

COGS 108: Data Science in Practice
Spring 2017 MWF 11:00-11:80
Peterson Hall 108

Instructor: Bradley Voytek (bvoytek@ucsd.edu)

Teaching Assistants:

Thomas Donoghue <tdonoghue@ucsd.edu>

Office hours: Mondays 10:00-10:50 or by appointment (CSB 169)

TA Office hours:

TBD

Final exam: Friday, 2017 June 16 11:30-14:30 **Note this may be subject to change!**

Grading: Six assignments (10% each) + Final project (30%) + Participation (10%)

***Critical note for Spring 2017:** Note this is the very first offering of this course in the history of UCSD, and this is the first time I've tried teaching a class this way. This will be especially tricky given how large the expected enrollment is (up to 300 students!), which is significantly larger than my first trial run of COGS 9 back in Fall 2014 (21 students). Thus, you are all guinea pigs for teaching a class structured this way.*

Additionally, the number of assignments stated above (6) may end up being fewer as we tweak the structure of the course throughout the quarter. In such cases, the benefit will always go to the students—we won't penalize you if we alter the course on your midway through the quarter.

Ultimately what you learn (and don't learn) in this course will help me shape future versions of COGS 108 to be better. But for now, caveat emptor.

Course Background and Overview: Who cares about data? We all should! In COGS 9 (Intro to Data Science) you learned why data and data science are important. The goal of that class was to give you an appreciation for *what can be done with data and where data can even lead you astray*.

In this course we take the educational view that “sometimes the best way to learn something is by doing it,” or, more importantly as author Neil Gaiman says, “sometimes the best way to learn something is by doing it wrong and looking at what you did.”

In this course, we aim to teach you the joys and frustrations of the *practice* of data science. We won't have you dive deeply into the methods or proofs of machine learning, clustering, etc. but we *will* have you try and implement various methods. At times we will even ask you to implement techniques we explicitly *haven't* taught you yet, as there may be times in your data science career you'll be asked to do so. We want you to build a technical toolkit as well as a skeptical mindset and “data intuition”—that nebulous sense that something in a dataset is “off”.

Topics Covered:

- . Python
- . The shape of data
- . Finding and extracting data
- . Data-mining
- . Text-mining and analytics
- . Geospatial analysis
- . Clustering and classification
- . Graph analysis
- . Data visualization and storytelling
- . Data privacy

Grades: There will be six assignments worth 10% each and a final project worth 30%. 10% of your grade is for class participation (attendance taken during guest lectures). Late assignments earn fractional credit (75% within one week late; 50% otherwise).

Final Project: The final project *will be group based, even if you don't want to work with groups. No exceptions.* The reality of working in technical fields is that you will need to work with others. Thus you need to learn how to communicate with groups effectively, handle disparities in knowledge and skill sets, manage time, and organize.

This year's final project will be *extra special* as it will be judged by a guest star, so I strongly encourage you to put some serious thought and effort into it. You will be briefed well ahead of time.

Piazza Discussion Board: This term we will be using the Piazza website for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. Please assist your fellow classmates, but be sure to do so in a manner that is consistent with the academic integrity policy (discuss the issue, but do not write the code needed or the answer). In addition to encouraging peer teaching and learning, engagement on Piazza can be a deciding factor for borderline grading cases (e.g., B+ vs. A-) since active Piazza citizens can earn 1% extra credit.

Find our class piazza page at: <https://piazza.com/ucsd/spring2017/cogs108/home>

Readings: There are no reading materials to be purchased. We will provide all necessary materials online on TritonEd. You will be asked to read and occasionally watch video tutorials online before attending class or for further learning afterwards.

Additional Readings “for fun”:

- . Donoho D, *50 Years of Data Science*
- . Tukey JW, *Exploratory Data Analysis*
- . Buchanan M, *Depths of Learning, Nature Physics* 2015
- . Krzywinski M & Cairo A, *Points of view: Storytelling, Nature Methods* 2013