

SPREADSHEETS

(Data for this tutorial at www.peteraldhous.com/Data)

Spreadsheets are great tools for sorting, filtering and running calculations on tables of data. Journalists who know the basics can “interview” data to find stories and trends that others may miss.

Once you know your way around a spreadsheet, it’s fairly easy to pick up the skills needed to work with many other tools for data analysis.

TIP! Whenever you get some new data, save it under a new name! Then if you mess things up, you can always go back to the original.

So, let’s open the file **City Budget.xls** and save it under a new name.

Formatting, basic calculations and sorting

Spreadsheets use a simple co-ordinate system, with letters for columns and numbers for rows. You can make calculations on data using these coordinates.

This data was provided by [Mark Horvit](#), executive director of Investigative Reporters and Editors, and has been used by IRE for training on using spreadsheets in journalism.

1. Formatting data

First, let’s format the columns as currency. Select columns **B** and **C** by clicking on the letters while pressing the **Ctrl** key; then select **Format > Format Cells** from the ribbon in the Home tab; when the dialog box pops up, select **Currency**, **\$** and **0** decimal places.

1	Department	This year	Next year
2	City Council	105,207	95,452
3	City Clerk	70,778	75,234
4	City Manager	182,540	190,321
5	Municipal Clerk	195,645	222,409
6	Personnel	191,116	195,400
7	Planning and Zoning	305,607	315,640
8	Fire Department	2,456,789	2,537,901
9	Police Department	3,101,345	3,545,367
10	Public Works	2,156,987	2,104,866
11	Community Development	85,654	90,447
12	General City	1,234,823	1,482,950
13	Parks and Recreation	1,400,400	1,235,674
14	Health	1,033,188	1,179,243
15			
16	Totals	12,520,079	13,840,904
17			
18			
19	Revenues	This year	Next year
20	General Procurement	2,500,234	2,661,234
21	Sales Taxes	3,967,138	4,020,112
22	Other local	2,623,456	2,846,980
23	Licenses and Permits	264,675	289,620
24	Fees and Surcharges	330,345	350,900
25	Revenue from Investments	1,096,345	1,234,341
26	Building Revenue	188,000	192,000
27	Miscellaneous	158,782	162,234
28	Interest	1,231,209	1,323,456
29	Fines	234,645	250,786
30			
31	Totals	12,594,829	13,331,663
32			

So, now we have everything in \$, but we still can't see everything in column A, so let's format the column width.

Select the square at the top left, which selects all the data in the sheet. Then **Format>Autofit Column Width**

TIP! If ever you see a column filled with #####, it just means it contains a number that is wider than the column. Use the same trick to fix the problem.

2. Performing basic calculations

Now let's calculate how the budget has changed from this year to last year. All calculations in Excel start by entering = into the cell. So, the formula for first cell is =(C2-B2).

(At this stage, we may need to format the cells in the column so that they are also in \$.)

