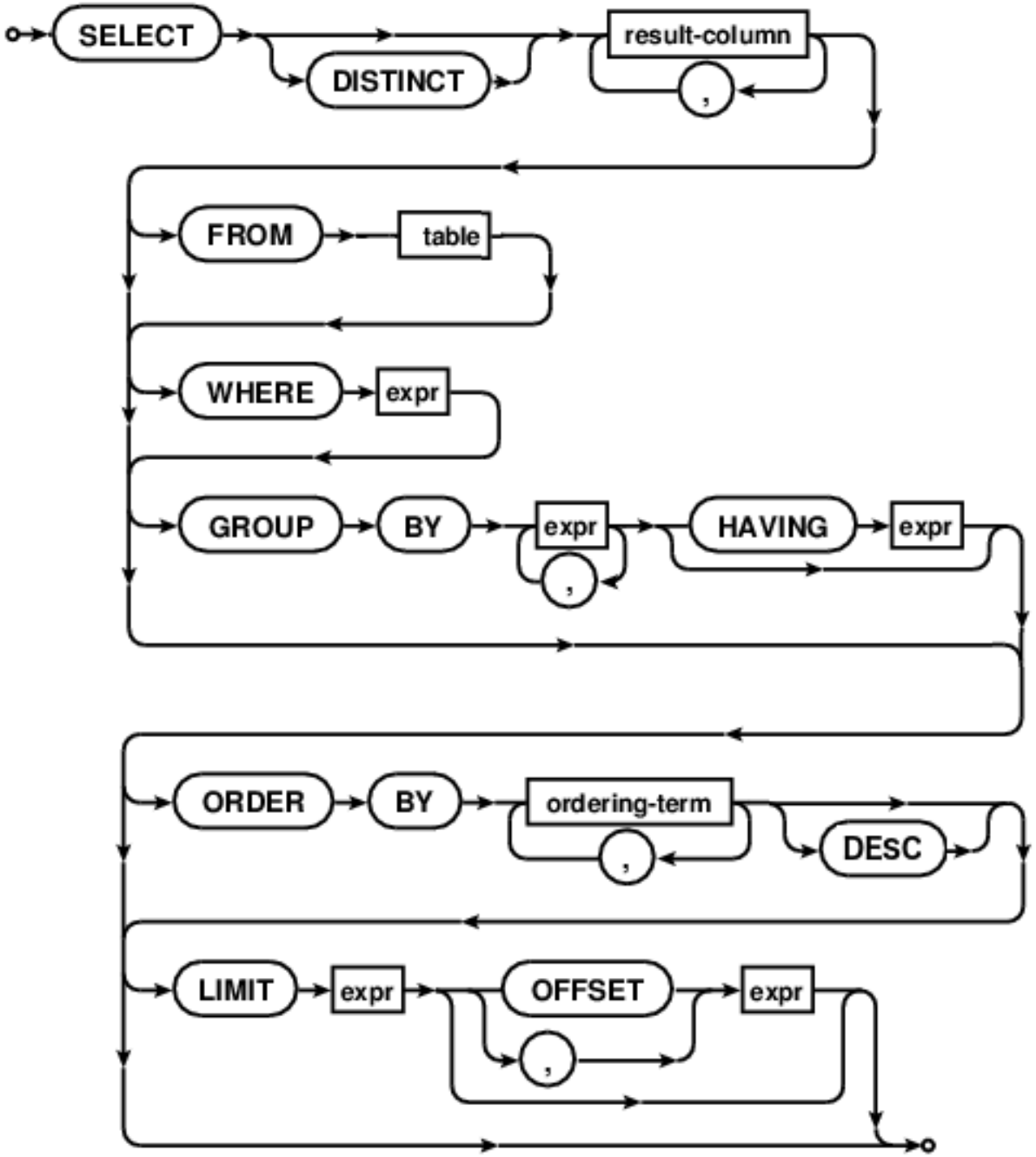
CS 217 Data Management and Information Processing

Structured Query Language 2


SELECT
DISTINCT
result-column
,
FROM
table
WHERE
expr
GROUP
BY
expr
,
HAVING
expr
ORDER
BY
ordering-term
,
DESC
LIMIT
expr
OFFSET
expr
,

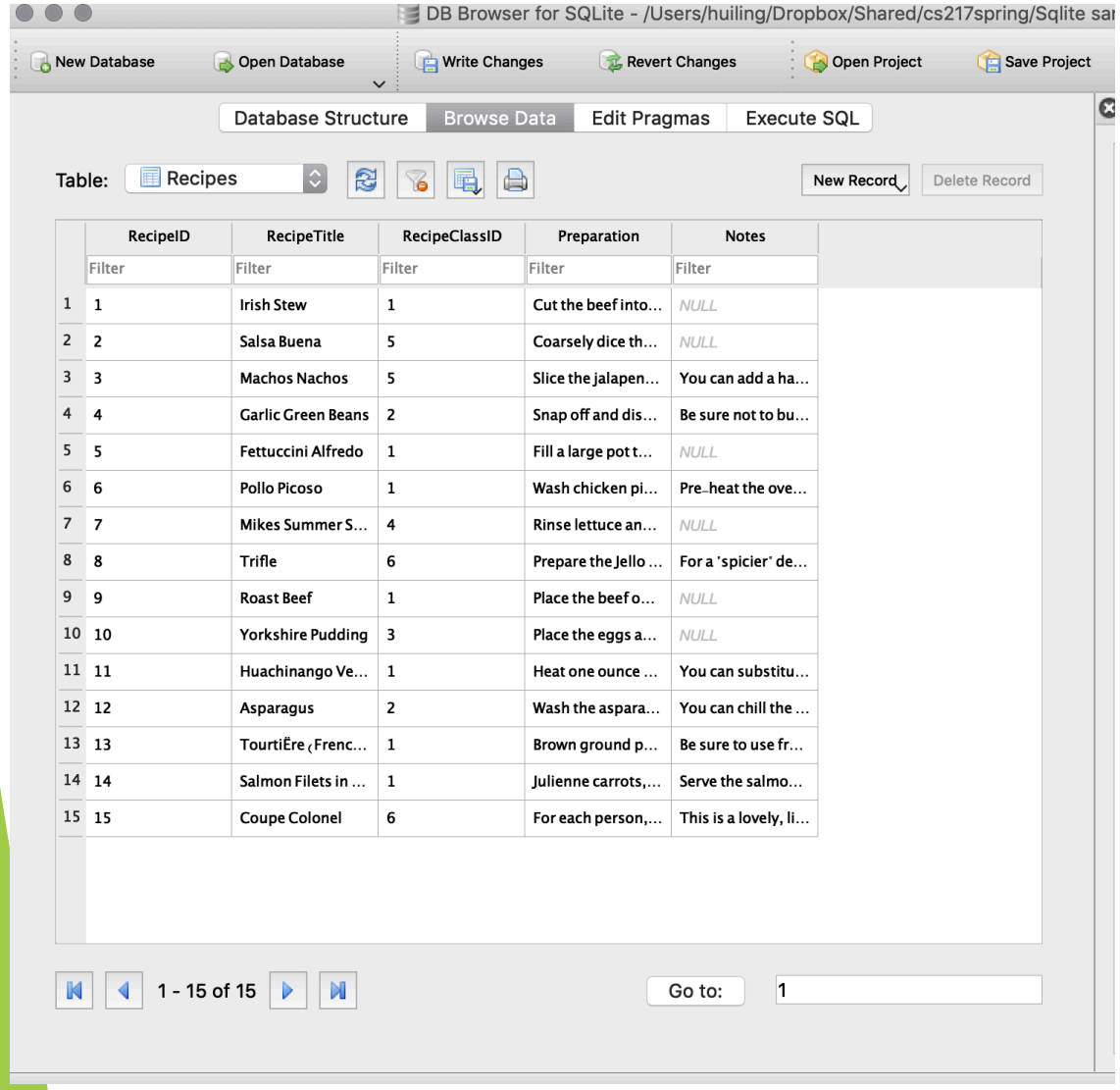
Accessing Databases

Tools to Inspect Database Content

- DB Browser for SQLite
  - Download it here: https://sqlitebrowser.org/dl/
  - Available on all platforms.
  - Download and install it!

