Biostats 597E

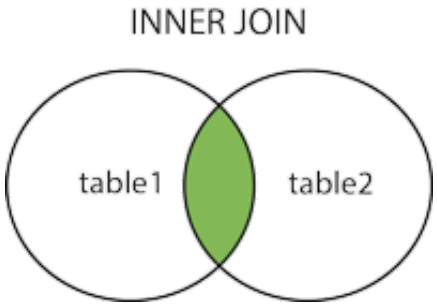
Week 3 - Introduction to SQL

SQL Joins

- Using **WHERE** with list of tables generate only records with matches.
- What if we want to keep those unmatched records?
- Here, we introduce SQL joins:
  - **INNER JOIN**: Returns all rows when there is at least one match in BOTH tables (same as using **WHERE**)
  - **LEFT JOIN**: Return all rows from the left table, and the matched rows from the right table
  - **RIGHT JOIN**: Return all rows from the right table, and the matched rows from the left table
  - **FULL JOIN**: Return all rows when there is a match in ONE of the tables
- Joins use **ON** clause to specify conditions

Inner Join

The INNER JOIN keyword selects all rows from both tables as long as there is a match between the columns in both tables.



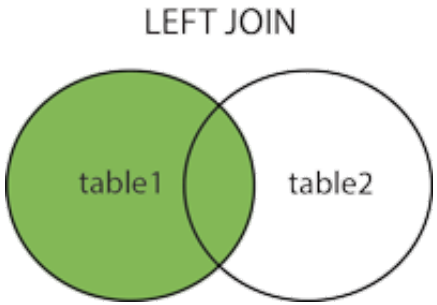
- Example: Find coordinates of country capitals

```
select Name, Capital, Latitude, Longitude
  from countries a inner join worldcitycoords b
  on a.Capital = b.City and a.name = b.Country
```

- You do not need a.Capital if column name Capital appears in only table a

Left Join

The LEFT JOIN keyword returns all rows from the left table (table1), with the matching rows in the right table (table2). The result is NULL in the right side when there is no match.

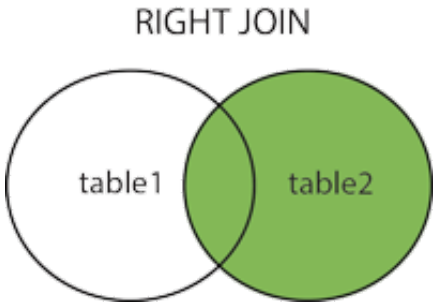


- Example: Find coordinates of country capitals, also list capitals without matched coordinates

```
select Name, Capital, Latitude, Longitude
  from countries a left join worldcitycoords b
  on a.Capital = b.City and a.name = b.Country
```

Right Join

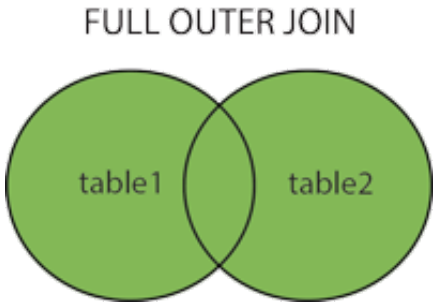
The RIGHT JOIN keyword returns all rows from the right table (table2), with the matching rows in the left table (table1). The result is NULL in the left side when there is no match.



- Right join is not implemented in SQLite
- Right join is available in other SQL system such as SAS
- Right join is less often used
- You can achieve right join using left join (how?)

