

IST 387/687 Intro to Applied Data Science

Welcome to Intro to Applied Data Science! This course introduces you to the full data science workflow – from managing and transforming data to analyzing it, visualizing it, and building machine learning models – all using R, a powerful programming language built for statistics and data science. Through hands-on projects and real-world problem-solving, you'll develop the technical skills, ethical judgment, and practical intuition to know when and how data science can make a difference. No prior programming experience? No problem – you'll build these skills gradually throughout the course.

Class Meetings

Each week the entire class will meet on **Thursdays, 9:30AM - 10:50AM** in **Eggers Hall 010** for lecture. This lecture will be led by **Prof. Christopher Dunham** (cmdunham@syr.edu).

In addition to the weekly large lecture, everyone will enroll in one lab section which is comprised of either all undergraduate students (i.e., taking IST 387) or all graduate students (i.e., taking IST 687). The following table contains information for each lab section.

Lab	Time & Location	Instructor
387 M002	Tu 11:00AM - 12:20PM; Huntington 101	Yuheun (Rachel) Kim; ykim72@syr.edu
387 M003	Tu 9:30AM - 10:50AM; Hinds 011	Bo Zhang; bzhang49@syr.edu
387 M004	Tu 11:00AM - 12:20PM; Hinds 011	Bo Zhang; bzhang49@syr.edu
687 M002	Th 11:00AM - 12:20PM; Hinds 011	Chris Dunham; cmdunham@syr.edu
687 M003	Th 12:30PM - 1:50PM; Hinds 021	Yuheun (Rachel) Kim; ykim72@syr.edu

Attendance is required for both the large lecture and lab meeting. Each lecture will include a short quiz that can only be completed if you are in the lecture room. Each lab meeting will feature a hands-on activity for credit that will only count if you attend the lab. Labs are not simply recitations of lecture materials; they provide an opportunity for you to practice real data science skills with guidance from a skilled instructor. More information about these assignments is provided later in the syllabus.

Office Hours

My regular office hours are **Thursdays, 1:00PM - 2:00PM** in **Hinds Hall 309**. There is no need to make an appointment for regular office hours; just drop in. I understand that this time may not work for everyone, so alternative times can be arranged. The easiest way to schedule an alternative meeting time is to email me a few days in advance of when you want to meet and include a list of days and times that you will be available. Note that office hours aren't just for problems with homework or feedback on an exam; sometimes you just want to chat or seek guidance about classes to take in a future semester. If you think you may need a letter of recommendation from me in the future, office hours are a good way for me to get to know you. **Never hesitate to drop into my office hours or reach out for an appointment!**

Course Format

This course is designed with a high-structure approach, meaning that your learning is supported by consistent, required practice before, during, and after every class meeting. Research on learning shows that effort distributed across many smaller activities — rather than crammed before an exam — leads to deeper understanding and better outcomes. A high-structure approach is especially valuable in this class because the programming skills we learn each week build directly on what came before, and with consistent practice, what feels unfamiliar at first will start to feel like second nature.

The course consists of one larger class plus one smaller lab section each week. During lecture, we will explore key concepts and examine their application in R. To get the most out of lecture, you are expected to complete the assigned readings beforehand and practice the coding examples from the book. Each lecture will feature a **short quiz** that tests your knowledge of key concepts from both the readings and the lecture. The more prepared you are, the easier — and more enjoyable — lecture will be.

Weekly lab sessions give you dedicated time to practice R coding, ask questions, and work through real data science problems with guidance from a skilled instructor. A **lab assignment** is due the day of each lab meeting, and you must attend to receive credit. These sessions are not simply recitations of lecture material; they are where you build hands-on skills. Weekly **homework assignments** extend what you practiced in lab and are due the day before the next lecture.

There will also be an **exam** covering lecture material and R coding — paper based, closed book, no computer. Coding questions are designed so that a computer isn't necessary. Finally, a **group project** toward the end of the semester gives you a chance to apply everything you have learned to a real-world data problem, where your task is to understand the domain and the data available in order to extract meaningful insight.

