Data Science, Undergraduate (DATA)

Courses

DATA C4AC Data and Justice 4 Units

Terms offered: Spring 2025

This course engages students with fundamental questions of justice in relation to data and computing in American society. Data collection, visualization, and analysis have been entangled in the struggle for racial and social justice because they can make injustice visible, imaginable, and thus actionable. Data has also been used to oppress minoritized communities and institutionalize, rationalize, and naturalize systems of racial violence. The course examines key sites of justice involving data (such as citizenship, policing, prisons, environment, and health). Along with critical social science tools, students gain introductory experience and do collaborative and creative projects with data science using real-world data.

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture and 1.5 hours of discussion per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternative to final exam.

Formerly known as: Data Science, Undergraduate 4AC

Also listed as: STS C4AC

DATA 6 Introduction to Computational Thinking with Data Science and Society 4 Units

Terms offered: Fall 2025, Fall 2024

This foundational Data Science course combines inferential and computational thinking as applied to the fundamentals of quantitative social inquiry. Apply critical concepts and skills in computer programming to conduct quantitative social science research in various contexts, including economic outcomes, public health, environmental justice, privacy, bioethics, and social networks. Understand the process of using data for quantitative analysis and how to develop a variety of figures, combined with text, to communicate their findings. The focus is on data exploration and identifying patterns relevant to social concepts, rather than inferences and predictions. The course can serve as a precursor to Data 8: Foundations of Data Science.

Objectives & Outcomes

Course Objectives: C6 also includes quantitative reasoning concepts that aren't covered in Data 8. These include certain topics in: principles of data visualization; simulation of random processes; and understanding numerical functions through their graphs. This will help prepare students for computational and quantitative courses other than Data 8.
C6 takes advantage of the complementarity of computing and quantitative reasoning to enliven abstract ideas and build students' confidence in their ability to solve real problems with quantitative tools. Students learn computer science concepts and immediately apply them to plot functions, visualize data, and simulate random events.

Foundations of Data Science (CS/Info/Stat C8, a.k.a. Data 8) is an increasingly popular class for entering students at Berkeley. Data 8 builds students' computing skills in the first month of the semester, and students rely on these skills as the course progresses. For some students, particularly those with little prior exposure to computing, developing these skills benefits from further time and practice. C6 is a rapid introduction to Python programming, visualization, and data analysis, which will prepare students for success in Data 8.

Student Learning Outcomes: Appreciate the interdisciplinary nature of data science.
Create and use visualizations to understand univariate data and to identify associations or causal relationships in bivariate data.
Formulate questions about data and perform exploratory data analysis.
Perform basic computations in Python, and be able to work with tabular data.
Run and understand basic probabilistic simulations.
Understand the syntactic structure of Python code.
Use good practices in Python programming.

Rules & Requirements

Credit Restrictions: Students will receive no credit for DATA 6 after completing DATA C8, DATA C88C, or COMPSCI 61A.

Hours & Format

Fall and/or spring: 15 weeks - 2 hours of lecture, 0 hours of discussion, and 2 hours of laboratory per week

Summer: 6 weeks - 6 hours of lecture, 2 hours of discussion, and 2 hours of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

DATA C8 Foundations of Data Science 4 Units

Terms offered: Fall 2025, Summer 2025 8 Week Session, Spring 2025, Fall 2024, Spring 2023, Fall 2022, Spring 2022, Fall 2021, Summer 2021 8 Week Session, Fall 2020

Foundations of data science from three perspectives: inferential thinking, computational thinking, and real-world relevance. Given data arising from some real-world phenomenon, how does one analyze that data so as to understand that phenomenon? The course teaches critical concepts and skills in computer programming and statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, document collections, geographical data, and social networks. It delves into social and legal issues surrounding data analysis, including issues of privacy and data ownership.

Rules & Requirements

Prerequisites: This course may be taken on its own, but students are encouraged to take it concurrently with a data science connector course (numbered 88 in a range of departments)

Credit Restrictions: Students will receive no credit for DATA C8\COMPSCI C8\INFO C8\STAT C8 after completing COMPSCI 8, or DATA 8. A deficient grade in DATA C8\COMPSCI C8\INFO C8\STAT C8 may be removed by taking COMPSCI 8, COMPSCI 8, or DATA 8.

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture and 2 hours of laboratory per week

Summer: 8 weeks - 6 hours of lecture and 4 hours of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Formerly known as: Computer Science C8/Statistics C8/Information C8

Also listed as: COMPSCI C8/INFO C8/STAT C8

DATA 36 Data Scholars Seminar 1 Unit

Terms offered: Not yet offered

Data 36 is a seminar for Data Scholars who are concurrently taking Data C8: Foundations of Data Science. Data Scholars is a cohort-model program to provide support in exploring and potentially declaring a Data Science major for students with little to no computational or statistical background prior to coming to the university. The primary role of Data 36 is to provide technical instruction to review and reinforce concepts in Data C8, in order to support Data Scholars' individual learning and success in the Data C8 course. Data 36 also provides an introduction to the Data Science curriculum at UC Berkeley, and the overall Data Science landscape in both industry and academia.

