INFO 2950: Introduction to Data Science

Fall 2023

Course website: https://canvas.cornell.edu/courses/55145

Faculty Name(s): Prof. Allison Koenecke
Faculty Email: koenecke@cornell.edu
Faculty Office Hours: Gates 227, Tuesdays 11:00am-12:00pm
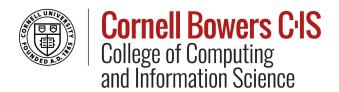**Instructor TA Name & Email:** Roz Thalken, ret85@cornell.edu

Credits and Credit Hour Options: 4.0 Credits, Letter Grade

Prerequisites/Corequisites: Prerequisites include MATH 1710 or equivalent, CS 1110 or CS 1112, or permission of instructor.

Time and Location: This course consists of lectures on Mondays and Wednesdays from 1:25-2:40pm in Ives 305, as well as weekly discussion sections on Fridays during one of the following time blocks: 8-8:50am (Hollister 312), 9:05-9:55am (Snee 1120, Hollister 314), 10:10-11am (Phillips 403), 11:15am-12:05pm (Hollister 362; Snee 1150), 12:20-1:10pm (Hollister 368, Hollister 312), or 1:25-2:15pm (Phillips 213).

Course Description

INFO 2950 is an applied introductory course on the foundations of data science, focusing on using data to identify patterns, evaluating the strength and significance of relationships, and generating predictions using data. Topics covered include the core principles of statistical programming (such as data frames, Python/R packages, reproducible workflows, and version control), univariate and multivariate statistical analysis of small and medium-size datasets, regression methods, hypothesis testing, probability models, basic supervised and unsupervised machine learning, data visualization, and network analysis. Students will learn how to use data to make effective arguments in a way that promotes the ethical usage of data. Students who complete the course will be able to produce meaningful, data-driven analyses of real-world problems and will be prepared to begin more advanced work in data-intensive domains.

Course Objectives/Student Learning Outcomes

By the end of the course, a student will be able to:

- Identify, acquire and clean project-relevant data from diverse sources. Analyze data fitness for purpose and evaluate solutions for data quality issues such as missing records.
- Use data ethically, including by applying best practices for data sourcing, documentation, and privacy preservation.
- Create data science projects using reproducible methods, common tools, and standard software libraries.
- Apply methods to transform tabular data through operations such as filtering, aggregation, and pivoting.
- Analyze data-generating processes via simulation.
- Apply exploratory data analysis through visualization, summary statistics and outlier identification. Use exploratory results to generate research hypotheses.
- Identify, measure, and evaluate the significance of patterns in large datasets.
- Select and apply common learning algorithms to make predictions about unobserved data.
- Create compelling, reliable arguments informed by data.

Course Materials

Course presentations and readings will be posted on Canvas in the Modules section.

Method of Assessing Student Achievement

- **Basis of Grade Determination:**

Assignment/Assessment	Percentage of Grade/ Points
Homework	30%
Final Project Checkpoints	15%
Midterm Exam	10%
Final Project	25%
Final Exam	15%
Professionalism	5%

All homework assignments will count equally regardless of their individual grading. For convenience in grading, we may grade one homework out of a total of 35 points and another out of 108, but both will count equally).

- **Grading Scale:** This class adheres to Cornell's grading scale:

A+	98-100%	4.3
A	93-97%	4.0
A-	90-92%	3.7
B+	88-89%	3.3
B	83-87%	3.0
B-	80-82%	2.7
C+	78-79%	2.3
C	73-77%	2.0
C-	70-72%	1.7
D	60-69%	1.3
F	Below 60%	0.0

Collaboration and academic integrity

All students should abide by the Cornell University Code of Academic Integrity, and all writing submitted should be one's own writing. While discussing course concepts with other students is highly encouraged, plagiarism (including use of ChatGPT) will result in zero credit and/or a referral to the Office of Judicial Affairs. Please reach out if additional questions arise on what is or is not permitted.

Students with Disabilities

Your access in this course is important to me. Please give me your Student Disability Services (SDS) accommodation letter early in the semester so that we have adequate time to arrange your approved academic accommodations. If you need an immediate accommodation for equal access, please speak with me after class or send an email message to me and/or SDS at sds_cu@cornell.edu. If the need arises for additional accommodations during the semester, please contact SDS. You may also feel free to speak with Student Services at Cornell Tech who will connect you with the university SDS office.

If you have, or think you may have a disability, please contact Student Disability Services for a confidential discussion: sds_cu@cornell.edu, 607-254-4545, sds.cornell.edu.
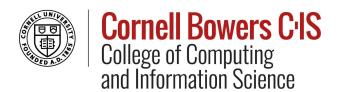
You must request your [SDS accommodation letter](#) **no later than the add/drop deadline for the semester**.

- **Students currently registered with SDS:** Once you request your accommodation letter and it is approved by SDS, it will be emailed to both you and me. Processing time can be up to 48-hours.
- **Students not registered with SDS:** The registration process for new accommodations can take up to three weeks. Once you are approved by SDS for accommodations, you will be able to request your accommodation letter for this course.
- **If you are approved for accommodations later in the semester:** you must request your accommodation letter as soon as possible.

Students with Exam Accommodations

In addition to requesting your accommodation letters, this course will be participating in the **Alternative Testing Program (ATP).** All exams will be centrally managed and supported by the ATP Testing Coordinator in the Office of Student Disability Services and accessible in your SDS student portal. **ATP support includes:**

- **Scheduling of accommodated exams:**
  - *Daytime exams:* Students are expected to start their accommodated exam at the same time as the general exam administered during class time.
    - Students with extended time accommodations who have an academic conflict (i.e. a course immediately following this class) may select to take their accommodated exam at 8 a.m. and/or 5 p.m. on the same day as the general exam. More information is available here at [sds.cornell.edu/atp](#)
  - *Evening prelims:* Evening prelim exams will begin promptly at 6:30 p.m.
- **Notification of accommodated exam logistics:**
  - All exam logistics are managed by the ATP and will ONLY be communicated to you via email from [sds-testing@cornell.edu](#) and accessible in your SDS student portal. **Please do not contact the instructors with questions about exam logistics, as we will not be able to answer them.**
  - *10-days prior to the exam date:* ATP will automatically send an email with the exam date, time, location.
  - *48-hours prior to the exam date:* ATP will send a reminder email about the exam.

- **Coordination of make-up exams** (i.e., for students who have been granted prior permission by me to take the exam on a day other than the scheduled date of the main exam) will be handled by the instructors – contact us at info2950-instructors@cornell.edu. The ATP will not be involved in the logistics for any make-up exams. If you miss your scheduled accommodated exam, you should notify me, not the ATP.

For students with other academic accommodations (not testing-related), please follow up with the instructors at info2950-instructors@cornell.edu to discuss the necessary logistics of your accommodation(s).

Mental Health & Well-being

Your health and wellbeing are important to me. There are services and resources at Cornell designed specifically to bolster undergraduate, graduate, and professional student mental health and well-being. Remember, your mental health and emotional well-being are just as important as your physical health. If you or a friend are struggling emotionally or feeling stressed, fatigued, or burned out, there is a continuum of campus resources available to you: https://mentalhealth.cornell.edu/get-support/support-students. Help is also available any time day or night through Cornell's 24/7 phone consultation (607-255-5155). You can also reach out to me, your college student services office, or Cornell Health for support.

Attendance Policy

Attendance is required for lectures and discussion sections; attendance will be tracked. If you cannot attend class due to extenuating circumstances (e.g., illness, religious observances, etc.), please reach out to the instructors ahead of class via email.