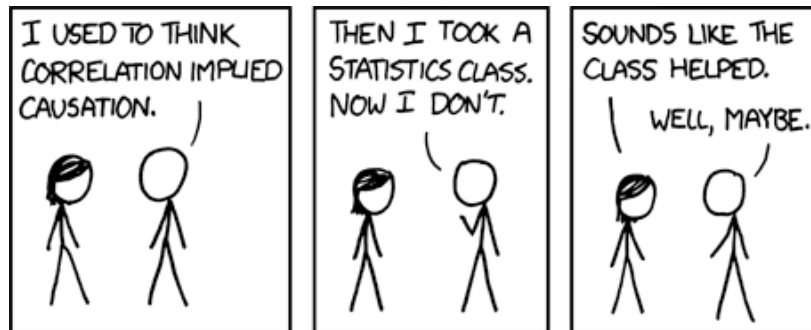


## ANT 388 - Applied Data Analysis

*Practical Statistics, Programming, and Data Science Tools for the Natural and Social Sciences*

### Course #: ANT 388

Tuesdays and Thursdays, 2:00 to 3:30 pm, SAC Room 5.112



**Instructor:** Dr. Anthony Di Fiore  
Department of Anthropology  
SAC Room 4.100H

**Contact:** [anthony.difiore@austin.utexas.edu](mailto:anthony.difiore@austin.utexas.edu)  
512-232-2183

**Office Hours:** Tuesdays and Thursdays, 2:00 to 3:30 and by appointment

**Websites:** Canvas (<https://utexas.instructure.com/>)  
ADA 2019 (<https://difiore.github.io/applied-data-analysis/>)

Check *Canvas* regularly for announcements, updated class information, readings, and supplementary materials

### I. Course Description

This course provides an overview of methods and tools for applied data analysis. It is geared toward research in biological anthropology and evolutionary biology, but the material covered is applicable to a wide range of natural, social science, and humanities disciplines. Students will receive practical, hands-on training in various data science workflows, including digital data acquisition and management, exploratory data analysis and visualization, and statistical analysis and interpretation. Statistical topics to be covered include basic descriptive and inferential statistics, hypothesis testing, regression, and general linear modeling, as well as additional specific methods based on student interest (e.g., geospatial data analysis, phylogenetic comparative methods, social network analysis, corpus construction and text mining, population genetic analysis). The course emphasizes the development of data science skills, focusing on the practical side of data manipulation, analysis, and visualization. Students will learn to use the

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
$\geq 0.1$	

statistical programming language **R** as well as other useful software tools (e.g., shell scripts, text editors, databases, query languages, and version control systems).

## II. Prerequisites

At least one semester of introductory statistics is recommended. Prior programming experience is not expected, but would be helpful.

## III. Required Texts

There are no required texts for this course, but we will be covering good bit of material from several of the following books, all of which are useful resources for learning basic to intermediate level statistics and R programming.

- Davies, T.M. (2016). *The Book of R: A First Course in Programming and Statistics*. San Francisco: No Starch Press.
- Crawley, M.J. (2014). *Statistics: An Introduction Using R* (Second Edition). Chichester, UK: John Wiley & Sons, Inc.
- Kabacoff, R.I. (2015). *R in Action: Data Analysis and Graphics with R* (Second Edition). Shelter Island, NY: Manning Publications Co.
- Wickham, H. and Grolemund, G. (2017). *R for Data Science*. Sebastopol, CA: O'Reilly Media, Inc.

Other useful texts and links are noted on **Resources** page of the course website.

## IV. Readings

Most of our readings for the semester are select chapters from the books listed above or on the **Resources** page of the *Canvas* website. Versions of all REQUIRED readings (apart from those from the primary textbook) will be posted as PDF files to the *Canvas* site so that you can download them for printing and reading.

For those with further interest in a topic, I will also often post additional, OPTIONAL readings for some weeks. Reading assignments and accompanying programming examples and exercises should be completed before class. I strongly recommend that you carefully work through code examples included in the readings as you prepare for class.