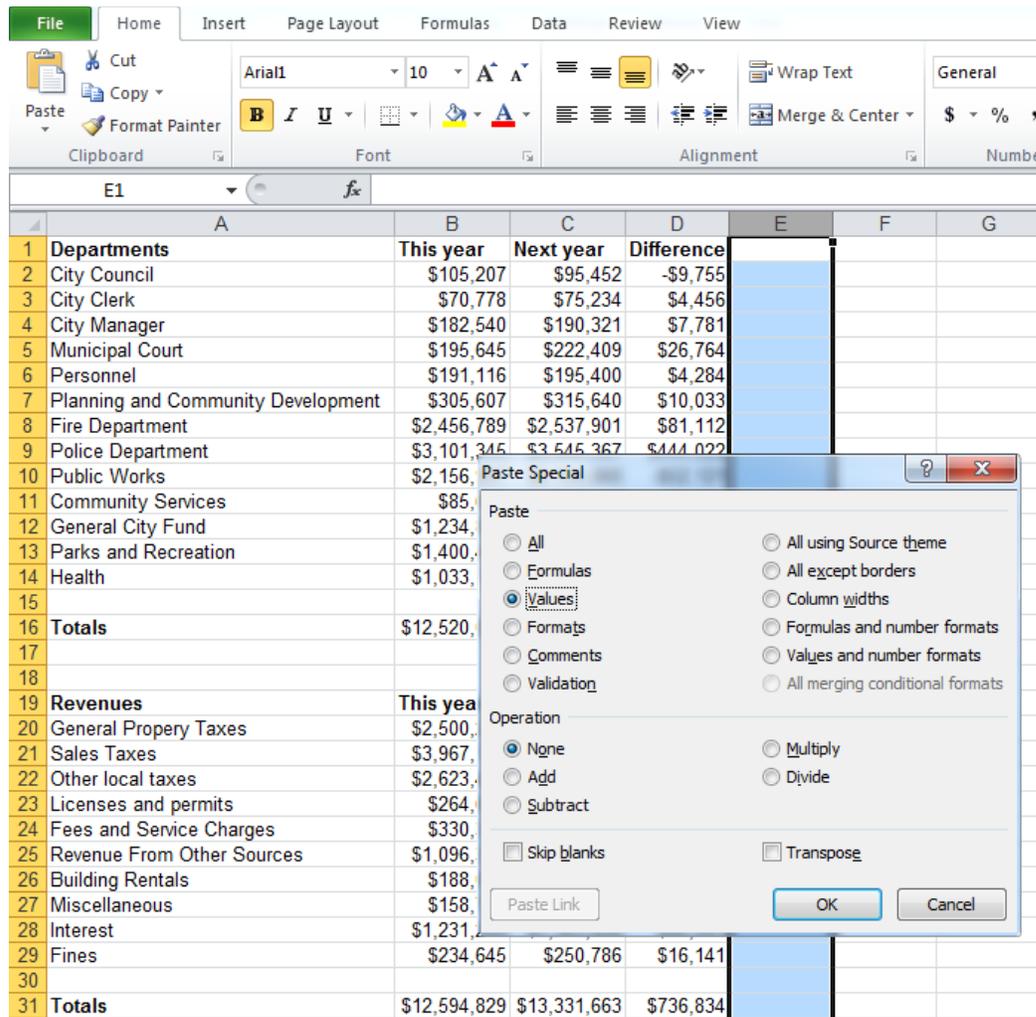
Move the cursor to the bottom right hand corner of the cell until you see it change to a cross, then double click. Excel autofills the column, performing the same calculation for each row until it hits a space.

TIP! Excel can recognize other patterns in your data and autofill in the same way. To number a list of records, for instance, type 1, 2, 3, into the first three cells next to a column of data, select those three cells and autofill.

You can also **Copy (Ctrl-C)** the first cell and **Paste (Ctrl-V)** all the way down the column. See how the "marching ants" form around the cell you've copied. (This will give 0s where the cells are blank, and #VALUE! errors as Excel tries to perform a calculation on text. These can simply be deleted.)

TIP! You can lose your work if you later delete data that was used to calculate values that you are still working with. If this is a possibility, Copy the calculated data and re-enter it by selecting Paste>Paste Special and clicking the Values button

So let's do that here:



We may need to format our new column to currency once more. See how the cells now contain the numbers, not the formulas used to calculate them.

Then delete the old column D by selecting the cells, then **Delete>Delete Cells** and selecting the button for **Shift Cells Left**.

Now let's calculate the percentage change in budget, rather than the absolute \$ value.

TIP! When adding a longish title to a column or row, go to **Format Cells**, select the **Alignment** tab and **Wrap text** to keep the formatting neat

The formula for the first cell is
 $=((C2-B2)/B2)$ or $=(D2/B2)$.

So let's enter this, and copy down the column as before. Again, we will need to **Format** the cells to have the results displayed as a percentage, rather than a decimal.

3. Sorting data

Let's sort by Percent change. To do so, highlight the range of data you want to sort, then

Select **Sort** in the **Data** tab. When the dialog box pops up, select **Sort By Percent Change** and **Order by Largest to Smallest**.

TIP! When sorting, make sure Excel has recognized that your data has a header row if there is one present, otherwise that will be sorted as well.

TIP! Make sure when sorting to select the full range of data you want to sort, and no more. Missing rows or columns from sorts, or including extraneous data, are common ways to scramble your data. Use the **Undo** arrow  if you make a mistake. In this case we want to sort separately on the **Departments** and **Revenues** tables, Taking care not to include the **Totals**.

