GEO 866 (Fall 2019) - Spatial Data Analysis
Website: http://www.msu.edu/~ashton/classes/866

I. Instructors
Dr. Ashton Shortridge
235 Geography Building
phone: 432-3561 / email: ashton@msu.edu
office hours: Tuesday 10-noon, Wednesday 10-11 am, or by appointment. No drop-ins, please.

Ms. Rui Zhang, Teaching Assistant
19 Geography Building (basement)
email: zangr50@msu.edu
office hours: Tuesday & Thursday 2:30-4:00.

II. Time and Place
Tuesday / Thursday 8:30 - 9:50 am, 126 Geography Bldg (Lecture)
Lab: Tues 12:40-2:30 pm / Thur 10:20-12:10 pm, 201 Geography Bldg (Computer Lab)
Final Exam Slot: Tuesday, December 10, 7:45-9:45 am

III. Course Objectives
Spatial is special, and special forms of analysis are required for handling spatial data. Spatial
statistics is a cover-all term for a diverse set of methods that describe and model characteristics of
spatial data. In some cases spatial location is the only factor being analyzed (disease point
pattern). In others the primary interest concerns an attribute present everywhere but sampled only
at a subset of locations (digital elevation models). A third set involves the analysis of data
collected and stored in spatial zones (U.S. Census data). While these three cases are by no means
exhaustive, they do represent the wide range of applications that we will deal with in this course.

Our objective is to learn fundamental and applied statistical approaches to describe, model, and
analyze these three basic types of data. These techniques include:

- **point pattern analysis**
  (kernel density, K-family statistics, identifying hot spots)
- **methods for continuous data**
  (interpolation, trend surface analysis, geostatistics - kriging, variography)
- **spatial zone analysis**
  (OLS regression assumptions, autocorrelation measures, spatial regression models)

Some machine learning methods will also be introduced. You will develop a diverse and
powerful set of analytical techniques for gaining insight into geographical processes and patterns.

IV. Grades
5% **Class Participation**: Contribute to discussion on readings, ask and answer questions,
don't skip or come late to class.

25% **Final Project**: Proposal, Presentation, and Paper.
20% Exams: Two exams, Cover material from lecture and lab.

50% Homework: Lab-based exercises
   1. You obtain data + R script & questions in lab.
   2. You run the script, analyze results, and write short report.
   3. Reports typically due by the following week's lab section.

Writeups are expected to answer every question in a brief report-style format. Grading will be based on quality of your responses to the questions, measured by their correctness and clarity. Barring special circumstances, late homeworks, project preproposals, etc. will not be graded.

Plagiarism is the use of others' ideas without identifying the source. It is one of the most serious academic offenses. Carefully read and understand MSU's policy on this matter, and understand that if I find evidence of plagiarism, improperly attributed group work, or cheating on an exam, I will issue a failing grade on the assignment or for the course, and report the conduct to University authorities. For more, see: [https://msu.edu/unit/ombud/academic-integrity/plagiarism-policy.html](https://msu.edu/unit/ombud/academic-integrity/plagiarism-policy.html)

V. The Book
No textbook is required; useful resources are listed below. There will be occasional readings.

VI. The Software
We will be using the R statistical package. R is really a specialized programming language that has two major advantages over most other options:
   • It is open-source, free software, so anyone can install and use it at no cost or restrictions.
   • Hundreds of specialized extensions have been developed to make R even more versatile.
   We will be using some of these add-on libraries to employ powerful spatial statistics

R definitely has a learning curve associated with it. A primary objective of this course is to gain experience at using R to explore and analyze data. More details about R, including very brief installation instructions, are available from the course web page, and at the official R site.

VII. The Project
The project involves original work using some spatial data and several of the techniques discussed in class. I can envision two separate types of projects:
   • You may have a specific problem in mind, and some data you want to analyze. For example, if your current research interest is studying spatial yield variation across a corn field, and you intend to use this class to help you develop techniques to tackle this problem, then it would make sense to use some existing yield data for your project.
   • You may be interested in exploring a statistical method more deeply. For example, you might investigate the sensitivity of, say, simple kriging to variability in the spatial covariance model (this particular example will make more sense later).

In either case, limiting the size of the problem is a good idea. Some people discover they have serious data collection or input problems rather late in the semester, and end up with limited time to perform the analysis, do the writeup, and develop the presentation. Most importantly, the project must concern methods covered in class.

Projects can be individual or group. Groups consist of two or possibly more people; each member will receive the same grade.
Project Deadlines

Wednesday, November 20, 5pm: Project Preproposals Due
Monday, December 1, 5 pm: Project Proposals Due
Tuesday, December 10: Presentations
Wednesday, December 11, 5 pm: Paper Due

Proposal (5% overall grade)
The proposal should be a 1-2 page typed document that accomplishes four things:
1. identifies the research problem and the research team. About five cited references are expected.
2. indicates the data required to work on the problem, whether you have it or not, etc.
3. outlines techniques and methods you will use on the data. If you will not be using R, explain why and list the package(s) you will use
4. indicates a time line for the project (data input, analysis, writeup, presentation design)

Large project deviations from what you proposed must be explained in the final report.

Presentation (10% overall grade)
We will employ a conference-style format, a rigorously enforced time limit. Laptop and LCD projector will be provided. I will expect a professional, polished, and rehearsed performance. Students from previous GEO 866 courses have gone on to present their work at major academic conferences, and that is the standard I am shooting for this semester.

Paper (10% overall grade)
The paper should be 5-8 pages in length. It is a research report; write in an appropriate style, include relevant references, and express yourself clearly and succinctly. The paper should indicate your research problem and significance, describe your methodology, and report on your findings. It is particularly important that you describe the statistical techniques you employ and explain why you are using them. Figures and tables should be included if they are helpful. If your report includes ideas from other people or works, you need to cite them. Failure to indicate sources is plagiarism. If you are unsure, reference or talk to me about it.

VIII. Other Resources
A decent statistics text (e.g. Ott, An Introduction to Statistical Methods & Data Analysis)
Online R interest group mailing lists, tutorials and manuals (see web page)

IX. Weekly Outline
A weekly outline of lecture notes will be available. This is not a substitute for attending lecture! Students who do not attend regularly fail this course!