\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***YEAR COURSE OFFERED:** 2023

**SEMESTER COURSE OFFERED:** Fall

**DEPARTMENT:** Political Science

**COURSE NUMBER:** 3316

**NAME OF COURSE:** Statistics for Political Science

**NAME OF INSTRUCTOR:**  Tom Hanna

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***This is a preliminary syllabus to give a course overview prior to the first week of class when a full updated syllabus will be issued**. Specifically, assignments and grade weights are likely to change and additional information on software, etc. may be provided either in the syllabus, in class, in Canvas, by email or any combination of these.

**The information contained in this class syllabus is subject to change without notice. Students are expected to be aware of any additional course policies presented by the instructor during the course**.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Welcome to Statistics for Political Science

Today the use of data and statistics is not limited to scientists, engineers and academics. Business, government, non-profits, and the media rely on big data for decision making. Unfortunately, much of the way statistics is used by lay people brings to mind Mark Twain’s famous remark that “There are three kinds of lies: lies, damned lies, and statistics.” As voters, we are faced with competing information from candidates often presenting the same statistics to push competing visions, proving Twain’s point! At the least by the end of this course you should be able to appropriately question whether a statistical argument is valid or not. You will also gain skills useful in political science and many professions.

This course is intended to introduce the use of statistical methods and software for the analysis of quantitative data for advanced undergraduate students with little to no previous background in statistics, mathematics, or programming. You will not need to use calculus or linear algebra. Course examples will be drawn heavily from political science particularly with some examples from the life sciences and other social sciences. But the use of these methods will serve students in many careers, and you will have the opportunity to explore the techniques with data of your own choosing from other fields. This is primarily a political science course, but it is also a core math/reasoning course and appropriate to other disciplines.

Data science is a rapidly growing profession in its own right and the tools of data science are being widely used in political science, other academic disciplines, government, non-profits, and business. As part of this course, you will learn the basics of one of the two most widely used data science programming languages, the R programming language. We will cover the fundamentals in a step by step fashion to better equip you for learning more later in formal settings or on your own. You will also learn to share your results using a public Github repository and develop a simple first project, giving you a widely recognized way to present yourself as a professional user of these tools.

## Specific course objectives include:

* Understanding basic descriptive statistics
* Understanding basic probability theory to prepare for learning more advanced topics
* Applying critical thinking skills to quantitative (numerical) data including the relationship between statistics and causal inference
* Learning the language of causal inference including how to properly present statistical results in the narrative
* Learning the basics of the R statistical programming language, R Studio development environment, and version control with Github
* Developing a Github profile and your first public repositories to demonstrate your statistics and coding proficiency to future employers or graduate schools

# Class Policies:

**My basic expectation is that as a student preparing to enter the workforce, school is your profession while you are here.** Many of you, of course, work at other jobs as well. The point is that this is a professional environment, not recreation. Professional courtesy toward your fellow students and the instructor are a bare minimum to succeed in this class.

* Respect for other people is non-negotiable. This includes other students, faculty, staff, and guests. This also includes people not present, such as human subjects in data sets we use, and even authors whose ideas we may disagree with.
* **Academic honesty is non-negotiable. If you are uncertain whether something is not appropriate, ask me.** **If a GroupMe is created for this class without providing the instructor access, I will assume the purpose is academic dishonesty.**
* Quizzes will be regular and unannounced with a quiz during many class periods. The time of quizzes may vary from beginning to middle to end of class.
* If arriving to class early or leaving late, please take a seat near the door and do so with a minimum of noise.
* If you need to take a break, please keep disruption to a minimum.
* The grade structure includes generous extra credit. Other opportunities including correction of errors on assignments, quizzes, and tests may be offered. Take advantage of this when it is offered.
* To get an A, regardless of extra credit, you must earn at least an 90 points on the regular work listed below.
* Use of electronic devices is required for the class, but must be limited to class material and sound should be turned off.
* Regrading: If you are not satisfied with any grade, you may request an official regrade after a 24-hour cooling off period and within one week following the grade. Regrade requests must be in writing, by email, and provide a reasoned argument why a better grade is justified. Any regrade will be a full re-examination of all aspects of the work and may result in an increase, decrease, or no change in the grade.

## Email:

Emails must come from your @cougarnet.uh.edu email account which you access in Outlook and must be directed to [tlhanna@central.uh.edu](mailto:tlhanna@central.uh.edu) . I will not respond to Canvas messages.

