DS 202 (Fall 2020): Data Acquisition and Exploratory Data Analysis

Delivery method: Flipped. Modules available on Canvas, and real-time class meetings on Zoom.

Online meeting times: Tue Thu 8:00 – 9:20 am

Instructor: Xiongtao Dai
    Email: `xdai@iastate.edu`
    Student hours: Thu 8:50 – 9:20 am, and by appointment

Teaching assistant: Lynn Huang
    Email: `lwhuang@iastate.edu`
    Student hours: Thu 8:00 – 8:50 am, and Fri 11:00 am – 12:00 pm

Prerequisites: DS 201

Course description: Data acquisition: file structures, web-scraping, database access; ethical aspects of data acquisition; types of data displays; numerical and visual summaries of data; pipelines for data analysis: filtering, transformation, aggregation, visualization and (simple) modeling; good practices of displaying data; data exploration cycle; graphics as tools of data exploration; strategies and techniques for data visualizations; basics of reproducibility and repeatability; web-based interactive applets for visual presentation of data and results. Programming exercises.

What you will learn: Be able to

- acquire and read data in different formats and from different sources;
- implement a basic data processing pipeline;
- explore data;
- visualize complex data in appropriate forms; and
- communicate your findings in a reproducible form.

How you will learn: The course will be taught in a flipped format, with everything delivered remotely (so there is no in-person meetings). In a typical week, you will

- study recorded lecture videos which "cover the contents" by Monday;
- finish an individual quiz by Monday, and an identical group quiz in class on Tuesday;
- during Tuesday's meeting, participate in demonstrations, group discussions, case studies, and lab activities which aims at applying the newly-learned knowledge to analyze a real-world dataset;
- work on weekly homework or lab assignments in Thursday's meeting, when the instructor and TA will be available to answer your questions;
- hand in your assignment by the next Monday.

Computer software: This class will be teaching the `R` language (https://www.r-project.org/). Making proper use of version control systems `git` and GitHub will be required for the homework, lab, and final project.

Optional reading materials:

- *R for Data Science*, Garrett Grolemund and Hadley Wickham, https://r4ds.had.co.nz/
- Chapter 1–2, *Pro Git*, Scott Chacon and Ben Straub, https://git-scm.com/book/en/v2

Course webpage: We will post course materials and announcement on Canvas. Much of the materials are developed by Dr Heike Hoffman.

Piazza: We will use Piazza as the Discussion board for our course. It is a website that allows students and instructors to post code, images, and formula very conveniently. Piazza is integrated into Canvas and so one can access it there; alternatively, visit (`www.piazza.com/iastate/fall2020/ds202`) in your browser. All questions regarding course work should be posted there, so that anyone in the class including the instructor, the TA, and peer students can contribute to a quick and complete answer. The response time is 24 hours. Emails to the TA and the instructor are reserved only for personal matters.

When posting on Piazza, please follow adequate netiquettes:

- Be polite and respectful to others.
- Search before you post. Your question may have already been asked and answered.
- When you post a question, please explain the context and give an example of what you have issue with. Posting screenshots and asking "What is going wrong?" is unacceptable.
- Posting short snippet of code is fine, but please refrain from posting a complete solution to a question.

Learning groups: You will be assigned to a learning group that lasts throughout the semester. The learning group is the primary point where you can turn to for help and discussion.

Quizzes: There will be weekly individual quizzes posted on Canvas. The quiz questions are supposed to be relatively simple and can be immediately answered after studying the lecture videos. The individual quizzes are due Monday nights, and will be immediately graded with the total score released but not the per-question scores.

Students will be able to correct mistakes and attempt a higher score in the group quiz in Tuesday's meeting, which will be the same as the individual quiz. Students will discuss within the learning group and arrive at a single answer. One of the group members will submit the quiz answers on Canvas, representing the group. If a question is answered incorrectly in the individual quiz but correctly in the group quiz, the student will gain back the deducted scores. Taking the group quiz will never decrease your quiz score. Only students who participated in the group quiz will receive credits.

Homework and Labs: There will be five homework and five lab assignments throughout the semester except for the midterm and preparation weeks. Homework are to be finished individually, and lab assignments are finished in pairs. One of the study groupmate will be assigned as your partner for each lab. You are encouraged to discuss the assignments with your groupmates or anyone else in the class. However, you (and your partner) must write the assignments individually. **Plagiarism detection will be strictly enforced** using the Measure Of Software Similarity (MOSS, `https://theory.stanford.edu/~aiken/moss/`).

Midterm exam: There will be one midterm exam. The midterm will be an open-everything take-home exam. You cannot obtain help from anyone else, however. The midterm exam will test on your understanding of R and git commands, data wrangling, graphics production, and real data analysis. There will be no final exam.

Final project: The final project is a team-based data analysis project. The project teams will be the study groups. The project consists of a preliminary investigation proposal, a written report, and an oral presentation during class time (dead week and finals' time). More information is available on Canvas.

Participation: We expect active participation in class, in the student hours, and on Piazza.

Grading: Letter grades will be assigned by the instructor. The grade may be curved, but only in a direction beneficial to the students as compared to the standard grading scheme (90% A-range, 80% B-range, etc). The graded components are

- 10% weekly quizzes
- 40% homework and labs
- 20% midterm exam
- 25% final project (5% proposal, 10% presentation, 10% report)
- 5% participation (primarily Piazza and team feedback)

Academic dishonesty: The class will follow Iowa State University's policy on academic dishonesty. Anyone suspected of academic dishonesty will be reported to the Dean of Students Office:

http://www.dso.iastate.edu/ja/academic/misconduct.html

COVID-19 health and safety requirements: Please find up-to-date information regarding the university's safety policy for COVID-19 on https://web.iastate.edu/safety/updates/covid19, in case you will participate in in-person activities with your teammates.

Classroom disruption policy: The class will follow university's Classroom Disruption Policy outlined here www.studentassistance.dso.iastate.edu/faculty-and-staff-resources/disruption.

Accessibility Statement: Iowa State University is committed to assuring that all educational activities are free from discrimination and harassment based on disability status. Students requesting accommodations for a documented disability are required to work directly with staff in Student Accessibility Services (SAS) to establish eligibility and learn about related processes before accommodations will be identified. After eligibility is established, SAS staff will create and issue a Notification Letter for each course listing approved reasonable accommodations. This document will be made available to the student and instructor either electronically or in hard-copy every semester. Students and instructors are encouraged to review contents of the Notification Letters as early in the semester as possible to identify a specific, timely plan to deliver/receive the indicated accommodations. Reasonable accommodations are not retroactive in nature and are not intended to be an unfair advantage. Additional information or assistance is available online at www.sas.dso.iastate.edu, by contacting SAS staff by email at accessibility@iastate.edu, or by calling 515-294-7220. Student Accessibility Services is a unit in the Dean of Students Office located at 1076 Student Services Building.

Other course policies and accommodation: www.celt.iastate.edu/teaching/preparing-to-teach/recommended-iowa-state-university-syllabus-statements/