INTRODUCTION TO DATA SCIENCE

DATA 110 Spring 2025
The University of North Carolina at Chapel Hill

COURSE INFORMATION

Credit Hours: 3
Pre or Co-Requisites: None
Target Audience: All Undergraduates
Meeting Pattern: MW lectures + F recitations. Lectures are led by instructors and recitations are led by TAs and involve more small group activities.
Instructional Format: In-Person
Classroom or Location:
DATA110.003: MW 1:25pm-2:15pm Chapman 0201 (Lee)

INSTRUCTOR INFORMATION

Name: Harlin Lee
Email Address: [link removed]
Office Location: on Canvas
Office Hours: on Canvas
Zoom Room ID: N/A
Teaching Philosophy: To create an inclusive learning community that uses a hands-on application of concepts that develop understanding of content, capacity for teamwork and communication & critical thinking skills.

COURSE CONTENT

Course Description
The Introduction to Data Science course is a broad, high-level survey of the major aspects of data science including ethics, best practices in communication (e.g. data visualization), mathematical/statistical concepts, and computational thinking. Students will gain an understanding of the fundamentals of data science to support more in-depth, advanced coursework that are requirements for the BA and BS in Data Science. The curriculum and format are designed specifically for students who are considering a major in data science and may not have taken statistics or computer science courses. The course is a requirement of the BA and BS in Data Science.

Course Texts & Materials
No required textbooks but below are supplementary materials that may be helpful and can be freely accessed online. We will add more to reading list as we go through the semester.

1. Computational and Inferential Thinking: The Foundations of Data Science by Ani Adhikari, John DeNero, David Wagner. https://inferentialthinking.com
2. Official Python 3 tutorial. https://docs.python.org/3/tutorial/index.html
3. Python Data Science Handbook by Jake VanderPlas. https://jakevdp.github.io/PythonDataScienceHandbook/
4. Python for Data Analysis by Wes Mckinney. https://wesmckinney.com/book/

Class Expectations

Regular attendance and active class participation. Lectures will *not* be streamed live or recorded unless there is a formal request from the Equal Opportunity and Compliance office (EOC). Any materials shared during class, e.g. lecture slides, worksheets, handouts, videos, readings, will be available on the course website after class.

Course Goals & Student Learning Outcomes (SLOs)

The goal of this course is to lay the foundation for subsequent courses required for the BS and BA in Data Science, as well as to introduce core concepts and ideas in the field of data science to any student, regardless of major. The course provides a high-level survey of current and emerging concepts in key data science domains, including computational thinking, mathematics/statistical skills, data management, communication best practices, and ethics. The course will offer hands-on analysis of real-world datasets, exposing students to the type of insights and problem-solving that the field of data science can deliver.

Accessibility and Equity: Students from all backgrounds should be able to take Intro to Data Science. As such, no prerequisites in statistics or programming are required for the course; only basic high-school algebra are necessary.

Diversity: Intro to Data Science can be taken by students from any major across campus and should be acceptable as a potential pre-requisite for statistics, math, or computing many majors.

Pedagogical Clarity: Intro to Data Science is designed to first teach introductory programming, then statistics through a computation lens, and ultimately concludes with basic methods in inference.

Core concepts and Learning Outcomes

- Data Management: Describe differences in types of data and the ways in which individuals and organizations store, manage, and interact with data. Identify and appropriately acknowledge sources of data. Apply basic data cleaning techniques to prepare data for analysis.
- Mathematical and Statistical Foundations: Select and use appropriate data analytics and statistical techniques to discover new relationships, deliver insights into research problems or organizational processes, and support decision-making. Draw accurate and useful conclusions from data analysis.
- Computational Thinking: Build and understand algorithms for analyzing large data sets and accurate numerical modeling for problems.
- Communication: Convey data analyses through written and oral communication skills as well as select the appropriate tools to visually display data.

- Responsible Data Science: Identify security, privacy protection, governance, and ethical considerations in data management. Differentiate between ethical and unethical uses of data science.

This course is designed to meet the general education requirement of Quantitative Reasoning. Below are the corresponding learning outcomes and student questions from UNC Chapel Hill's IDEAs in Action General Education Curriculum.

IDEAs in Action General Education Curriculum
FC-QUANT
Student Learning Outcomes:
1. Summarize, interpret, and present quantitative data in mathematical forms, such as graphs, diagrams, tables, or mathematical text.
2. Develop or compute representations of data using mathematical forms or equations as models and use statistical methods to assess their validity.
3. Make and evaluate important assumptions in the estimation, modeling, and analysis of data, and recognize the limitations of the results.
4. Apply mathematical concepts, data, procedures, and solutions to make judgments and draw conclusions.
5. Synthesize and present quantitative data to others to explain findings or to provide quantitative evidence in support of a position.