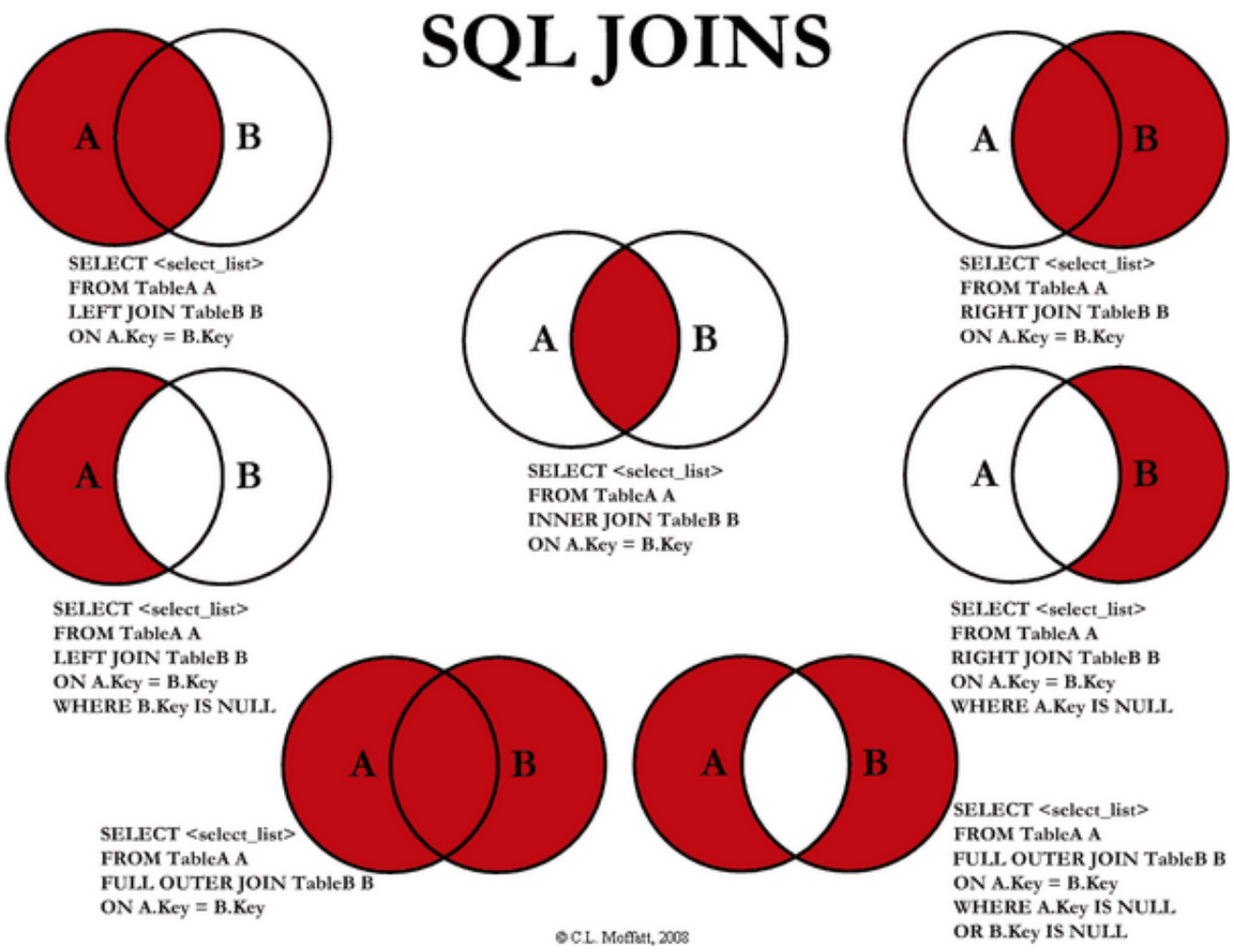
Full Join

The FULL OUTER JOIN keyword returns all rows from the left table (table1) and from the right table (table2). The FULL OUTER JOIN keyword combines the result of both LEFT and RIGHT joins.



- Full join is not implemented in SQLite
- Full join is available in other database system such as SAS
- Full join is rarely used in my experience

Summary


SQL JOINS
A
B
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL
© C.L. Moffatt, 2008

/

Excercise

- Using **oilprod** and **oilrsrvs** to find countries that have oil production but do not have oil reserve

Combine Join and Sub-query

- Many data retrieving tasks can be achieve by combining joins and sub-queries
- Example: For each country, find what proportion of its population over the total population of its corresponding continent

```
select a.Name, a.continent,
       a.population / b.total_population as percentage
  from countries a inner join
       (select Continent, sum(population) as total_population
          from countries group by Continent) b
    on a.Continent = b.Continent
```