

Simple Subqueries

SULIA EVANS
@bork

SQL queries run
in this order

FROM + JOIN



WHERE



GROUP BY



HAVING



SELECT (window functions
happen here !)



ORDER BY



LIMIT

<https://jvns.ca/blog/2019/10/03/sql-queries-don-t-start-with-select/>

Subqueries are useful when you want to filter or make some type of transformation prior to the main selecting, filtering, or calculating. A subquery is a nested (inner) query that could be run on its own.

This pattern -- of running a bit of inner code before running additional outer code -- is common in SQL, as well as Excel and Python, when the work to be done becomes more complex. Remember to always evaluate from the inside out!

One or more subqueries can be included in the SELECT, FROM, or WHERE clause of a query.

What do you think the following query will return?

```
3 → SELECT company, landed, new_jobs
1 → FROM (SELECT * FROM ecd WHERE county = 'Williamson') as williamson_county_jobs
2 → WHERE EXTRACT(year FROM landed) = '2018' ;
```



Subqueries in the FROM clause must be aliased

1. Get rows that have data for Williamson County
2. Keep just the data for 2018
3. Select the company, landed date, and new jobs for these rows

	<div>company</div> <div>text</div>	<div>landed</div> <div>date</div>	<div>new_jobs</div> <div>integer</div>
1	CKE Restaurants Holdings, Inc.	2018-06-28	145
2	SpecialtyCare	2018-03-15	124
3	DaVita Inc.	2018-05-31	145
4	Delek US Holdings, Inc.	2018-01-25	175
5	Pet Sense	2018-12-31	37

Let's try another one:

```
SELECT county, company, landed, capital_investment
FROM ecd
WHERE CAST(capital_investment as decimal) >
      (SELECT AVG(CAST(capital_investment as decimal)) FROM ecd);
```

What happens first?

	county text	company text	landed date	capital_investment money
1	Shelby	ALSAC St Jude Children's	2016-11-30	\$1,000,000,000.00
2	Montgomery	Hankook Tire Co., Ltd	2013-10-14	\$800,000,000.00
3	Gibson	Tyson Foods, Inc.	2017-11-20	\$320,000,000.00
4	Blount	Denso Manufacturing Tenness...	2017-10-06	\$1,000,000,000.00
5	Sullivan	Eastman Chemical Company	2013-05-29	\$1,600,000,000.00
6	Davidson	AllianceBernstein L.P.	2018-05-02	\$71,518,867.00
7	Davidson	Bridgestone Americas, Inc.	2014-11-11	\$232,600,000.00
8	Maury	GM Spring Hill Manufacturing	2011-11-21	\$244,000,000.00

- Format your queries by lining up SELECT, FROM, WHERE, GROUP BY and ORDER BY keywords.
- Annotate your queries with comments if the logic is complex
 - `--` for single line comments and
 - `/*` for multiline comments `*/`
- Indent your *subqueries*

```
SELECT
    col1,
    col2,
    col3
FROM table1
WHERE col1 IN
    (SELECT id
     FROM table2
     WHERE year = 1991);
```

- Consider whether a subquery is necessary (subqueries require computing power)

Exercises

1. Write a subquery in the WHERE clause to select the top 5 capital investment amounts from the ecd table. Be sure to exclude NULL capital investments in the subquery. Next write an outer query to return the company, landed date, county, and capital investment for the top 5 investment amounts.
2. Use a subquery in the FROM section to find observations in county tier 4 with capital investment greater than \$10,000,000 in the ecd database. Remember to alias your subquery! Next write an **outer** query to find the average number of new jobs and the average amount of total grants where an *economic development* grant is included in the total for this group.