Learning Goals

- Explain what it means for a dataset to be "tidy" (one variable per column, one observation per row) and recognize common signs of untidy data.
- Use `separate()` and `unite()` to split one column into multiple variables and combine multiple variables into one column.
- Use `pivot_wider()` and `pivot_longer()` to reshape data.
- Perform multi-step tidying workflows that combine pivoting and separating to extract multiple variables from column names.
- Combine datasets using `left_join()`, `right_join()`, `inner_join()`, and `full_join()`, and explain how each join changes rows and missing values.

Key Definitions / Functions

- `tidy data`:
- `separate()`:
- `unite()`:
- `pivot_longer()`:
- `pivot_wider()`:
- `left_join() / right_join()`:
- `full_join()`:

Practice Problems

For each task below, write the R code you would use and briefly describe what you expect the output to look like.

1. Create the dataset below. Use `separate()` to split `student_id` into `cohort` and `number` using the dash as the separator.

```
df <- data.frame(student_id = c("F24-001","F24-014","S25-003","S25-011"),
                 score = c(88, 91, 77, 84))
```

2. Using the dataset below, use `unite()` to create a new column called `section` by combining `building` and `room` with a dash (example: `A-101`).

```
df <- data.frame(building = c("A","A","B","B"),
                 room = c(101, 102, 201, 202),
                 seats = c(30, 28, 35, 40))
```

3. Create the dataset below. Use `pivot_wider()` so that each `team` is one row and each `game` becomes a column. The cells should contain `score`.

```
df <- data.frame(team = c("Lions","Lions","Tigers","Tigers"),
                 game = c("G1","G2","G1","G2"),
                 score = c(21, 17, 14, 28))
```