## V. Learning Objectives

By the end of this course, you should...

- be familiar with **key concepts and methods** in applied data science for acquiring and managing data, conducting exploratory data analyses, testing statistical hypotheses, building models to classify and make predictions about data, and evaluating model performance;
- have a facility with **modern tools for data analysis**, (e.g., the Unix command line, version control systems, the R programming environment, web APIs) and be able to apply "best practices" in data science;
- know how to **interact with both local and remote data sources** to store, query, process, and analyze data presented a variety of common formats (e.g., delimited text files, structured text files, various database systems);

- be comfortable **writing simple computer programs** for data management, statistical analysis, visualization, and more specialized applications;
- know how to design and implement **reproducible data science workflows** that take a project from data acquisition to analysis to presentation and be able to organize your work using a version control system;
- and be able to apply all of these tools to **questions of interest in the natural and social sciences**.

## VI. Grading and Assessment

Course grades will be based on attendance and participation in class and on several types of assignments as follows:

### Regular attendance and participation in class - 15%

Your attendance and your participation in class constitute a combined 15% of your final grade. Simply put, you **MUST** attend and contribute to this course in order to get a good grade. I expect you come to class with questions and comments about the days readings and to participate in class discussion. Attendance will be taken regularly during throughout the semester, so absences will be noted and penalized.

### Programming problem sets - 5 x 5% = 25%

Homework and programming assignments are due 24 hours before the start of the class period following the one in which they are assigned.

### One individual data analysis replication assignment - 25%

This is an individual assignment in which, in consultation with the instructor, you will choose a published paper and dataset from the primary literature and replicate the analyses presented in the paper. The goals of this assignment are to give you hands-on experience in accessing and analyzing data and in producing a reproducible data analysis workflow.

### One group presentation and written R vignette - 25%

This is a group assignment in which, in consultation with the instructor, you will select a particular statistical method you would like to explore further and prepare a presentation for the class and an R “vignette” exemplifying and applying the method.

### Extra credit

You can also earn up to 10% extra credit for developing an individual interactive data visualization created as a Shiny web app. We will discuss this in more detail later in the course.

Finally, I will occasionally give you "start of class" quizzes to gauge your understanding of material from lecture, class, and the readings, but these do not impact your grade at all.

## VII. Other Logistical Information

For this class, I will use *Canvas* – a Web-based course management system with password-protected access at <https://utexas.instructure.com/> – to distribute course materials, to communicate with you, and to post assignments and grades. I may also ask you to use *Canvas* to submit some assignments. Additionally, I will post information about resources and modules we will cover in the class on the course website, hosted on GitHub: *Applied Data Analysis 2019* (<https://difiore.github.io/applied-data-analysis/>)

## ***Behavioral Expectations***

### Attendance

Inasmuch as students have voluntarily sought admission to the University, I expect you to attend all class meetings, including all lectures and all meetings of associated discussion sessions. Students may be excused for documented medical or personal emergency and will receive reasonable accommodation for the observance of religious holidays (see below). In these cases, they should contact me in advance or, in cases of emergency, as soon as is practicable. Students are responsible for making up any material or assignments they miss.

### Classroom Decorum

Please remember that the classroom is a space for free and open inquiry and for the critical evaluation of ideas, and it should be free of personal prejudice. Every student has the *right* to learn as well as the *responsibility* not to deprive others of that right, and every student is accountable for his or her actions. Students and instructors alike have an obligation to all members of the class to create an educational atmosphere of mutual trust and respect in which differences of opinion can be subjected to deliberate and reasonable examination without animus. During classroom and online discussions and interactions with one another, please treat your fellow classmates with courtesy, civility and respect.

As a matter of courtesy to me and to your fellow students, I expect you arrive at class on time, prepared and ready to participate. Please do not schedule other engagements during this class time.