- Use sqlite package in python

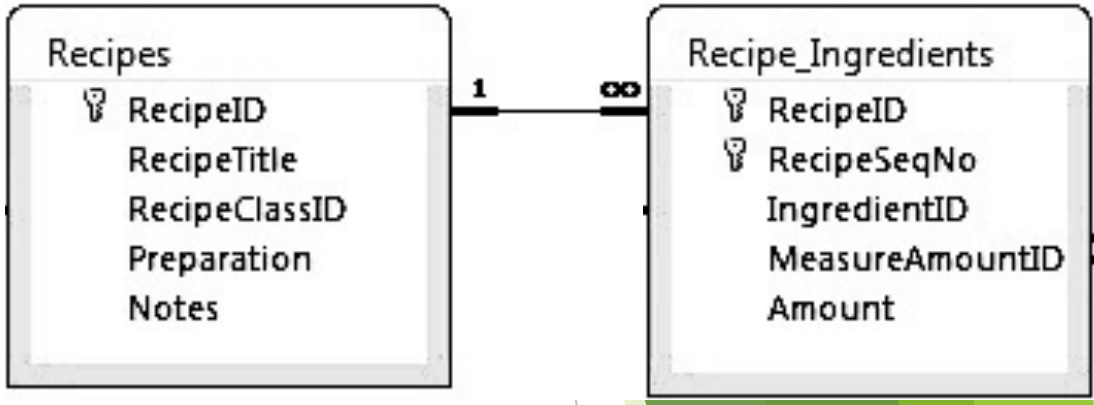
DB Browser for SQLite



Sqlite datasets and schema uploaded on Canvas!

Aggregation, Operators and Subquery

Aggregation functions



- COUNT, SUM, MIN, MAX, AVG
- **Case-insensitive**
- These can be used to print out values that depend on multiple rows.
- For example, how much ingredients are used for Irish stew?
  - We have to add up the "Amount" from many rows to get this answer:

```
SELECT SUM(Amount) FROM Recipe_Ingredients
    WHERE RecipeID=1;
```

  - ("Irish stew" corresponds to RecipeID=1)

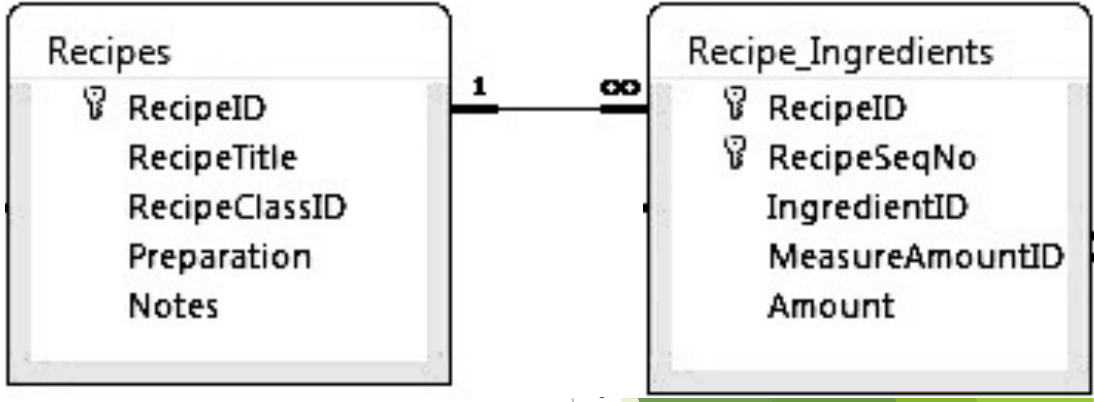
- Normally, aggregation applies to all the rows, but…
- GROUP BY causes aggregations to occur on subsets of rows, where rows are grouped according to some rule.
  - Each group contains rows having the same value for the grouping expression
  -
```
SELECT SUM(Amount) FROM Recipe_Ingredients
  GROUP BY RecipeID;
```
  - Same as above, but list amounts for Recipes

More GROUP BY Next week!