Audience

This course is intended for both undergraduate and graduate students with an interest in learning about data science and how to use data to solve real-world problems. No prior knowledge of data science or experience with programming languages is required. In fact, most students who take this course are new to programming.

Credits

3 credit hours

Learning Objectives

After taking this course, students will be able to:

1. Identify a problem and the data needed for addressing the problem
2. Perform basic computational scripting using R and other optional tools
3. Transform data through processing, linking, aggregation, summarization, and searching
4. Organize and manage data at various stages of a project lifecycle
5. Determine appropriate techniques for analyzing data

Shared Competencies

TBD ... include note that this is for undergraduates only

Textbook (required)

The primary textbook for this class is *Data Science for Business with R (2021)*, by Jeffrey S. Saltz and Jeffrey M. Stanton (ISBN: 978-1544370453). The book is available in paperback and electronic version at the University bookstore and on Amazon.com:

<https://www.amazon.com/Data-Science-Business-Jeffrey-Saltz/dp/1544370458>.

Software

We will use the R programming language (<https://cran.r-project.org/>) in this course, via RStudio Desktop (<https://posit.co/download/rstudio-desktop/>) or Posit Cloud (<https://www.posit.cloud>). These are free software/services. Be sure you have a functioning laptop with access to R and RStudio on the first day of class.

Grading

Assessment	Notes	Grade
In-class quizzes	Each of the 12 lectures will feature a 4-6 question multiple-choice quiz based on readings and the lecture. You must be present to take the quiz. <i>Your lowest two quiz scores will be dropped.</i>	10 * 1% = 10%
In-class lab assignments	While you are in your weekly lab session, you will work on an in-class assignment. At the end of each lab session, you will submit a copy of your work to Blackboard. You will receive full credit for your submission if you have given a reasonable effort (i.e., it is okay if the code was not perfect; your lab is for learning R coding concepts). I will not accept late lab assignments; however, <i>your lowest two lab grades will be dropped.</i> Feedback on lab submissions will be minimal since these are designed to be completed in class with the help of your lab instructor.	10 * 1% = 10%

Homework assignments	<p>Homework assignments are designed for you to practice the necessary skills in carrying out data processing, analysis, and management tasks. The code you create for the homework assignments will be instrumental in helping you to complete the exam and the project successfully. I will not accept any late homework assignments; however, <i>your lowest two homework grades will be dropped.</i></p> <p>Homework assignments will be graded according to the following rubric: 0: no submission, 1: most key questions unanswered or incorrect, 2: most key questions answered correctly but with some errors or omissions, 3: all questions are complete and correct or only very minor errors or omissions. You will receive written feedback on any errors that affect your grade.</p>	<p>$10 * 3\% = \mathbf{30\%}$</p>
Exam	<p>The exam is designed to take 80 minutes to complete and will be administered in your lab section (see schedule for date). No outside materials will be allowed for this exam (closed book, no notes, no computer). If you have a documented accommodation for extended time or an alternative testing location, please register in advance with the testing center.</p>	<p>25%</p>
Group Project	<p>This is a predictive analytics effort that will be completed during the later part of the course, after the exam. Groups will be assigned by your lab instructor. The project will include both a formal report and a presentation. More information will be made available around the time of the exam.</p>	<p>25%</p> <p>Presentation: 10% + Report: 15%</p>

All work will be graded on the scale as specified for the component. All individual grades are summed at the end of the semester, and the final letter grade is determined using the following scale. Note: A grade of D is not available to graduate students.

