

Introduction to DATA 606

Statistics & Probability for Data Analytics

Jason Bryer, Ph.D., Angela Lui, Ph.D., and Christian Martinez

Fall 2026

Agenda

- About your instructor
- Syllabus
- Class meetups
- Course Schedule
- Assignments (how you will be graded)
 - Participation
 - Labs
 - Data Project
 - Exams
- Software
 - The DATA606 R Package
 - Using R Markdown

A little about me...

- Assistant Professor at CUNY in Data Science and Information Systems
- 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](#)



Built a van...



Runner...



And photographer.



A little about Angela...



HUNTER



UNIVERSITY AT ALBANY
State University of New York

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

**CUNY School of
Professional Studies** | **CUNY**



Diagnostic Assessment & Achievement of College Skills



Teaching Experience

- Introduction to Statistics in Social Sciences
- Special Issues in Testing
- Evaluation
- Motivation in Education
- Introduction to the Psychological Processing of Schooling
- Educational Psychology in Adolescent Development

Homeowner



Christian Martinez



- Adjunct Lecturer at the City University of New York
- Founder and Director of the NYC Open Data Lab.
- My work centers on data science, reproducible research, open-source software, and helping people use public data more effectively.
- Created the nycOpenData R package and an expanding ecosystem of OpenData packages that make government open data easier to access and analyze directly in R.
- My open-source and community work was recognized with the inaugural 2026 Posit Impact Award in Open Source Contribution.

When I'm not doing any of this fun stuff, I'm on the handball court—always ready for a new challenger.



Syllabus and course materials are here: <https://fall2026.data606.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 or link to the built HTML (e.g. Rpubs, [Github](#))

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://fall2026.data606.net/course-overview/software/>

Meetups

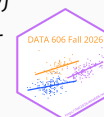
We will have meetups on Monday evenings at 8:00pm.

Meetups will be recorded and made available the next day on the [course website](#).

Though attending live is not strictly required, **We expect everyone to watch the lectures during the week**. I use the class meetups to convey important information and announcements. Very often we 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](#)



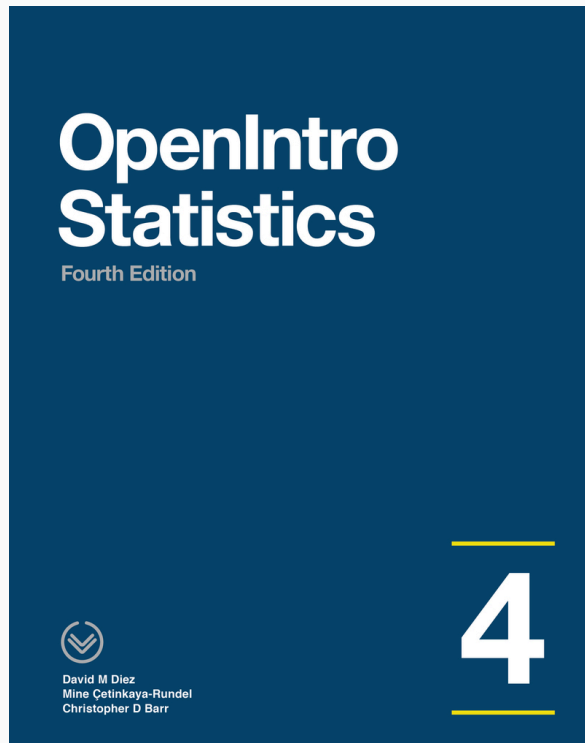
Schedule

Start	End	Topic
Monday, February 02, 2026	Sunday, February 15, 2026	Chapter 1 - Intro to Data, R, and RStudio
Monday, February 16, 2026	Sunday, February 22, 2026	Chapter 2 - Summarizing Data
Monday, February 23, 2026	Sunday, March 01, 2026	Chapter 3 - Probability
Monday, March 02, 2026	Sunday, March 08, 2026	Chapter 4 - Distributions
Monday, March 09, 2026	Sunday, March 15, 2026	Chapter 5 - Foundation for Inference
Monday, March 23, 2026	Sunday, April 05, 2026	Chapter 6 - Inference for Categorical Data
Monday, April 06, 2026	Sunday, April 12, 2026	Chapter 7 - Inference for Numerical Data
Monday, April 13, 2026	Sunday, April 19, 2026	Chapter 8 - Linear Regression
Monday, April 20, 2026	Sunday, April 26, 2026	Chapter 9 - Logistic Regression
Monday, April 27, 2026	Sunday, May 03, 2026	Chapter 9 - Multiple Regression
Monday, May 04, 2026	Sunday, May 10, 2026	Intro to Bayesian Analysis
Monday, May 11, 2026	Sunday, May 17, 2026	Final Exam



