

INTRODUCTION TO DATA SCIENCE

DATA 1301

Unique Number: 88541

Fall 2020

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/in-person by appointment

Class start/end: Aug 26, 2020 – Dec 8, 2020
Lecture meeting times: Mo – We 4:00PM – 5:20PM
Lecture meeting place: SEIR 294, Teams Virtual Room

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

COVID-19 Notes:

Given the current global pandemic situation, in-person meetings will be limited. All lectures, HW, and quizzes can be also done and submitted online with no contact with the instructor. Students will be allowed to attend the classes and submit their HW/Quizzes completely online if they are concerned about their safety and health or the health of their family members. In such case, please inform the instructor of your intention to attend the classes online.

COURSE OBJECTIVES / ACADEMIC LEARNING GOALS

This is the first course in the series of DATA SCIENCE courses offered by the Data Science program of the College of Science at the University of Texas Arlington.

This course offers a survey of contemporary approaches to data-driven discovery and will be the required entry point for students majoring in Data Science and will be designed to be accessible to all others interested in the modern data revolution. The course will include discussions of ethical principles involving privacy, data security, and broader societal implications. In particular, various human and cognitive biases that can affect scientific conclusions and discoveries. The fundamental principles of visualization will be also discussed. Simple statistical concepts and techniques will be discussed via such simple computing tools as Excel. An introduction to computer programming will be also given.

COURSE SCHEDULE

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

- The three pillars of Scientific Inference
- A brief history of Computers, Scientific Programming
- A brief history of Data Science and the digital world revolution
- The Mathematical and statistical foundations of Data Science
 - Set theory and basic logic
 - Random numbers, probability
 - What is a distribution?
 - Likelihood and the Bayes Theorem
 - What is a vector?
 - ...
- Types of data
 - sequential
 - time-series
 - spatial
 - ...
- 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?
 - How can we test the accuracy of hypotheses about data?
- Data Visualization
 - Different types of visualization, plots, and graphs
 - How to not put your audience into sleep instantly?
- Big Data
 - How do Facebook and Twitter handle the massive amounts of information exchange on their websites?
- Demystifying Machine learning and Deep Learning and their connections with each other and with Data Science
- Principles of success in Data Science
 - 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
 - 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 TEXTBOOKS

No textbook is required for this course. Online 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>

COURSE LOGISTICS

Grading:

Homework: 33% (Assignments might not be weighted equally)

Quizzes: 33%

Final Exam: 34%

Homework Policy:

There will be approximately one homework per week or 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 weekly or 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/>.