Introduction to IS 381

Statistics and Probability with R

Jason Bryer, Ph.D.

Fall 2025

Agenda

- About your instructor
- Syllabus
- Class meetups
- Course Schedule
- Assignments (how you will be graded)
  - Participation
  - Labs
  - Data Project
  - Exams
- Software
  - Using R Markdown

A little about me...

- Assistant Professor at CUNY in Data and Information Sciences

- Principal Investigator for a Department of Education Grant to develop and test the Diagnostic Assessment and Achievement of College Skills (www.DAACS.net)

- Authored over a dozen R packages including:

  - likert
  - ShinyQDA
  - DTedit
  - login

- Specialize in propensity score methods. Three new methods/R packages developed include:

  - multilevelPSA
  - TriMatch
  - PSAboot

Also a Father...

Runner...

And photographer.

Syllabus

Syllabus and course materials are here: https://fall2025.is381.net

The site is built using Quarto and hosted on Github. Each page of the site has a "Edit this page" link at the bottom right, use that to start a pull request on Github.

We will use Brightspace primary for submitting assignments only. Please submit a PDF.

PDFs are preferred for the homework as there is some LaTeX formatting in the R markdown files. The `tinytex` R package helps with install LaTeX, but you can also install LaTeX using MiKTeX (for Windows) and BasicTeX (for Mac) See this page for more information: https://fall2025.is381.net/course-overview/software/

Assignments / Grading

Labs (50%)

Data Project (25% Total; Proposal 5%, Final Presentation 20%)

Final Exam (15%)

Participation (10%):

- DAACS SRL (https://cuny.daacs.net) and Google Form (only once, at the beginning of the semester)
- Weekly One-Minute Papers

Meetups

We will have meetups on Monday evenings at 6:30pm to 7:30pm.

Meetups will be recorded and made available the next day on the course website https://fall2025.IS381.net.

Though attending live is not strictly required, **I expect everyone to watch the lectures during the week.** I use the class meetups to convey important information and announcements. Very often I will cover some topics not in the textbook. Students who attend the meetups tend to do well on the assignments.

One Minute Papers - Complete the one minute paper after each Meetup (whether you watch live or watch the recordings). It should take approximately one to two minutes to complete. This allows me to 1) verify you have attended/watch the meetup and 2) get feedback about what you learned and what you may still be unclear.

Please note: *Students who participate in this class with their camera on or use a profile image are agreeing to have their video or image recorded solely for the purpose of creating a record for students enrolled in the class to refer to, including those enrolled students who are unable to attend live. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate exclusively using the "chat" feature, which allows students to type questions and comments live.* Click here for CUNY's camera use policy
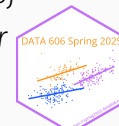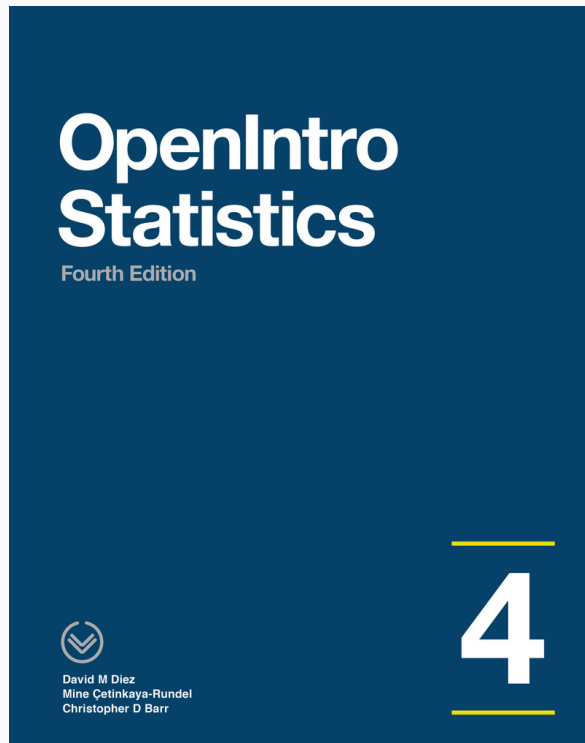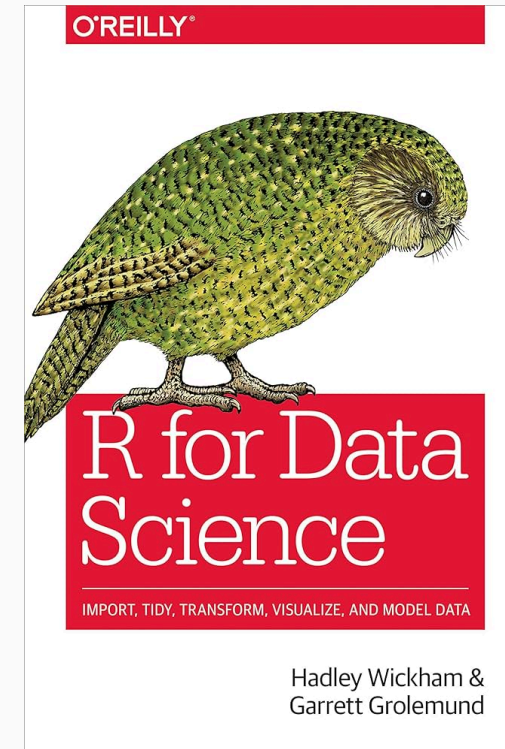
Schedule

