PP275: Spatial Data and Analysis
Semester: Fall 2015

Lecture: Wednesday 2-5 PM, room 105 Goldman School of Public Policy
Section: Friday 2-4 PM, room 250 Goldman School of Public Policy

Instructor: Solomon Hsiang (shsiang@berkeley.edu)
Office Hours: Monday 12:00-2:00

GSI: Felipe Gonzalez (fgonzalez@econ.berkeley.edu)

Optional “required” text:
"Spatial data analysis" by Christopher Lloyd, Oxford University Press

Recommended texts (depending on students' interests):
"The Visual Display of Quantitative Information" by Edward Tufte
"Statistics for Spatio-Temporal Data" by Noel Cressie and Christopher Wikle

Class structure

Because the detailed analysis of spatial data is a relatively young and informal field, there is no "corpus" of training that is considered standard (and there are very few textbooks). Recognizing this fact, the modest aim of this class is to help students "think spatially" so they can understand and engage with spatial data and studies of spatial data intuitively and creatively. The course is designed to train students to think clearly and rigorously about spatial data and its analysis, to be informed consumers of spatial analyses, and to be capable of conducting their own analysis and display of spatial data for policy or research. Usefulness and creativity in analyses are emphasized alongside accuracy, clarity, and aesthetic appeal in the display of results. Exposure to diverse applications, through both readings and labs, give students a sense of breadth and utility of analyses they may undertake using tools from the class.

Foundations: The first half of the course teaches students the foundations of spatial data and its analysis. Each class introduces new data structures and various analytical or statistical measures applied to these structures. Associated readings, discussions and lab assignments reinforce students understanding and help them see how these ideas are applied. Students use Matlab for all assignments so they may understand how foundational concepts build on one another and develop an understanding of the underlying mathematical procedures. The week students have to prepare for the midterm exam gives them the opportunity to review these foundations.

Advanced topics: The second half of the course exposes students to advanced topics, such as high-dimensional data, optimization, remote sensing, and more advanced analytical techniques. Students will also be introduced to Q-GIS.

Final project: During the second half of the course, students will develop a final project on a topic of their choosing. The project will give the students an opportunity to apply the tools they have learned throughout the course to a problem they are interested in. Through feedback on their proposal and presentation, students are encouraged to undertake analyses that are thoughtful, clear and that they find challenging.

Lectures: 3-hr Wednesday classes are broken into complementary elements that provide topical immersion while avoiding mental fatigue (or at least trying to).

The first 1.5-2 hrs of class focuses on analytical topics.
The mid-class break gives students an opportunity to informally discuss and critique the design/aesthetics of graphical displays.
The last 45 min of class will be a critical discussion of assigned readings.

Discussion section: 2-hr session on Fridays led by the GSI with a focus on programming and implementing the analytical techniques presented in the lectures.

Deliverables

Readings: Students are responsible for completing readings for class so that they are prepared to discuss the material insofar as it relates to the structure, analysis and display of spatial data. For the last 45 min of each class (approx) students may be called on to lead a discussion of individual readings.

Lab Assignments: Seven lab assignments will be assigned during the course of the semester. In general, labs are due Sunday evening. They will provide hands-on experience with data, analytical techniques, computing, and plotting tools. The labs are designed to introduce students to the Matlab computing language and environment. Matlab is available for students in the GSPP computer lab, as well as numerous computer labs around campus. Students can also purchase a low cost student license (with Mapping Toolbox) from Mathworks (recommended).

Midterm: An exam emphasizing the "fundamentals" will be given in class on Oct 7. Students will not be explicitly examined on the more advanced materials presented in the second half of the term.

Final project (proposal/presentation/paper): Students will conduct an analysis of data on a relevant topic of their choosing. A 1 page proposal describing the topic of interest and potential data sources is due Oct 14. Students will do a 10 minute presentation of their results during the final two classes on Dec 2 & Dec 9. A final analysis (roughly 8-10 pages) is due Dec 16.

