GEOG 8104: Spatial Methods for Health and Population Research
Fall 2015 - Tuesdays 2:15 – 5:00
Derby Hall Room 1116
Prof. Elisabeth D. Root

General Introduction
This seminar surveys current topics in health and population research with a focus on spatial statistical methods such as spatial regression, geographically weighted regression, multilevel modeling and cluster analysis. Throughout the seminar we will review the broad field of spatial data analysis and the range of issues that arise when analyzing georeferenced data. We will read and discuss selections from the current literature and critique how spatial data and statistical methods are being integrated into population and health research.

By the end of this course, students will be able to: a) critique spatial analytic methods in selected papers, b) present arguments about appropriate and inappropriate spatial research methods for a given research problem, c) recognize the complexities inherent in using spatial data and choose appropriate methods for data analysis, and d) implement spatial statistical methods using simple population or health datasets.

I have not designed this as a GIS course but throughout the semester you will have plenty of opportunity to learn *ArcGIS* and other software, namely *SatScan* (cluster analysis), *GeoDa* (now called *OpenGeoda*), and *R* (spdep) for spatial analysis. *GeoDa* is a program that facilitates exploratory spatial data analysis and can be used for spatial regression modeling, while *R* has all these functions in addition to geographically weighted regression, cluster methods and much much more. All of these software packages are free, so we don't need to do any specific work in the lab. You may use any other software you wish (such as Stata or SAS).

What I think about teaching
My philosophy of teaching is that students need to learn to learn. This may sound strange, after all haven't we been in school a really long time? Don't we already know how to learn? The answer to this question is yes, we do know how to learn in some ways but college should require greater depth and breadth of thinking, especially as a graduate student. What I found out during the course of my PhD is that I needed to **teach myself** a great deal because there were often no classes on the methods or subjects in which I became interested. Sometimes, there wasn't a professor on campus who knew anything about my topic! So, I think it's very important for you to learn how to find the information you need through research and how to apply it to real world problems. The other thing I found out after working in both the private sector and at a University is that very little research is done by one person. Most of the best science is conducted by teams of researchers who brainstorm and discuss and learn together. Learning will therefore come about through research, reading, listening, communication and collaboration with other students.

This course will be conducted using a modified **problem based learning (PBL)** approach. The PBL allows for student collaboration in an active learning environment. PBL requires students

to activate prior knowledge and elaborate on that knowledge in small group discussions. This knowledge and knowledge gained through focused reading is then integrated around relevant problems and information. Students will be assigned to a group which will be their group for the entire semester. In this group, students will be given problems to solve. While the class will involve some "mini lectures" on specific topics as well as whole group discussion, a certain portion of our classes will be designated to the PBL groups.

When the PBL method is used and a problem is distributed, discussion groups will work on the problem in class and outside of class:

1. Students are presented with a problem (a set of research paper(s) or a "problem"). The whole class will discuss the main points of the papers. We may have a speaker who gives a "mini-lecture" on the topic.
2. Students then break into groups and clarify unknown terms and concepts and devise a problem description. Groups then organize their ideas and previous knowledge related to the problem, and attempt to define the broad nature of the problem.
3. Throughout discussion, students pose questions, called "learning issues," on aspects of the problem that they do not understand. These learning issues are recorded by the group. Students are continually encouraged to define what they know - and more importantly - what they don't know.
4. Students rank, in order of importance, the learning issues generated in the session. They decide which questions will be followed up by the whole group, and which issues can be assigned to individuals, who later teach the rest of the group. Students and instructor also discuss what resources will be needed in order to research the learning issues, and where they could be found or whether they need to be taught using a mini-lecture format.
5. Students research their assigned learning issue(s) through self-directed study outside of class. If new learning issues arise during the course of research, be sure to record these and bring them back to the group.
6. When students reconvene, they explore the previous learning issues, integrating their new knowledge into the context of the problem. Each group member "teaches" other group members what they've learned through self-directed research and shares resources. Often, new learning issues are defined as the group progresses through the problem.
7. Groups then devise a "work plan" which outlines how your group will conduct the small-scale research project. Work plans will typically be due at the end of the second class period of the module. The work plan should include:
    - A clearly stated research question
    - The method(s) that will be used to answer the research question – be sure to note what additional learning is necessary to carry this out
    - Ideas for datasets that will be used for the analysis
    - Roles and responsibilities for each group member
    - Feedback for the instructors on any "mini-lectures" or demonstrations that may be needed to enhance skills/knowledge
8. Groups work on their research project and write a research paper which presents their "answer" to the problem.

