UNT | Geography and the Environment

GEOG 5190
Advanced Quantitative Techniques
Spring, 2018

Thursday 6:00-8:50pm
ENV 340

Catalog Description: Application of advanced statistical procedures including multivariate techniques to analysis of point and areal patterns and spatial data.

My Description: Geography is marked by a diversity in subject matter that includes physical (environmental), human (socio-economic), and integrated (human-physical) topics of inquiry. Therefore, there are a variety of advanced analytical methods that geographers can employ in their studies. This course introduces some of the most commonly employed of these advanced techniques.

The course focuses specifically on multivariate and spatial statistics. It examines multivariate methods used widely by geographers and other scientists, including an introduction to elementary matrix algebra, multiple correlation and regression, factor analysis, cluster analysis, multidimensional scaling, and other terminology whose use in social settings can be very impressive.

This course mixes opportunities for hands-on experience with extensive periods devoted to gaining practical understanding. You should leave this course with confidence in the methods we have discussed, and an appreciation for how these statistical methods are applied in geographic research.

Instructor: Dr. Murray Rice
Class: Thursday, 6:00 to 8:50 pm
Classroom: ENV 340
Office: ENV 310G
Office Hours: Tuesday, 10:00 am to 12:00 pm
Office Telephone: 565-3861
E-Mail: rice@unt.edu

Class Web Page:
http://www.murrayrice.com/geog-5190.html

Course Philosophy: I am here to organize the course and introduce you to the topics and readings we will examine. I don't have all the answers and I don't pretend to have all the answers, but I will share with you from what I know. I will do my best to make the course interesting, relevant, and challenging.

This being said, it's important that you understand that you have the most important role in making GEOG 5190 a success for you. You will determine how much you actually get out of this course. Doing the readings outlined, and coming to class and labs ready to think and participate in group discussions puts you in the best position to benefit from what this course offers. I encourage you to make full use of the learning opportunities that this class presents.

Grading: It is not essential to pass any particular element to pass the course, but relative success in each will affect your final grade.

Labs	As outlined in class schedule, 5 labs x 10% each	50%
Class Participation	Attendance, Questions, Discussion Participation	10%
Project	Proposal (due in Week 8)	5%
	Report (due in Week 14)	25%
	Presentation (occurring in Weeks 14 and 15)	10%

The University of North Texas makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with an accommodation letter to be delivered to faculty to begin a private discussion regarding your specific needs in a course. You may request accommodations at any time, however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the Office of Disability Accommodation website at http://www.unt.edu/oda. You may also contact them by phone at 940.565.4323.

My Promise to You: There are some things that are important for you to understand in terms of how you and I will communicate this semester. I'm saying this upfront because I want to be clear about what I can promise to do for you, and what I cannot promise to do.

> **(a)** I promise to gladly answer focused, specific questions you have related to the course. A focused question might be something like "I don't understand this comment you made on my last lab, can you please help me understand?", or "Can you explain to me again what the difference is between cluster analysis and factor analysis?"
>
> **(b)** I cannot promise to answer broad, open-ended questions, even if they relate directly to the course. An open-ended question would include something like "Here is a draft of my answers for the spatial statistics lab. Can you give me some comments?" or "I don't understand multiple regression, can you explain it again?". Unfocused questions like these don't show a lot of thought on the part of the person doing the asking, or a lot of respect for the time of the person to whom the question is directed. Please think through your questions before asking.

(c) Related to (b) above, I will not read through draft versions of labs or reports (or revised versions of material already submitted) when you simply want general feedback and suggestions. Please do not hand in entire drafts or re-done work for me to read and comment.

However: feel free to bring a draft or some re-done work if you have questions, but make sure you have thoughtful, specific questions to ask when you do so.

I really do welcome (and encourage) questions and discussion, despite the rather formal language I use above.

Required Text: The details of how to access the required readings for this course will be discussed in the first class of the semester. However, please note that several required readings come from a very readable guide to multivariate statistics assembled by Dr. Mike Wulder, a research scientist with the Canadian Forest Service Branch of Natural Resources Canada. The Wulder readings were originally accessible through Dr. Wulder's work website, but are now offline. You can access these readings together with the other required readings for the course through our GEOG 5190 directory on the CSAM "class" network drive (accessible from our CSAM lab computers). In the course schedule I refer to the Wulder readings as "Wulder, section on (fill in the topic)". Other readings may be identified as needed.

Optional Readings: The following are good, comprehensive resources relating to the various statistical methods we will discuss in this course, and are available at the UNT library.

