

February 2, 2021

11.524: Spatial Statistics Workshop

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

- Lecture: Tu, 2:30-4pm
- Lab: Th, 2:30-4pm
- Zoom¹

Description

The broad availability of spatial data on and in cities means that planners can paint pictures of both ‘what is where’ and ‘what was where, when’ with an unprecedented level of detail. However, ‘where’ questions often produce more questions than answers. Maps are evocative, but they are unable to answer questions that are crucially important to planners: how are phenomena interrelated, clustered, and interdependent? Spatial statistics offer one analytical approach for getting at these complex questions that are often key to understanding urban environments.

Even as these methods enhance the analytical repertoire of urban planners, we must remember that methods don’t come from nowhere; they cannot be fully divorced from the conditions of their emergence. Even methods are placed; they come from somewhere. As such, we will be digging into the history of spatial statistics methods and approaches, producing biographies of our methods.

¹<https://mit.zoom.us/j/97345722035?pwd=ZEVjbnhpRHVHWkYwa3Jwcis5Q0lDUT09>



Figure 1: Geoff Dutton's American Graph Fleeting (1979), an animated map hologram produced at the Harvard Laboratory for Computer Graphics and Spatial Analysis. See it in action in a video posted to YouTube by Keith Clarke, <https://www.youtube.com/watch?v=t1160bBcmbA>.

What Will We be Learning?

Students will develop the technical skills necessary to ask spatial questions, statistically. We will be covering spatial autocorrelation (including local indicators), interpolation and kernel density methods, and spatial regression. Students will also learn to *communicate* about spatial statistics, producing illustrated guides to individual methods. As such, students should emerge from the class able to perform sophisticated spatial statistical analyses *and* able to contextualize specific methods.

How Will We be Learning?

This is a quantitative class in which our methods will be empirical and statistical; our tools will be computational. As such, much of our time will be spent in the weeds, learning how to build models that take space into account, and ask spatial questions, statistically. However: 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.

We will be using freely accessible software. We will be using entirely free and open source software. QGIS 3.10 will be our GIS of choice. Our spatial statistics work will be carried out in the R statistical computing environment. These tools will be sufficient to carry out almost any spatial analysis task... for free!

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 or tested out of 11.205: Introduction to Spatial Analysis and 11.220: Quantitative Reasoning and Statistical Methods for Planning (or its equivalent). We will assume a familiarity with GIS fundamentals (projections, overlay operations), spatial analysis, basic inferential statistics, and linear modeling. For most undergraduates, general MIT requirements will prove sufficient. Students from other institutions hoping to cross-register should have experience with GIS, quantitative methods, and will find past exposure to the R statistical computing environment beneficial.

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
Weekly Exercises	50%
Demystification Guide	20%

Assignment	Weight
Project	20%
Attendance	10%

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. **You will be permitted to drop a single exercise.**

Demystification Guides

- Due: Tuesday, 6 May at 8:00am.

Statistics have a nasty way of making things less lucid even as they are intended to produce clear understandings of the world. This is partially due to the opaque, mathematical language generally used to describe methods and results, and to ensure their rigor. The cynic might also say that statisticians and data scientists—who command high salaries due to their mastery of an arcane tongue—are not incentivized to make their approaches transparent. Your challenge, should you choose to accept it, is to take one of the methods we cover this semester and make it as radically accessible as—with drawings, with pithy and clear text. You will then revise this demystification guide based on feedback from myself and your classmates; you will also have to provide feedback! The product will be included in a volume of demystifiers that will become an asset to future students.

Project

- Due: Tuesday, 20 May at 8:00am.

For your final assignment in this course, you will be applying one method covered in the course to an area of inquiry that is of interest to you. However, the outcome of this project is not a conventional paper or presentation. Instead, you will produce an R markdown notebook that thoroughly documents and explains your approach.

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, this is an extraordinary year. I recognize that there may be a variety of reasons for you to miss a lecture or lab session. I am recording the class lectures so that you can catch up asynchronously when necessary. If you need to in. If you are unable to attend the regularly scheduled lectures and labs for an extended period of time (due to, say, a 12-hour time difference between yourself and the east coast of the United States), let me know as soon as possible—I'll do our best to accommodate your situation.

Deadlines

- Weekly Exercise 4: Thursday, 6 May at 8:00am.

