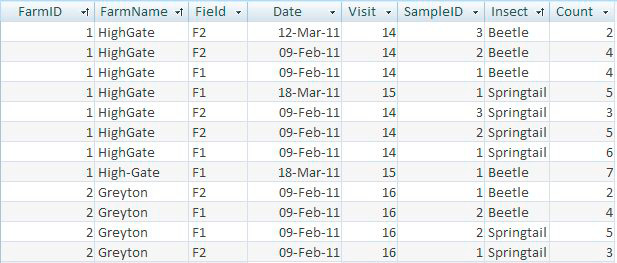
David Thomsen

Exercise 8-1.

[Example 1-3](https://learning-oreilly-com.cobalt.champlain.edu/library/view/beginning-database-design/9781430242093/9781430242093_Ch01.xhtml#ex1-3) back in [Chapter 1](https://learning-oreilly-com.cobalt.champlain.edu/library/view/beginning-database-design/9781430242093/9781430242093_Ch01.xhtml) is a good real–life example of unnormalized data. To recap: Farms are visited and a number of samples are taken from different fields. The number of each species ( just Springtail and Beetle for now) in each sample is recorded. A version of the data is shown in [Figure 8-14](https://learning-oreilly-com.cobalt.champlain.edu/library/view/beginning-database-design/9781430242093/9781430242093_Ch08.xhtml#F14).



[**Figure 8-14.**](https://learning-oreilly-com.cobalt.champlain.edu/library/view/beginning-database-design/9781430242093/9781430242093_Ch08.xhtml#FG14) Unnormalized version of insect data

Consider the following questions:

a) What are some of the updating problems that could occur with the table in [Figure 8-14](https://learning-oreilly-com.cobalt.champlain.edu/library/view/beginning-database-design/9781430242093/9781430242093_Ch08.xhtml#F14)?

* Inconsistent Data
  + Within Farm Name there is a data discrepancy between “HighGate” and “High-Gate”
* Too many fields
  + The table above holds the information for both the Farm information as well as the Sample information. For ease of joining and querying, this should be split into two tables, a farm table and a sample table.
* The Visit data is currently inconsistent as the visit number is supposed to detail the number of the visit which can be cross checked with the date. In this table it says that visit 14 took place on both the 12th of March as well as the 9th of February. Visit 16 seems to be standardized but I believe there is too much information within this one table and the best move is to split it up.
* The formatting for the date column is also not properly formatted as the proper DATE YYYY-MM-DD syntax. The table’s formatting also does not align with the DATETIME YYYY-MM-DD HH:MI:SS syntax. This could cause issues with comparing the dates as these are likely stored as string values which cannot be compared as easily as the others mentioned.
  + Source: <https://www.w3schools.com/sql/sql_dates.asp>

b) Which of the following functional dependencies hold for the insect data?

FarmID → FarmName?

FarmID → Visit?

Visit → Date?

Date → Visit?

Visit → FarmID?

Sample → Field?

(Sample, VisitID) → Field?

(Sample, Insect) → Count?

Of those dependencies only three of them are functional and hold for the insect data.

**FarmID → FarmName**

The FarmID directly determines the Farm Name so this is a functional dependency.

**Visit → FarmID**

The Visit is dependent on the Farm ID as each visit needs to be at a farm, but one farm does not NEED any visits. This is a functional dependency.

**(Sample, VisitID) → Field**

Both a sample and a visit must take place in a Field. This is a functional dependency.

c) (VisitID, Sample, Insect) → Count? (VisitID, Sample, Insect) is suggested as an appropriate primary key. Can you determine all the other values from knowing the values of these three fields? Would it be a suitable primary key?

This is an acceptable primary key as there is a way to definitively identify the row of the data based on exclusively these fields. For example, if the primary key for an entry is (14, 2, beetle), there is only one row that this corresponds to and this is true for any combination of these keys. In the table as it stands, this is a suitable key but with many additional issues in this table, I would propose rebuilding the table into two separate tables; One holding visit information and the other holding insect information.

d) Using the fields in Part C as a primary key use the normalization rules to decompose the table in [Figure 8-14](https://learning-oreilly-com.cobalt.champlain.edu/library/view/beginning-database-design/9781430242093/9781430242093_Ch08.xhtml#F14) into a set of tables in third normal form.

We have created three main tables to decompose the insect data in a more manageable way.

Below is how we decided to break this table down:

Visit, SampleID, Insect, Date, Count : PK – Visit, SampleID, Insect

Visit, SampleID, Field : PK – Visit, SampleID

Visit, FarmID, FarmName : PK – Visit

