Table joins

Cedric Arisdakessian

2022-09-09

One example

qPCR data:

SampleID	Site	Time	qPCR_16S
IKE_001	Hawaii Kai	t0	1e6
IKE_002	Hawaii Kai	t1	1e6
IKE_003	Kualoa	t0	3e6
IKE_004	Kualoa	t2	3.5e6
IKE_005	Waimea	t1	1e5
IKE_006	Waimea	t2	2e5

Site metadata:

Site	Temperature	Salinity
Kualoa	15	0.2
Waimea	20	0.3

Temporal metadata:

Timepoint	Month	Year
t0	February	2021
t1	February	2022
t2	April	2022

A few concepts

Different types of keys

- Unique key: uniquely identifies a row
- Foreign key: A key matching another table's unique key

Can you see the unique and foreign keys in the previous tables?

Different types of join

- inner join: intersection of keys -> `dplyr::inner_join()`
- left join: all keys from left table -> `dplyr::left_join()`
- right join: all keys from right table -> `dplyr::right_join()`
- outer join: union of all keys -> `dplyr::outer_join()`

https://cdn.educba.com/academy/wp-content/uploads/2019/11/joins-in-mysql-1.png.webp

Back to our example

```
site_metadata <- read.csv(...)
time_metadata <- read.csv(...)
qpcr <- read.csv(...)

data <- qpcr %>%
  right_join(site_metadata, by="Site") %>%
  left_join(temporal_metadata, by=c("Timpoint"="Time"))
```

How many rows are there in the final `data` table?

In other languages

SQL:

```sql
SELECT x FROM table1
  LEFT JOIN table2 ON table1.key1 = table2.key2
```

Python:

```python
data1.merge(
  data2,
  left_on="key1", right_on="key2",
  how="inner"
)
```