I expect emails to be courteous and respectful of my time. I will be courteous and respectful of your outside time as well, so I expect you to read my emails to you and act on them as needed. When emailing me, please include the course and section number in the email subject, along with the reason for the email. I am not particular about format, but a well written, brief email is easier to respond to quickly. I am a Ph.D. candidate not a full Ph.D., so I am you can refer to me as Tom, Mr. Hanna, or Professor during this course, whatever you are comfortable with. I will respond to emails within 24 hours barring emergencies during normal business days and sooner if possible. **You should bear in mind that a degree of professionalism in your emails is an important career skill you should start learning now and that professionalism is a graded component of this course.** Again, as long as you are courteous and respectful of my time, I am not concerned about precise format, but here is a great reference you should consider reading before emailing any professor: https://marktomforde.com/academic/undergraduates/Email-Etiquette.html

# Major Assignments/Exams

The grade for this course will be based on professionalism and participation, quizzes, problem sets, mid-term and final exams, and a final project as follows (Note: the total points exceed 100 while the course grade is based on a point scale of 0-100. In other words, receiving full credit in any one category can count as “extra credit” to make up for other categories.):

## Professionalism, participation, and syllabus quiz 0-10:

Passing the Syllabus Quiz in Canvas is required to continue with the course. Failing to take the Syllabus Quiz by August 30 may result in being dropped from the course. The minimum expectations are that you follow course policies, attend regularly, and participate in the course. Up to two absences, late arrivals, and early departures (total) may be exempted from professionalism grade without excuse. A portion of each class period will be devoted to answering questions and resolving questions from the previous problem sets and readings. A portion of each lab period will be spent covering errors from the previous session. The issue you ask about today may be encountered by someone else next week, so asking appropriate questions is encouraged as a major portion of participation. Because some errors take research to resolve, your *participation* grade is not limited to in class participation, but includes participation in the class Microsoft Teams channel, sending questions to the instructor ahead of class time, and participation in virtual or in person office hours. That said, participation through Office hours, Teams, and email contact is a bonus that works with attendance, not a substitute for it.

## Quizzes 0-10:

Quizzes will cover prior lecture material and topics of discussion from previous problem sets. Quizzes will be very short answers or calculations during class time, or R exercises during lab time. They will typically last 5 minutes or less. Up to two quizzes may be exempted without excuse. Answers will be posted, so make-ups will not be allowed.

## Problem/reading sets 0-40:

There will be a total of 8 problem/reading sets each worth 5 points. These assignments will include a combination of written responses, statistical calculations, and exercises in R. For statistical calculations and R programming exercises, it is expected that students will run into errors. This is a critical part of learning advanced skills. Written responses detailing the reasoning used, attempts to correct errors, and showing the work done will count for a substantial portion of the grade, up to full credit for high quality work and writing. You may work together on problem/reading sets but must do so through the official Teams channel or provide notice to the instructor of who you are working with and how. **If you use a GroupMe for work in this course, the instructor must be given access or academic dishonesty will be assumed.** Late work will be accepted with 50% deducted unless a valid reason for an extension is provided at least 24-hours before the assignment is due. No late assignments will be accepted after answers are posted.

## Midterm exam/Final Exam 20 points each (40 total):

Both exams will be 90 minutes. The Final Exam will be cumulative. Students will be allowed to bring a one-page, single sided handwritten reference sheet containing any material they find useful. The exams will be a combination of short answers, statistical calculations, and interpretation of output from R. Calculations should be simple enough to finish by hand, but handheld calculators will be allowed. No other electronic devices will be allowed – phones may not be used as calculators. There will be a 30-minute exam review session at the end of the previous class session, with time allowed after class for those with additional questions. The Midterm is tentatively scheduled for Tuesday, July 26 at 2:00 PM. The Final exam is tentatively scheduled for Wednesday, August 10, at 2 PM. Makeup exams will be given only for documented emergencies.

## Project 20 points:

At the beginning of the course, you will learn to setup a Github account, install R and R Studio, and integrate them with Github. As the course progresses, you will complete a short statistical analysis on a subject of your choice applying the methods from the course. You will compile the results and write a short paper explaining the results of the analysis in narrative form and including nicely formatted tables of results. This paper should be 1 to 2 pages of text with some tables and graphs. Though the final product will be due at the end of the semester, it will actually consist of compiling your lab work in your chosen data set from throughout the semester. You will share this project and your corrected problem sets through a Github repository which you may choose to make public to showcase your work to future employers.

