Relational Data

Nick Ulle

What is Relational Data?

Relational data are split across multiple tables.

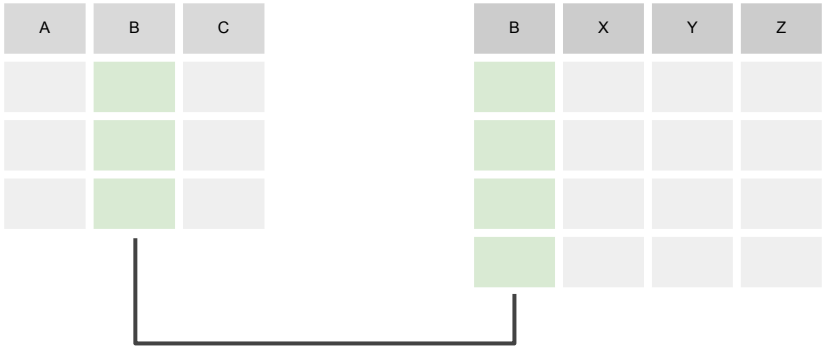
For example, a grocery store inventory might have tables for:

- Store information
- Product information
- Product quantities at each store

The relational format reduces redundancy.

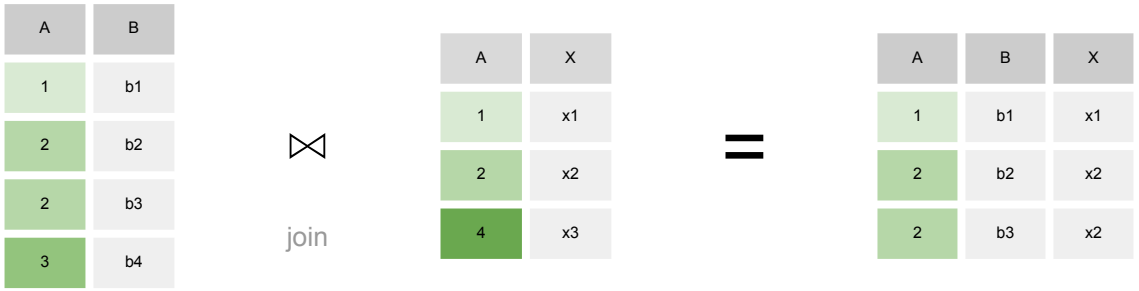
Relations Between Tables

A *key* column relates one table to others.



Combining Tables

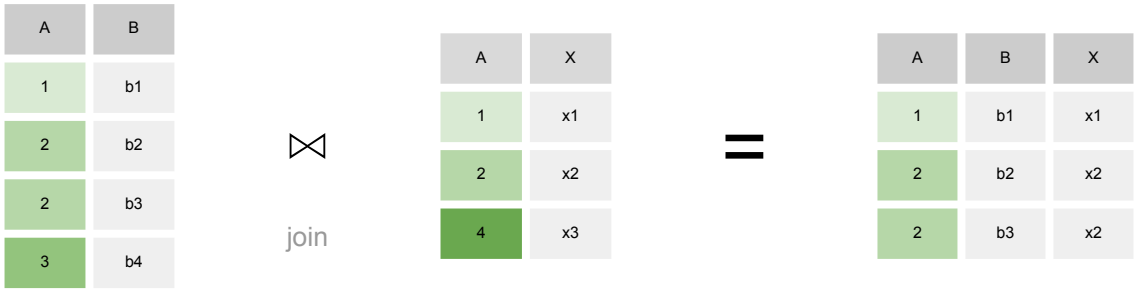
A *join* combines two tables based on a key.



New rows are created by combining rows where the keys match.

