

INTA 4803/6450: Data Analytics and Security

Fall 2024

(The readings in this syllabus are subject to change)

Instructor: Austin Beacham (austinbeacham@gatech.edu)

Lectures: Tuesday/Thursday 5-6:15pm

Location: Habersham G17

Office Hours: TBD (will discuss with class)

Online Content: <https://gatech.instructure.com/courses/421432>

Overview: The use of data analytics has exploded across almost every facet of life, from private businesses to policymaking to academic research. We have new techniques and methods to answer questions both old and new.

This course has two primary objectives. First, it introduces students to a wide range of methods and tools for empirical data analysis in social science research. Second, it aims to develop students' skills in identifying good social science research questions and creating research designs that can best answer those questions, particularly focused on research topics related to Comparative Politics and International Relations. In other words, the goal of the course is to help you identify good questions to ask about the social world, and know what methods you can use to answer those questions. Examples of tools that we will discuss include geospatial data analysis, remote sensing, text as data, large language models (LLMs), and visual analytics. To be clear, you are not expected to master any of these tools, but rather become familiar with them enough to be able to more easily apply them in your own life should you choose to dedicate more time to them. We will also be reading and discussing published academic research that uses the methods that we will learn, to provide students examples of what research can look like.

By the end of this seminar, students will be proficient in the basics of analysis of social science data using a number of tools, able to identify good research questions and designs, able to clearly and effectively communicate their results, and have the necessary knowledge to take additional seminars on more advanced topics not offered by the Nunn School.

Prerequisites: There are no formal prerequisites for this course, but I recommend that students have taken INTA 2010: Empirical Methods or a similar course before taking this seminar. While the mathematical demands are fairly light, the course's focus on research design and interpreting results of data analysis means that having familiarity with the basics of empirical methods is helpful. As this is an upper division/Master's course, it is assumed that students have the mathematical background to progress through the material a bit faster than a purely

introductory course. If you have concerns about this, please talk to me before deciding the course is not right for you.

Software: We will use a number of software tools in this course, and I will give you a heads up on when you will need them during the semester. However, we will definitely be using the open-source statistical software R, as well as the open-source geospatial analysis software QGIS. I recommend downloading and playing around with these tools sooner rather than later, as they can require some troubleshooting to get working on some operating systems.

Reading: There are no required textbooks for this course. I will provide PDFs for all articles that are assigned, which are available through the Georgia Tech library. I will note that this course is relatively light on reading, so students are expected to read **in full, before class**, and to **carefully think** about the research design and methods used in the articles. Also note that I will sometimes only assign part of an article — pay close attention to the Canvas file name.

Grading and Requirements

The course will consist of the following:

- 3 Homework Assignments (15% each, 45% total)
- Research Proposal (40%)
- Participation (15%)

Homework assignments: The homework assignments are designed to ensure that students are staying up to date with the reading and understanding the course material. They also help me better tailor the rest of the course depending on how students perform. I will release the homework assignments two weeks before the due date to ensure that you have plenty of time to complete them. All assignments are to be submitted electronically through the Canvas course portal.

Research Proposal: Students will apply the tools from this course to identify a question related to Comparative Politics or International Relations that is interesting to them. They will describe their theorized answer to the question, and explain **in detail** how they would set up a research design to answer the question/test their theory. The research proposal should look like a fleshed out academic journal-style research article that is missing the results section (students may show preliminary results if they have them). This means that students should include descriptive visualizations of the data that they would analyze further in a full paper. Students may collect their own data or propose to use existing data in a novel way. As co-authorship is common in the field, students may form groups of no more than three co-authors to submit a single final paper. Expectations for polish and quality in co-authored papers will be commensurately higher. More detail will be given about the research proposal as the course progresses.

Participation and Attendance: This is an in-person course and as such there will not be remote attendance options nor recorded lectures (with some scheduled exceptions). We will have in-depth class discussions throughout the semester about research design and tools, as well as the specifics of individual articles that we read. I will provide multiple avenues to participate in these discussions. Students are expected to have completed the readings **prior** to class. Preparedness and level of engagement are both factors of the participation grade. I reserve the right to call on individuals at random or administer reading quizzes, although I prefer not to. Additionally, attendance is part of the participation grade. You are allowed three unexcused absences; after this, failure to attend will negatively impact your participation grade.

Class Policies

Late Assignments Late homework assignments will be accepted with a 5 point (out of 100) reduction for each day that it is late, up to one week. Late research proposals cannot be accepted due to the turnaround time with final grading at the end of the semester. If you are sick or have extenuating circumstances, please let me know. I am happy to work with you if I am aware of issues ahead of time.

Diversity and Inclusivity Statement The Institute does not discriminate against individuals on the basis of race, color, religion, sex, national origin, age, disability, sexual orientation, gender identity, or veteran status in the administration of admissions policies, educational policies, employment policies, or any other Institute governed programs and activities. The Institute's equal opportunity and non-discrimination policy applies to every member of the Institute community. The Institute's affirmative action program, Title IX program, and related policies are developed in compliance with applicable law. Pursuant to Title IX, the Institute does not discriminate on the basis of sex in its education programs and activities. As such, the Institute does not tolerate any kind of gender-based discrimination or harassment, which includes sexual violence, sexual harassment, and gender-based harassment. Inquiries concerning the Institute's application of or compliance with Title IX may be directed to the Title IX Coordinator, Burns Newsome, burnsnewsome@gatech.edu, 404-385-5151.

