

INTRODUCTION TO DATA SCIENCE

DATA 1301
Unique Number: 87614
Fall 2021

Instructor: Amir Shahmoradi
office: SEIR 365
e-mail: a.shahmoradi@uta.edu
e-mail: shahmoradi@utexas.edu
e-mail: a.shahmoradi@gmail.com
Office/Lab/Help hours: **dataCAVE** (computer lab in the UTA Main Library)
Online by appointment

Class start/end: Aug 25, 2021 – Dec 6, 2021
Lecture meeting times: MWF 9:00AM – 9:50PM
Lecture meeting place: LS 101 (and if needed, Teams Virtual Room)

Teaching Assistants:	TBA
office:	TBA
e-mail:	TBA
office hours:	TBA

Course Purpose Statement

This course satisfies the University of Texas at Arlington Core Curriculum requirement in Mathematics. This course offers a survey of contemporary approaches to data-driven discovery and is designed and accessible to every student interested in the modern data revolution. This course focuses on qualitative statistical literacy in logic, data collection, visualization, and relationships, as well as the fundamentals of scientific inference. The course also includes discussions of the ethical principles involving privacy, data security, and broader societal implications. In particular, various human and cognitive biases can affect scientific conclusions and discoveries. The fundamentals of data visualization will also be discussed. Simple statistical concepts and techniques will be addressed via simple computing tools such as Excel or programming languages like Python. An introduction to computer programming history will also be given.

Textbook and Materials

No textbook is required for this course. Online and class lecture notes will be used as reference. However, a list of textbooks for those who are interested to self-educate themselves or go beyond class syllabus is provided below,

- *Calling Bullshit: The Art of Skepticism in a Data-Driven World*, by Carl Bergstrom and Jevin West.
[Available here.](#)
- Computational and Inferential Thinking,
<https://www.inferentialthinking.com/chapters/intro#the-foundations-of-data-science>

Some activities of the course will require students to have access to computing device. UTA Library offers free access to computers and laptops for students of this course.

Additional Materials for This Course

None.

Attendance Policy

Attendance is not mandatory. However, all absences should be coordinated with the instructor.

Grading Scale

Grades will be computed based on the following distribution. Grades are rounded up accordingly.

Percentage	Letter
90 — 100%	A
80 — 89%	B
70 — 79%	C
60 — 69%	D
Below 60%	F

Grade Calculation

Assignments and Course Requirements	Percent of Grade
Attendance/Participation	0%
Biweekly Homework	33%
Biweekly Quizzes	33%
Final Exam	34%

- No homework or quiz grade will be dropped unless it is justified at the discretion of the instructor.

COURSE SCHEDULE

The following is a tentative outline of topics to be covered:

Intro to Data Science/Course (3 lectures)

- Lecture 1:
 - Homework 1: Data collection and visualization
 - The three pillars of Scientific Inference (0.5 lecture)
 - A brief history of Data Science and the digital world revolution
- Lecture 2: A brief history of Computers, Scientific Programming

Intro to Data/Visualization (6 lectures)

- Lecture 3: A basic workflow for scientific inference
 - Data Collection
 - Types of Data
 - sequential
 - time-series
 - spatial

- ...
 - Data Reduction
- Lecture 4:
 - Homework 2: Data summarization, visualization
 - Data Exploration
 - Data Analysis
 - Computing the mean, standard deviation, ...
 - What is data variability, uncertainty and statistical confidence?
 - Thinking in more than three dimensions (Are there four-dimensional super creatures?)
 - What is correlation and its connection to causation?
- Lecture 5:
 - Data Visualization
 - Different types of visualization, plots, and graphs
 - How to not put your audience into sleep instantly?
- Lecture 6: Hypothesis Formation and Model Construction
 - Homework 3-4: Logic

Logic, Probability, ... (6 lectures) digression

- (The Mathematical and statistical foundations of Data Science)
 - Boolean Logic
 - Bayesian Logic
 - Set Theory
 - A Brief History of Probability
 - What is a distribution?
 - How Cognitive Biases Affect Scientific Discoveries?

“Models” (3 Lectures)

- Homework 5: models
- Model Calibration
 - What is a vector?
 - Likelihood and the Bayes Theorem
- Model Validation
 - How can we test the accuracy of hypotheses about data?
- Model Selection
- Prediction
 - Uncertainty Quantification

“Domain” (6 lectures)

- Big Data
 - Midterm exam.
 - Homework 6-7: Domain specific problems repeating previous concepts
 - How do Facebook and Twitter handle the massive amounts of information exchange on their websites?
- More on domain science

Machine Learning (6 lectures)

- Homework 8: some simple applications
- Demystifying Machine learning and Deep Learning
 - Supervised Learning
 - Unsupervised Learning
- Web-based DL model training (e.g. TensorFlow playground)
- NVIDIA Digits (interface to Caffe2), teachable machine

- Homework 9: semester project

Principles of success in Data Science (1 Lecture)

- Workflow and reproducibility
 - Data Science is all about organization and discipline
- How to choose between Data Science projects and career options?
 - What is the connection between Data Science and other natural sciences and engineering fields?

The Ethics of Data Science (3 Lectures)

- Human Cognitive biases
- Instrumental and data collection biases
- Methodological biases
- Work pressure and misconduct
- Data privacy and confidentiality protection
- Refuting nonsense and fake news
- ...

COURSE LOGISTICS

Homework Policy:

There will be approximately one homework biweekly. Assignments will be due every Monday before the lecture begins and should be added to an online repository determined by the instructor. No late assignments will be accepted. No exceptions to the homework policy will be made without prior instructor approval.

Examinations:

There will be no midterm exams but one final exam.

Quizzes:

There will be biweekly quizzes.

Attendance:

Regular attendance is expected. Any absence requires prior approval from the instructor, or compelling evidence of illness or an official letter from the university administration. Student attendance will be randomly checked.

Scholastic dishonesty: All students are responsible for upholding the University rules on scholastic dishonesty. Students who violate University rules on scholastic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on scholastic dishonesty will be strictly enforced.

Other matters: The University of Texas at Arlington provides, upon request, appropriate academic adjustments for qualified students with disabilities. Any student with a documented disability (physical or cognitive) who requires academic accommodations should contact the UTA's Office for Students with Disabilities as soon as possible to request an official letter outlining authorized accommodations. For visit <https://www.uta.edu/disability/>.