Chapter 4

Using Pivot Tables in Excel (Live Exercise with Data)

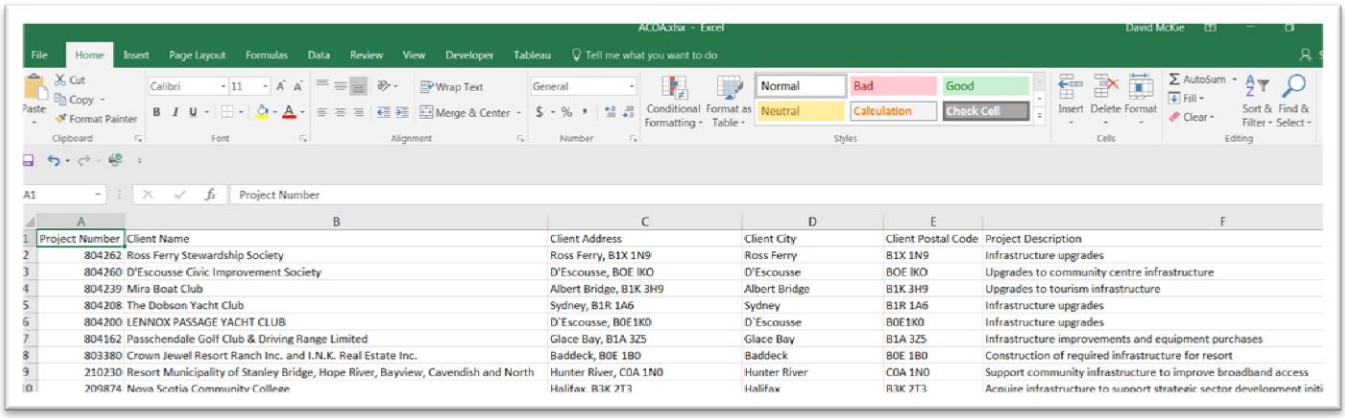
In chapter four, we used B.C.'s political donations data to learn how to build pivot tables, which group elements in your data and summarize the information using totals and subtotals. Because pivot tables feature so prominently in the work of journalists, we'll use a live dataset to explore some of 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 Atlantic Canada Opportunities Agency (ACOA) 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.

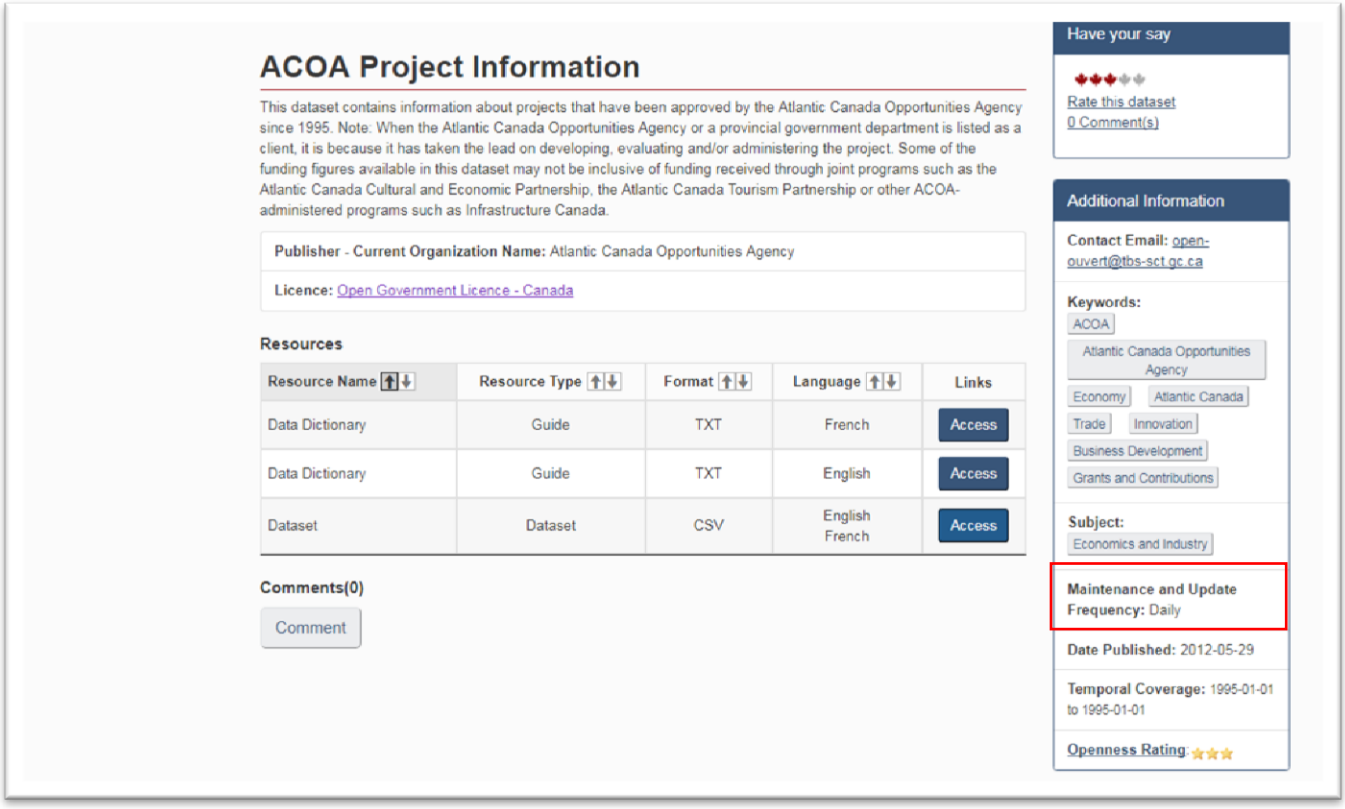
Task one: How to download tables from an open-data site



