

# Sorting and Removing Duplicates

## Objectives

- Write queries that display results in a particular order.
- Write queries that eliminate duplicate values from data.

Data is often redundant, so queries often return redundant information. For example, if we select the quantities that have been measured from the `survey` table, we get this:

```
%load_ext sqlitemagic
```

```
%%sqlite survey.db  
select quant from Survey;
```

```
rad  
sal  
rad  
sal  
rad  
sal  
temp  
rad  
sal  
temp  
rad  
temp  
sal  
rad  
sal  
temp  
sal  
rad  
sal  
sal  
rad
```

We can eliminate the redundant output to make the result more readable by adding the `distinct` keyword to our query:

```
%%sqlite survey.db
select distinct quant from Survey;
```

```
rad
sal
temp
```

If we select more than one column—for example, both the survey site ID and the quantity measured—then the distinct pairs of values are returned:

```
%%sqlite survey.db
select distinct taken, quant from Survey;
```

```
619 rad
619 sal
622 rad
622 sal
734 rad
734 sal
734 temp
735 rad
735 sal
735 temp
751 rad
751 temp
751 sal
752 rad
752 sal
752 temp
837 rad
837 sal
844 rad
```

Notice in both cases that duplicates are removed even if they didn't appear to be adjacent in the database. Again, it's important to remember that rows aren't actually ordered: they're just displayed that way.

## Challenges

1. Write a query that selects distinct dates from the `Site` table.

As we mentioned earlier, database records are not stored in any particular order. This means that query results aren't necessarily sorted, and even if they are, we often want to sort them in a different way, e.g., by the name of the project instead of by the name of the scientist. We can do this in SQL by adding an `order by` clause to our query:

```
%%sqlite survey.db
select * from Person order by ident;
```

```
danforthFrank    Danforth
dyer      William Dyer
lake      Anderson Lake
pb        Frank    Pabodie
roe       ValentinaRoerich
```

By default, results are sorted in ascending order (i.e., from least to greatest). We can sort in the opposite order using `desc` (for "descending"):

```
%%sqlite survey.db
select * from person order by ident desc;
```

```
roe       ValentinaRoerich
pb        Frank    Pabodie
lake      Anderson Lake
dyer      William Dyer
danforthFrank    Danforth
```

(And if we want to make it clear that we're sorting in ascending order, we can use `asc` instead of `desc`.)

We can also sort on several fields at once. For example, this query sorts results first in ascending order by `taken`, and then in descending order by `person` within each group of equal `taken` values:

```
%%sqlite survey.db
select taken, person from Survey order by taken asc, person desc;
```

```
619 dyer
619 dyer
622 dyer
622 dyer
734 pb
734 pb
734 lake
735 pb
735 None
735 None
751 pb
751 pb
751 lake
752 roe
752 lake
752 lake
752 lake
837 roe
837 lake
837 lake
844 roe
```

This is easier to understand if we also remove duplicates:

```
%%sqlite survey.db
select distinct taken, person from Survey order by taken asc, person
desc;
```

```
619 dyer
622 dyer
734 pb
734 lake
735 pb
735 None
751 pb
751 lake
752 roe
752 lake
837 roe
837 lake
844 roe
```

## Challenges

1. Write a query that returns the distinct dates in the `Visited` table.
2. Write a query that displays the full names of the scientists in the `Person` table, ordered by

family name.

## Key Points

- The records in a database table are not intrinsically ordered: if we want to display them in some order, we must specify that explicitly.
- The values in a database are not guaranteed to be unique: if we want to eliminate duplicates, we must specify that explicitly as well.

---

**Email** ([admin@software-carpentry.org](mailto:admin@software-carpentry.org)) **Twitter** (<https://twitter.com/swcarpentry>)  
**RSS** (<http://software-carpentry.org/feed.xml>) **GitHub** (<https://github.com/swcarpentry>) **IRC** (<irc://moznet/sciencelab>)  
**License** ([../LICENSE.html](http://software-carpentry.org/02-sort-dup.md))  
**Bug Report** (<mailto:admin@software-carpentry.org?subject=bug%20in%20novice/sql/02-sort-dup.md>)