Fundamentals of Statistical Data Science

Data & Web Technologies for Statistical Analysis

The major themes of the class are:

- Collecting and cleaning data from diverse, non-tabular sources.
- Creating and evaluating interactive visualizations for data.
- Scientific computing with the Python programming language.

Tentative Schedule

For the first 3 weeks we'll learn about Python libraries for scientific computing. Then we'll spend 4 weeks learning strategies to collect data from a variety of sources. In weeks 8 and 9, we'll learn about how to create interactive visualizations and websites to host them. Finally, in week 10, each project group will present their preliminary work to the class.

Week	Topics
1	Python; Jupyter; Version Control
2	Numpy; Pandas
3	Static Visualization; Matplotlib
4	Web APIs; JSON format
5	Web Scraping; XML format
6	Natural Language Processing
7	Databases; SQL
8	Interactive Visualization
9	Interactive Visualization
10	In-class Presentations

Contacts

Name	@ucdavis.edu	Role
Nick Ulle	naulle	Instructor
Xiaodong Wang	xidwang	Lead TA
Shan Jiang	shjiang	TA
Chunhan Li	boxli	Reader

We **use email only for private matters** (grading, emergencies, etc.). Please do not email us about class material – post on Piazza instead.

Piazza

Piazza (www.piazza.com) is the class' online forum. You should have already received an email inviting you to Piazza. If you did not, or have any problems accessing Piazza, please email me or a TA as soon as possible. Note: the Piazza access code is `"sta141b"`.

Announcements, references, office hours, and other class information will be posted to Piazza. Canvas will only be used for grading.

Post your questions about class material on Piazza. Anyone in the class can answer your question, so you're likely to get an answer quickly. When you use Piazza:

- Be polite and respectful to others.
- Search before you post. Your question may have already been asked and answered.
- When you post a question, explain the context and give an example of what you mean.

Grading

Participation (10%) In order to do well in this class, you must be an active participant. Active participation means asking or answering questions. Attendance does not count as active participation.

You can participate in lecture, discussion, office hours, or Piazza. You are not required to participate in all of these, but it may help if you are trying to get an A+.

In order to make the participation grade transparent, you will log your in-person participation through a Google Form. A link to the form is posted on the class website. Each time you participate in class or office hours, use the form to record it (within 1 week of when you participated). You do not need to log participation on Piazza.

Participation is the best way to the best way to get help and verify that you understand the material.

Assignments (50%) There will be 5 assignments.

Short assignments will be graded for correctness. Long assignments will be graded with a rubric that measures the quality of your writing, graphics, and code. The rubric is available on Canvas.

Assignment grades are never dropped. If you can't turn in an assignment on time and have a legitimate excuse (such as medical emergency), email me as soon as possible to arrange an extension or alternative.

Final Project (30%) For the final project, you will work in teams of 3-4 people. Groups of 4 will be held to a higher standard than groups of 3.

More details about the project will be released near the beginning of the quarter. The project will be due in finals week.

Final Presentation (10%) In week 10, each group will present preliminary results from their project to the class. Every group member must speak during the presentation.

You are also required attend and provide feedback on presentations from some of the other teams.

More details about presentations will be released in week 8.

Academic Honesty

Professional programmers talk to their coworkers and use references to help solve programming problems, so I encourage you to:

- Discuss the problems with your classmates.
- Search for references online and in books.
- Adapt short pieces of code ($\leq$ 10 lines) you find on Piazza or online. When you do this, you must **cite the source**. For Piazza, cite the post number. For other sources, cite the title, author, and URL.

That said, **all writing and graphics must be your own work. At least 75% your code must be your own work.** If you're unsure whether something is okay, please ask!

The university code of academic conduct (sja.ucdavis.edu/files/cac.pdf) applies to this class. In particular:

> Students are responsible to know what constitutes cheating. Ignorance is not an excuse.

Waitlist

This class has a long waitlist. If you decide you don't want to take the class, please drop immediately to make room for others. The drop deadline is October 9th.

The Statistics Department PTA policy is online at statistics.ucdavis.edu/courses/pta-policy. If you have any questions adding or dropping, contact Kim McMullen (stat-advising@ucdavis.edu).