

CAS CS 105: Introduction to Databases and Data Mining
Boston University
Fall 2018

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: see the staff page of the course website
office: Psychology Building (PSY), room 228D
64 Cummington Mall (behind Warren Towers)

Teaching Fellow

Ryan Yu (ryu1@bu.edu)
See the staff page of the course website for office hours.

Meeting Times and Places

Lecture: MWF, 1:25-2:15 pm, CAS 211
Lab: a weekly, one-hour session; see your schedule for the time and location.

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

Requirements and Grading

1. Nine problem sets (30% of the final grade)
2. A final project (10%). This 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 quizzes (20%)
4. Final exam (30%)
5. Preparation and participation (10%)

The final exam will replace your lowest problem-set 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.

Collaboration Policy

You are strongly encouraged to collaborate with one another in studying the lecture materials and preparing for quizzes and exams.

You may discuss ideas and approaches to the assignments with others (provided that you acknowledge doing so in your solution), 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 complete the actual solutions on your own.**

Academic Misconduct

We will assume that you understand BU's Academic Conduct Code:

<http://www.bu.edu/academics/policies/academic-conduct-code>

Prohibited behaviors include:

- copying all or part of someone else's work, even if you subsequently modify it; this includes cases in which someone tells you what you should write for your solution
- viewing all or part of someone else's work
- showing all or part of your work to another student
- consulting solutions from past semesters, or those found online or in books
- posting your work where others can view it (e.g., online).

Incidents of academic misconduct will be reported to the Academic Conduct Committee (ACC). The ACC may suspend/expel students found guilty of misconduct. ***At a minimum, students who engage in misconduct will have their final grade reduced by one letter grade (e.g., from a B to a C).***

Other Policies

Laptops: Students taking CS courses are expected to have a laptop capable of running a currently supported version of Microsoft Windows, Mac OS X, or Linux. See this page for more info: <https://www.bu.edu/cs/undergraduate/undergraduate-life/laptops>

Late problem sets: Problem sets must be submitted by the date and time listed on the assignment (typically by 11:59 p.m.). There will be a 10% deduction for submissions 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 so you will have ample time to ask questions and obtain assistance from the course staff.

Pre-lecture preparation: To help you prepare for lecture, you will typically be required to complete an assigned reading and/or watch one or two short videos. You will also be required to perform some type of brief task (an online quiz or other exercise) to demonstrate that you have completed the necessary preparation. Your work on these tasks will not typically be graded for correctness, but it should demonstrate that you have adequately prepared for lecture. The pre-lecture tasks must be submitted by the specified date and time. **Late pre-lecture work will not be accepted.**

The *attendance/participation* portion of your grade will be based on your consistent attendance at the lectures and labs, and on your participation in the activities for each class. In particular, you must participate in small-group activities during lecture in which you will discuss questions with other students and "vote" on the answers. These activities are designed to deepen your understanding of the material, and you will be graded on your participation, *not* on the correctness of your answers. To accommodate unavoidable absences or lateness, you will receive full credit for attendance as long as you make at least 85% of the votes over the course of the semester and attend 85% of the lab sessions. **Voting from outside of the classroom or voting for someone else is not allowed, and will result in a lowering of the participation grade of the students involved.**

The final grades are *not* curved. The performance of the class as a whole is taken into account when assigning letter grades, but this can only improve your grade, not harm it.

Extensions and makeup 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 except in extraordinary circumstances.

Course Materials

Textbook: There is no required textbook for the course. We will provide detailed lecture notes and optional supplemental readings.

In-class software: We will be using the Top Hat platform for in-class activities and attendance. More detail will be provided in class.

Schedule (tentative)

week	lecture dates	topics, exams, and special dates
0	9/5, 9/7	Course overview and introduction Database fundamentals <i>No labs this week</i>
1	9/10, 9/12, 9/14	The relational model: How is data organized in a typical database? SQL: a language for creating, modifying, and extracting data from a database
2	9/17, 9/19, 9/21	SQL (cont.) <i>9/17: last day to add a class</i> Problem Set 1 due on 9/20
3	9/24, 9/26, 9/28	SQL (cont.) Problem Set 2 due on 9/27
4	10/1, 10/3, 10/5	Programming in Python: intro., working with numbers Problem Set 3 due on 10/4
5	10/9 , 10/10, 10/12	More Python: making decisions <i>No lecture on 10/8 (Columbus Day)</i> <i>Lecture on 10/9 (Mon. schedule)</i> <i>10/9: last day to drop without a 'W'</i> Quiz 1 on 10/10 Problem Set 4 due on 10/12 (Fri)
6	10/15, 10/17, 10/19	More Python: working with strings/text and lists; accessing a database Problem Set 5 due on 10/18
7	10/22, 10/24, 10/26	More Python: working with data stored in a text file Problem Set 6 due on 10/25
8	10/29, 10/31, 11/2	Data visualization: how to create compelling and useful data graphics Problem Set 7 due on 11/1
9	11/5, 11/7, 11/9	Data mining: what is it? How can a computer learn to categorize things? <i>11/9: last day to drop a class with a 'W'</i> Quiz 2 on 11/7
10	11/12, 11/14, 11/16	More data mining: categorizing things (cont.), finding patterns in numeric data Problem Set 8 due on 11/15 Final-project proposal due on 11/18
11	11/19	More data mining: discovering arbitrary relationships in data <i>No lectures on 11/21, 11/23 (Fall break)</i> <i>No labs this week.</i>
12	11/26, 11/28, 11/30	More data mining: preparing data for mining, case study Problem Set 9 due on 11/29

13	12/3, 12/5, 12/7	Case studies and/or project sessions Quiz 3 on 12/5 Work on final project
14	12/10, 12/12	Case studies and/or project sessions Final projects due on 12/12 <i>12/13-16: Study period</i>
15		Final exam: Fri, 12/21, 12:30-2:30 pm <i>The exam must be taken at this time, so please plan your travel accordingly!</i>