

Group by exercises.

Hello everyone ... GROUP BY queries are really quite simple.

You can group by one (or more attributes).

In the final query answer, you can include any of the grouping attributes plus any aggregate applied to any of the attributes in the table. These aggregates are computed against the rows in a group – resulting in one row in the query answer for each group.

Here are a few sample queries, followed by query answers to the three “French fries warmup” exercises that you see at the end of Lecture 2 slides. These are the exercises you used with the actual “French fries” – the small pieces of data that were in the envelope.

Note that the data for the French fries exercise is available in the Postgres phpPgAdmin query page. Just look in the Sailors database for the table names Store. You can practice your GROUP BY queries using this data.

Lois Delcambre

1. For each store, find out how many prices are listed in the Store table. Note since none of the rows in this table have a NULL price, and since SQL has bag semantics – with duplicates – the count of prices is the same as count(*) – the count of rows for each group. Try it!

```
select store, count(price) from store group by store
```

Query Results

| store | count |
|--------------|-------|
| Safeway | 6 |
| 7-11 | 9 |
| Circle K | 8 |
| Plaid Pantry | 6 |

4 row(s)

2. For each store, find out how many distinct prices are used.

```
select store, count(distinct price) from store group by store
```

Query Results

| store | count |
|--------------|--------------|
| 7-11 | 3 |
| Circle K | 7 |
| Plaid Pantry | 5 |
| Safeway | 5 |

4 row(s)

- Here's an example where we GROUP BY two attributes – store and beverage. For each store/beverage combination, find how many distinct prices are used.

select store, beverage, count(distinct price) from store group by store, beverage

Query Results

| store | beverage | count |
|--------------|--------------------------------|--------------|
| 7-11 | 2% Milk | 1 |
| 7-11 | Chocolate Milk | 1 |
| 7-11 | Coke | 2 |
| 7-11 | Diet Caffeine-Free Cherry Coke | 1 |
| 7-11 | Diet Coke | 2 |
| 7-11 | Whole Milk | 2 |
| Circle K | Coke | 2 |
| Circle K | Diet Coke | 2 |
| Circle K | Whole Milk | 4 |
| Plaid Pantry | 2% Milk | 1 |
| Plaid Pantry | Diet Pepsi | 1 |
| Plaid Pantry | Pepsi | 2 |
| Plaid Pantry | Whole Milk | 2 |
| Safeway | 2% Milk | 2 |
| Safeway | Diet Coke | 1 |
| Safeway | Diet Pepsi | 1 |
| Safeway | Skim Milk | 2 |

17 row(s)

- Answer to the first “French fries warmup”:

What is the size of the largest size item that each store sells?

select store, max(size) from store group by store

NOTE: this answer is NOT what you would expect because the sizes are entered as text strings, e.g., "16oz" rather than as numbers.

Query Results

| store | max |
|--------------|------|
| Safeway | 64oz |
| 7-11 | 8oz |
| Circle K | 8oz |
| Plaid Pantry | 64oz |

4 row(s)

5. Answer to the second "French fries warmup":
For each store that sells at least 4 distinct beverages*, what is the cost of the highest price item that the store sells?

```
select store, max(price)
from store
group by store
having count(distinct beverage) > 3
```

Query Results

| store | max |
|--------------|------|
| 7-11 | 1.1 |
| Plaid Pantry | 3.2 |
| Safeway | 2.75 |

row(s)

6. Answer for the 3rd "French fries warmup"

For each beverage that comes in at least 3 sizes (ignoring what store it is sold in), what is its minimum cost (ignoring size) ?

```
select beverage, min(price)
from store
group by beverage
having count(distinct size) > 2
```

Query Results

| beverage | min |
|------------|------|
| Coke | 0.65 |
| Diet Coke | 0.6 |
| Whole Milk | 0.6 |

3 row(s)

7. Here's an example of a GROUP BY query that also uses a WHERE clause.

For each store, find the average price of all beverages that cost more than 0.97. Note the answers are not quite exact because I defined the price attribute as a real number and real arithmetic isn't as precise (as using integer arithmetic, for example).

```
select store, AVG(price) from store where price > 0.97 group by store
```

Query Results

| store | avg |
|--------------|------------------|
| Circle K | 2.0071428162711 |
| Safeway | 1.80000001192093 |
| Plaid Pantry | 1.79999999205271 |
| 7-11 | 1.08000001907349 |

8. Here's an example of a GROUP BY query that uses a WHERE clause and a HAVING clause. You can see that 7-11 drops out of the answer (compared to the previous query) because it's average price (of prices above 0.97) is only \$1.08

For each store where the average price of beverages that cost more than 0.97 is greater than \$1.50, list the store and the average price of beverages that cost more than 0.97.

```
select store, AVG(price) from store where price > 0.97 group by store having  
AVG(price) > 1.5
```

Query Results

| store | avg |
|----------|-----------------|
| Circle K | 2.0071428162711 |

Safeway 1.80000001192093
Plaid Pantry 1.79999999205271

3 row(s)