CS 100: Data Fluency for All

Professor Amy Greenwald
Mondays, Wednesdays, and Fridays at 1pm in Salomon Center 003

Course Description and Goals

This course will introduce students to a variety of statistical and computational techniques that data scientists use to tell stories. The subject matter for such stories might range from the American slave trade to the 2016 presidential election. As an example of the former, data scientists have designed powerful visualizations demonstrating the growth and demise of the slave trade over time, highlighting the extent of the human displacement. For the latter, one could imagine mining Twitter feeds to measure the public’s relative interest in various presidential candidates, thereby predicting a winner.

Data fluency can be understood to encapsulate both data literacy and data communication. Data literacy includes the basics of statistics and machine learning. Data communication relies heavily on principles of design. Students will be taught to apply statistical, machine learning algorithms (clustering, regression, and classification) to data sets in order to extract meaningful information from them. They will also be taught elements of design, and to use visualization tools to graphically display potentially complex relationships in a comprehensible way.

Syllabus

0. What is Data Science?
1. Introduction to Spreadsheets
2. Measures of Central Tendency
3. Exploratory Data Analysis
4. Visualizing Qualitative vs. Quantitative Data
5. Introduction to R
6. Clustering
7. Measures of Dispersion
8. Central Limit Theorem
9. Regression
10. Classification

11. Tools for working with Text

12. Visualizing Structured Data (Maps and Networks)

Course Format

CSCI 0100 lectures are held on Mondays and Wednesday at 1pm. Students are expected to attend all lectures; attendance is taken regularly with iClickers. For future reference, all lecture slides will be posted on the course web page, after class. There are also required weekly readings that reinforce the lecture materials.

In addition to weekly lectures, there are weekly two-hour studio sessions, which offer students hands-on environments in which to practice the techniques they are taught in lecture. Students are required to register for and attend one studio per week. Four studio sections will be offered, likely on Wednesdays and Thursdays.

On Fridays at 1pm, the TAs will run “section”. Attendance at these sections is optional, but encouraged. TAs will briefly review the week’s lecture and studio materials, as well as ongoing and recently-submitted homework assignments, after which they will open up the floor for questions. These sections are a great place to gain insights into how to complete the bi-weekly homework assignments.

There are no exams in CSCI 0100. Students are evaluated based on their performance in studio, five bi-weekly homework assignments, and a final project. The final project involves writing as well as programming, and students are assessed on both of these dimensions (and others, like creativity).

Students should expect to spend two hours per week in lecture, two hours per week in studio, and one hour per week in section. In addition to these five hours of instruction, there will be bi-weekly homework assignments (5-10 hours over the course of two weeks), and optional short readings every week (1-2 hours). The final project is open-ended, but 25 hours, over the course of four weeks, should suffice to produce adequate work. In sum, students should expect to spend 10-12 hours per week working on this course.

Grading

Grading rubrics in CS 100 are developed by the professor in conjunction with undergraduate TAs. TAs then grade all assignments, although the professor assigns final grades and reviews assignments as necessary (e.g., in all borderline cases). Grade complaints on individual assignments should be addressed to the relevant TA, but final grade complaints can be addressed to the professor. The grading breakdown is as follows:
Assignment Percentage

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>5%</td>
</tr>
<tr>
<td>Studios</td>
<td>25%</td>
</tr>
<tr>
<td>Homeworks</td>
<td>35%</td>
</tr>
<tr>
<td>Project</td>
<td>35%</td>
</tr>
</tbody>
</table>

### Late Policy

For assignments that are to be turned in electronically, students will be granted three free late days, which can be applied, as needed, over the course of the semester. In the unfortunate situation that these three free late days are all used up, late day penalties will apply: -10% within 24 hours, and -25% within 48. No assignments will be accepted electronically more than 48 hours beyond their due date.

For assignments that are to be graded interactively (meaning students have a set time at which they will be meeting a TA), the following late penalties always apply: if the student is late by 10 minutes or less, -10%; 10 to 20 minutes, -20%; more than 20 minutes counts as a “no show”, for which the penalty is -50%. This same penalty schedule applies recursively to rescheduled interactive gradings following a no show. Last-minute email requests to reschedule interactive gradings must be sent to the relevant grader(s) and to the head TAs at least 2 hours prior to the scheduled meeting time to avoid any penalties.

For group projects that are graded interactively, if some members show up for the grading session while others do not, the grading will proceed, and those who do not appear will receive a grade of 0 for that portion of the project, while those who appear late will be penalized according to the aforementioned penalty schedule.

Extensions may be granted by the professor in extreme circumstances. If you are ill, please do not ask for an extension without a note from health services.

### Collaboration Policy

Students are encouraged to collaborate with their peers in CSCI 0100. Studios are pair-programmed, with each student finding a different partner each week.

When working on homework assignments, students may consult one another; but students are required to list the names of all students with whom they discussed an assignment on their submitted work. Unnatural similarities among students’ submissions with other students whose names are not listed will be forwarded to
the Dean of the College’s office for review, to assess whether or not there has been a violation of Brown’s Academic Code.

If you have any questions about this policy, please ask the course staff for clarification. Not understanding our policy is not grounds for not abiding by it.

**Diversity and Inclusion**

The computer science department is committed to diversity and inclusion, and strives to create a climate conducive to the success of women, students of color, students of all (or no) sexual orientations, and any other students who feel marginalized for any reason.

If you feel you have been mistreated by another student, or by any of the course staff, please feel free to reach out to one of the CS department’s Diversity and Inclusion Student Advocates, or to Professor Greenwald or Professor Cetintemel (the CS department chair). We, the CS department, take all complaints seriously.

**Accommodations**

If you feel you have any disabilities that could affect your performance in the course, please contact SEAS, and ask them to contact the course staff. We will support accommodations recommended by SEAS.

**Harassment**

Please review Brown’s [Title IX and Gender Equity Policy](#). If you feel you might be the victim of harassment (in this course or any other), you may seek help from any of the resources listed [here](#).

**Alternative Brown Courses**

This course has no prerequisites. It teaches elementary statistics and elementary computer science, assuming no background whatsoever in either. Students who do have experience in these areas are encouraged to consider taking Data Science (CSCI 1951A) in the Computer Science department, which puts much more of an emphasis on databases; or Big Data (ECON 1660) in the Economics department, which has more of an emphasis on modeling causality. Both of these courses also involve programming, from scratch, various machine learning algorithms.
In CSCI 0100, we make use of off-the-shelf machine learning algorithms that are readily available in software libraries.

This course also differs from Introduction to Computation for the Humanities and Social Sciences (now CSCI 0030). The primary difference is that CSCI 0100 has a greater emphasis on statistics, while CSCI 0030 teaches more programming. In CSCI 0100 students learn R, a statistical software package, and many basic statistical concepts, such as regression. In CSCI 0030, students learn some Python and basic computer programming constructs, such as iteration. Nonetheless, there is significant overlap (for example, in visualization principles), so students are encouraged to take only one of these two offerings. Neither course has any prerequisites.