

# Welcoming learners to data science with the tidyverse

Mine Çetinkaya-Rundel

Professor of the Practice and Director of Undergraduate Studies, Department of Statistical Science,  
Duke University

Computing makes up a large and growing component of data science and statistics courses. Many of those courses, especially when taught by faculty who are statisticians by training, teach R as the programming language. A number of instructors have opted to build much of their teaching around use of the tidyverse. The tidyverse, in the words of its developers, "is a collection of R packages that share a high-level design philosophy and low-level grammar and data structures, so that learning one package makes it easier to learn the next" (Wickham et al. 2019). The shared principles have led to the widespread adoption of the tidyverse ecosystem. No small part of this usage is because the tidyverse tools have been intentionally designed to ease the learning process and cognitive load for users as they engage with each new piece of the larger ecosystem. Moreover, the functionality offered by the packages within the tidyverse spans the entire data science cycle, which includes data import, visualisation, wrangling, modeling, and communication. The tidyverse provides an effective and efficient pathway to data science mastery for students at a variety of different levels of experience. In this talk, I will introduce the tidyverse from an educator's perspective, touching on the what (a brief introduction to the tidyverse), the why (pedagogical benefits, opportunities, and challenges), the how (scoping and implementation options), and the where (details on courses, curricula, and student populations).