BASEBALL SALARIES: EXCEL FILTERS AND PIVOT TABLES

Sarah Cohen, The Washington Post

Some reporters choose to use Excel as a little database. The reason is that it's simple and flexible for summarizing and filtering small data sets.

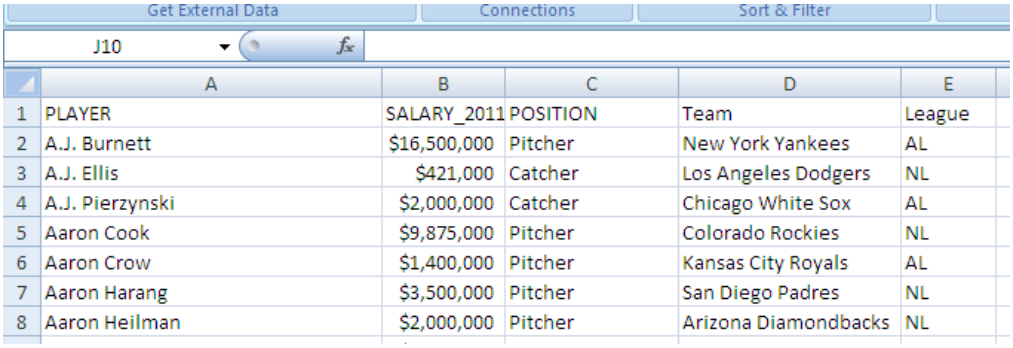
The two key tools are filters and pivot tables. Filters let you see just the items you want, without changing the underlying data. Pivot tables are powerful summary tools, much like statistical software for large databases. You'll use them to create crosstabs.

SETTING UP YOUR SPREADSHEET

There are two steps to get ready for a filter or pivot table. One you have to do. The other is optional, but will make your life easier.

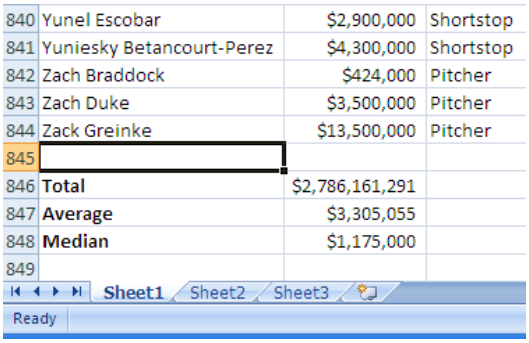
Make sure it looks like a database

To make a filter or pivot table work, you need your data in columns all next to one another, and rows without blanks. You also need field or column names directly above the data in a single cell:



	A	B	C	D	E
1	PLAYER	SALARY_2011	POSITION	Team	League
2	A.J. Burnett	$16,500,000	Pitcher	New York Yankees	AL
3	A.J. Ellis	$421,000	Catcher	Los Angeles Dodgers	NL
4	A.J. Pierzynski	$2,000,000	Catcher	Chicago White Sox	AL
5	Aaron Cook	$9,875,000	Pitcher	Colorado Rockies	NL
6	Aaron Crow	$1,400,000	Pitcher	Kansas City Royals	AL
7	Aaron Harang	$3,500,000	Pitcher	San Diego Padres	NL
8	Aaron Heilman	$2,000,000	Pitcher	Arizona Diamondbacks	NL

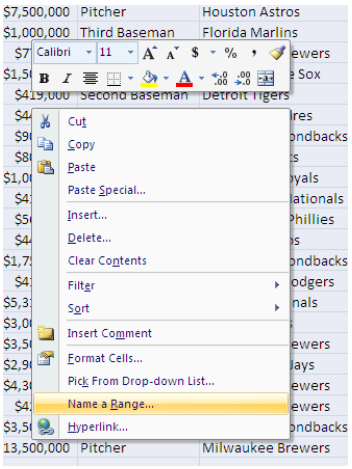
Be sure any totals, averages, notes or other words you put at the bottom are separated by a blank row:



