



THE UNIVERSITY OF TEXAS AT AUSTIN
SCHOOL OF
information

INF 310D: Introduction to Human-Centered Data Science

FACULTY INFORMATION:

Tony H. Grubestic (Instructor)
Geoinformatics & Policy Analytics Lab (GPAL)
1616 Guadalupe Suite #5.202
Austin, TX 78701
Email: grubestic@utexas.edu

Jake Nelson (Instructor)
Geoinformatics & Policy Analytics Lab (GPAL)
1616 Guadalupe Suite #5.530
Email: jake.nelson@austin.utexas.edu

COURSE MEETING TIMES AND LOCATION:

Room: Nope... Online!!
Time: Tuesday and Thursday, 11:00am – 12:30pm

OFFICE HOURS AND INSTRUCTOR COMMUNICATION:

Tony: By appointment
Jake: By appointment

This course uses a “Three Before Me” policy for student to faculty or teaching assistant communications. When questions arise during the course of this class, please remember to check **three** sources for an answer **before** emailing your instructors:

1. Course Syllabus
2. Announcements in Canvas
3. The Discussion/Q & A Forum in Canvas

This policy will help you in potentially identifying answers before either Jake or I can get back to you, and it also helps us from answering similar questions or concerns multiple times.

Although your classmates may be helpful in answering questions, please be aware that nothing in the Canvas forum constitutes official course policy. In the event of a conflict between information on the Q & A forum and information in the syllabus, always refer to the syllabus unless the information in the syllabus was formally changed by the instructors through an official course

announcement. Please note that your instructors will respond to your Q&A Forum posts and to your emails within 24-48 hours on business days.

All instructor correspondence will be sent to your “@utexas.edu” e-mail account. Forwarded emails to and from your UT to a personal account is not recommended because course related emails are frequently “lost” in cyberspace. **UT faculty will not respond to any e-mail address other than UT account addressed emails.**

NOTE: When emailing your instructors, please include the course number “INF 310” in the subject of your email, so we can prioritize it over other emails that are not course related.

COURSE DESCRIPTION:

Introduction to Human-Centered Data Science provides students with the skills and knowledge needed to explore and analyze `big data' with an understanding of technical, social, and ethical considerations. It also provides students with an opportunity to better appreciate how data and analytics can be used to develop strategies, plans, and policies to improve both business and community operations. Lastly, our class will favor (only slightly) spatial data.

COURSE PREREQUISITE(S):

Introductory class in statistics

LEARNING GOALS:

Upon successful completion of the course, students should be able to:

1. Characterize the process for data acquisition, storage, visualization, and clustering as well as summative/predictive analytics;
2. Identify the societal and ethical challenges and concerns for effectively harnessing 'big' data;
3. Evaluate data science methods for tasks including regression, cluster analysis, and exploratory spatial data analysis;
4. Enhance data analysis skills using a range of software packages, Python and R;
5. Design, implement and communicate about a self-chosen human-centered data science project by proposing a project idea, designing and executing experiments to support the proposed idea, writing a paper about the project and presenting the project to peers at the end of the course.

REQUIRED BOOKS:

Provost, F. and T. Fawcett. (2013). *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. O'Reilly Media. ISBN: 978-1449361327

OTHER REQUIRED RESOURCES:

A functioning computer, up-to-date within the last 5 years. Please note, mobile devices such as tablets or smartphones are not recommended for use with Canvas. The Canvas mobile app can be used for

occasionally checking on the course, but the mobile app will not display the course in its entirety. You must access the course regularly from a full-featured laptop or desktop computer.

