Quick analysis in Excel

Jaimi Dowdell, IRE

Reporter David Bergman analyzed data from the World Bank's fraud and corruption department which lists people and companies ineligible to receive contracts from the World Bank. His findings in the Bangladesh publication New Age include:

> "The World Bank in the last three years has barred from bidding for bank-financed contracts more organisations and individuals from Bangladesh than from any other country.
>
> An analysis of a World Bank database containing details of all current suspensions shows that since the beginning of 2009, 22 individuals and 16 organisations based in Bangladesh have been de-barred for either a three or four year period.
>
> Eleven of the suspensions took place in 2011.
>
> All the Bangladeshi organisations and individuals were debarred because of their involvement in 'fraudulent practices' comprising either a 'misrepresentation' of facts or 'collusion.'"

The data is current on the World Bank's website.   It contains the firm or individual's name, address, country, ineligibility period and grounds for debarment (See Figure 1). Refer to the World Bank's procurement guidelines to find explanations of the different grounds for debarment.

Figure 1

Firm Name	Address	Country	Ineligibility Period From	To	Grounds
REEF BUILDING SYSTEMS LIMITED (REEF)*158	MILIMANI FLATS, MILIMANI ROAD, OFF STATE HOUSE, P.O. BOX 10439-00400 NBI, NAIROBI	Kenya	09-JUL-2013	08-JUL-2016	Procurement Guidelines, 1.14(a)(ii)
CECEP NEW HOPE & TARLROAD (BEIJING) ENVIRONMENTAL TECHNOLOGY CO., LTD.*157	FLOOR 6, BLOCK A, JIENENG BUILDING, 42 XIZHIMEN NORTH ST., HAIDIAN DISTRICT, 100082, BEIJING	China	28-JUN-2013	28-JUN-2015	2006 Procurement Guidelines, para. 1.14(a)(ii)
YAVUZER İNŞAAT MAKINE VE MALZEMELERİ SANAYI VE TICARET LTD. ŞTI*156	ISTIKLAL MAHALLESI SANTRAL SOKAK NO:1/ 23, ESKIŞCHIR	Turkey	18-JUN-2013	17-JUN-2015	Procurement Guidelines, 1.14(a)(ii)
SCIENTIFIC ENERGY AND ENV. MGT SYSTEMS (SEEMS)*155	23B, IXORA DRIVE, M.K.O. GARDENS, ALAUSA, IKEJA, LAGOS	Nigeria	17-JUN-2013	17-JUN-2015	2006 Consultant Guidelines, para. 1.22(a)(ii)
M/S WIJEKAMAL ENGINEERING PVT. LTD.*154	E-52, UDUKUMBURA, NELUNDENIYA	Sri Lanka	11-JUN-2013	10-JUN-2017	Procurement Guidelines, 1.14(a)(ii)
GHD PTY LTD., in its own name and in its capacity doing business as GHD INDONESIAN OPERATING CENTRE *152	L15 133 CASTLEREAGH ST., SYDNEY	Australia	10-JUN-2013	09-JUN-2014	2004 Consultant Guidelines, Paragraph 1.22(a)(ii)
PT GLOBAL HUTAMA DESAIN (PT GHD) *153	4F WIRAUSAHA BUILDING, SUITE 401, JL. HR RASUNA SAID KAV C-5, JAKARTA 12940	Indonesia	10-JUN-2013	09-JUN-2014	2004 Consultant Guidelines, Paragraph 1.22(a)(ii)

To analyze the data, Bergman did what many reporters – including me – have done before: He counted. He went through each and every line and counted up the number of times each country was represented. After he'd been through the hundreds of records he compared the totals for each company. It worked for his story, but it took more time than necessary.

We can do this same analysis using pivot tables in Excel. By tossing aside your pencil and scratch paper and having Excel do all the work you'll increase your speed and accuracy.