Accommodation for Students with Disabilities Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with the Office of Disability Services at Suite 123, Smithgall Student Services Building, 353 Ferst Drive, 404-894-2563 (Voice); 404-894-1664 (TDD). For more information on Georgia Tech's policy on working with students with disabilities, please review the Office of Disability Service's web page at <http://disabilityservices.gatech.edu/>. The Office of Disability Services collaborates with students, faculty, and staff to create a campus environment that is usable, equitable, sustainable and inclusive of all members of the Georgia Tech community. Disability as an aspect of diversity that is integral to society and Georgia Tech. If

students encounter academic, physical, technological, or other barriers on campus, the Disability Services team is available to collaboratively find creative solutions and implement reasonable accommodations. I am very happy to meet during my office hours to discuss your needs.

Academic Integrity Academic dishonesty in the form of cheating or plagiarism will not be tolerated. In brief, plagiarism is defined, for the purposes of this class, as: copying, borrowing, or appropriating another person's work and presenting it as your own in a paper or oral presentation, deliberately or by accident. Acts of plagiarism will be reported in accordance with the Honor Code. In order to avoid being charged with plagiarism, if you use the words, ideas, phrasing, charts, graphs, or data of another person or from published material, then you must either: 1) use quotation marks around the words and cite the source, or 2) paraphrase or summarize acceptable using your own words and cite the source. The plagiarism policy is not restricted to books, but also applies to video and audio content, websites, blogs, wiki's, and podcasts. Plagiarism includes putting your name on a group project to which you have minimally contributed. For information on Georgia Tech's Academic Honor Code, please visit <https://policylibrary.gatech.edu/student-life/academic-honor-code>. Any student suspect of cheating or plagiarizing on an assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations. The student will also receive a grade of zero on the assignment at the instructor's discretion.

Use of Generative Artificial Intelligence (AI) Tools You may use generative AI programs, e.g. ChatGPT, to help generate ideas and brainstorm. You should be aware that the material generated by these programs may be inaccurate, incomplete, biased or otherwise problematic. Also, use of these tools may stifle your own independent thinking and creativity.

Generative AI derives its output from previously created texts from other sources that the models were trained on yet doesn't cite sources. Per GaTech's Honor Code, you may not submit any work generated by an AI program as your own. If you include material generated by an AI program, it should be cited like any other reference material (with due consideration for the quality of the reference, which may be poor). When/if you use AI platforms in your assignments, please write a note to clarify where in your process you used AI, include the prompt used to generate the material, and which platform(s) you used. See this article for how to cite AI properly: <https://apastyle.apa.org/blog/how-to-cite-chatgpt>. NOTE : you should **not** use AI to generate the text of your final research proposal. You will be reported to the Office of Student Integrity if AI use is detected or suspected.

Course Schedule

	Tuesday	Thursday
Week 1	August 20: Introduction	August 22: Fundamentals of research design “The Formula” Reading: Clark (2020)
Week 2	August 27: Fundamentals of research design How to ask a good research question Reading: Zinnes (1980)	August 29: Wrapping up research design Research in the Wild 1
Week 3	September 3: Cross national, cross sectional (CSNS) Visualization 1: simple scatterplots in R Sample articles: Battig and Bernauer (2009)	September 5 (Recorded lecture, instructor traveling for conference): CSNS continued Research in the Wild 2
Week 4	September 10: Cross national time series Sample articles: Gartzke (2007)	September 12: Visualization 2: time series visualization
Week 5	September 17: Other forms of data: dyads Sample articles: Chatagnier and Kavaklı (2017)	September 19: Other forms of data: hazard models Visualization 3: decay curves Sample articles: Bernauer et al. (2013)
Week 6	September 24: Other forms of data: network analysis Visualization 4: Networks Sample articles: Fowler youtube video , Kinne (2018)	September 26: Networks, continued Research in the Wild 2
Week 7	October 1: Introduction to geospatial data	October 3: QGIS Visualization 5: Maps
Week 8	October 8: Geospatial data, continued	October 10: Geospatial data, continued Research in the Wild 3
Week 9	October 15 (No class — school holiday)	October 17: Remote sensing: Google Earth Explorer Visualization 6: Satellite imagery
Week 10	October 22: Remote sensing, continued: Google Earth Engine	October 24: Text as data Sample readings: Hulme (2024)

	Visualization 7: False color bands	
Week 11	October 29 (Recorded lecture, instructor attending conference): Text as data, continued Visualization 7: Word clouds	October 31: Visual analytics Tableau Visualization 8: Tableau! Research in the Wild 4
Week 12	November 5 (No class — election day, go vote)	November 7: Visual analytics, continued
Week 13	November 12: Large language models	November 14: Large language models, continued
Week 14	November 19: Large language models + visual analytics	November 22: Other research methods: grab bag Research in the Wild 5
Week 15	November 27: Final project presentations	November 29 (No class — school holiday)
Week “15”	December 3: Final project presentations	

FINAL RESEARCH PROPOSAL DUE 11:59pm ET on Tuesday, December 10th