

COGS 9 | Introduction to Data Science

Meeting: Fall 2024, TuTh 9:30a-10:50a (all times Pacific)

Location: Solis 107

Instructor: Professor Bradley Voytek (<https://voyteklab.com>)

Course podcast: https://podcast.ucsd.edu/watch/fa24/cogs9_a00

Course Piazza*: <https://piazza.com/ucsd/fall2024/cogs9>

Course GitHub: <https://github.com/IntroDataSci>

*You will be able to post anonymously on Piazza; however, you will only be anonymous to your classmates. Your Instructor and TAs will be able to see who you are.

One-on-one Q&As:

| Date & Time | Location | Instructional Staff | email |
|----------------------------|-------------------------------|----------------------|--|
| *** Tuesdays 11:00a-12:00p | CSB 169 | Prof. Bradley Voytek | bvoytek@ucsd.edu |
| * Thursdays 8:15 - 9:15 am | Zoom (link) | TA: Eena Kosik | ekosik@ucsd.edu |
| ** Mondays 9:00-10:00a | Mandeville Coffee Cart | TA: Yueying Dong | yud070@ucsd.edu |
| * Thursdays 4:00-5:00p | Zoom (link) | TA: Kaushik Madhavan | kmadhavan@ucsd.edu |
| Tuesdays 3:00-4:00p | Mandeville Coffee Cart | IA: Hannah Song | hsong@ucsd.edu |
| Tuesdays 2:00-3:00p | Mom's Café | IA: Jingyi Wang | jiw109@ucsd.edu |
| Fridays 12:00-1:00p | CENTR 222 | IA: Abhay Anand | a7anand@ucsd.edu |
| Tuesdays 1:00-2:00p | PC Gameroom, sofas | IA: Shivani Kedila | skedila@ucsd.edu |

Sections

| Section | Day | Time | Location | Staff |
|--------------|-----|--------------|--|---------|
| A01 (535379) | W | 9:00-9:50a | CENTR 222 | Holly |
| A02 (535380) | W | 10:00-10:50a | CENTR 222 | Eena |
| A03 (535381) | W | 11:00-11:50a | CENTR 222 | Eena |
| A04 (535382) | * F | 12:00-12:50p | Virtual-only Zoom (link) | Kaushik |
| A05 (535385) | F | 3:00-3:50p | PETER 103 | Holly |

* Virtual-only

** Also available for virtual appointments by request

*** Available at other times, by request

COURSE OBJECTIVES

- Define terminology for core concepts in data science.
 - Learn to think critically about data, and how to approach problems with a “data-first” mindset.
 - Introduce the basics of data visualization and practice basic storytelling about data.
 - Inspect and work through problems demonstrating “p-hacking”, related to ethical data science.
 - Discuss data privacy and ethics considerations, using real-world examples.
 - Explain examples of real-world data science projects that have been pivotal for understanding aspects of human behavior, language, and society that have helped scientific progress, and business.
 - Give students first-hand experience with common pitfalls of data analyses and how to avoid them.
-

COURSE MATERIALS

- There is no textbook
 - All materials will be provided through Canvas
-

GRADING & ATTENDANCE

Grading:

| | Points | % of Total Grade |
|--------------------|--|------------------|
| (4) Assignments | 40 pts each; 160 total | 40 |
| (2) Exams | 50 pts each; 100 total | 25 |
| (5) Readings | 20 pts each (one dropped); 80 total | 20 |
| (1) Final Project | 40 total | 10 |
| (n) Guest Lectures | 20/n pts each; 20 total | 5 |

| Letter Grade | From | To |
|--------------|-------|--------|
| A+ | 97.00 | 100.00 |
| A | 93.00 | 96.99 |
| A- | 90.00 | 92.99 |
| B+ | 87.00 | 89.99 |
| B | 83.00 | 86.99 |
| B- | 80.00 | 82.99 |
| C+ | 77.00 | 79.99 |
| C | 73.00 | 76.99 |
| C- | 70.00 | 72.99 |
| D+ | 67.00 | 69.99 |
| D | 63.00 | 66.99 |
| D- | 60.00 | 62.99 |
| F | 0.00 | 59.99 |

Notes:

- Final exam date: No final exam, only final project deadline.
- There are 400 possible points to be earned in this course. To determine your final grade, you will add up all of the points for the above categories and divide your grade by 4. Your letter grade will be determined using the standard grading scale. Grades are *not* rounded up.

Semi-synchronous lecture and synchronous discussion sections

Lectures will be given live during the normal class time, however they will also be recorded for later viewing.