Project Number	Client Name	Client Address	Client City	Client Postal Code	Project Description
804262	Ross Ferry Stewardship Society	Ross Ferry, B1X 1N9	Ross Ferry	B1X 1N9	Infrastructure upgrades
804260	D'Escousse Civic Improvement Society	D'Escousse, B0E 1K0	D'Escousse	B0E 1K0	Upgrades to community centre infrastructure
804239	Mira Boat Club	Albert Bridge, B1K 3H9	Albert Bridge	B1K 3H9	Upgrades to tourism infrastructure
804208	The Dobson Yacht Club	Sydney, B1R 1A6	Sydney	B1R 1A6	Infrastructure upgrades
804200	LENNOX PASSAGE YACHT CLUB	D'Escousse, B0E1K0	D'Escousse	B0E1K0	Infrastructure upgrades
804162	Passchendale Golf Club & Driving Range Limited	Glace Bay, B1A 3Z5	Glace Bay	B1A 3Z5	Infrastructure improvements and equipment purchases
803380	Crown Jewel Resort Ranch Inc. and I.N.K. Real Estate Inc.	Baddeck, B0E 1B0	Baddeck	B0E 1B0	Construction of required infrastructure for resort
210230	Resort Municipality of Stanley Bridge, Hope River, Bayview, Cavendish and North	Hunter River, C0A 1N0	Hunter River	C0A 1N0	Support community infrastructure to improve broadband access
209874	Nova Scotia Community College	Halifax, B3K 2T3	Halifax	B3K 2T3	Acquire infrastructure to support strategic sector development initi

Download the "Dataset" from the federal government's open data website, as well as the data dictionary 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 "Data Dictionary" (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.