Hours & Format

Fall and/or spring: 15 weeks - 1.5 hours of seminar per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Offered for pass/not pass grade only. Final exam not required.

DATA 88 Data Science Connector 2 - 4 Units

Terms offered: Spring 2022, Spring 2021, Fall 2020
Designed to be taken in conjunction with the Foundations of Data Science (COMPSCI/INFO/STAT C8) course, each connector course will flesh out data science ideas in the context of one particular field. Blending inferential thinking and computational thinking, the course relies on the increasing availability of datasets across a wide range of human endeavor, and students' natural interest in such data, to teach students to work actively with data in a field of their interest and to interpret and critique their analyses of data. Topics vary by field, and several topics will be offered each term.

Objectives & Outcomes

Course Objectives: Discuss how to formulate and substantiate an argument with evidence
Explain a variety of analytic and visualization techniques
Explore approaches to effective communication
Explore the challenges with working with primary and secondary data

Student Learning Outcomes: Apply data analysis to evaluate everyday problems
Communicate effectively in written, spoken, and graphical form about specific issues
Interpret statistical results
Know how to locate and use primary data sources
Obtain and/or collect relevant data using specific qualitative and/or quantitative research methods
Understand how to use empirical evidence to evaluate an argument

Rules & Requirements

Prerequisites: Instructors may require students to enroll concurrently or have completed Data 8 (COMPSCI/STAT/INFO C8)

Repeat rules: Course may be repeated for credit without restriction. Students may enroll in multiple sections of this course within the same semester.

Hours & Format

Fall and/or spring: 15 weeks - 2-4 hours of seminar per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternative to final exam.

DATA 88E Economic Models 2 Units

Terms offered: Fall 2025, Fall 2024, Fall 2023
This class aims to motivate and illustrate key concepts in economics through a series of exercises and examples that use Python Jupyter notebooks. The class covers concepts from introductory economics, microeconomic theory, econometrics, development economics, environmental economics and public economics. The course provides data science students a pathway to apply Python programming and data science concepts within the discipline of economics. The course will also gives economics students a pathway to apply programming to reinforce fundamental concepts and to advance the level of study in upper division coursework and possible thesis work.

Objectives & Outcomes

Course Objectives: Demonstrate how to construct understanding of concepts in economics by developing and coding examples
Illustrate topics in economics through coding applications
Motivate basics of econometrics from a data science perspective

Student Learning Outcomes: Programmatically create and interpret graphs of simple equations used in microeconomics
Reason about and solve simple equations used in microeconomics through coding
Understand basic concepts in economics

Rules & Requirements

Prerequisites: You must have taken Data C8 or be concurrently enrolled in Data C8 to take this course. That being said, we are able to make exceptions if you have prior programming or data science experience; please email the course staff if you have any questions. Prior economics knowledge may be helpful but is not necessary

Hours & Format

Fall and/or spring: 15 weeks - 2 hours of lecture per week

Summer:
6 weeks - 5 hours of lecture per week
8 weeks - 4 hours of lecture per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternative to final exam.

DATA C88C Computational Structures in Data Science 3 Units

Terms offered: Fall 2025, Summer 2025 8 Week Session, Spring 2025, Spring 2023, Fall 2022

Development of Computer Science topics appearing in Foundations of Data Science (C8); expands computational concepts and techniques of abstraction. Understanding the structures that underlie the programs, algorithms, and languages used in data science and elsewhere. Mastery of a particular programming language while studying general techniques for managing program complexity, e.g., functional, object-oriented, and declarative programming. Provides practical experience with composing larger systems through several significant programming projects.

Objectives & Outcomes

Course Objectives: Develop a foundation of computer science concepts that arise in the context of data analytics, including algorithm, representation, interpretation, abstraction, sequencing, conditional, function, iteration, recursion, types, objects, and testing, and develop proficiency in the application of these concepts in the context of a modern programming language at a scale of whole programs on par with a traditional CS introduction course.

Student Learning Outcomes: Students will be able to demonstrate a working knowledge of these concepts and a proficiency of programming based upon them sufficient to construct substantial stand-alone programs.

Rules & Requirements

Credit Restrictions: Students will receive no credit for DATA C88C after completing COMPSCI 61A.

