Grouping the aggregated rows

- So, an aggregate function performs a calculation for multiple rows so that the end result is a single value
- If the result table always contains just a single row, how can we write a query such as, *what's the average grade from each course?*
- To achieve this, we need to *group* the rows and perform the aggregate function for each group separately
- This can be done using the `GROUP BY` statement

The GROUP BY statement

- The `GROUP BY` statement uses a column or a group of columns to form groups of rows which the aggregate function operators on:

```
-- what's the average grade from each course?
SELECT course_code, AVG(grade) as average_grade FROM CourseGrade
-- form the groups based on the course_code
GROUP BY course_code
```

The GROUP BY statement

- The result table will have a row for *each distinct column value* of the `GROUP BY` column
- Each row has the corresponding aggregate function result for that group
- In the example's case the average grade for each course code:

course_code	average_grade
a290	2
a450	3
...	...

The GROUP BY statement

- As mentioned, the `GROUP BY` statement can have multiple columns
- In this case the result table will have a row for *each distinct combination of column values* of the `GROUP BY` columns

```
-- what's the average grade from each course instance?
SELECT course_code, instance_number, AVG(grade) as average_grade FROM CourseGrade
-- form the groups based on the course_code and instance_number
GROUP BY course_code, instance_number
```

The GROUP BY statement

- It is worth noting that in the `SELECT` statement we can only select columns that are either aggregate functions or columns used in the `GROUP BY` statement:

```
-- ❌ student_number is not an aggreagate function, nor it is in the GROUP BY statement.
-- This will lead into an error
SELECT course_code, student_number, AVG(grade) as average_grade FROM CourseGrade
GROUP BY course_code
```

- This causes the following error:

> Column 'CourseGrade.student_number' is invalid in the select list because it is not contained in either an aggregate function or the GROUP BY clause