Students are reminded that cell phone use is forbidden in class. Please shut off all cellular telephones and pagers before class starts and refrain from checking messages and texts during class time. If this becomes an issue, students violating this rule will be asked to leave the class. Students should also refrain from eating in class. You are welcome to bring in water or another non-alcoholic beverage to drink during class.

Except in cases of emergency, I expect you to remain in the classroom for the duration of the lecture or section meeting. If you know that you will need to leave a particular class session early, please let me know beforehand. If it is necessary to leave or enter a room once class has begun, please do so quietly and with as little disruption as possible.

If you have trouble hearing or concentrating on the lecture or media presentation because of distractions around you, quietly ask those responsible for the distraction to stop. If the distraction continues, please let me know. Disruptive classroom behavior may be subject to disciplinary sanction.

Finally, please be aware that no audio or video recording of my lecture is permitted without prior, written approval, and I do not give assignments for extra credit, so please be sure to pay close attention to due dates of assignments.

### Completion of Assignments

Students are expected to submit course work on time and to retain copies of their work until a final grade has been received for the course. Instructors are not obliged to accept late work and may assign a failing or reduced grade to such assignments.

Students who encounter sudden and incapacitating illness or other comparably grave circumstance beyond their control that prevents them from completing an examination or assignment in a course should see me immediately to discuss the situation. Under appropriate circumstances, the student may be granted a temporary mark of *Incomplete*. To receive an *Incomplete*, students must have completed all other requirements for the course, including satisfactory attendance, and there must be a strong likelihood they will pass the course when all work is completed.

## ***Religious Holy Days***

By University of Texas at Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, I will give you an opportunity to complete the missed work within a reasonable time after the absence.

### ***Student Feedback***

During this course I will periodically be asking you to give me feedback on your learning in informal as well as formal ways, including through anonymous surveys about how my teaching strategies are influencing your learning. It is very important for me to know your reaction to the material we are covering in class and on graded materials and assessments, so I encourage you to respond to these surveys, ensuring that together we can create an effective environment for teaching and learning.

### ***Important Policies***

#### Late assignments

Assignments turned in late will be marked down 10% for each day or portion of a day that they are late.

#### Make-up assignments

Make-up assignments will not be permitted, except in the case of medical emergency (for which documentation from your physician's office within 2 days of the due date is required) or for observance of religious holy days (in which case you must let me know two weeks in advance).

#### Review of grades

Please review all grades assignments immediately after they are returned. If you feel an error has been made in how you were graded or if there is anything you do not understand about why you received a particular grade, you must come to office hours or come see me in an individual appointment to bring that to my attention within one week of the assignment being returned. Email is not an acceptable venue for requesting a review of your grade.

## **VIII. Tentative Course Schedule**

Below is a VERY TENTATIVE schedule of weekly topics, which represents my current plans and objectives. As we go through the semester, those plans are LIKELY to change somewhat to enhance the class learning opportunity. Such changes are not unusual and should be expected.

Changes, revisions, and updates to the course schedule will be posted in a timely fashion on the course website and/or noted in class. PDF files of the readings for each class, apart from those from the primary text, will also be posted on the course *Canvas* site, and you should check the course website and *Canvas* sites regularly for announcements and to be prepared for class.

Week	Topic
1	An Introduction to R
2	Data Science Preliminaries
3	Statistics and Programming Fundamentals
4	Inferential Statistics and Hypothesis Testing
5	Reshaping, Graphing, and Plotting for Exploratory Data Analysis
6	Simple Linear Regression and Analysis of Variance

7	Multiple Regression and Generalized Linear Modeling
8	Classification and Categorical Data Analysis
9	Working with Geospatial Data
10	Tree Building and Phylogenetic Data Analysis
11	Scraping Data from the Web, Using APIs, Scraping Data from the Web, and Working with Text
12	Linking R to other Data Analysis Ecosystems (RMDBs, Python)
13	Group Presentations and Vignettes
14	Group Presentations and Vignettes
15	Group Presentations and Vignettes

NOTE: Depending on interest, we may also have one or two extra, optional “workshops” on other topics of possible interest, e.g., connecting R to the Python computing language for data manipulation and analysis.