Let's practice with an updated version of the World Bank data. You'll learn how to get this data from the web and into your spreadsheet in another lesson, but for now open the file WorldBank.xlsx. It should look like Figure 2 below.

Figure 2

Remember whenever you open up data you should take a moment and get familiar with it. Look at the different columns and take note of how many records the spreadsheet contains. I have 608 rows of information with one row for the headers which means I have 607 records of companies or individuals debarred from doing business with the World Bank. Keep track of that record count as you conduct your analysis.

SORT TO GET TO KNOW YOUR DATA

Let's look at this data just a little bit. What's the most recent record of disbarment? To find out, let's sort the data.

Remember when you sort that you need select your data. Click anywhere within your data and hold down shift+ctrl+8. (This will be different if you're using a Mac. Different Macs work differently – try using command+A and if that doesn't work use shift and your arrow keys.)

Once you've selected the data, click on the "Data" tab and select "Sort" as shown in Figure 3.

Figure 3
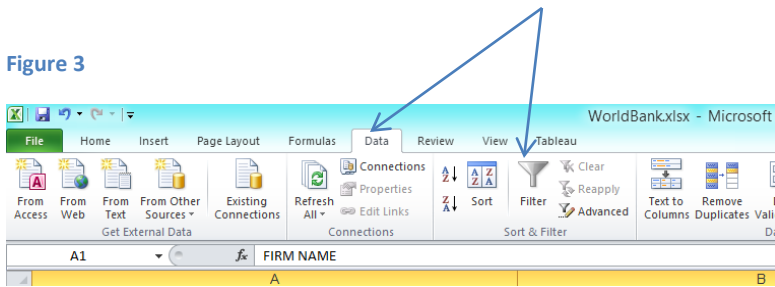
When the sort window appears, select "Ineligibility period from," "Values" and "newest to oldest" in the window (Figure 4). Then click OK

Figure 4

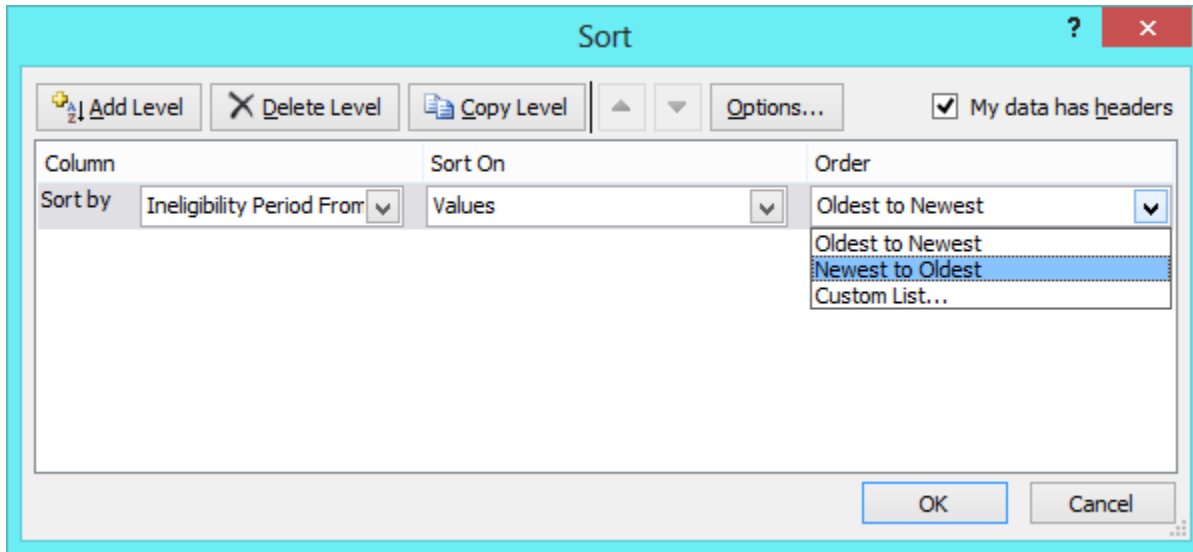You should find that the most recent debarment date according to this data is June 28, 2013. If you scroll to the bottom of your data or resort to show the oldest to the newest date, you should find March 16, 1999.