Groups

Problem discussion groups will be determined randomly and will consist of approximately 3 members. My job as "tutor" will be to clarify issues and to keep the groups on task. If lecture is needed to clarify some theories or concepts, it will be given. The second day of class, I will have you work in your groups to set out "ground rules". You will submit a copy to me and keep a copy for your team. I understand that sometimes group dynamics don't work. If you experience irreconcilable problems in your group, and an individual is violating the group's ground rules, there will be an option to "divorce" a group member, but only after a strict protocol is adhered to:

1. The group will choose one member to talk directly to the problem individual. During this discussion, every attempt will be made to resolve the conflict. Be sure to refer to the "ground rules" your group made and let the individual know which of those rules he or she has violated. Be nice! This is your team member, so don't gang up on him or her.
2. If the individual does not "clean up" his or her act, the group will prepare an email to me, cc'ing all group members, outlining the problem, which ground rules have been violated, and providing evidence that you have already tried to resolve the problem using the method in step 1 (above). I will review the email and ask the group to come meet with me during office hours.
3. If the individual still does not change his or her behavior, I will talk with that student one-on-one and attempt to resolve the issue. If it cannot be resolved the individual may be "divorced" from the group. This will have **serious consequences** for the group portion of that individual's grade.

Semester at a glance

During the semester, we will use 3-4 modules to learn how geographic thinking and spatial methods can be applied to population and public health problems and their solutions. I have chosen the topics for the first 2 modules:

Module 1: Introduction to Spatial Demography, Spatial Epidemiology, Health Geography
Module 2: Neighborhoods and Health

You will have the opportunity to choose the other module topics as a class (we will vote during the third class). I have listed several options below, though we will brainstorm additional ideas on the first day of class:

- Physical environment (natural/build) and chronic disease
- Environmental health
- Migration and health or social/educational/economic outcomes
- Residential segregation
- Urban environments and health or social/educational/economic outcomes
- Environmental justice

Each module will take approximately 3-4 weeks and will consist of a combination of whole-class discussions, mini-lectures, and group work. The ultimate goal of each module is to conduct a

small-scale research project and write a paper that *could* (with revision/expansion) be submitted to a peer-reviewed journal. Along the way, you will have time to reflect on the process of problem-based learning, the new skills, methods and techniques you have learned, and problems and questions you feel are unresolved.

Assessment

You will be assessed on both your individual work in the class and on group work. The importance of each component will be as follows:

Group papers	50%
Individual presentation	10%
Individual "reflective portfolio"	30%
Participation and engagement	10%

Group papers

Most PBLs will result in a research paper. Many journals have a special category of "brief articles" which compress a full research article into a few pages. Each group will work together as a team to prepare a "brief article" that could be submitted to a peer-reviewed journal (or expanded later and submitted as a full article). As part of this exercise, your group will identify which journal (and audience) your work is situated for. Papers are expected to have all the usual parts: Introduction, Data and Methods, Results, Discussion (or whatever is the norm for the journal you have identified). These papers and associated group materials (e.g., work plans) will be 50% of your grade.