Hours & Format

Fall and/or spring: 15 weeks - 2-2 hours of lecture, 2-2 hours of laboratory, and 0-1 hours of supplement per week

Summer: 8 weeks - 4-4 hours of lecture, 4-4 hours of laboratory, and 0-2 hours of supplement per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Instructors: Ball, Culler, DeNero

Formerly known as: Computer Science 88

Also listed as: COMPSCI C88C

DATA C88S Probability and Mathematical Statistics in Data Science 3 Units

Terms offered: Spring 2025, Spring 2024, Summer 2023 8 Week Session, Fall 2022

In this connector course we will state precisely and prove results discovered while exploring data in Data C8. Topics include: probability, conditioning, and independence; random variables; distributions and joint distributions; expectation, variance, tail bounds; Central Limit Theorem; symmetries in random permutations; prior and posterior distributions; probabilistic models; bias-variance tradeoff; testing hypotheses; correlation and the regression model.

Rules & Requirements

Prerequisites: Prerequisite: one semester of calculus at the level of Math 16A, Math 10A, Math 1A, or Math 51. Corequisite or Prerequisite: Foundations of Data Science (COMPSCI C8 / DATA C8 / INFO C8 / STAT C8)

Credit Restrictions: Students will receive no credit for DATA C88S after completing STAT 134, STAT 140, STAT 135, or DATA C102.

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture and 2 hours of discussion per week

Summer: 8 weeks - 6 hours of lecture and 4 hours of discussion per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Formerly known as: Statistics 88

Also listed as: STAT C88S

DATA 94 Special Topics in Data Science 1 - 4 Units

Terms offered: Spring 2025, Fall 2024, Spring 2021

Topics will vary semester to semester.

Hours & Format

Fall and/or spring: 15 weeks - 1-3 hours of lecture, 0-2 hours of discussion, and 0-2 hours of laboratory per week

Summer: 8 weeks - 2-6 hours of lecture, 0-4 hours of discussion, and 0-4 hours of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

DATA C100 Principles & Techniques of Data Science 4 Units

Terms offered: Fall 2025, Summer 2025 8 Week Session, Spring 2025, Summer 2024 8 Week Session, Fall 2022, Fall 2021, Fall 2020
In this course, students will explore the data science lifecycle, including question formulation, data collection and cleaning, exploratory data analysis and visualization, statistical inference and prediction, and decision-making. This class will focus on quantitative critical thinking and key principles and techniques needed to carry out this cycle. These include languages for transforming, querying and analyzing data; algorithms for machine learning methods including regression, classification and clustering; principles behind creating informative data visualizations; statistical concepts of measurement error and prediction; and techniques for scalable data processing.

Rules & Requirements

Prerequisites: DATA C8 or STAT 20 with a C- or better, or Pass; and COMPSCI 61A, COMPSCI/DATA C88C, or ENGIN 7 with a C- or better, or Pass; Corequisite: MATH 54, 56, 110, EECS 16A, PHYSICS 89 or equivalent linear algebra (C- or better, or Pass, required if completed prior to Data C100)

Credit Restrictions: Students will receive no credit for DATA C100\STAT C100\COMPSCI C100 after completing DATA 100. A deficient grade in DATA C100\STAT C100\COMPSCI C100 may be removed by taking DATA 100.

Hours & Format

Fall and/or spring: 15 weeks - 3-3 hours of lecture, 1-1 hours of discussion, and 0-1 hours of laboratory per week

Summer: 8 weeks - 6-6 hours of lecture, 2-2 hours of discussion, and 0-2 hours of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Instructors: Gonzalez, Nourozi, Perez, Yan

Formerly known as: Statistics C100/Computer Science C100

Also listed as: COMPSCI C100/STAT C100

DATA C101 Data Engineering 4 Units

Terms offered: Fall 2025, Spring 2025
This course will cover the principles and practices of managing data at scale, with a focus on use cases in data analysis and machine learning. We will cover the entire life cycle of data management and science, ranging from data preparation to exploration, visualization and analysis, to machine learning and collaboration, with a focus on ensuring reliable, scalable operationalization.

Rules & Requirements

Prerequisites: COMPSCI 61B, or INFO 206B, or equivalent courses in programming with a C- or better, or Pass; and COMPSCI C100 / DATA C100 / STAT C100, or COMPSCI 189, or INFO 251, or DATA 144, or equivalent upper-division course in data science with a C- or better, or Pass

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture and 1 hour of discussion per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Instructors: Hellerstein, Jain, Parameswaran, Yan

Formerly known as: Data Science, Undergraduate 101

Also listed as: COMPSCI C187

DATA C102 Data, Inference, and Decisions 4 Units

Terms offered: Fall 2025, Spring 2025, Fall 2024
This course develops the probabilistic foundations of inference in data science, and builds a comprehensive view of the modeling and decision-making life cycle in data science including its human, social, and ethical implications. Topics include: frequentist and Bayesian decision-making, permutation testing, false discovery rate, probabilistic interpretations of models, Bayesian hierarchical models, basics of experimental design, confidence intervals, causal inference, Thompson sampling, optimal control, Q-learning, differential privacy, clustering algorithms, recommendation systems and an introduction to machine learning tools including decision trees, neural networks and ensemble methods.

Rules & Requirements

Prerequisites: Math 54 or 56 or 110 or Stat 89A or Physics 89 or both of EECS 16A and 16B with a C- or better, or Pass; Data/Stat/CompSci C100 with a C- or better, or Pass; and any of EECS 126, Data/Stat C140, Stat 134, IndEng 172, Math 106 with a C- or better, or Pass. Data/Stat C140 or EECS 126 are preferred

Credit Restrictions: Students will receive no credit for DATA C102 after completing STAT 102, or DATA 102. A deficient grade in DATA C102 may be removed by taking STAT 102, STAT 102, or DATA 102.

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture, 1 hour of discussion, and 1 hour of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Formerly known as: Statistics 102

Also listed as: STAT C102

DATA C104 Human Contexts and Ethics of Data - DATA/History/STS 4 Units

Terms offered: Fall 2025, Summer 2025 8 Week Session, Spring 2025, Spring 2022, Fall 2020, Spring 2020
This course teaches you to use the tools of applied historical thinking and Science, Technology, and Society (STS) to recognize, analyze, and shape the human contexts and ethics of data. It addresses key topics such as doing ethical data science amid shifting definitions of human subjects, consent, and privacy; the changing relationship between data, democracy, and law; the role of data analytics in how corporations and governments provide public goods such as health and security to citizens; sensors, machine learning and artificial intelligence and changing landscapes of labor, industry, and city life. It prepares you to engage as a knowledgeable and responsible citizen and professional in the varied arenas of our datafied world.

Rules & Requirements

Credit Restrictions: Students will receive no credit for DATA C104\HISTORY C184D\STS C104D after completing DATA 104. A deficient grade in DATA C104\HISTORY C184D\STS C104D may be removed by taking DATA 104.

Hours & Format

Fall and/or spring: 15 weeks - 3-3 hours of lecture and 0-1.5 hours of discussion per week

Summer:
6 weeks - 7.5-7.5 hours of lecture and 0-3.5 hours of discussion per week
8 weeks - 6-6 hours of lecture and 0-3 hours of discussion per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Formerly known as: History C184D/Science and Technology Studies C104D

Also listed as: HISTORY C184D/STS C104D

DATA C131A Statistical Methods for Data Science 4 Units

Terms offered: Fall 2025, Summer 2025 8 Week Session, Fall 2024, Fall 2023

This course teaches a broad range of statistical methods that are used to solve data problems. Topics include group comparisons and ANOVA, standard parametric statistical models, multivariate data visualization, multiple linear regression, logistic regression and classification, regression trees and random forests. An important focus of the course is on statistical computing and reproducible statistical analysis. The course and lab include hands-on experience in analyzing real world data from the social, life, and physical sciences. The R statistical language is used.

Rules & Requirements

Prerequisites: DATA/COMPSCI/INFO/STAT C8 or STAT 20; and MATH 1A, MATH 51, MATH 16A, or MATH 10A/10B. Strongly recommended corequisite: STAT 33A or STAT 133

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture and 2 hours of laboratory per week

Summer: 8 weeks - 6 hours of lecture and 4 hours of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Formerly known as: Statistics 131A

Also listed as: STAT C131A

DATA C140 Probability for Data Science 4 Units

Terms offered: Fall 2025, Spring 2025, Fall 2024, Spring 2024

An introduction to probability, emphasizing the combined use of mathematics and programming. Discrete and continuous families of distributions. Bounds and approximations. Transforms and convergence. Markov chains and Markov Chain Monte Carlo. Dependence, conditioning, Bayesian methods. Maximum likelihood, least squares prediction, the multivariate normal, and multiple regression. Random permutations, symmetry, and order statistics. Use of numerical computation, graphics, simulation, and computer algebra.

Objectives & Outcomes

Course Objectives: Data/Stat C140 is a probability course for Data C8 graduates who have taken more mathematics and wish to go deeper into data science. The emphasis on simulation and the bootstrap in Data C8 gives students a concrete sense of randomness and sampling variability. Data/Stat C140 capitalizes on this, abstraction and computation complementing each other throughout. Topics in statistical theory are included to allow students to proceed to modeling and statistical learning classes without taking a further semester of mathematical statistics.

Student Learning Outcomes: Understand the difference between math and simulation, and appreciate the power of both
Use a variety of approaches to problem solving
Work with probability concepts algebraically, numerically, and graphically

Rules & Requirements

Prerequisites: DATA/COMPSCI/INFO/STAT C8, or both STAT 20 and one of COMPSCI 61A or COMPSCI/DATA C88C with C- or better, or Pass; and one year of calculus at the level of MATH 1A-1B or MATH 51-52 or higher, with C- or better, or Pass. Corequisite: MATH 54, MATH 56, EECS 16B, MATH 110 or equivalent linear algebra (C- or better, or Pass, required if completed prior to enrollment in Data/Stat C140)

Credit Restrictions: Students will receive no credit for STAT C140 after completing STAT 134, or EECS 126.