Look at the dates found in the column "Ineligibility period to." What do you notice?

That's right, not all of the entries are dates. Let's practice using filters to see how many records don't have dates.

FILTER TO QUICKLY FIND INFORMATION

Remember to turn on the filter you'll simply click somewhere in your data, click the "Data" tab and select the "Filter" icon as shown in Figure 5.

Figure 5
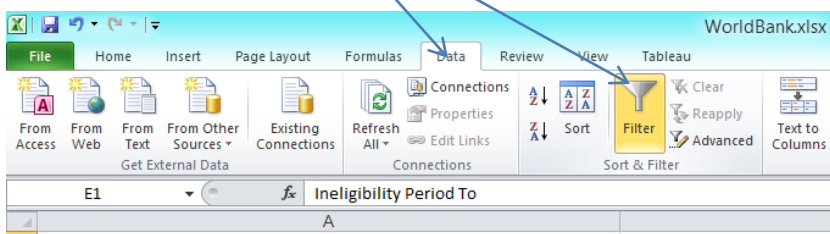
You should notice that dropdown arrows have appeared next to all of your column headers. Click on the dropdown next to "ineligibility period to." You should see a list of dates, or years (the individual dates are collapsed by year) and then "permanent." (See figure 6)

Figure 6
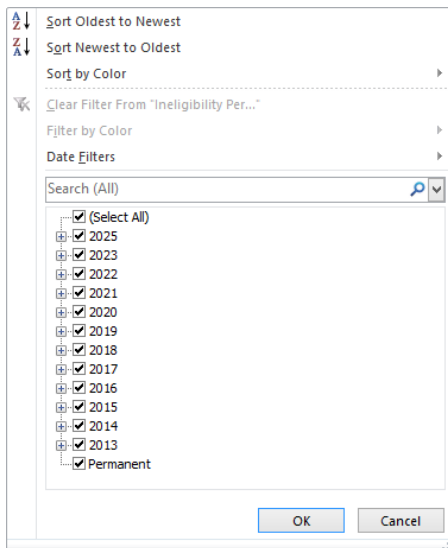
Click on the checkmark next to "(Select All)" to uncheck everything. Then click on the box next to "Permanent" (Figure 7).

Figure 7
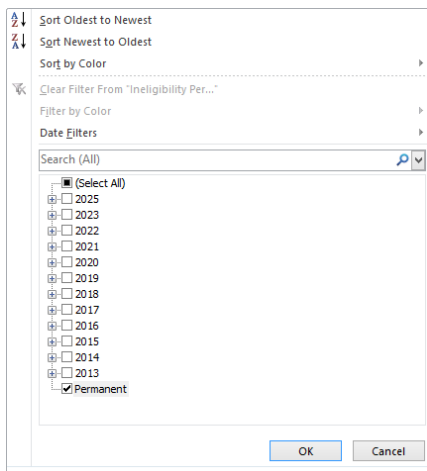
Click OK and you should now only see the records that list "Permanent" for the "Ineligibility period to" date. Look to the lower lefthand corner of your screen and Excel should show you the number of records returned from your filter, "94 of 607 records" as shown in Figure 8.

Figure 8

Notice that the row numbers along the left are now blue. This means that Excel is in filter mode. The rest of the records are just hiding and you can e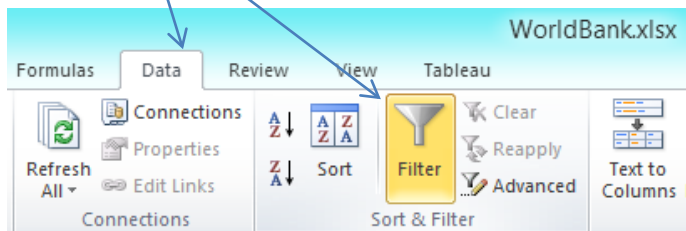asily get them back by changing that column's filter to "Select All" again or turning the filters off. To turn the filters off click on the "Filter" icon once again found under the "Data" tab (Figure 9).