Questions for Students:
1. What is the role of mathematics in organizing and interpreting measurements of the world?
2. How can mathematical models and quantitative analysis be used to summarize or synthesize data into knowledge and predictions?
3. What methodology can we apply to validate or reject mathematical models or to express our degree of confidence in them?

COURSE ASSIGNMENTS & ASSESSMENTS
Assignment Descriptions

Weekly Lab Exercises [10%]
Lab assignments are small group exercises that are intended to be completed during Friday recitations. You should submit individually at the end of 50 minutes. At the latest, it must be submitted by 11:59pm of the same day. Labs will be graded for participation and *reasonable* progress, and attendance is required. One lowest score out of the 11 labs will be dropped automatically. Excused absences policy and request form: [link removed]. Do not request extension or excused absences via email.

Homework [30%]
You may discuss problems with other students/course staff, but complete and submit independently. See Late Work policy below.

HW1 (5%): Introduction to Data Science
HW2 (5%): Introduction to Programming and Pandas
HW3 (5%): Data Visualization and Communication
HW4 (5%): Probability, Statistics, and Sampling
HW5 (5%): Inference & Hypothesis Testing
HW6 (5%): Machine Learning

Exams [35%]
Exam 1 (15%) will cover week 1 through week 7 (summary statistics and boxplot) material.
February 24 (Monday) during the usual lecture time in the usual lecture room.

Exam 2 (20%) will cover week 7 (probability and uniform distribution) through week 14 material.
April 16 (Wednesday) during the usual lecture time in the usual lecture room.

Group Project [25%]
Poster session presentation + peer review during final exam block:
May 05, Monday 4pm-7pm for Mon/Wed/Fri 1:25 p.m. class

Group project will be assessed based on:
1. Group poster presentation
2. Project proposal
3. Group write-up:
    a. problem statement
    b. data visualization
    c. data analysis
4. Group codebase
5. Teamwork assessment
6. Feedback to other teams

Extra Credit [up to 1%]
Based on in-class and online participation

Grading Scale & Schema
Late Work
Due dates for every Homework assignment are provided on the course syllabus and course
schedule. Unless otherwise stated, assignments are due on those days at 11:59pm. Submit them
to Gradescope, which can be accessed via Canvas.
You have 3 late days that you can use throughout the semester. They must be requested at least 24
hours in advance by filling out the form. Do not request extension via email.

Regrade Request
For all labs, homeworks and exams, regrade requests will be open on Gradescope for one week
after the grade is released to the students. After that period, no requests for regrade will be
accepted via Gradescope, email, or in person.

Grading Scale

In general, this class will *not* grade on a curve unless in extreme situations. Even in those rare cases, the curve will never lower a student's grade. For example, if you have received an 88%, you are guaranteed a B+. Please see a full explanation of the grading system at the [Registrar's website](#).

Numeric Grade (%)	Letter Grade
94 and above	A
90 – 93.9	A-
87 – 89.9	B+
83 – 86.9	B
80 – 82.9	B-
77 – 79.9	C+
73 – 76.9	C
70 – 72.9	C-
67 – 69.9	D+
60 – 66.9	D
below 60	F

Table a: Grading Scale Table

COURSE SCHEDULE

Class schedule may change over the course of the semester.

[Introduction to Data Science]

Week	Date	Lec	Topic(s)	Release	Due
1	Jan 8 (W)	1	Syllabus, What is data science?		
	Jan 10 (F)		*Lab 1*		
2	Jan 13 (M)	2	Data modality, Tables	HW1	

Table b: Unit/Module 1 Schedule

[Python Programming]

Week	Date	Lec	Topic(s)	Release	Due
	Jan 15 (W)	3	What is programming? Variables, data types		
	Jan 17 (F)		*Lab 2*		
3	Jan 20 (M)		**No Class**		
	Jan 22 (W)	4	Python control flow (if, for)		**HW 1**
	Jan 24 (F)		*Lab 3*		
4	Jan 27 (M)	5	Table column operations	HW2	
	Jan 29 (W)	6	Table row operations		
	Jan 31 (F)		*Lab 4*		

Table c: Unit/Module 2 Schedule

[Data Communication]

Week	Date	Lec	Topic(s)	Release	Due
5	Feb 3 (M)	7	Guest lecture: Odum Institute & UNC libraries Data lifecycle and ethics (if time)		
	Feb 5 (W)	8	Line plots, Scatter plots		**HW2**
	Feb 7 (F)		*Lab 5*		
6	Feb 10 (M)		**No Class**	HW3	
	Feb 12 (W)	9	Bar charts, Histograms		

