

Explore data and R

Cleaning data

- ✓

Video: Cleaning up with the basics
7 min
- 📖

Reading: File-naming conventions
10 min
- 📖

Reading: More on R operators
10 min
- ▶

Video: Organize your data
6 min
- 📖

Practice Quiz: Hands-On Activity: Cleaning data in R
1 question
- 📖

Reading: Optional: Manually create a data frame
10 min
- ▶

Video: Transforming data
5 min
- 📖

Reading: Wide to long with tidyr
10 min
- ⌕

Ungraded Plugin: Clean, organize, and transform data with R
10 min
- 📖

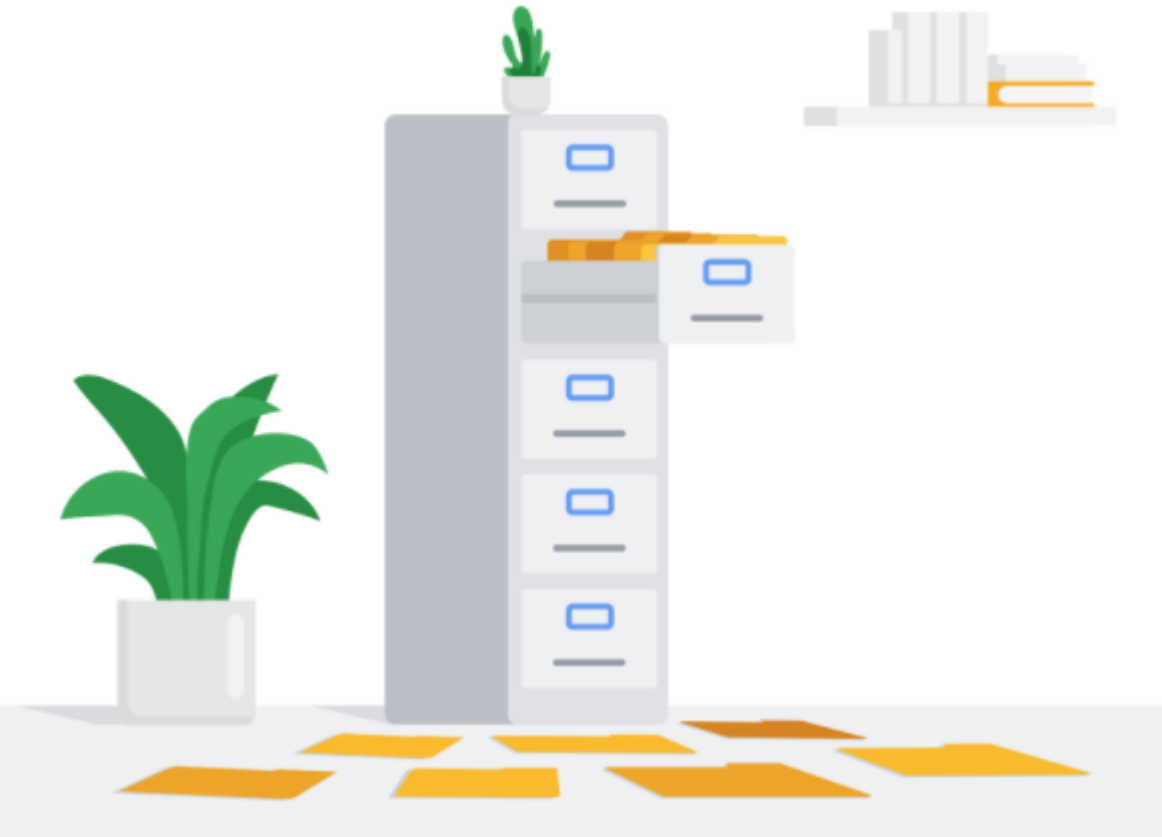
Practice Quiz: Test your knowledge on cleaning data
3 questions

Take a closer look at the data

Weekly challenge 3

File-naming conventions

An important part of cleaning data is making sure that all of your files are accurately named. Although individual preferences will vary a bit, most analysts generally agree that file names should be accurate, consistent, and easy to read. This reading provides some general guidelines for you to follow when naming or renaming your data files.



What’s in a (file)name?

When you first start working with R (or any other programming language, analysis tool, or platform, for that matter), you or your company should establish naming conventions for your files. This helps ensure that anyone reviewing your analysis–yourself included–can quickly and easily find what they need. Next are some helpful “do’s” and “don’ts” to keep in mind when naming your files.

Do

- Keep your filenames to a reasonable length
- Use underscores and hyphens for readability
- Start or end your filename with a letter or number
- Use a standard date format when applicable; example: YYYY-MM-DD
- Use filenames for related files that work well with default ordering; example: in chronological order, or logical order using numbers first

Examples of good filenames
2020-04-10_march-attendance.R
2021_03_20_new_customer_ids.csv
01_data-sales.html
02_data-sales.html

Don't

- Use unnecessary additional characters in filenames
- Use spaces or “illegal” characters; examples: &, %, #, <, or >
- Start or end your filename with a symbol
- Use incomplete or inconsistent date formats; example: M-D-YY
- Use filenames for related files that do not work well with default ordering; examples: a random system of numbers or date formats, or using letters first

Examples of filenames to avoid
4102020marchattendance<workinprogress>.R
_20210320*newcustomeridsforfebonly.csv
firstfile_for_datasales/1-25-2020.html
secondfile_for_datasales/2-5-2020.html

Additional resources

These resources include more info about some of the file naming standards discussed here, and provide additional insights into best practices.

- **How to name files** [↗](#): this resource from Speaker Deck is a playful take on file naming. It includes several slides with tips and examples for how to accurately name lots of different types of files. You will learn why filenames should be both machine readable and human readable.
- **File naming and structure** [↗](#): this resource from the Princeton University Library provides an easy-to-scan list of best practices, considerations, and examples for developing file naming conventions.

Mark as completed