840	Yunel Escobar	$2,900,000	Shortstop
841	Yuniesky Betancourt-Perez	$4,300,000	Shortstop
842	Zach Braddock	$424,000	Pitcher
843	Zach Duke	$3,500,000	Pitcher
844	Zack Greinke	$13,500,000	Pitcher
845			
846	**Total**	$2,786,161,291	
847	**Average**	$3,305,055	
848	**Median**	$1,175,000	
849			

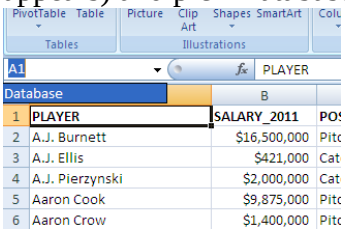
Name the database

You don't have to do this, but it makes later work easier.
Select the data, including the titles but excluding the totals at the bottom.



Right click on the selected data and select "Name a range" and in the new window type in a name. I usually call it "Database." Now whenever you want to refer to that area, you can refer to the name Database instead of its address.

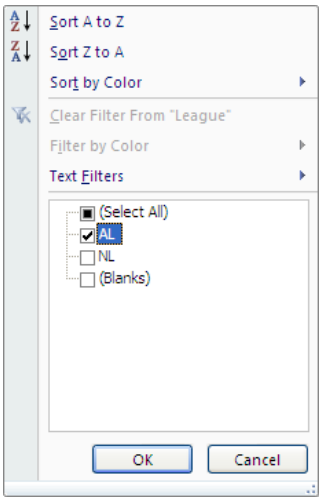
Try it. Select any cell on your worksheet. Now drop down the menu on the upper left corner of your worksheet (where the cell address usually appears) and pick Database:



You'll be taken to your data.

FILTERING

Select your database using its name, and click on the Data tab/menu. Click on Filter – this looks like a funnel – to get drop-down menus next to each of the column headings:



Word Filtering

Use the drop-down menu to show only rows with certain words, like AL here:

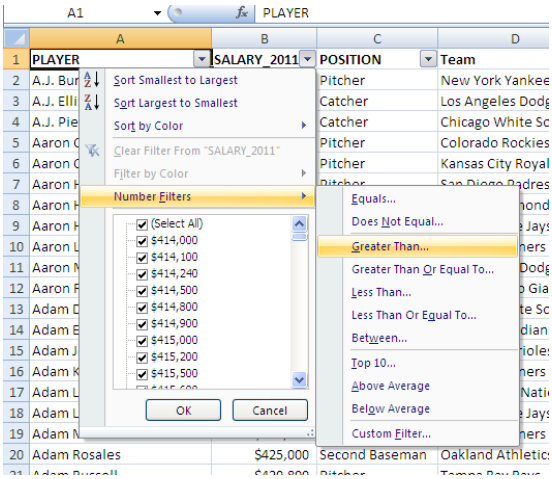
Your row numbers will turn blue to give you a visual clue that you're filtering. The drop-down arrow turns blue next to League to show that it's the field that's filtered.

To return to all rows, drop down the filter again and choose (Select All).

Number Filtering

To choose numbers, you'll usually want to choose values greater than or less than a certain value.

Drop down your menu in the appropriate column and choose Number Filter. Then Greater than or less than and type in a value.



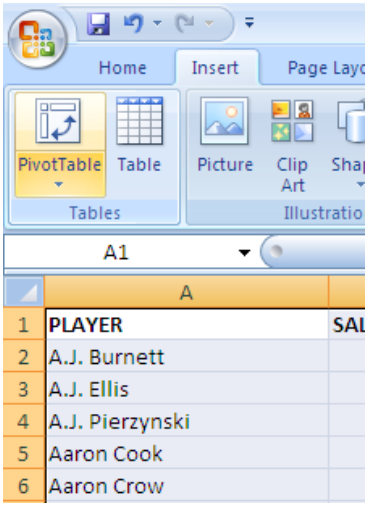
REMOVING A FILTER

Make sure to get rid of any filters before you sort or make any new calculations. To remove the filter, click on the Data tab/menu and click on Filter, or the funnel icon to remove the check-mark.