> *Note: Group members will evaluate the relative contributions of each group member to each project. These evaluations will be anonymous and will require a grade and an explanation of the grade. Thus, you will grade yourself and each other member of the group for each problem. (For example, if the group earns a 90 on the paper and member A earns all A's from her group peers, she will receive a 90. However, if member B receives Ds from his group peers, he will receive a D for the paper).*

Individual presentation

Students will individually present the results of one group paper during the semester and are expected to prepare a PowerPoint presentation of their group's work. Presentations should be approximately 10-15 minutes in length and the format modeled off a presentation for PAA, AAG, APHA or some other scientific conference. The presentation will be 10% of your grade.

Individual "reflective portfolio"

Throughout the semester, you will be asked to make a journal entry two or three times each week. Journal entries can be about anything: reflections on the PBL, comments on what you find during your self-directed research, discussions of difficulties or successes in the course or with group work, etc., clarification or connecting concepts within this course or with those in other courses, or reflect on their feelings toward this course or statistics/geography in general. I will collect and respond to the journals on a regular basis.

For the end of the semester, you will prepare a reflective portfolio that will be graded. The purpose of the portfolio assignment is to allow you to highlight your own selections of your work and give an analysis of them in your own words. I will focus on two specific things in evaluating your portfolio: (1) an understanding of some key concepts in spatial methods and how they apply to population/health research, and (2) a self-awareness of your journey (where you started from, where you went, and where you are now).

Select three pieces of work from this semester to include in your portfolio. These pieces can include journal writing, group or individual work plans, class notes, statistical analyses, or any other pieces of work you have produced in this class. Your analysis should explain your reasons for picking these pieces. As examples, you might consider a selection which shows the development of your understanding of one key concept, or a selection which shows your growing appreciation of and proficiency with spatial statistical methods, or a selection which shows your connection of two or more key concepts.

The most important part of the portfolio is your reflection on why you chose the pieces you did, how they show your understanding of some key concept(s), and how they show a self-awareness of your journey through this class. You should definitely write more than one paragraph but no more than five pages. The portfolio will be 30% of your grade.

Participation and engagement

Due to the PBL structure of the class, students are expected to attend all classes. **One absence will be allowed**. After that, any student missing class (without an extremely good excuse) will receive a D on the group assignment. Attendance, as well as my assessment of how engaged you are in the PBL process, are 10% of your grade.

Reading Materials

Since we will use a PBL approach, most of the readings you do during the course of the semester will be self- or group-directed. That means we won't all be reading the same thing. I will assign several key articles per module as required reading for this course and will suggest chapters in books that pertain to relevant tools/methods. Below, I have some suggestions for books and workbooks that may be useful for you, especially if you want more complex "textbook style" readings or step-by-step instructions for how to conduct specific analyses.

Workbooks

- Anselin L. 2005. *Spatial Regression Analysis in R: A Workbook.* University of Illinois, Urbana-Champaign: Spatial Analysis Laboratory. Available online: http://geodacenter.asu.edu/system/files/rex1.pdf

- Anselin L. 2005. *Exploring Spatial Data with GeoDa: A Workbook* UC Santa Barbara, CA: Center for Spatially Integrated Social Science. The workbook, software and many related materials are available online at the current GeoDa homepage: http://geodacenter.asu.edu/.

- Harris R. 2009. Introduction to Geographically Weighted Regression. University of Bristol: School of Geographical Sciences & CMPO. Available online: http://www.bris.ac.uk/cmpo/events/2009/segregation/gwr.pdf. And here's the data: http://www.bris.ac.uk/cmpo/events/2009/segregation/southeastdata.csv.

- Kurland KS, Gorr WL. 2007. *GIS Tutorial for Health*. ESRI Press: Redlands, CA. ($$)

- Parker RN, Asencio EK. 2008. *GIS and Spatial Analysis for the Social Sciences: Coding, Mapping, and Modeling*. Routledge/Taylor & Francis: New York, NY. ($$)

The expectation is that if you need to you work thorough these workbooks at your own pace and use these as a way to build up your confidence and abilities in handling geospatial data. If you want, I can identify the "must do" tutorials. There is some duplication of general areas across these workbooks and there is no need to do all exercises (except for honing your own skills). However, one should note that while there is overlap there are also differences.