Week	Topic	Start	End
1	Introduction to R and RStudio	August 25, 2025	August 31, 2025
2	R coding basics	September 01, 2025	September 07, 2025
3	Data (importing and structure)	September 08, 2025	September 14, 2025
4	Data visualization with ggplot2	September 15, 2025	September 21, 2025
5	Reshaping Data	September 22, 2025	September 28, 2025
6	Probability	September 29, 2025	October 05, 2025
7	Distributions	October 06, 2025	October 12, 2025
8	Foundation for inference	October 13, 2025	October 19, 2025
9	Central limit theorem	October 20, 2025	October 26, 2025
10	Inference for proportions	October 27, 2025	November 02, 2025
11	Inference for two-way tables	November 03, 2025	November 09, 2025
12	Inference for numerical data	November 10, 2025	November 16, 2025
13	Analysis of variance	November 17, 2025	November 23, 2025
	NO CLASS - Thanksgiving	November 24, 2025	November 30, 2025
14	Correlation	December 01, 2025	December 07, 2025
15	Wrap up / Final Presentations	December 08, 2025	December 14, 2025

Textbooks

Diez, D.M., Barr, C.D., & Çetinkaya-Rundel, M. (2019). *OpenIntro Statistics (4th Ed).*

This will be our primary textbook for most of the semesters. Our goal is to cover all the chapters.

Wickham, H., Cetinkaya-Rundel, M., & Grolemund, G. (2023). R for Data Science (2nd ed).

Communication

- Slack Channel: https://cuny-msds.slack.com

    - Click here to join the group

- Email: jason.bryer@cuny.edu

- Phone/Zoom: Please email to schedule a time to meet.

- Office hours by appointment.

Software

This is an applied statistics course so we will make extensive use of the R statistical programming language.

Install R and RStudio on your own computer. I encourage everyone to do this at some point by the end of the semester. I have instructions on the course website here: https://fall2025.data606.net/course-overview/software/

You will also need to have LaTeX installed as well in order to create PDFs. The `tinytex` R package helps with this process:

```r
install.packages('tinytex')
tinytex::install_tinytex()
```

Use of Artificial Intelligence (AI)

First, AI is a marketing term. I prefer to be more specific regarding what we are doing:

1. Machine Learning (ML) - This course, along with IS382, will provide the foundations for how ML algorithms work. Generally speaking, the goal is to predict some known (and sometimes unknown in the case of unsupervised learning models) outcome.

2. Large Language Models (LLM) - This is often what people mean when they say AI. This includes producsts like ChatGPT, Antrhopic, Google Gemini, etc. LLMs generate text, images, videos, etc. from a prompt.

The goal of this course is for *you to develop the foundation knowledge and skills to do statistics*. Using chat bots to do the assignments subverts this goal. **The content generated by LLMs is often wrong!** If you use LLMs to assist in completing the assignments, **you must include the prompt and response in your submission**.

Next steps...

Before August 31st:

- Complete this Google form: https://forms.gle/sXrP9X7ZK2dXv2Lu6
- Go to https://cuny.daacs.net and complete the self-regulated learning assessment
- Join the Slack channel

Then:

- Start *Intro to R Lab* (due August 31st)

Good luck with the semester!

jason.bryer@cuny.edu
cuny-msds.slack.com
@jbryer
@jbryer@vis.social
fall2025.is381.net