3 solutions for dealing with duplicate data in Excel

Duplicate data in Excel can throw a monkey wrench into your information processing works. Here are three solutions for handling duplicate data in Excel 2010 and 2013.

1. Find duplicates

If you want to find duplicate data items in a column that should have only unique values, try Conditional Formatting. Select all the data in the column. Then, on the Home tab, Styles group, choose Conditional Formatting. From the Highlight Cells Rules fly-out menu, you’ll see Duplicate values near the bottom. Choose this, and the type of formatting you would like to use.

This won’t change or delete anything, but it will make it easier to spot the potentially offending items. This only applies to data items in a column, not records across multiple columns.

2. Remove duplicates

Remove duplicates goes beyond a single column and allows you to examine all data items in the row or certain ones to detect and delete a duplicate. For example, if you were trying to narrow your list down to households, rather than individual names in a list where multiple names might appear for a single address, you could use Remove Duplicates. From the Data tab, Data Tools groups, select the Remove Duplicates button. Uncheck any references to individual names, leaving only the address fields selected. It will keep the first entry of a set of duplicates and delete any remaining ones.

It’s important to make sure that the one you might want to keep, if you have a preference, is the first one in the list. One way to accomplish this is to add a column to your list and fill all values with a 2. Then, look for the ones you want to keep and change the 2 to a 1. Then sort your list in ascending order by that field. You can delete this extra column before or after remove duplicates once you’ve sorted it.
3. Finding duplicates with VLookUp

Remove Duplicates is great, but it does not tell you which ones have been deleted. If this is important to your work process, you can create a deleted items list using VLookUp. First, create a new column to the left of your list and name it something like Index. Type a 1 in the first cell of this new column and a 2 in the second. Select both cells you’ve just filled, and fill it down the rest of your list, creating a sequential “ID” number.

Now duplicate the worksheet in the same workbook. (Right-click the worksheet tab, choose Move or Copy..., check the Create a copy box.) Run Remove Duplicates on one of your worksheets, unchecking your Index column. Now perform a VLookUp in the original list to find all the numbers that match in the modified list. Those that read #N/A were deleted in the Remove Duplicates operation. That formula will look something like: