# Using pivot tables in Excel (live exercise with data)

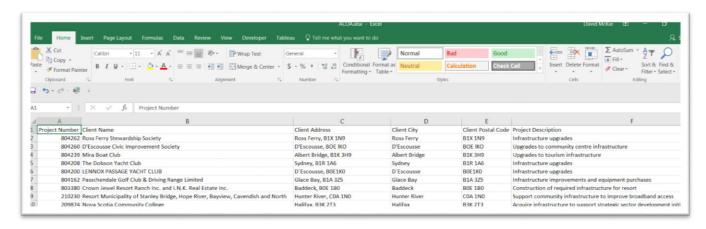
In chapter four, we used B.C.'s political donations data to learn how to build pivot tables, which group elements in your data and summarize the information using totals and subtotals. Because pivot tables feature so prominently in the work of journalists, we'll use a live dataset to explore some of the ways we can group, sort, count and sum data in order to come up with great stories.

For this tutorial, we'll take a deep dive into the database that the <u>Atlantic Canada Opportunities Agency (ACOA)</u> uses to record the money it distributes to institutions that apply.

#### What we will learn:

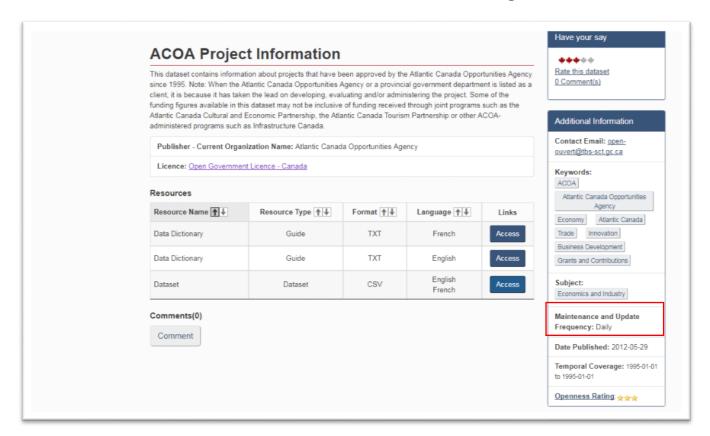
- 1) How to download tables from an open-data site;
- 2) How to prepare for a pivot table;
- 3) How to create a pivot table
- 4) How to SUM, COUNT and group by year
- 5) How to filter by project length.

## Task one: How to download tables from an open-data site



Download the "Dataset" from the federal government's open <u>data</u> <u>website</u>, as well as the <u>data dictionary</u> that explains the dataset's contents.

Before we download, let's understand what we've looking at.



The information under the title "ACOA Project Information" provides the bare-bone details of the table that you're about to download. Before working with any dataset, it's important to know as much about the institution, organization or agency that maintains it, which is usually that body's home page.

Next, check out the menus to the right, which tell us more about the data. Among other things, the "Additional Information" section indicates how frequently the dataset is updated; in this case, daily. This means that your table should contain MORE records when compared to the data featured in THIS tutorial.

Under the "Resources" section, you'll find the "<u>Data Dictionary</u>" (also called a "Readme" file in other open-data sites) that explains what's in the table. Be sure to click the "Access" tab to save this file, which is in text format <TXT>.

To review the checklist of questions you should asking yourself before downloading any dataset from open-data websites, please review pages 28 to 30 of The Data Journalist.

The ACOA dataset, which is updated daily, is rich with possibilities, allowing journalists to quickly discover who is receiving the most lucrative federal grants, how long it has taking them to repay the money, and whether the agency's total spending is decreasing or increasing year over year.

This dataset is also a typical example of one that is ideal for a pivot table; that is, every row contains the same information, which then allows the pivot table to group on each value.

Still, there are a few things we should routinely do when preparing the data to be analyzed in a pivot table.

Let's begin with the dates. We should determine which date field is the most important, and pull or parse it into the component parts: year, month and day.

Task two: How to prepare for a pivot table

Once you've downloaded the table, which is in csv format, open it, copy the website's URL and paste it into the first available cell in the first row, which in this case is R1.

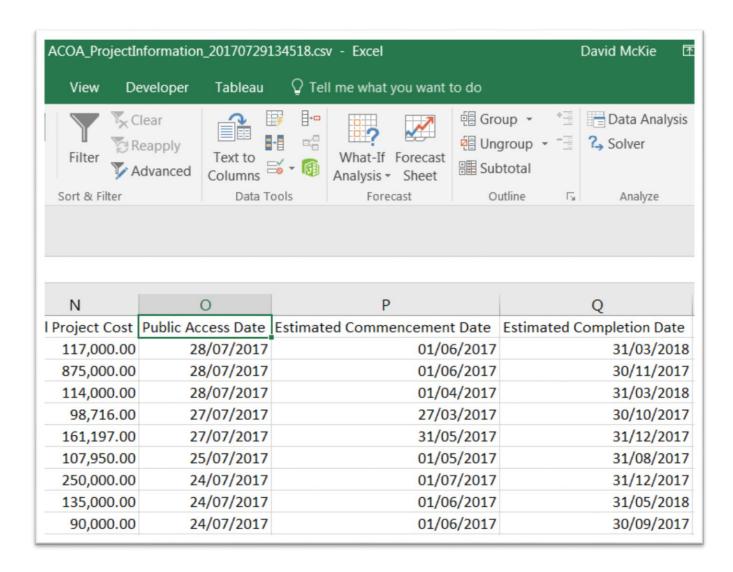
Save the csv file as an Excel table.

