Importing data into R

Cheatsheet

2024-07-28

> **Note**
>
> This cheatsheet does not have working examples as it is intended to be used as a reference guide. If you wish to practice, download the two files below and try importing them into R using the code snippets provided in the cheatsheet.

> **Download data**
>
> We have two separate datasets. The first dataset is part of the possums dataset used in BIOL2022 labs. It contains two numerical variables: `ExpBLUP` and `AactiveTBLUP`. The data is available in the file `possums-blup.csv`.
> The second dataset, `penguins.csv`, contains data collected by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER. Details about the dataset can be found here.

File paths

- **File paths** specify the location of a file on your computer. They show the route to the file, starting from the root of the file system, passing through folders and subfolders, and ending with the file name.
- For reproducibility, consider using **projects** in RStudio, which standardises the working directory. Alternatively, use `setwd()` but note that absolute paths may not work on other computers as they are specific to your computer.

Working directory

The **working directory** is the folder where R will look for files by default. If you use absolute paths, you don't need to set the working directory but your paths will not be reproducible on

other computers. Use `getwd()` as a first step to check the current working directory and get your bearings, as one is always set when you open R.

```
getwd()                           # Get current working directory
setwd("C:/path/to/your/folder")   # Set working directory
data <- read.csv("file.csv")      # File in the current working directory
data <- read.csv("data/file.csv") # File in a subdirectory
data <- read.csv("../file.csv")   # File in a parent directory
```

Importing data into R

The most common data formats and how to import them into R are listed below. For other formats, see the **More resources** section.

CSV files (.csv)

You can either use base R's `read.csv()` or the `readr` package to import CSV files.

```
df <- read.csv("file.csv") # Base R

library(readr) # readr package (faster, more robust)
df <- read_csv("file.csv")
```

Excel files (.xlsx, .xls)

You can use the `readxl` package to import Excel files. Use `?read_excel` to view more options e.g. `sheet`, `range`, etc.

```
library(readxl) # readxl package
df <- read_excel("file.xlsx", sheet = "Sheet1")
```

Tab-delimited files (.tsv)

You can use base R's `read.delim()` or the `readr` package to import tab-delimited files.

```
df <- read.delim("file.txt") # Base R

library(readr) # readr package
df <- read_tsv("file.tsv")
```

RDS Files (R's native format)

You can use the `readRDS()` function to import RDS files.

```
df <- readRDS("file.rds")
```

Windows file paths

- Windows file paths use backslashes (\) instead of forward slashes (/), which can cause issues in R.
- R deals with this by "escaping" the backslashes. For every backslash in a file path, you need to use two backslashes.
- Use `r(...)` to automatically escape Windows backslashes in file paths. For example:
  - `r"(C:\Users\jd\Documents\folder\file.csv)"` converts to
  - `"C:\\Users\\jd\\Documents\\folder\\file.csv"`
- Combining `r()` with `read_csv()`:

```
readr::read_csv(r"(C:\Users\jd\Documents\folder\file.csv)")
```

More resources

- File Paths in R for Epidemiology – a detailed guide on file paths in R.
- Datacamp's How to Import Data Into R – a comprehensive tutorial on importing data into R including databases, APIs, and web scraping.

License