DSCI 549: Introduction to Computational Thinking and Data Science 32430D — Fall 2021

Instructors Information

Name: Deborah Khider
Email: khider@usc.edu
Office Hours: M-W - 4:00-5:00pm PT (7:00-8:00pm ET) - Zoom: 917 9463 3637

Name: Valentino Crespi
Email: crespi@usc.edu
Office Hours: T-W - 6:30-7:30am PT (9:30-10:30am ET) - Zoom: 931 8767 5656

Graders Information

Name: TBA

Class Information

Dates: 8/27/2020 – 12/3/2020
Time and classroom: F 1:00pm PT – 4:20pm PT, THH 210 AND/OR ONLINE

All class communications, materials dissemination, and announcements will occur via Blackboard.

Course Description

Introduction to data analysis techniques and associated computing concepts for non-programmers. Topics include foundations for data analysis, visualization, parallel processing, metadata, provenance, and data stewardship.

Expanded Course Description

This course will teach non-programmers to think in computing terms about modern topics, and to approach real-world phenomena through data science. The course is intended for students in disciplines outside of computer science, so no prior experience with computer science is assumed. The course topics will be particularly relevant to students interested in physical, biological, and social sciences. In particular, the course will cover different kinds of data and corresponding approaches to data analysis, including geospatial data, time series, networks, and multimedia data.

Learning Objectives and Outcomes

Students learn to run multi-step analysis through a graphical workflow interface and will experience firsthand complex concepts in data science such as parallel computing, provenance, and visualization. Students also learn to use ontologies and logic representations to capture metadata and other knowledge about complex data. The course includes practical lessons to use workflow

and ontology development toolkits, as well as best practices for data stewardship and dissemination.

In particular, the course will enable students to:

- Acquire computational thinking skills that will enable students to represent and reason about complex problems in the digital arena
- Understand different kinds of data in terms of their possibilities and limitations to approach complex problems cast in terms of the emerging field of data science
- Familiarize with data science scholars' best practices in data documentation and dissemination

Prerequisite(s):
None
Co-requisite(s):
None
Recommended preparation:
Mathematics, Statistics, and Logic undergraduate course.

Textbooks and Software

There is no textbook. All required software is freely available for students to install on their personal computers or to access through a web interface.

Lectures

Lecture materials will be pre-recorded and made available at least one week prior to class. Students are expected to listen to the material before the Friday class. Class time will be used for hands-on practicals.

Homework assignments

There will be seven homework assignments throughout the course. The homework assignments will test student understanding of the concepts learned in class but will be independent of the main class project. The assignments must be submitted individually and students will receive individual scores. These assignments should be treated as quizzes in the sense that only clarifying comments about the questions can be asked to the instructors.

Students may work in groups to complete the tasks. The homework assignments are expected to take 3-4 hours. Each assignment is graded on a scale of 0-100 and the grading criteria will be specified in each assignment. **All answers must be justified. Answers with no justification will not be marked.**

Homework assignments are due at **12:00pm Pacific** on the due date and should be submitted via Blackboard. Homework will be accepted up to one week late as long as the student requested a late submission ahead of the deadline, and, in that case, the assignment will be graded at 20% less than the total possible points for the assignment. After one week, the assignment will not be graded.

You will be allowed to drop the two lowest grades, including any assignments not submitted within the 1-week grace period.

Project

Students will be asked to complete a project based on the materials learned in class. The project must be submitted individually and students will receive individual scores. The project replaces midterm/final for the class. Only clarifying comments about the questions can be asked from the instructors.

Students may work in groups; **however** the dataset/questions/methods must be unique to each project. Projects that are too similar in nature will be given a grade of zero. **All answers must be justified. Answers with no justification will not be marked.**

The first part of the assignment will be due on **Monday Nov 1st, 2021** at **5:00pm Pacific** while the second part will be due on **Wednesday Dec 15th, 2021** at **5:00pm Pacific**. No late report will be accepted except in the case of a **documented emergency**. For the second submission, students must request an Incomplete if they are unable to submit by the deadline.

Grading

The course grade is determined by the following components:

Homeworks	500 points (7x100 points, 2 grades dropped)
Project	500 points (2x250 points)

Grade Scale

Final grades will be assigned according to the following scale:

A	940–1000	C+	770–799
A–	900 – 939	C	740–769
B+	870–899	C–	700–739
B	840–869	D	640–699
B–	800–839	F	0 – 639

No rounding up will be applied.

Schedule

Lecture materials will be pre-recorded and made available on Blackboard at least one week prior to class. Class time will be dedicated to practicals. These lectures will either be online or in-person (see schedule below) and recorded. These recordings will be made available through Blackboard. The practicals will not be graded and will consist of short quizzes to test the understanding of the materials, tutorials on software, and examples of typical data science tasks. Answer keys will be distributed through Blackboard.