You might consider using the following software packages for this course:

- R (v. 3.6.3 or later: <https://www.r-project.org/>)
- **Blue Sky** (<https://www.blueskystatistics.com/>) (**GUI for R**)
- PSPP (<https://www.gnu.org/software/pspp/>) (Open source SPSS)
- Python (v. 3.86 or later: <https://www.python.org/>)
- **GeoDa** (v. 1.16 or later: <http://geodacenter.github.io/>)
- **Presentation software** (PowerPoint, Keynote, or similar)
- **Word processing software** (Word, Pages, or similar)
- **Web browsers** (Chrome, Mozilla Firefox, or Safari)
- **Adobe Acrobat Reader** (free)

LEARNING ACTIVITIES:

1. **Exams (40%):** There will be two exams in this class, each worth 20% of your final grade. Each examination consists of multiple choice and short answer/essay/application questions. The multiple-choice questions are designed to test your knowledge of lecture and text material. These questions are often difficult and require careful reading on the exam. The short answer/essay questions are designed to evaluate your comprehension of methods, their analytical foundations and the implications of their use for people and society. Clarity of expression is a factor in assessing these questions. There are no make-up exams.
2. **Exercises (20%)**
There will be five exercises in this class that will pertain to the techniques and methods covered in our lectures. Exercises may include the submission of statistical output, its interpretation, a visualization or map, a small essay on the preliminary results, or an essay regarding data uncertainty, errors, constraints, or implications for public use. This is an opportunity for you to express yourself. Take advantage of it. Grading will adhere to the rubric posted in Canvas.
3. **Team Mini Project (20%)**
This is ½ course-long project that will require a final report (**60 points**) and presentation (**30 points**) to the class. Do not ask us what we want your team's project to be about. This is an opportunity for you to select a topic that interests you. A one-page proposal of your paper is due on the 6th week of class (**10 points**). The proposal should be written in sentence form (not an outline) and should contain:
 - A working title
 - A brief description of what topic your team plans to investigate
 - A brief description of how your team plans to go about investigating that topic
 - A brief hypothesis of what your team expects to find for the results
 - A bibliography of sources your team has already found on this topic (10 minimum)
 - A comprehensive description of where your team is going to get data for analysis

The actual project report should adhere to the following guidelines:

- Typed, double-spaced, 1-inch margins on all sides, 11 or 12 point Calibri font. You must also have a separate cover sheet that includes names and contact information (UT email).
- The paper must be referenced (APA style); including complete and specific within-paper citations of all books, articles, and Internet sites consulted (if I can't look these references up... don't use them).

4. **Article Summaries (20%):** Five journal articles will be assigned for the class to read during the semester. The topics of these articles will relate to the subject matter of the lecture and/or class for that particular week. Students are required to write a detailed summary of each article (approx. 2-3 pages) discussing the subject matter and statistical or visualization techniques applied (with an emphasis on the former). Be critical!! Grading will adhere to the rubric posted in Canvas.

SUBMITTING ASSIGNMENTS AND LATE ASSIGNMENTS

All assignments are to be submitted by the due date. Late submissions will not be accepted unless:

1. an extension was granted by the instructor in advance of the due date;
2. documentation of an unforeseeable excused absence (as assessed in the instructor's sole discretion) is provided;
3. Your instructors retain the right of refusal on assignments. In other words, if you are turning in late work, we may not accept it.

Additionally:

4. All assignments, unless otherwise announced by the instructor, **MUST** be submitted via Canvas. Each assignment will have a designated place to submit the assignment. Do not submit an assignment via email.
5. Late submission of assignments is subject to penalty (see below). There are no exceptions. Schedule accordingly.

<u>Number of days late</u>	<u>Percentage points taken off</u>
1	20%
2	40%
3	60%
4	80%
>=5	100%

SUMMARY OF GRADED ASSESSMENTS:

Grading Policy: The final grade is based on the proportion of cumulative points you have earned over the semester. There are a total of 500 points.