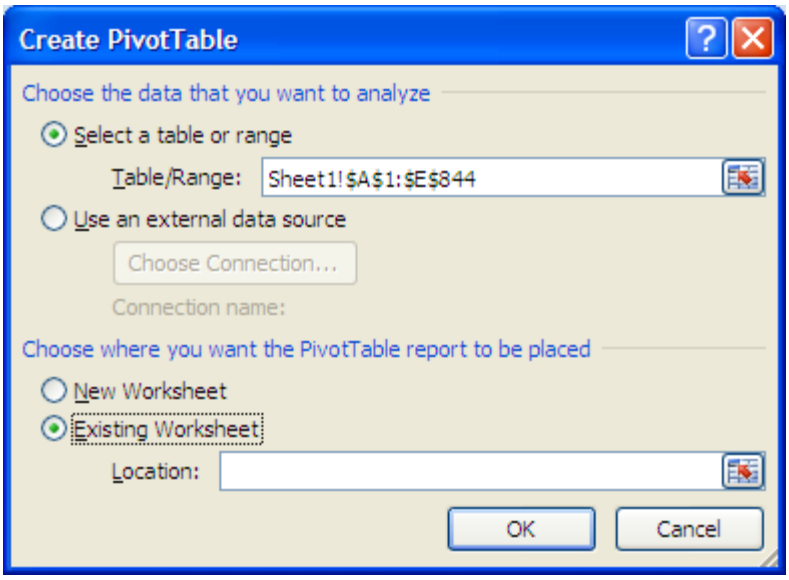
SETTING UP A PIVOT TABLE

You have powerful summary tools with a Pivot Table. It's much like a Totals, Group By or Summary query in a database manager, but it allows even more flexibility.

To start a Pivot Table, select your data, including the headings. Then choose the Insert tab/menu and select PivotTable.

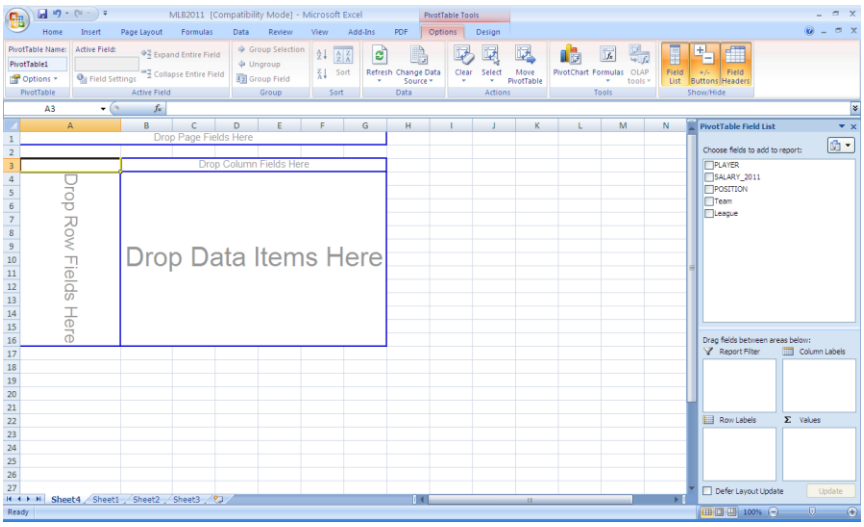


You'll get a Pivot Table Wizard. Most of the time, you'll use a simple Excel list or database and create just a Pivot Table, as shown below, so you can just fly right by the first question by pressing the OK button.