Davis, J.C. (1986) *Statistics and Data Analysis in Geology*, 2nd Edition. Wiley.
(UNT Library Call Number: QE48.8 .D38 1986 c.2)

Rogerson, P.A. (2001) *Statistical Methods for Geography*, 2nd Edition. Sage.
(UNT Library Call Number: G70.3 .R64 2001)

I would view these as reference materials to consult as you wish, rather than texts that are absolutely necessary to review each week.

Labs: Along with your project, labs form the core of your work in this course. The labs will provide you with access to SPSS, a powerful statistical software package for PCs. The assignments will be based on a particular geographic issue or problem and your task will be to apply a particular statistical approach to solving that problem. Labs will involve both "number crunching" and a written report of your efforts. Therefore, your grading will be balanced to reflect the importance of each.

Late Policy: Anything handed in late* will be subject to a 10% penalty. Late work will not be accepted after graded work has been handed back. Graded work is usually returned one week after the due date. I will grant exceptions to the above if you provide documentation substantiating a valid personal emergency. I am the judge of what constitutes a "valid personal emergency".

---

* Late = "after the beginning of class on the assigned due date".

Project: You will complete a major project as a key component of your course work. Your task in the first two weeks of the semester is to identify a geographical *theory* and *database* that can be analyzed using the methods we will be examining in class. Three elements comprise the project.

1. Proposal/Overview: Your project proposal/overview summarizes your project topic choice. This one- to two-page document needs to:

1. define the general topic area for your project (population geography, climatology, zooarchaeology, etc.)
2. explain, in basic terms, why anyone would be interested in this topic area
3. discuss the database you will use: source of data, any characteristics of the data you know now, availability of the database to you (can you access it for this course?)
4. give your best assessment of which technique from class will apply to your analysis
5. be readable

2. Written Report: Your report documents the analysis you have completed. See a brief summary of report guidelines provided on the following page (more in class).

3. Presentation: This 15 minute (plus 5 minutes for questions) presentation summarizes your report findings and gives you experience at public speaking in a research context.

Key point you must understand as you plan for your project: the project must focus on application of one of the techniques taught in this class. There are many other wonderful analytical tools not taught in this class, but I want you to use a method we cover in this course. There are no exceptions to this class policy.

Attendance Policy: Full attendance is essential and required in this course.

- Unexcused, undocumented absences will result in a 5% penalty per absence from your final class percentage grade: this deduction is taken from your 10% "class participation" grade, so two unexcused absences = a grade of zero for your class participation grade.
- More than two unexcused absences will result your automatic withdrawal from the course.
- It is *your* responsibility to provide reasons/documentation for absences for there to be any possibility of having an absence counted as "excused" (I will not approach you to ask).

Class Dates: Presentation and due dates are final and will not be changed regardless of student circumstances (except for emergencies as outlined above). It is your responsibility to plan so outside activities will not conflict with class dates.

Extra Credit: The Department of Geography does not allow extra credit assignments (work not specified on a course syllabus).

Guidelines for the Term Project Report: Below are a few key things that you should note about the paper portion of your term project.

1. The paper must be 12 to 22 double-spaced pages in length (one double-spaced page is approximately 250 words). Maps, diagrams, and illustrative material are NOT included in this length. However, please note that such material must be incorporated into, and discussed in, the body of the text (as in a book, article, or thesis). Do not simply place this material at the end of the paper.

2. The paper should have a title page, table of contents, and references. These are also not included in the above page figure. The references should be referred to in the text, and listed at the end of the paper, using *The Professional Geographer*'s referencing system (look at a recent copy of it in our library or via our library website if you are not familiar with this respected geography journal).

3. The paper should have an introduction and conclusion. The general idea is simple: introduce a problem/issue, solve/analyze it using your data set, and discuss what your results mean.

4. The paper should include section headings to identify the various components of your report (such as introduction, results, and conclusion). These sections do not have to be numbered, but if you do number them, use a numbering system like 1.0, 1.1, 1.1.1, 1.2, 1.3, 2.0, 2.1, etc.

5. The major emphasis of your paper should be on the analysis of your data set. You must analyze your data using some technique or techniques that we will cover in this course. The data set can come from a book, another course, a professor, a friend, or be collected just for this class. The data set should be geographical, and it should be of interest to you. As soon as you're thinking of adopting a data set, drop me a quick e-mail – I'm interested in knowing what you are doing. Please note: I completely appreciate that coming up with a dataset may be challenging, but do not complain to me that this data selection exercise is unfair. You are a UNT graduate student, and working with complex datasets is at the core of your experience here. If you are having a lot of trouble identifying a suitable dataset, then maybe you should not be enrolled in a graduate program.

6. The paper should contain a section explaining how the data were obtained and a discussion of their characteristics – especially those characteristics that might cause problems for the analysis. You should explain why you are doing the analysis (i.e. why anyone would be interested in it), and what you expect to find based on the theory you are using and your knowledge of the subject.

7. You should explain all steps of the statistical procedure and why you are using that particular procedure. You should include all computational steps, formulae, and/or relevant output. You should explain the results of the analysis and what they mean in substantive terms. Relate this discussion to the data, not just to the statistical techniques you use.

8. An attractive presentation is important. Treat the paper as a consultant's report being presented to a corporate or governmental client.

9. Computer output (from SPSS or other statistical/GIS software) should be integrated into the text (not in appendices at the end). You only need to include the critical information, not the entire output. Selecting what is critical, and what is not important, is an important part of the exercise.

10. A conclusion that summarizes your results and their meaning must be included at the end of the report. Invest the time to make this a good, coherent discussion.

Courtesy: You can expect to be treated with respect as a student in this course. I promise to do my best to make sure this happens for everyone in the class. I expect the same in return from you, for both me and the other students around you. A few specifics to consider include the following:

- Please arrive before the class begins.
- If you are late, please disturb the other students as little as possible when you arrive.
- Please avoid talking or engaging in other behavior during the class that disturbs those around you.
- If you cannot stay awake, please stay home and sleep.
- Please do not come to class do anything other than participate fully in the class.
- Leave during class only if it is really necessary (i.e. restroom breaks).
- Please turn off your cell phone before class begins.

Academic Dishonesty: Academic dishonesty in this course will be penalized according to University of North Texas rules and regulations, including a potential mark of 0 on a test or assignment, a grade of "F" in the class, or suspension or expulsion from the university, depending on the nature and circumstances of the dishonesty. Learning what is dishonest and how to stay away from such conduct is good preparation for a successful career.

To help you avoid academically dishonest behavior, the Center for Student Rights and Responsibilities at the University of North Texas has developed a definition of academic dishonesty and a set of strategies to protect yourself from being accused of academically dishonest behavior.

The following is a summary of definitions and strategies from CSRR:

Forms of Academic Dishonesty

- **Cheating:** intentionally using or attempting to use unauthorized materials, information, or study aids in any academic exercise. The term academic exercise includes all forms of work submitted for credit or hours.
- **Plagiarism:** the deliberate adoption or reproduction of ideas, words or statements of another person as one's own without acknowledgement.
- **Fabrication:** intentional and unauthorized falsification or invention of any information or citation in an academic exercise.
- **Facilitating academic dishonesty:** intentionally or knowingly helping or attempting to help another to violate a provision of the institutional code of academic integrity.

Proactive strategies to protect yourself from charges of academic dishonesty:

1. Prepare thoroughly for examinations and assignments.
2. Take the initiative to prevent other students from copying your exam or assignments, e.g., shield your answer sheet during examinations, do not lend assignments to be turned in to other students.
3. Check your faculty member's course syllabus for a section dealing with academic dishonesty for that course. There may be special requirements. If you cannot find a written section in the syllabus, ask the faculty member what his/her expectations are.
4. Consult the Code of Student Conduct for a detailed definition of academic dishonesty.
5. Do not look in the direction of other students' papers during examinations.
6. Utilize a recognized handbook for instruction on citing source materials in papers.
7. Consult with individual faculty or academic departments when in doubt.
8. Utilize the services of the University Writing Center, located in room 105 of the Auditorium Building, for assistance in preparing papers.
9. Discourage dishonesty among other students.
10. Refuse to assist students who cheat.

Course Schedule

Please read the following schedule carefully. Note the weeks when labs are to be done or when student work is due. The weeks marked "LAB" will be completely lab-based (no lecture). Lab reports are due at the beginning of class the week following the lab. Please recall from the attendance policy outlined earlier in this syllabus that full attendance is required in this course.