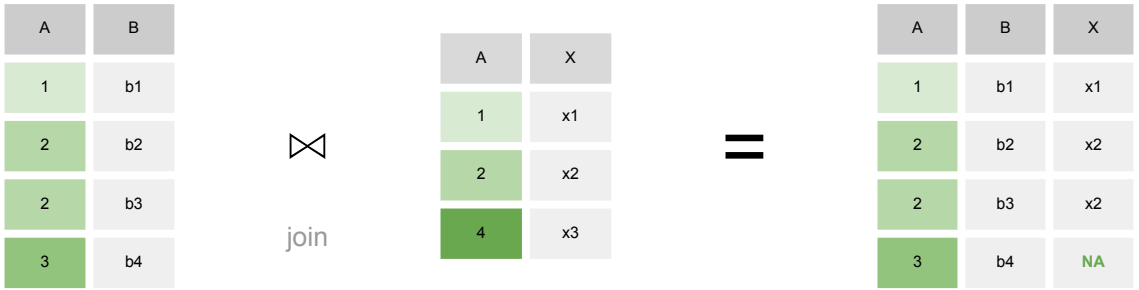
Inner Join

An *inner join* keeps only rows that match:



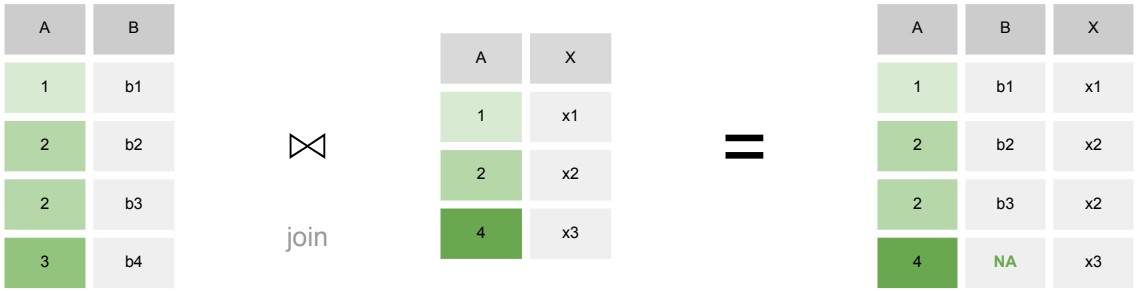
Left Join

A *left join* keeps all rows in the left table, matching rows in right table:



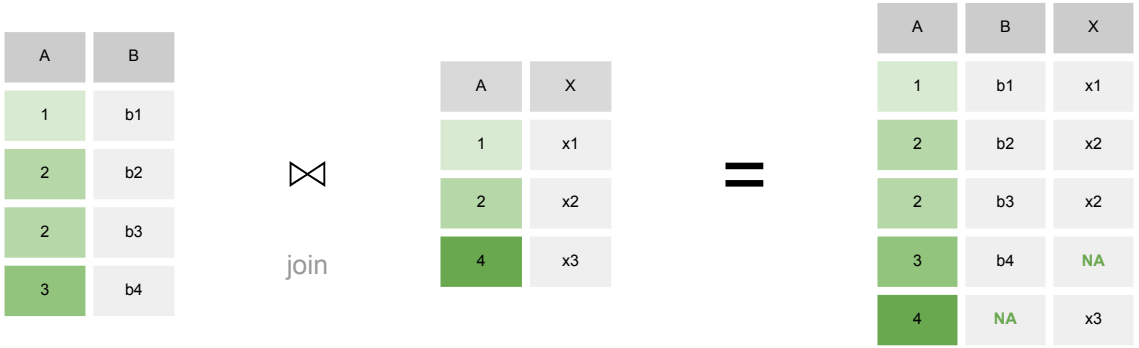
Right Join

A *right join* keeps matching rows in left table, all rows in right table:



Full Join

A *full join* keeps all rows:



Joins

with dplyr

Nick Ulle

STAT 33
Wrap-up

Nick Ulle

What Now?

- **DATA 8**: For everyone. Basic statistics skills. A chance to learn Python.
- **DATA 100**: If you want a data science career. Round out your skill set: databases, visualizations, modern statistical methods.
- **STAT 133**: If you want to know more about (how statisticians use) R.