Instructor:	Jonathan "Nate" Wells	**Email:**	wellsj@reed.edu
Classroom:	Library 389	**Office:**	Library 392
In-person Office Hours:	M 3-4pm, W 10-11am, F 10-11am		
Virtual Office Hours:	T 2-3pm		
Zoom Link:	https://zoom.us/my/wellsj392		

Course Description: This course is an overview of modern approaches to analyzing large and complex data sets that arise in a variety of fields from biology to marketing to astrophysics. The most important modeling and predictive techniques will be covered, including regression, classification, clustering, resampling, and tree-based methods. There will be several projects throughout the course, which will require significant programming in R.

Prerequisites: MATH141, or Instructor Consent.

Distribution Requirements: This course can be used towards your Group III, "Natural, Mathematical, and Psychological Science," requirement. It accomplishes the following learning goals for the group:

> Use and evaluate quantitative data or modeling, or use logical/mathematical reasoning to evaluate, test or prove statements; Given a problem or question, formulate a hypothesis or conjecture, and design an experiment, collect data or use mathematical reasoning to test or validate it; Collect, interpret and analyze data

This course does not satisfy the "primary data collection and analysis" requirement.

Textbooks:

- (Primary) *An Introduction to Statistical Learning*, 2nd Edition by James, Witten, Hastie, and Tibshirani. Free online access to the textbook is available via SpringerLink
- (Secondary) *Applied Predictive Modeling*, 1st Edition by Kuhn and Johnson. We will use select content from this text to supplement *ISL*. Free online access to the textbook is available at SpringerLink

Course Resources: The following web-based resources will be used for communicating class information:

- Slack `reedmath391fall2021.slack.com` (*announcements, discussions, direct messaging*)
- Course Website `https://Reed-Stat-Learning-Fall-2021.github.io` (*documents, a daily schedule, assignments*).
- GitHub Classroom `https://classroom.github.com/classrooms/88862064-reed-stat-learning-fall-2021-classroom` (*homework submission*)

Technology: You are encouraged to bring a personal computer to class each day for notetaking and live coding. Access to a computer with webbrowser will be required for homework completion and submission. Computing & Information Services offers programs for long-term laptop loan: `https://www.reed.edu/cis/facilities/student-technology-equipment-program.html`

We will make very frequent use of the R programming language to create statistical models, run simulations, and implement stat learning algorithms. All homework will be completed using the RStudio IDE. R and RStudio are free to use, and can either be installed locally on your computer, or can be accessed using the Reed RStudio Server: `https://rstudio.reed.edu/`

Throughout the term, we will use GitHub to manage and submit assignments. GitHub is a hosting service to house Git-based projects online, and is designed to assist with version control and collaboration on big projects. `https://github.com/`

Communication: If you would like to contact me, I can most easily be reached via Slack message weekdays between 8am and 6pm. While I try to answer messages as soon as possible, in some cases, I may not be able to respond until the following school day. If you'd prefer to talk live, send me a message and we can schedule a time to chat on zoom.

Course Outcomes: By the end of the course, a student should be able to:

1. Articulate and compare the different philosophical approaches to prediction, statistical inference, classification, and clustering.
2. Create valid statistical models, perform data analysis using software, and communicate results in non-technical language using reproducible methods in order to answer a particular research question.
3. Implement simulation and randomization algorithms in order to demonstrate and assess properties of statistical models.
4. Assess and compare the performance of a variety of statistical models, and select appropriate models according to suitable criteria.
5. Apply statistical learning techniques to real-world data and problems.
6. Justify and describe properties of particular statistical learning methods by appealing to mathematical theory.