DISTINCT

- Select only distinct values

```
SELECT DISTINCT <columns>

FROM <table>
```

Aggregation functions

- Work on the columns of a table:

  `SELECT FUNCTION_NAME(columns_name) <etc.>`

- `AVG` – calculate average
- `MIN` – find minimum value
- `MAX` – find maximum value
- `SUM` – sum the values
- `COUNT` – count the number of selected rows

Aggregation functions

- e.g. find max average high:

```
SELECT MAX(AvgHigh) FROM Temperature
```

- e.g. also find the city where the max average high occurs:

```
SELECT City, MAX(AvgHigh) FROM Temperature
```

- e.g. find number of cities with an average low of less than 0

```
SELECT COUNT(City) FROM Temperature WHERE AvgLow
< 0
```

GROUP BY

- Often used with aggregate functions
- Group records according to the value of a given column/variable

- Example: you may want to know the maximum AvgHigh for each province, rather than overall

GROUP BY

```
SELECT <cols>, FUNCTION(<col>)

FROM <table>

WHERE <query conditions>

GROUP BY <col>
```

GROUP BY

```sql
SELECT Province, MAX(AvgHigh)

FROM Geography JOIN Temperature ON
Geography.City = Temperature.City

GROUP BY Province
```

GROUP BY

```
>>> run_query(db, 'SELECT Province, MAX(AvgHigh)
FROM Geography JOIN Temperature ON
Geography.City = Temperature.City GROUP BY
Province')
```

[('Alberta', 10.3), ('BC', 13.9), ('Manitoba', 8.1), ('NWT', -0.8), ('New Brunswick', 11.0), ('Newfoundland', 8.6), ('Nova Scotia', 10.7), ('Ontario', 12.6), ('PEI', 9.5), ('Quebec', 10.9), ('Saskatchewan', 8.9), ('Yukon', 4.1)]