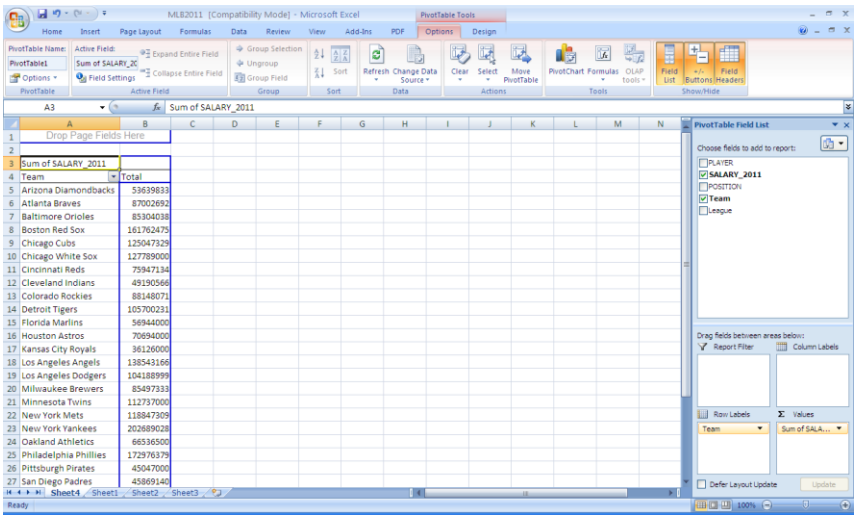


In the newest versions of Excel, (2007 and later), you will be taken to the layout screen. See earlier versions of this handout for handling PivotTables in Excel 2003 or earlier.

This layout is the meat of your pivot table. You'll drag the buttons for each element you want to see from the right-hand side to the boxes.



(When you drag a column that contains only numbers into the center, Data, area, Excel assumes you want to add up the values. Otherwise it assumes you want to count up the lines, or salaries in this case.)

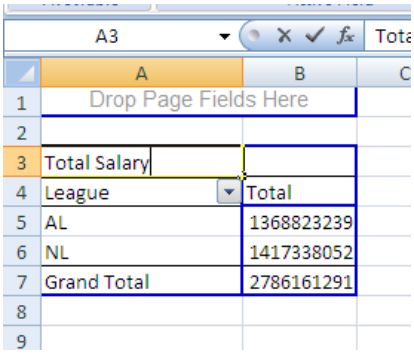


2		
3	Sum of Salary	
4	League	Total
5	AL	942216102
6	NL	1020625712
7	Grand Total	1962841814
8		

EDITING A PIVOT TABLE

To change column or row headings

Type right over what Excel gave you:

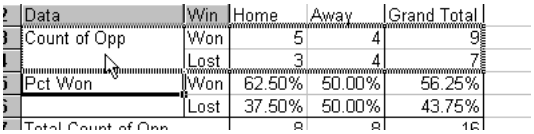


To add or delete elements

Click anywhere on the PivotTable to launch the layout along the right. You can now edit your PivotTable. (Note: For earlier versions of Excel, simply click on the Pivot Table Wizard button that appears when you click on the Pivot Table. Then select layout.)

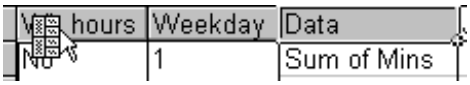
MOVING PIECES AROUND

Move an entire row or column by moving its title. The rest of the Pivot Table row or column moves with you.