The csv file is now your backup.

Work with the Excel file, which unlike csv files, accommodates multiple worksheets.

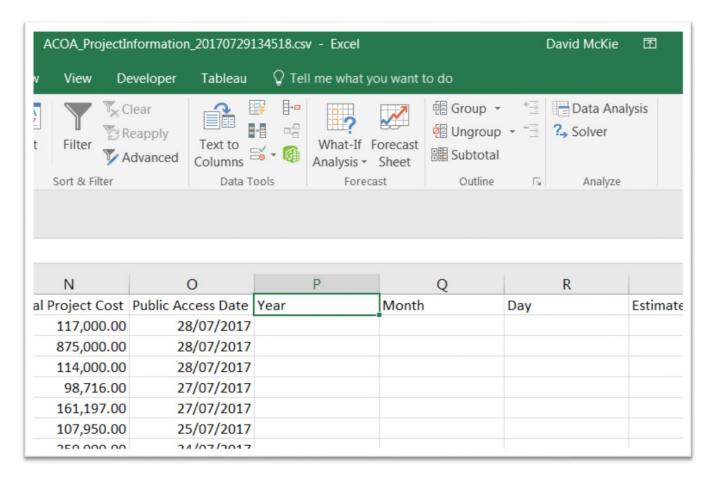
The last three columns are the dates.

Sort column O, "Public Access Date", in descending order to determine the most recent entry.

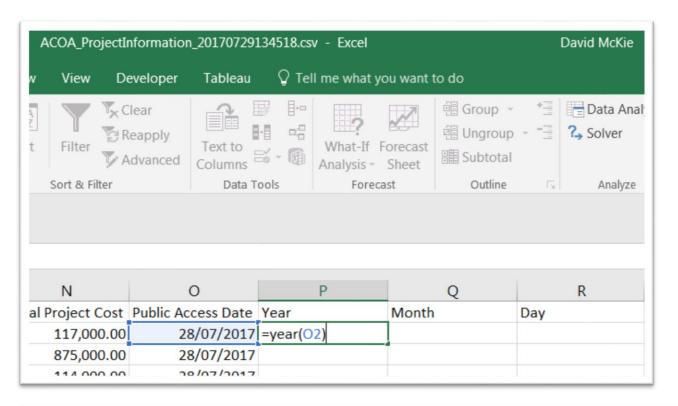


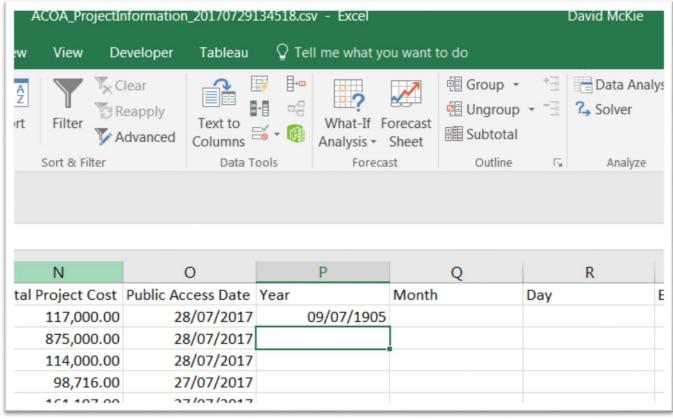
According to the data dictionary that we downloaded at the beginning of this task, the Public Access Date is "usually 60 days following client acceptance of written offer." The Estimated Commencement Date, is the "date the project is expected to begin." And the Estimated Completion Date is "the date the project is expected to be completed."

Chose the "Public Access Date". Create three columns to the right: year, month, day.

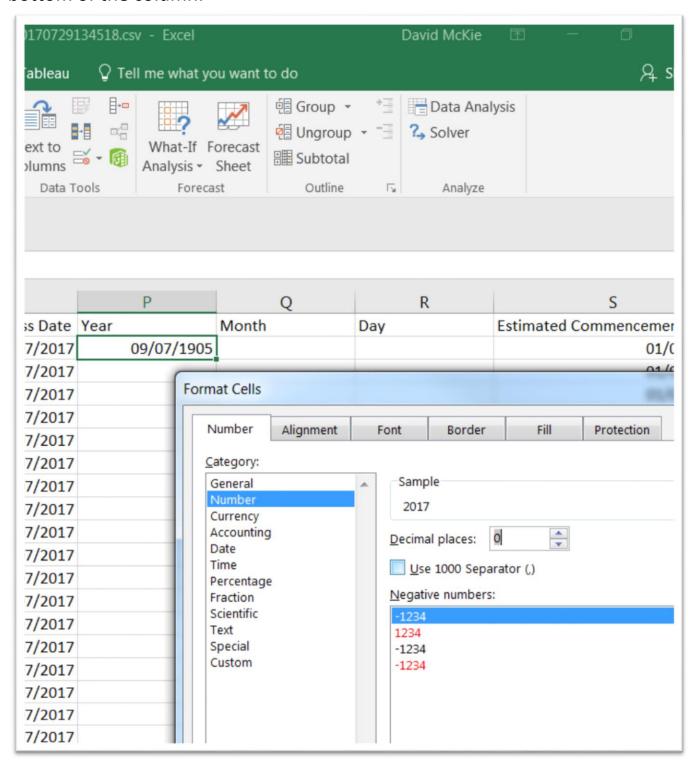


We'll use the Year function <=YEAR(cell reference)> to extract the year -- a task explained on page 14 of The Data Journalist's on-line tutorial called "Working with Specialized Functions and Formulas in Excel" -- to pull the year out of the date.

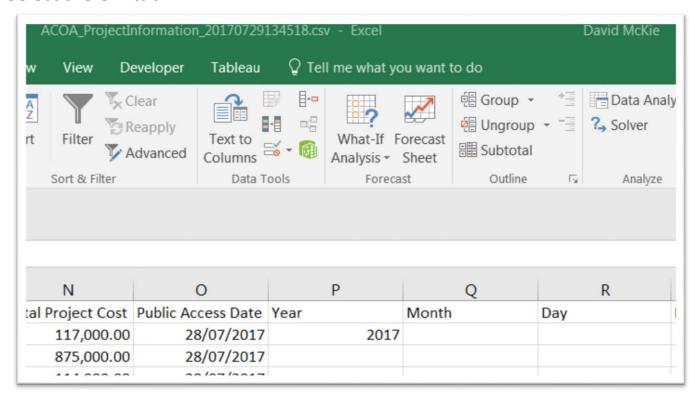




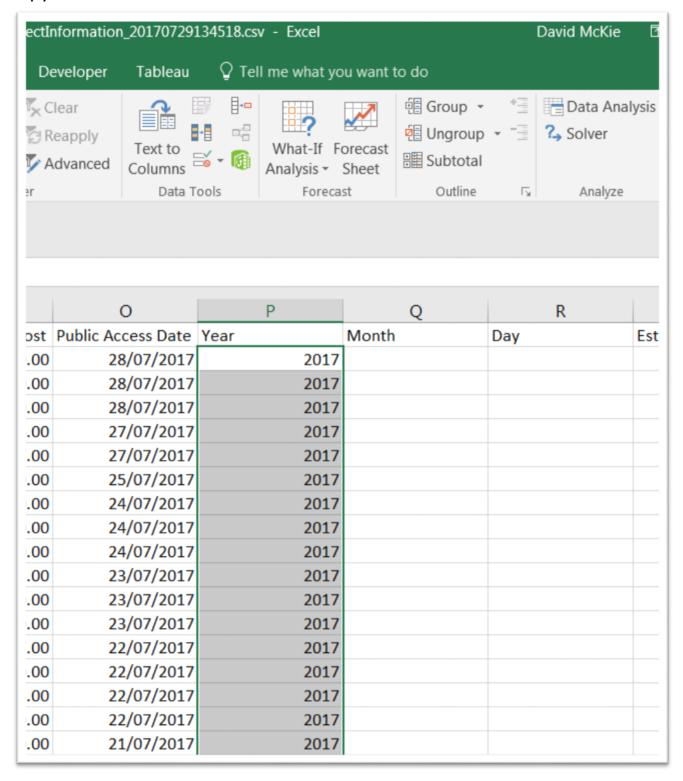
Move your cursor back up to P2, reformat the value as general, or a number with no decimal points, and then copy the result to the bottom of the column.



#### Select the OK tab.

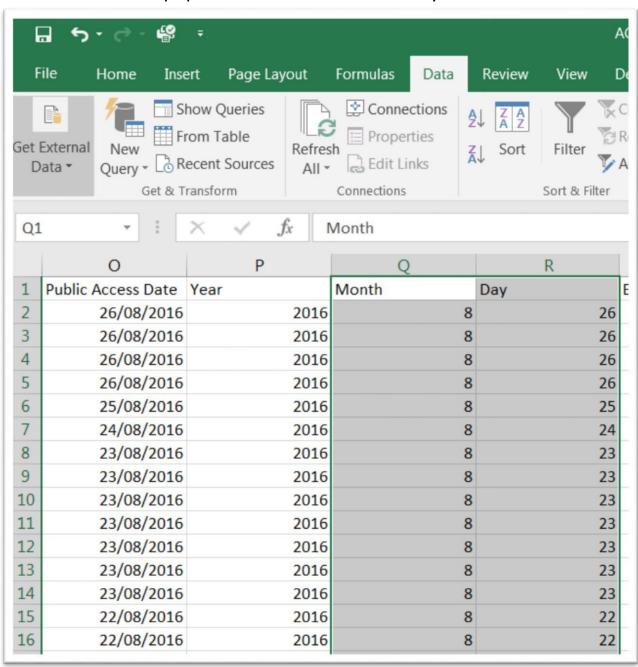


### Copy the new value in P2 to the bottom of the column.



When copying the formula, always scroll to the bottom to make sure all the cells have been populated. Remember, spreadsheets stop copying if they encounter a blank cell.

Following the same process, use the MONTH function <=MONTH(cell reference)> to fill column Q with days of the month, and the DAY function <=DAY> to populate column R with the days of the week.

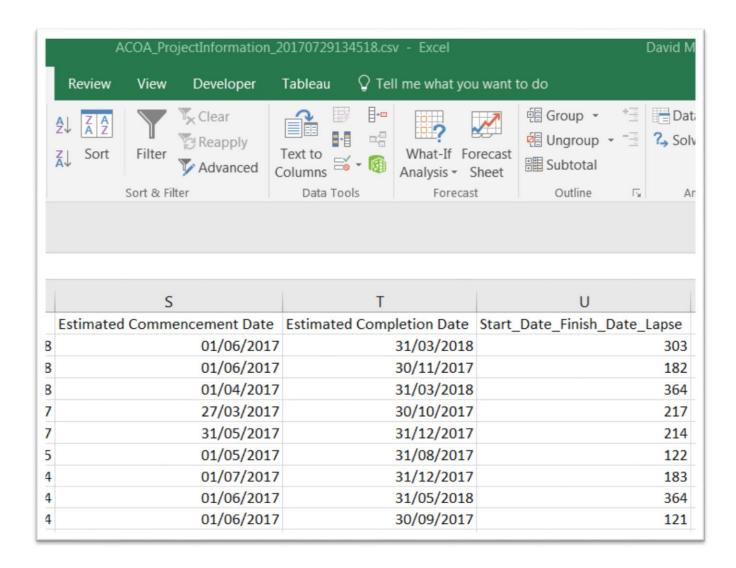


As we learned in this table's <u>data dictionary</u>, the Estimated Commencement Date and Estimated Completion Date are also interesting.

Create a new column to the right of T: "Start\_Date\_Finish\_Date\_Lapse".

As we learned in chapter four of The Data Journalist, because a date in Excel is really a number, we can subtract the oldest date from the most recent one, and then format the result as a number with no decimal places in order to determine the days in between.

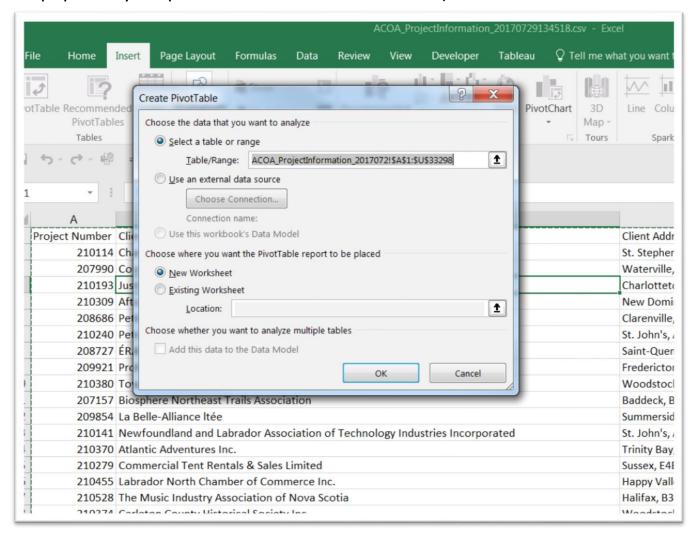
This comes in handy when describing length of time, which might indicate delays, negligence or corruption.



### Task three: How to create a pivot table

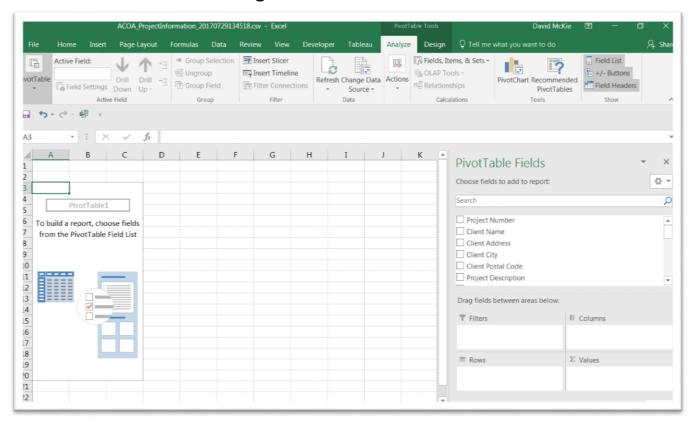
First, let's create the pivot table. Place your cursor anywhere within the table. Go the "Insert" menu of the ribbon, select the "Pivot-Table" option, which produces a dialogue box that identifies the range of cells to be included. Pay attention to the range to ensure that all the cells are covered. Remember, Excel stops at blank cells. If your table has a blank cell, or row that you failed to detect, you'll have missing values. (NOTE: Mac users be sure to select the "manual option". The automatic option

#### will populate your pivot table with random values.)



You'll notice that the table's entire range: <\$A1\$:\$U\$33298>. The dollar signs "anchor" the two extreme cells – A1, and U33298 – in the range. Towards the bottom of the dialogue box, you have the option to display the pivot table in a "New Worksheet", or "Existing Worksheet". Chose the first option, then OK.

(NOTE: You can edit the cell range, by click on the icon to the right of the "Table/Range:" rectangular box, and then manually typing the cell references in the range, being sure to use the dollar sign to anchor the values at each end of the range.

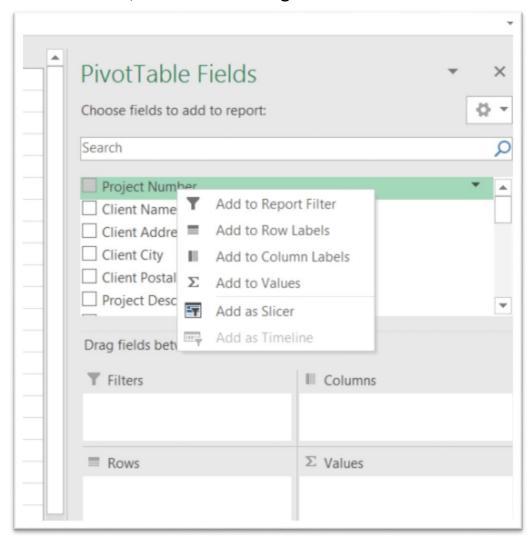


The worksheet is divided into three sections: the "Pivot Table Fields" list; the box containing Filters, Columns, Rows and Values; and the interface to the left.

