Course Description and Goals: Spatial data – which is prevalent in climatology, meteorology, geology, oceanography, ecology and many other fields – requires specialized statistical tools for analysis due to the inherent dependence which exists between samples taken at nearby locations (or times). In this introductory course in spatial statistics, we'll cover central topics including spatial point processes, lattice models, and geostatistics. Theory (mathematical models and background) will be presented in concert with inference and prediction (statistics and estimation) with an emphasis on numerical examples to provide intuition and identify commonly used tools for EDA including visualization and important summary statistics. Students should take this course if they are interested in learning more about tools to handle the complex datasets where each datum may depend in some part upon (all) others.

Prerequisites: Stat 110; Stat 111 (may be taken concurrently); MATH 21a, 21b or equivalent; Stat 139 is highly recommended but not required

Course Webpages: Canvas for assignments and check-ins and ed for questions and additional resources

Course Staff

Instructor: Alex Young
Email: alexander_young@fas.harvard.edu
Office: Science Center 604

Teaching Fellow: David Harshbarger
Email: davidharshbarger@g.harvard.edu

Course Meetings, Attendance, and Communication

Lectures: Tuesday/Thursday: 10:30 - 11:45 in Science Center 309A
Attendance: Attendance is highly encouraged to enable students to ask questions in real time and to join their peers in actively working on content during the lesson. Copies of the lecture notes will be posted on Canvas after class, but students are responsible for any announcements made during the class period.
Office Hours: To be determined after surveying enrolled students.
Sections: This course will not include a section. Students are encouraged to attend the office hours to work on the problem sets in with course staff and their peers.
Ed: A collaborative discussion board on Ed will be open throughout the semester. Students are encouraged to post questions and answers to problem sets and lectures. Any helpful outside examples you find are also welcome.
Communicating with staff: Communication is critical so do not hesitate to reach out to anyone on the course staff should you have questions or concerns about course assignments, content, due dates, or unexpected circumstances which are having a deleterious effect on your ability to get the most out of this class. Ed, email, office hours, or extra meetings are all acceptable depending on your preference.

Course Materials

Recommended Texts: There is no required text for the course; class notes should be sufficient to complete all assignment. The lecture notes are based off of the following books which are great references and can provide additional detail beyond what we cover in class:

Applied spatial data analysis with R by Bivand, Pebesma, and Gómez-Rubio
Spatio-Temporal Statistics with R by C. Wikle, A Zammit-Mangion, N. Cressie
Spatial Point Patterns: Methodology and Applications with R by Baddeley, Rubak, and Turner

Technical and Computational Requirements: To balance the technical discussions and ideas presented in class students will be expected to follow guided assignments using R code and to interpret the results. As such, each student should have access to a computer with R and RStudio installed and with the ability to install additional R packages as needed including *tidyverse* and *sf*.

All problem sets should be typeset with supporting code/figures where applicable. As such, familiarity with R and Markdown/LaTeX is required. Accommodations will be made to assist students develop proficiency with these tools. Tutorials for R and R Markdown may be found at https://www.rstudio.com/online-learning/ and for LaTeX at https://www.latex-tutorial.com/. Numerical examples will be covered in class, and students will have access to these examples including source code via Canvas.

Grading: Problem Sets (60%), Check-Ins (5%), Final Exam (35%)

Problem Sets: Seven problem sets will be assigned as RMarkdown files. Each assignment will contain written and coding problems. Students will be asked to complete open portions of the RMarkdown file with answers or code as required and generate a pdf file which they will submit via Canvas. Handwritten solutions will not be accepted, and students are responsible for ensuring their assignments are appropriately typeset. The lowest problem set score will be dropped. *Extensions will not be given. Instead, a late penalty will be applied to all assignments submitted after the deadline which lowers the maximum possible grade by 1% for every hour (rounded up) the assignment is late.*

Check-Ins: At the beginning of each class, there will be a short check-in comprised of one or two questions that are expected to take 3-5 minutes to complete. The content of the check-in will come from the previous lecture material, and there will be time set aside at the beginning of class to do complete them. However, students will have a 24-hour window in which to complete the check-in to allow students to complete asynchronously as needed. The lowest 10 check-ins will be dropped.