Section Exams:

Exam 1	100
Exam 2	100
Exercises	100
Mini-Project	100
Article Summaries (n = 5)	<u>100</u>
Total	500

FINAL COURSE GRADING SCALE:

Final Grades will be assigned using the following breakdown:

A	460 - 500
A-	450 - 459
B+	440 - 449
B	410 - 439
B-	400 - 409
C+	390 - 399
C	360 - 389
C-	350 - 359
D+	340 - 349
D	310 - 340
D-	300 - 309
F	299 or below

ACADEMIC INTEGRITY:

1. Academic Dishonesty – In addition to academic performance, students are expected to demonstrate the qualities of honesty and integrity. All submissions by a student are expected to be the original work product of the submitting student. Material that violates this requirement in any way, or that constitutes any form of dishonesty, cheating, fabrication, the facilitation of academic dishonesty, and/or plagiarism, may result in the student receiving a failing grade in the course ('F') and in appropriate disciplinary action being initiated. The possible sanctions include, but are not limited to, grade penalties, permanent record on your transcript of academic dishonesty, loss of registration privileges, disqualification, and dismissal. Please see the official UT Policy on Cheating and Plagiarism (Policy Statement 08-02) for more details about the consequences of academic dishonesty. It is accessible online at:

<https://deanofstudents.utexas.edu/conduct/academicintegrity.php>

2. Electronic Review – To ensure compliance with academic integrity policies, written submissions will be submitted to TurnItIn, or a similar plagiarism detection program. Submission of any writing

assignment in the course to the instructors constitutes consent by the student for the instructors to upload the paper to check against any anti-plagiarism database.

3. Citation Style – Students are required to use the citation style and format of either the American Psychological Association (APA).
4. Avoiding Plagiarism – Some students truly do not understand what plagiarism is, and therefore plagiarize unwittingly or unintentionally. But ignorance is not an excuse for unethical academic conduct. To combat such ignorance, here are rules and resources to help you avoid any problems with plagiarism. Of course, these rules apply regardless of the citation form or style you may be using.
 - a. Direct Quotations – Whenever you directly quote someone else, you must provide a citation to the source of the material from which you are quoting. Moreover, you must put the material in quotation marks or otherwise set it off in an indented quote so the reader knows what words are yours and what words are quoted. It is unacceptable to use the words of others and only partially quote the original source. This is true even if you provide citation to the source both in text and in your references section!
 - b. Paraphrasing/Indirect Quotations – Whenever you indirectly quote someone else (i.e., you paraphrase the work of another), you must provide a citation to the source of the material from which you are paraphrasing. Simply changing the structure of a sentence, or a few words in a sentence so that the sentence you write is not an exact quote from the original source does not mean a citation is not needed. This is because the idea you are expressing is not your own, but rather someone else's.
 - c. Using Other's Ideas – Even if you compose an entire paragraph of writing in your own words (i.e., neither quoted nor paraphrased), if the idea you are expressing in that paragraph is not your own, original idea, you must provide a citation to the source from which you obtained this idea.
 - d. Collaborative Work – If you collaborate on any work with someone else and fail to acknowledge that collaboration, you are guilty of plagiarism. If you have received permission from you professor to collaborate on some assignment, be sure that all the contributor's names appear on the submission.
 - e. Altering or Revising Another's Work – If you alter or revise the work done by someone and submit that work as your own, you have plagiarized. Similarly, if you allow someone else to alter or revise work that you have done and then allow that person to submit it as his or her own work, you are both guilty of plagiarism. Work that is not entirely your own must be credited by citation, both in text and in your references page.
 - f. Altering or Revising Your Own Prior Work – You should also be aware that altering or revising your own work that was prepared for another class or another professor, and not bringing it to the attention of the professor to whom you are submitting the revised work is also academic dishonesty. If, for example, you have two classes that require a term paper, and you can write one paper that meets the requirements of both classes, you may not submit that paper to both professors unless you get permission to do so in advance from both professors. Similarly, if you wrote a paper several semesters ago that can be revised and

submitted in satisfaction of a paper requirement for a course in which you are currently enrolled, doing so is academic dishonesty unless you get the advanced permission of your professor to do so. The reason this is dishonest is that it is not an original work prepared in satisfaction for the requirements on the course you are currently taking.