Potential extra credit: (PEC) assignments are to upload an image file of "the most interesting and beautiful display of spatial data" that you can find (be sure to include a caption if

necessary). I will pick the top five (or so) and the class will vote on the best during the mid-class break. The winner each week gets a point added to their final grade.

Grading: 10% class participation, 45% labs, 15% midterm, 5% final project proposal, 5% final project presentation, 20% final project write-up

Class policies

Laptops should not be used during class, except for during any portion of class in which we are explicitly working with software.

I will do my best to respond to email inquiries within 48 hours, but do not expect me to respond to deadlines shorter than that timeframe. Plan ahead and email me questions in advance.

Late assignments are accepted but penalized a full grade.

Classes:

1) Introduction: Motivation, Examples, History and Preliminaries (8/26)

Discussion of objectives, expectations, deliverables and schedule
Some motivating examples of spatial data and analysis
Notational preliminaries
Coordinate systems

Lab 1 assigned

2) Space: Measurement and Representation (9/2)

Measurement, error, approximation
Dimensionality
Distance, speed and velocity
Latitude, longitude and distances on a sphere
A very brief history of spatial data and analysis
Sources of spatial data
Satellites and the basic physics of remote sensing
Design of maps: messaging, clarity, accuracy, aesthetics

Readings for class discussion (choose 2 that you are prepared to explain to the class):

- Kinney et al. (2011) "Traffic Impacts on PM2.5 Air Quality in Nairobi, Kenya," Environmental Science and Policy
- Lu et al. (2012) "Predictability of population displacement after the 2010 Haiti

earthquake" PNAS
- Rothstein (2007) "Does Competition Among Public Schools Benefit Students and Taxpayers? Comment" American Economic Review

Due at midnight the night before class: Upload PEC1
Lab 1 due in class (hard copy)
Lab 2 assigned

3) Point Processes (9/9)

Point processes
Attributes
Clustering
Histograms
Heterogeneity in point processes

Readings for class discussion (choose 2 that you are prepared to explain to the class):

- Sakaki et al., "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," WWW (2010)
- de Boer et al., "Understanding spatial differences in African elephant densities and occurrence, a continent-wide analysis," Biological Conservation (2013)
- Currid & Williams, "The geography of buzz: art, culture and the social milieu in Los Angeles and New York," Journal of Economic Geography (2010)
- Block et al. "Fast Food, Race/Ethnicity, and Income: A Geographic Analysis," American Journal of Preventive Medicine (2004)

Due at midnight the night before class: upload PEC2
Lab 2 due at noon (upload)
Lab 3 assigned

4) Lines, Polygons and Networks (9/16)

Lines and buffers
Polygons and boundaries
Area and Boolean logic
Centroids and weighting
Networks in space
Examples

Choose 2 readings that you are prepared to discuss in class (Bao & Chambers are recommended):

- Bao (2012) "Dams and Intergovernmental Transfer: Are Dam Projects Pareto Improving in China?"
- Chambers et al. (2010) "A Measure of Bizarreness" Quarterly Journal of Political Science
- Olmstead & Rhode (2011), "Adapting North American wheat production to climatic challenges, 1839–2009" PNAS
- Phithakkitnukoon et al. (2012) "Tracking Trash"
- Ratti et al, "Redrawing the Map of Great Britain from a Network of Human Interactions", PlosOne 2010
- Cai et al (2013) "Using kernel density estimation to assess the spatial pattern of road density and its impact on landscape fragmentation," International Journal of Geographical Information Science

Due at midnight the night before class: upload PEC3
Lab 3 due at noon (upload)
Lab 4 assigned

5) Fields (9/23)

Fields, rasters and grids
Isotropy
Gradients
Contours
Interpolation
Point-wise functions

Readings for class discussion (choose 2):

- "The velocity of climate change", Loarie et al. Nature 2009
- "Geography and macroeconomics: New data and new findings," Nordhaus, PNAS 2006
- "Geographic distribution of global agricultural lands in the year 2000," Ramankutty, Global Biogeochemical Cycles 2008
- Global Water Resources: Vulnerability from Climate Change and Population Growth, Vorosmarty et al, Science 2000