WEEK	TOPIC
1 (Jan 18)	**Introduction to Course / Discussion of Term Projects** *Reading:* Simmons et al., "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant"; Kerski, "Understanding data: It is critical!" (see syllabus and handouts page on course website for all reading links)
2 (Jan 25)	**Why Multivariate Statistics? / A Quick Tour of Matrix Algebra** *Reading:* Hair et al., Chapter 1 (pages 1-17); Lang, Chapter 2 (pages 42-58); Griffith and Amrhein, Chapter 3 (pages 55-58); *Data Screening Handout* document (see handouts page on course website)
3 (Feb 1)	**Introduction to SPSS** **LAB** *Reading:* Field, Chapter 2 (pages 37-62)
4 (Feb 8)	**Spatial Statistics: Spatial Pattern and Autocorrelation** *Reading:* Ebdon, Sections 7.4 and 7.5 (pages 143-163); Wulder, Section on Spatial Autocorrelation; Tenenbaum and Waters, "Spatial patterns of subprime mortgages by local banks, nonlocal banks, and independents in the continental US" (see syllabus and handouts page on course website) *Supplementary Reading (Not Required, but Helpful):* Boots and Getis (*Point Pattern Analysis* E-Book), Odland (*Spatial Autocorrelation* E-Book)
5 (Feb 15)	**Spatial Autocorrelation** **LAB**
6 (Feb 22)	**Multiple Correlation and Regression** *Reading:* Field, Sections 5.5 - 5.7 (pages 157-184); Wulder, Section on Multiple Correlation and Regression; *Dealing with Normality Handout* document (see syllabus and handouts page on course website)
7 (Mar 1)	**Multiple Regression** **LAB**

8 (Mar 8) **Spatial Pattern: Smoothing and Trend Surface Analysis**
(Project Proposal Due Today)
Reading: Wren, "Trend Surface Analysis – A Review" (see syllabus and handouts page on course website)

Supplementary Reading (Not Required, but Helpful): Unwin, *An Introduction to Trend Surface Analysis* (dated but classic E-Book)

Mar 12-16 Spring Break: No Classes (Enjoy Your Week!)

9 (Mar 22) **Principal Components and Factor Analysis**
Reading: Wulder, Section on Principal Components and Factor Analysis (web); Field, Chapter 15 (pages 619-637); Demšara *et al*., "Principal component analysis on spatial data: An overview"

10 (Mar 29) **Factor Analysis/Multidimensional Scaling** **LAB**
The AAG annual conference is happening this week, which involves many people from our department. In class we will discuss the arrangements we will make for everyone this week, whether or not you are attending.

11 (Apr 5) **Cluster Analysis**
Reading: Wulder, Section on Cluster Analysis; Griffith and Amrhein, Chapter 8 (pages 207-232); *Cluster Analysis Handout* document and Esri *Tapestry Segmentation Reference Guide* document (see syllabus and handouts page on course website)

12 (Apr 12) **Project work week**
No formal class meeting this week, but I expect you to be making specific progress toward the completion of your term project during this time. The AAG conference is this week, so some in the class may also be traveling.

13 (Apr 19) **Cluster Analysis** **LAB**

14 (Apr 26) **First Week of Project Presentations**
(Project Reports Due Today)
Attendance is mandatory, unless excused for a very good reason (foreseeable reasons must be presented for approval in advance)

15 (May 3) **Second Week of Project Presentations**
(Graded Project Reports Returned)
Attendance is mandatory, unless excused for a very good reason (foreseeable reasons must be presented for approval in advance)

As explained in the attendance policy outlined earlier in this document, unexcused absences from any class in this course will result in a 5% deduction from your final course grade. Absences will be excused only in the most severe of circumstances for the final project presentation meetings in this course (scheduled here for weeks 14 and 15).

Reading References

The following are full references for the articles, books, and other readings listed in the preceding course schedule. Links for many of these references are also provided on the course website.

Boots, Barry N., and Art Getis (1987) *Spatial Autocorrelation*. Beverly Hills: Sage.

Demšara, Urška, Paul Harris, Chris Brunsdon, A. Stewart Fotheringham, and Sean McLoone (2013) "Principal component analysis on spatial data: An overview." *Annals of the Association of American Geographers* 103(1), 106-128.

Ebdon, David (1987) *Statistics in Geography*, 2nd Edition. Oxford: Blackwell.

Esri (2012) *Tapestry Segmentation Reference Guide*. Redlands, CA: Esri.

Field, Andy (2005) *Discovering Statistics Using SPSS*, 2nd Edition. London: Sage.

Griffith, Daniel A. and Carl G. Amrhein (1997) *Multivariate Statistical Analysis for Geographers*. Upper Saddle River: Prentice Hall.

Hair, Joseph F., Rolph E. Anderson, Ronald L. Tatham, and William C. Black (1995) *Multivariate Data Analysis*, 4th Edition. Upper Saddle River: Prentice-Hall.

Kerski, Joseph (2015) Understanding data: It is critical! *Spatial Reserves*, November 22. Accessible at: https://spatialreserves.wordpress.com/2015/11/22/understanding-your-data-it-is-critical/.

