CS-181 and DA-210, Data Systems

> Syllabus, Fall 2020

Summary Information

Professor	Thomas C. Bressoud (Dr. B)
Office	Olin Hall 210
Email	bressoud@denison.edu
Class Hours	MWF: 10:20 a.m. to 11:10 a.m. T: 8:30 a.m. to 9:20 a.m. Zoom Class
Office Hours	MWF: 8:30 a.m. to 9:20 a.m. TR: 2:50 p.m. to 3:40 p.m. Zoom Office Hours
Final Exam	after Thanksgiving; time TBD
Course Page	https://notebowl.denison.edu
Weekly Schedule	CS-181 Weekly Schedule on Google Drive
Credit Hours	4 — *This course adheres to Denison's Academic Credit Policy*
Teaching Assistants	Henri Hegemier (hegemi_h1@denison.edu) Dan Seely (seely_d1@dension.edu) TA Session Zoom

COVID-19 Summary

- Primarily classes will be held in person and on campus.
  - Professor is planning on teaching in person, at least until such time that health and safety are an issue.
- All on-campus students are expected for all in-person class sessions.
  - No "split" for who, on campus, is expected to attend particular class sessions.
  - Students in quarantine, or those in isolation that are asymptomatic or able to manage will move to remote interaction until they are cleared.
  - To better enable breakouts and group interaction, the Tuesday sessions may move to entirely remote, but we will decide that as we progress.
- For social distancing, class will be in a large classroom space.
- All waiting list students were allowed to enter the class, and so the total class size is 30, and of these, five have elected to be remote, and almost all of these are +11 or +12 hour time difference from US Eastern time.
- All class sessions will involve all participating students using Zoom, and these Zoom sessions will be recorded to give students options for viewing asynchronously.
- Participation will be a part of the grade, which can be satisfied, in part, by being part of the synchronous class sessions, or by checking in during office hours.

- Office hours will all be held via Zoom, whether students are on-campus or remote.
- To support remote students and the time zone difference, MWF office hours will be at 8:30 am, with TR hours in the afternoon for on-campus students unwilling to be active at that time.
- To make test assessments lower stakes, we will have four **quizzes** during the semester at approximately 3 week intervals, instead of the normal 2 or 3 midterms, and each of these will be held during the Tuesday morning session so that on-campus and remote students can all take the quizzes synchronously.
- In-semester projects will be reduced to a single project with multiple options for fulfilling, and the end-of-semester assessment can either be a culminating synthesis project or a cumulative final exam.
  - With the revised academic schedule, the final exam or final project will be remote and completed after Thanksgiving.
- Given the size of the class and the challenging semester expected, we have been approved for *two* teaching assistants, who will hold Zoom sessions for "walk-in" tutoring hours and will run the auto-grading part of homework assignments.
- **All** students are expected to have their own computing device (laptop), capable of running Python and the Anaconda development environment, and to have that device at every class and quiz session. (This is not new for the COVID situation, but is worth emphasizing at the onset.)

Description

The Data Systems course provides a broad perspective on the access, structure, storage, and representation of data. It encompasses traditional Database systems, but extends to other structured and unstructured repositories of data and their access/acquisition in a client-server model of Internet computing. Also developed are an understanding of data representations amenable to structured analysis, and the algorithms and techniques for transforming and restructuring data to allow such analysis. The primary programming language used in the course will be *Python*.

Course Content and Learning Goals

We have divided the course content into six units, scaffolding the learning goals and starting from the basis of the CS 11X introductory course and having each unit building from the last and each introducing additional complexity in Data Systems. Units may take more or less class time to treat adequately, so the durations listed for each unit in the table below are only approximate.

Unit	Name	Description
00	*Foundations*	After covering an overview of data systems, we begin with building a foundation in Python to be used in the remainder of the course. We review CS11X topics such as working with files and file objects, representing data in lists and dictionaries, and then progress to useful and possibly new topics like list comprehensions, lambda functions, and regular expressions. Duration: 2 weeks.
01	*Tabular Data Model*	The focus of this unit is on single, rectangular table/data frame datasets. We will use Python facilities for generating/transforming lists and work toward structured data frames supported by the pandas module. Data will be input from URLs and files and organized in rectangular text formats like CSV and variations. The data model implied by so-called "tidy" data arrangements will be covered. Duration: 2+ weeks.