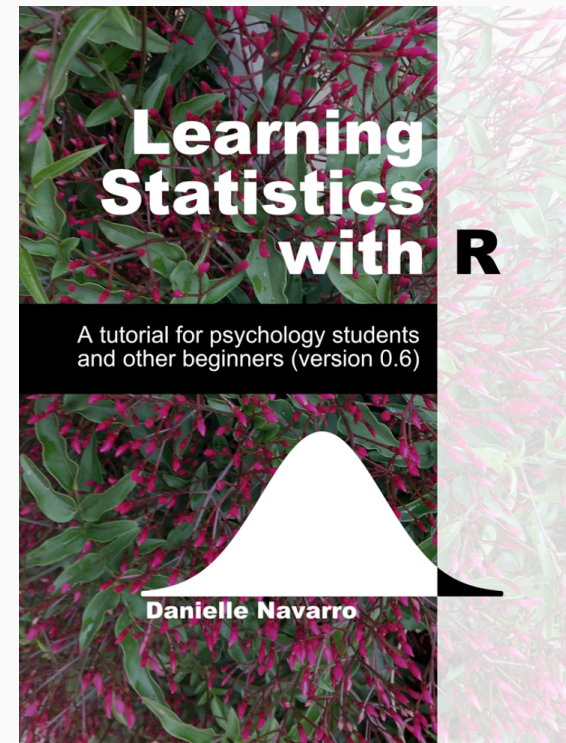
Diez, D.M., Barr, C.D., & Çetinkaya-Rundel, M. (2019). *OpenIntro Statistics (4th Ed)*.

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



Navarro, D. (2018, version 0.6). *Learning Statistics with R*

This textbooks has a chapter on Bayesian analysis that we will use at the end of the semester.



Assignments

- Participation (10%)
 - DAACS
 - One Minute Papers
- Labs (35%)
 - Labs are designed to introduce to you doing statistics with R.
 - Answer the questions in the main text as well as the "On Your Own" section.
- Data Project (30%)
 - This allows you to analyze a dataset of your choosing. Projects will be shared with the class. This provides an opportunity for everyone to see different approaches to analyzing different datasets.
 - For the final presentation, **you will be required to present during one of several 90 minute sessions.**
- Exams
 - Midterm (10%)
 - Final exam (15%)

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 products like ChatGPT, Anthropic, 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.**

Communication

- Slack Channel: <http://cuny-msds.slack.com>
 - [Click here to join the group](#)
- Email: jason.bryer@cuny.edu, angela.lui@cuny.edu, and c.martinez0@outlook.com
- Phone/Zoom: Please email to schedule a time to meet.
- Office hours by appointment.

Why R?

There are many languages data scientists use. **R** is specifically designed for statistics. We will leverage many R packages that are specifically designed to conduct, teach, and communicate statistical analysis.

To be a well rounded data scientists, I believe you need to have experience in both R and Python. For this course we will use R. Future courses will focus on Python.



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://fall2026.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:

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


The **DATA606** R package contains many data sets and functions we will use throughout the semester. It also has a `startLab` function that will copy each of the labs to your current working directory. Use the following commands to install the package (only necessary once per R installation):

```
remotes::install_github('jbryer/DATA606')
```

To start the first lab...

```
DATA606::startLab('Lab1')
```

This will copy the R markdown file and any supporting files to your current working directory. Use the "Knit" button in R Studio to build a PDF of the document.

Next steps...



Before Sunday (September 6th):

- Complete this Google form: <https://forms.gle/7QKPCx3YPFyHJnKQ6>
- Go to <https://cuny.daacs.net> and complete the self-regulated learning assessment
- [Join the Slack channel](#)

Then:

- Start Lab 1 (due September 6th)



Good luck with the semester!

✉ jason.bryer@cuny.edu

✉ angela.lui@cuny.edu

✉ c.martinez0@outlook.com

⚙ http://cuny-msds.slack.com

🐙 @jbryer

✉ @jbryer@vis.social

🔗 fall2026.data606.net