Importing data & tidy data

*with the **readr** and **tidyr** packages*

reshaping a data frame

`wide <-> long`

read_csv()
pivot_longer()
pivot_wider()

Every tool from the last two chapters assumes your data is tidy.

Tidy data

Three rules. A data frame that follows all three works with every **ggplot2** and **dplyr** function you have already met.

Three Rules

- Each **variable** is its own column.
- Each **observation** is its own row.
- Each **value** is its own cell.
- Untidy data is not wrong, it is just not ready yet.

read_csv()

Import a **comma-separated values** file into R as a data frame. From the **readr** package, not the base R function of a similar name.

Import a Spreadsheet

read_csv() is not read.csv(): the underscore version is the tidyverse one.

read_csv()

Example

what to call the result

```
dem_score <- read_csv("https://moderndive.com/data/dem_score.csv")
```

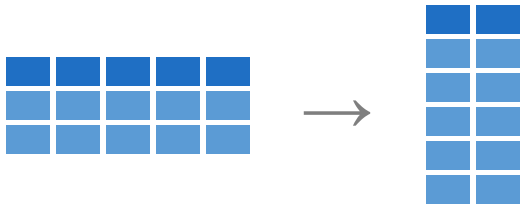
the verb (function)

where the file is: a path or a URL

wide format

One row per case, with a **column for every year** or category. Easy to read on a page, awkward to plot: the variable you want on the x-axis is spread across the column names.

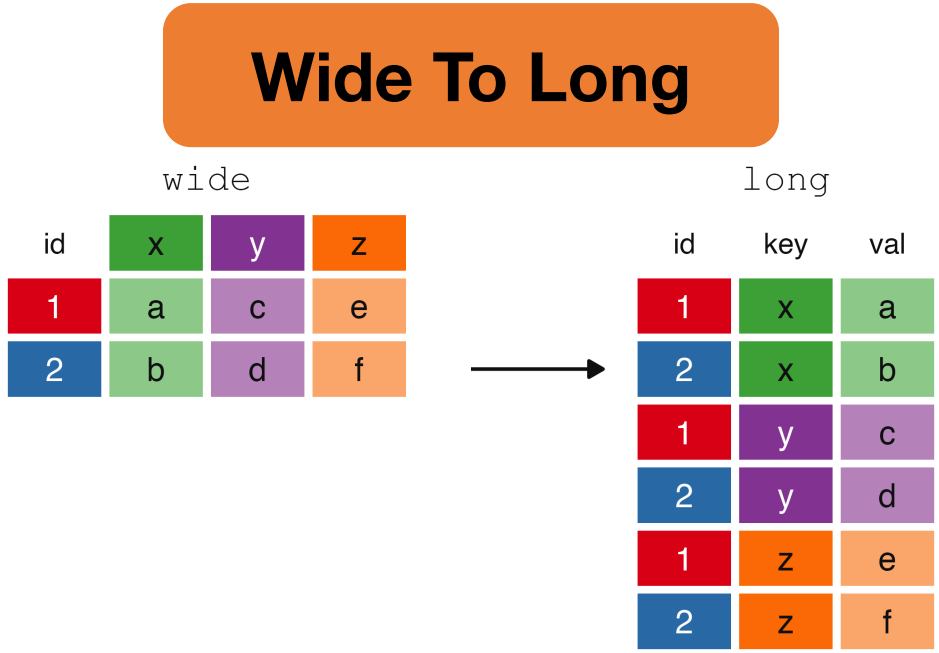
Values Hidden In Headers



pivot_longer()

Diagram adapted from Garrick Aden-Buie's tidyexplain

Turn many columns into **two**: one holding the old column names, one holding the values. This is what makes a wide table tidy.



pivot_longer()

Example

the data frame

```
drinks_smaller_tidy <- drinks_smaller |>
  pivot_longer(names_to = "type", values_to = "servings", cols = -country)
```

the verb (function)

new column for the old headers

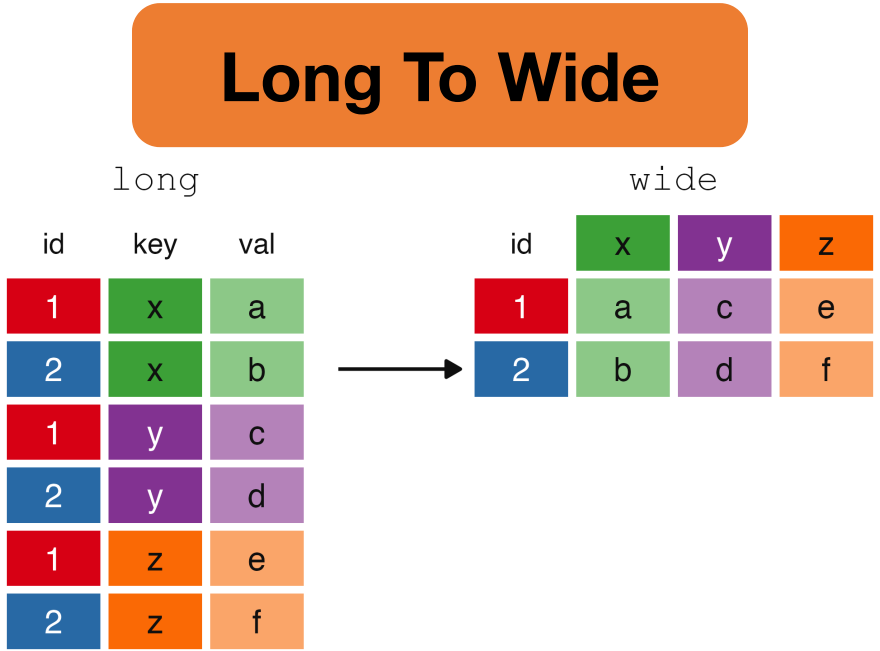
new column for the values

which columns to reshape

pivot_wider()

Diagram adapted from Garrick Aden-Buie's tidyexplain

The inverse of **pivot_longer()**. Spread one column of names back out into many columns. Useful for presenting a table, rarely for analysis.



pivot_wider()

Example

```
drinks_smaller_tidy |>
  pivot_wider(names_from = type, values_from = servings)
```

the verb (function)

column supplying the new headers

column supplying the new values

tidyverse

One package that loads the others.

The Umbrella Package

- **library(tidyverse)** loads ggplot2, dplyr, readr, tidyr and more.
- **readr** imports. **tidyr** reshapes.
- **dplyr** wrangles. **ggplot2** plots.
- They share a design, so what you learn on one transfers to the rest.