

C: Basic database joins

Combining datasets by matching fields

- Terminology: **joining** or **merging** data

Example:

Dataset 1: State FIPS codes and names (names data)

FIPS	Name
09	Connecticut
23	Maine
25	Massachusetts
33	New Hampshire
44	Rhode Island
...	...

- Will often refer to as a **table**
- Terminology from **SQL**, a core database language

Dataset 2: FIPS codes and populations (pop data)

FIPS	Pop
25	6830193
09	3581504
33	1343622
23	1332813
44	1056611

50	624977
...	...

Want to **merge** or **join** (SQL) *population* **onto** the *names* by FIPS code:

FIPS	Name	Pop
09	Connecticut	3581504
23	Maine	1332813
25	Massachusetts	6830193
33	New Hampshire	1343622
44	Rhode Island	1056611
50	Vermont	624977
...

Name:	From name data
Pop:	From pop data
FIPS:	Shared by both (join key)

Conceptual approach:

Step 1: build a new table:

- Has names data in **left** columns
- Has new empty columns on **right** for pop data

FIPS	Name	Pop
09	Connecticut	
23	Maine	
25	Massachusetts	

33	New Hampshire	
44	Rhode Island	
50	Vermont	
...	...	

Step 2: loop through rows of new data

- Get the value of **FIPS** in **left table**
- Find the row in **right table** with the same value of **FIPS**
- Add data from **right table onto** the matching **left row**

New Table (left, L)

FIPS	Name	Pop
09	Connecticut	
23	Maine	
25	Massachusetts	
33	New Hampshire	
44	Rhode Island	
50	Vermont	
...	...	

Pop Data (right, R)

FIPS	Pop
25	6830193
09	3581504
33	1343622
23	1332813
44	1056611
50	624977
...	...



New Table (R joined onto L)

FIPS	Name	Pop
09	Connecticut	3581504
23	Maine	
25	Massachusetts	

33	New Hampshire	
44	Rhode Island	
50	Vermont	
...	...	