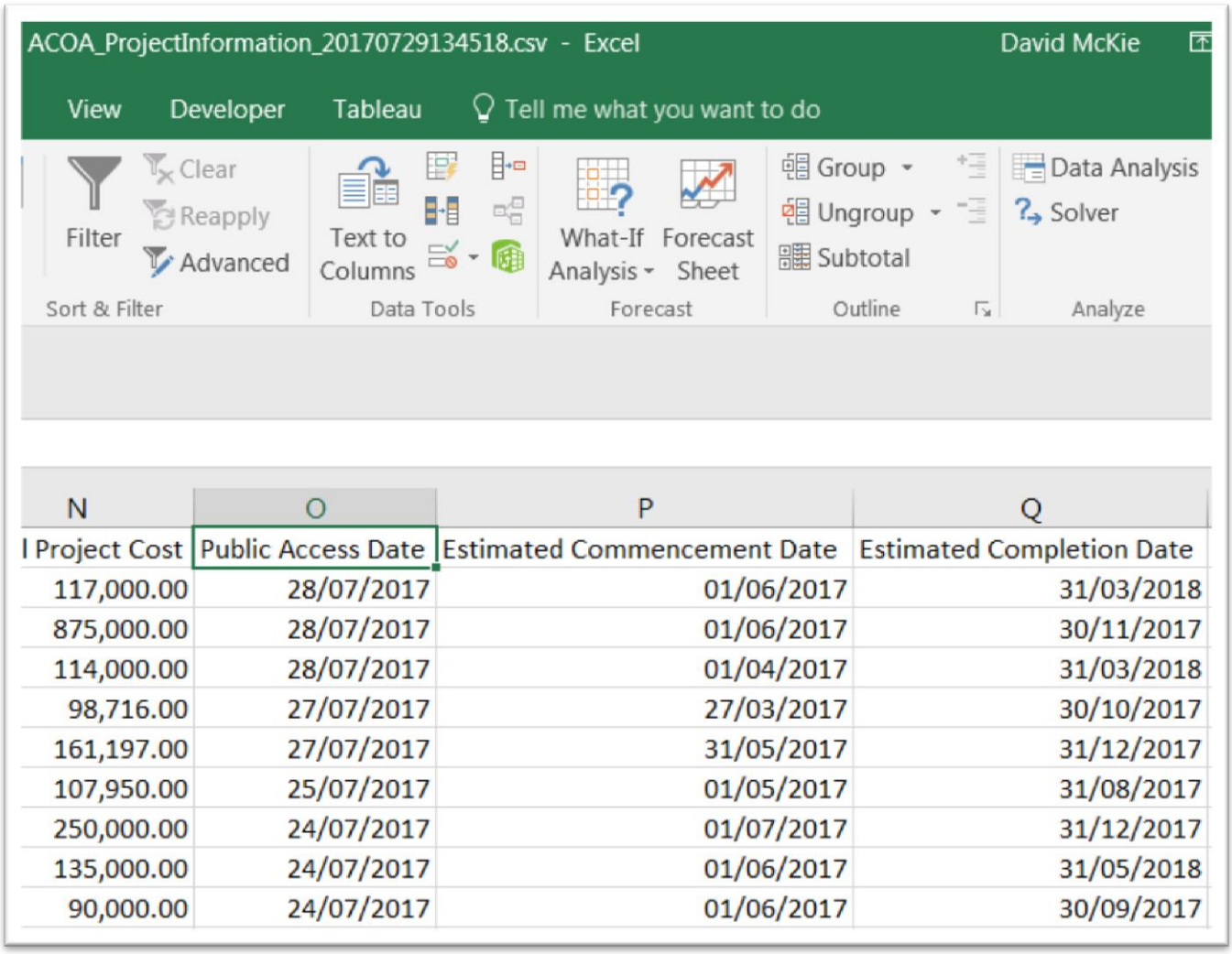
Save the csv file as an Excel table.

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.