	Lecture Topic	In-class Practicum	Assignment due
27-Aug		Introduction	
3-Sep	Data Science Data Life Cycle	Introduction to Jupyter Notebook and Google Colab Data Exploration	
10-Sep	Data Analysis Software/Workflow	Describing your data science project as a workflow WINGS Practicum	Homework 1
17-Sep	Experimental Design Probability	Observational vs experimental studies Understanding bias in experimental design [ONLINE]	Homework 2
24-Sep	Inferential Statistics	Inference through simulations [ONLINE]	
1-Oct	Data Analytics	Machine learning practicum [ONLINE]	Homework 3
8-Oct	Network and Timeseries	Predicting ENSO using deep learning [ONLINE]	Homework 4
15-Oct		NO CLASS - FALL BREAK	
22-Oct	Natural Language Processing Multimedia	NLP practicum [ONLINE]	
29-Oct	Parallel and distributed computing	NO CLASS - Work on projects	Homework 5
5-Nov	Semantic Metadata/Ontology	Designing an ontology Protégé practicum	Project 1
12-Nov	Provenance and data stewardship	Introduction to GitHub, Zenodo, Figshare Using Provo to report on data science study	Homework 6
19-Nov	Visualization	Designing a dashboard	Homework 7
26-Nov		NO CLASS - Thanksgiving	
3-Dec	Communicating data science	Presentation drills	
15-Dec			Project 2

Attendance

Students are expected to attend synchronously in-person or online depending on the lecture. All lectures will be recorded and posted through Blackboard. Students who cannot attend synchronously are required to review the recorded material on their own time, in addition to the pre-recorded lectures.
Please, **DO NOT** attend class if you are feeling sick or have received a positive COVID-19 test as mandated by University policy.
Students attending class lectures in-person are **required** to wear a mask at all times. There will be **absolutely no eating** during class time. Students should put their mask back on after drinking **immediately**.

Office Hours

All material related questions will be covered in office hours. Office hours will be held at the regular times indicated for each instructor at the beginning of this syllabus and are available on a first-come, first-served basis. Other times are on a case-by-case basis by appointment contingent on instructor's availability.
Please come prepared with specific questions about the materials covered in the pre-recorded lectures and in-class exercises. Please remember that the homework assignments and project represent your understanding of the material and we will therefore not answer questions such as "Am I doing this right?", "Can you tell me if this is the right answer?".

There will be no Slack channel for this class. Email should be limited to administrative questions and requests for extensions under the conditions explained above.

Honor Code

In response to recommendations made by the Academic Integrity Task Force to the Dean, the USC Viterbi School of Engineering now has an Honor Code. The Code was developed by Viterbi students, and its text is as follows:

> Engineering enables and empowers our ambitions and is integral to our identities. In the Viterbi community, accountability is reflected in all our endeavors.
>
> Engineering+ Integrity. Engineering+ Responsibility. Engineering+ Community. Think good. Do better. Be great.
>
> These are the pillars we stand upon as we address the challenges of society and enrich lives.

During your time here at Viterbi, please know that academic and personal resources are available to help you:

- The student-driven and student-written Honor Code is here: http://viterbi.usc.edu/academics/integrity/.
- An introductory video is posted at https://myviterbi.usc.edu/ under the link "Academic Integrity Introduction" and serves as a reminder of the school's emphasis in maintaining a high level of academic integrity.
- Master's and PhD students can contact the GAPP office in OHE 106 (https://gapp.usc.edu/) for other helpful resources.
- The Viterbi Academic and Resource Center (VARC) (http://viterbi.usc.edu/students/undergrad/varc) has a variety of services available.

Academic Integrity

The Viterbi School takes academic integrity violations seriously. Most of the violations that have been reported in the past fall into four categories: unauthorized collaboration, plagiarism, code sharing, and cheating on an exam. Specifically:

- Unauthorized collaboration - Unauthorized collaboration on a project, homework or other assignment. (section 11.14.B) All homework assignments must be individually developed. Students that collaborate on assignments will be referred to the Academic Integrity Coordinator.
- Plagiarism - presenting someone else's ideas as your own, either verbatim or recast in your own words - is a serious academic offense with serious consequences.
- Code sharing - Obtaining for oneself or providing for another person a solution to homework, a project or other assignment, without the knowledge and expressed consent of the instructor. (section 11.14.A)
- Cheating in an exam - this may involve a number of violations, such as looking at class notes during the exam, looking at other student's exam, "texting" with other students during the exam. See the section titled Two Exams for a list of specific violations.

Please note that that these are only the basic violations that we have encountered in the past, and there are many more. Please familiarize yourself with the discussion of plagiarism in SCampus in Section B.11.00, Behavior Violating University Standards and Appropriate Sanctions available at https://scampus.usc.edu/b/11-00-behavior-violating-university-standards-and-appropriate-sanctions/. All academic integrity violations will be referred to the Academic Integrity Coordinator of the Viterbi School of Engineering. The process for adjudicating these cases is available in SCampus, Part B, Section 13.

Other Misconduct

Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct/.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity http://equity.usc.edu/ or to the Department of Public Safety http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety whole USC community. Another member of the university community - such as a friend, classmate, advisor, or faculty member - can help initiate the report, or can initiate the report on behalf of another person. The Center for Women and Men http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the American Language Institute http://dornsife.usc.edu/ali which sponsors courses and workshops specifically for international graduate students. The Office of Disability Services and Programs http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, USC Emergency Information http://emergency.usc.edu/ will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.

Diversity

The diversity of the participants in this course is a valuable source of ideas, problem solving strategies, and engineering creativity. The instructors encourage and support the efforts of all of our students to contribute freely and enthusiastically. As members of an academic community, it is our shared responsibility to cultivate a climate where all students and individuals are valued and where both they and their ideas are treated with respect, regardless of their differences, visible or invisible.

Students with disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m. - 5:00 p.m., Monday

through Friday. Website and contact information for DSP: http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html, (213) 740-0776 (Phone), (213) 740-6948 (TDD only), (213) 740-8216 (FAX), ability@usc.edu.

Emergency Preparedness/Course Continuity in a Crisis

In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies.