Introduction to Data Exploration with R

STS 98

This is a gentle course on programming, basic statistics, and working with data.

We'll begin with an introduction to the R programming language, and learn how to use it for basic tasks such as loading a data set to inspect its structure. From there, we'll discuss what kinds of data are common in practice, what it means for something to be "random," and strategies for summarizing data. Visualization is a major component of exploratory data analysis, so we'll also spend several weeks learning what qualities make visualizations effective, using R's graphics and mapping libraries to practice these skills. Finally, we'll examine what makes data "messy" and how to clean it up.

Here's the schedule of topics by week:

1	git and R	6	Midterm
2	R Navigation and Graphics	7	Messy Data
3	Basic Statistics	8	Maps and Data Shape
4	Grouped Data	9	Advanced Graphics
5	Customized Graphics / Review	10	Advanced Topics / Review

Contacts & Meeting Times

A small army has been mobilized to help you learn:

	Name	@ucdavis.edu	Office Hours		
AI	Nick Ulle	naulle	F	1–3	Shields 360
TA	Shaheen Amirebrahimi	samirebrahimi	W	2–4	Shields 360
	Clark Fitzgerald	rcfitzgerald	T	—	Shields 360
	Shaun Geer	sgeer	M	5–7	Shields 360
Intern	Nick Alonzo	nalonzo	—		
	Somayh Saleh	somsaleh	—		

Lectures are TTh 3:10–4:30pm in Wellman 26. Discussions are in Shields 360, at the following times:

	am		pm	
Monday	10:00	11:00	2:10	4:10
Wednesday			1:10	

Grading

Assignments (50%) There will be two short homeworks and three projects. They will be graded according to the class rubric, which emphasizes writing and graphics. Writing is an important step in the data exploration process, to reflect on your results and to present them to others. Use graphics to motivate and support your writing. Graphics should never appear alone, without a written interpretation, and should not be redundant or irrelevant. It's your responsibility to be familiar with the rubric.

Assignments will be submitted using git. See the lecture notes for details.

Exams (30%) There will be two exams, both in Wellman 26:

- A midterm on May 3

- A final on June 6 from 10:30am – 12:30pm

They will have written-answer and multiple choice questions based on material from class. You will not be asked to write code on the exams, so you don't need to memorize the details every R function. However, you may be asked to read code and discuss what it does. You may also be asked to reason about, interpret, or critique results for a given data set. There will be no make-up exams unless you have proof of a medical emergency. If you require SDC accommodations, let me know before the end of week 2.

Participation (20%) It's critical that you participate in class and on Piazza. Asking and answering questions is the best way to avoid falling behind and verify your understanding of the material. We'll also have peer reviews at the end of some assignments. These are an opportunity to get additional feedback on your work and learn new things from your classmates.

Piazza (`www.piazza.com`) is the class' online forum. All announcements will be posted to Piazza. Plan to visit daily. You should also **post questions about class material on Piazza** rather than emailing them to me or a TA (emails about grading or private matters are okay). Typically you'll get an answer much faster, because everyone in the class has the opportunity to answer your question. Posting to Piazza also allows others to benefit from your questions and answers.

Please be polite and respectful to others when using Piazza. Before posting, search Piazza to see if your question has already been asked, because duplicate postings make it more difficult to find answers. When you do post a question, explain the context and, if possible, give a small example of what you mean. Otherwise, you may be asked to clarify your question before receiving an answer.

Academic Honesty

Be honest at all times! All writing, graphics, and output must be your own original work. However, for assignments you may:

- Discuss the problems with your classmates.

- Search for relevant information online.

- Use short pieces of code (≤ 20 lines) you find on Piazza or online, as long as you cite the source. This **does not include** your classmates' GitHub repositories.

The university code of academic conduct (`sja.ucdavis.edu/cac.html`) applies to this class. One of the many things listed there is:

> Know the rules - ignorance is no defense. Those who violate campus rules regarding academic misconduct are subject to disciplinary sanctions, including suspension and dismissal.

If you're unsure whether something is allowed, please ask!