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.

Lecture Recording

To accommodate folks who may have trouble joining class at the usual time, I will be recording class sessions and making them available. To ensure that everyone is as comfortable as possible, however, I intend to prioritize student and instructor privacy. In practice this means a few things:

1. The recordings will not be made publicly available; access will be limited to members of the class.
2. Any student can ask for us to pause the recording at any time, no questions asked.
3. Students are not required to have their camera on during class time.

Office Hours

You can find me in a Zoom room between...

- 9:30-11:10 AM Eastern (UTC-04:00) on Tuesdays and Fridays and
- 4:30-6:10 PM Eastern (UTC-04:00) on Thursdays.

I find it very helpful if you book sessions in advance through the Calendly² application, though this is not absolutely mandatory. We're all suffering from Zoom fatigue to varying degrees, so make it pleasant/fun, whatever that means for you! Bring a cup of tea. Wear a funny hat. Pet an animal. If the scheduled 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.

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.

²<https://calendly.com/robkyhuntley/office-hours>

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.

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 resources³ 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.

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

³<https://medical.mit.edu/services/mental-health-counseling>

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

“MIT acknowledges Indigenous Peoples as the traditional stewards of the land, and the enduring relationship that exists between them and their traditional territories. The land on which we sit is the traditional unceded territory of the Wampanoag Nation. We acknowledge the painful history of genocide and forced occupation of their territory, and we honor and respect the many diverse indigenous people connected to this land on which we gather from time immemorial.” (Statement developed by the MIT Indigenous Peoples Advocacy Committee (IPAC) in partnership with MIT’s American Indian Science and Engineering Society (AISES), Native American Student Association (NASA) and other Indigenous MIT students/alumni.)

Schedule

W01: What is Special about Spatial?

In *lecture this week* (Tuesday, 6 April), we will be introducing the set of problems we will be confronting for the rest of the semester. Observations of spatial phenomena can almost never be assumed to be independent—they’re autocorrelated. Spatial phenomena, aggregated by areal unit, are highly sensitive to the chosen unit of enumeration—areas are modifiable. It can be tempting to make inferences from aggregate units to disaggregated units—ecological fallacies lurk! And patterns can appear different depending on if the unit of analysis is the neighborhood, city, region, etc.—they’re scale-dependent!

In *lab this week* (Thursday, 8 April) we will be setting up our computing environment and introducing R as a spatially-literate statistical computing (and even mapping!) language.

Readings

- O’Sullivan, David, and David J. Unwin. 2010. “The Pitfalls and Potential of Spatial Data.” In *Geographic Information Analysis*, 33–54. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9780470549094.ch2>.

Assignments

- **Weekly Exercise 1:** Thursday, 15 April at 8:00am.

W02: Global Spatial Autocorrelation

In *lecture this week* (Tuesday, 13 April), we will be discussing spatial autocorrelation in some depth; on the one hand, autocorrelation is a problem to be dealt with. After all, it causes much mischief with classical statistics! However, it also allows us to quantify, for example, clustering and dispersion in a spatial data set using e.g., Moran’s I and Geary’s C statistics. Eric will be away on Thursday, due to the Passover holiday.

In *lab this week* (Thursday, 15 April), we will be implementing global spatial autocorrelation tests—Moran’s I, Geary’s C—and learning to interpret them.

Deadlines

- Chi, Guangqing, and Jun Zhu. 2020. “Exploratory Spatial Data Analysis.” In *Spatial Regression Models for the Social Sciences*, 21–53. Advanced Quantitative Techniques in the Social Sciences Series 14. Thousand Oaks: SAGE.
- O’Sullivan, David, and David J. Unwin. 2010. “Area Objects and Spatial Autocorrelation.” In *Geographic Information Analysis*, 187–214. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9780470549094.ch7>.

Assignments

- **Weekly Exercise 2:** Thursday, 22 April at 8:00am.
- **Demystification Guide:** Thursday, 6 May at 8:00am.

W03: Local Spatial Autocorrelation

There will be no *lecture this week* (Tuesday, 20 April).

In *lab this week* (Thursday, 22 April) we will be playing catch-up. We'll be covering local statistics—a family of statistics that allows for spatial variance, including the Getis-Ord G_i^* and geographically weighted regression—and implementing them in R. Along the way, we will confront the enormous difficulty of making inferences from these statistics.