Figure 9

The dropdown arrows next to each column should disappear and you should now see all records.

SUMMARIZE DATA WITH PIVOT TABLES

Now that we've gotten a bit familiar with our data, let's answer the question that Bergman used for the basis of his 2012 story "Bangladesh tops WB procurement black list."

Which country has the most firms and individuals on the debarred list?

You could count up each country individually and then compare your numbers. You could filter for every country listed and keep a tally. The easiest and quickest way is to get Excel to do all of this for us with a pivot table.

Pivot tables allow us to take a large chunk of individual records and summarize them by groups. In this situation, we want to count the number of records found for each country. The resulting list should just have a country name and a number beside it. You can answer similar questions using pivot tables and different data

Here are some examples of questions pivot tables can help you answer:
-Which state received the most grants?
-Which team pays the most in total salaries?
-Which county has the greatest number of hunting accidents?
-Which candidate received the most money?

To start a pivot table, first select your data. If you're using a PC you'll click somewhere in the data and hold down the shift+ctrl+8 keys. Notice that your data is now highlighted out to the first blank column on the right and the first blank row on the bottom.

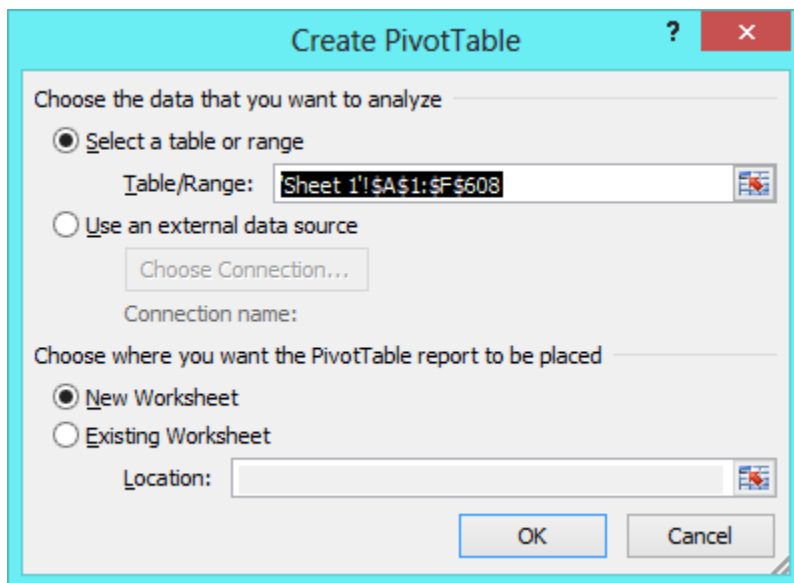Now, to start the pivot table click under the "Insert" tab and look to your far left for the "Pivot table" icon as shown in Figure 10 below. Depending on the version of Excel you have, this may be different. If you're using Excel for the Mac or PC versions from 2003 or earlier you'll find pivot tables under the "Data" menu item.
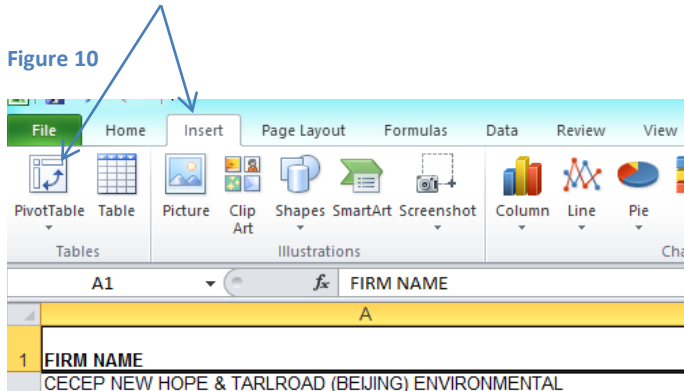
Figure 10

A window should appear (Figure 11).

Figure 11

In this window Excel wants to know where the data is that will be used in the pivot table and where we'd like to put the new pivot table. Since we highlighted the data the "Table/Range" information should be correct so you can leave it as it is. Make sure you have "New Worksheet" select as where the pivot table will be placed. Then click OK and a new sheet should appear with our pivot table tools. This is where we'll create our data summary. (See Figure 12)

Some different versions of Excel work most like the one I'm showing you here (Excel 2010 for the PC). If you're using a different version, Excel may launch a Wizard with three steps. Step 1 is similar to the window described in Figure 11 – just click "OK." Step 2 points to the data. If you've highlighted your data click "OK." Step 3 shows a finish line flag, but we're not quite finished. In this window, click "Layout." That will take you to a window that, although it looks a bit different, will function similarly to the tools shown in Figure 12. Once you've worked out the layout you can click "OK" and "Finish."

Figure 12
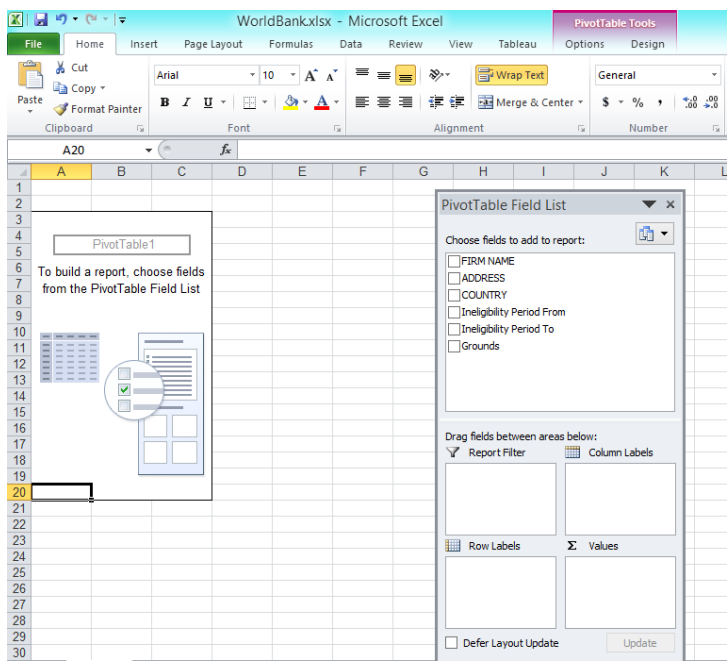
You'll build your pivot table using the tools in the box on the right. Notice you see a list of all of the columns names, or fields, that are found in your original data. Our question is which country has the most debarred firms/individuals. What we want to count is the different countries. You'll take that field from the list and drop it under the "Row Labels" portion of the PivotTable Field List.

Anything you put in the "Row Labels" portion will appear along the left-hand side of your PivotTable. Each country should only be listed one time. There should be no duplicates for items you place in a PivotTable and the items will be listed alphabetically. See Figure 13.

Figure 13
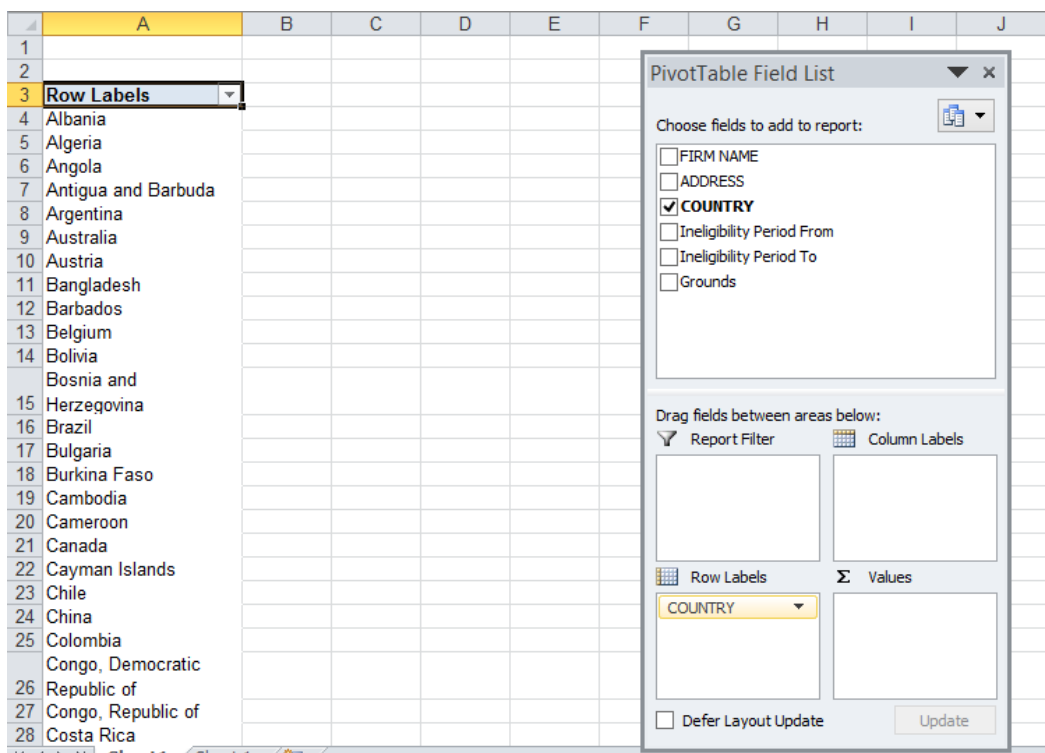
Now we have a list of countries, but we want to know how many times each country was listed in the database. To do that, we need to count the number of records for each country. Any time you want to do math in a PivotTable, you'll work with the "Values" box of the PivotTable Field List window. Since we're counting the number of times each country is listed, we'll take the country column header once again and drop it under the "Values" box. See Figure 14
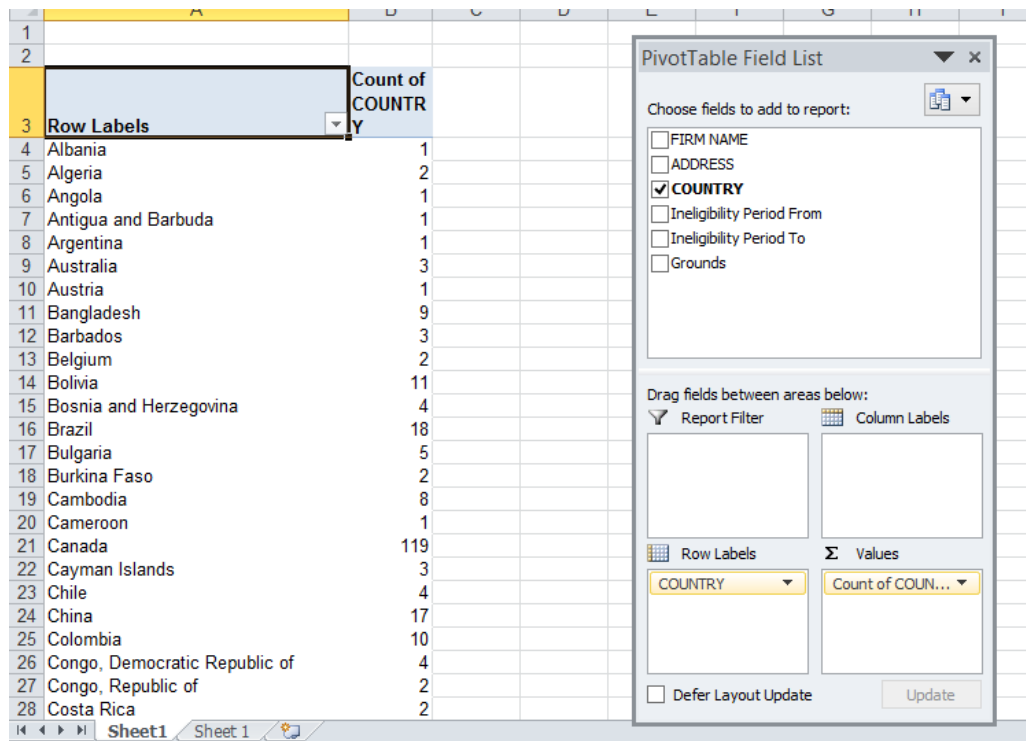
A number should appear by each country. That number represents the number of records for each country – in this case, the number of individuals or firms debarred from each nation. Scroll to the bottom and notice that Excel has calculated a grand total in your PivotTable. This total should match the number of records in your original data (607 in this case). If it doesn't, something is wrong.

Since we're journalists, we typically like to see things ordered from greatest to least or least to greatest. We want to know who is at the top and who is at the bottom and to do that we need to sort. PivotTables are sorted differently than the way you've been sorting in spreadsheets.
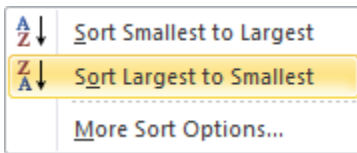
For the newer versions of Excel right click on any of the numbers listed next to the country names. When you do that, a window should appear (Figure 15).
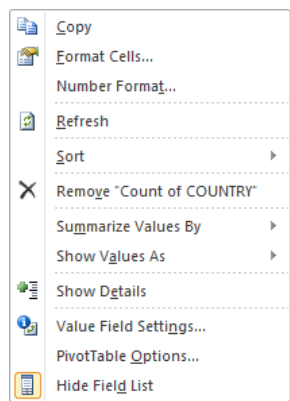
Click on sort and then select "Sort Largest to Smallest" (Figure 16)

Figure 16

From this result you can see that Canada is now at the top of the list followed by the United States, Indonesia and the United Kingdom.

FILTERING IN PIVOT TABLES

Remember when we looked at the "Ineligibility period to" column in the original data? We saw 94 records of firms or individuals who are permanently debarred. Which country has the most of those permanently debarred?

To find out this answer we can work with the PivotTable we already have going. (You can also start a new one if you like.)

Notice that there are four windows below the Field List. So far we've worked with "Row Labels" and "Values." This time we're going to add information to "Report Filter." We know that we only want to see the permanently debarred records and we find that in the "Ineligibility period to" column. Drag that column to the "Report Filter" box as in Figure 17.

Figure 17
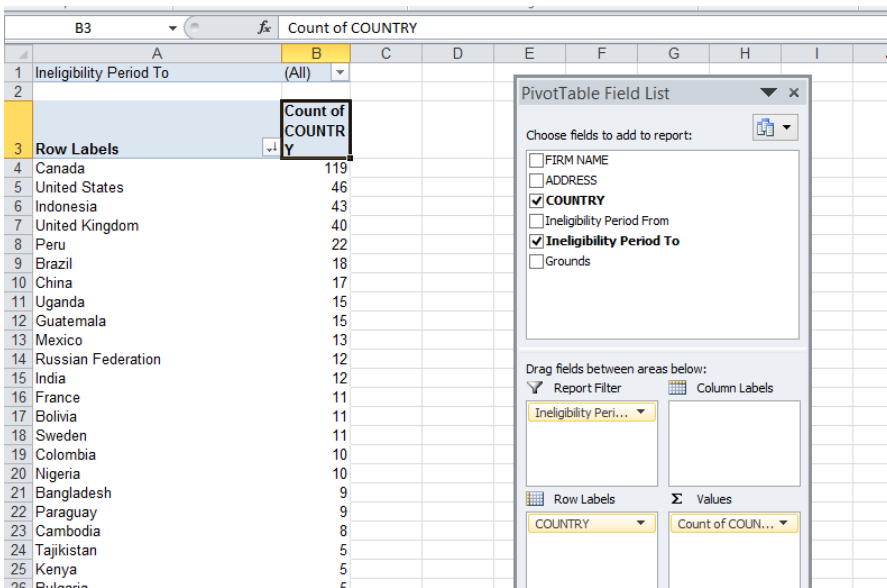
Notice that the field name can now be found in cell A1 just above the PivotTable. This is a filter of that column and works much like filters normally work in spreadsheets. By default, all values for "Ineligibility Period To" are shown. You know this as in cell "B1" it says "(ALL)." See Figure 18.