# Grading Policy

**Earning the following will guarantee the associated grade. Depending on class performance overall, I reserve the right to adjust the scale in favor of the students:**

• A: 94.000 Points and above

• A- : 90.000 to 93.999 Points

• B+ : 87.000 to 89.999 Points

• B: 84.000 to 86.999 Points

• B -: 80.000 to 83.999 Points

• C+: 77.000 to 79.999 Points

• C: 74.000 to 76.999 Points

• C-: 70.000 to 73.999 Points

• D+: 67.000 to 69.999 Points

• D: 64.000 to 66.999 Points

• D-: 60.000 to 63.999 Points

• 59.999 and below Points: F

# In class labs and required Software

This course will use the R statistical programming language, the R Studio development environment for R, and Github version control. All of these are available for free. The second class period will be dedicated to the basics of Github, R, and R Studio. **Those who have created a Github account, and pre-installed these programs on their own laptop configured to work with the Github account will get more out of the first lab session including an opportunity for extra credit. For those who want to forego extra credit opportunities, I will also go over how to use Posit Cloud, an online version of R Studio to complete assignments.**

A portion of each class period will be dedicated to learning the R statistical programming language. We will cover the topic in small bite sized pieces, allowing time between classes to deal with software errors, questions, and concerns. At the end of the semester, those who invest the time will have the beginnings of their first public Github repository to demonstrate these skills (extra credit opportunities).

Errors in code are one of the most frustrating parts of learning R or any other programming language. Learning to deal with them is also one of the most important parts of the process. **Your grade will not depend on error free R output if you use your own device, as I would rather work with you to resolve errors than have you not turn in work and learn nothing. Thos using Posit Cloud will have less leeway for errors.** You will be expected to be able to interpret R results on both the mid-term and final exams.

**Installing R, R Studio, and Git: The main point to remember is that installing things in order will reduce the potential for errors. Following is an excellent resource for installing R and R Studio, setting up a Github account, and making them all work together. If you encounter problems, don’t hesitate to contact me by email or Teams so we can devote class time to substance:**

<https://happygitwithr.com/index.html> (Ch. 4, 5, 6, 7 and 12 are most important.)

**Another option I highly recommend** is the following Coursera course. It can easily be completed during their two week free trial and includes a certificate which can be linked to your LinkedIn profile. (I completed it in a few hours. Those with less computer experience may take longer, but may benefit even more.):

<https://www.coursera.org/learn/data-scientists-tools?specialization=jhu-data-science>

# Microsoft Teams & Github

This class has a Microsoft Teams group and Github repository. Participation in the Teams discussion channels is encouraged and will add to your class participation grade. All required readings, lecture slides, written materials, and additional recorded video examples in R will be posted to Teams. Lab materials and R exercises for problem sets will be posted to Github for easy integration to R.

# Required Reading

We will use R exercises either directly from or derived from:

*An Introduction to Statistical Learning with Applications in* *R*

Gareth James • Daniela Witten • Trevor Hastie • Robert Tibshirani

Second Edition, Corrected Printing June 2023

[This book is available free online from the authors.](https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023.pdf) You do not need to read the entire very advanced book, but should have it as a reference.

Other required readings will be provided through Teams and/or Github.

# Recommended Reading

While all topics will be covered in lectures, lecture slides, and provided readings, some of the works I am using in preparing this course and which you may find useful are:

* *Introductory Statistics: A Conceptual Approach Using R* by Miller, Ferron, and Ware. ISBN-13: 9780415996006 DIGITAL ISBN-13: 9781136870101 ISBN-10: 0415996007
* *Even You Can Learn Statistics & Analytics* by Levine and Stephan **ISBN-13:** 9780133382662 **DIGITAL ISBN-13:** 9780133382679 **ISBN-10:** 133382664
* *The Cartoon Guide to Statistics* by Larry Gonick and Wolcott Smith

I**SBN-10 ‏ : ‎**0062731025 **ISBN-13 ‏ : ‎**978-0062731029

* Statistics All-in-One For Dummies 1st Edition, Kindle Edition

ISBN-13: 978-1119902560 (Used hardcover and paperback editions are available but cost more than the Kindle edition and may not include the online learning features)

The first two books are available through the university bookstore in print and digital editions. The others are available through the Kindle store on Amazon, or available used at lower prices.

# List of discussion/lecture topics

Generally, each class will be begin with a lecture, discussion, and practice with the assigned material followed by a 30-minute lab period using R. The first several session will be:

August 21: Course introduction: Policies, Procedures, and Overview

August 23: Introduction to R and R Studio (Lab – bring laptop with R and R Studio installed with Github (for full extra credit) or Posit Cloud account ready for basic credit.)