IMPORTANT COURSE POLICIES

1. Required behavior standards are listed in the [Student Code of Conduct and Student Disciplinary Procedures](#), [Campus IT Policies](#), and outlined by the [Office of Student Rights and Responsibilities](#). Anyone in violation of these policies is subject to sanctions.
2. Students are entitled to receive instruction free from interference by other members of the class. When a student's behavior disrupts the educational process, the instructor may ask the student to leave the classroom.
3. The Office of Student Rights and Responsibilities accepts [incident reports](#) from students, faculty, staff, or other persons who believe that a student or student organization may have violated the Student Code of Conduct.
4. In accordance with [Note Taking Services](#), written permission must be secured from the instructor of the class in order to sell the instructor's oral communication in the form of notes. Notes must have the notetaker's name as well as the instructor's name, the course number, and the date.
5. No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets, may be shared online or with anyone outside of the class unless you have my explicit, written permission. Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. I am well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course.
6. This class is a [Wikipedia Free Zone](#). Under no circumstances are students to use Wikipedia, Infoplease, About.com or other sites that are considered open "information commons" for assigned work or term papers. Failure to observe this rule will result in a 10% penalty, **FOR EACH INSTANCE**, on a given assignment.
7. Please notify the instructors **BEFORE** an assignment is due if an urgent situation arises and the assignment will not be submitted on time. Published assignment due dates (Central Time) are firm. Please follow the appropriate University policies to request an [accommodation for religious practices](#) or to accommodate a missed assignment due to University-sanctioned activities.
8. All assignments, unless otherwise announced, **MUST** be submitted to the designated area of Canvas unless otherwise noted.
9. There is a limited timeline to [drop or add the course](#). Consult with your advisor and notify your instructors to add or drop this course. If you are considering a withdrawal, review the following UT policies: [Withdrawal from Classes](#), [Medical/Compassionate Withdrawal](#), and a [Grade of Incomplete](#).
10. Grade disputes must first be addressed by discussing the situation with the instructors. If the dispute is not resolved with the instructor, the student may appeal to the department chair per the [University Policy for Student Appeal Procedures on Grades](#).

SPECIAL LEARNING NEEDS:

The University of Texas at Austin provides upon request appropriate academic accommodations for qualified students with disabilities. To determine if you qualify, please contact the Dean of Students at 471-6259; 471-4641 TTY. If they certify your needs, we will work with you to make appropriate arrangements.

DUE DATES, ABSENCES, AND EXCUSES:

Planned Excused Absences - It is important to complete your work on time, both so you can stay on track and so you can collaborate and interact with your fellow students. Late assignments will not be accepted. You will receive a zero if you do not submit your assignment on time. We use the Canvas upload time as the official time of the submission, so if you submit your assignment 5 seconds after the deadline, Canvas will mark your assignment late and you will receive a zero for that assignment. Not cool, but we have no control over this.

Please make sure you submit your assignments well in advance; computer problems, network outages, and other “technical” issues will not be considered a valid excuse for submitting an assignment late, so make sure you submit your work early. **Plan ahead.**

If you know that a future event will interfere with your ability to complete course work, you are encouraged to contact the instructors to make arrangements for completing the assignments. Provided that a student informs the instructor of the reason for the absence at least one week in advance of the absence, the instructors will attempt to make arrangements, only when absences are due to any of the following three reasons:

- a. religious reasons;
- b. jury duty or similar governmental obligation (a copy of the summons or other official paperwork must be provided); and
- c. university-sanctioned and/or university-approved activities (e.g., artistic performances, participation in research conferences, intercollegiate athletic activities, student government, required class field trips, etc.).

MAKE-UP WORK:

If you are seriously ill or you are injured in a manner that prevents you from completing your work, you should email and notify the instructor to let him know that you are sick or injured as soon as possible. The same is true for the death, injury, or serious illness of an immediate family member. Please be advised that for missed work based on an emergency situation to be excused, you may be required to submit corroboratory documentation, such as a note from a physician documenting the illness or injury (notes from family members will not suffice), a death certificate or other form of proof of a family member’s death, etc.