Book and other materials

- Bivand, R, EJ Pebesma and V Gomez-Rubio. *Applied Spatial Data Analysis with R*. New York: Springer.
  Supplementary material: http://www.asdar-book.org/

- Cromley E and S McLafferty. 2011. *GIS and Public Health* (2nd ed). New York: The Guilford Press.

- Fotheringham AS, Brunsdon C, Charlton M. 2002. *Geographically Weighted Regression*. West Sussex, England: John Wiley and Sons.
  Supplementary material: http://ncg.nuim.ie/ncg/GWR/software.htm

- Singer JD and JB Willett. 2003. *Applied Longitudinal Data Analysis.* Oxford, New York: Oxford University Press.
  Supplementary material: http://gseacademic.harvard.edu/alda/

- de Smith MJ, Goodchild MF, Longley PA. 2006-2008. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Available online at: http://www.spatialanalysisonline.com/. This is perhaps the most comprehensive single on-line source for material on both concepts and methods available.

- Waller LA and CA Gotway. 2004. *Applied Spatial Statistics for Public Health Data*. New York: John Wiley and Sons.
  Supplementary material: http://www.sph.emory.edu/~lwaller/WGindex.htm

- Ward MD and KS Gleditsch. 2008. *Spatial Regression Models*. Thousand Oaks, CA: Sage Publications.

Datasets

Below, I have links to some datasets that have geographic data associated with them. You may choose to use one or more of these datasets for your research papers. You are also encouraged to use your own datasets.

Measure DHS: http://www.measuredhs.com/
These are comprehensive, population-based surveys that collect a wide range of demographic and health data. Measure also has several HIV and biomarker datasets. GIS data are available for many countries and most datasets. You do have to submit a short approval application, which usually only takes a few days to process.

CARES: https://mycares.net/
The CARES (Cardiac Arrest Registry to Enhance Survival) dataset contains all cases of cardiac arrest in a defined geographic area. The ultimate goals of CARES is to help local EMS administrators and medical directors identify who is affected, when and where cardiac arrest events occur, which elements of the system are functioning properly and which elements are not, and how changes can be made to improve cardiac arrest outcomes. I have access to this data if you'd like to use it.

SEDAC: http://sedac.ciesin.columbia.edu/data/sets/browse
The Socioeconomic Data and Applications Center at Columbia has many spatially explicit, sub-national datasets on population processes and socioeconomic status from around the world. You can get data on poverty, infant mortality, population change, urban areas, EPA TRI and hazardous waste sites and much much more. I highly recommend browsing this site.

LA FANS: http://lasurvey.rand.org/about/
The Los Angeles Family and Neighborhood Survey (L.A.FANS) is a study of adults, teens, children, and neighborhoods in Los Angeles County. Our goal is to understand: how neighborhoods affect a variety of outcomes, including children's development and well-being and stress and health among children and adults. You can't get actual GIS data without an IRB approval, but they have public use datasets with contextual neighborhood data.

WHO-SAGE: http://www.who.int/healthinfo/systems/sage/en/index.html
The WHO Multi-Country Studies unit developed the Study on Global AGEing and Adult Health (SAGE) as part of a Longitudinal Survey Program to compile comprehensive longitudinal information on the health and well-being of adult populations and the ageing process. The core SAGE collects data on respondents aged 18+ years, with an emphasis on populations aged 50+ years, from nationally representative samples in six countries (China, Ghana, India, Mexico, Russia and South Africa). If you are interested in using these data come talk with me. They did collect GIS data, and I may be able to obtain it for you if you have a good study question.

