Instructor: Alex Young, email: alexander_young@fas.harvard.edu
Teaching Fellow: Dieyi Chen, email: dieyi.chen@g.harvard.edu
Lectures: Monday/Wednesday: 12 - 1:15 ET
Course Webpage: canvas.harvard.edu/courses/85296
Office Hours: To be determined after surveying enrolled students.

Sections: This course will not include a section. Rather, the course staff will focusing on holding office hours over a wide range of times to accommodate students from all time zones.

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 depends in some part upon all others.

Recommended Texts: There is no required text for the course; class notes should be sufficient to complete all assignment. However, the following books are great references:

Applied spatial data analysis with R by Bivand, Pebesma, and Gómez-Rubio - Available through Hollis.
Spatio-Temporal Statistics with R by C. Wikle, A Zammit-Mangion, N. Cressie - Available as a free e-book
Spatial Point Patterns: Methodology and Applications with R by Baddeley, Rubak, and Turner - a thorough treatment of spatial point pattern data
Statistics for Spatial Data by N. Cressie - A classic introduction to the field
Statistics for Spatio-Temporal Data by N. Cressie and C. Wikle - A modern introduction to data with spatial and/or temporal dependence
Theory of Spatial Statistics by M.N.M. van Lieshout - A short, theoretical introduction to probabilistic models of spatial data

Attendance: All lectures will be recorded and posted on Canvas so attendance is not required. However, attendance to live lectures is highly encouraged to enable students to ask questions in real time and will include regular breakout rooms allowing students to join their peers in actively working on content during the lesson under supervision of the course staff. Special consideration in the scheduling of office hours will be given to accommodate students in time zones which prevent their attendance in class.

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

Grading: Problem Sets (60%), Check-Ins (15%), Final Exam (25%)

Problem Sets: A total of six 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. Collaboration with other students is encouraged, but students must write their own solutions in their own words. The lowest homework score will be dropped. *Extensions will be handled on a case by case basis, but will not be accepted if solutions for the assignment have already been posted.*

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 Thursday April 29th and will include theoretical and computational problems. Students will have until 11:59PM on Friday May 7th 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.*

Technical and Computational Aspects of the Course: To balance the technical discussions and ideas presented in class and problem sets, students will be expected to follow guided assignments using R code and to interpret the results. All assignments should be typeset with supporting code/figures where applicable. As such, familiarity with R and LaTeX is required. However, 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/.

Challenges and Community under COVID-19: Remote instruction challenges our ability to develop and support one another in many ways such as the brief moments of conversation before and after class and the natural development of study groups. It is the goal of the course staff to support and facilitate these types of interactions through alternative formats and assignments so that we can recreate the innumerable often unseen advantages of in-person classes. In particular, we will make use of the following tools, assignments, and format this semester:

> **Slack:** A Slack channel for the course, which can be accessed through Canvas, has been created with specific channels allowing students and course staff to interact on questions related to the lectures and assignments.
>
> **Check-ins:** The daily check-ins will focus on important details from the previous lecture. The goals is to provide course staff will additional insight into the progress of each student that would otherwise be available indirectly during in person instruction.
>
> **Lecture format:** To recreate the feel and pacing of a traditional lecture, content will be presented via with handwritten notes rather than prepared slides. Breakout rooms will be used occasionally for small in-class problem sessions and peer-to-peer conversations. Additional numerical examples and case-studies will also be presented. Associated scripts/code to reproduce the results will be posted on Canvas.

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.

Important Dates: All dates subject to change in the event that the University alters the schedule for Spring 2021
Tentative Problem Set Due Dates: February 12, 26; March 12; 26; April 9, 23
Final Exam: Posted: April 29, 2021, Due: May 7, 2021

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 25 - January 29

- Monday: Syllabus, Overview, Objectives, Motivation, and Examples
- Wednesday: Point Processes Theory I, Definitions and Examples, 1st and 2nd order properties

Week 2: February 1 - February 5

- Monday: Point Processes Theory II, CSR, Poisson Processes, Simulation
- Wednesday: Point Processes Inference - G,F, K functions

Week 3: February 8 - February 12

- Monday: Point Processes Model Fitting - Parametric (MLE) and Nonparametric
- Wednesday: Point Processes Additional Topics -Cluster Processes, Cox Processes
- Friday: pset#1 due at 11:59PM ET

Week 4: February 15 - February 19

- Monday: Point Processes - Case Studies
- Wednesday: Areal Processes Theory - General Definitions and Examples, Lattices/Graphs/Neighbors, Simulation

Week 5: February 22 - February 26

- Monday: Areal Processes Theory II - Markov Random Fields, Besag Factorization, Hammersley-Clifford
- Wednesday: Areal Processes Inference - EDA and tests for Spatial Dependence
- Friday: pset#2 due at 11:59PM ET

Week 6: March 1 - March 5

- Monday: No class (Wellness day)
- Wednesday: Areal Processes Model Fitting - MLE and Limitations

Week 7: March 8 - March 12

- Monday: Areal Processes Model Fitting II - Validation, Bootstrapping, and Model Selection
- Wednesday: Areal Processes Model Fitting III - Validation, Bootstrapping, and Model Selection
- Friday: pset#3 due at 11:59PM ET

Week 8: March 15 - March 19

- Monday: Areal Processes - Case Studies
- Wednesday: Geostatistics Theory - General Definitions and Examples, Stationarity, Variograms

Week 9: March 22 - March 26

- Monday: Geostatistics Theory II - Gaussian Processes, Mean and Covariance Functions, Bochner's Theorem
- Wednesday: Geostatistics Inference - Identifying Spatial Dependence, Fitting Variograms
- Friday: pset#4 due at 11:59PM ET

Week 10: March 29 - April 2

- Monday: Geostatistics Inference II - Identifying Spatial Dependence, Fitting Variograms
- Wednesday: No class (Wellness day)

Week 11: April 5 - April 9

- Monday: Geostatistics Prediction - Simple, Ordinary, Universal Kriging
- Wednesday: Geostatistics Prediction II - Universal Kriging, Co-kriging
- Friday: pset#5 due at 11:59PM ET

Week 12: April 12 - April 16

- Monday: Geostatistics - Case Studies
- Wednesday: Hierarchical Methods

Week 13: April 19 - April 23

- Monday: Hierarchical Methods
- Wednesday: Spatio-Temporal Processes
- Friday: pset#6 due at 11:59PM ET

Week 14: April 26 - April 30

- Monday: Spatio-Temporal Processes
- Wednesday: Spatio-Temporal Processes
- Thursday: Final Exam released at 12:01 AM ET, Due May 9th at 11:59PM ET