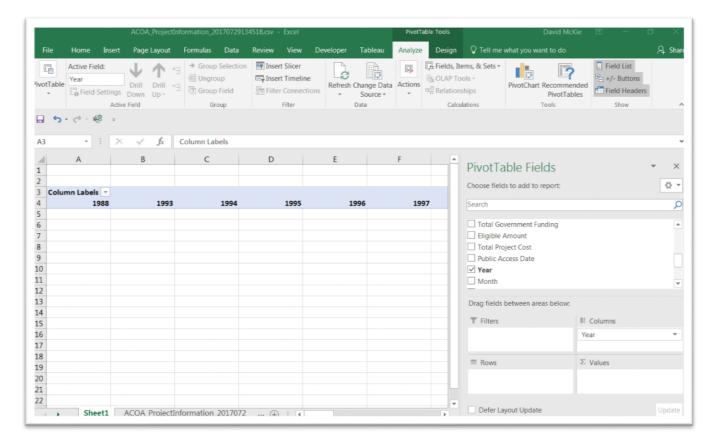
(Note: In a Mac, the box to the right is called a "Pivot Table Builder".)

If you click outside the pivot table, the "Pivot Table Fields" list, or the "Pivot Table Builder" disappears. To get it back, click inside the pivot table.

We can set up the actual layout for the pivot table by either clicking on and dragging the field names to one of the four boxes, or clicking on each of the columns to produce a short-cut menu that contains one of the four boxes, as in the screen grab below.

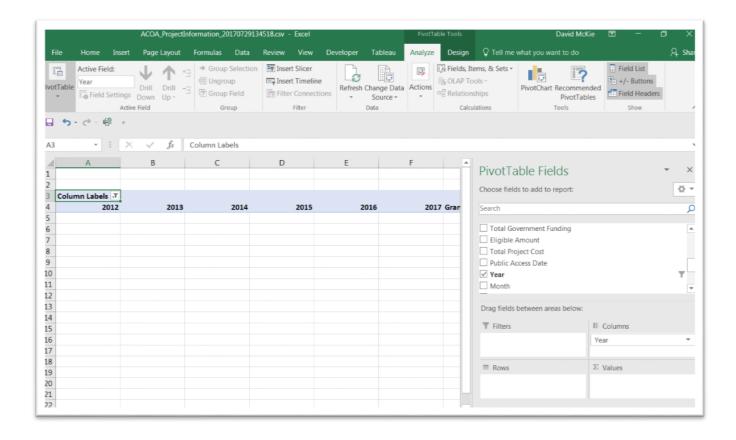


Add "Year" to the "Columns" section, using any of the methods described above.



We can see from the dates, that this dataset stretches back into the past. Unless we're looking for something specific, we may only be interested in five full years worth of data, which in this case is 2012 to 2016. Since 2017 is not yet complete, we wouldn't use it in any analysis that compares year-to-year results.

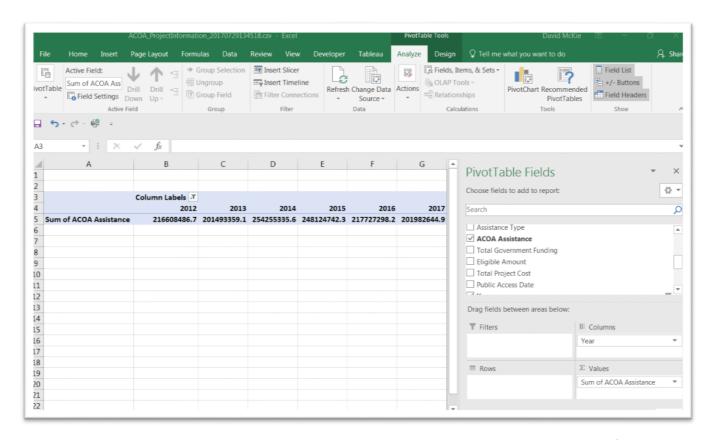
De-select all the years before 2012 from the filter, leaving 2017 for now. To de-select the dates, we'll need to get the short cut menu by selecting the arrow to the right of "Column Labels". From the short cut menu, de-select the unwanted dates.



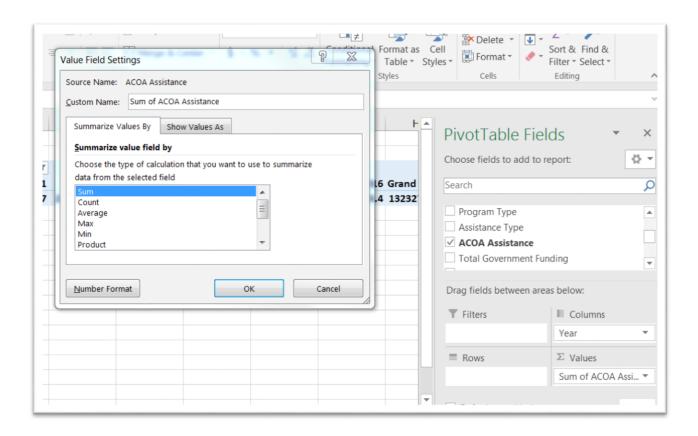
### Task four: How to SUM, COUNT and group by year

Now we want to sum the total number of ACOA grants by year in order to determine if the dollar values are increasing or decreasing.