August 28: Introductory Concepts in Probability and Statistics

August 30: No class – workday for syllabus quiz and homework

Important dates:

September 4: Labor Day, no class

September 6: Last to drop without receiving a grade

October 11: Midterm (tentative)

November 15: Last day to drop with a W

November 20: We will have class

November 22-25: Thanksgiving Holiday, no class

November 27: We will have class

December 2: Last day of class

December 5-13: Finals week. Final exam per University schedule

Course outline:

* Descriptive and summary statistics
* Choosing your project data set
* Presenting descriptive statistics graphically and in tables with R
  + With assigned data
  + With your project data
* Probability
* Probabilities in R
* Frequency distributions and Z-scores
* Frequency distributions in R
* Conducting simulations in R
* Correlation
* Plotting correlations in R
  + With assigned data
  + With your project data
* Statistics and causal inference
* Hypothesis testing and confidence intervals
  + Z-tests, T-tests, Chi sqare, ANOVA
* Hypothesis testing in R – interpretation
* Conducting hypothesis testing in R
  + With assigned data
  + With your project data
* Regression (Ordinary Least Squares)
* Modeling simple linear regression in R
  + With assigned data
  + With your project data
* Presenting linear regression in R
  + With assigned data
  + With your project data
* Multiple Regression – Control variables and multiple explanatory variables
* Regression with multiple variables in R

# Academic Integrity

Cheating and plagiarism will not be tolerated and will be penalized. Each student in this course is expected to abide by the University of Houston’s policies against cheating and plagiarism. The University’s statement on academic honesty is available here: <https://www.uh.edu/provost/policies-resources/honesty/>

You are encouraged to work, study, and discuss information and concepts covered in class and readings with other students. Cooperation between students should never result in one student possessing a copy of work done by someone else, either in electronic form or hard copy. Cooperation in studying for tests ends when the test begins. Should copying occur, both the student who copied the work and the student who provided the copied material will automatically receive no points for that test. Repeated and/or egregious violations will be dealt with more severely, and may include failing the course and facing further disciplinary action.

# Honor Code Statement

Students may be asked to sign an honor code statement as part of their submission of any graded work including but not limited to projects, quizzes, and exams:

“I understand and agree to abide by the provisions in the University of Houston Academic Honesty Policy. I understand that academic honesty is taken very seriously and, in the cases of violations, penalties may include suspension or expulsion from the University of Houston."

# Accommodations for Students with Disabilities

The Americans with Disabilities Act is a federal law providing comprehensive protections, including reasonable academic accommodations, for persons with disabilities. The University of Houston is committed to providing reasonable accommodations for students who have learning disabilities, psychiatric disabilities, and health impairments. Any student in this course who has a disability that may prevent him or her from fully demonstrating his or her abilities should inform me personally. If you believe you have a disability which is eligible and requires accommodation, please contact the Center for Students with Disabilities (CSD). For detailed information, including documentation requirements, listings of available academic support services, and test administration policies, please visit <http://www.uh.edu/csd/>

# Statement on Counseling and Psychological Services

Counseling and Psychological Services (CAPS) can help students who are having difficulties managing stress, adjusting to college, or feeling sad and hopeless. To reach CAPS (www.uh.edu/caps), call 713-743-5454 during and after business hours for routine appointments or if you or someone you know is in crisis.

No appointment is necessary for the “Let's Talk” program, a drop-in consultation service at convenient locations and hours around campus. <http://www.uh.edu/caps/outreach/lets_talk.html>

# Excused Absence Policy

Regular class attendance, participation, and engagement in coursework are important contributors to student success. Absences may be excused as provided in the University of Houston Undergraduate Excused Absence Policy (<https://uh.edu/provost/policies-resources/student/excused-absence-policy/>) for reasons including: medical illness of student or close relative, death of a close family member, legal or government proceeding that a student is obligated to attend, recognized professional and educational activities where the student is presenting, and University-sponsored activity or athletic competition.

Additional policies address absences related to military service, religious holy days, pregnancy and related conditions, and disability.

# Recording of Class

Students may not record all or part of class, livestream all or part of class, or make/distribute screen captures, without advance consent of the instructor. If you have or think you may have a disability such that you need to record class-related activities, please contact the Center for Students with DisABILITIES.

If you have an accommodation to record class-related activities, those recordings may not be shared with any other student, whether in this course or not, with any other person, or on any other platform. Classes may be recorded by the instructor. Students may use instructor’s recordings for their own studying and notetaking. Instructor’s recordings are not authorized to be shared with anyone without the prior approval of the instructor. Failure to comply with requirements regarding recordings will result in a disciplinary referral to the Dean of Students Office and may result in disciplinary action.

# Syllabus Changes

Please note that the instructor may need to make modifications to the course syllabus to ensure the health and safety of the University community and may do so at any time. Notice of such changes will be announced as quickly as possible on Teams.