

Course Outline

SO 8243
Spatial Analysis of Social Data
Semester: TBA

Dr. Frank M. Howell

Office: 324-A Etheredge Hall

Office Hours: TBA

(other times by appointment only; e-mail or call for a time)

Class meetings: Building, Room, Date, Time TBA

Phone: 662.325.7872

E-mail: fmh1@ra.msstate.edu

Dept. Receptionist: 662.325.2495

TEXTBOOKS (Required):

Bailey, Trevor and Anthony Gatrell. 1995. *Interactive Spatial Data Analysis*. New York: Wiley.

Fotheringham, Stewart, Chris Brunson, and Martin Charlton. 2002. *Geographically Weighted Regression*. New York: Wiley.

Lee, Jay and Davis Wong. 2001. *Statistical Analysis with ArcView GIS*. New York: Wiley.

O'Sullivan, David and David J. Unwin. 2002. *Geographic Information Analysis*. New York: Wiley.

Note: additional required texts available for download on the Internet are shown later in this syllabus.

Required Listserv Subscriptions:

You are required to subscribe to the following Internet Listservs during the semester. These subscriptions must be completed by the end of the first week of class. Note that each has a "digest" format as well as web-based archives.

Listsrv for Openspace -- Open source spatial analysis software tools:

<http://sal.agecon.uiuc.edu/mailman/listinfo/openspace>

Listsrv for R-sig-Geo -- R Special Interest Group on using Geographical data and Mapping:

<http://sal.agecon.uiuc.edu/csiss/#rgeo>

COURSE OVERVIEW:

Catalog Description. SO 8243. Spatial Analysis of Social Data. (3) (Prerequisite: SO 8283 or equivalent, or consent of the instructor). Three hours lecture. Spatial theories of society; relevant digital databases; procedures for visualizing data; exploratory spatial data analysis; local and global spatial regression models.

Overview. Most social data have or can have a spatial element to them—crimes are committed at a location, unemployment is tracked by city or county, the location of hazardous waste sites occur in some spatial proximity to human populations, consumers purchase products at stores located in specific places, populations grow and decline in relative proximity to larger urban centers, disease and death are distributed unevenly across space, and social inequality is spatially-situated—but conventional procedures of social data analysis do not make use of this important information. The National Science Foundation recently sponsored the Center for Spatially-Integrated Social Science (CSISS) to serve as a catalyst for the better understanding of social phenomena using spatial analysis procedures (see Goodchild et al. 2000). The social sciences have varied widely in how they have adopted spatial technology for studying social phenomena. However, the roots of map use and making inferences about collective social

behavior in Sociology can be traced back to at least 1915 when the rural sociologist, C. J. Galpin, used hand-drawn maps to plot the “ruts” in the dirt roads in rural Wisconsin which helped demarcate the boundaries of small communities in proximity to regional population centers. As Robert Park noted in 1929 (Park 1929), Galpin’s work led the Chicago School researchers to replicate them in the city. In contemporary work, Abbott (1997) has argued for the enduring importance of this perspective. The work of the students of Park and others in the Chicago School, Shaw and McKay, on delinquency rates in Chicago was another strand of the use of spatial analysis during that period (Shoemaker 1996). There are also a number of other sociological theories identified in the course outline imbued with spatial thinking. Theoretical issues will be an integral part of the discussion of the methods articulated during the semester. The development and expression of your ideas on how “space” fits into contemporary Sociology is also a welcomed part of the class (see Gieryn 2000; Goodchild et al. 2000 for some inspiration).

The course is designed to give you advanced training in the spatial analysis of sociological data. Successfully completing it fulfills an advanced course requirement in the Ph.D. program in Sociology. Other graduate programs may permit this course to qualify for quantitative training requirements. While this course introduces the use of modern techniques of spatial analysis applied to social data, the explosion in quantitative social science methods during the past 25 years or so makes the effective learning of these methods a process that must transpire over a period of several years. It is therefore important for students enrolled in this course to accept the necessity of continuing their quantitative training beyond actual course work as a commitment to their professional development. This course should perhaps be best viewed as just *one* important step in that process of professional development.

The objectives of the course are to acquaint graduate students in Sociology with methods to analyze spatially-referenced data with the procedures appropriate for the field’s theoretical base. The course contact hours are organized into three major foci: (a) essential theoretical concepts and the constituent reference to Census Bureau demography (including the digital TIGER database); (b) the visualization of social data facilitated by Geographic Information Systems software; and (c) techniques to construct or analyze social point, line, and polygon data using exploratory and confirmatory approaches. Weekly homework, performed largely in the Department-maintained Computer Laboratory, will emphasize applications of these procedures to actual social theories and data using a pedagogical model of: application, interpretation, and presentation of empirical analyses by students. The semester-length paper is a key instrument by which students will demonstrate competence in spatial analysis methods covered in the course.

Organization. The class is a combination of lecture, homework assignments in the computing lab, and seminar formats. A strict requirement for each class meeting is to complete the assigned readings or tasks before coming to class. The in-class period will largely consist of a lecture, generally using examples with actual social data, but will be combined with some interactive discussion by members of the class. There will usually be some “hands-on” activity during or after the class meeting. There will be a great deal of material on the Internet with which students will be required to become familiar. Students are encouraged to become familiar not only with the Departmental Computing Laboratory (and the Teaching Assistants who manage this facility), but also the MSU computing labs across campus (see <http://www.its.msstate.edu/Information/PublicLabs/index.php> for locations and hours). While the University labs have GIS and related software on their computers, the Departmental Computing Lab will have all of the software used in the course installed and updated. Much of this software is offered under a “free” license model, so students can download most of the programs for use on their personal computers as well.

Students are required to attend all class meetings. Please notify me or the Departmental

receptionist (see cover page) if you are to be absent from a class meeting. You are responsible for acquiring notes, handouts, and so forth from any classes that you do not attend from other students. Academic policies contained in the current *MSU Bulletin* apply to this course. Students in the class are required to have a current MSU e-mail address (NetID) to foster course participation and to check this account regularly (see <http://www.its.msstate.edu/helpdesk/mail/> for details). Some course information and assignments, for instance, may only be disseminated through electronic mail.

Let me note that students enrolled in my classes have the top priority to see me at my office in Etheredge Hall. Graduate student advisees come second, with my Department Head coming in third. (He knows this.) In balancing the demands of teaching, research, and service, I stay very busy. So that I can make you a priority this semester, I have to rely on office-hour blocks or appointments to meet with students. If you have a brief question, sending e-mail to me is fine. If you need to meet with me outside of my scheduled office hours, you will need to e-mail or call for an appointment. I can usually see you fairly quickly but it will need to be by appointment so that all students have equal access. Thank you for understanding policy but I want to emphasize that as a student in one of my classes, *you* have top priority on my schedule this semester.

Grading Procedures. The requirements of the course include attendance at all class sessions, completion of all assigned readings prior to the relevant class meeting, completion of all homework assignments in a timely manner, and active participation in class (both in-class and via electronic mail). The examinations will be organized around the major foci of the course as shown below. There will be a semester-length research project resulting in both an electronic presentation in a colloquium setting using Microsoft Powerpoint (or equivalent) and a written paper. The topic and scope must be approved by the Instructor. This research project must use spatial data to address some theoretically-driven research problem and be at least 20 typed pages (excluding figures and tables) in the format of the *American Sociological Review*. I will entertain proposals for research-team projects (not more than three students on a team) if a clear delineation of equitable work-loads is presented in writing with a clear commitment to an equal grade on the paper for each member of the team. This must be done at the time the paper abstract is submitted (announced TBA by the Instructor). A breakdown of the final course grade is as follows:

Exam on social demography and databases	15%
Exam on visualizing social data	20%
Exam on ESDA	20%
Final exam on spatial regression models	20%
Research paper and presentation	25%
Total	100%

The course grades will be assigned letter grades using a conventional breakdown of final averages: A=90-100; B=80-89; C=70-79; D=60-69; F=below 60.

Academic Misconduct and Dishonesty. Mississippi State University treats academic misconduct and dishonesty very seriously, as do I. It is incumbent upon you to become familiar with University policies and procedures regarding this topic. They can be reviewed at the University's website, accessible through the web address of <http://www.msstate.edu/web/security.html>.

COURSE TOPICS AND GENERAL SCHEDULE:

(Note: this schedule is subject to change at the discretion of the Instructor. The projected timing for each topic is shown at left. The sequence will generally be followed verbatim. Key readings are shown in parenthesis to the right of each entry. A reference list is included at the end of this course outline.)

- Week 1.** Social theories involving space, place, and reflexive relationships (3 hours)
1. C.J. Galpin and rural communities in Wisconsin (Galpin 1915)
 2. The Chicago School of urban change (Abbott 1997)
 3. Hawley's Social Ecology of Space (Hawley 1950; Duncan et al. 1959)
 4. Kaufman-Wilkinson Social Field Theory (Kaufman 1959; Wilkinson 1991)
 5. Spatial theories of crime (Shaw & McKay, 1969; Green et al. 1998; Anselin, Cohen, Cook, Gorr and Tita 2000)
 6. Population deconcentration and the rural-urban dynamic (Lichter & Fuguitt 1982)
 7. Other relevant social science theoretical perspectives
- Weeks 2 & 3.** Social Demography Concepts and Databases (6 hours)
8. Census Bureau geographic hierarchy (Census GARM)
 9. Major Spatial Data Sources: TIGER, USDOT, USGS
(Census TIGER Documentation; USGS 1995; plus the following websites:
GeoData Website [<http://www.geodata.gov>];
Federal Geographic Data Committee clearinghouse
<http://clearinghouse1.fgdc.gov>
National Map Atlas [<http://nationalatlas.gov>]
U.S. Bureau of Transportation Statistics
<http://www.geocomm.com/links/linksdata.html>
U.S. EPA Envirofacts Data Warehouse
<http://www.epa.gov/enviro/>)
 10. Selected commercial sources of spatial data
(Geography Network [<http://www.geographynetwork.com>])
(Geocommunity Data Depot <http://gisdatadepot.com/>)
- Weeks 4 & 5.** Visualizing Social Data Using Geographic Information Systems (6 hours)
11. Elementary concepts of maps and GIS (O'Sullivan & Unwin 2003)
 12. Spatial concepts, operations, and data types (O'Sullivan & Unwin 2003)
 13. Accessing Census and other datasets, extracting spatially-referenced TIGER data, and building GIS coverages (TIGER Documentation; Breslin et al. 1998; handout)
 14. Creating point coverages using GPS and geocoding location information (O'Sullivan & Unwin 2003; Breslin et al. 1998)
 15. Common spatial operations used to construct social variables (O'Sullivan & Unwin 2003)
- Weeks 6-8.** Exploratory Spatial Data Analysis (ESDA) of Social Data (9 hours)
16. Preparing social data for ESDA: buffers, aggregation, point-in-polygon, Thiessen polygons, and spider diagrams (Anselin 2003b,d; Breslin et al. 1998; Fotheringham et al. 2000; O'Sullivan & Unwin 2003)
 17. Construction and evaluation of spatial weights: contiguity, distance, and nearest-neighbors (Anselin 2003a,b,c)
 18. Central tendency and dispersion of point and line data (Anselin 2003d; Lee and Wong 2001; Mardia 1972; Mardia & Jupp 2000; O'Sullivan & Unwin 2003)

19. Detecting spatial outliers, linked brushing of maps (Anselin 2003b)
20. Computation of counts and rates standardization and smoothing procedures (Anselin, Kim, and Syabri 2003; Anselin 2003a; Bailey & Gatrell 1995: Chapter 8)
21. Identification and visualization of spatial autocorrelation patterns using Moran's I, Local Indicators of Spatial Association (LISA), and Moran scatterplots linked to LISA maps (Anselin 2003a,b; Anselin, Syabri, and Smirnov 2002)

Weeks 9-12. Spatial Regression Models of Social Data (12 hours)

22. Classes of spatial regression models (1 hour)
 - a. Continuous surface vs. lattice models (Anselin & Bera 1998)
 - b. Global vs. local models (Fotheringham et al. 2002)
 - c. Space vs. time models (Bailey & Gatrell 1995: 122-133; Kuldorff 1997; Levine 2002)
23. Global spatial regression models (5 hours)
 - a. Conventional OLS regression models of demographic data (Anselin 1988)
 - b. Diagnostic tests for spatial dependence and heterogeneity in OLS models (Anselin 2002)
 - c. Specifying and estimating spatial lag and spatial error models (Anselin & Cho 2002a,b; Anselin & Morena 2002)
24. Local spatial regression models (4 hours)
 - a. Spatial heterogeneity as a condition of social activity
 - b. Estimating local regression models using geographically-weighted regression (GWR) (Fotheringham et al. 2002)
 - c. Data requirements for GWR: Gaussian, binomial, Poisson estimators (Fotheringham et al. 2002)
 - d. Estimating GWR models using fixed vs. adaptive kernel density (Fotheringham et al. 2002)
 - e. Visualizing spatially heterogeneous slopes using GIS (Fotheringham et al. 2002)
25. Space-time models (2 hours)
 - a. Examining spatial and temporal patterns (Bailey & Gatrell 1995; Kuldorff 1997, 2003)
 - b. Identifying statistically significant clusters of events or outcomes (Kuldorff 1997, 2003; Levine 2002; Bailey & Gatrell 1995)
 - c. Modeling space and time patterns and interactions (Bailey & Gatrell 1995)