Table d: Unit/Module 3 Schedule

[Statistics]

Week	Date	Lec	Topic(s)	Release	Due
	Feb 14 (F)		*Lab 6*		
7	Feb 17 (M)	10	Summary Statistics, Boxplot		
	Feb 19 (W)	11	Probability, Uniform distribution		**HW3**
	Feb 21 (F)		*Exam 1 review*		
8	Feb 24 (M)		**Exam 1**	HW4	
	Feb 26 (W)	12	Sampling		
	Feb 28 (F)		*Lab 7*		
9	Mar 3 (M)	13	Association, Correlation, Causality		
	Mar 5 (W)	14	Gaussian distribution, Inference		
	Mar 7 (F)		*TBD*		**HW4**
10	Mar 10 (M)		**No Class**		
	Mar 12 (W)		**No Class**		
	Mar 14 (F)		**No Class**		
11	Mar 17 (M)	15	Hypothesis testing	HW5	
	Mar 19 (W)	16	Hypothesis testing		
	Mar 21 (F)		*Lab 8*		

Table e: Unit/Module 4 Schedule

[Prediction]

Week	Date	Lec	Topic(s)	Release	Due
12	Mar 24 (M)	17	Data Science life cycle, Modeling Group project details	Project proposal	
	Mar 26 (W)	18	Intro to machine learning		**HW5**
	Mar 28 (F)		*Lab 9*		
13	Mar 31 (M)	19	Linear regression	HW6	**Project proposal**

Week	Date	Lec	Topic(s)	Release	Due
	Apr 2 (W)	20	Decision trees		
	Apr 4 (F)		*Lab 10*		
14	Apr 7 (M)	21	Underfitting and overfitting		
	Apr 9 (W)	22	KNN		**HW6**
	Apr 11 (F)		*Lab 11*		
15	Apr 14 (M)		Exam 2 review		
	Apr 16 (W)		**Exam 2**		

Table f: Unit/Module 5 Schedule

[More on Data Science]

Week	Date	Lec	Topic(s)	Release	Due
	Apr 18 (F)		**No class**		
16	Apr 21 (M)	23	K-means Clustering		
	Apr 23 (W)	24	Big data, Next steps for data scientists		
	Apr 25 (F)		*Office hours for final group project*		
17	Apr 28 (M)		Time for final group project		
	May 5		**Final "exam" (poster presentation)**		

For Final Exam information, please refer to the University's official <u>Final Examination Schedule</u>.

POLICY STATEMENTS

Academic Policies

University Class Attendance Policy

University Policy: As stated in the University's <u>Class Attendance Policy</u>, no right or privilege exists that permits a student to be absent from any class meetings, except for these University Approved Absences:

1. Authorized University activities: <u>University Approved Absence Office (UAAO) website</u> provides information and <u>FAQs for students</u> and <u>FAQs for faculty</u> related to University Approved Absences

2. Disability/religious observance/pregnancy, as required by law and approved by <u>Accessibility Resources and Service</u> and/or the <u>Equal Opportunity and Compliance Office</u> (EOC)

3. Significant health condition and/or personal/family emergency as approved by the <u>Office of the Dean of Students</u>, <u>Gender Violence Service Coordinators</u>, and/or the <u>Equal Opportunity and Compliance Office</u> (EOC).

Instructors may work with students to meet attendance needs that do not fall within University approved absences. For situations when an absence is not University approved (e.g., a job interview, illness/ flu or club activity), instructors are encouraged to work directly with students to determine the best approach to missed classes and make-up assessment and assignments.

Honor Code Statement

All students are expected to follow the guidelines of the UNC Honor Code. In particular, students are expected to refrain from "lying, cheating, or stealing" in the academic context. If you are unsure about which actions violate the Honor Code, please see me, or consult studentconduct.unc.edu.

Syllabus Changes

The instructor reserves the right to make changes to the syllabus including project due dates and test dates. These changes will be announced as early as possible.

Acceptable Use Policy

By attending the University of North Carolina at Chapel Hill, you agree to abide by the University of North Carolina at Chapel Hill policies related to the acceptable use of IT systems and services. The Acceptable Use Policy (AUP) sets the expectation that you will use the University's technology resources responsibly, consistent with the University's mission. In the context of a class, it's quite likely you will participate in online activities that could include personal information about you or your peers, and the AUP addresses your obligations to protect the privacy of class participants. In addition, the AUP addresses matters of others' intellectual property, including copyright. These are only a couple of typical examples, so you should consult the full Information Technology Acceptable Use Policy, which covers topics related to using digital resources, such as privacy, confidentiality, and intellectual property. Additionally, consult the Safe Computing at UNC website for information about data security policies, updates, and tips on keeping your identity, information, and devices safe.

