

For the final project you will identify some publicly available datasets, build a workflow, and create a PostGIS database to answer an interesting spatial question of your choosing. You will also write a brief paper and create a map that describes your database. This assignment will give you the opportunity to use all of the fundamental concepts and skills that you have learned in the course. You may choose any topic and datasets that interest you, but please avoid extraordinarily large datasets (i.e. millions of records).

For example, in my work I was interested in understanding the geographic distribution of public libraries in the United States, to see if there are meaningful differences in library distribution and accessibility across the country. My workflow involved:

1. Identifying and downloading data from two sources (the IMLS and the US Census Bureau)
2. Processing the data to remove unwanted attributes and records, and geocoding some records to improve locational precision
3. Loading shapefiles of census geographies
4. Loading coordinate data of libraries and block group population centroids and building geometry
5. Re-projecting all layers so they shared a spatial reference system that was appropriate for the analysis
6. Using spatial joins to assign block groups the attributes of larger regions in which they were located
7. Using distance functions and proximity analysis to find the closest library to every block group
8. Calculating population-weighted distances and writing GROUP BY statements to summarize the distance and population data for larger geographic areas like states and regions
9. Creating spatial tables and views to map data in QGIS
10. Exporting data out to Excel and Geoda to create additional summaries and generate spatial statistics

2 Deliverables

I will examine all four of your deliverables as individual pieces and as a whole in order to assess whether you met the goals for the assignment. I can only make my assessment based on the materials you submit, so make sure that the database and the paper clearly address all of the criteria.

- Presentation (Due Tue May 14th 5:30pm) - Powerpoint or PDF to the Box
- Project (Due Tue May 21st 9:30pm) - in a ZIP file
  - SQL backup file of your PostGIS database
  - Paper / write up (Doc or PDF format)
  - Map (image or PDF format)

2.1 Presentation (2 points)

On Tue May 14 in class, you will give a brief presentation (5 minutes maximum, strictly enforced) that outlines the basics of your project. You must address:

1. The research question you seek to answer (i.e. the goal of the database)
2. The data that you're going to use
3. Your planned workflow
4. Your planned analysis

Your project does not have to be finished at this stage, but you should have a solid plan in place for executing it. The presentation will give you the opportunity to receive feedback from your classmates and the instructor.

2.2 Database (12 points)

These are minimum requirements - you must do *all* of the following:

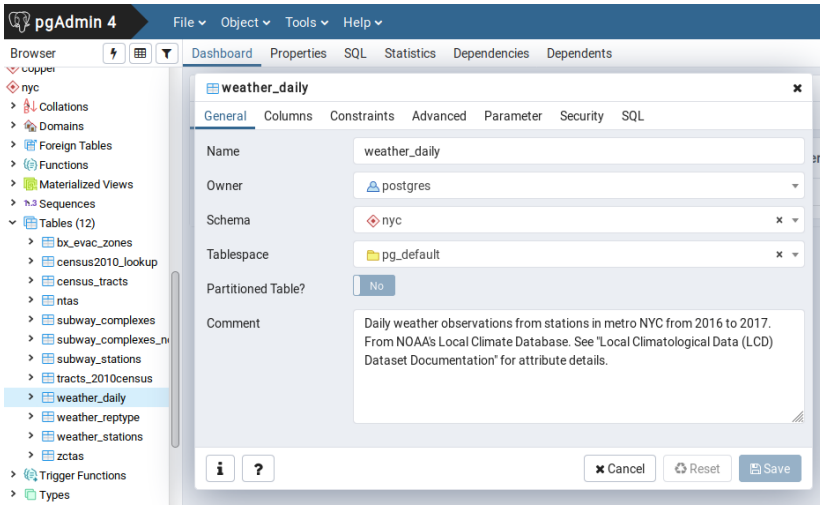
1. All data must be stored in a PostGIS database in a schema named for your project
2. Database must have at least 4 source tables (data that you loaded into the database)
   - At least 3 tables must NOT be part of any of the data that you were given or worked with during the course
   - At least 2 tables must be spatial
   - Use appropriate names for tables and columns, assign proper data types, and designate primary keys
   - Include a few examples of enforcing constraints (with checks, keys, or domains)
3. Create at least 1 view that provides a relevant summary of data
4. Perform a spatial analysis (proximity analysis or spatial selection) within the database
   - Spatial tables must be in a spatial reference system that is appropriate for the geographic area and the goals
   - Data and spatial data should be at the appropriate scale and precision for the analysis
   - If you save the results of your spatial analyses or any derivative data you create as views, I will be able to see the full statements you wrote. If you save the results in new tables, you will need to provide some additional description in your paper so that I know what you did to create the data