TBA Final Exam (take-home)
TBA Research Project Colloquium

The following required texts are available for download on the Internet as noted:

- Anselin, Luc. 2003a. *GeoDa 0.9 User's Guide*. Center for the Spatial Integration of Social Sciences and Spatial Analysis Laboratory, University of Illinois. Urbana, IL: University of Illinois. (<http://sal.agecon.uiuc.edu/csiss/pdf/geoda093.pdf>).
- Anselin, Luc. 2003b. *An Introduction to EDA with GeoDa*. Spatial Analysis Laboratory, University of Illinois. (<http://sal.agecon.uiuc.edu/csiss/pdf/quicktour.pdf>).
- Anselin, Luc. 2003c. *An Introduction to Spatial Autocorrelation Analysis with GeoDa*. Spatial Analysis Laboratory, University of Illinois. (<http://sal.agecon.uiuc.edu/csiss/pdf/spauto.pdf>).
- Anselin, Luc. 2003d. *An Introduction to Point Pattern Analysis using CrimeStat*. Spatial Analysis Laboratory, University of Illinois. (<http://sal.agecon.uiuc.edu/csiss/pdf/points.pdf>).
- Anselin, Luc. 2003e. *An Introduction to Spatial Regression Analysis in R*. Spatial Analysis Laboratory, University of Illinois. (<http://sal.agecon.uiuc.edu/csiss/pdf/spdepintro.pdf>).
- Kuldorff, Martin. 2003. *SaTScan User Guide for version 4.0*. (Available online at <http://www.satscan.org>).
- Levine, Ned. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 2.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. May 2002. (Available online at <http://www.icpsr.umich.edu/NACJD/crimestat.html>).
- U.S. Bureau of the Census. 1994. *Geographic Areas Reference Manual*. Washington, DC: The Bureau. (available online at <http://www.census.gov/geo/www/garm.html>)
- U.S. Bureau of the Census. 2000. *Census 2000 TIGER/Line Files Technical Documentation*. Washington, DC: The Bureau. (<http://www.census.gov/geo/www/tiger/tiger2k/tiger2k.pdf>)

Recommended reference works:

- Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Norwell, MA: Kluwer.
- Breslin, Pat, Nick Frunz, Eileen Napoleon and Tim Ormsby. 1998. *Getting to Know ArcView GIS*. Redlands, CA: ESRI Press. (or equivalent)
- Cressie, Noel. 1993. *Statistics for Spatial Data, Revised Edition*. New York: Wiley.
- Fotheringham, A. Stewart, Chris Brunson, and Martin Charlton. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Thousand Oaks, CA: Sage.
- Michael F. Goodchild and Donald G. Janelle. In press, December 2003. *Spatially Integrated Social Science*. New York: Oxford University Press.
- Gieryn, Thomas F. 2000. "A space for place in sociology." *Annual Review of Sociology* Vol. 26: 463-496.
- Haining, Robert. 1990. *Spatial data analysis in the social and environmental sciences*. Cambridge: Cambridge University Press.
- Liverman, Diana, Emilio F. Moran, Ronald R. Rindfuss, and Paul C. Stern (eds.). 1998. *People and Pixels: Linking Remote Sensing and Social Science*. Washington, DC: National Academic Press. (available for reading online at <http://www.nap.edu/catalog/5963.html>).