Figure 18
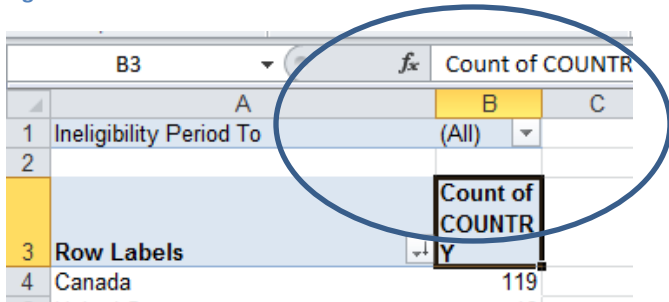
Click on the dropdown next to "(ALL)" and select "permanent." We won't be using it in this exercise, but notice that there is an option to select more than one value at the same time. Click OK.

Figure 19
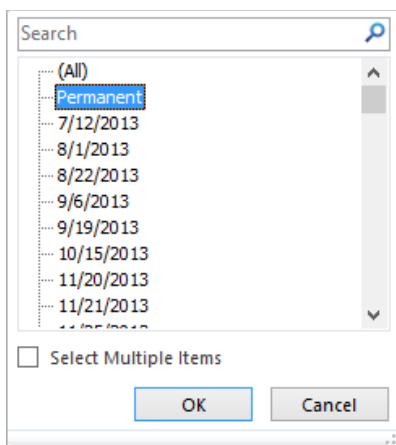
Now you are filtering your PivotTable. Look at the grand total – it should say "94." That matches the total number of records for the same filter in the main spreadsheet. You can now see which country has the most permanently debarred individuals/firms.

COMMON HICCUPS

I've lost my box that has the column names listed, how can I get it back?
-To get your PivotTable Field List window back, all you need to do is click somewhere in your PivotTable.

I added a new column and it doesn't show up in my field list.
-Under the "Options" tab click on the "Refresh" icon.
-If that doesn't work, delete the sheet with your PivotTable, save the workbook and start a new PivotTable.

I can't get the PivotTable to launch. It keeps giving me some weird error about the field name not being valid.
-Each of your columns must have labels. If one column header is blank the PivotTable will refuse to launch.

I know I'm supposed to have unique values in my PivotTable, but I see repeats. What's going on?
-Although two words may look exactly the same to our eyes leading or trailing spaces can through off Excel. If you see this happening, add a new column to your data and populate it using the Trim function. For example, let's say my original data that is showing duplicates in the PivotTable is found in column B. I'll add a new column, label it and then write this formula for the second row: =trim(B2)

The result will look the same to you, but all of the extra spaces will be gone.

Remember that PivotTables are space hogs.
-Once you've got what you like in your PivotTable, I recommend you highlight the table, copy it, go to a new sheet, right-click and select "Paste special" and then "values." This will just give you the data and not the PivotTable tools, etc. You can then delete the sheet from where you copied the PivotTable.