

# Applied Statistical Modeling and Inference: Bayesian APSTA-GE 2123, Spring 2020 (last seven weeks)

Instructor: Ben Goodrich ( [benjamin.goodrich@columbia.edu](mailto:benjamin.goodrich@columbia.edu) )  
*Verify that the date below is recent! Syllabus subject to change!*

March 29, 2020

Course Time: Mondays 8:55-12:00 officially in room 307 of building 194M but online due to the coronavirus  
Teaching Assistant: Christina Ho

Office hours: Online by appointment

## Course Description

An introduction to Bayesian statistical methods with applications to the social sciences. The primary software used will be Stan, which students do not need to be familiar with in advance. We will access the Stan library via R, so some experience with R is necessary.

## Prerequisites

Steinhardt students presumably have taken an introduction to probability in the fall and a course on frequentist inference in the first seven weeks of the spring semester. Any non-Steinhardt student interested in taking this course should have taken similar courses, which would cover basic probability (with calculus), linear regression, and some computer programming in R.

## Grading

Grading will be based  $\frac{1}{6}$  on class participation,  $\frac{1}{2}$  on homework assignments, and  $\frac{1}{3}$  on the final project. Asking one (public) conceptual question on CampusWire per week (separate from any specific questions you have about your homework) or substantially discussing another student's question is considered good class participation.

## CampusWire

CampusWire is a beta version of a tool that is available at <https://campuswire.com/c/GB57F2ED0> that you will be emailed an invitation to before the semester starts. Rather than emailing questions directly to the professor or TAs, you should post on CampusWire. That way, other students can answer your question, benefit from an answer that the professor or TA provides, ask follow-up questions, etc. There is also Reddit-style upvoting and the statistics collected by CampusWire go into the participation portion of your grade. Students should not ask questions in office hours that have not first be posted on CampusWire.

If your question pertains to an ongoing homework assignment, your grades, or similar, then you should click on the option to make your post only visible to "Instructors and TAs". Otherwise, you should post to "Everyone in the class" and avoid direct messaging the instructor and TAs. There is an option to post in Stealth Mode, in which case no

one will know it was you that asked the question, but doing so obviously cannot count toward the class participation component of your course grade.

There are Notification options under User Settings (in the bottom left) where you can control how often you receive emails about activity on CampusWire. You can turn some or all of those off but are still responsible for reading posts by other students.

## Textbook

*Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, second edition, by Richard McElreath, published by CRC Press in 2020, which can be accessed online from <https://ebookcentral-proquest-com.proxy.library.nyu.edu/lib/nyulibrary-ebooks/detail.action?docID=6133700>. McElreath's [website](#) has links to his course videos and other material.

## Coronavirus

The class will take place entirely online in order to reduce the spread of coronavirus. Although we will undoubtedly have to adapt as the quarter moves along, the initial plan is to run a “flipped” class where everyone is expected, in advance, to do the reading and watch the corresponding videos that Richard McElreath made to go along with the second edition of his textbook. These videos are somewhat informal and were intended primarily for graduate students in biology and similar fields. I want to primarily use our online class time to work through problems that push beyond that so that the course is appropriate for graduate students in applied statistics. The biggest difference in emphasis is that we will focus on writing code in the Stan language, whereas the textbook emphasizes writing R code that utilizes McElreath's rethinking package (whose `ulam` function generates Stan code).

## Outline

### Week 1: Introduction and Probability

- *Statistical Rethinking*, chapters 1 and 2. If the hardcopy of your book has not arrived yet, you can read these two chapters [here](#).
- McElreath videos: [The Golem of Prague](#) and [Garden of Forking Data](#)
- Grant Sanderson video: <https://youtu.be/HZGCoVF3YvM>

### Week 2: Modern Markov Chain Monte Carlo

- *Statistical Rethinking*, chapter 9
- McElreath video: [Markov Chain Monte Carlo](#)

### Week 3: Linear Models with `rstanarm`

- *Statistical Rethinking*, chapters 4, 5, and 7
- McElreath videos: [Geocentric Models](#), [Spurious Waffles](#), [Ulysses' Compass](#)

### Week 4: Generalized Linear Models with `rstanarm`

- *Statistical Rethinking*, chapters 10 and 11
- McElreath videos: [Maximum entropy and GLMs](#) and [God Spiked the Integers](#)

## Week 5: Generalized Linear Models with brms

- “brms: An R Package for Bayesian Multilevel Models using Stan” by Paul-Christian Bürkner (2017), *Journal of Statistical Software*, 80:1 [Link](#)
- “brms Reference Manual” by Paul-Christian Bürkner. [Link](#)
- *Statistical Rethinking with brms, ggplot2, and the tidyverse* by A. Solomon Kurz (2018) [Link](#) chapters 9 and 10 (note that these currently refer to the FIRST edition of *Statistical Rethinking* but it is the same content as in the previous week)

## Week 6: Multilevel Models with rstanarm and brms

- *Statistical Rethinking*, chapters 13 and 14
- McElreath videos: [Multilevel Models](#), [Multilevel Models 2](#), and [Adventures in Covariance](#)
- *Statistical Rethinking with brms, ggplot2, and the tidyverse* by A. Solomon Kurz (2018) [Link](#) chapters 12 and 13 (note that these currently refer to the FIRST edition of *Statistical Rethinking* but it is the same content as in the previous bullet point)

## Week 7: Writing your own Stan Programs with rstan

- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). “Stan: A Probabilistic Programming Language”. *Journal of Statistical Software*, 76(1), 1 - 32. [Link](#). Do not worry too much about references to the “command line”; we will be using the rstan R package to interface with Stan and can obtain all the same information.
- *Statistical Rethinking*, chapter 6
- McElreath video: [Haunted DAG](#)
- *Causal Models: Guide to gbiqq* by Macartan Humphries and Alan Jacobs (2020) [Link](#)