Unit	Name	Description
02	*Relational Data Model*	The single most important form of Data System, grounded in nearly 50 years of development of both underlying theory of structure and efficient implementations, is the Relational Database system. This unit will explore the data model associated with relational databases and the operations and organization involved in sound design of the set of tables for a relational database. We will also gain skills in querying existing database and creating databases using the Structured Query Language (SQL) through both direct commands and through programmatic use. Duration: 3 weeks.
03	*Networking, HTTP, HTML*	As we expand how we acquire data from various sources, we must start with learning how networks operate and how data is requested and returned to client applications. In this unit, our goals are to understand and program client applications over the network using HTTP, the protocol of the web, and to understand client server requests for static data in files or as HyperText Markup Language (HTML). Duration: 2 weeks.
04	*Hierarchical Data Model*	When we interact with data systems on the Internet, many providers of data structure and represent provided data in a flexible form that is based on a tree data model. Dominant specific examples include XML and JSON data formats, but HTML trees will also be discussed. This unit will focus on the operations involved in "traversing" such structures and formats and explore both declarative and programmatic means to process the data and build data frames in our programs to allow for analysis. Duration: 2 weeks.
05	*APIs*	Systems providing data (servers) must be built in ways that allow client applications to follow established protocols and conventions that allow the client to determine and specify what data is desired. Further, systems may need to identify particular users and/or client applications to determine and distinguish valid requests from unauthorized access. The protocols and conventions for network-based access of data are known as the Application Programming Interface (API). This unit will cover some of the most common types of APIs and associated authentication, and look at the protocols and network communication that underlie these APIs. This unit will occupy the remainder of the semester.

Integrated Course Aspects

- Visualizations: For each unit, we use Tableau, a cloud and desktop based tool that facilitates the design and construction of appealing visualizations from well structured datasets.
- Real-world Data Sources: The world is filled with messy data, sometimes incomplete and often repurposed and combined from multiple sources. One of the goals of this course is to enable students to adapt and compensate for the messiness of data and to structure and transform it into a form amenable for analysis and use. So projects in each instructional unit ask students to acquire data from primary sources and adapt it to their needs.

Resources

1. **Introduction to Data Systems** by Bressoud and White: This is a book being published by Springer Nature and written by your Denison professors for this class. Publication is in process and, by copyright, we are not allowed to distribute PDF, so we have engaged the copy center and bookstore to create book copies in three-ring binders for your use.

2. **Online resources**: The Data Systems, as treated in this course, cover a very broad spectrum. The new textbook brings many of the topics together into a single resource, but no single text could hope to address all aspects of the

material, particularly with regard to programming packages and their extensive interfaces. Therefore a significant portion of our reading and exploration of the topics of this course will involve investigation through tutorials, instructional materials, and other online sites and resources. One of the goals of this class is to get students comfortable enough in this kind of exploration that, when faced with a need to acquire and structure data from a new data source, they can learn about the system and "come up to speed" in figuring out how to solve the problems before them.

Coursework

The primary forms of coursework in CS-181 include:

- **Classwork and Participation**

  Several times per unit, we will engage in hands-on tutorial and in-class sessions working with iPython notebooks. These will be submitted and checked for due-diligence completion and to evaluate basic correctness.

  In addition, students are expected to attend class and to be actively engaged, participating in discussions and asking questions. Students will be expected to share their screen and "drive" as we work through in-class activities.

- **Homework Exercises and Reading Questions**

  Homework assignments are designed to facilitate student mastery of the building block concepts and skills needed in data systems. Expect a homework assignment three times a week. These almost always take the form of iPython notebooks. Homework assignments are comprised mostly of programming exercises, which are mostly Python coding, often with a set of assert statements for feedback on the correctness of each answer, and the exercises build from material covered in class. The reading questions cover material from assigned readings *before* class coverage, and are graded more leniently and help facilitate that reading has been done, and yield more effective class time sessions.

  It is a violation of academic integrity to share or exchange answers to homework exercises and reading questions.

  Late homework submissions will lose 15% of the entire assignment grade per subsequent 24 hour period for the first two days. Then assignments are graded and the grace period is over.

  The lowest two homework grades will be dropped, to allow for extenuating circumstances encountered during the semester.

- **Projects**

  There will be one project halfway through the semester and the option of a final project instead of a final exam at the end of the semester. Students will use and combine the knowledge gained from the exercises, the class coverage, and their own exploration from the Internet to synthesize solutions to project problems. Projects involve the basic steps of acquiring data from one or more Internet sources, cleaning, processing and combining data, and designing and producing a visualization to answer exploratory questions about the data. These projects will be completed individually. The last project will be a synthesis project that involves APIs, multiple data systems providers, and security.

- **Quizzes**

  There will be 4 quizzes spread throughout the semester. Each will assess students' mastery of new material, but some cumulative material may also be expected because of the way the units build from one to the next. Note that all quizzes count toward the final grade. Quizzes may not be "made up" after the fact, but some accommodation to take one early will be arranged in the case of a University sanctioned excuse.