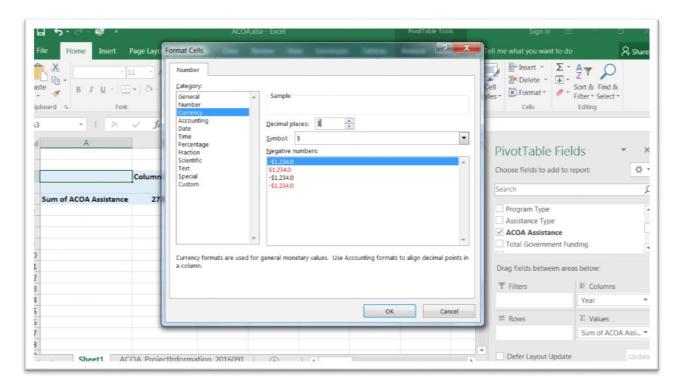
Drag the "ACOA Assistance" row into the "Values" section.



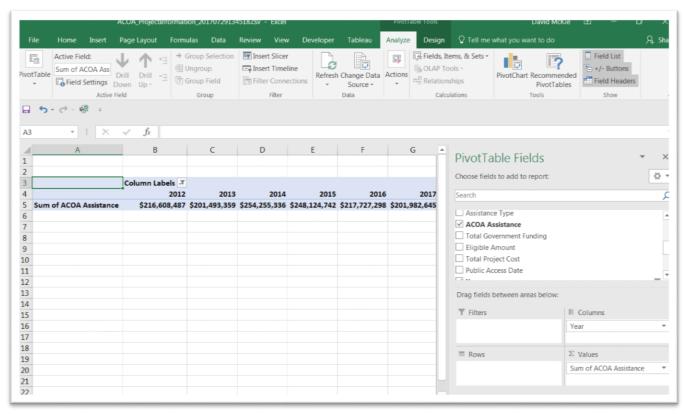
Format the numbers as currency by selecting the arrow to the right of the SUM label for PC users -- or the icon for Mac users that resembles the letter <i> -- to produce a value field setting dialogue box that looks like this.



The pivot table guessed correctly that you want to sum the amount. Reformat the number by selecting the "Number Format" tab, which for Mac users will be on the right-hand side.



Select currency and no decimal places and return to the pivot table.

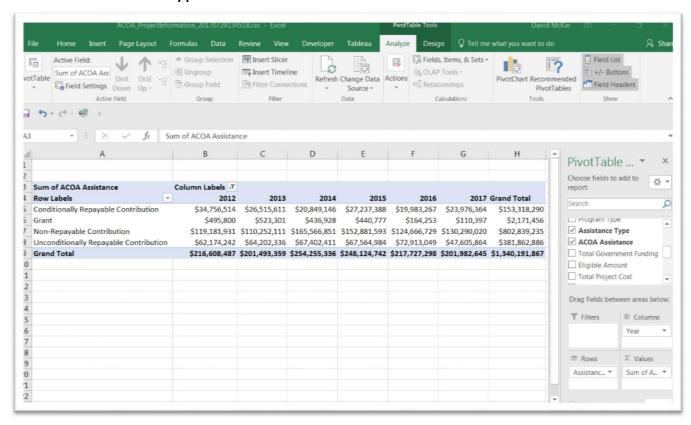


If this was all we were interested in, we could simply copy and use the paste-special option that we learned in the similarly named <u>tutorial</u> that goes with this chapter, and calculate the percent changes. However, from simply eyeballing the numbers, there doesn't seem to be anything worth pursuing.

Let's keep building the pivot table in order to take a deeper dive into the data.

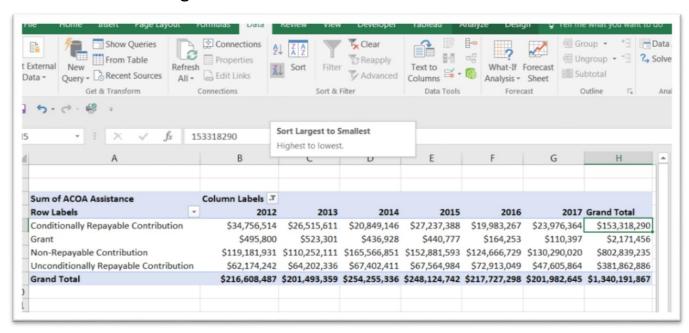
The "Row Labels" section should contain the column on which you want to do your analysis. Put another way, the column you care the most about.

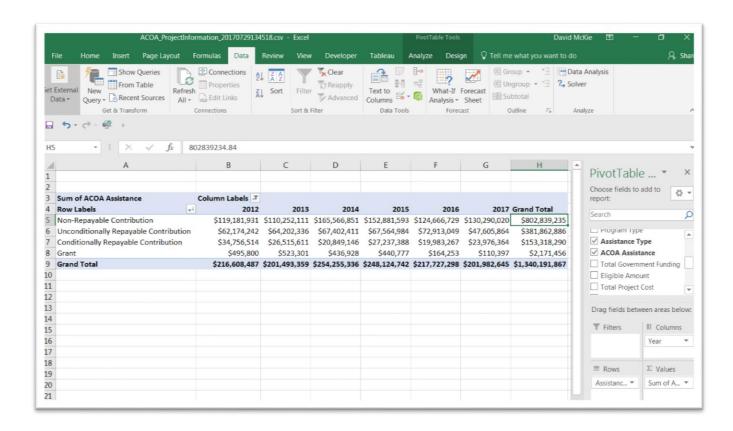
Select "Assistance Type".