STUDENT CONDUCT:

The University and your instructors expect all students to comport themselves professionally. Toward that end, we expect everyone to abide by the following rules of etiquette:

- Although you are welcome to disagree with the professor or your classmates, everyone is expected to participate respectfully. Students who engage in personal attacks; who use rude, insulting, or disrespectful language; or who engage in other disruptive or threatening behavior may face disciplinary action.
- Please use proper grammar, spelling, and punctuation in all course-related communications, whether that is in the discussion boards, emails to the instructor, or assignments. “Text lingo” and other invented spellings are not allowed.
- What To Call Us – We don’t care much for titles, so “Tony” and “Jake” are fine with us. If that freaks you out, “Dr. G” or “Dr. Grubestic”; Dr. J or Dr. Nelson work too.

SEXUAL VIOLENCE AND HARASSMENT:

Title IX is a federal law that provides that no person be excluded on the basis of sex from participation in, be denied benefits of, or be subjected to discrimination under any education program or activity. Both Title IX and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at <https://titleix.utexas.edu/campus-resources>.

TECHNICAL SUPPORT:

For technical assistance 24 hours a day, 7 days a week, please contact the Information Technology Office Help Desk:

Phone: 512-471-4357

Web: <https://education.utexas.edu/about/college-offices/information-technology-office>

If you are in the Ischool you can also email the ischool help desk: help@ischool.utexas.edu

SYLLABUS CHANGES:

We view this syllabus as an educational contract between us and the students in the course. Accordingly, we will make every effort to avoid changing the course schedule, but the possibility exists that unforeseen events will make syllabus changes necessary. We therefore reserve the right to make changes to the syllabus, other than to the attendance and grading policies.

STUDENT SUCCESS

Students are expected to engage with the material in-class and outside of class. Furthermore, those who do not read the assigned material carefully will do poorly. All assigned readings and lecture material are examinable. If you are having problems with the course material or with the assignments, or if you just seem to be getting behind, please see the course instructor as soon as possible.

To summarize:

- Check the course daily
- Read announcements
- Read and respond to course email messages as needed
- Complete assignments by the due dates specified
- Communicate regularly with your instructor and peers
- Create a study and/or assignment schedule to stay on track

COVID GUIDANCE

To help keep everyone at UT and in our community safe, it is critical that students report COVID-19 symptoms and testing, regardless of test results, to [University Health Services](#). Please see this [link](#) to understand what needs to be reported.

CLASS RECORDINGS

Class Recordings: Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

TITLE IX

Beginning January 1, 2020, Texas Senate Bill 212 requires all employees of Texas universities, including faculty, report any information to the Title IX Office regarding sexual harassment, sexual assault, dating violence and stalking that is disclosed to them. Texas law requires that all employees who witness or receive any information of this type (including, but not limited to, writing assignments, class discussions, or one-on-one conversations) must be reported. If you would like to speak with someone who can provide support or remedies without making an official report to the university, please email advocate@austin.utexas.edu. For more information about reporting options and resources, visit <http://www.titleix.utexas.edu/>, contact the Title IX Office via email at titleix@austin.utexas.edu, or call 512-471-0419.

Although graduate teaching and research assistants are not subject to Texas Senate Bill 212, they are still mandatory reporters under Federal Title IX laws and are required to report a wide range of behaviors we refer to as sexual misconduct, including the types of sexual misconduct covered under Texas Senate Bill 212. The Title IX office has developed supportive ways to respond to a survivor and compiled campus resources to support survivors

COURSE SCHEDULE:

Please note: date, topics, and assignments listed are “best estimates”; they may be adjusted from time to time. In addition, although we start slowly, the course picks up steam (significantly) during weeks 4 and 5 and there is no looking back. Manage your time carefully. Work ahead when you can.

WEEK 1 (01/18 – 01/22)

Lecture: (1) Syllabus; (2) Why Should We Care About Data Science (+ Humans)?
Reading: Chapter 1 (Provost & Fawcett)

WEEK 2 (01/25 – 01/29)

Lecture: Data Science Problems and Solutions
Reading: Chapter 2 (Provost & Fawcett)

Article Review 1 (Due 01/28):
Cao, L. (2017). Data science: challenges and directions. *Communications of the ACM*, 60(8), 59-68.

