# Aggregation

## Lecture 7

Robb T. Koether

Hampden-Sydney College

Mon, Jan 29, 2018

# Aggregate Functions

- MySQL supports five aggregate functions.
  - `COUNT()`
  - `SUM()`
  - `AVG()`
  - `MAX()`
  - `MIN()`

# Aggregate Functions

## Aggregate Functions

```
SELECT COUNT(*), SUM(salary), AVG(salary), MIN(salary), MAX(salary)
FROM employees;
+----------+-------------+-------------+-------------+-------------+
| COUNT(*) | SUM(salary) | AVG(salary)  | MIN(salary) | MAX(salary) |
+----------+-------------+-------------+-------------+-------------+
|       12 |   595000.00 | 49583.333333 |    25000.00 |   100000.00 |
+----------+-------------+-------------+-------------+-------------+
```

- Each function applies to the tuples in the table produced by the `SELECT` query.
- The functions `SUM()`, `AVG()`, `MAX()`, and `MIN()` must be applied to specific attributes.

# Aggregate Functions

## Aggregate Functions

```
SELECT COUNT(*), SUM(salary), AVG(salary), MIN(salary), MAX(salary)
FROM employees WHERE dept = 2;
+----------+-------------+-------------+-------------+-------------+
| COUNT(*) | SUM(salary) | AVG(salary)  | MIN(salary) | MAX(salary) |
+----------+-------------+-------------+-------------+-------------+
|        5 |   265000.00 | 53000.000000 |    35000.00 |    80000.00 |
+----------+-------------+-------------+-------------+-------------+
```

- Each function applies to the tuples in the table produced by the `SELECT` query.
- The functions `SUM()`, `AVG()`, `MAX()`, and `MIN()` must be applied to specific attributes.

# The GROUP BY Clause

## The GROUP BY Clause

```
SELECT select_attribute_list
FROM table_name
WHERE condition
GROUP BY group_attribute_list
```

- We can use the GROUP BY clause to group the tuples according to one or more attributes.

# The GROUP BY Clause

## Group by Sex

```
SELECT sex, COUNT(*), MIN(bdate), MAX(bdate)
FROM employees
GROUP BY sex;
+------+----------+------------+------------+
| sex  | COUNT(*) | MIN(bdate) | MAX(bdate) |
+------+----------+------------+------------+
| F    |        6 | 1959-03-31 | 1985-12-02 |
| M    |        6 | 1955-03-17 | 1986-06-12 |
+------+----------+------------+------------+
```

- For example, we can group the employees by their sex.
- In such queries, we should also select the group-by attributes, but we should not select any attributes that vary within groups.

# The GROUP BY Clause

## Group by Sex and Department

```
SELECT dept, sex, COUNT(*), MIN(bdate), MAX(bdate)
FROM employees
GROUP BY dept, sex;
+------+------+----------+------------+------------+
| dept | sex  | COUNT(*) | MIN(bdate) | MAX(bdate) |
+------+------+----------+------------+------------+
|    1 | M    |        1 | 1974-02-15 | 1974-02-15 |
|    2 | F    |        2 | 1968-05-22 | 1985-12-02 |
|    2 | M    |        3 | 1966-11-24 | 1986-06-12 |
|    3 | F    |        4 | 1959-03-31 | 1985-10-12 |
|    3 | M    |        2 | 1955-03-17 | 1966-08-21 |
+------+------+----------+------------+------------+
```

- We can group by more than one attribute.
- This will create subgroups within groups.

# The HAVING Clause

## The HAVING Clause

```
SELECT select_attribute_list
FROM table_name
WHERE tuple_condition
GROUP BY group_attribute_list
HAVING group_condition
```

- The HAVING clause is the same as the WHERE clause, except that it applies to groups, not tuples.
- The WHERE clause applies only to tuples.

# The HAVING Clause

## Average Salary by Department

```
SELECT dept, AVG(salary)
FROM employees
GROUP BY dept
HAVING AVG(salary) >= 50000;
+------+----------------+
| dept | AVG(salary)    |
+------+----------------+
|    1 | 100000.000000 |
|    2 |  53000.000000 |
+------+----------------+
```

- Find all departments with an average salary of at least $50,000.

# The HAVING Clause

## Average Salary by Department

```
SELECT dept, AVG(salary)
FROM employees
WHERE salary >= 50000
GROUP BY dept
HAVING AVG(salary) >= 50000;
+------+----------------+
| dept | AVG(salary)    |
+------+----------------+
|    1 | 100000.000000 |
|    2 |  63333.333333 |
|    3 |  70000.000000 |
+------+----------------+
```

- How does this example differ from the previous example?

# The HAVING Clause

## Group by Social Security Number

```
SELECT fname, lname, COUNT(*)
FROM employees AS e, dependents AS d
WHERE e.ssn = d.ssn
GROUP BY e.ssn
HAVING COUNT(*) >= 2
ORDER BY lname;
+----------+---------+----------+
| fname    | lname   | COUNT(*) |
+----------+---------+----------+
| Frank    | Gilbert |        2 |
| James    | Green   |        2 |
| Jennifer | Wallace |        3 |
+----------+---------+----------+
```

- Find all employees who have at least 2 dependants, ordered by last name.

# Practice

- Find all female employees who have at least 2 dependents.
- Find all employees who have no dependents.
- Display the average number of children among the employees.