	A	B	C	D	E	F	G	H	I
1	Departments	This year	Next year	Difference	Percent change				
2	City Council	\$105,207	\$95,452	-\$9,755	-9.27%				
3	City Clerk	\$70,778	\$75,234	\$4,456	6.30%				
4	City Manager	\$182,540	\$190,321	\$7,781	4.26%				
5	Municipal Court	\$195,645	\$222,409	\$26,764	13.68%				
6	Personnel	\$191,116	\$195,400	\$4,284	2.24%				
7	Planning and Community Development	\$305,607	\$315,640	\$10,033	3.28%				
8	Fire Department	\$2,456,789	\$2,537,901	\$81,112	3.30%				
9	Police Department	\$3,101,345	\$3,545,367	\$444,022	14.32%				
10	Public Works								
11	Community Services								
12	General City Fund								
13	Parks and Recreation								
14	Health								
15									
16	Totals								
17									
18									
19	Revenues								
20	General Property Taxes								
21	Sales Taxes								
22	Other local taxes								
23	Licenses and permits								
24	Fees and Service Charges								
25	Revenue From Other Sources	\$1,096,345	\$1,234,341	\$137,996	12.59%				
26	Building Rentals	\$188,000	\$192,000	\$4,000	2.13%				
27	Miscellaneous	\$158,782	\$162,234	\$3,452	2.17%				
28	Interest	\$1,231,209	\$1,323,456	\$92,247	7.49%				
29	Fines	\$234,645	\$250,786	\$16,141	6.88%				
30									
31	Totals	\$12,594,829	\$13,331,663	\$736,834	5.85%				
32									

Sorting by Percent change and Difference, for both Departments and Revenues can produce ideas for story angles. What might the reporter have pursued in this case?

TIP! Always check any calculations that you haven't done yourself

The mantra of journalism should be: "If your mother says she loves you, check it out!" We should be as skeptical of numbers as we are of human sources. So let's just check that the totals we were given add up correctly.

To check the Total for This Year, Departments, the formula is
=SUM(B2:B14).

Now **Copy** this cell, and **Paste** into columns C and D. What's our reporting plan now?

4. Performing anchored calculations

In each calculation we've performed so far, we've moved down or across the spreadsheet performing the equivalent calculation for that row or column. But sometimes that isn't what we want to do.

Let's calculate the increase in revenue from each source as a percentage of the total increase in revenue.

Use the formula

=(D20/\$D\$31), and then **Autofill**.

See that the \$ signs anchor the calculation. To anchor to a specific cell, you need a \$ in front of both coordinates.

Importing data, filtering, subtotals and pivot tables

Now we're going to look at a larger dataset, used in reporting this story (<http://www.newscientist.com/article/dn18806-revealed-pfizers-payments-to-censured-doctors.html>) to show how you can use a spreadsheet to quickly drill down to selections from the data, and to perform useful summary calculations.

Data are often provided as text (.txt) or comma separated values (.csv) files, so we'll first learn how to import a text file into Excel.

TIP! Databases should always be able to export data as .txt or .csv files, so bear this in mind whenever an organization says it can't provide data because it's in a special format.

