INTRODUCTION TO DATA SCIENCE
DATA 1301
Unique ID: 27294
Spring 2024

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:	January 17, 2024 – April 29, 2024
Lecture meeting times:	MW 4:00 PM – 5:20 PM
Lecture meeting place:	PKH 103 (and, if needed, Teams Virtual Room)

Teaching Assistants:	Asirvadam Nicki
office:	Teams or on campus by prior schedule
e-mail:	nicki.asirvadam@mavs.uta.edu
office hours:	TBA

COVID-19 NOTES:

All classes will be held in person (or other modalities if needed). Please avoid attending lectures in person if you do not feel well for any reason. The instructor will ensure students can access all materials, homework, and quizzes if they cannot attend the lectures in person due to sickness or other reasons. PLEASE TAKE YOUR HEALTH AND OTHERS IN THIS CLASS SERIOUSLY BY PAYING ATTENTION TO COVID19 PROTOCOLS. FAILURE TO DO SO COULD COST YOUR FRIENDS, CLASSMATES, OR YOUR INSTRUCTOR'S LIVES. Wearing a mask and vaccinating is not mandated but strongly encouraged. Your instructor is fully vaccinated and will wear (K)N95 and/or double masks throughout the class sessions.

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 critical entry point for students majoring in Data Science. It will be designed to be accessible to everyone interested in the modern data revolution. The course includes discussing the ethical principles of privacy, data security, and broader societal implications. 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 will also be given.

COURSE SCHEDULE

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

Intro to Data Science/Course (3 lectures)
- o Lecture 1:
    - ▪ Homework 1: Data collection (Google), and visualization
    - ▪ The three pillars of Scientific Inference (0.5 lecture)
    - ▪ A brief history of Data Science and the digital world revolution
- o Lecture 2: A Brief History of Computers, Scientific Programming

Intro to Data/Visualization (6 lectures)
- o 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 TEXTBOOKS

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 in self-educate themselves or going 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:
Quizzes: 35%
Homework: 30% (Assignments might not be weighted equally)
Final Exam: 20%
Presentation: 10%
Attendance: 5%

Homework Policy:
There will be approximately one homework assignment 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:

Attendance is an essential part of this course. Students are expected to attend all classes and arrive on time. To ensure that students keep up with the course material, each student will be allowed to have up to **three absences** during the semester without penalty. If you have difficulty attending the lectures, please talk to the course instructor at the beginning of the semester. If a student misses a lecture due to an illness or emergency, they should notify the instructor immediately and provide appropriate documentation. Excused absences will not count against the five allowed absences. Please note that attendance is not only for the purpose of receiving credit but also to gain knowledge and engage in the learning process. Being present in class ensures that students have access to important information, discussions, and activities that contribute to their overall understanding of the course material.

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 accommodation 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/.