HRSA ARF: http://arf.hrsa.gov/
The Area Resource File, maintained by the Health Resources and Services Administration, is a database containing more than 6,000 variables for each of the nation's counties. ARF contains

information on health facilities, health professions, measures of resource scarcity, health status, economic activity, health training programs, and socioeconomic and environmental characteristics. In addition, the basic file contains geographic codes and descriptors which enable it to be linked to many other files and to aggregate counties into various geographic groupings. If you want to use it, we'll buy it for the class.

EPA TRI: http://www.epa.gov/tri/tridata/preliminarydataset/
The EPA compiles and maintains data on regulated facilities and handle and emit toxic chemicals in the U.S. The goal of the Toxics Release Inventory Program is to provide communities with information about toxic chemical releases and waste management activities and to support informed decision making at all levels by industry, government, non-governmental organizations, and the public. Data have a latitude/longitude associated with them so they are easily mapped in a GIS. This is a historical database, so you can download data for many years.

U.S. Poverty Data:
http://www.hsph.harvard.edu/thegeocodingproject/webpage/monograph/povdata.htm
The Public Health Disparities Geocoding Project has ready-made data on census tract-level poverty in the use for 1980, 1990 and 2000 derived from the U.S. Census.

U.S. Census: http://www.census.gov/
Don't forget the U.S. Census! Tons of demographic data available for census blocks, block groups, tracts and counties. The Decennial Census is not the only source of data. The new American Community Survey has started to publish datasets for low levels of geography.

Minnesota Population Center IPUMS: http://www.ipums.org/
The Minnesota Population Center maintains the Integrated Public Use Microdata System. There is an international and U.S. system. IPUMS (US) has individual- and household-level census data for multiple years. IPUMS-International is an effort to inventory, preserve, harmonize, and disseminate census microdata from around the world. The data are coded and documented consistently across countries and over time to facilitate comparative research. IPUMS data are available free of charge through a web dissemination system. There is some geographic information, but it is not any lower than the county-level.

COURSE SCHEDULE

**** Expectations for each new module ****

At the beginning of each module, I will assign a set of articles to read. Before you come to class you are expected to have read these articles AND find one additional article on the module topic that you read, evaluate/critique and bring to class.

Module 1: Introduction to Spatial Demography, Health/Medical Geography

Assignment

You should spend a couple of hours on this. Go to your favorite online resource for searching journal articles (e.g., PubMed, POPLine, GeoBase, Web of Science, etc.). Conduct a literature search using terms and key words you feel describe concepts covered in this course. You can try two tactics:

1) Try to keep the keywords general so you find review articles that broadly discuss the "state of the science". Examples include: "spatial epidemiology", "spatial demography", "population geography", "medical geography", or "neighborhoods and health".
2) Search for specific topics (maybe the ones that interest you), such as: "obesity" and "neighborhood" or "cardiovascular" and "physical environment" or "migration" and "education".

Read the abstracts of the papers you find (you don't need to read the full papers, though you may want to read a few you find particularly interesting). As you read the abstracts think about the following questions:

- What study questions are researchers addressing?
- What methods are they using?
- What data do they use and what level of geography is it at?
- Are there certain diseases or population processes on which researchers appear to focus?
- What do the researchers mean by "place", "space" or "geography"?
- What are the major limitations they discuss?

Write down a set of your own keywords that specifically address the questions above. In the end, you should have a list of concepts that "describe" the major theories, methods and tools used when geography or space are important aspects of the research question.

You'll need two things for class: 1) Your list of keywords and an understanding of what they all mean…we'll be making a concept map in class, and 2) A copy of the article you thought was the most interesting one you found during your search.