Lang, Serge (1986) *Introduction to Linear Algebra*. New York: Springer-Verlag.

Odland, John (1987) *Spatial Autocorrelation*. Beverly Hills: Sage.

Simmons, J.P., J.D. Nelson, and U. Simonsohn (2011) "False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological Science* 20(10), 1-8.

Tenenbaum, H., and N.M. Waters (2011) "Spatial patterns of subprime mortgages by local banks, nonlocal banks, and independents in the continental US." *Environment and Planning A* 43(8), 1761-1778.

Unwin, David J. (1978) *An Introduction to Trend Surface Analysis*. CATMOG 5. University of East Anglia, UK.

Wulder, Mike (2009) *A Practical Guide to the Use of Selected Multivariate Statistics*. Formerly Online, Now Archived in the GEOG 5190 directory (CSAM Computer Labs).

Wren, Easton (1973) "Trend Surface Analysis – A Review". *Canadian Journal of Exploration Geophysics*, December: 39-44.

Things You Should Already Know

The following are concepts covered in UNT Geography's GEOG 4185/5185 and in similar statistics courses in other geography and social science departments. I'm expecting you to already be familiar with these concepts as a starting point for our discussion in GEOG 5190. We will not necessarily be using each and every one of these terms explicitly in this course, but you will be at a disadvantage in learning the new concepts presented in GEOG 5190 if you do not understand a large number of the concepts listed here.

If you are not familiar with one or more of these concepts, please let me know so I can point you to some resources to help you catch yourself up. If you are not familiar with a large number of these concepts, then I strongly recommend that you withdraw from this course and take GEOG 5185 to gain the background you need to succeed in GEOG 5190 in a future semester. We will not spend any substantial class time reviewing the concepts listed below in this course.

Basic Statistical Concepts

Uses of Statistics
Measurement Scales: Nominal, Ordinal, Interval, Ratio
Probability Distributions

Description

Central Tendency
Dispersion
Skewness and Kurtosis

Samples and Sampling

Sampling Methods
Estimates from Samples
Sample Size

Comparative Methods and Significance Testing

Steps involved in a formal statistical test
K-S Test
Runs Test
U Test
t Test
Chi Square Test
H Test

Relationships

Correlation (basic concept in a two-variable context)

Trends

Simple Linear Regression (two variable)

Brief Summary of Major Methods Covered in GEOG 5190

Spatial Autocorrelation

A statistical test for the existence of some form of spatial pattern. Spatial autocorrelation answers the question: does a given distribution have a statistically-significant spatial pattern (dispersed or clustered), or is it statistically indistinguishable from a random pattern? We will look at spatial distributions involving points and spatial distributions involving areas.

Correlation and Regression*

Correlation: how strong is the relationship between two variables?
Regression: prediction of one variable based on another

Multiple Correlation and Regression

Examine correlation and regression relationships among three or more variables. Use of these methodologies allows us to create highly complex and valuable predictive models, as well as to understand the relationships that characterize complex datasets with up to 10 variables.

Spatial Pattern: Smoothing and Trend Surface Analysis

Uncover and better represent overall spatial trends. For example, these methodologies are useful if you need to simplify and extract the major features of a complex spatial distribution across a region. Trend surface analysis is conceptually related to regression.

Factor Analysis

Simplify highly complex, multivariate situations by identifying major influences. This methodology is particularly valuable with datasets encompassing 10 or more variables.

Multidimensional Scaling**

Simplify and understand a complex database by analyzing dissimilarities among database observations and creating a representation ("map") of similar and dissimilar observations.

Cluster Analysis

A grouping or classification methodology. Cluster analysis facilitates the identification of similar and dissimilar objects, people, places, etc. where each object is characterized by a lengthy list of characteristics (e.g. identify groupings of similar census tracts across the U.S. where each census tract has dozens or hundreds of attributes associated with it)

---

* Discussed briefly as part of an introduction to multiple correlation and regression, but you should already be familiar with the basics of these methods when you begin GEOG 5190.

** Discussed briefly as part of the introduction to Factor Analysis.

My goal in this course is for your time here to be of great value to you. This course provides a "tool-kit" of concepts and skills you should find to be a helpful starting-point for careers in geography and related earth- and social-science disciplines. My hope is that this course will provide you with knowledge and confidence in statistics that can help you in many ways, regardless of where your life and career takes you in the future.

Please do not hesitate to let me know if there is anything else we can do, beyond what you see in this course package, that could be useful in preparing you for the challenges you will face in your future.