Spatial Database Management (GEP 664 / GEP 380)

Spring 2019

Last revision to syllabus: January 11, 2019

1 The Basics

Meeting: Tue 6pm - 9:30pm, Gillet Hall Room 322 (GISc Lab)

Instructor: Frank Donnelly

Email: francis.donnelly@baruch.cuny.edu

Office: Baruch College, Newman Library. 151 E 25th St, Rooms 951-952

Office phone: 646-312-1657

Office hours: Email for appointment

Course website: `https://spatdb.commons.gc.cuny.edu/`

I'm the Geospatial Data Librarian at Baruch College CUNY. I'm available for appointments at the GIS Lab on Baruch's campus Wednesdays through Fridays. Email me to schedule an appointment.

2 Catalog Description

Spatial Database Management with a focus on managing spatial data within a relational database for use with Geographic Information Systems. In addition to learning relational database concepts and Structured Query Language (SQL), students will learn how to create and manage a spatial database, manage database security, maintain data integrity, model spatial relationships within the database, and work within a multiuser editing environment.

Prerequisite for GEP 664: GEP 505 (Principles of Geographic Information Science)

Prerequisite for GEP 380: GEP 204 (Basics of Mapping) or GEP 205 (Principles of GIS)

It is expected that students in this course have already completed one of these prerequisites. Fundamental knowledge and experience with GIS and basic computer skills are presumed.

3 Learning Goals

After successfully completing this course, you are expected to be able to:

- Understand fundamental relational database concepts
- Understand and execute SQL (Structured Query Language) and spatial SQL
- Manage a spatial database using database management software (PostgreSQL & PostGIS)
- Model relationships and manage data integrity within a spatial database
- Prepare, process, and load data into a database
- Perform spatial analysis in a spatial database and in conjunction with desktop GIS software

4 Structure of the Class

This course meets once a week in the GIS lab for approximately 3.5 hours. Lab assignments from the previous week are due at the beginning of class. The course will typically begin with lecture and discussion, followed by hands-on exercises. For shorter topics we will pivot back and forth between lecture / discussion and exercises. Time will be alloted at the end of class to work on lab assignments.

5 Assessment

Your understanding of the course material will be evaluated through lab and homework assignments, a midterm quiz, a final project, and in-class discussion and participation.

Attendance

Attendance is mandatory. A large portion of your grade is based on lab and homework assignments that we will cover in detail in class. We only have 14 sessions and most of the material is cumulative, as each topic builds on what was previously covered. Missing class will impede your ability to progress in the course, and will be detrimental to your grade in terms of assignments and participation.

With the exception of documented emergencies or illnesses more than two missed classes will result in an automatic F grade. If you miss two or fewer classes, your participation grade will be impacted and you will still be responsible for learning the material (on your own) and turning assignments in on time. If you're not going to be in class I expect that you will notify me in advance. If you have a series of documented emergencies that stretch over an extended period of time, it may be necessary for you to withdraw from the course.

Please come to class on time. Excessive lateness will count against your participation grade.

Grading

Grades will not be curved and no grades will be dropped.

Grade Weighting

- 60% Assignments
- 20% Final Project
- 10% Midterm Quiz
- 10% Participation

Grade Scale (GEP 664)

- A 93-100, A- 90-92
- B+ 87-89, B 83-86, B- 80-82
- C+ 77-79, C 73-76, C- 70-72
- F <70

Grade Scale (GEP 380)

- A 93-100, A- 90-92
- B+ 87-89, B 83-86, B- 80-82
- C+ 77-79, C 73-76, C- 70-72
- D+ 67-69, D 60-66
- F <60

Assignments

Homework will be assigned after most sessions to reinforce what we have covered in class, and to give you the opportunity to work with the concepts and applications so you can learn them. Assignments will constitute the largest portion of your grade.

- Do the reading assigned for a particular class before that class.
- All assignments will be posted on the course website.
- Use the homework template for completing assignments.
- All homework assignments are due at the beginning of class. Assignments not turned in at that time will be penalized 25% from the original total grade for each day the assignment is not handed in:

- Handed in at the beginning of class, max possible grade is 100%
- Not handed in at the beginning and up to 1 day late, max possible grade is 75%
- Two days late, max possible grade is 50%
- Three days late, max possible grade is 25%
- Four days late: 0%, and participation grade is impacted

Midterm Quiz

A short quiz will be given towards the middle of the semester to test your knowledge of key concepts that have been covered thus far. The quiz will consist of short answers and problem solving, where you will be asked to write SQL queries given a table and a question, or re-construct queries given a table and a result.

