February 1, 2022

11.523: Fundamentals of Spatial Database Management

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

- Lecture: Tu, 4–6:00pm
- Lab: Th, 4–6:00pm

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 to develop varied representations of urban environments requiring large quantities of data. 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. Technically, it also implies that planners will benefit from a familiarity with formal spatial databases and query languages, including SQL.

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 historically-situated datasets.

Figure 1: MIT and its surroundings, from the Atlas of the City of Cambridge *of 1916, published by G.W. Bromley and Co. of Philadelphia.*

Using materials from the Leventhal Map & Education Center at the Boston Public Library, we will be 1) creating historical GIS models of Boston-area maps—including recently digitized materials produced by the Boston Redevelopment Authority during and after the height of urban renewal, zoning maps, and fire insurance atlases; and 2) building standardized database models for zoning maps across Massachusetts municipalities.

What Will We be Learning?

Students will develop the technical skills necessary to design, build, and interact with spatial databases using PostgreSQL, PostGIS, QGIS, and the SQL query language. Students will also learn to write highly contextual metadata ('data biographies'). Students will be prepared to perform database maintenance, modeling, and digitizing tasks, and to critically evaluate and document data sources using rich metadata.

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, we will also be working with archival documents, doing the more conceptual work of devising ways of making those documents speak to each other. Finally: my hope is that no one will be intimidated because 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, we will be proceeding assuming a baseline of shared knowledge. We will assume a familiarity with GIS and spatial analysis fundamentals (projections, spatial data types, geoprocessing operations). MCP students hoping to enroll in this course must have taken `11.205`: Introduction to 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…

Assessment and Assignments

Assessment

Assignment	Weight	Due
Weekly Exercises	20%	Weekly
Digitized Atlas	30%	3 March
Digitized Zoning Map	30%	17 March
Data Biographies	10%	17 March
Attendance	10%	-

Weekly Exercises (30%)

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

Digitized Atlas (30%)

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! However, the more detailed a spatial database, the more difficult its upkeep. There is always a trade-off between complexity and tractability! 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. Subsequently, you will be expected to populate your database with vector data based on features depicted in maps—this process is generally called 'digitizing'—as well as georeference the maps themselves.

Data Biographies (for both above projects)

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 with specific subject positions, under specific historical conditions. Data have histories, which we'll take the liberty (following Heather Krause) of calling these '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 work of contextualizing and historicizing it for us.

Attendance (10%)

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. As such, I will be taking attendance in lectures and labs. That said, these continue to be extraordinary circumstances—I

recognize that there may be a variety of reasons for you to miss a lecture or lab session.

If you are unable to attend the regularly scheduled lectures and labs for an extended period of time, let me know as soon as possible—I'll do our best to accommodate your situation.

Why Attendance and Not Participation?

Great question! 'Participation' is a vague category that allows instructors to evaluate students based on unstated and tacit criteria—it's also a nasty way to let various biases into one's evaluation of students. For this reason, I use attendance. We all know what it means!

Late 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 B). After that point, late assignments will receive no credit and will not be accepted.

There will of course be situations where I am willing to bend this rule. This is particularly true now. 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. I also feel that it is important to say that I promise to greet you with the assumption of your honesty.

Office Hours

You can find me in office hours at…

- 9–10:30 AM on Tuesdays and Thursdays and
- 4–6 PM Eastern on Wednesdays.

I find it very helpful if you book sessions in advance through the Calendly[1] application, though this is not absolutely mandatory. If the scheduled time does

[1] https://calendly.com/robskyhuntley/office-hours

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.

Communication

We have seen an explosion of platforms in use by educators for channeling class-related communication. I will be sticking to email because, honestly, Slack messages piling up gives me indigestion. If you want to chat amongst yourselves, I am happy to set up a Slack/Teams/Discord/whatever channel. But I will not be checking it.

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 5pm in the morning on 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.

On 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 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.

[2] https://medical.mit.edu/services/mental-health-counseling

Graduate Students: GradSupport

As a graduate student, a variety of issues may impact your academic career including funding, faculty/student relationships, and interpersonal concerns. In the Office of Graduate Education (OGE), GradSupport provides consultation, coaching, and advocacy to graduate students on matters related to academic and life challenges. If you are dealing with an issue that is impacting your ability to attend class, complete work, or take an exam, you may contact GradSupport by email at gradsupport@mit.edu or via phone at (617) 253-4860.

Special Accommodations for Students with Disabilities

If you need disability-related accommodations, I encourage you to meet with me early in the semester. If you have not yet been approved for accommodations, please contact Student Disability Services at sds-all@mit.edu. I look forward to working with you to assist you with your accommodations!

