

# Lab 5, Making Sense of Large Datasets

CS 0131, Software for Personal Computing  
Timothy J Parenti

18 February 2013

So far, we’ve seen that Excel is good for arbitrary calculations — and powerful ones at that — but sometimes we have large datasets to deal with. In these cases, formulas can get unwieldy, but structuring our data as a table can help us to better manage it. As we mentioned before, tables have *records* which run horizontally and refer to a single instance of a dataset, as well as *fields* which run vertically and refer to a single type of data included in all records.

It might help to think of a table as a large filing drawer filled with index cards. Each index card represents a record, a single entry in the table. The fields, then, are the individual bits of data written on each card. Perhaps these fields are written in the same format on each card. This lends structure to the data: The data in many fields constitutes a record, while many records constitute an entire dataset. We’ll come back to these ideas when we talk about databases, but it’s good to start thinking this way now.

## 1 Tracking Transactions

You are the accountant at The Spa Experts, and you’re going over the transactions made by your sales representatives last month. In order to more easily make sense of the raw data, you decide to convert it into an Excel table so that you can refer to fields within records using symbolic structured references. You also decide to augment the data with conditional formatting to make important records stand out.

1. Download `spa.xlsx` from today’s entry in the “Schedule” page of the course website and open the workbook.
2. In order to make our data easier to work with, we’ll turn it into a table with individual records and fields:
  - a. Select any cell in the data, then click `Insert >> Table`. Notice that the create table dialog box appears and the range **A5:J75** is selected as the data source for your table.
  - b. Make sure the checkbox that says “My table has headers” remains checked, then click `OK`.
3. Reduce the widths of columns **A**, **C**, and **D** so that the data fits but the headings can wrap.
4. Scroll down toward the bottom of the table. Notice that once the column headings of the table scroll off the top of the screen, they replace the default column headings (that is, the letters A through J) where they normally appear.
5. You need to apply a style to this table that matches the corporate brand image. Under `Table Tools >> Design >> Table Styles`, select *Table Style Medium 6*.
6. Sort the data by selecting the filter arrow next to *Sales Rep Last Name* and selecting “Sort A to Z”.
7. Now that your data is sorted by name, you realize that you’d actually like the sales for each person to be sorted by the amount of the sale:

- a. While a cell in the table is active, select **Home** > **Editing** > **Sort & Filter** > **Custom Sort**.
- b. Under the last name sort, add a new level of sorting by *Amount*, sorting the values from largest to smallest.
- c. Click **OK**.

**Note:** Always check your data to determine how many levels of sorting you need to apply. If your table, for example, contains several people with the same last name but different first names, you would first sort by the last name, then by first name. To ensure that Excel “breaks ties” in the sequence you desire, use the Sort dialog box rather than sorting the columns individually.

8. One of the nice things about converting our data into a table with records and fields is the ability to refer to our fields symbolically using “structured references,” which more clearly indicate what type of data is being used in the calculations:

- a. In cell **I6**, call the IF() function using from the library of Logical functions on the **Formulas** tab. A dialog box appears requesting the three arguments, *logical\_test*, *value\_if\_true*, and *value\_if\_false*.
- b. For *logical\_test*, type **[Payment Type]="Paid in Full"**.

**Huh?** What are you doing here? The square brackets around “Payment Type” are telling Excel you’re referring to this record’s field with the corresponding name; namely, the values in column **F**. You’re then asking as part of the logical test condition whether that value is equal to the string of text “Paid in Full”.

- c. If the condition is true, this means the client paid the full amount in the Amount column, so the formula for *value\_if\_true* should just be **[Amount]**.
- d. Otherwise, the client only paid the down payment, which is 25% of the amount. The formula for *value\_if\_false* should thus be **[Amount]\*\$D\$3**.
- e. Click **OK** to enter the formula. Note that, because you are entering the formula into a table, Excel copies the formula down automatically.

**Note:** One nice thing about structural references is that formulas are often much clearer when you can see exactly what types of data are being manipulated to achieve your results. In addition, since each calculation is effectively constrained to a single record, Excel does a lot of the referential heavy-lifting for us.

9. Likewise, in cell **J6**, type the formula **= [Amount] - [Down Payment]**. Notice that, as soon as you type the opening square bracket, Excel presents you with the list of fields to choose from.
10. You’d like a visual way to see at a glance which sales are the bigger ones:
  - a. Select the range **H6:H75**, which contains the total sale amounts.
  - b. Click **Home** > **Styles** > **Conditional Formatting** > **Data Bars**.
  - c. Select the *Red Data Bar* under the Gradient Fill heading.
11. You’re now interested in the spa sales with more than \$8,000 owed. Since you’re satisfied with your current sort, you’d rather bring attention to these records in another way using Conditional Formatting:
  - a. Select the range **J6:J75**, which contains the amounts owed.
  - b. Click **Home** > **Styles** > **Conditional Formatting** > **New Rule**.
  - c. Select “Use a formula to determine which cells to format”.
  - d. In the rule description, enter the formula **=J6>8000**. Note that references in this formula are relative to the first cell in the range you selected.
  - e. Click the **Format** button to specify the formatting for the cells that match this criterion.
  - f. On the “Fill” tab of the Format Cells dialog that appears, select the orange color at the top right.

- g. Click  to confirm the formatting, then click  to apply the conditional formatting rule.
12. We now want to focus only on sales made between 16 and 31 January 2013, hiding the others from view:
- a. Select the filter arrow next to *Date* and select  .
  - b. In the Custom AutoFilter dialog, specify that you wish to view rows where the date is after or equal to 1/16/2013, and before or equal to 1/31/2013.
  - c. Click  and observe that the records which fail to match this criterion are hidden from view.
13. Type your name in cell **G3** and save all changes to your workbook.

## Submission

To receive credit for these exercises, call over the instructor, who will check that you have completed the assignment. Then, log into CourseWeb and use the “Assignment Submission” section to submit your “spa.xlsx” file for **Lab 5**.

You should be able to complete this lab in the allotted class time; however, if you are running low on time, the instructor will give you further instructions for completing the rest of this lab at home.

**Your lab must be checked by the instructor BEFORE you leave the room AND you must submit your files to CourseWeb in order to receive credit for the lab! Don’t forget!**