Deadlines

- Weekly Exercise 2: Thursday, 22 April at 8:00am.

Readings

- O'Sullivan, David, and David J. Unwin. 2010. "Local Statistics." In *Geographic Information Analysis*, 215–38. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9780470549094.ch8>. 14 May

Assignments

- Weekly Exercise 3: Thursday, 29 April at 8:00am.
- Final Project: Thursday, 20 May at 8:00am.

W04: Describing Fields

In *lecture this week* (Tuesday, 27 April), we will cover interpolation methods for prediction/estimation, and various means for statistically describing fields. We will also cover kriging methods, which account for trends over space—albeit somewhat superficially.

In *lab this week* (Thursday, 29 April), we will be implementing the interpolation methods covered this week in R, focusing on their sensitivity to, for example, bandwidth and semivariogram model fit.

Deadlines

- Weekly Exercise 3: Thursday, 29 April at 8:00am.

Readings

- O’Sullivan, David, and David J. Unwin. 2010. “Describing and Analyzing Fields.” In *Geographic Information Analysis*, 239–76. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9780470549094.ch9>.
- O’Sullivan, David, and David J. Unwin. 2010. “Knowing the Unknowable: The Statistics of Fields.” In *Geographic Information Analysis*, 277–313. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9780470549094.ch10>.

Assignments

- Weekly Exercise 4: Thursday, 6 May at 8:00am.

W05: Spatial Regression I

In *lecture this week* (Tuesday, 4 May), we’ll be diving into spatial regression. If previously in the semester, we’d been discussing how to measure spatial autocorrelation, we now learn how to account for it in the context of linear modeling. This first week, we’ll be focusing on *spatial lag modeling*, which assumes that the dependent variable is dependent upon adjacent values of the dependant variable.

In *lab this week* (Thursday, 6 May), we will be implementing spatial lag modeling in the R statistical computing environment, introducing both frequentist and Bayesian approaches.

Deadlines

- Weekly Exercise 4: Thursday, 6 May at 8:00am.
- Demystification Guide: Thursday, 6 May at 8:00am.

Readings

- Chi, Guangqing, and Jun Zhu. 2020. “Models Dealing with Spatial Dependence.” In *Spatial Regression Models for the Social Sciences*, 55–83. Advanced Quantitative Techniques in the Social Sciences Series 14. Thousand Oaks: SAGE.
- Ward, Michael D., and Kristian Skrede Gleditsch. 2019. “Spatially Lagged Dependent Variab14 Mayative Applications in the Social Sciences 155. Los Angeles, CA: SAGE.

Assignments

- Weekly Exercise 5: Thursday, 13 May at 8:00am.

W06: Spatial Regression II

In *lecture this week* (Tuesday, 11 May), we'll be diving into *spatial error modeling*; if spatial lag modeling, as covered last week, assumes autocorrelated dependent variables, the error model assumes that it is the model's error term that is autocorrelated. This is useful when it is assumed that not all explanatory variables are included in the model.

In *lab this week* (Thursday, 13 May), we will be implementing R workflows for spatial error modeling—again in both frequentist and Bayesian paradigms—and introducing methods for model comparison.

Deadlines

- Weekly Exercise 5: Thursday, 13 May at 8:00am.

Readings

- Chi, Guangqing, and Jun Zhu. 2020. "Models Dealing with Spatial Dependence." In *Spatial Regression Models for the Social Sciences*, 55–83. Advanced Quantitative Techniques in the Social Sciences Series 14. Thousand Oaks: SAGE.
- Ward, Michael D., and Kristian Skrede Gleditsch. 2008. "Spatial Error Model." In *Spatial Regression Models*, 75–86. Los Angeles, CA: SAGE.

Assignments

- Weekly Exercise 6: Thursday, 20 May at 8:00am.

W07: Bonus: Cluster Identification

In *lecture this week* (Tuesday, 18 May), we will be debriefing the semester and sharing our work on the demystification guides and projects.

There will be no *lab this week* (Thursday, 20 May), as classes end on Tuesday 18 May.

Deadlines

- Weekly Exercise 6: Tuesday, 20 May at 8:00am.
- Final Project: Thursday, 20 May at 8:00am.