

Computing infrastructure and curriculum design for introductory data science

Mine Çetinkaya-Rundel, *Duke University* + *RStudio*

Contact: cetinkaya.mine@gmail.com

The goal of this workshop is to equip educators with concrete information on content and infrastructure for designing and painlessly running a modern data science course. This is a three-part workshop. Part 1 will outline a curriculum for an introductory data science course and discuss pedagogical decisions that go into the choice of topics and concepts as well as the choice of programming language (R) and syntax (primarily tidyverse), and the emphasis on literate programming for reproducibility (with R Markdown). Part 2 will discuss infrastructure choices around teaching data science with R: RStudio as an integrated development environment, cloud-based access with RStudio Cloud and Server, version control with Git, and collaboration with GitHub. Part 3 will focus on classroom management on GitHub (with ghclass). Workshop attendees will work through several exercises from the course and get first-hand experience with using the tool-chains and techniques described above. While the workshop content will focus on usage of R, many of the pedagogical takeaways will be language agnostic. All workshop content, including teacher facing documentation and student facing course materials, will also be available to participants via datasciencebox.org. Please bring a laptop with you.

Keywords: data science; pedagogy; reproducibility; R; version control

DOI: <https://doi.org/10.1145/3287324.3287556>