Reading:

1. Voss, P. 2007. Demography as a spatial social science. *Population Research and Policy Review* 26: 457-476.

2. Goodchild, MF, L Anselin, RP Appelbaum and B.H. Harthorn. 2000. Toward Spatially Integrated Social Science. *Int Regional Science Review* 23(2): 139-159.
3. Reibel, M. 2007. Geographic Information Systems and Spatial Data Processing in Demography: A Review. *Population Research and Policy Review* 26: 601-608.
4. Entwisle, B, RR Rindfuss, SJ Walsh, TP Evans and SR Curran. 1997. Geographic Information Systems, spatial network analysis, and contraceptive choice. *Demography* 34(2): 171-187.
5. Rezaeian, M, G Dunn, S St Leger, L Appleby. 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *J Epidemiol Community Health* 61: 98-102.
6. Pearce, N. 1996. Traditional Epidemiology, Modern Epidemiology and Public Health. *Am J Pub Health* 86(5): 678-683.
7. Elliott, P and D Wartenberg. 2004. Spatial Epidemiology: Current Approaches and Future Challenges. *Env Health Persp* 112(9): 998-1006.
8. Glass, GE. 2000. Update: Spatial Aspects of Epidemiology: The Interface with Medical Geography. *Epidemiol Rev* 22(1): 136-139.
9. Rushton, G. 2003. Public Health, GIS and Spatial Analytic Tools. *Annu Rev Public Health* 24: 43-56.
10. Rosenberg, MW. 1998. Medical or Health Geography? Populations, Peoples and Places. *Int J Pop Geog* 4: 211-226.

Questions:

- In population science, demography and public health why do you think the "geographic perspective" (in general) and spatial analysis (in particular) took a back-burner for so long? What has changed?
- Are there similarities between spatial demography, spatial epidemiology, and medical geography? Differences?
- How do different disciplines seem to define "place" or "space"?
- Is the "ecological fallacy" always a fallacy or are there uses for ecological analyses?

Module 2: Neighborhoods and Health

Mini-lectures:

- (Spatial) Multi-level models
- Spatial cluster analysis

Reading:

1. Diez Roux, AV and C Mair. 2010. Neighborhoods and health. *Ann N Y Acad Sci* 1186:125-45.
2. Northridge, ME, ED Sclar and P Biswas. 2003. Sorting Out the Connections between the Built Environment and Health: A Conceptual Framework for Navigating Pathways and Planning Healthy Cities. *J Urban Health* 80(4): 556-568.

3. Cummins, SCJ, S Curtis, AV Diez-Roux, S. Macintyre 2007. Understanding and representing "place" in health research: a relational approach. *Soc Sci Med* 65:1825-38.
4. Oakes, J.M. 2004. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 58 (10):1929.
5. Diez Roux, AV. 2004. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med* 58 (10):1953.
6. Chaix, B, J Merlo, D Evans, C Leal, S Havard. 2009. Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Soc Sci Med* 69: 1306–1310.
7. Kwan, M. 2009. From place-based to people-based exposure measures. *Soc Sci Med* 69 (9):1311-1313.
8. McGinn, AP, KR Evenson, AH Herring, SL Huston, DA Rodriguez. 2007. Exploring associations between physical activity and perceived and objective measures of the built environment. *J Urban Health* **84:** 162–184.
9. Murray, ET, AV Diez Roux, M Carnethon, PL Lutsey, H Ni, ES O'Meara. 2010. Trajectories of Neighborhood Poverty and Associations with Subclinical Atherosclerosis and Associated Risk Factors: The Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol* 171(10): 1099-110.
10. Kling, JR, JB Liebman, LF Katz. 2007. Experimental Analysis of Neighborhood Effects. *Econometrica* 75(1): 83-119.
11. Morenoff, JD. 2003. Neighborhood mechanisms and the spatial dynamics of birth weight1. *American Journal of Sociology*, *108*(5), 976-1017.
12. Sampson, RJ, & WB Groves. 1989. Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology*, 774-802.

Questions:

- What relationships and processes are neighborhoods and health studies attempting to understand?
- How do researchers link social environments and health?
- What is Michael Oakes' major methodological criticism of multilevel modeling?
- What do you see as the biggest (or most interesting) challenge to N&H studies? How would you use spatial methods to address this challenge?
- Are the ways in which people model neighborhood effects truly capturing spatial relationships? If so, how?