Now we can see the categories of financial assistance. Because as journalists we like seeing who got the most, sort the "Grand Total"

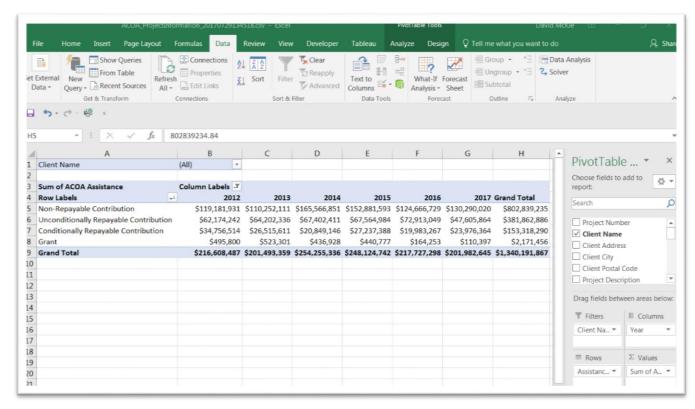
### section in descending order.





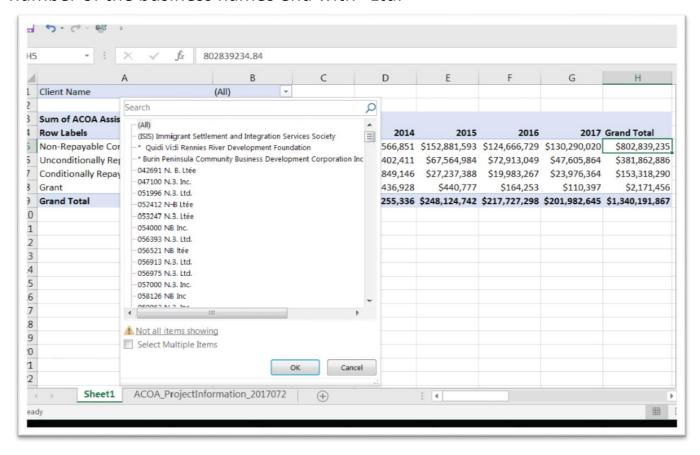
The "Non-Repayable Contribution" category takes top spot. We can also see how the other types behave year after year. Be sure to consult the <u>data dictionary</u> for a definition of each category.

Now it's time for a deeper dive. What if we wanted to define the kinds of institutions receiving these non-repayable contributions. Drag the "Client Name" column into the "Filters" section, which you can see in cells A1 and B1.

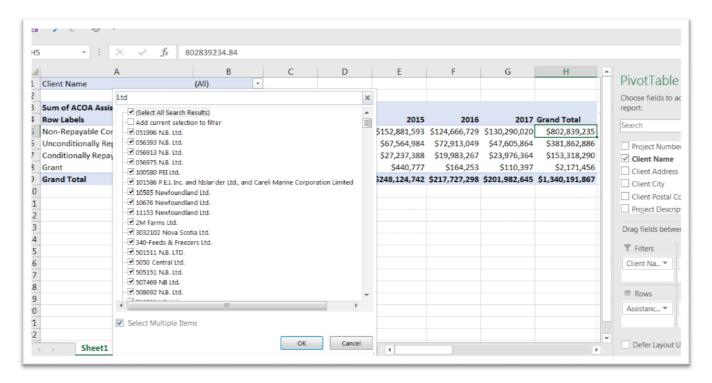


Click on the arrow to obtain the drop-down menu. From perusing the categories, it's obvious that many businesses receive assistance. A

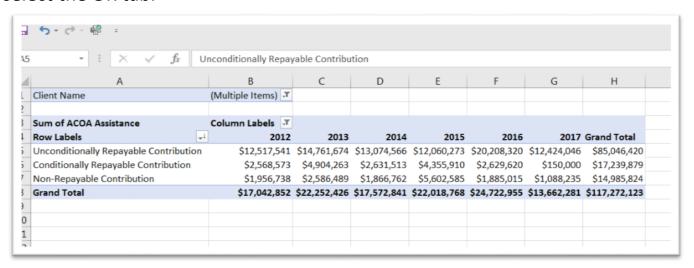
#### number of the business names end with "Ltd."



There are too many values to display in this filter box. If there is a particular company or organization of interest, then you can search the original table, and then return to the pivot table to plug the names into the search box that you see in the screen grab above.



Select "Ltd.", the acronym for the designation limited. Once you've made the selection, in this case the companies with the "Ltd." suffix, select the OK tab.

