11.523: Fundamentals of Spatial Database Management

Eric Robsky Huntley
Massachusetts Institute of Technology
ehuntley@mit.edu

Lab: Tu, 4-7pm, 9-554 | Lecture: Th, 5-6:30pm, 9-450A

Description

Advances in urban science, the rise of “big data”, the drive to build smarter cities, and the widespread embrace of the open data movement are coalescing into new opportunities for planners for make data actionable through analysis and visualization. Even as these contemporary discourses and innovations proceed apace, urban archives are richer and more available than ever before. This poses challenges to planning, as a community of practice: to be more contextual, even as naive empiricism becomes ever-more-tempting; to be more historical, even as the present demands ever more of our attention.

Indeed, data produced and distributed in a vacuum is worthless, and worse: it might lead us to think that evidence can be divorced from its place and context. As such, we will strive to produce contextually rich and well-situated datasets that are responsive to the needs of local stakeholders. We will be working with the Leventhal Center for Maps & Education at the Boston Public Library to create historical GIS models of Boston-area maps and atlases, as well as research into the provenance of those same data sources.
What Will We be Learning?

Students will develop the technical skills necessary to design, build, and interact with spatial databases using PostgreSQL, PostGIS, and the SQL query language. Students will also learn to write data biographies and highly contextual metadata. As such, students should emerge from the class able to perform basic database maintenance and data collection tasks, as well as able to critically evaluate the stories of data maintained and collected.

How Will We be Learning?

This is a data-based (!) class in which our methods will be empirical and our tools will be computational. As such, much of our time will be spent in the weeds, learning how to design and ask questions of spatial databases and how to build models of urban environments. However: my hope is that no one will be intimidated be-
cause they’re not “data scientists”. To the extent that it is possible, I will be seeking to make these methods approachable and accessible.

**Are There Prerequisites?**

Again, accessibility is a priority. However, it is best to proceed with a baseline of shared knowledge. MCP students hoping to enroll in this course must have taken 11.205: Introduction to Spatial Analysis. We will assume a familiarity with GIS fundamentals (projections, geoprocessing operations) and introductory spatial analysis. For most undergraduates, general MIT requirements supplemented with 11.188 will prove sufficient. Students from other institutions hoping to cross-register should have experience with GIS.

**Am I Required to Buy the Texts?**

No! All readings are uploaded to the course Stellar site. In fact, many, if not all, of the texts are available digitally to MIT affiliates through the libraries. Just remember this favor when you ask yourself how much of the reading to complete…

Also, I recognize that the preponderance of these readings are by men; this is, of course, a problem. To cite two of my colleagues, Carrie Mott and Daniel Cockayne: “the choices we make about whom to cite – and who is then left out of the conversation – directly impact the cultivation of a rich and diverse discipline, and the reproduction of geographical knowledge itself. To cite narrowly… does a disservice not only to researchers and writers who are othered by white heteromasculinism, but also to the prevailing impression of geography upon those who may be less familiar with the discipline, most notably, our students” (Mott and Cockayne 2017, 955). In future iterations of this class, I would like my primary texts to do more to intervene in the field and to challenge dominant representations of “GIScience” and its practitioners… for now, mea culpa, and as I locate other resources I will do my best to call your attention to them.

**Assessment and Assignments**

**Assessment**

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly Exercises</td>
<td>10%</td>
</tr>
</tbody>
</table>
### Assignment Weight

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Model</td>
<td>20%</td>
</tr>
<tr>
<td>Data Model Revisions</td>
<td>10%</td>
</tr>
<tr>
<td>Data Biography</td>
<td>20%</td>
</tr>
<tr>
<td>Data Biography Revisions</td>
<td>10%</td>
</tr>
<tr>
<td>Digitized Atlas Sheet(s)</td>
<td>20%</td>
</tr>
<tr>
<td>Contribution to Mapathons</td>
<td>10%</td>
</tr>
</tbody>
</table>

### Weekly Exercises

- Due: Weekly on Thursday at 8:00am.

These are straightforward exercises intended to keep you caught up with the material. They should not take you more than an hour to complete and will be tightly coupled to each week's readings.

### Data Biography

- Draft Due: Tuesday, 18 February at 8:00am.
- Revision: Tuesday, 25 February at 8:00am.

Data (and the maps for which they form the basis) do not emerge from thin air. They are collected for specific purposes, by specific actors, under specific historical conditions. Data have histories, which we’ll take the liberty of calling biographies. So, in addition to building a data model to represent the information available in your maps, you will have to tell us about the maps as data sources, doing the hard work of contextualizing and historicizing it for us. You will then revise a draft of this biography in response to feedback, and present it to the class.