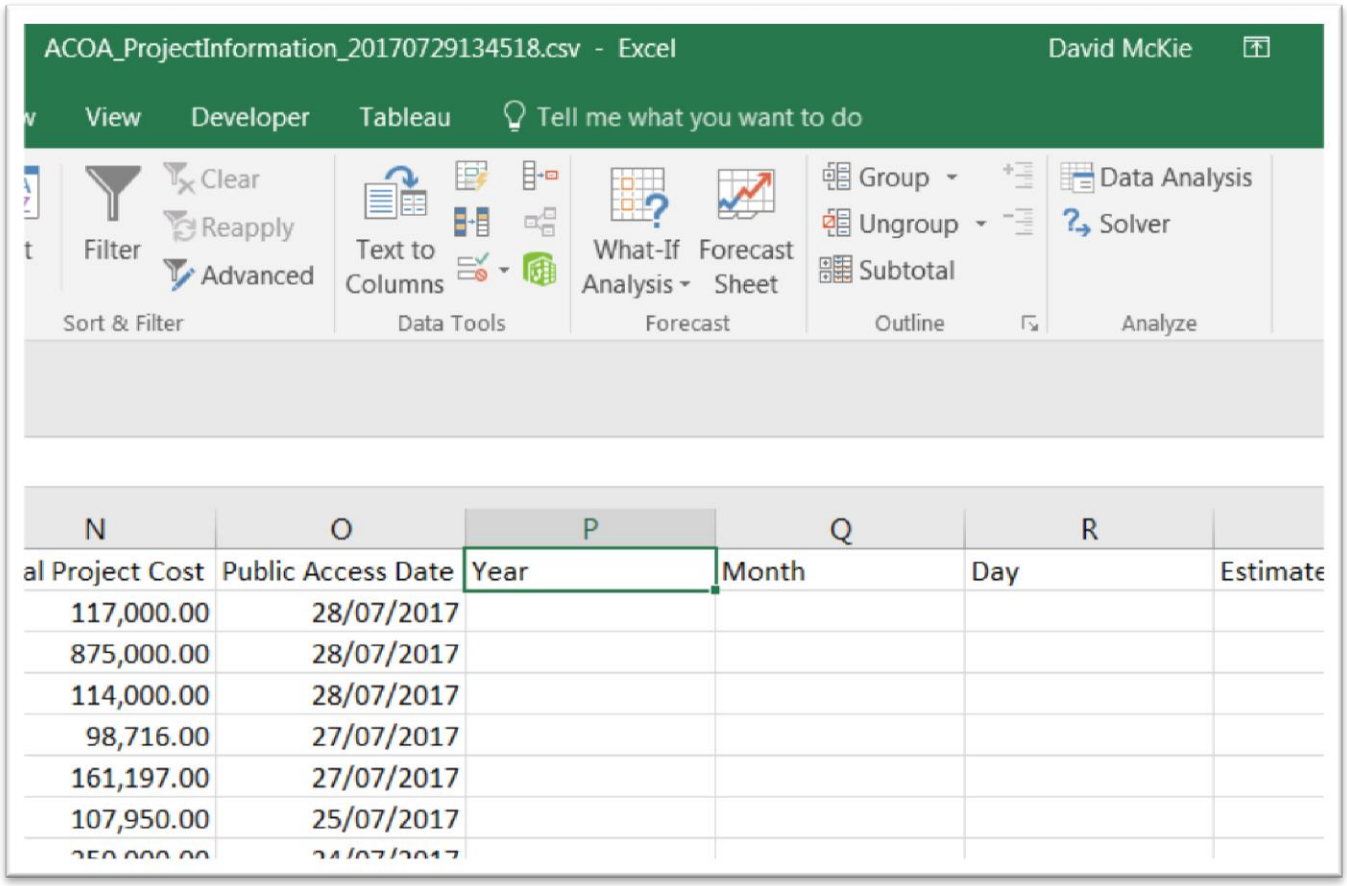
Sort column O, "Public Access Date", in descending order to determine the most recent entry.



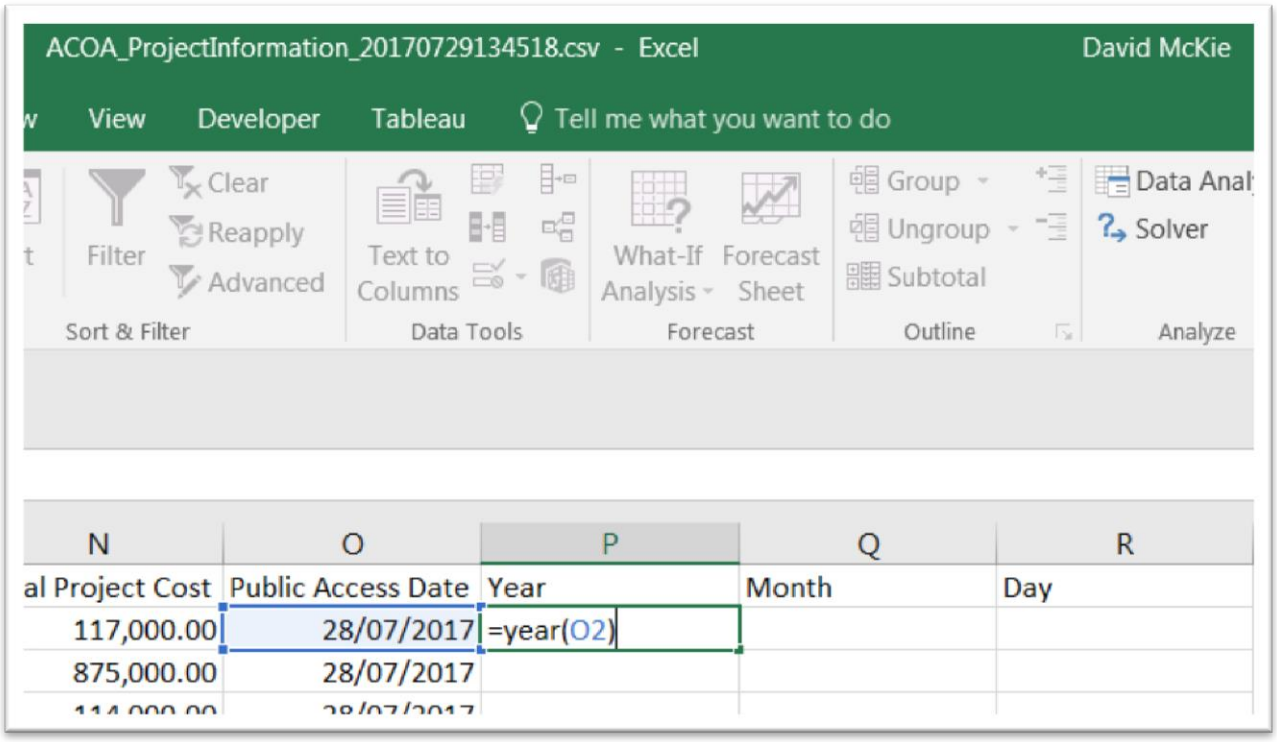
N	O	P	Q
l Project Cost	Public Access Date	Estimated Commencement Date	Estimated Completion Date
117,000.00	28/07/2017	01/06/2017	31/03/2018
875,000.00	28/07/2017	01/06/2017	30/11/2017
114,000.00	28/07/2017	01/04/2017	31/03/2018
98,716.00	27/07/2017	27/03/2017	30/10/2017
161,197.00	27/07/2017	31/05/2017	31/12/2017
107,950.00	25/07/2017	01/05/2017	31/08/2017
250,000.00	24/07/2017	01/07/2017	31/12/2017
135,000.00	24/07/2017	01/06/2017	31/05/2018
90,000.00	24/07/2017	01/06/2017	30/09/2017

