CAS CS 105: Introduction to Databases and Data Mining
Boston University
Fall 2016

Syllabus

Description: Databases and other collections of data are everywhere. Retailers use data about customers and their purchases to make decisions that increase profits. Researchers analyze data about the human genome to find treatments for diseases. Policymakers analyze socioeconomic data to gain insights that guide their decisions. Online music and video services perform data mining to deliver customized recommendations. How does all this work? CS 105 explores the ways in which collections of data are organized, stored, and analyzed. Topics include relational databases and the SQL query language, the writing of simple programs to process data, data visualization and the graphical display of data, and data-mining techniques for discovering patterns in data. Applications from a variety of domains (including business, the arts, the life sciences, and the social sciences) are used to illustrate the course's key concepts. The course counts towards the math/CS divisional studies requirement.

Prerequisites: none

Instructor
David G. Sullivan, Ph.D. (dgs@cs.bu.edu)
office hours: Wed., 11-noon; Thurs., 9-11; Fri., 2-3 p.m.; and by appointment
office: Psychology Building (PSY), room 228D
64 Cummington Mall (behind Warren Towers)

Teaching Fellow
Weicheng Ma (weicheng.ma1991@gmail.com)
See the staff page of the course website for office hours.

Meeting Times and Places
lectures: MWF, 1-2 p.m., LSE B01
labs: Wednesdays, 2-3 p.m. (A2)
       Tuesdays, 3-4 p.m. (A3)
       Wednesdays, 4-5 p.m. (A4)
All labs meet in the CS teaching lab, EMA 304.

Course Website: http://www.cs.bu.edu/courses/cs105
Requirements
1. Nine problem sets
2. A final project, which will involve using the techniques covered in the course to organize and analyze a collection of data that interests you, to draw conclusions based on your analysis, and to present your results in a clear and compelling way. **The project will be completed in teams of three.**
3. Three closed-book quizzes
4. A closed-book final exam
5. Attendance at and participation in both the lectures and labs

Policies
Lateness policy: Assignments must be submitted by the date and time listed on the assignment. There will be a 10% deduction for homework that is up to 24 hours late. **We will not accept any homework that is more than 24 hours late.** Plan your time carefully, and don’t wait until the last minute to begin an assignment. Starting early will give you ample time to ask questions and obtain assistance from members of the course staff.

Determining the final grade:
- attendance/participation 10%
- problem sets 30%
- final project 10%
- quizzes 20%
- final exam 30%

The final exam will replace your lowest assignment grade if doing so helps your final grade. The final exam will also replace your lowest quiz grade if doing so helps your final grade.

Extensions and makeup quizzes/exams will only be given in documented cases of serious illness or other emergencies.

You cannot redo or complete extra work to improve your grade.

Incompletes will not be given.

Collaboration Policy
- You are strongly encouraged to collaborate with one another in studying the textbook and the lecture materials.
- You may discuss ideas and approaches to the assignments with others, but such discussions should be kept at a high level, and should not involve actual details of the code or of other types of answers. You must also acknowledge clearly in the appropriate portion of your work (e.g., in the comments of your code) people with whom you discussed ideas for that portion.
- **The work that you submit must be entirely your own, which you must complete without looking at other people's work.** This includes the work of other students in the class, the work of people from outside the class, and
solutions found online. If you are having trouble, we encourage you to ask questions on Piazza, to email the course staff, or to come to office hours.

- **You must not permit other students to look at your work.**

**Violations of Collaboration Policy**

Violations of the collaboration policy — whether they involve basing your work on the work of someone else or allowing another student to look at your work — will result in a grade of 0 and will be reported to the Academic Conduct Committee (ACC). The ACC may suspend or expel students found guilty of plagiarism or other forms of misconduct. We will assume that you understand the University’s Academic Conduct Code, which is available here: [http://www.bu.edu/academics/policies/academic-conduct-code](http://www.bu.edu/academics/policies/academic-conduct-code)

**Textbooks (optional)**

We will provide lecture notes that fully cover all of the material you are expected to learn as part of the course. However, you may wish to consider purchasing:


**Schedule (tentative)**

<table>
<thead>
<tr>
<th>week</th>
<th>lecture dates</th>
<th>topics, exams, and special dates</th>
<th>optional readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9/7, 9/9</td>
<td>Course overview and introduction</td>
<td>K (Kroenke):</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Database fundamentals</td>
<td>chap. 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No lab meetings this week</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9/12, 9/14, 9/16</td>
<td>The relational model: How is data organized in a typical database? SQL: a language for creating, modifying, and extracting data from a database</td>
<td>K: chap. 2, 3</td>
</tr>
<tr>
<td>2</td>
<td>9/19, 9/21, 9/23</td>
<td>SQL (cont.) 9/19: last day to add a class</td>
<td>K: chap. 2, 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem Set 1 due on 9/22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9/26, 9/28, 9/30</td>
<td>SQL (cont.) case study: an astronomical database</td>
<td>see above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem Set 2 due on 9/29</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10/3, 10/5, 10/7</td>
<td>Programming in Python: intro., working with numbers</td>
<td>Z (Zelle): chap. 1, 2, 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem Set 3 due on 10/6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10/11, 10/12, 10/14</td>
<td>More Python: making decisions No lecture on 10/10 (Columbus Day) Lecture on 10/11 (Mon. schedule) 10/12: last day to drop without a 'W' Quiz 1 on 10/12</td>
<td>Z: sect. 7.1-7.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem Set 4 due on 10/14 (Friday)</td>
<td></td>
</tr>
</tbody>
</table>

David G. Sullivan, Ph.D.
<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Topics and Assignments</th>
</tr>
</thead>
</table>
| 6    | 10/17, 10/19 | More Python: working with strings/text and lists; accessing a database  
No lecture on 10/21  
**Problem Set 5 due on 10/20** |
| 7    | 10/24, 10/26, 10/28 | More Python: working with data stored in a text file  
**Problem Set 6 due on 10/27** |
| 8    | 10/31, 11/2, 11/4 | Data visualization: how to create compelling and useful data graphics  
**Problem Set 7 due on 11/3** |
| 9    | 11/7, 11/9, 11/11 | Data mining: what is it? How can a computer learn to categorize things?  
11/10: last day to drop a class with a 'W'  
**Quiz 2 on 11/9** |
| 10   | 11/14, 11/16, 11/18 | More data mining: categorizing things (cont.), finding patterns in numeric data  
**Problem Set 8 due on 11/17**  
**Final-project proposal due on 11/20** |
| 11   | 11/21 | More data mining: discovering arbitrary relationships in data  
No lectures 11/23 or 11/25 (Fall recess)  
No labs this week.  
**Work on final project** |
| 12   | 11/28, 11/30, 12/2 | More data mining: preparing data for mining, case study  
**Problem Set 9 due on 12/1** |
| 13   | 12/5, 12/7, 12/9 | Case studies and/or project sessions  
**Quiz 3 on 12/7**  
**Work on final project** |
| 14   | 12/12 | Wrap-up  
12/13-12/15: Study period  
**Final projects due on 12/12** |
| 15   |            | **Final exam: Mon., 12/19, 3-5 p.m.**  
The exam must be taken at this time, so please plan your travel accordingly! |