### Data Model

- Draft Due: Tuesday, 3 March at 8:00am.
- Implementation Due: Tuesday, 10 March at 8:00am.

One of the key skills that this class will be focusing on is planning and implementing object-relational data models. Such a model for an neighborhood might include separate tables describing neighborhood boundaries, street centerlines, cadastral (i.e., parcel) boundaries, building footprints, property owners, real estate
transactions, tree locations, utility line locations… you can see how these models can get very, very detailed! In order to build a historical database you will have to make non-trivial decisions about, for example, what features to include and what attributes of those features are salient. Before undergoing the technical work of implementing a model, you will be expected to design the model. You will then revise it in response to feedback.

Digitized Atlas Sheet(s)

- Tuesday, 10 March at 8:00am.

You will be expected to populate your database with vector data based on features depicted on at least one (1) map sheet—this process is generally called “digitizing”. Depending on how densely populated the map sheet is with features, you may be asked to digitize more than one sheet.

Contribution to Mapathons

- Events will be week of 16 March - 20 March.

Once you’ve gone through the exercise of digitizing an Atlas sheet, it will be time to lead others in doing the same! We’ll be convening public mapathons that will encourage people to learn about the history of their neighborhoods; these will hopefully have the added benefit of populating the database much more quickly than would be possible if we were all digitizing alone.

Attendance

For a very long time, I was against taking attendance. But here’s the thing… this course will be a collective effort, and the quality of our work together might be seriously taxed by mid-semester apathy. So: feel free to take one unexcused absences. Beyond that, each unexcused absence will result in a stepwise reduction in your final grade (e.g., A becomes A-). Please note, though, that an absence does not imply that I must provide you with a walkthrough of the happenings of the day!

Office Hours

Office hours will be set together in the first week of class.
I find it very helpful if you book sessions in advance through the DUSP office hours portal\(^1\), though this is not absolutely mandatory. No time slot can possibly please everyone. As such, if this time does not work for you, we can make arrangements to meet at another time. However, please be conscientious! I set this time aside each week for office hours and I really do try to manage my time.

**Email**

I reserve the right to take up to 24 hours to respond to your emails during the week. I will respond to emails sent after 5 pm the following day. I do not respond to emails on Saturday and do so on Sunday only at my discretion. The labor movement fought long and hard to secure your weekend! It is truly remarkable that we live in a world in which an email can travel to space and divebomb from the exosphere to our pockets in a matter of seconds; this does not imply that our response must be equally instantaneous.

**Lateness Policy**

Turning in assignments promptly is important both for keeping current with the subject matter, which is cumulative, and to keep all students on a relatively level playing field. A late assignment will be accepted up until one week after the original due date for a loss of one letter grade (e.g., an A becomes an A-). After that point, late assignments will receive no credit and will not be accepted.

**On Exceptions...**

There will of course be situations where I am willing to bend these rules. These are regulatory ideals, not absolutes. I am a human being; you are human beings. The contingencies and exigencies that condition your life are real. I will strive to respect them if you similarly strive to not take advantage of my inclination to respect them. We’re all in this weird boat together—let’s be good to each other.

**A Note on Graduate School and Mental Health**

Academic environments are taxing places. For reasons structural, institutional, financial, and interpersonal, they do not always lend themselves to what most reasonable people would think of as human flourishing. I went to graduate school. In

\(^1\)https://dusp.mit.edu/oh
fact, I went to graduate school twice. I also went to college. Without dwelling on the issue, I will say that I am intimately familiar with the toll that institutions of higher education can exact on our mental health and wellbeing.

I have two points here: 1) MIT offers a range of counseling and mental health resources\(^2\) for students. I would really encourage you to be proactive about taking advantage of them; and 2) do not hesitate to let me know if you’re struggling. It is not my intention to mine for the details of your private lives! It is only to let you know that I am sensitive to the distinctive difficulties of the environment we inhabit and that help is available.

Schedule

W01: Welcome!

There is no lab this week (Tuesday, 4 February).

In lecture this week (Thursday, 6 February), we will discussing my expectations for you this semester, and what you can expect of me. We will also be doing a high-level survey of the historical work we will be undertaking this semester and why it is an excellent venue in which to develop spatial data modeling skills.

Assigned

- Weekly Exercise 1: Thursday, 13 February at 8:00am.