Integer vs. floating point division

- Computers store numbers in two basic ways:
  - Integers are whole numbers (0, 3, -40,921)
  - Floating Point numbers (floats) can be fractional (1.234, 0.0, -9.9×10^{-4})
- When doing arithmetic on two integers, an integer is always produced.
  - 1+1 = 2, 2-1=1, 4*3=12, 13/4=3
- Integer division is weird – it always rounds down: 2/3 = 0,-5/2 = -3
- When doing arithmetic involving at least one float, a float is produced.
  - 1.0 + 1.0 = 2.0, 1.5 * 2 = 3.0, 13/4.0=3.25
- Usually you need floating-point (not integer) division in your queries.
  - Just precede the expression with a floating point operation to force the division to be floating point: 1.0 * -5 / 2 = -2.5

Commonly used operators



- AND, OR

```
SELECT * FROM customers WHERE City = "Chicago” AND
(State = "Illinois” OR State = "IL");
```

- IN, NOT IN

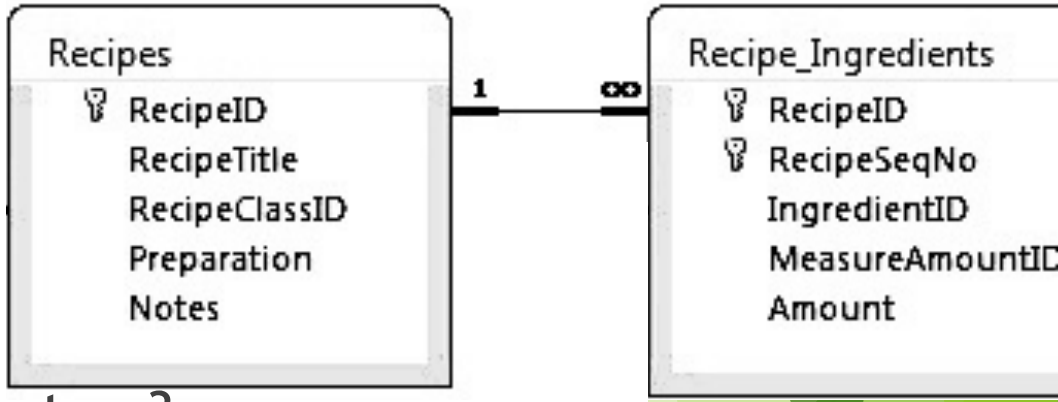
```
SELECT * FROM Recipe_Ingredients WHERE IngredientID
NOT IN (2, 10);
```

- LIKE
  - To search for a specified pattern in a column.

```
SELECT * FROM Recipes WHERE RecipeTitle LIKE "% %";
```

Subqueries

- Any single value, list of values, or table can be replaced by a subquery
- A subquery is a query that appears inside of parentheses.
  - The subquery is computed first and its result is “plugged into” the parent expression.



- E.g., how much ingredients are used for Irish stew?
  - `SELECT SUM(Amount) FROM Recipe_Ingredients WHERE RecipeID=1;`
  - `SELECT SUM(Amount) FROM Recipe_Ingredients WHERE RecipeID=`

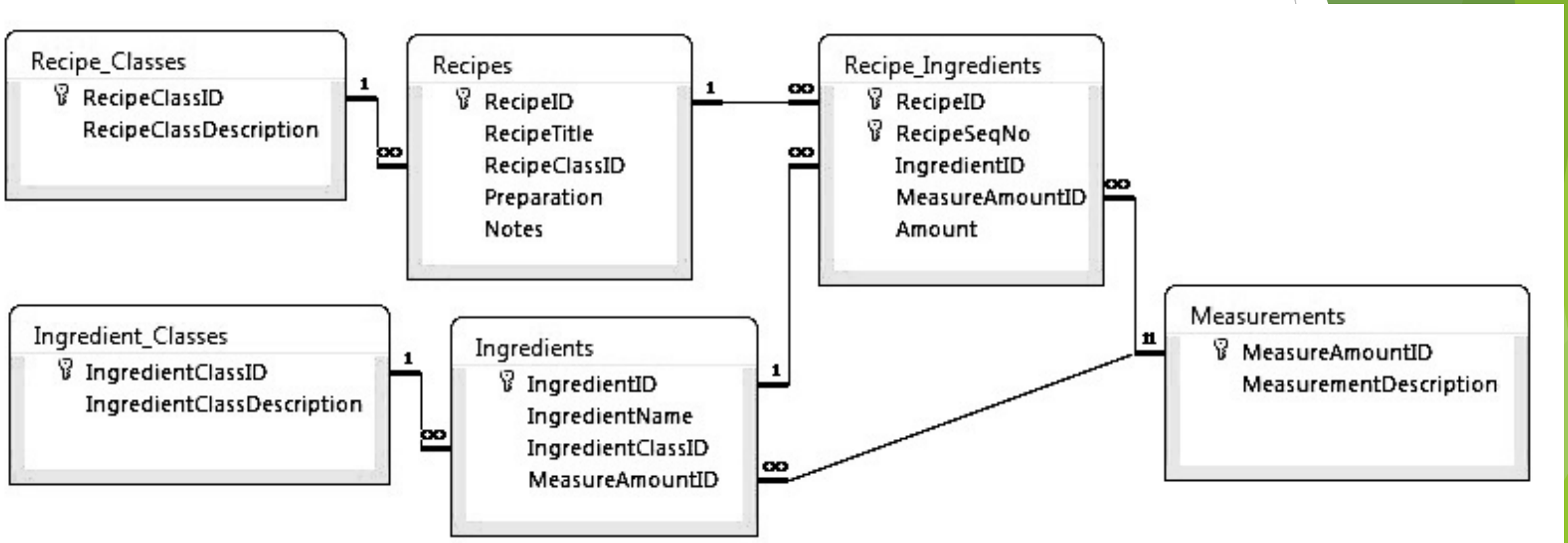
```
(SELECT RecipeID FROM Recipe WHERE RecipeTitle="Irish stew");
```