Data	Win	Home	Away	Grand Total
Count of Opp	Won	5	4	9
	Lost	3	4	7
Pct Won	Won	62.50%	50.00%	56.25%
	Lost	37.50%	50.00%	43.75%
Total Count of Opp		8	8	16

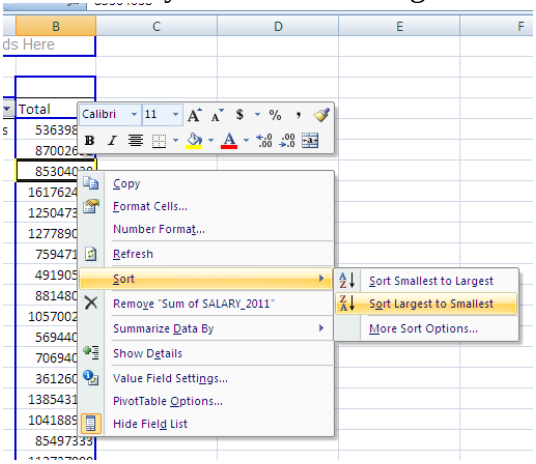
Drag the gray buttons to any position you want.



You can see which way the page is turned when you drag.

SORTING PIVOT TABLES

This can be kind of dicey in earlier versions of Excel, but Microsoft has made it easier from 2007 onward. All you need to do is right click on one item in the column you want to sort, then select sort.



A WORD ON PIVOT TABLE SETUPS

Pivot tables are easy to read when they follow Philip Meyer's rule of thumb: Put the independent field as columns, the dependent field as rows, and calculate a column percent.

Translation: Put whatever comes first in time in the COLUMNS area, put whatever comes last in time in ROWS and change "Normal" to "% of Column" in the Show Data As area.

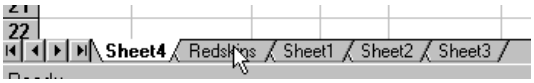
That way you can read it normally, from left to right.

You can use this for all kinds of stories: Do women with expensive insurance get c-sections more frequently than those covered under Medicaid or charity cases? Do upper-income minority homeowners get refused loans more frequently than similarly situated whites? How frequently do public officials use their cell phones during work hours?

TIPS, TRAPS AND FAQS

How do I get back to my original data?

At the bottom of your screen are little tabs, like index markers. Click on the appropriate one. If you haven't named it, choose Sheet 1.



How do I get Pivot Tables to calculate medians?

You don't. It's a key element missing from both Pivot Tables and database management programs. There are ways to calculate medians for groups, but they're quite difficult.

How do I get a Pivot Table to update data I've changed in the original database?

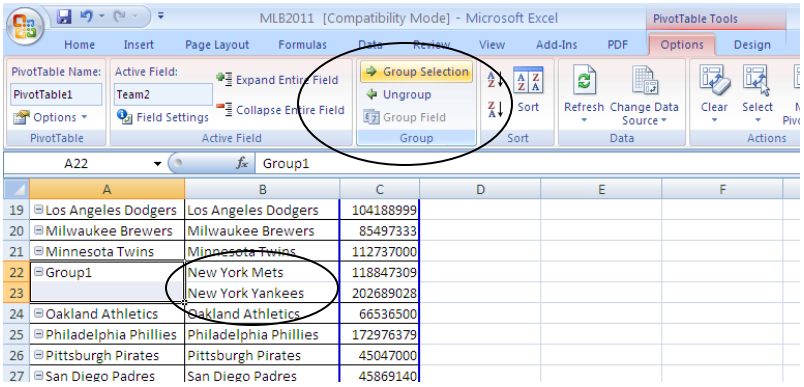
Click the Refresh icon under the Options tab while you have any part of the pivot table selected.

I get an error when I try to create the pivot table.

You probably have a missing heading. Check to make sure that the row above your data is filled in with titles.

Creating bigger categories:

Rather than using formulas, it may be simplest to create an "Other" category or other groupings through a pivot table.
Put the values you want together next to one another. Then select them, and choose the Group button under the Options tab.



Hiding details:

You can hide details by simply moving the field into the page element, usually in A1. Look for the little symbol that looks like a few pages together when you move it up.

To ignore the certain values, double-click on the field name that holds the information you want to ignore. Click on each one you want to be excluded from your analysis