

CAS CS 105: Introduction to Databases and Data Analysis

Boston University

Fall 2009

Syllabus

Description: This course will examine how collections of data are organized, stored, and analyzed. Topics include: relational databases and the SQL query language; the process of writing simple programs to analyze data using the Python programming language; and data-mining techniques for discovering patterns in data. Applications from a variety of domains (including business, medicine, the arts, and the sciences) will be used to illustrate the concepts of the course through lectures and hands-on labs.

This course is a Math/CS divisional studies course.

Prerequisites: none

Instructor

David G. Sullivan, Ph.D. (dgs @ cs . bu . edu)

office hours: Mon., 3-5 p.m.; Wed., 2-4 p.m., and by appointment

office: Psychology Building (PSY), room 228D

64 Cummington Street (behind Warren Towers)

Teaching Fellow

Joseph Akinwumi (akin @ cs . bu . edu)

See the staff page of the course website for his office hours.

Meeting Times and Places

lectures: MWF, 1-2 p.m.

labs: Wednesdays, 2-3 p.m. (A2)

Wednesdays, 4-5 p.m. (A3)

All labs meet in the CS teaching lab, EMA 304.

Course Website: <http://www.cs.bu.edu/courses/cs105>

Please consult this site *frequently* (at least once per week) for important announcements and updates to the course materials handed out in lecture.

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
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, and a 20% deduction for homework that is 24-48 hours late. **We will not accept any homework that is more than 48 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%

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*

The collaboration policy for this class is as follows.

- You are strongly encouraged to collaborate with one another in studying the textbook and the lecture material.
- As long as it satisfies the following conditions, collaboration on the homework assignments is encouraged and will not reduce your grade:
 1. Before discussing each homework problem with anyone else, you must give it an honest half-hour of serious thought.
 2. You may discuss ideas and approaches with other students in the class, but not share actual code or solutions to other types of problems. In other words, the work you submit must be entirely your own, which you must complete without looking at other people's work, and you must not permit others to copy your work. You must also **acknowledge clearly in the appropriate portion of your solutions** (e.g., in the comments of your code) people with whom you discussed ideas for that portion.
 3. You may get help from the teaching staff and from tutors in the lab for specific homework problems. Don't expect them to do it for you, however.
 4. If you get *really* stuck with a bug in a program (defined roughly as over an hour of frustration), you are allowed to get help from a friend **as long as you acknowledge that help clearly in your solutions** (e.g., in the comments of your code).
 5. You may not work with people outside this class (but come and talk to us if you have a tutor), seek online solutions, get someone else to do it for you, etc.

- You are not permitted to collaborate on exams.

The last point is particularly important: if you don't make an honest effort on the homework but always get ideas from others, your exam score will reflect it.

Violations of Collaboration Policy*

Violations of collaboration policy fall into two categories: ones that are *acknowledged* at the time they occur (for example, in clearly marked comments in your code) and ones that are *unacknowledged*.

Acknowledged violations (e.g., using someone else's code for a method you didn't know how to write yourself, and stating clearly in your code that this is not your own work) will result in an appropriate reduction in the grade, but will not be considered cheating.

Unacknowledged violations of the collaboration policy—for example, not stating the names of your collaborators, or any other attempt to represent the work of another as your own—will result in an automatic failing grade and will be reported to the Academic Conduct Committee (ACC). The ACC often suspends or expels students deemed guilty of plagiarism or other forms of cheating. We will assume that you understand the CAS Academic Conduct Code, copies of which are available in room CAS 105.

If you are uncertain as to whether a particular kind of interaction with someone else constitutes illegal collaboration or academic dishonesty, please ask Dr. Sullivan *before* taking any action that might violate the rules; if you can't reach him in time, then at the very least include a clear explanation of what happened in your homework write-up to avoid being treated as a cheater. Citing your sources is usually the easiest way out of trouble.

Textbooks

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:

- *Database Concepts, 3rd edition* by David M. Kroenke (Prentice Hall, 2008)
- *Python Programming: An Introduction to Computer Science* by John Zelle (Franklin Beedle Publishing, 2004)

These books will be available for purchase at Barnes & Noble.

* Thanks to Prof. Leo Reyzin, who wrote the original versions of the sections describing the collaboration policy and violations to that policy. I have made only minor modifications.

Schedule (tentative)

week	lecture dates	topics, exams, and special dates	optional readings
1	9/2, 9/4	Course overview and introduction Database fundamentals <i>No lab meetings this week</i>	K (Kroenke): chap. 1
2	9/9, 9/11	The relational model <i>No lecture on 9/7 (Labor Day)</i>	K: chap. 2, 3
3	9/14, 9/16, 9/18	SQL Problem Set 1 due on 9/16 <i>9/16: last day to add a class</i>	K: chap. 2, 3
4	9/21, 9/23, 9/25	SQL (cont.) Database modeling and design Problem Set 2 due on 9/23	K: chap. 4, 5
5	9/28, 9/30, 10/2	Database modeling and design (cont.) Problem Set 3 due on 9/30 Quiz 1 on 10/2	K: chap. 4, 5
6	10/5, 10/7, 10/9	Programming in Python: intro., working with numbers Problem Set 4 due on 10/7 <i>10/7: last day to drop without a 'W', and to change from credit to audit status</i>	Z (Zelle): chap. 1, 2, 3
7	10/13 , 10/14, 10/16	Programming in Python: making decisions, working with strings and lists Problem Set 5 due on 10/14 <i>No lecture on 10/12 (Columbus Day)</i> <i>Lecture on Tues. 10/13 (Mon. schedule)</i>	Z: sect. 7.1-7.3, sect. 4.1-4.5
8	10/19, 10/21, 10/23	Programming in Python: accessing a database, working with files Problem Set 6 due on 10/21	Z: sect. 4.6
9	10/26, 10/28, 10/30	Programming in Python: files (cont.) Quiz 2 on 10/30	
10	11/2, 11/4, 11/6	Data visualization & information design Problem Set 7 due on 11/4	
11	11/9, 11/13	Data mining: intro., classification learning Problem Set 8 due on 11/13 <i>No lecture on 11/11 (Veteran's Day)</i> <i>No lab meetings this week</i> <i>11/12: last day to drop a class with a 'W'</i>	
12	11/16, 11/18, 11/20	Classification learning, numeric estimation, preparing the data Final-project proposal due on 11/18	
13	11/23	Preparing the data, association learning <i>No lectures 11/25 or 27 (Fall recess)</i> <i>No lab meetings this week</i>	

14	11/30, 12/2, 12/4	Association learning, case study Problem Set 9 due on 12/2 Quiz 3 on 12/4	
15	12/7, 12/9, 12/11	Case studies and/or project presentations Final projects due on 12/11	
		Final exam: Mon., 12/21, 12:30-2:30 p.m.	