Examples and Demos

Recipes.sqlite

SalesOrders.sqlite

Recipes.sqlite (download it from Canvas)



Recipes.sqlite (download it from Canvas)

- Print an alphabetically sorted list of ingredients (hint: ORDER BY).
- How many times is butter used as an ingredient?
- How many ingredients are in the Yorkshire Pudding recipe?
- What percentage of ingredients are vegetarian? Vegan?
- How many recipes have multi-word names? Nine-letter names?

Print an alphabetically sorted list of ingredients (hint: ORDER BY).

How many times is butter used as an ingredient?

How many ingredients are in the Yorkshire Pudding recipe?

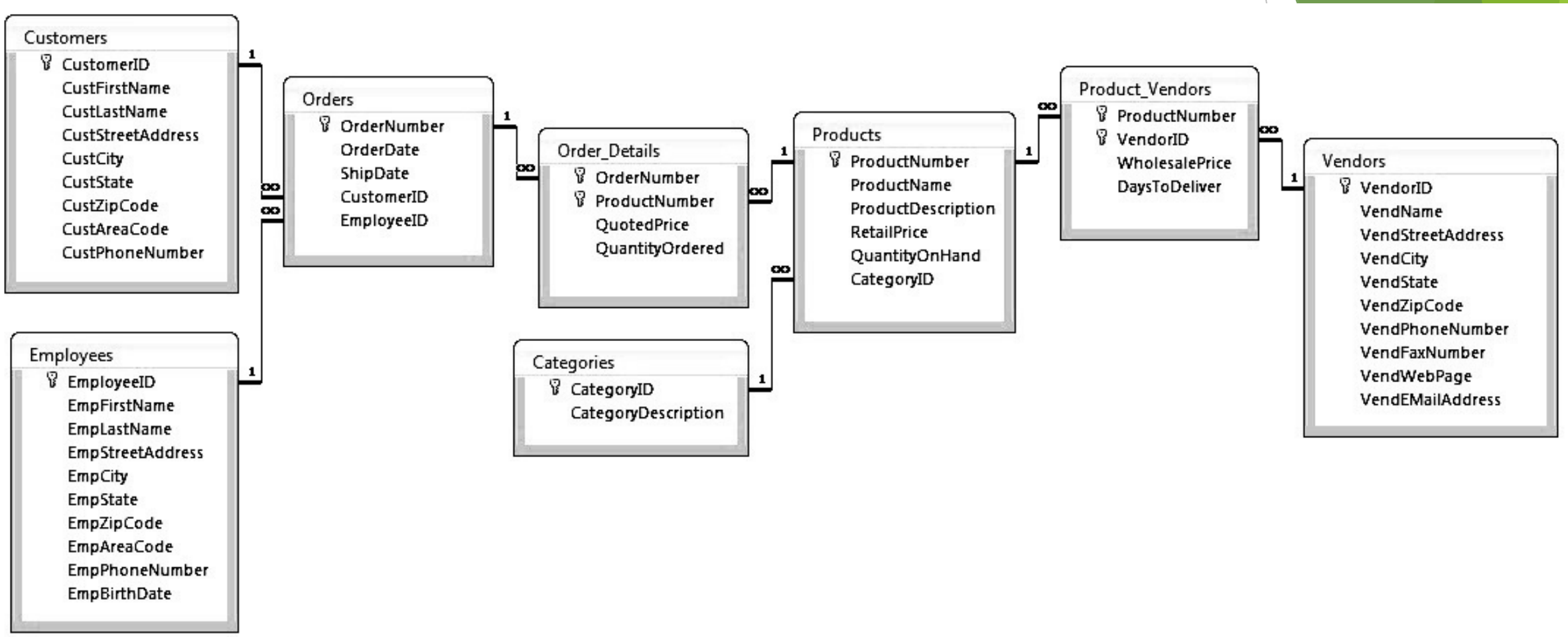
What percentage of ingredients are vegetarian?

