Calculating New Values

Objectives

- Write queries that calculate new values for each selected record.

After carefully re-reading the expedition logs, we realize that the radiation measurements they report may need to be corrected upward by 5%. Rather than modifying the stored data, we can do this calculation on the fly as part of our query:

```
%load_ext sqlitemagic
```

```
%%sqlite survey.db
select 1.05 * reading from Survey where quant='rad';
```

```
10.311
8.19
8.8305
7.581
4.5675
2.2995
1.533
11.8125
```

When we run the query, the expression `1.05 * reading` is evaluated for each row. Expressions can use any of the fields, all of usual arithmetic operators, and a variety of common functions. (Exactly which ones depends on which database manager is being used.) For example, we can convert temperature readings from Fahrenheit to Celsius and round to two decimal places:

```
%%sqlite survey.db
select taken, round(5*(reading-32)/9, 2) from Survey where quant='temp';
```

```
734 -29.72
735 -32.22
751 -28.06
752 -26.67
```

We can also combine values from different fields, for example by using the string concatenation

operator `||`:

```
%%sqlite survey.db
select personal || ' ' || family from Person;
```

```
William Dyer
Frank Pabodie
Anderson Lake
Valentina Roerich
Frank Danforth
```

It may seem strange to use `personal` and `family` as field names instead of `first` and `last`, but it's a necessary first step toward handling cultural differences. For example, consider the following rules:

Full Name	Alphabetized Under	Reason
Liu Xiaobo	Liu	Chinese family names come first
Leonardo da Vinci	Leonardo	"da Vinci" just means "from Vinci"
Catherine de Medici	Medici	family name
Jean de La Fontaine	La Fontaine	family name is "La Fontaine"
Juan Ponce de Leon	Ponce de Leon	full family name is "Ponce de Leon"
Gabriel Garcia Marquez	Garcia Marquez	double-barrelled Spanish surnames
Wernher von Braun	von *or* Braun	depending on whether he was in Germany or the US
Elizabeth Alexandra May Windsor	Elizabeth	monarchs alphabetize by the name under which they reigned
Thomas a Beckett	Thomas	and saints according to the names by which they were canonized

Clearly, even a two-part division into "personal" and "family" isn't enough...

Challenges

1. After further reading, we realize that Valentina Roerich was reporting salinity as percentages. Write a query that returns all of her salinity measurements from the `Survey` table with the values divided by 100.

2. The `union` operator combines the results of two queries:

```
%%sqlite survey.db
select * from Person where ident='dyer' union select * from Person where ident='roe';
```

```
dyer    William   Dyer
roe     Valentina Roerich
```

Use `union` to create a consolidated list of salinity measurements in which Roerich's, and only Roerich's, have been corrected as described in the previous challenge. The output should be something like:

619 0.13

622 0.09

734 0.05

751 0.1

752 0.09

752 0.416

837 0.21

837 0.225

3. The site identifiers in the `Visited` table have two parts separated by a '-':

```
%%sqlite survey.db
select distinct site from Visited;
```

```
DR-1
DR-3
MSK-4
```

Some major site identifiers are two letters long and some are three. The "in string" function `instr(X, Y)` returns the 1-based index of the first occurrence of string Y in string X, or 0 if Y does not exist in X. The substring function `substr(X, I)` returns the substring of X starting at index I. Use these two functions to produce a list of unique major site identifiers. (For this data, the list should contain only "DR" and "MSK").

Key Points

- SQL can perform calculations using the values in a record as part of a query.