Inclusive Classroom

MIT values an inclusive environment. I hope to foster a sense of community in this classroom and consider this classroom to be a place where you will be treated with respect. I welcome individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations – and other visible and nonvisible differences. All members of this class are expected to contribute to a respectful, welcoming, and inclusive environment for every other member of the class. If this standard is not being upheld, please feel free to speak with me.

Indigenous Land Acknowledgement

> We acknowledge Indigenous Peoples as the traditional stewards of the land, and the enduring relationship that exists between them and their traditional territories. The lands which MIT occupies are the traditional unceded territories of the Wampanoag Nation and the Massachusett Peoples. We acknowledge the painful history of genocide and forced occupation of these territories, as well as the ongoing processes of colonialism and dispossession in which we and our institution are implicated. Beyond the stolen territory

> which we physically occupy, MIT has long profited from the sale of federal lands granted by the Morrill Act, territories stolen from 82 Tribes including the Greater and Little Osage, Chippewa, and Omaha Peoples. As we honor and respect the many diverse Indigenous people connected to this land from time immemorial, we seek to Indigenize our institution and the field of planning, offer space, admit more Native students, hire more Indigenous faculty, and work together towards the visionary, power-shifting project of decolonization: Land Back. –-Adapted from the MIT Land Acknowledgement by Catherine D'Ignazio, Eric Robsky Huntley, Janelle Knox-Hayes and DUSP faculty, and previously building off work of MIT's American Indian Science and Engineering Society (AISES), Native American Student Association (NASA) and other Indigenous MIT students/alumni. For more information, contact aises-request@mit.edu or nasa-exec@mit.edu.

Schedule

`W01`: Welcome!

In *lecture this week* (Tuesday, 1 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 work we will be undertaking this semester and why insurance and zoning maps are excellent venues in which to develop spatial data modeling skills.

In *lab this week* (Thursday, 4 February), we will get our computing environments set up for the semester; this means installing software (QGIS, PostgreSQL, PostGIS) and making sure it is working smoothly. We'll also be learning to perform basic queries on data using **S**tructured **Q**uery **L**anguage (SQL). You may not know it, but if you've taken an intro to GIS class, you've used SQL!

Assigned

- `Weekly Exercise 1`: Thursday, 9 February at 12:00pm.

`W02`: 'Unbinding the Atlas'

In *lecture this week* (Tuesday, 8 February), we will ground ourselves in questions of how data are produced, collected, and documented… and how they might be

otherwise. We'll also be introducing a couple of key examples of digitized and georeferenced geodata.

In *lab this week* (Thursday, 10 February), we will be continuing to build on the SQL queries we introduced last class. Namely, we will introduce table joins, as well as grouping and aggregating operations.

Deadlines

- `Weekly Exercise 1`: Thursday, 10 February at 12:00pm.

Readings

- Lipton, Belle. 2019. "Opening access to historical urban atlases of Boston." North American Cartographic Information Society Conference. Tacoma, WA. https://www.youtube.com/watch?v=y6wolT1jbso.
- Knutzen, Matthew Allen. 2013. "Unbinding the Atlas: Moving the NYPL Map Collection Beyond Digitization." *Journal of Map & Geography Libraries* 9 (1–2): 8–24. https://doi.org/10.1080/15420353.2012.726204.
- Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2020. "Datasheets for Datasets." ArXiv:1803.09010 [Cs], January. http://arxiv.org/abs/1803.09010.
- D'Ignazio, Catherine, and Lauren Klein. 2020. "Chapter Eight: Teach Data Like an Intersectional Feminist!"[3] In *Data Feminism*. Cambridge, MA: The MIT Press.
- Krause, Heather. 2019. "An Introduction to the Data Biography." *We All Count*. January 21. https://weallcount.com/2019/01/21/an-introduction-to-the-data-biography/.

Assigned

- `Weekly Exercise 2`: Thursday, 17 February at 12:00pm.

`W03`: Data Models

In *lecture this week* (Tuesday, 15 February), we will have a visit by folks from the Leventhal Map and Education Center at the Boston Public Library. They'll discuss

[3] https://data-feminism.mitpress.mit.edu/pub/czq9dfs5/release/2

both the collection we'll be working with and how they're thinking about problems of critical data contextualization. We will be discussing data models—basically, how to think about the world like a database. We will be also introducing a series of questions around classification and categorization that will guide how we think about models not just as technical artifacts but as entities with social and political lives.

In *lab this week* (Thursday, 17 February), we will conclude our introduction to non-spatial SQL by examining subqueries, how to create tables using SQL, and how to populate databases by importing data stored in other formats.

Deadlines

- `Weekly Exercise 2`: Thursday, 17 February at 12:00pm.