Hours & Format

Fall and/or spring:
15 weeks - 3-3 hours of lecture, 1-1 hours of discussion, 1-1 hours of supplement, and 0-1 hours of voluntary per week
15 weeks - 3-3 hours of lecture, 2-2 hours of discussion, 0-0 hours of supplement, and 0-1 hours of voluntary per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Formerly known as: Statistics C140/Data Science, Undergraduate C140

Also listed as: STAT C140

DATA 144 Data Mining and Analytics 3 Units

Terms offered: Fall 2025, Fall 2024, Fall 2023

Data Mining and Analytics introduces students to practical fundamentals of data mining and emerging paradigms of data mining and machine learning with enough theory to aid intuition building. The course is project-oriented, with a project beginning in class every week. The in-class portion of the project is meant to be collaborative and a time for the instructor and GSIs to work closely with project groups to understand the objectives, help work through software logistics, and connect project work to lecture. Lectures will introduce theories, concepts, practical contexts, and algorithms. Students should expect to leave the class with hands-on, contemporary data mining skills they can confidently apply in research and industry.

Objectives & Outcomes

Course Objectives: Conduct manual feature engineering (from domain knowledge) vs. machine induced featurization (representation learning)
Develop intuition in various machine learning classification algorithms (e.g. decision trees, feed-forward neural networks, recurrent neural networks, skip-grams) and clustering techniques (e.g. k-means, spectral)
Foster critical thinking about real-world actionability from analytics
Provide an overview of issues in research and practice that will affect the practice of data science in a variety of domains

Student Learning Outcomes: Develop capabilities in a range of data mining techniques
Gain the ability to solve problems in data mining research and practice
Think critically about how to assess analytics
Use data mining and analytics in a domain of application

Rules & Requirements

Prerequisites: Corequisite: Data/CompSci/Stat C100 (C- or better, or Pass, required if completed prior to Data 144)

Credit Restrictions: Students will receive no credit for DATA 144 after completing INFO 154. A deficient grade in DATA 144 may be removed by taking INFO 154.

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternative to final exam.

Instructor: Pardos

DATA 145 Evidence and Uncertainty 4 Units

Terms offered: Not yet offered

When we learn about the world from data, how much can we rely on the conclusions we draw? How do we know if we could do better? This course will cover the statistical theory required to measure and control our uncertainty when we analyze large and complex modern data sets. We will use mathematical and computational lenses to examine optimality properties and error bounds. Topics include the Bayesian and frequentist paradigms, asymptotic and finite-sample methods, parametric and nonparametric techniques, causality, and multiple testing.

Objectives & Outcomes

Course Objectives: The course is primarily intended for students interested in machine learning and artificial intelligence, whether in industry or academia. It will also be helpful preparation for students who want to study statistics at the graduate level. It will examine approaches to defining and modeling uncertainty, and will identify connections and differences between the frequentist and Bayesian paradigms. The emphasis will be on situations where classical statistical methods do not apply and only minimal distributional assumptions can be made. In such settings, computational solutions might be feasible if the mathematics becomes intractable.

Student Learning Outcomes: Students will understand the need for statistical inference in data science and why the Bayesian viewpoint is so pervasive in modern data analysis. They will recognize the power and limitations of classical methods and newer computationally intensive approaches. They will appreciate the optimality or near-optimality properties of some asymptotic methods, and learn how to work in finite-sample settings where asymptotic methods do not apply. Throughout, they will use mathematics and computation as needed for problem solving. Upon leaving the course, students should be able to follow upcoming developments in the field without extensive further education in statistical inference.

Rules & Requirements

Prerequisites: Math 53, Data C100, and either Data C140 or EECS 126, with a C- or better or Pass

Credit Restrictions: Students will receive no credit for DATA 145 after completing STAT 210A. A deficient grade in DATA 145 may be removed by taking STAT 210A.

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture and 1 hour of discussion per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Final exam required.

Instructors: Adhikari, Fithian

DATA C146 Foundations for Computational Precision Health 3 Units

Terms offered: Fall 2025
Students will build expertise in developing machine-learning tools to address challenges in health care. The course emphasizes both “how to formulate useful computational health problems”, and “how to develop computational solutions”. On the health side, we’ll get clinical guest lectures exploring challenges across diverse areas of healthcare (e.g., cardiology, cancer, primary care). On the computational side, the course will cover machine learning and deep learning foundations, state-of-the-art neural networks, and then advanced research topics. The course will emphasize rigorous evaluation, algorithmic bias, deployment, and auditing. The class will culminate in an open-ended final project, integrating skills learned in the course.

Objectives & Outcomes

Course Objectives: Articulate the key challenges in diverse areas of healthcare, including cancer, cardiology, and emergency care.
Develop machine learning methods to leverage, text, images, volumes and time-series data
Formulate precise computational research questions to improve healthcare.
Understand and perform clinically-informed evaluation analyses of predictive ML tools
Understand the role of the various information modalities (e.g., radiology, pathology, labs) in health care. This means understanding why the various modalities are acquired, what they physically capture, and what decisions they enable.

Rules & Requirements

Prerequisites: Data C100 and Data C140

Hours & Format

Fall and/or spring: 15 weeks - 2 hours of lecture per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternate method of final assessment during regularly scheduled final exam group (e.g., presentation, final project, etc.).

Instructors: Yala, Chen

Also listed as: CPH C100

DATA C182 Designing, Visualizing and Understanding Deep Neural Networks 4 Units

Terms offered: Fall 2025, Spring 2025, Fall 2024, Spring 2008
Deep Networks have revolutionized computer vision, language technology, robotics and control. They have growing impact in many other areas of science and engineering. They do not however, follow a closed or compact set of theoretical principles. In Yann Lecun's words they require "an interplay between intuitive insights, theoretical modeling, practical implementations, empirical studies, and scientific analyses." This course attempts to cover that ground.

Objectives & Outcomes

Student Learning Outcomes: Students will come to understand visualizing deep networks. Exploring the training and use of deep networks with visualization tools.
Students will learn design principles and best practices: design motifs that work well in particular domains, structure optimization and parameter optimization.
Understanding deep networks. Methods with formal guarantees: generative and adversarial models, tensor factorization.

Rules & Requirements

Prerequisites: MATH 53, MATH 54, and COMPSCI 61B; COMPSCI 70 or STAT 134; COMPSCI 189 is recommended

Credit Restrictions: Students will receive no credit for COMPSCI 182 after completing COMPSCI W182, or COMPSCI L182. A deficient grade in COMPSCI 182 may be removed by taking COMPSCI L182, COMPSCI W182, COMPSCI W182, or COMPSCI L182.

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of lecture and 1 hour of discussion per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternative to final exam.

Instructor: Gonzalez

Formerly known as: Computer Science 182

Also listed as: COMPSCI C182

DATA 188 Advanced Data Science Connector 2 - 4 Units

Terms offered: Spring 2024
Designed to be taken concurrently with or after Principles and Techniques of Data Science (Data C100) or Probability for Data Science (Data C140) or both, each connector course consists of an intensive study of data science ideas in a particular field. Topics include the development of the theory of data science, the application of data science in a variety of domains, its human and social contexts, and anything else. Topics vary by field, and more than one topic may be offered in a semester.

Objectives & Outcomes

Course Objectives: Develop theoretical mastery in data science topics, address the challenges of gathering data and converting it to usable formats, develop skills in selecting appropriate data science methods, explore approaches to decision-making and effective communication.

Student Learning Outcomes: Understand and apply theory in an area of data science, or follow the data science life cycle in a domain of application from question formulation to the use of advanced data science methods and the communication of results.

Rules & Requirements

Prerequisites: Prerequisites or corequisites may vary depending on topic. Consult the Schedule of Classes or department website for details

Repeat rules: Course may be repeated for credit when topic changes. Students may enroll in multiple sections of this course within the same semester.

Hours & Format

Fall and/or spring: 15 weeks - 2-3 hours of seminar and 0-1 hours of discussion per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternate method of final assessment during regularly scheduled final exam group (e.g., presentation, final project, etc.).

DATA H195A Data Science Honors Thesis Seminar 2 Units

Terms offered: Prior to 2007
The senior honors thesis seminar gives students an opportunity to experience firsthand what it means to do data science research. Over two semesters, students will learn to formulate a research problem, design a research strategy, collect evidence, and write up the findings and analysis. The first semester focuses primarily on the preparation and implementation of a research proposal, as well as data management strategies. During the second semester, we will emphasize analysis and writing. The final result will be a hybrid product with a 20-25 page research paper, with data visualizations and analysis tables, along with a documented data source, annotated code, well documented Github repository, and open science posting of the project.

Objectives & Outcomes

Course Objectives: Assist students with project organization and management.
Convey approaches to effective writing and visual communication.
Discuss how to formulate and substantiate an argument with evidence.
Explain approaches to designing a research question and project.
Explore a variety of analytic and visualization techniques and discuss their appropriateness to different research questions.
Identify the challenges in data acquisition and management.

Student Learning Outcomes: Communicate effectively in written, spoken, and graphical form.
Develop an understanding of data availability, constraints, and ethics.
Develop data management skills.
Develop reproducible research and interpret results.
Formulate a proposal for a research project.
Learn how to develop a research question and project.
Understand how to organize empirical work into a written document.
Understand how to use empirical evidence to construct an argument.

Rules & Requirements

Prerequisites: There are no specific prerequisites. Students must be accepted into the data science honors program in order to take this course. Students must complete H195A in order to enroll in H195B

Hours & Format

Fall and/or spring: 15 weeks - 2 hours of seminar per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternative to final exam.

DATA H195B Data Science Honors Thesis Seminar 2 Units

Terms offered: Spring 2020

The senior honors thesis seminar gives students an opportunity to experience firsthand what it means to do data science research. Over two semesters, students will learn to formulate a research problem, design a research strategy, collect evidence, and write up the findings and analysis. The first semester focuses primarily on the preparation and implementation of a research proposal, as well as data management strategies. During the second semester, we will emphasize analysis and writing. The final result will be a hybrid product with a 20-25 page research paper, with data visualizations and analysis tables, along with a documented data source, annotated code, well documented Github repository, and open science posting of the project.

Objectives & Outcomes

Course Objectives: Assist students with project organization and management.
Convey approaches to effective writing and visual communication.
Discuss how to formulate and substantiate an argument with evidence.
Explain approaches to designing a research question and project.
Explore a variety of analytic and visualization techniques and discuss their appropriateness to different research questions.
Identify the challenges in data acquisition and management.

Student Learning Outcomes: Communicate effectively in written, spoken, and graphical form.
Develop an understanding of data availability, constraints, and ethics.
Develop data management skills.
Develop reproducible research and interpret results.
Formulate a proposal for a research project.
Learn how to develop a research question and project.
Understand how to organize empirical work into a written document.
Understand how to use empirical evidence to construct an argument.

Rules & Requirements

Prerequisites: There are no specific prerequisites. Students must be accepted into the data science honors program in order to take this course. Students must complete H195A in order to enroll in H195B

Hours & Format

Fall and/or spring: 15 weeks - 2 hours of seminar per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Letter grade. Alternative to final exam.

DATA 197 Field Studies in Data Science 0.5 - 6 Units

Terms offered: Fall 2019

Students take part in organized individual field sponsored programs with off-campus organizations or tutoring/mentoring relevant to specific aspects and applications of data science on or off campus. Note Summer CPT or OPT students: written report required. Course may not count toward major requirements but will be counted in the cumulative units toward graduation.

Rules & Requirements

Prerequisites: Consent of instructor (see department advisor). Upper-division standing

Repeat rules: Course may be repeated for credit with advisor consent.

Hours & Format

Fall and/or spring: 15 weeks - 1.5-18 hours of fieldwork per week

Summer:
8 weeks - 3-34 hours of fieldwork per week
10 weeks - 2.5-27 hours of fieldwork per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Offered for pass/not pass grade only. Alternative to final exam.

DATA 198 Directed Group Studies for Advanced Undergraduates 1 - 4 Units

Terms offered: Spring 2025, Fall 2024, Spring 2024

Written proposal must be approved by a faculty sponsor, who will serve as Instructor of Record. Seminars for the group study of selected topics, which will vary from semester to semester. Topics may be initiated by students.

Rules & Requirements

Prerequisites: Instructors may require students to enroll concurrently or have completed Data 8 (COMPSCI/STAT/INFO C8). Upper-division standing and consent of instructor

Repeat rules: Course may be repeated for credit without restriction. Students may enroll in multiple sections of this course within the same semester.

Hours & Format

Fall and/or spring: 15 weeks - 1-4 hours of directed group study per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Offered for pass/not pass grade only. Final exam not required.

DATA 199 Supervised Independent Study and Research 1 - 4 Units

Terms offered: Prior to 2007
Independent study and research by arrangement with faculty or staff. This course allows students to obtain course credit for participation in undergraduate research. Students may opt either to participate in a semester-long series of workshops which provide a guided research experience with project milestone assignments and regular feedback, or they may opt to work independently with supervision from one faculty research mentor.

Objectives & Outcomes

Student Learning Outcomes: Develop and refine skills acquired in other courses in a hands-on, self-directed research project.
Identify how to properly manage data and describe best practices in programming and analytics.
Integrate feedback from an instructor into research on a regular basis.
Learn how to structure and complete a research project working independently.

Rules & Requirements

Prerequisites: Instructors may require students to enroll concurrently or have completed Data 8 (COMPSCI/STAT/INFO C8). Upper-division standing and consent of instructor

Repeat rules: Course may be repeated for credit without restriction.

Hours & Format

Fall and/or spring: 15 weeks - 3-12 hours of independent study per week

Summer:
6 weeks - 7.5-30 hours of independent study per week
8 weeks - 5.5-22.5 hours of independent study per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Undergraduate

Grading/Final exam status: Offered for pass/not pass grade only. Alternative to final exam.