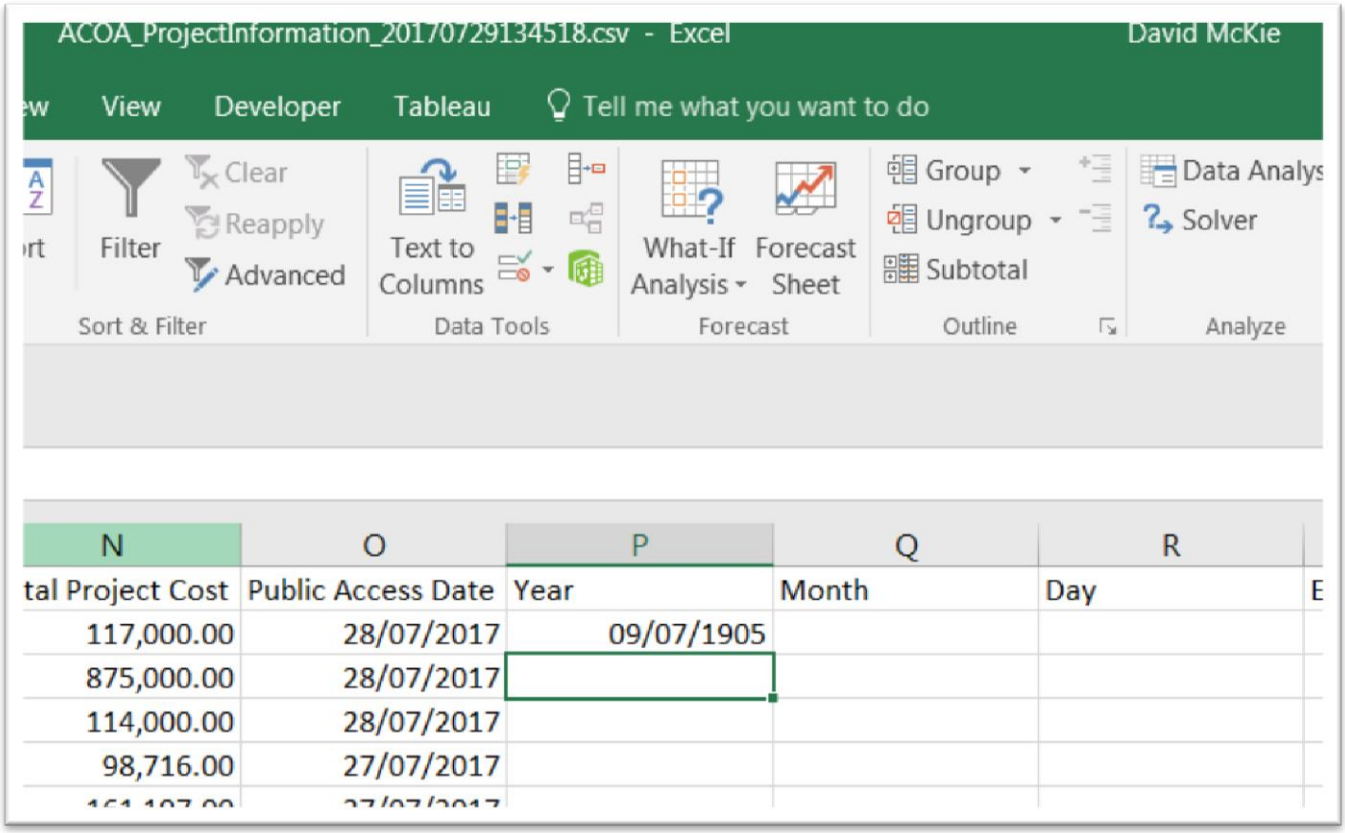
According to the data dictionary that we downloaded at the beginning of this task, the Public Access Date is "usually 60 days following client acceptance of written offer." The Estimated Commencement Date, is the "date the project is expected to begin." And the Estimated Completion Date is "the date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