Readings

- Bowker, Geoffrey and Susan Leigh Starr. 1999. "Introduction: To Classify is Human." In *Sorting Things Out: Classification and Its Consequences*, 1-32. Cambridge, MA: The MIT Press.
- D'Ignazio, Catherine and Lauren Klein. 2020. "What Gets Counted Counts."[4] In *Data Feminism*. Cambridge, MA: The MIT Press. Links to an external site.
- Yeung, Albert K. W., and G. Brent Hall. 2007. "Database Models and Data Modeling." In *Spatial Database Systems: Design, Implementation, and Project Management*, 55–92. The GeoJournal Library 87. Dordrecht, The Netherlands: Springer.

Assigned

- `Weekly Exercise 3`: Thursday, 24 February at 12:00pm.
- `Digitized Atlas Sheet`: Thursday 3 March at 12:00pm.

`W04`: Spatial Data Models

There will be *no lecture period this week* (Tuesday, 22 February), due to MIT's operating on a Monday schedule this Tuesday.

[4] https://data-feminism.mitpress.mit.edu/pub/h1w0nbqp/release/2

In *lab this week* (Thursday, 24 February), we will be making up some time. I will lecture on what makes spatial data models distinctive and how to think through the problems of collecting and representing spatial data. We will also begin 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.

Deadlines

- `Weekly Exercise 3`: Thursday, 24 February at 12:00pm.

Readings

- Yeung, Albert K. W., and G. Brent Hall. 2007. "Spatial Data and Spatial Database Systems." In *Spatial Database Systems: Design, Implementation, and Project Management*, 93–128. The GeoJournal Library 87. Dordrecht, The Netherlands: Springer.
- Schuurman, Nadine. 2005. "Social Perspectives on Semantic Interoperability: Constraints on Geographical Knowledge from a Data Perspective." *Cartographica: The International Journal for Geographic Information and Geovisualization* 40 (4): 47–61.
- (Skim!) Herring, John R., ed. 2011. "OpenGIS® Implementation Standard for Geographic Information - Simple Feature Access - Part 1: Common Architecture." Open Geospatial Consortium Inc.

Assigned

- `Weekly Exercise 4`: Thursday, 3 March at 8:00am.

`W05`: Well-Known Ontologies

In *lecture this week* (Thursday, 1 March), we will be examining several influential data schema. We will also be welcoming a visitor to talk about the problems of working between municipal zoning maps. We'll also be discussing Monica Stephens's well-known piece on the representation of gendered spaces on OSM, which shows us exactly how dramatically gendered participation in modeling can affect models and ontologies.

In *lab this week* (Thursday, 3 March), we will be continuing our work on spatial queries and learning to implement our models using both raw `SQL` code and GUI interfaces for PostGIS. We will also be covering how to work with QGIS tools to create rich spatial and attribute data stored in a PostGIS database. This will lead directly into next week's exercise—collecting spatial data from your plan or map artifacts.

Deadlines

- `Weekly Exercise 4`: Thursday, 3 March at 8:00am.
- `Digitized Zoning Map`: Thursday, 17 March at 12:00pm.

Readings

- Bronin, Sara C. "How to Make a Zoning Atlas: A Methodology for Translating and Standardizing District-Specific Regulations." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, December 30, 2021. https://doi.org/10.2139/ssrn.3996609.
- The OpenStreetMap Map Features documentation. "Map Features - OpenStreetMap Wiki." n.d. Accessed February 3, 2020. https://wiki.openstreetmap.org/wiki/Map_Features.
- The New York Public Library's Space/Time directory schema. "NYC Space/Time Directory." n.d. Accessed February 3, 2020. http://spacetime.nypl.org/#data.
- Apple's Indoor Mapping Data Format. "Home - Indoor Mapping Data Format (1.0.0.Rc.1)." n.d. Accessed February 3, 2020. https://register.apple.com/resources/imdf/.
- Stephens, Monica. 2013. "Gender and the GeoWeb: Divisions in the Production of User-Generated Cartographic Information." *GeoJournal* 78 (6): 981–96.

Assigned

- `Weekly Exercise 5`: Thursday, 10 March at 12:00pm.

`W06`: Implementing Our Models

In lieu of a formal *lecture this week* (Tuesday, 8 March), we will be working on our digitizing projects.

In *lab this week*, (Thursday, 10 March), we will be learning how to stand up and deploy PostGIS databases on servers.

Deadlines

- `Weekly Exercise 5`: Thursday, 10 March at 12:00pm.

`W07`: Map!

In lieu of a *lecture this week*, we will be learning how to stand up and deploy tile-servers.

In *lab this week*, we will be sharing our results!

Deadlines

- `Digitized Zoning Map`: Thursday, 17 March at 12:00pm.