Grade Determination

Category	Weight
Homework	25%
Project	10%
Quiz	30%
Final Exam or Final Project	25%
Classwork/Reading Questions/Participation	10%

Policies

Academic Integrity

Proposed and developed by Denison students, passed unanimously by DCGA and Denison's faculty, the Code of Academic Integrity requires that instructors notify the Associate Provost of cases of academic dishonesty. Cases are typically heard by the Academic Integrity Board which determines whether a violation has occurred, and, if so, its severity and the sanctions. In some circumstances the case may be handled through an Administrative Resolution Procedure. Further, the code makes students responsible for promoting a culture of integrity on campus and acting in instances in which integrity is violated.

Academic honesty, the cornerstone of teaching and learning, lays the foundation for lifelong integrity. Academic dishonesty is intellectual theft. It includes, but is not limited to, providing or receiving assistance in a manner not authorized by the instructor in the creation of work to be submitted for evaluation. This standard applies to all work ranging from daily homework assignments to major exams. Students must clearly cite any sources consulted—not only for quoted phrases but also for ideas and information that are not common knowledge. Neither ignorance nor carelessness is an acceptable defense in cases of plagiarism. It is the student's responsibility to follow the appropriate format for citations. Students should ask their instructors for assistance in determining what sorts of materials and assistance are appropriate for assignments and for guidance in citing such materials clearly. For further information about the Code of Academic Integrity, see http://denison.edu/academics/curriculum/integrity.

> In this class, you may discuss problems with other students in the class, at the level of drawing pictures and abstract solutions on the whiteboard, but written and typed work, including homework code, must be your own. Under no circumstances is it ok for exchange of files between students, nor for one student to examine another student's code on their screen. While it may be easy for students to pass files around in violation of this policy, it is also very easy for me to run programs to measure the degree of similarity between students' submissions, and I do so on a regular basis. Also note that these "plagiarism detectors" are not fooled by changing variable names and moving things around that some students think will fool the professor and disguise such cheating. See, for example, the MOSSMeasure of Software Similarity) system provided by Stanford University.

Students found responsible for breaches of academic integrity will earn a failing grade for the course and may receive even more severe consequences, per University policy

Attendance and Participation Policy and Responsibility

I expect a high level of participation by all the students in this class. True learning requires an active role and regular attendance in order to engage and master the material. Thus, I expect your attendance at every class meeting. Besides the direct grade-effect on quizzes and and classwork activities, there is the 4% participation part of the grade that I will use in instances of lack of attendance or participation during class.

You are responsible for all reading assignments, whether or not the material is also covered during class time, as well as for announcements and decisions made relative to coursework that may occur during classtime. If you miss class, it is *your responsibility* to get notes and information on the material that was conveyed during any class you might miss.

Accommodations

Students with a documented disability who wish to request reasonable academic accommodations based on the impact of a disability should complete a Request for Academic Accommodations/Faculty Notification form with the Academic Resource Center (ARC) located in 020 Higley Hall and contact me privately as soon as possible to discuss specific needs and arrangements. I rely on the Academic Resource Center (ARC) to verify the need for reasonable accommodations based on the documentation on file in that office. Reasonable accommodations cannot be applied retroactively and therefore ideally should be enacted early in the semester as they are not automatically carried forward from a previous term and must be requested every semester.

Appropriate Use of Course Materials

As an institution which strives to inspire and educate our students to become discerning moral agents and active citizens of a democratic society, we are committed to complying with all laws regarding copyright throughout the University. This syllabus and all course materials used in this course may be copyrighted and accordingly will be governed by the provisions of the U.S. copyright law (for an overview see http://copyright.gov/circs/circ01.pdf and for fair use guidelines see http://copyright.gov/fair-use/more-info.html). In particular, posting any course materials on commercial sites or creating a bank of materials for distribution to other students may be considered a violation of the University's Code of Academic Integrity as well as a breach of copyright law. If you have any questions about these guidelines, please speak with your instructor.

Reporting Sexual Assault

Project write-ups, and other coursework submitted for this class are generally considered confidential pursuant to the University's student record policies. However, students should be aware that University employees are required by University policy and Title IX guidance to report allegations of discrimination based on sex, gender, gender identity, gender expression or sexual orientation, including sexual misconduct, sexual assault and suspected abuse/neglect of a minor, occurring on campus and / or involving current students at Denison University when they become aware of possible incidents in the course of their employment, including via coursework or advising conversations. There are others on campus to whom you may speak in confidence, including counselors at the Whisler Center for Student Wellness, and clergy. More information on Title IX and University policy guidance on gender identity / expression bias and sexual misconduct / assault, including support resources, how to report, and prevention and education efforts, can be found at denison.edu/titleix; students may also contact Steve Gauger, Campus Title IX Coordinator, in Doane Administration 001, by email at gaugers@denison.edu, or by phone at 740-587-8660.