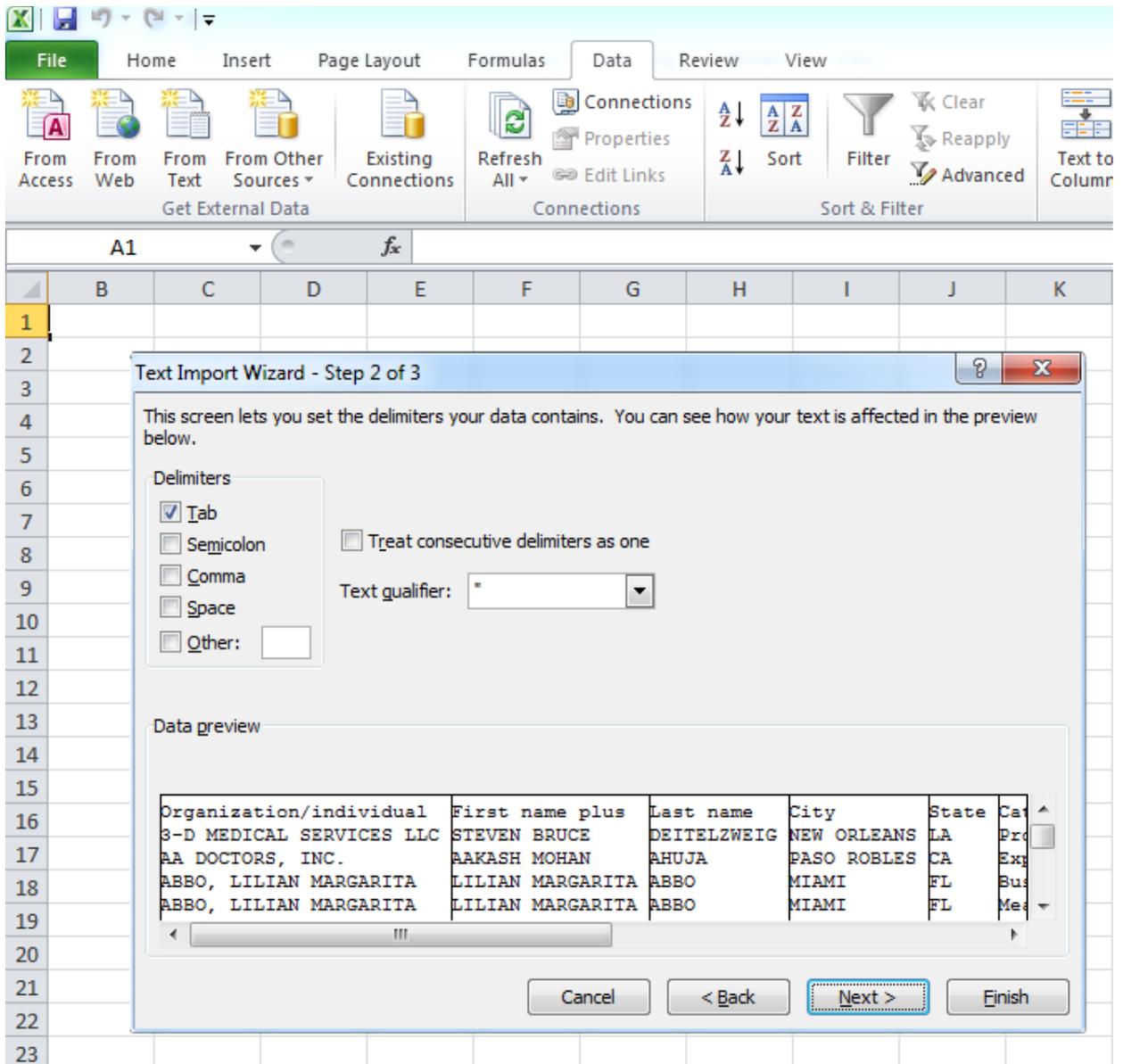
1. Importing a text file

Find the file **Pfizer payments.txt** and open it in Notepad or another text editor. See that individual entries are separated by Tabs and some of the entries are surrounded by double quote marks.

Now open a new document in Excel by selecting **Blank Workbook** under the **File** tab.

Then select **From Text** under the **Data** tab.

Browse for the file **Pfizer payments.txt** and double click on it. Excel should recognize that our file is delimited. When we click through to the next dialog box we should see that Excel has recognized that fields in the data are separated by a Tab, but we can select the correct delimiter if it hasn't. Excel should also have correctly selected the double quotes as a qualifier for entries that will be treated as text.



The next dialog box allows us to format each column. We want to select **Text** for all apart from the columns with money, for which we will select **General**.

Finally, check that the data is going to be entered into the top left hand corner of the spreadsheet and click OK.

Format the last three columns as currency, make the header row bold, and keep it present when you scroll through the spreadsheet by selecting **Freeze Panes>Freeze Top Row** under the **View** tab.

Save this file as **Pfizer payments**, and then again under a new name, by selecting **Save As** under the **File** tab, making sure to save as an **Excel Workbook**.

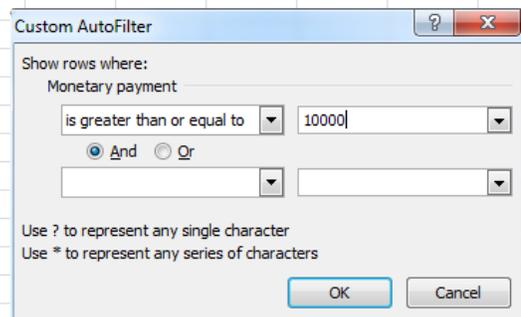
See that we have a list of more than 10,000 payments, which in this form doesn't tell us very much.

2. Filtering data

So let's set up the spreadsheet so we can filter for what we're interested in, by selecting **Filter** under the **Data** tab. See that each of the columns now has a button with a little triangle.

Let's say we want a list of all doctors in California who were paid more than \$10,000 to run expert-led forums. Click on the button for **State**, select **CA**, then **Category** and select **Expert-Led Forums**. Finally, Click on the **Total** button and select **Number Filters>Greater Than Or Equal To**, and select **Is greater than or equal to 10000**, as follows:

	G	H	I	J	K	L	M	N	O	P
	Monetary payment	Non-monetary payment	Total							
	\$1,000	\$0	\$1,000							
	\$850	\$0	\$850							
	\$1,000	\$0	\$1,000							
	\$3,000	\$0	\$3,000							
	\$1,000	\$0	\$1,000							
	\$6,750	\$0	\$6,750							
	\$2,000	\$0	\$2,000							
	\$4,500	\$0	\$4,500							
	\$3,500	\$0	\$3,500							
	\$2,200	\$0	\$2,200							
	\$750	\$0	\$750							
	\$20,500	\$0	\$20,500							
	\$4,500	\$0	\$4,500							
	\$1,000	\$0	\$1,000							
	\$1,500	\$0	\$1,500							
	\$400	\$0	\$400							
	\$6,250	\$0	\$6,250							
	\$3,750	\$0	\$3,750							
	\$4,000	\$0	\$4,000							
	\$21,000	\$0	\$21,000							
	\$1,000	\$0	\$1,000							



Click OK, and we have a list of 31 doctors.

TIP! If you need to do further calculations on a filtered subset of the data, select the data, Copy it, select Sheet 2 and paste into this new worksheet. To rename the sheet for future reference, right click on the Sheet tab and select Rename

Select **Filter** again under the **Data** tab to remove filters once more

3. Calculating subtotals

Let's now say we want to know how much money went to each state. First, **Select** all the data and **Sort** by state under the **Data** tab.

Then, with all the data selected, select **Subtotal** under the **Data** tab.

When the dialog box pops up, fill in like this:

At each change in: State

Use function: Sum

Add subtotal to: Total

E	F	G	H	I
State	Category	Monetary payment	Non-monetary payment	Total
AK	Expert-Led Forums	\$1,750	\$0	\$1,750
AL	Business Related Travel	\$0	\$314	\$314
AL	Expert-Led Forums			,750
AL	Professional Advising			,000
AL	Expert-Led Forums			,750
AL	Meals			\$58
AL	Business Related Travel			\$317
AL	Expert-Led Forums			,000
AL	Meals			\$56
AL	Business Related Travel			\$423
AL	Expert-Led Forums			,500
AL	Meals			\$222
AL	Business Related Travel			,002
AL	Expert-Led Forums	\$,250
AL	Meals			\$644
AL	Business Related Travel			\$330
AL	Expert-Led Forums			,200
AL	Meals			\$138
AL	Business Related Travel	\$0	\$123	\$123
AL	Educational Items	\$0	\$67	\$67
AL	Expert-Led Forums	\$3,500	\$0	\$3,500
AL	Meals	\$0	\$100	\$100

Subtotal ? X

At each change in:

State ▼

Use function:

Sum ▼

Add subtotal to:

- City
- State
- Category
- Monetary payment
- Non-monetary payment
- Total ▼

Replace current subtotals

Page break between groups

Summary below data

When the calculation is done, click on the **2** at top left to hide the data and just leave the subtotals (**1** hides everything except the grand total.)

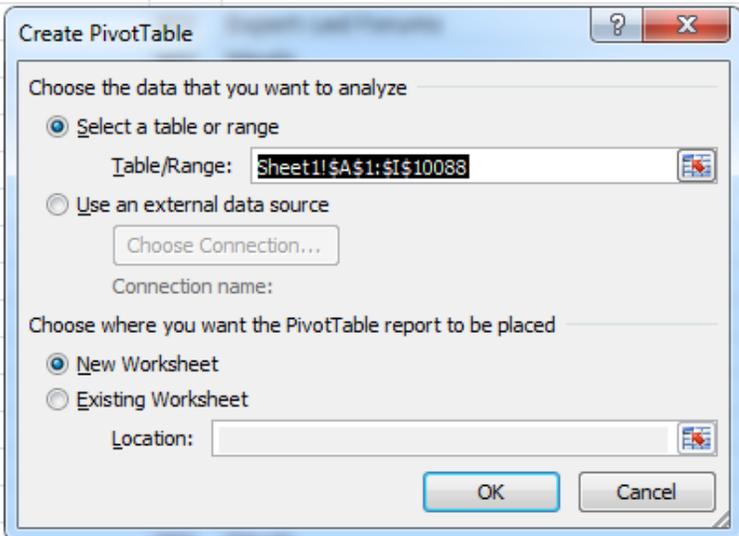
Select **Subtotal** and click **Remove All** to clear the subtotals.

4. Making pivot tables

What if we want to know the totals by state and by category of payment? We can do that in one step, using a pivot table.

Select **Pivot Table** under the **Insert** tab, and make sure the table is going to be created in a new worksheet.

City	State	Category
CHARLESTON	WV	Meals
CHARLESTON	WV	Expert-Led Forums
CHARLESTON	WV	Meals
SCOTT DEPOT		
SCOTT DEPOT		
MORGANTOWN		
HURRICANE		
BECKLEY		
BECKLEY		
BECKLEY		
MORGANTOWN		
HUNTINGTON		
HUNTINGTON	WV	Meals
HUNTINGTON	WV	Professional Advising
MORGANTOWN	WV	Business Related Travel



In the menu that opens up on the right, select State, Category and Total, and drag State from **Row Labels** to **Column Labels**

PivotTable Field List

Choose fields to add to report:

- Organization/individual
- First name plus
- Last name
- City
- State**
- Category**
- Monetary payment
- Non-monetary payment
- Total**

Drag fields between areas below:

 Report Filter

 Column Labels

Report Filter area (empty)

Column Labels area: State

 Row Labels

 Values

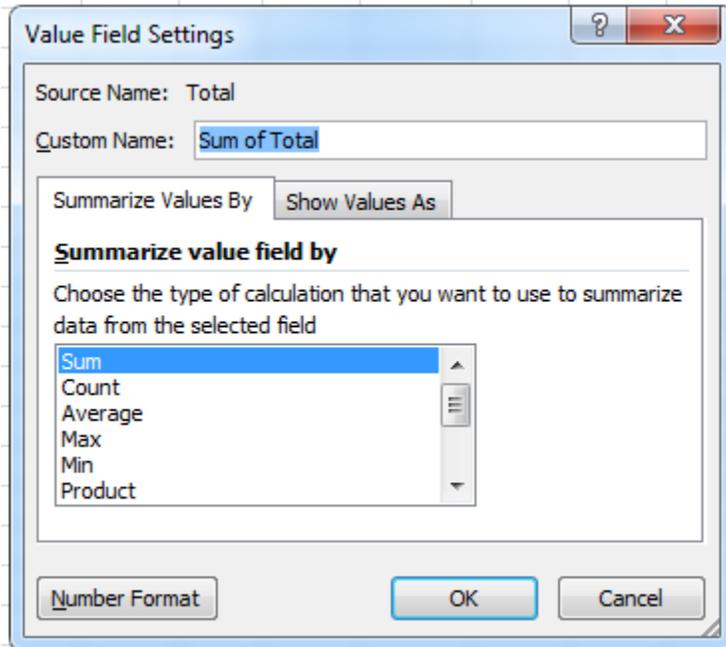
Row Labels area: Category

Values area: Sum of Total

Defer Layout Update

Update

Note the calculation for **Values** has defaulted to Sum, which is what we want, but we can select other summary statistics such as Count or Average by clicking the little triangle to the right and selecting Value Field Settings.



We can improve the formatting by selecting all the data and **Format Cells** to currency and **AutoFit Column Width**.

Finally, **Freeze Panes > Freeze First Column** under the **View** tab, so that the first column will remain visible as we scroll right.