

May Institute on computation and statistics for mass spectrometry and proteomics

April 30, 1:30pm – May 2, 5:00pm 2025. **Intermediate R and data visualization**

Venue:

Northeastern University main campus, West Village H, first and second floor rooms: 110, 108, 210A/B, 212. Please see the [annotated map](#).

Lead instructor:

[Kylie Bemis](#), k.bemis@northeastern.edu

Guest speakers:

[Olga Vitek](#)

Teaching assistants:

[Sai Lakkimsetty](#), [Anshuman Raina](#), [Yinyue Zhu](#)

Description:

This course is for R users who are looking to level up their R programming skills. It is designed for scientists who can do basic analysis using R scripts, and want to write more performant, maintainable, and scalable R programs for their research, institution, or public distribution.

On day 1, we will begin by breaking down R's fundamental philosophy of programming, including variables, objects, functions, functional programming, and other language features.

On day 2, we will focus on writing correct and efficient code. We will discuss strategies for writing more performant R code as well as how to identify bottlenecks and bugs through profiling and debugging. We will also discuss statistical methods, statistical considerations, and strategies for scaling algorithms to larger-than-memory datasets.

On day 3, we will discuss the development, distribution, and maintenance of R packages and the basics of contributing to open-source projects like Bioconductor.

Each topic will include a mix of lecture and hands-on examples with live coding.

References:

The course will combine lectures and practical hands-on exercises. The discussion of programming with R is based on the following textbooks:

- Wickham. [Advanced R](#), Chapman and Hall/CRC, 2014
- Wickham. [R packages](#), O'Reilly, 2015
- Grolemund & Wickham. [R for Data Science](#), O'Reilly, 2017

Target Audience:

Target audience are experimental scientists, bioinformaticians, computer scientists, data scientists, statisticians or engineers, with a minimal prior exposure to R (e.g., at the level of the course 'Beginners R and beginners statistics') is expected.

Requirements:

Participants should be familiar with programming equivalent to the [Beginner's R](#) session.

We will use the latest versions of [R 4.5](#) with [Bioconductor 3.21](#). Please install or update to these versions if you wish to follow along with the examples.

Course materials:

[Intermediate R, data visualization and statistics](#)

Tentative schedule (updated 11/1)

Wednesday April 30, 2025

12:30 p.m. Registration
1:30 p.m. **Introduction to Statistics**, Olga Vitek
3:00 p.m. Refreshments
3:30 p.m. **Functions and objects**, Kylie Bemis
5:00 p.m. Q&A and adjourn

Thursday, May 1, 2025

9:00 a.m. **Debugging in R**, Kylie Bemis
10:30 a.m. Refreshments
11:00 a.m. **Principles of statistical inference**, Olga Vitek
12:30 p.m. Lunch on your own
1:30 p.m. **Benchmarking and performance**, Kylie Bemis
3:00 p.m. Refreshments
3:30 p.m. **Scalability and algorithms**, Kylie Bemis
5:00 p.m. Q&A and adjourn

Friday, May 2, 2025

9:00 a.m. **Experimental design. Class discovery and class prediction**, Olga Vitek
10:30 a.m. Refreshments
11:00 a.m. **R packages**, Kylie Bemis
12:30 p.m. Lunch on your own
1:30 p.m. **Github repositories**, Kylie Bemis
3:00 p.m. Refreshments
3:30 p.m. **Advanced R (Q&A)**, Kylie Bemis
5:00 p.m. Wrap-up and adjourn

Link to participant lunch order pickup

If you would like us to pick up your prepaid lunch order from Tatte or Anna's Taqueria, fill this form: <https://forms.gle/5E8apFjvW3qZhgQA>

Course evaluation

Please help us improve the program in the future by filling in this form:

<https://forms.gle/s6bbgAUGhdDy6GjG9>