Grade	Points	Grade	Points	Grade	Points	Grade	Points
A	93-100	B+	87-89	C+	77-79	D	60-69
A-	90-92	B	83-86	C	73-76	F	0-59
		B-	80-82	C-	70-72		

Course Specific Policies

- It is unethical to allow some students additional opportunities, such as extra credit assignments, without allowing the same options to all students
- Lab assignments are due the day of the lab without exception; however, the two lowest lab grades will be dropped
- Students are expected to complete their lab work *in lab*; you must attend your lab section to receive credit for your weekly lab assignment
- Homework assignments are due the day before the following lecture without exception; however, the two lowest homework grades will be dropped
- There are no make-up quizzes, but the lowest two quiz grades will be dropped
- The drop policy — two lowest scores dropped in each of quizzes, labs, and homework — is built-in flexibility for when life gets in the way. Rather than navigating excused absence requests, you can simply use your drops when you need them and move on.

How to Succeed in This Course

- Complete your weekly assigned readings before lecture
- Bring your laptop to every lecture and lab session
- Take notes during lecture and ask questions
- Start assignments early and never hesitate to ask for help
- Practice coding often and build a code repository
- Cultivate an interest in and passion for data science!

Use of Artificial Intelligence in Course

Based on the specific learning outcomes and assignments in this course, artificial intelligence is permitted on take-home assignments and the final project. See each assignment's instructions for more information about what artificial intelligence tools are permitted and to what extent, as well as citation requirements. If no instructions are provided for a specific assignment, then no use of any artificial intelligence tool is permitted. Any AI use beyond that which is detailed in course assignments is explicitly prohibited except when documented permission is granted.

Extra Credit Opportunity (1pt)

As soon as you join the course, send an email to your lab instructor and cc me (unless I *am* your lab instructor). Introduce yourself—e.g., your general background, your major/program and year of study—and share your reasons for taking the course. Share any prior experience with coding—whether it be in R or Python or anything else. Mention if you have a conflict that would prevent you from attending my regular office hours. Confirm that you have downloaded R and RStudio (or signed up for an account with Posit Cloud).

Course Schedule

Week / Date of Lecture	Lecture Topic	Required Readings	Focus of Lab	Assignments
01 / Aug 27	Data Science Overview	Ch 1: Begin at the Beginning With R	Using R	HW 1 - Assigned
02 / Sep 03	Using R to Manipulate Data	Ch 2: Rows & Columns Ch 3: Data Munging	Working with data frames	HW 2 - Assigned
03 / Sep 10	Descriptive Statistics & Functions	Ch 4: What's my Function? Ch 5: Beer, Farms & Peas	Descriptive statistics & writing functions	HW 3 - Assigned
04 / Sep 17	Sampling & Inferential Statistics	Ch 6: Sample in a Jar	Sampling	HW 4 - Assigned
05 / Sep 24	Connecting with External Data Sources	Ch 7: Accessing Data (storage wars)	Using JSON	HW 5 - Assigned
06 / Oct 01	Introduction to visualization	Ch 8: Pictures vs numbers	Using ggplot2	HW 6 - Assigned
07 / Oct 08	Working with map data	Ch 9: Map Mash-Up	Using ggplot2 for maps	HW 7 - Assigned
08 / Oct 15	Linear Modeling (LM)	Ch 10: Lining Up Our Models	Working with LM	HW 8 - Assigned
09 / Oct 22	Supervised Learning	Ch 11: What's your vector, Victor?	Support vector machines and trees	HW 9 - Assigned
10 / Oct 29	Unsupervised Learning & Ethics in Data Science	Ch 12: Hi Ho, Hi Ho-Data Mining We Go	Association rules mining	HW 10 – Assigned
Exam: Nov 05	<i>No lecture (study)</i>	-	<i>Exam in lab</i>	-
11 / Nov 12	Interactive Apps & Deep Learning	Ch 14: Shiny Web Apps	Shiny	HW 11 – Assigned
12 / Nov 19	Text Mining	Ch 13: Word perfect	Word clouds and sentiment analysis	HW 12 – Assigned
	<i>Thanksgiving Break</i>	-	-	
13 / Dec 03	<i>No Lecture (project)</i>		Presentations	-

Note: Fall break on Tue, Oct 13 means IST-387 students will not meet for lab after week 7's lecture. You will still be responsible for submitting your lab on time. Seek help early if needed.