In conducting your workflow or doing your spatial analysis as described above, you must demonstrate that you did *at least 3* of the following within the database:

- Build geometry or geography from coordinate-based data
- Transform spatial reference systems
- Perform a spatial join operation
- Summarize (count, sum, avg, etc) spatial features based on their spatial relationship with other features
- Decompose geometry (create centroids, envelopes, dissolve features, calculate intersections, etc)
- Generate spatial measurements (area, length, perimeter, distance, etc)

You should follow 1st normal form; if you break it you must explain why in your write-up. You do not have to follow 2nd or 3rd normal form, but should keep the rules in mind and apply them judiciously. Creating a diagram might be helpful in planning your database, but you do not have to turn one in.

Use PgAdmin to add comments to all tables and views to indicate what they are. Select a table or view, right click, choose Properties, and type the comment.



Your database should be PostGIS enabled (with CREATE EXTENSION postgis;), and you should store all of your objects under one schema for your project. Use postgres as the username and password. Your final database should not contain any temporary or extraneous tables or views.

To create a backup file, select the database in the tree, right click, and choose Backup. Leave the format as custom and set the rolename to postgres. Don't modify or change any of the options on the other tabs. Once the backup file has been created, you should verify that it works. Create a blank, PostGIS enabled database and try restoring the backup file to it. Refer back to the course handout on Accessing PostgreSQL & PostGIS Databases for details on backup and restore.

2.3 Paper (4 points)

Using paragraphs and complete sentences, write a (minimum) three to four page paper that explains the goal of your database and summarizes the process you used to build it. While your write-up must describe the database and discuss the steps you took, you also need to go beyond mere description and discuss your rationale and decision-making process. Address the following points:

1. List the objects in the database with a brief description (this can be a bullet or enumerated list)
2. What was the goal of the database
3. What data did you use and what were the sources
4. Outline your workflow with examples (this can be a bullet or enumerated list)
5. What was your rationale in structuring the database (the tables, their relationships and constraints)
6. How did you conduct the analysis and what was the result

In your workflow (number 4), for each step where you used SQL to accomplish a task, provide *one* sample statement that illustrates what you did. Exceptions: you do *not* have to provide examples of CREATE TABLE for creating empty tables or CREATE VIEW statements, as I can see these in the database.

2.4 Map (2 points)

Provide one sample map that visualizes data directly from your database. The map can: display all the database features, or can be an example of a spatial query, or can show the result of an analysis, or can be a thematic or reference map. Choose whatever you wish, but I must be able to understand what the map depicts based on its contents or some accompanying description. The map should be a formal, finished product saved as a static image (PNG, JPG) or PDF file. Do not submit a project file (qgs or mxd).

3 Checklist

Database	
Created SQL backup file of database AND tested it to insure it loads properly	
Met minimum requirements for creating tables and views	
Conducted a spatial analysis and saved results as views or tables	
Did at least three of the specified tasks as part of your workflow / analysis	
Added comments in PgAdmin to all tables and views	
Paper	
Finished the paper, addressed all six points, and provided the necessary context for understanding the database	
Map	
Created a final, finished map and saved it as a static image or PDF	
Final Submission	
Zipped your SQL backup file, paper, and map into one zip file, renamed the zip file so it is: yourname_gep664_finalproj.zip, and submitted by the due date.	

4 Data Sources

A short (and certainly not exhaustive) list of suggestions for finding free spatial and attribute data.

Geolode : global directory of websites that provide free spatial data
`http://geolode.org/`

Natural Earth : global spatial data for countries
`http://www.naturalearthdata.com/`

UNdata : global attribute data for countries
`http://data.un.org/`

Data.gov : large US repository of spatial and attribute data
`https://www.data.gov/`

American Factfinder : US repository for recent census data
`https://factfinder.census.gov`

Census TIGER files : US national spatial files
`https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html`

NYC Bytes of the Big Apple : NYC Dept City Planning spatial data
`https://www1.nyc.gov/site/planning/data-maps/open-data.page`

NYC Open Data : NYC repository of spatial and attribute data
`http://opendata.cityofnewyork.us/`

Baruch Geoportal : CUNY repository with mix of NYC, national, and global spatial data
`https://www.baruch.cuny.edu/confluence/display/geoportal`