Lecture attendance

Our goal is to make lecture and discussion section worth your while to attend. However, you have the choice to *not* attend lectures *or* discussions. That said, attendance is required on guest lecture days, and that attendance will constitute your Guest Lecture attendance grade.

Grades

Grades for assignments and exams will be released on Canvas approximately a week after the submission date.

Everyone in the class begins the quarter **two bonus points** that amount to a 0.5% boost to your overall, final grade. These are "fudge" points to get folks who are *very close* to the next grade level up and over the mark at the end of the quarter. ***It is possible to have these bonus points taken away.*** Any student that messages me, the TAs, or IAs with requests asking for free points, or to "bump up" their grade "just a little bit" because you're "so close", or to get another chance on an exam, assignment, or reading that wasn't already cleared with me for a verifiable excuse will be ignored.

Furthermore, *the free bonus points that I gave will be removed from any student who makes these kinds of grade-begging requests.* These kinds of requests are unfair to other students who do not make such unreasonable requests, and it would be very unfair to everyone else who wouldn't get the same special treatment.

Note that questions to get clarification, or regarding possible grading errors or mistakes, or regrade requests are *not unreasonable nor inequitable*, and will not be penalized! I welcome these kinds of questions.

Check your grades and get in touch if any are missing or if you think there is a problem and I will work quickly to correct any errors.

Assignment regrades

We will work hard to grade everyone fairly and return assignments quickly. But we know you also work hard and want you to receive the grade you've earned. Occasionally, grading mistakes do happen, and it's important to us to correct them.

If you think there is a mistake in your grade for an assignment, submit a regrade request on Gradescope within 72 hours of receipt of the grade. This request should include evidence of why you think your answer was correct (i.e., a specific reference to something said in lecture) and should point to the specific part of the assignment for us to reconsider.

Discussion sections

Discussion sections will be used to review content from lectures, discuss readings, and guide your assignments. You should be signed up for a section for which you can attend. However, if you are unable to attend the section for which you are signed up, you are free to attend a different section any given week than the one in which you're assigned.

COURSE TOPICS & ASSIGNMENTS

This class is *a survey course* intended to get you all excited about becoming data scientists! Data are everywhere and they're being used in tried-and-true — as well as in new, awesome, and creative — ways. This course will introduce you to topics in data science, discuss what it means to be a data scientist, and get you on your way to thinking like a data scientist. To see what topics will be introduced in this course, see the calendar at the end.

Assignments

There are four assignments. Assignments will focus on applying the concepts covered in lecture and ensuring you're on the right track for your final project.

Assignments will be posted on the Thursdays of the week before they're due, and you have until *Friday at 11:59 PM* of the following week (so a little over 8 days) to complete each assignment. Assignments will be released on Canvas and submitted on Gradescope. *Assignments will always be due Fridays at 11:59 PM.* Assignments 1, 3, and 4 are submitted individually. The second assignment (and your final project) will be turned in *as a group*. You will receive feedback along with a grade within a week from the due date. *Feedback from A2 should be incorporated into your final project.*

It is your responsibility to ensure your assignments are submitted on time, without errors, in the correct format. This means you need to immediately check to make sure your submission uploaded correctly and without errors. If you accidentally submit the wrong thing, or if there are submission errors, or you're having internet troubles, you can try again before the deadline. If errors continue, get in touch with Prof. Voytek right away (bvoytek@ucsd.edu), and attach your assignment to that email to ensure that you get credit for submitting it on time. Note that this might mean you have to take photos of your assignment pages and use a cell phone to submit your assignment on time.

Late assignments earn fractional credit (75% within one week late; no late assignments accepted after one week).

Final project

The final project is a report on how you would handle a complicated data science project. It's essentially a culmination of the four assignments all tied together in a nice report. This report will include all the nitty gritty, whys, and hows of the analysis you have chosen. You'll specify your data science question, find data that could be used to answer the question, and describe the analysis you *would* carry out to answer your question of interest. You WON'T have to actually perform the analysis to answer the question; you'll just write about it. This will be turned in as a PDF.

You can choose your final project groups of 4-5 people. If you do not have a group, Professor Voytek will assign one. There will be time to work on and discuss your second assignment and projects in section, so we recommend (but do not require) you form groups within the section you plan to attend.

Exams

There will be two exams covering material in lecture (including guest lectures!) and the readings discussed in section. They will be closed notes; you may not use any outside resources. Exams will be primarily multiple choice with a few short answer questions. See schedule below for in-class exam dates.

If you know you will miss an exam, contact Prof. Voytek right away. It is possible to take a makeup exam, but only if adequate notice is given so that Prof. Voytek can arrange your testing with the Triton Testing Center (TTC). See the TTC section below for more details.