Data Security & Privacy

UNC-Chapel Hill Privacy Statement: Canvas Discussion Forum, Assignments, DropBox, Gradebook, and Tests & Quizzes tools are designed to share FERPA-protected information privately between instructors and individual students.

Grade Appeal Process

If you have any concerns with grading and/or feel you have been awarded an incorrect grade, please discuss it with me as soon as possible. If we cannot resolve the issue, you may talk to our director of undergraduate studies or the dean.

Generative AI Usage Guidance

The following sections provide the philosophy and specific guidelines for using these tools and features (increasingly, generative AI capabilities will be integrated with everyday applications). Unless I provide other guidelines for an assignment or exam, you should follow these guidelines. Please follow all the guidelines listed on the Generative AI Usage page provided by the provost.

Services & Student Support Policies

Accessibility Resources & Services (ARS)

Accessibility Resources and Service (ARS – ars@unc.edu) receives requests for accommodations, and through the Student and Applicant Accommodations Policy determines eligibility and identifies reasonable accommodations for students with disabilities and/or chronic medical conditions to mitigate or remove the barriers experienced in accessing University courses, programs and activities. ARS also offers its Testing Center resources to students and instructors to facilitate the implementation of testing accommodations. Faculty and instructors with any concerns or questions about accommodations and/or their implementation, are invited to reach out to ARS to discuss.

Counseling & Psychological Services (CAPS)

UNC-Chapel Hill is strongly committed to addressing the mental health needs of a diverse student body. The Heels Care Network website is a place to access the many mental health resources at Carolina. CAPS is the primary mental health provider for students, offering timely access to consultation and connection to clinically appropriate services. Go to their website https://caps.unc.edu/ or visit their facilities on the third floor of the Campus Health building for an initial evaluation to learn more. Students can also call CAPS 24/7 at 919-966-3658 for immediate assistance.

Title IX Resources

Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Reports can be made online to the EOC at https://eoc.unc.edu/report-an-incident/ or by contacting the University's Title IX Coordinator (Elizabeth Hall, titleixcoordinator@unc.edu) or the Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu). Confidential resources include Counseling and Psychological Services and the Gender Violence Services Coordinators (gvsc@unc.edu). Additional resources are available at safe.unc.edu.

Policy on Non-Discrimination

The University is committed to providing an inclusive and welcoming environment for all members of our community and to ensuring that educational and employment decisions are based on individuals' abilities and qualifications. Consistent with this principle and applicable laws, the University's Policy Statement on Non-Discrimination offers access to its educational programs and activities as well as employment terms and conditions without respect to race, color, gender, national origin, age, religion, genetic information, disability, veteran's status, sexual orientation, gender identity or gender expression. Such a policy ensures that only relevant factors are considered, and that equitable and consistent standards of conduct and performance are applied. If you are experiencing harassment or discrimination, you can seek assistance and file a report through the Report and Response Coordinators (email reportandresponse@unc.edu or see additional contact info at safe.unc.edu) or the Equal Opportunity and Compliance Office at https://eoc.unc.edu/report-an-incident/.

Diversity Statement

I value the perspectives of individuals from all backgrounds reflecting the diversity of our students. I broadly define diversity to include race, gender identity, national origin, ethnicity, religion, social class, age, sexual orientation, political background, and physical and learning ability. I strive to

make this classroom an inclusive space for all students. Please let me know if there is anything I can do to improve. I appreciate any suggestions.

Undergraduate Testing Center
The College of Arts and Sciences provides a secure, proctored environment in which exams can be taken. The center works with instructors to proctor exams for their undergraduate students who are not registered with ARS and who do not need testing accommodations as provided by ARS. In other words, the Center provides a proctored testing environment for students who are unable to take an exam at the normally scheduled time (with pre-arrangement by your instructor). For more information, visit http://testingcenter.web.unc.edu/.

Learning Center
Want to get the most out of this course or others this semester? Visit UNC's Learning Center at http://learningcenter.unc.edu to make an appointment or register for an event. Their free, popular programs will help you optimize your academic performance. Try academic coaching, peer tutoring, STEM support, ADHD/LD services, workshops and study camps, or review tips and tools available on the website.

Writing Center
For free feedback on any course writing projects, check out UNC's Writing Center. Writing Center coaches can assist with any writing project, including multimedia projects and application essays, at any stage of the writing process. You don't even need a draft to come visit. To schedule a 45-minute appointment, review quick tips, or request written feedback online, visit http://writingcenter.unc.edu.