References

- Abbott, Andrew. 1997. "Of time and space: the contemporary relevance of the Chicago School." *Social Forces* 75 (4): 1149-1182.
- Anselin, Luc. 2002. "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." *Agricultural Economics*. (forthcoming). (Available online at <http://agec221.agecon.uiuc.edu/users/anselin/papers/hood.pdf>)
- Anselin, Luc and Anil Bera. 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," In *Handbook of Applied Economic Statistics*, eds. Aman Ullah and David E. Giles. New York: Marcel Dekker, pp. 237-289.
- Anselin, Luc and Rosina Morena. 2002. Properties of Tests for Spatial Error Components. (Available online at http://agec221.agecon.uiuc.edu/users/anselin/papers/am_revised.pdf).
- Anselin, L. J. Cohen, D. Cook, W. Gorr, G. Tita. 2000. "Spatial Analyses of Crime," Pp. 213-262 in David Duffee (ed.), *Criminal Justice 2000: Volume 4. Measurement and Analysis of Crime and Justice*. Washington, DC: National Institute of Justice.
- Anselin, Luc, Ibnu Syabri and Oleg Smirnov. 2002. "Visualizing multivariate spatial correlation with dynamically linked windows." Paper presented at the annual meetings of the Association of American Geographers, Los Angeles, March 20-23. (available online at http://agec221.agecon.uiuc.edu/users/anselin/papers/multi_lisa.pdf).
- Anselin, Luc, Yong Wook Kim and Ibnu Syabri. 2002. "Web-Based Spatial Analysis Tools for the Exploration of Spatial Outliers," in *GIScience 2002, The Second International Conference on Geographic Information Science*, Boulder, CO, Sept. 25-28, pp. 12-15.
- Anselin, Luc and Wendy K. Tam Cho. 2002a. "Spatial Effects and Ecological Inference." *Political Analysis* 10: 276-297.
- Anselin, Luc and Wendy K. Tam Cho 2002b. "Conceptualizing Space: Reply." *Political Analysis* 10: 301-303.
- Duncan, Otis D. and Leo Schnore. 1959. "Cultural, behavioral, and ecological perspectives in the study of social organizations." *American Journal of Sociology*, 65: 132-46.
- Lichter, Daniel T. and Glenn V. Fuguitt. 1982. "The Transition to Nonmetropolitan Population Deconcentration." *Demography* 19 (May): 211-21.
- Galpin, C.J. 1915. *The Social Anatomy of an Agricultural Community*. Research Bulletin 34. Madison, WI: Agricultural Experiment Station, University of Wisconsin, May.
- Goodchild, Michael F., Luc Anselin, Richard P. Applebaum, and Barbara Herr Harthorn. 2000. "Toward Spatially Integrated Social Science." *International Regional Science Review* 23:139-159.
- Hawley, A.H. (1950). Human ecology. *International encyclopedia of the social sciences*, Vol. 4. New York: MacMillan, 328-336.
- Kaufman, Harold F. 1959. "Toward an interactional conception of community." *Social Forces* 38 (1): 8-17.
- King, Gary 2002. "Isolating Spatial Autocorrelation, Aggregation Bias, and Distributional Violations in Ecological Inference: Comment on Anselin and Cho." *Political Analysis* 10: 298-300.
- Kuldorff, M. 1997. "A spatial scan statistic." *Communications in Statistics-Theory and Methods* 26: 1487-1496.
- Green, Donald, Dara Z. Strolovitch, and Janelle S. Wong, "Defended Neighborhoods, Integration,

- and Racially Motivated Crime," *American Journal of Sociology* 104, no. 2 (1998): 372-403.
- Mardia, K.V. 1972. *Statistics of directional data*. New York: Academic Press.
- Mardia, K.V. and P.E. Jupp. 2000. *Directional Statistics*. New York: Wiley.
- Park, Robert E. 1929. "Urbanization as measured by newspaper circulation." *American Journal of Sociology* 35 (1): 60-79.
- Shaw, Clifford R. and McKay, Henry D. *Juvenile Delinquency and Urban Areas*. (Chicago: The University of Chicago Press, 1969)
- Shoemaker, Donald J. 1996. *Theories of Delinquency*. New York: Oxford University Press.
- U.S. Geological Survey. 1995. *Geographic Names Information System Data Users Guide 6, Fourth Printing (revised)*. Available online at http://geonames.usgs.gov/gnis_users_guide_toc.html.
- Wilkinson, Kenneth P. 1991. *The Community in Rural America*. Middleton, WI: Social Ecology Press.