Readings

There will be five weeks where readings are assigned. Just like the Assignments, Readings will be posted on the Thursdays of the week before they're due, and you have until *Friday at 11:59 PM* of the following week (so a little over 8 days) to complete the reading quiz assignment *on Canvas*. You are not timed. **You *must* click submit** to submit your reading quizzes. Your **most recent submission** will be graded – **you only get three attempts** (with unlimited attempts it would be possible to figure out all the answers simply through trial-and-error). If you fail to finish and submit your quiz answers before the deadline, it will not be graded **and you will receive an automatic zero**. Your lowest reading quiz score will be dropped.

No late credit will be given if Reading quiz assignments are submitted after the due date. You will be able to view the correct answers for one week after the due date, beginning at 12:01a a few minutes after the due date.

Planned readings:

All of the below readings are available on the class GitHub page at:

<https://github.com/IntroDataSci/Readings>

- R1: Donoho D, *50 Years of Data Science*
- R2: Keyes O, Hutson J, & Durbin M, *A Mulching Proposal*
- R3: Wickham H, *Tidy Data*
- R3: Woo K & Broman K, *Data in Spreadsheets*
- R4: Peck, E, Ayuso S, & El-Etr O, *Data Is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania*
- R5: Angwin J, Larson J, Mattu S & Kirchner L, *Machine Bias*

OTHER GOOD STUFF

Class conduct

In all interactions in this class, you are expected to be respectful. This includes following the [UC San Diego principles of community](https://ucsd.edu/about/principles.html): <https://ucsd.edu/about/principles.html>

This class will be a welcoming, inclusive, and harassment-free experience for everyone, regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, ethnicity, religion (or lack thereof), political beliefs/leanings, or technology choices.

At all times, you should be considerate and respectful. Always refrain from demeaning, discriminatory, or harassing behavior and speech. Last of all, take care of each other.

If you have a concern, please speak with Dr. Voytek, your TAs, or IAs. If you are uncomfortable doing so, that's ok! The [OPHD](#) (Office for the Prevention of Sexual Harassment and Discrimination) and [CARE](#) (confidential advocacy and education office for sexual violence and gender-based violence) are wonderful resources on campus.

OPHD: <https://blink.ucsd.edu/HR/policies/sexual/OPHD.html>

CARE: <https://care.ucsd.edu/>

Academic integrity

Don't cheat.

You are encouraged to (and at times will have to) work together and help one another. However, you are personally responsible for the work you submit. For assignments, it is also your responsibility to ensure you understand everything your group has submitted and to make sure the correct file has been uploaded, that the upload is uncorrupted, and that it renders correctly. Projects may include ideas and code from other sources—but these other sources must be documented with clear attribution. Please review academic integrity policies at: <https://academicintegrity.ucsd.edu/>

We anticipate you all doing well in this course; however, if you are feeling lost or overwhelmed, that's ok! Should that occur, we recommend: (1) attending discussions and leveraging the time there, (2) attending weekly one-on-one hours with Dr. Voytek and the course TAs and IAs and/or, (3) browsing Piazza.

Cheating and plagiarism have been and will be strongly penalized. If, for whatever reason, Canvas is down or something else prohibits you from being able to turn in an assignment on time, immediately contact me by emailing your assignment by email (bvoytek@ucsd.edu), or else it will be graded as late.

Accommodations and disability access

Students requesting accommodations must provide a current Authorization for Accommodation (AFA) letter. These letters are issued by the Office for Students with Disabilities (OSD), which is in *Pepper Canyon Hall Suite 300*. Please contact Dr. Voytek privately to arrange accommodations.

Contacting the OSD can help you further:

858.534.4382 (phone)

osd@ucsd.edu (email)

<http://disabilities.ucsd.edu>

Satisfactory Academic Progress (SAP)

Satisfactory Academic Progress (SAP) refers to the academic standards students must maintain to remain eligible for federal, state, and institutional financial aid. If you are receiving financial aid, please ensure you review the SAP requirements and the appeals process: <https://fas.ucsd.edu/forms-and-resources/sap/index.html>

Triton Testing Center (TTC)

The Triton Testing Center (TTC) is in Pepper Canyon Hall room 364 and co-located with the Academic Integrity Office: <https://tritontesting.ucsd.edu>

The TTC is used for students who are approved to have extra time on exams, who will take their exams on the scheduled exam days in the TTC.

The TTC is also used for rare cases where students have arranged, well ahead of time, with Prof. Voytek to take an exam on a different day due to a verified, excused absence.