Final Project

For the final project each student will create a curated database using publicly available data, and in doing so will display competency with the various topics covered in the course that includes: database design, extraction, transformation, and loading, SQL data definition and data manipulation queries, and spatial analysis.

Participation

Asking questions and participating in class discussions is strongly encouraged. Excessive lateness and missing classes and assignments will count against your participation grade.

6 Course Materials

Course website : This syllabus, the class schedule, and links to databases, software, resources, and readings are available on the course website at `https://spatdb.commons.gc.cuny.edu/`.

Textbooks

These textbooks are *required* for this class. We will use both of them extensively for readings and exercises both in class and for homework.

Practical SQL by Anthony DeBarros, No Starch Press, 2018.

PostGIS in Action 2nd edition by Regina Obe and Leo Hsu, Manning Publications, 2015.

Other Readings

I have placed a selection of other required readings on e-reserve at the library. I'll provide you with the password for accessing the materials. The e-reserves page is located here: `http://libguides.lehman.edu/er.php`.

Resources

PostgreSQL Documentation: `https://www.postgresql.org/docs/10/index.html`

SQL Commands Reference: `https://www.postgresql.org/docs/10/sql-commands.html`

PostGIS Documentation: `http://postgis.net/docs/manual-2.4/`

Software

We will primarily be using the PostgreSQL 10 relational database with the PostGIS module and the pgAdmin 4 graphical user interface - for the sake of brevity I'll refer to this collection of software as "PostGIS". The software is free and open source and is available for all operating systems. The software is available in the lab, but since it's free I strongly encourage all students who have their own machines to install it, so you can work on assignments both inside and outside the lab.

I've created a handout that's posted on the course website with instructions for downloading and installing PostGIS. Please follow the instructions closely and install the software within the first week of class. If you have problems make arrangements to speak with me immediately - do not wait until we are a number of weeks into the class. Software installation trouble will not be considered as a valid reason for not handing in assignments.

There are a variety of ways to access a PostGIS database: through a text-based command line program (psql), through a table-based graphic user interface (pgAdmin), and through desktop GIS software. We will experiment with all three. For a desktop GIS package you should install a long term release (LTR) of QGIS: either 3.4 Madeira or 2.18 Las Palmas. QGIS is free, open source, and cross platform, and allows you to write regular and spatial SQL queries directly and see the results. We will *not* be using ArcGIS in this class (although it can be used to connect to PostGIS databases).

Course Databases

We will use several datasets for this class:

1. A sample database called gep664_2019 that I created for this class
2. The tutorial database that comes with our textbook *PostGIS In Action*
3. Tutorial data from our other textbook *Practical SQL*

The gep664 and *PostGIS In Action* databases are available as a back-up or SQL dump files that can be loaded on individual machines, both in the lab as well as on your own laptop or personal computer. Since the database software is open source and the databases are readily available, you should be able to do your work from multiple locations, but will need to have a plan for keeping your data "in synch". Databases are complex structures and can't simply be copied and pasted like regular files. I've created a handout that demonstrates the various tools for connecting to databases and the steps you would take for backing-up and restoring them. The handout and links to the databases are available on the course website.

7 Course Policies

Contacting me: please consult the syllabus, the assignment, slides, readings, software documentation, and your notes prior to emailing me with questions. If you have questions please explain the issue as clearly and thoroughly as possible - screen shots are helpful. For complex issues it's better to make an appointment. Do not expect instantaneous responses to email, particularly late at night or on weekends.

Cell phone use: be considerate: no talking or texting in class. Turn your phones off or to vibrate.

Computer use: during class you will be expected to follow along with lectures, participate in discussion, and work on class exercises. You may not use the lab computers or your personal devices to work on other projects or to engage in leisure activities.

Lab rules: drinking or eating in the lab classroom is not allowed. Please adhere to this and other policies as posted in the lab.