WEEK 3 (02/01 – 02/05)

Lecture: Reproducibility and Accountability + Prepping Data and Models
Reading:

Johnson, Isaac L., et al. (2016). Not at home on the range: Peer production and the urban/rural divide. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.

Exercise 1: (Due 02/04):

WEEK 4 (02/08 – 02/12)

Lecture: Math Boot Camp + Scientific Method + Correlation + Bivariate Regression
Reading: TBD

Article Review 2 (Due 02/11):
Yamashita, T., & Kunkel, S. R. (2010). The association between heart disease mortality and geographic access to hospitals: county level comparisons in Ohio, USA. *Social Science & Medicine*, 70(8), 1211-1218.

WEEK 5 (02/15 – 02/19)

Lectures: Predictive Modeling (Multivariate Regression Analysis)
Reading: Chapter 3 (Provost & Fawcett)

Exercise 2 (Due 02/18)

Article Review 3 (Due 02/18):
Florida, R. (2002). The economic geography of talent. *Annals of the Association of American Geographers*, 92(4), 743-755.

HUGE WEEK 6 (02/22 – 02/26)

Lectures: Logistic Regression + Overfitting
Reading: Chapters 4 and 5 (Provost & Fawcett)

Exercise 3 (Due 02/25)

Article Review 4 (Due 02/25):

Grubestic, T. H. (2003). Inequities in the broadband revolution. *The Annals of Regional Science*, 37(2), 263-289.

Mini Project Proposal (Due 02/25)

WEEK 7 (03/01 – 03/05)

Lecture: Spatial Analysis and Spatial Autocorrelation

Reading:

Fotheringham, A. S., & Brunsdon, C. (1999). Local forms of spatial analysis. *Geographical Analysis*, 31(4), 340-358.

Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37(4), 491-496.

WEEK 8 (03/08 – 03/12)

EXAM 1 (Due 03/11)

Lecture: Exam Review (Tuesday)

Reading: None...

WEEK 9 (03/15 – 03/20) Spring Break!

WEEK 10 (03/22 – 03/26)

Lectures: Spatial Regression

Reading:

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247-267.

Chi, G., & Zhu, J. (2008). Spatial regression models for demographic analysis. *Population Research and Policy Review*, 27(1), 17-42.

Exercise 4 (Due 03/25)

ANOTHER HUGE WEEK 11 (03/29 - 04/02)

Lecture: Cluster Analysis + Fuzzy Clustering

Reading: Chapter 6 (Provost & Fawcett)

Article Review 5 (Due 04/01):

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 41(3), 359-364.

Interim Project Report (Due 04/01)

WEEK 12 (04/05 – 04/09)

Lecture: Advanced Machine Learning + Visualization
Reading: Chapter 7 (Provost & Fawcett)

Exercise 5 (Due 04/08)

Live with Jake – TBD, open session (60 minutes) for project help/guidance (completely optional, on Zoom)

WEEK 13 (04/12 – 04/16)

EXAM 2 (Due 04/15) – Exam Review (Tuesday) and Exam 2 Via Canvas (Thursday)

WEEK 14 (04/19 – 04/23)

Project Presentations on Zoom!! (04/20) + (04/22)

WEEK 15 (04/26 – 04/30)

Flex Week: Final Project Presentation (If needed, 04/27)
Bonus Lecture (Practical Statistics for Urban Informatics, 4/29)

Final Project Report (Due Friday, 04/30)

WEEK 16 (05/03 – 05/07)

***Exam Makeup Week:** Full course recap (Tuesday) and completely optional makeup exam (Thursday).

***Please note:** You are only eligible to take the third exam if you completed both Exam 1 and Exam 2.

Exam 3 is comprehensive (a full semester worth of material). However, it will allow you to replace your lowest score from Exam 1 or Exam 2. There is no penalty for taking this exam. In other words, if you score lower on Exam 3 (when compared to Exam 1 or 2, you keep your highest scores).

FINALS WEEK (05/10 – 05/14)

No Final Exam.