Final Exam: A take home final exam will be posted at 12:01AM on May 4th and will include theoretical and computational problems. Students will have until 11:59PM on Wednesday May 10th to complete the final and post their solutions on Canvas. Collaboration of any type on the final are not allowed. However, additional office hours will also be provided by the course staff. *Extensions will not be given except in the most extenuating situations and only at the discretion of the instructor.*

Academic Integrity and Collaboration: Please familiarize yourself with Harvard's policy on academic integrity in the Undergraduate Handbook. Stated therein: "It is expected that all homework assignments, projects, lab reports, papers, theses, and examinations and any other work submitted for academic credit will be the student's own." However, collaborating with other students is a powerful learning tool and is also highly encouraged. Here is a bit of advice on how to collaborate effectively without running afoul of the academic integrity policies.

Check-Ins: There will generally not be sufficient time to collaborate with fellow students on these check-ins. Instead focus on answering independently with the aid of your notes from the previous lecture.

Problem Sets: Prior to collaborating with other students you should make an honest effort to work through each pset problem. When collaborating discuss plans or approaches for solving problems without directly comparing solutions. Finally, write up your solutions independently from classmates and do not check them against others' answers or outside materials.

Final Exam: You are encouraged to use the course materials freely and to contact course staff with any questions. However, the final exam should be entirely free of collaboration with other students or outside resources.

Important Dates: Any changes to the following dates will be announced in class.
Tentative Problem Set Due Dates: January 27; February 10, 24; March 10, 31; April 14, 26
Final Exam: Posted: May 3; Due: May 10

Tentative Schedule: This course will begin with an overview of spatial point processes where the focus is in on the random location of events. In the second portion of the course, the focus will be on the dependence of real-valued random samples taken at regular fixed locations (lattices/areal data). We will consider the case of real valued data on continuous locations (geostatistics) with a focus on Gaussian processes, anistrophy, and prediction (kriging). If time permits, select topics will be revisited from a modern perspective which incorporates (Bayesian) hierarchical modeling, advanced computation, and temporal dependence. A more detailed (tentative) calendar is shown below.

Week 1: January 23 - January 27

- Tuesday: Syllabus, Overview, Objectives, Motivation, and Examples
- Thursday: Point Processes - Theory
- Friday: pset#1 due at 11:59 PM

Week 2: January 30 - February 3

- Tuesday: Point Processes - Theory

- Thursday: Point Processes - EDA

Week 3: February 6 - February 10

- Tuesday: Point Processes - EDA
- Thursday: Point Processes - Modeling and Inference
- Friday: pset#2 due at 11:59PM ET

Week 4: February 13 - February 17

- Tuesday: Point Processes - Modeling and Inference
- Thursday: Point Processes - Case Study

Week 5: February 20 - February 24

- Tuesday/Thursday: Areal Processes - Theory
- Friday: pset#3 due at 11:59PM ET

Week 6: February 27 - March 3

- Tuesday: Areal Processes - Theory/EDA
- Thursday: Areal Processes - EDA

Week 7: March 6 - March 10

- Tuesday/Thursday: Areal Processes - Modeling and Inference
- Friday: pset#4 due at 11:59PM ET

Spring Break: no class

Week 8: March 20 - March 24

- Tuesday: Areal Processes - Modeling and Inference
- Thursday: Areal Processes - Case Study

Week 9: March 27 - March 31

- Tuesday/Thursday: Geostatistics - Theory
- Friday: pset#5 due at 11:59PM ET

Week 10: April 3 - April 7

- Tuesday/Thursday: Geostatistics - EDA

Week 11: April 10 - April 14

- Tuesday: Geostatistics - Modeling, Inference, and Prediction
- Friday: pset#6 due at 11:59PM ET

Week 12: April 17 - April 21

- Tuesday: Geostatistics - Case Study
- Thursday: Hierarchical Modeling

Week 13: April 24 - April 26

- Tuesday: Hierarchical Modeling
- Wednesday: pset#7 due at 11:59PM ET

Final Exam Period

- Final exam released on May 4th at 12:01 AM
- Final exam due on May 10th by 11:59 PM