

# Practical Data Management with R

## Online course learning objectives

This course teaches learners how to use R to manage data in a wide variety of formats, in a reproducible manner, at scale.

### This course will help learners to:

- An understanding of Basic R commands and data structures for manipulating data.
- The ability to read data from multiple formats in and out of R.
- Proficiency using loops, conditional statements, and functions to automate common data management tasks.
- Familiarity with R's package system for extending its functionality.
- The skills to clean and manage multiple complex datasets.
- The ability to clean and manipulate textual data.
- An understanding of basic web scraping techniques, for both standard web pages and the Twitter API.
- An overview of the techniques and hardware necessary to manage large datasets efficiently.

**Language:** English

**Time to complete:** 30 hours

**Instructor:** Matthew Denny

## Online course full syllabus

### MODULE ONE: INTRODUCTION TO R AND RSTUDIO

You will learn about: installing R and RStudio, basic R programming skills and five basic data structures.

- How to install R and RStudio.
- Basic R programming skills, such as how to write commands in an R script.
- How to understand the core data structures you need to manage a huge variety of data.

### MODULE TWO: R PROGRAMMING FUNDAMENTALS

You will learn about: Data I/O and packages, looping and conditional statements, and functions.

- Data I/O and packages so you can extend the functionality of R.
- Looping and conditional statements so you can automate wildly complex tasks.
- Functions so you don't have to write the same code over and over again for similar tasks.

# Practical Data Management with R

## MODULE THREE: DATA MANAGEMENT IN R

You will learn how to: manage multiple datasets by example, work with text data, convert long- and wide-format data and deal with messy data.

- Manage multiple datasets by example.
- Convert long and wide format data.
- Deal with poorly formatted data and/or missing data.
- Automate tasks using functions Work with and manipulate text data.

## MODULE FOUR: AUTOMATED DATA COLLECTION

You will be given: an overview of web/text scraping and the legal considerations, a basic web scraping example and an understanding of scraping Twitter.

- An overview of web/text scraping and the related legal considerations.
- A basic web scraping example so you can learn how to treat a webpage as a messy text document.
- An understanding of scraping Twitter.

## MODULE FIVE: PERFORMANCE AND SCALABILITY

You will be introduced to High Performance Computing and Big Data and learn about performant programming.

- Giving you an overview of big data and high-performance computing (HPC).
- Teaching you about performant programming.
- Wrapping up with a range of next steps and ways to extend your skills.