# Applied Statistics for Data Scientists with R

Class 06: Data Wrangling with Base R and dplyr

# Learning Objective

1. Base R data manipulation vs dplyr

2. dplyr: select, filter, arrange, mutate, group by, summarize

# Key Data Wrangling Tasks

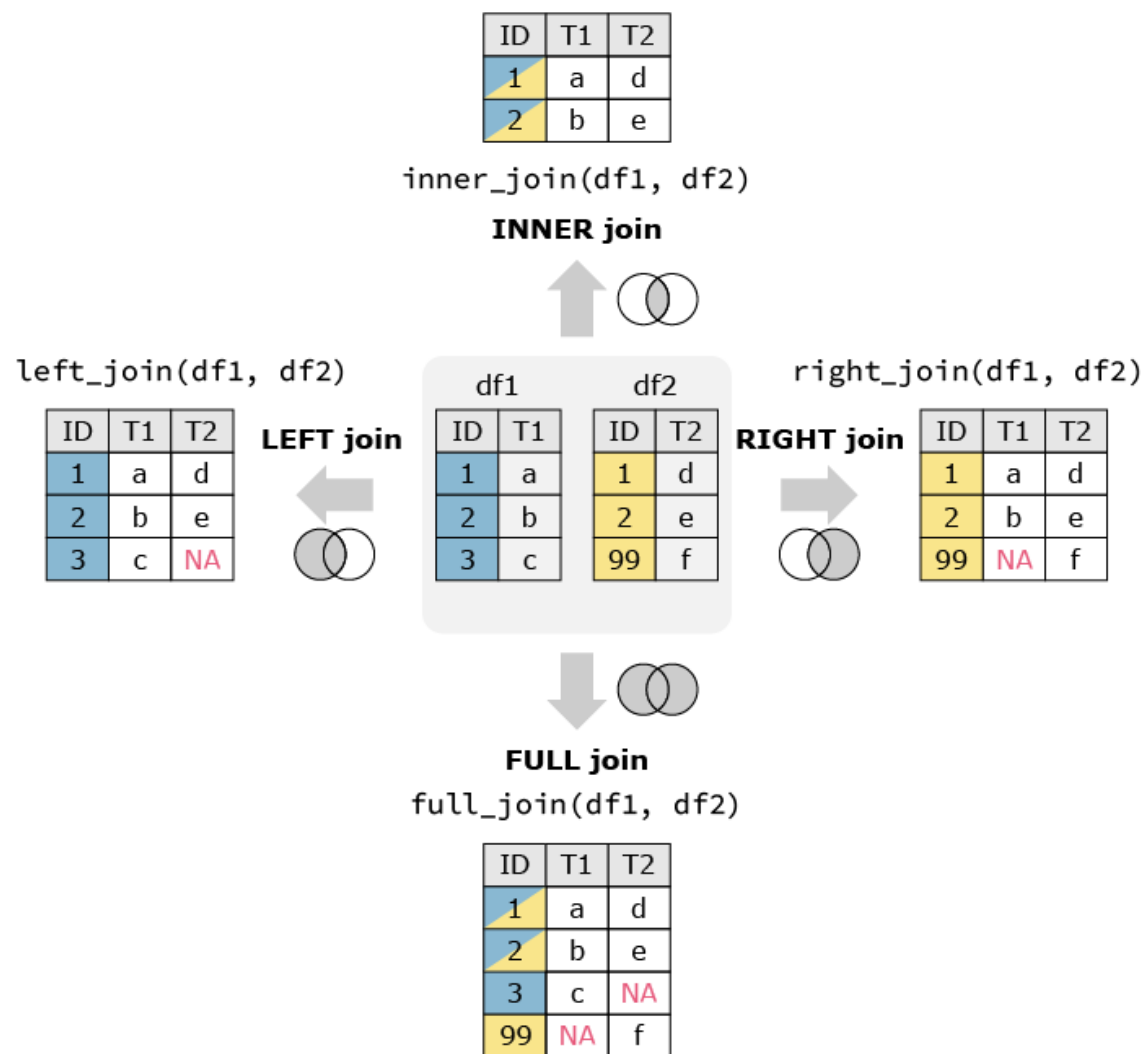Data wrangling means transforming raw data into a structured and usable format for analysis.

- Filter rows

- Select columns

- Create / remove columns

- Edit columns / rows

- Arranging

- Data reshaping

- Join data frames

# Reshape Data Frames

"Long" format

| country | year | metric |
|---------|------|--------|
| x | 1960 | 10 |
| x | 1970 | 13 |
| x | 2010 | 15 |
| y | 1960 | 20 |
| y | 1970 | 23 |
| y | 2010 | 25 |
| z | 1960 | 30 |
| z | 1970 | 33 |
| z | 2010 | 35 |

"Wide" format

| country | yr1960 | yr1970 | yr2010 |
|---------|--------|--------|--------|
| x | 10 | 13 | 15 |
| y | 20 | 23 | 25 |
| z | 30 | 33 | 35 |

# Joining Datasets

# Some Add-ins to Make Life Easier

- questionr
- Styler
- reprex