Collaboration: I encourage all members of the class to talk to one another, ask each other questions, and help one another. However, all of the assignments must be completed by you individually and submitted work must be your own.

Academic dishonesty: will not be tolerated. Academic dishonesty includes, but is not limited to: cheating, plagiarizing (including cutting and pasting or paraphrasing information from the internet without proper citation), fabricating information or citations, facilitating acts of academic dishonesty by others, submitting work of another person or papers written for other courses, or tampering with the academic work of other students. For further clarification, please read CUNY's policy on academic integrity in the most recent edition of the graduate or undergraduate bulletin at `http://www.lehman.edu/academics/`.

Students with disabilities: Lehman College is committed to providing access to all programs and curricula to all students. Students with disabilities who may need classroom accommodations are encouraged to register with the Office of Student Disability Services. For more information, please contact the Office of Student Disability Services, Shuster Hall, Room 238, tel #: 718-960-8441.

8 Schedule

Schedule is subject to revision throughout the semester.

In the schedule, "Practical" refers to our textbook *Practical SQL* while "Action" refers to our other text, *PostGIS in Action*. Other readings indicated in parentheses are available on the library's E-reserve page at `http://libguides.lehman.edu/er.php`.

Readings should be completed *before* the class for which they are assigned. Exercises are due at the *beginning* of the following class

Week 1 (Jan 29) - Introduction

Topics: Course overview, technology setup, interfaces, database test drive

Readings: DMS article (Duckham)

Assignment 1: Database installation

Week 2 (Feb 5) - SQL DML

Topics: Database fundamentals, data manipulation language, aggregates, joins

Readings: Practical Introduction, Chap 1 through 4

Optional Reading : SQL data manipulation (Connelly)

Assignment 2: Intro to SQL Part 1

HOLIDAY - No class on Feb 12
Assignment 2 due by 9:30pm

Week 3 (Feb 19) - SQL DDL

Topics: Database fundamentals, data definition language, data types, constraints

Readings: Practical Chap 5 through 8

Optional Reading : SQL data definition (Connelly)

Assignment 3: Intro to SQL Part 2

Week 4 (Feb 26) - Database design

Topics: Entity relationships, normalization, keys, modeling

Readings: Database Design (Kreibich), Entity Relationship Model (Song)

Assignment 4: Design and normalization

Week 5 (Mar 5) - Data processing

Topics: Data formats, processing, loading (ETL)

Readings: Practical Chap 9

Assignment 5: Data cleaning & SQL review

Week 6 (Mar 12) - Spatial databases

Topics: Spatial data fundamentals, spatial data types

Readings: Action preface, Chap 1, 2, & 5

Assignment 6: Spatial data basics

Week 7 (Mar 19) - Spatial reference systems

Topics: General overview, loading spatial data, coordinate transformation

Readings: Action Chap 3, 4, & 6

Assignment 7: Coordinate systems

Week 8 (Mar 26) - Spatial relationships & analysis

Topics: Geographic relationships, indexes, spatial indexes

Readings: Action Chap 9 & 11

Assignment 8: Spatial analysis

Week 9 (Apr 2) - Proximity analysis

MIDTERM QUIZ: At beginning of class

Topics: Buffers, nearest neighbor, coordinate system considerations

Readings: Action Chap 10

Assignment 9: Proximity analysis

Week 10 (Apr 9) - Organizing spatial data

FINAL PROJECT: Assigned

Topics: Spatial data design

Readings: Action Chap 14, Practical Chap 18,

Assignment 10: Spatial data design

Week 11 (Apr 16) - Database management, SQL DCL

Topics: Users and permissions, data control language, maintenance and performance

Readings: Practical Chap 15 & 17

Optional Reading : SQL data definition (Connelly)

SPRING BREAK - No class on Apr 23

Week 12 (Apr 30) - Rasters / Other database formats

Topics: Rasters, SQLite / Spatialite, ArcGIS formats

Readings: Action Chap 7

Week 13 (May 7) - Database programming / Final project prep

Topics: Python for databases, final project help

Week 14 (May 14) - Presentations / Final project prep

FINAL PROJECT: Presentations

Week 15 (May 21) - Exam Day - Class does NOT meet

FINAL PROJECT: Due by 9:30pm