DS 202 (Fall 2021): Data Acquisition and Exploratory Data Analysis

Delivery method: In-person lectures

Meeting times and location: Tuesday and Thursday 8:00 am – 9:20 am, Gilman 1352

Instructor: Xiongtao Dai
Email: `xdai@iastate.edu`
Office hours: Thursday 1:00 – 2:00 pm, and by appointment
– Physical location: On the turf south of the Parks Library
– Zoom link: https://iastate.zoom.us/j/93009772568?pwd=aWlWWHN0SmN2U011K3BiejhFK3E4Zz09

Teaching assistant: Chengpeng Zeng
Email: `czeng@iastate.edu`
Office hours: Wednesday and Friday 3:00 – 4:00 pm, and by appointment
– Physical location: Snedecor 1205
– Zoom link: https://iastate.zoom.us/j/93009772568?pwd=aWlWWHN0SmN2U011K3BiejhFK3E4Zz09

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: In a typical week, you will

- attend lectures on Tuesday and Thursday;
- post questions on Piazza and attend office hours throughout the week;
- finish a weekly individual quiz by Thursday; and
- hand in your assignment by the end of Saturday.

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 Hofmann. Past recordings of lecture videos will be posted on Canvas to facilitate your learning. The contents covered by the past recordings will be mostly the same as what we cover in class, except for Module 14. The presentation of materials and examples might differ.

Whiteboard notes: Occasional notes made in class will be posted here: https://iowastate-my.sharepoint.com/:o:/g/personal/xdai_iastate_edu/EvzyiJQqTRlEsJIbF3phE6sBrdEuiSVKHdyzVOz5tco1ew?e=y8Bp1g. (Bookmark this page!)

Piazza: We will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the TA and the instructor, you are highly encouraged to post your questions on Piazza. Piazza is integrated into Canvas so you can access it there; alternatively, visit piazza.com/iastate/fall2021/ds202 in your browser. The response time is 24 hours. Email the TA and the instructor 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.

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 learning from the two weekly lectures. The individual quizzes are due by Thursday at 11:59 pm, and will be immediately graded. You have two attempts. The higher score from the two attempts will be kept.

Homework and Labs: There will be 6 homework and 5 lab assignments throughout the semester, except for the midterm and preparation weeks. Homework are to be finished individually, and lab assignments can be finished either individually or (optionally) in a team of two. Feel free to discuss the assignments with anyone in the class, including your classmates, TA, and the instructor. However, you (and your lab mate) 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 on October 21 from 7:30 am to 9:30 am (no lecture on that day). The midterm will be an open-book open-Internet 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 is no final exam.

Final project: The final project is a data analysis project. You can optionally work in a team of two formed among yourself. The project consists of an investigation proposal, a written report, and a short oral presentation. More information is available on Canvas.

Participation: Active participation in class, office hours, and on Piazza are highly encouraged. If you work in a team, you should actively contribute to the group work. If a team member does not respond in a timely fashion, the other team member reserves the right to submit the group work as he/her individual work, giving no credits to the delaying team member.

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)

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

Face masks encouraged: Because of the continuing COVID-19 pandemic, all students are encouraged—but not required—to wear face masks, consistent with current recommendations from the Centers for Disease Control and Prevention. Further information on the proper use of face masks is available at: https://www.cdc.gov/coronavirus/2019-ncov/your-health/effective-masks.html

Vaccinations encouraged: All students are encouraged to receive a vaccination against COVID-19. Multiple locations are available on campus for free, convenient vaccination. Further information is available at: https://web.iastate.edu/safety/updates/covid19/vaccinations Vaccinations may also be obtained from health care providers and pharmacies.

Physical distancing encouraged for unvaccinated individuals: Classrooms and other campus spaces are operating at normal capacities, and physical distancing by faculty, staff, students, and visitors to campus is not required. However, unvaccinated individuals are encouraged to continue to physically distance themselves from others when possible.

Free expression: Iowa State University supports and upholds the First Amendment protection of freedom of speech and the principle of academic freedom in order to foster a learning environment where open inquiry and the vigorous debate of a diversity of ideas are encouraged. Students will not be penalized for the content or viewpoints of their speech as long as student expression in a class context is germane to the subject matter of the class and conveyed in an appropriate manner.

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

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