

DATA 1010 SYLLABUS

FALL 2021

BROWN UNIVERSITY

SAMUEL S. WATSON & KARIANNE BERGEN

EMAIL sswatson@brown.edu
karianne_bergen@brown.edu

Please CC both instructors on any course-related emails.

WEBSITE <https://data1010.github.io>

CLASS MEETING Monday, Wednesday, and Friday from 10:00 to 11:50 in CIT 227 (or Zoom for remote students).

OFFICE HOURS Watson: TBD
Bergen: Mondays 3:30-5:00pm (Zoom or in-person)

COURSE DESCRIPTION An introduction to the mathematical methods of data science through a combination of computational exploration, visualization, and theory. Students will learn programming basics, topics in numerical linear algebra and scientific computing, mathematical probability (probability spaces, Bayes theorem, and the central limit theorem), statistics (point estimation, confidence intervals, hypothesis testing, maximum likelihood estimation, density estimation, bootstrapping, and cross-validation), and machine learning (regression, classification, and dimensionality reduction, including linear regression, decision trees, support vector machines, neural networks, principal component analysis, t-SNE, Bayes nets, MCMC, Bayesian methods and probabilistic programming).

TEXTBOOK The course content will have been made available as a sequence of free *Mathigon* courses at <https://mathigon.org/data-gymnasia>.

For those who would like a traditional textbook with more examples and problems, we recommend *A first course in probability* by Sheldon Ross, *All of Statistics* by Larry Wasserman, and *Pattern Recognition and Machine Learning* by Christopher Bishop.

STUDENT SATISFACTION & INCLUSION Our top priority is for you to have an excellent learning experience in this course. We intend to set clear learning objectives and equip you with the right tools to achieve them. We invite comments, criticisms, concerns, and suggestions at any time. If you perceive that you are not doing as well as you'd like, please see us right away. We can help with math/programming/stats concepts, of course, but we are also happy to help you troubleshoot your approach to studying, problem solving, etc. We will grant you the assumption that you are doing your best to learn, and we have zero interest in making you feel judged for where you are in the learning process. *You should have every expectation that you can translate sustained hard work into a high level of course success.*

Data science is a new, multidisciplinary field which benefits from the perspectives and contributions of those from diverse disciplinary and personal backgrounds. As your instructors, We are committed to fostering an inclusive and intellectual environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ...)

VIDEO LECTURES In advance of each class session, students will watch lecture videos (recorded during a previous year's run of the course), answer accompanying questions, and complete assigned Data Gymnasia sections.

CLASS SESSIONS Each class will be conducted as a guided problem solving session. We will discuss follow-up questions about the prepared videos, recap the main ideas, and solve new exercises. The format will incorporate individual work, group work, and whole-group instruction.

Class will be mediated by Prismia (for chat <https://prisima.chat>) and Zoom (for remote students).

Prismia data on student participation will be used to generate a participation score. Students who answer all or almost all of the questions asked on Prismia will receive full participation credit. Others will receive points in proportion the proportion of question answered.

Students who are not able to participate in live class sessions should request accommodation.

ASSIGNMENTS Written assignments will be due at 6 PM on each Friday. In recognition of the extra workload associated with reading, watching videos, and answering questions before class each day, the homework will involve a *small* number of computational explorations and expository exercises. These assignments will be viewed as writing exercises and assessed based on clarity.

Solutions will be submitted via Prismia.

There will be a 2-hour grace period to help you out in the event of technical difficulties. There is no score penalty, but your work will be marred with a red "late" indicator. No submissions are accepted after the grace period.

HOMEWORK POLICY There are no dropped homework assignments. Dates are coordinated with DATA 1030 and 1050, so curricular conflicts should not be an issue. In event of medical or family emergency, contact us for accommodation.

ARTICLES Students will write two in-depth Medium articles developing and explaining an extension of an idea from the course.

GRADING

Class preparation & participation	15%
Homework	35%
Exams	40%
Articles	10%

For the exams we will be using *standards-based grading*. There are 40 learning standards, and your goal will be to demonstrate mastery or proficiency on as many of them as possible. You will achieve a gold medal (mastery), a silver medal (proficiency), a bronze medal (good-faith-effort completion) for each standard as it is assessed. Every objective will be assessed on multiple exams, so you will have an opportunity to improve your medal in standards where you want to improve. Once you have achieved an exam gold medal on a given standard, you will no longer need to solve the problems keyed to that standard on future exams.

Your final exam grade will be determined by how many gold, silver, and bronze medals you accumulate. You must reach the prescribed numbers in all three categories to earn the given letter grade:

Grade	Gold	Silver (or higher)	Bronze (or higher)
A	30	34	38
B	24	30	37
C	20	28	36

PREREQUISITES	Prerequisites to this course include problem solving facility with sets and functions, basic programming skills, linear algebra, and some topics in multivariable calculus. We have prepared Data Gymnasia courses, aside from the ones we're using to develop the course content, to help you fill in or refresh any of this background.
COLLABORATION AND ACADEMIC INTEGRITY	You may collaborate on solving homework problems, but you must write your solutions entirely by yourself, and you may not do so with reference to notes taken while working in a group. Writing solutions based on notes which represent the ideas of others short-circuits the exercise and impedes your learning. On each homework submission, write a statement on the front page either listing collaborators or confirming that you did not collaborate. Using the internet to look up solutions to homework problems is not acceptable, although of course you may freely use any available resources to learn the material more generally.
COURSE-RELATED WORK EXPECTATIONS	Students will meet 2 hours per day in class (86 hours total), and readings for class will take about 1.5 hours per class day (64 hours total). Weekly written assignments and review for exams will take about 14 hours per week (168 hours total). In addition, the two Medium articles are expected to require 20 hours each.
DISABILITY SUPPORT	Please inform us if you have a disability or other condition that might require modification of these procedures. We are happy to accommodate your learning needs. You should also contact the Student and Employee Accessibility Services at 401-863-9588 or SEAS@brown.edu.