How many recipes have multi-word names? Nine-letter names?

Recipes.sqlite (answers)

- Print an alphabetically sorted list of ingredients (hint: ORDER BY).
  - `SELECT IngredientName FROM Ingredients ORDER BY IngredientName;`
- How many times is butter used as an ingredient?
  - `SELECT COUNT(*) FROM Recipe_Ingredients WHERE IngredientID=47;`
- How many ingredients are in the Yorkshire Pudding recipe?
  - `SELECT COUNT(*) FROM Recipe_Ingredients WHERE RecipeID=10;`
- What percentage of ingredients are vegetarian?
  - `SELECT 100.0* COUNT(*)/(SELECT COUNT(*) FROM Ingredients) FROM Ingredients WHERE IngredientClassID NOT IN (2, 10);`
- How many recipes have multi-word names? Nine-letter names?
  - `SELECT COUNT(*) FROM Recipes WHERE RecipeTitle LIKE "% %";`
  - `SELECT COUNT(*) FROM Recipes WHERE RecipeTitle LIKE "_________";`
  - `SELECT COUNT(*) FROM Recipes WHERE LENGTH(RecipeTitle) = 9;`

SalesOrders.sqlite



SalesOrders.sqlite

- List all customers in California (CA). Count them.
- List all customers in a west coast state (CA, OR, WA).
- Count the unique customer area codes in California (CA).
- What is the full address of customer John Viescas?
- What is the most expensive product? Cheapest? Cheapest 5?
- What is the total value of the products on hand? Bikes on hand?

List all customers in California (CA). Count them.

List all customers in a west coast state (CA, OR, WA).

Count the unique customer area codes in California (CA).

What is the full address of customer John Viescas?

What is the single most expensive product? Cheapest? Cheapest 5?

What is the total value of the products on hand? Bikes on hand?

SalesOrders.sqlite (answers)

- List all customers in California (CA).

```
SELECT CustFirstName || " " ||CustLastName FROM Customers WHERE CustState = "CA";
```

- List all customers in a west coast state (CA, OR, WA).

```
SELECT * FROM Customers WHERE CustState IN ("CA","OR", "WA");
```

- Count the unique customer area codes in California (CA).

```
SELECT COUNT(DISTINCT CustAreaCode) FROM Customers
    WHERE CustState = "CA";
```

- What is the full address of customer John Viescas?

```
SELECT CustStreetAddress || "
" || CustCity || " " || CustState || " " || CustZipCode AS FullAddress FROM Customers
WHERE CustFirstName = "John" AND CustLastName = "Viescas";
```

- What is the single most expensive product? Cheapest 5?
  - `SELECT ProductNumber, ProductName FROM Products WHERE RetailPrice = (SELECT MAX(RetailPrice) FROM Products);`
  - `SELECT ProductName, RetailPrice FROM Products ORDER BY RetailPrice LIMIT 5;`
- What is the total value of the products on hand? Bikes on hand?
  - `SELECT SUM(RetailPrice * QuantityOnHand) FROM Products;`
  - `SELECT SUM(RetailPrice * QuantityOnHand) FROM Products WHERE CategoryID=2;`