# TEACHING DATA SCIENCE PRINCIPLES WITH R

Uma Ravat
UC Santa Barbara, USA
umaravat@ucsb.edu

## INTRODUCTION

Educators creating introductory data science courses without prerequisites will find details about our course design decisions, tools, and adopted resources.

The 10-week introductory course encompassed three foundational modules: (1) R Programming, (2) Statistics, and (3) Databases. We chose Base R instead of tidyverse for its applicability in advanced courses, and since there is no significant difference in student experience with Introductory Statistics labs in formula or tidyverse syntax (McNamara 2022).

## PEDAGOGICAL CHOICES

The 5-credit course comprised two 75-minute lectures and two 50-minute lab sections each week. Lectures introduced topics with live RStudio coding, followed by "YourTurn" exercises or interactive learnr tutorials (Aden-Buie 2023) during which students were assisted by Undergraduate Learning Assistants. This structure offered practice, support, and instant feedback despite the large course size. Each lecture session was followed by a 50-minute lab led by a Teaching Assistant during which students worked on and submitted (for completion) an RMarkdown worksheet that aligned with the lecture topic. Weekly RMarkdown homework was due mid-week and was followed by an end-of-week quiz. Additional assessments included a multiple-choice midterm and final exam. This structure provided essential practice needed by novice R programmers.

## TECHNOLOGY CHOICES

RStudio via JupyterHub was employed as an alternative to paid PositCloud. This ensured each student's R and Rstudio setup was identical to the live-coding setup used by the instructor which allowed students to code along with the instructor in RMarkdown, and work with reproducible workflows and literate programming within 30 minutes of the first lecture itself. Upon logging in, each student's JupyterHub auto-fetched lecture slides, labs, and homework content hosted on GitHub. Resources in Çetinkaya-Rundel, M. (2021) offer an excellent first-day activity for introducing coding and reproducibility. Toward the end of the course, students installed both R and Rstudio locally on their computers in order to gain an important and valuable skill for budding data scientists.

An Ed-stem discussion group, Edstem (2023) was set up primarily for its ability to not only post questions about coding but also to write R code that can be run right in the discussion post in the browser. This was very valuable for beginner programmers as teaching staff were able to correct code without the need for having the student be physically present or doing a copy-paste online.

## INSTRUCTOR REFLECTIONS

Choosing R over tidyverse helped students gain valuable SQL and database skills that proved advantageous in job interviews. A drawback of interactive learnr tutorials (Aden-Buie 2023) is the inability to retain student-typed commands/code for future review. Students favored RMarkdown exercises for their ability to save and revisit work, addressing this limitation.

## REFERENCES

McNamara, A. (2022). Teaching modeling in introductory statistics: A comparison of formula and tidyverse syntaxes. *ArXiv*. /abs/2201.12960

Aden-Buie G, Schloerke B, Allaire J, Rossell Hayes A (2023). learnr: Interactive Tutorials for R. https://rstudio.github.io/learnr/, https://rstudio.github.io/learnr/.

Çetinkaya-Rundel, M. (2021). Let them eat cake (first)!, Invited Talk, https://speakerdeck.com/minecr/let-them-eat-cake-first-diz

Edstem (2023) An online discussion and communication tool. Retrieved from https://edstem.org/

  