A0: Sections & Assignments Overview

Overview & Goals

This is the hands on section of the course. The *practice* of data science largely boils down to writing code, and so, although this is not a programming class *per se*, the majority of the section and assignments will center around using Python to do data science.

Not everything you are expected to do will be explicitly mapped out, step by step. Not only would that level of direction be prohibitively long to prepare, it would ultimately be dishonest and unhelpful - data science is ultimately about figuring things out. If clear step-by-step instructions could be written down that always worked, data science would be automated and this class would be moot. The human level of figuring things out when it is not not entirely clear how to proceed is the actual practice here. You are going to have to be okay with trying things out that don't work, with getting stuck, with not quite knowing what is going on all the time. That's the job.

What we don't want you to do is get or feel totally lost. If you do find yourself wandering through a problem that you really have no idea about, 17 layers into stack overflow, Jupyter running at a sluggish pace due to an uncountable number of browser tabs - pause, and ask someone for help. Most of the time a quick intervention from someone who can identify the problem, give you an overview as to what is happening, and suggest where to look can save you tons of time (and frustration).

As instructors, we are here to help you, but keep in mind that the # students >>> # instructors. Thus we encourage you to help each other. If you are unsure about something, ask your fellow students. Everyone has different levels of experience in different domains. This is how {data science, programming, science, research, industry} works - teams of people with different backgrounds and expertise sharing knowledge and ideas and working toward a common goal. The best way to learn something, or to check if you really know it, is to teach it. So if you do think you know something, offer help! Sections and office hours are meant to be collaborative experiences.

Ask questions on piazza, and come to office hours. Please be aware that we may not know the answer to every question, at least not right away. Data science / programming is broad and dynamic and no single person can be on top of everything. If nothing else, we will try and give you pointers on what might be happening, and whether it is likely to be tractable, or potentially very messy and worth circumventing. Knowing which problems you can solve, and which you can't (at least in the context of a given project) is an important skill.

The goal of this class is not to teach you (all of) data science - we cannot possibly do that in 10 weeks. The goal is to give you a meaningful introduction and hands-on experience of what data science is, to teach you how to continue to learn data science, and to give you the desire and skills needed to do so. There are an amazing amount of resources out there for this material. The difficulty, as a newcomer, is to figure out what it is you're looking for, and how to find it - that is, figuring out what you don't know and what you need to learn. Technical experts do not know all the answers, they just know where to find the answers. The goal of these sections and assignments is to show you what is available, so you know where to look if you want to keep going in data science.

Administrative Details

Sections Details

There are 7 sections:
- Monday @ 9 am, CSB 004   - Tom
- Monday @ 1pm, CSB 004   - Tom
- Wednesday @ 9 am, CSB 004   - Larry
- Wednesday @ 10 am, CSB 004   - Larry
- Friday @ 10 am, CENTR 222   - Harshita
- Friday @ 12 pm, CSB 004   - Harshita
- Friday @ 3 pm, CSB 004   - Harshita

Office Hour Details

- Tom: 10-10:50 am & 2-2:50 pm, on Mondays
- Larry: 1-1:50 pm on Tuesdays & Thursdays
- Harshita: 2-2:50 pm & 4-4:50 pm on Fridays
- Tony: 2-3 pm on Tuesdays & Thursdays

Contact Details

If your question/request is general, please post on Piazza. If it is about grading, please contact us through the course e-mail (COGS108 {at} gmail.com). If you have a question specific for a TA/IA, use the following emails:

TA: Tom Donoghue   tdonoghue {at} ucsd.edu
TA: Larry Muhlstein   lmuhlstein {at} ucsd.edu
TA: Harshita Mangal   hmangal {at} eng.ucsd.edu
IA: Tony Chan   tcc004 {at} ucsd.edu

Regardless of who you contact please **include your student number**. This is how we are keeping track of students / assignments.

Supported Tools

Officially, we will be using, assignments will require, and we will support the use of:
- python3 with the anaconda platform
- Jupyter Notebooks
- git & Github (optionally using the SourceTree GUI)

The assignments must be completed using these tools. You are welcome to explore other tools as you explore these topics, and to use different tools for the project. Note however, that we offer no guarantees that we can help with other languages / modules / tools, etc.