Due at midnight the night before class: upload PEC4
Lab 4 due at noon (upload)
Lab 5 assigned

6) High-dimensional fields (9/30)

Vector fields

Dynamic fields
Intensive and extensive variables
Spatial averaging
Empirical orthogonal function analysis
Point-wise regression and corrections
Readings for class discussion (choose 2):

- Burchfield et al.: CAUSES OF SPRAWL: A PORTRAIT FROM SPACE, Quarterly Journal of Econ (2006)
- Hsiang et al: Civil conflicts are associated with the global climate, Nature (2011)
- Rodell et al: Satellite-based estimates of groundwater depletion in India, Nature (2009)
- Wesolowski et al. Quantifying the Impact of Human Mobility on Malaria, Science (2012)

Due at midnight the night before class: upload PEC5
Lab 5 due at noon (upload)

7) Midterm (10/7)

Midterm in class

8) 3D Data and Causal Inference (10/14)

3D data: points, fields and surfaces
Transects of data
Digital elevation models
Using space to colocate events
Causal inference

Readings for class discussion (read both):

- Wolfram Schlenker and Michael J. Roberts (PNAS) "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change"
- Jesse K. Anttila-Hughes and Solomon M. Hsiang (working paper) "Destruction, Disinvestment, and Death: Economic and Human Losses Following Environmental Disaster"

Final Project Proposal due at noon (upload)
Lab 6 assigned

9) Computation in Spatial Problems (10/21)

Minimum distance along a network
Optimizing policies in space

Spatial Markov chains

Readings for class discussion (choose 2):

- Keeling et al (Science, 2001) “Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape”
- Dell (American Econ Review, 2015) “Trafficking Networks and the Mexican Drug War”
- Hall et al. (Tellus, 2007) “Statistical modelling of North Atlantic tropical cyclone tracks”

10) Intro to Remote Sensing (10/28)

Spectra
Classification
Image sharpening
Cloud removal
Compositing

Readings for class discussion (choose 2):

- Fretwell et al (PLOS One, 2014) “Whales from Space: Counting Southern Right Whales by Satellite”
- Sibley et al. (Agronomy, Soils & Environmental Quality, 2013) “Testing Remote Sensing Approaches for Assessing Yield Variability among Maize Fields”
- Henderson et al (American Econ Review, 2012) "Measuring Economic Growth from Outer Space”

Lab 6 due at noon (upload)
Lab 7 assigned

11) Parameters, Probability, Stocks & Flows (11/4)

Possible Guest lecture: TBD

3D Fields
Variable coefficients as a function of space
Flux and divergence
Bayesian search theory

Readings for class discussion (choose 2):

- Baldocchi et al (Tree Physiology, 1995) “Energy and CO2 flux densities above and below a temperate broad-leaved forest and a boreal pine forest”
- Carroll et al (Nature, 2015) “Temporal and spatial analysis of the 2014–2015 Ebola virus

outbreak in West Africa"
- Brockfeld et al (Physical Review E, 2001) "Optimizing traffic lights in a cellular automaton model for city traffic"
- Ando et al (Science 1998) "Species Distributions, Land Values, and Efficient Conservation"

Veterans Day - no class (11/11)

Lab 7 due 11/13 at 5 pm (upload)

12) Regressions in Space (11/18)

Possible Guest Lecture: TBD

Spatial autocorrelation
Spatial regression models
Spatial lags
Spatially correlated errors
Issues arising from aggregation

Readings for class discussion (choose 2):

- Linden & Rockoff (American Econ Review, 2008) "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws"
- Nowacek et al. (Marine Mammal Science, 2001) "Short-term effects of boat traffic on bottlenose dolphins, Tursiops truncatus, in Sarasota Bay, Florida"
- Burke et al (2015) "Global Nonlinear Effect of Temperature on Economic Production"

Almost Turkey Day - no class (11/25)

13) Final project presentations (12/2)

Students present final projects.

14) Final project presentations con't (12/9)

Students present final projects.

Write-up of final project is due Dec 16 at 5 pm