DATA 200S Principles and Techniques of Data Science 3 Units

Terms offered: Fall 2025, Spring 2025, Fall 2024
Explores the data science lifecycle: question formulation, data collection and cleaning, exploratory, analysis, visualization, statistical inference, prediction, and decision-making. Focuses on quantitative critical thinking and key principles and techniques: languages for transforming, querying and analyzing data; algorithms for machine learning methods: regression, classification and clustering; principles of informative visualization; measurement error and prediction; and techniques for scalable data processing. Research term project.

Rules & Requirements

Prerequisites: DATA/COMPSCI/INFO/STAT C8; and COMPSCI 61A or COMPSCI/DATA C88C. Corequisites: MATH 54 or EECS 16A

Credit Restrictions: Students will receive no credit for DATA 200S after completing DATA C100, or DATA C200. A deficient grade in DATA 200S may be removed by taking DATA C200.

Hours & Format

Fall and/or spring: 15 weeks - 3-3 hours of lecture, 1-1 hours of discussion, and 0-1 hours of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Graduate

Grading: Letter grade.

DATA C200 Principles and Techniques of Data Science 4 Units

Terms offered: Spring 2025, Fall 2024, Spring 2024, Spring 2023, Spring 2022, Spring 2021, Spring 2020
Explores the data science lifecycle: question formulation, data collection and cleaning, exploratory, analysis, visualization, statistical inference, prediction, and decision-making. Focuses on quantitative critical thinking and key principles and techniques: languages for transforming, querying and analyzing data; algorithms for machine learning methods: regression, classification and clustering; principles of informative visualization; measurement error and prediction; and techniques for scalable data processing. Research term project.
Rules & Requirements

Prerequisites: COMPSCI C8 / INFO C8 / STAT C8 or ENGIN 7; and either COMPSCI 61A or COMPSCI 88. Corequisites: MATH 54 or EECS 16A

Credit Restrictions: Students will receive no credit for DATA C200\COMPSCI C200A\STAT C200C after completing DATA C100.

Hours & Format

Fall and/or spring:
8 weeks - 6-6 hours of lecture, 2-2 hours of discussion, and 0-2 hours of laboratory per week
15 weeks - 3-3 hours of lecture, 1-1 hours of discussion, and 0-1 hours of laboratory per week

Summer: 8 weeks - 6-6 hours of lecture, 2-2 hours of discussion, and 0-2 hours of laboratory per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Graduate

Grading: Letter grade.

Formerly known as: Statistics C200C/Computer Science C200A

Also listed as: COMPSCI C200A/STAT C200C

DATA C204 Human Contexts and Ethics of Data 4 Units

Terms offered: Spring 2025, Fall 2024, Spring 2024
This course teaches you to use approaches from the across the humanities and interpretive social sciences and tools of Science, Technology, and Society (STS) to recognize, analyze, and shape the human contexts, social implications, and ethics of data and data technologies, including data analytics, algorithmic decision systems, machine learning (ML), and artificial intelligence (AI).
Rules & Requirements

Prerequisites: Graduate standing or permission of the instructor. Graduate students without previous (undergraduate or graduate-level) preparation in the interpretive social sciences or humanities are encouraged to confer with the instructor before enrolling

Hours & Format

Fall and/or spring: 15 weeks - 2 hours of seminar per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Graduate

Grading: Letter grade.

Instructor: Carson

Also listed as: HISTORY C254/STS C204

DATA 298 Directed Group Studies for Graduates 1 - 4 Units

Terms offered: Prior to 2007
DATA 298 provides credit for directed group study by graduate students working closely with an instructor who is a faculty member. Students are responsible for finding an instructor to supervise their work, and they will meet with that instructor weekly or bi-weekly. Faculty members must commit to supervising and evaluating the students' work and be available to meet regularly as required by the guidelines.
Rules & Requirements

Prerequisites: Instructors may require students to enroll concurrently or have completed Data C8 (COMPSCI/STAT/INFO C8) or or Data C200 (COMPSCI C200A/STAT C200C). Graduate standing and consent of instructor

Repeat rules: Course may be repeated for credit without restriction. Students may enroll in multiple sections of this course within the same semester.

Hours & Format

Fall and/or spring: 15 weeks - 1-4 hours of directed group study per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Graduate

Grading: Letter grade.

DATA 375 Professional Preparation: Teaching of Data Science 2 Units

Terms offered: Spring 2025, Fall 2024, Spring 2024

Discussion and practice of techniques for effective teaching of student-centered learning, focusing on issues most relevant to teaching assistants in data science courses. Discussion, review and development of formative and summative assessments, guidance of laboratory classes, course development, supervised practice teaching, and culturally relevant pedagogy.

Rules & Requirements

Prerequisites: Concurrent Teaching Assistant appointment required

Hours & Format

Fall and/or spring: 15 weeks - 2 hours of lecture per week

Additional Details

Subject/Course Level: Data Science, Undergraduate/Professional course for teachers or prospective teachers

Grading: Offered for satisfactory/unsatisfactory grade only.