Aggregate functions

- Functions **count** (number of tuples), **sum** (sum), **avg** (average), **min** (minimum) and **max** (maximum) can be applied to a set of numbers given as a column of a relation.

- no other aggregate functions in SQL2

- If the keyword **distinct** is used in front of the aggregation attribute, first duplicates are eliminated before the aggregate is computed.

- examples:
  - How many students are there with different names?
    
    ```sql
    select count (distinct name) from students
    ```
  - How many professors have a name starting with “Sch”?
    
    ```sql
    select count (*) from professors where name like “Sch%”
    ```
  - What is the average number of semesters of all students?
    
    ```sql
    select avg (sem) from students
    ```
Grouping

- general form of the select-from-where clause
  
  ```sql
  select
  from
  [where <condition>]
  [group by <group-by-expression> [, <group-by-expression>]*
   [having <condition>]]
  [order by <order-expression>]
  ```

- **group by** clause
  - A “group-by-expression” is an expression that refers only to those attributes that are *not* used for computing the aggregate. Tuples with equal values for the specified value are summarized in groups (partitions).
  - For each group the query produces a new tuple in the result relation. Hence, only attributes with one value per group are permitted after the `select` clause.

- **having** clause
  - choice of groups with respect to a condition which may contain only arguments with one value per group
- Determine the number of hours per week in which professors have given lectures.

```sql
select held_by, sum(hpw) as number
from lectures
group by held_by
```

- Determine the number of hours per week of those lectures held by professors who predominantly give long lectures (> 2 hours per week on average).

```sql
select held_by, sum(hpw) as number
from lectures
group by held_by
having avg(hpw) > 2
```

- Determine the number of hours per week of those lectures held by C4 professors who predominantly give long lectures (> 2 hours per week on average).

```sql
select held_by, name, sum(hpw) as number
from lectures, professors
where held_by = pers-id and rank = "C4"
group by held_by, name
having avg(hpw) > 2
```