How to get your question(s) answered and/or provide feedback

It's *great* that we have so many ways to communicate, but it can get tricky to figure out who to contact or where your question belongs or when to expect a response. These guidelines are to help you get your question answered as quickly as possible *and* to ensure that we're able to get to everyone's questions.

That said, to ensure that we're respecting their time, TAs and IAs have been instructed they're only obligated to answer questions between normal working hours (M-F 9am-5pm). However, I *know* that's not when you may be doing your work. So, please feel free to post whenever is best for you while knowing that if you post late at night or on a weekend, you may not get a response until the next weekday. As such, do your best not to wait until the last minute to ask a question.

Finally...

If you have...

- **questions about course content:** these are awesome! We want everyone to see them and have their questions answered too....so post these to Piazza!
- **questions about course logistics:** first, check the syllabus. If you can't find the answer, ask a classmate. If still unsure, post on Piazza.
- **questions about a grade:** If for an assignment, submit a regrade request on Gradescope. For anything else, post as a question on Piazza, address it to "Instructors," and select the folder "regrades"
- **something super cool to share related to class:** feel free to email Dr. Voytek (bvoytek@ucsd.edu) or come to one-on-one hours. Be sure to include COGS9 in the email subject line and your full name in your message.
- **something you want to talk about in-depth:** meet during weekly one-on-one hours or schedule a time to meet by email. Be sure to include COGS9 in the email subject line. (bvoytek@ucsd.edu).

Schedule

Exact guest lecture dates subject to change, but assignment and exam dates are stable

| Week | Date | Title | Schedule |
|-------------|------------|--|-------------------------------|
| 0 | Thu Sep 26 | 01 - Introduction | class lecture |
| 1 | Tue Oct 01 | 02 - What is Data Science? | class lecture |
| 1 | Thu Oct 03 | 03 - Privacy and ethics | class lecture |
| 1 | Fri Oct 04 | -- | -- |
| 2 | Tue Oct 08 | 04 - Ethics (cont.) | class lecture |
| 2 | Thu Oct 10 | 05 - Data visualization | class lecture |
| 2 | Fri Oct 11 | -- | R1: Data Science quiz |
| 3 | Tue Oct 15 | 06 - Data and information | class lecture |
| 3 | Thu Oct 17 | 07 - Data and storytelling | class lecture |
| 3 | Fri Oct 18 | -- | A1: Data Viz due |
| 4 | Tue Oct 22 | 08 - Algorithms and crowds | class lecture |
| 4 | Thu Oct 24 | <i>Guest lecture: Jenni Tremaine, Ph.D. Director, Data Insights: Sony</i> | guest lecture 1 |
| 4 | Fri Oct 25 | -- | R2: Data Ethics quiz |
| 5 | Tue Oct 29 | EXAM 1 | exam 1 |
| 5 | Thu Oct 31 | 09 - Data wrangling | class lecture |
| 5 | Fri Nov 01 | -- | A2: Final Outline due |
| 6 | Tue Nov 05 | <i>Guest lecture: Emily Kubicek, Ph.D. Manager, Data Strategy: Disney</i> | guest lecture 2 |
| 6 | Thu Nov 07 | 10 - Hypothesis testing and EDA | class lecture |
| 6 | Fri Nov 08 | -- | R3: Tidy Data quiz |
| 7 | Tue Nov 12 | 11 - Prog and version control | class lecture |
| 7 | Thu Nov 14 | 12 - Stats and probability | class lecture |
| 7 | Fri Nov 15 | -- | A3: p-values due |
| 8 | Tue Nov 19 | <i>Guest lecture: Duncan Watson-Parris, Ph.D. Assistant Professor: UC San Diego Data Science</i> | guest lecture 3 |
| 8 | Thu Nov 21 | 13 - Inference and machine learning | class lecture (Eena Kosik) |
| 8 | Fri Nov 22 | -- | R4: Data Viz quiz |
| 9 - VIRTUAL | Tue Nov 26 | Pre-recorded 14 - Geospatial analysis and text-mining | class lecture (pre-recorded) |
| 9 | Thu Nov 28 | <i>Thanksgiving holiday: no class!</i> | holiday: no class! |
| 9 | Fri Nov 29 | -- | A4: ML due |
| 10 | Tue Dec 03 | EXAM 2 | exam 2 |
| 10 | Thu Dec 05 | 15 - The Future of Data Science | class lecture |
| 10 | Fri Dec 06 | -- | R5: Algorithms quiz |
| finals | Thu Dec 12 | Final Project Deadline (DO NOT SHOW UP) | Final Project deadline |