Section Topics & Hands-On Materials

Section topics will follow the topics outlined for the lectures. For each lecture topic outlined in the syllabus, there will be a Jupyter notebook covering that topic, with a basic demonstration. We may not cover all the section materials explicitly in section, we will prioritize covering material that is needed for the assignments - that is, sections will serve as hands-on sessions targeted at fulfilling the assignments.

The section materials are not intended, or written, as tutorials. Instead they are more like an index, or a map. For each topic, we aim to give a cursory overview and simple demonstration of what the topic under investigation is, and guide you to bigger and better resources to really dive into it. You are not expected to, and will not be able to, follow every link for every topic - pick the ones you are most interested in and/or the most helpful ones to get you unstuck for the assignments and/or project.

Section Attendance & Switching

Section attendance is not strictly mandatory but is recommended. Sections will be used for hands-on tutorials and overviews of the assignments - they are the best way to get detailed instructions on how to do the assignments. You should make sure you are enrolled in a section that you can attend, and plan to attend that section each week. If something comes up, and you have a conflict, you may instead attend another section. Each section, in a given week, will cover the same material.

Assignments

Assignments will be done in Jupyter Notebooks. They will be released on Github (https://github.com/COGS108) and submitted to TritonED.

Assignment Schedule (UPDATED)

The (tentative) schedule for assignments is as follows:

Name	Due Date
A1 - Set Up & Github (10%)	11:59 pm, Monday, April 17th (Week 3)
A2 - Data Exploration (10%)	11:59 pm, Tuesday, April 25th (Week 4)
A3 - Data Privacy (10%)	11:59 pm, Monday, May 8th (Week 6)
A4 - Project Proposal (10%)	11:59 pm, Monday, May 15th (Week 7)
A5 - Data Analysis & ML (15%)	11:59 pm, Monday, May 22nd (Week 8)

Note: The original plan was 6 assignments, worth 10% each, to a total of 60% of your total grade. With the update to the schedule, and dropping the last assignment, A5 will now be worth 15% (and be slightly larger) while the other 5% will be free points, maintaining a total of 60% of your grade coming from assignments.

Naming Conventions

We will be tracking assignments and files using your unique student ID. **You must follow the naming specifications we outline in the assignments.** This is necessary due to the size of the class - we have many students with the same first and/or last name. Assignments are auto-graded and these scripts may fail with improperly named files. Triple check your submitted files meet the specified naming conventions. If you fail to do so, it is likely your submission won't make it through the grading procedure and you will not receive a grade.

Using Jupyter Notebooks for Class Assignments

We will be using a system that allows for automatically grading notebook submissions. Notebooks will be released with step-by-step instructions on what code to enter. Follow these instructions for working with these notebooks:
- Whenever you see '# YOUR CODE HERE', replace it with code to answer the question
    - Also, remove the 'raise' line, or the notebook will raise an error
- Do not edit or delete and cell that has 'assert' lines in it
    - These lines are used to check your code. Editing them will be flagged as attempted cheating, and they will be reset to the original versions before grading
- You can add new cells, and write extra code, as long as you follow what is written above
- Your grade is partly graded on the public tests (the 'asserts') that are released with the assignment, and partly on a hidden set of tests.

Assignment Questions & Using Piazza

Please use Piazza for all general questions - for example, if you find a question unclear, or are totally lost and need some direction. When answering questions on Piazza, you may answer with suggestions on what topics / ideas / lectures to look into, and/or vague pseudocode but do **not** provide code that answers assignment questions. For questions that are tangential or unrelated to answering assignments, you may post minimal code segments. If you have specific questions about your assignment (for example, you are missing a grade), email the class email (COGS108 {at} gmail.com)

Grades

Grades will be released on TritonED a week after the submission date. We will try to send out automated alerts if we do not receive a submission and/or if it fails to be processed, but ultimately it is your responsibility to check your grades and get in touch if any are missing or you think there is a problem.

Assignment Regrades & Solutions

After grading is complete, we will release the assignment solutions at the same time that we release the grades (1 week after the submission deadline). These solutions will be our solutions and the full test-suite that was used for grading. Note that there may be multiple possible solutions, and variations may receive full credit, provided they pass the test criteria that are specified in the question set up. If you think there is a mistake or ambiguity (for example, your different solution meets the question specifications, but fails on an unexpected test) email the course email (COGS108 {at} gmail.com), and we will look into it.

Late Submissions

You may submit late assignments, up until assignment solutions are posted, but at a 25% penalty.