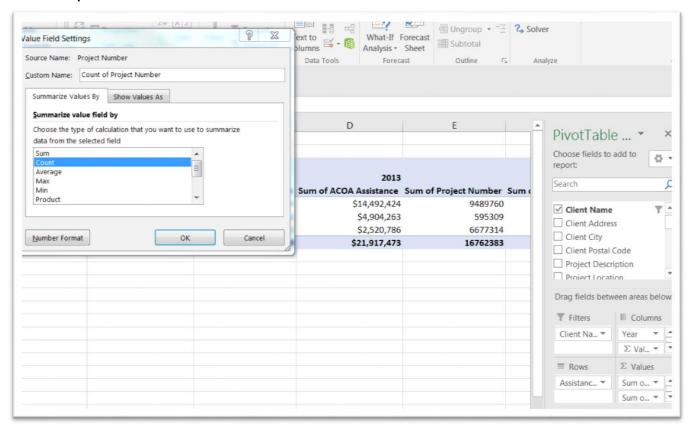


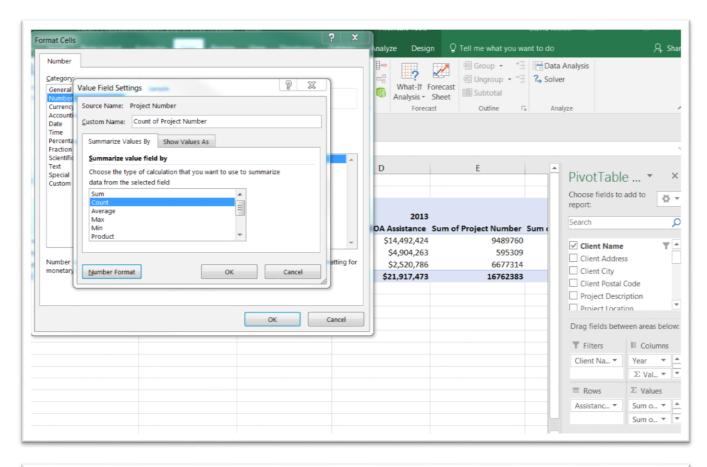
ACOA hands out most of its money to businesses with the "Ltd" suffix as in the form of unconditionally repayable contributions. If you're happy with this result, you can give the pivot table a specific name, and return to your original table to build another one. Or if you wanted to

perform some math on the values such as percent of total, or year-toyear changes, you could use the paste-special option that we learned in Chapter 4's fourth online tutorial to create a table in a new worksheet.

You could also add one more variable to the pivot table's "Values" section. In this case, we can COUNT the number of recipients.

Drag project number into values, and format it as a number with no decimal places.





2016		2017		Total Sum of ACOA Assistance	<b>Total Count of Project Number</b>
um of ACOA Assistance	<b>Count of Project Number</b>	Sum of ACOA Assistance	Count of Project Number		
\$19,858,320	57	\$12,224,046	35	\$83,615,098	29
\$2,629,620	6	\$150,000	1	\$17,239,879	3
\$1,800,247	39	\$1,063,235	24	\$14,312,073	24
\$24,288,187	102	\$13,437,281	60	\$115,167,050	56

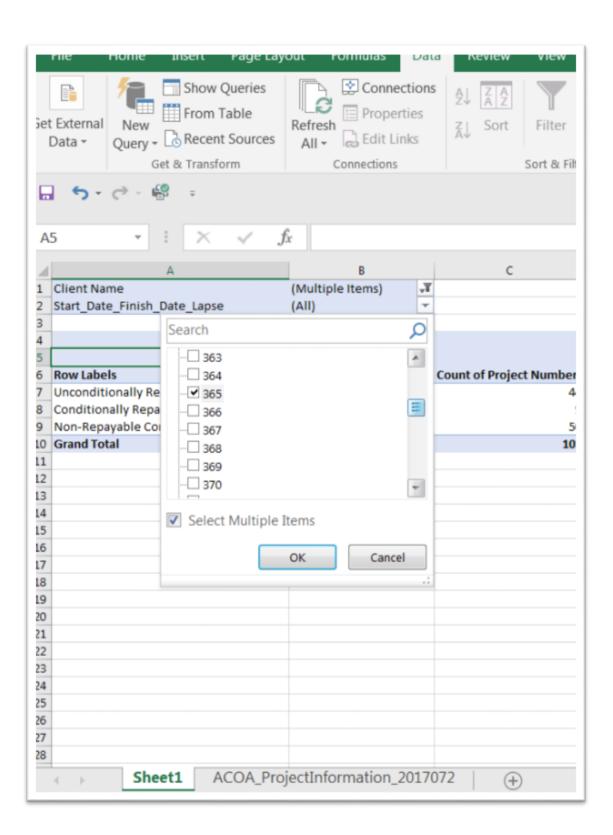
To the right of each year, and the grand total, the pivot table has added a column that COUNTS the number of recipients. In 2016, 57 corporations received \$19.8 million in unconditional repayable contributions.

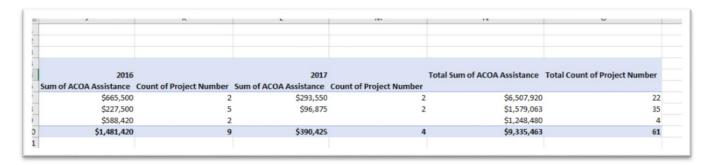
## Task five: How to filter by project length

We can also add a second column to the filter. You'll recall in task number one that we calculated the difference between the start and

finish dates. Let's limit our selection to projects that took, say, 365 days.

Drag the "Start\_Date\_Finish\_Date\_Lapse" column into the "Filters" section, and select 365 from the drop-down menu.





In 2016, two projects estimated to take 365 days to complete received, \$655,500 in "unconditional repayable contributions".

Since the drop-down menu sorts the amounts in descending order, we can simply select the lengths of time that we want, a potentially tedious process that could take a few minutes, but one that's worth while if it leads to a good story.

The possibilities with this dataset are endless, in large part because it's large and contains many categories that stretch over many years.

The pivot table allows us to take a deep dive into the data for ideas that could lead to original stories.