W02: "Unbinding the Atlas"

In lab this week (Tuesday, 11 February), we will be setting up our machines with software we’ll need for the semester. This means installing PostgreSQL, PostGIS, and QGIS, as well as learning to load and perform basic queries on data using Structured Query Language (SQL). You may not know it, but if you’ve taken an intro to GIS class, you’ve used SQL!

In lecture this week (Thursday, 13 February), we will be covering our project for the half-semester—digitizing atlases in the holdings of the Boston Public Library—and discussing precedents. We will also be discussing the why and the how of what some have begun to call data biographies, or contextually rich metadata.

\(^2\)https://medical.mit.edu/services/mental-health-counseling
Deadlines

• Weekly Exercise 1: Thursday, 13 February at 8:00am.

Readings

On digitizing historical atlases…


On data biographies…


Assigned

• Weekly Exercise 2: Thursday, 20 February at 8:00am.
• Data Biography Draft: Tuesday, 18 February at 8:00am.

W03: Data Models

In lecture this week (Thursday, 20 February), we will be discussing data models—basically, how to think about the world like a database. We will also be considering
how problems of data modeling become political by examining Monica Stephens’s excellent case study of OpenStreetMap.

There is no lab this week (Tuesday, 18 February) due to the Presidents’ Day holiday—the Institute will run on a Monday schedule on Tuesday. However, you should expect to spend some time outside of class building on the SQL exercises we began last week.

Deadlines

- Weekly Exercise 2: Thursday, 20 February at 8:00am.
- Data Biography Draft: Tuesday, 18 February at 8:00am.

Readings


Assigned

- Weekly Exercise 3: Thursday, 27 February at 8:00am.
- Data Biography Revisions: Tuesday, 25 February at 8:00am.

W04: Spatial Data Models

In lab this week (Tuesday, 25 February), we will be learning how to perform spatial queries using PostgreSQL’s PostGIS extension. This extension allows us to ask spatial questions—questions of adjacency, proximity, intersection—within our SQL queries. This is where spatial databases shine! We can do entire GIS projects in relatively few lines of easily interpretable code.

In lecture this week (Thursday, 27 February), we will be covering what makes spatial data models distinctive and how to think through the problems of collecting and representing spatial data. We will also be editing the OpenStreetMap database in order to better understand some of these questions.
Deadlines

- Weekly Exercise 3: Thursday, 27 February at 8:00am.
- Data Biography Revisions: Tuesday, 25 February at 8:00am.

Readings


Assigned

- Weekly Exercise 4: Thursday, 5 March at 8:00am.
- Data Model Draft: Tuesday, 3 March at 8:00am.
- Mapathon Planning Roles: Events will be week of 16 March - 20 March.

W05: Implementing Models

In lab this week (Tuesday, 3 March), we will be implementing our models using both raw SQL code and GUI interfaces for PostGIS. This will lead directly into next week’s exercise—collecting spatial data from your atlas.

In lecture this week (Thursday, 5 March), we will be examining several influential data schema. The model developed for the New York Public Library’s Space/Time directory, the model developed by OpenStreetMap contributors, and a specification, recently released by Apple, that allows for the representation of interior spaces. (!!!)

Deadlines

- Weekly Exercise 4: Thursday, 5 March at 8:00am.
- Data Model Draft: Tuesday, 3 March at 8:00am.
• Mapathon Planning Check-In: Events will be week of 16 March - 20 March.

Readings


Assigned

• Weekly Exercise 5: Thursday, 12 March at 8:00am.
• Data Model Implemented: Tuesday, 10 March at 8:00am.

W06: Digitizing and Implementation

In lab this week (Tuesday, 10 March), we will be covering how to work with desktop GIS tools to create rich spatial and attribute data stored in a PostGIS database.

In lecture this week (Thursday, 12 March), we’ll be continuing to work on digitizing our map sheets. This will serve to pre-populate our database in advance of the mapathons.

Deadlines

• Weekly Exercise 5: Thursday, 12 March at 8:00am.
• Data Model Implemented: Tuesday, 10 March at 8:00am.
• Mapathon Planning Check-In: Events will be week of 16 March - 20 March.

Readings

No readings this week.

Assigned

• Digitize Atlas Sheet: Tuesday, 10 March at 8:00am.
W07: Map!

In *lab this week* (Tuesday, 17 March), we will be covering how to make backups of PostGIS databases—as stable and secure as these are, they still live on mere mortal hard drives.

In lieu of a *lecture this week* (Thursday, 19 March), we will be meeting at Eric’s home in Somerville for a home-cooked meal celebrating your hard work.

*Deadlines*

- Digitized Atlas Sheet(s): Tuesday, 17 March at 8:00am.
- Mapathons! Throughout the week, as appropriate.