## IX. Academic Integrity

### *The University of Texas Honor Code*

The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community.

As a student at The University of Texas, you are expected to maintain the highest integrity in your academic work and to adhere to and abide by the University of Texas Honor Code all times. All work you submit in this course for academic credit **must be entirely your own work**. You are not permitted to collaborate with another student on a graded assignment without the express permission of the instructor (e.g., for collaborative assignments).

You are of course encouraged to study with your fellow students and to discuss information and concepts covered in class with one another. You are also able to give "consulting" help to or receive "consulting" help from your classmates. However, this permissible cooperation should never involve one student having possession of a copy of all or part of work done by someone else, in hard copy or digital form. Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero for the assignment.

Penalty for violation of The University of Texas Honor Code and for all of the violations of academic integrity discussed above can also be extended to include failure of the course and University disciplinary action.

## XI. Other University Notices and Policies

### *Use of Email for Official Correspondence to Students*

All students should become familiar with the University's official e-mail student notification policy. It is the student's responsibility to keep the University informed as to changes in his or her e-mail address. Students are expected to check e-mail on a frequent and regular basis in order to stay current with University-related communications, recognizing that certain communications may be time-critical. It is recommended that e-mail be checked daily, but at a minimum, twice per week. The complete text of this policy and instructions for updating your e-mail address are available at <http://www.utexas.edu/its/help/utmail/1564>.

### *Documented Disability Statement*

Any student with a documented disability who requires academic accommodations should contact Services for Students with Disabilities (SSD) at (512) 471-6259 (voice) or 1-866-329-3986 (video phone) or via the internet at <http://www.utexas.edu/diversity/ddce/ssd>. Faculty are not required to provide accommodations without an official accommodation letter from SSD.

- Please notify me as quickly as possible if the material being presented in class is not accessible (e.g., instructional videos need captioning, course packets are not readable for proper alternative text conversion, etc.).
- Please notify me as early in the semester as possible if disability-related accommodations for field trips are required. Advanced notice will permit the arrangement of accommodations on the given day (e.g., transportation, site accessibility, etc.).
- Contact Services for Students with Disabilities at 471-6259 (voice) or 1-866-329-3986 (video phone) or reference SSD's website for more disability-related information:  
[http://www.utexas.edu/diversity/ddce/ssd/for\\_students.php](http://www.utexas.edu/diversity/ddce/ssd/for_students.php).

### *Behavior Concerns Advice Line (BCAL)*

If you are worried about someone (e.g., a fellow student) who is acting differently, you may use the Behavior Concerns Advice Line to discuss by phone your concerns about another individual's behavior. This service is provided through a partnership among the Office of the Dean of Students, the Counseling and Mental Health Center (CMHC), the Employee Assistance Program (EAP), and The University of Texas Police Department (UTPD). Call 512-232-5050 or visit <http://www.utexas.edu/safety/bcal>.

### *Q Drop Policy*

The State of Texas has enacted a law that limits the number of course drops for academic reasons to six (6). As stated in Senate Bill 1231:

"Beginning with the fall 2007 academic term, an institution of higher education may not permit an undergraduate student a total of more than six dropped courses, including any course a transfer student has dropped at another institution of higher education, unless the student shows good cause for dropping more than that number."

### *Emergency Evacuation Policy*

Occupants of buildings on the UT Austin campus are required to evacuate and assemble outside when a fire alarm is activated or an announcement is made. Please be aware of the following policies regarding evacuation:

- Familiarize yourself with all exit doors of the classroom and the building. Remember that the nearest exit door may not be the one you used when you entered the building.
- If you require assistance to evacuate, inform me in writing during the first week of class.
- In the event of an evacuation, follow my instructions or those of class instructors.
- Do not re-enter a building unless you are given instructions by the Austin Fire Department, the UT Austin Police Department, or the Fire Prevention Services office.