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

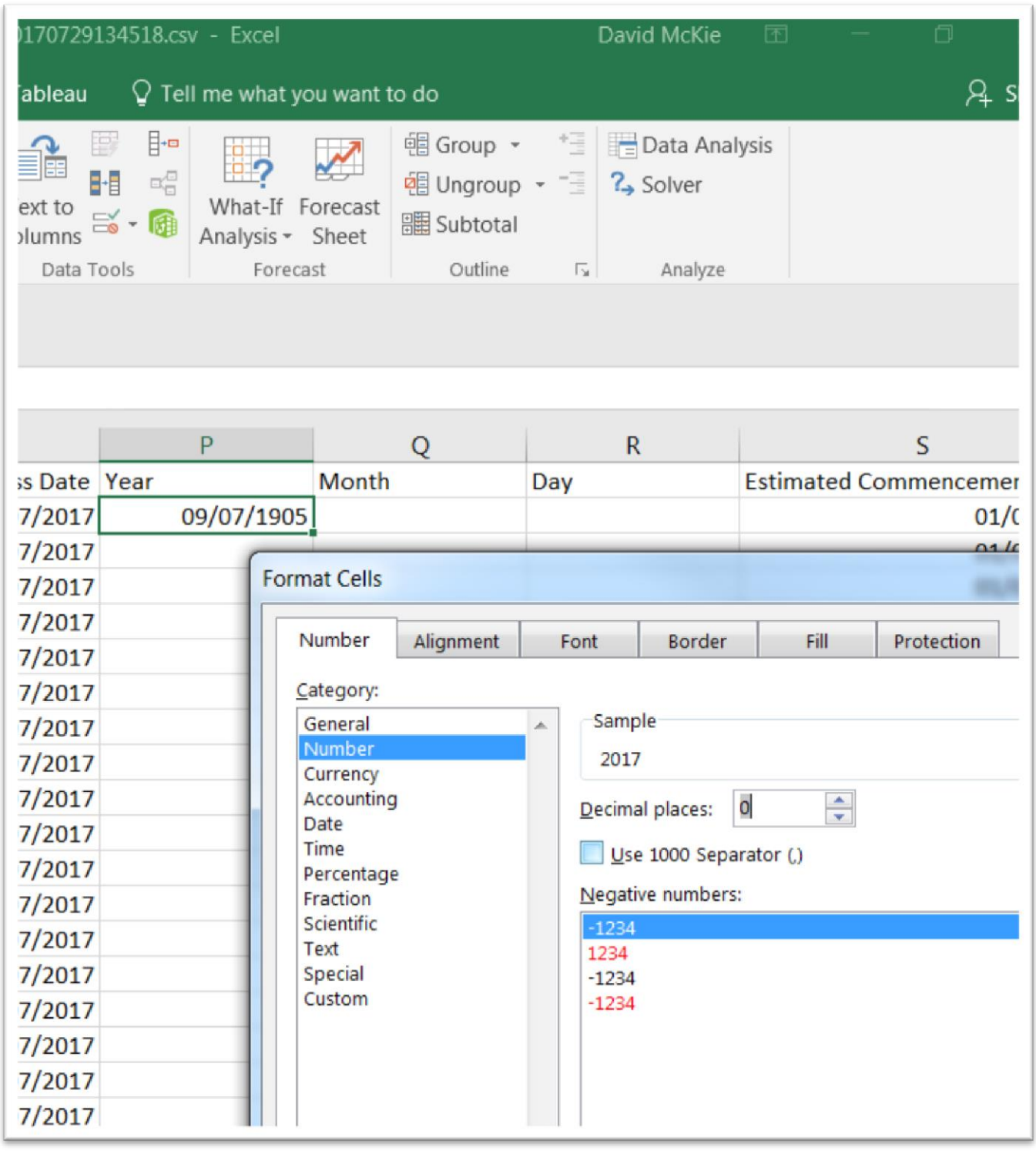


We'll use the Year function <=YEAR(cell reference)> to extract the year -- a task explained on page 14 of The Data Journalist's on-line tutorial called "Working with Specialized Functions and Formulas in Excel" -- to pull the year out of the date.

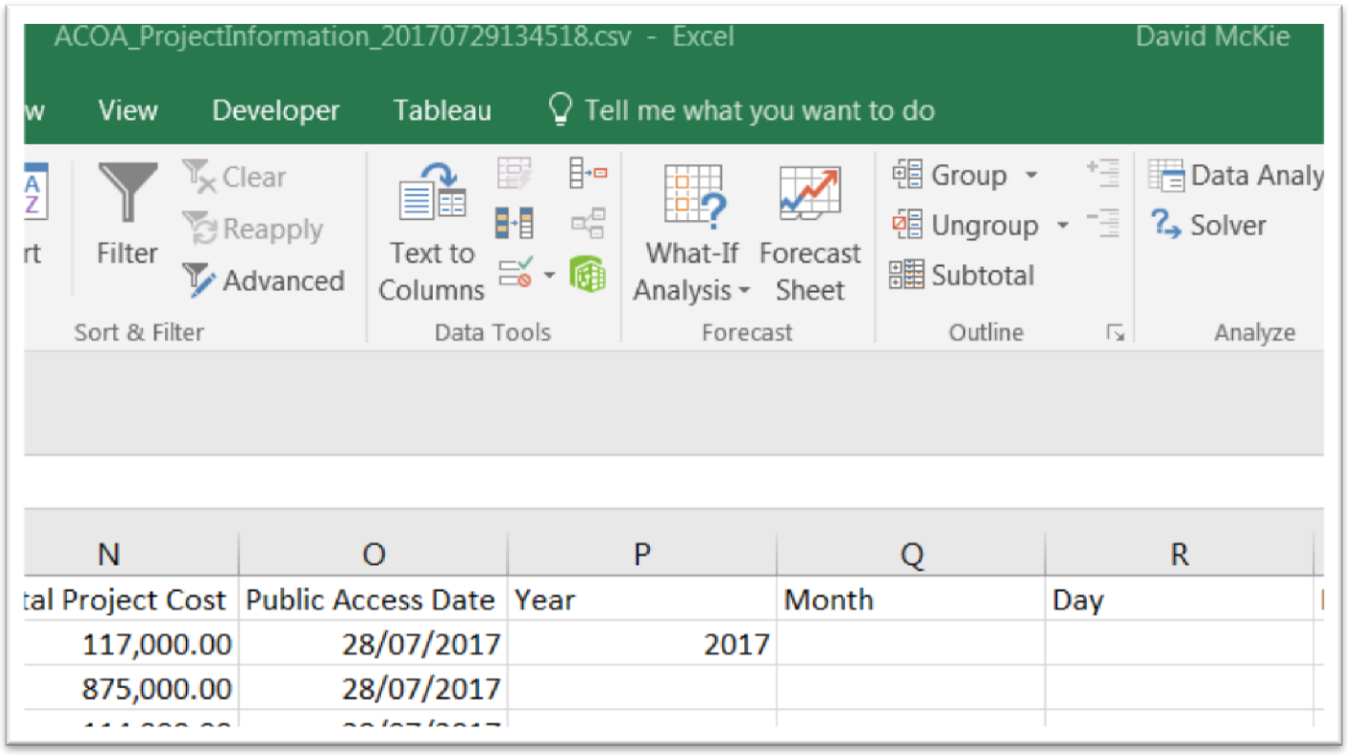

ACOA_ProjectInformation_20170729134518.csv - Excel
David McKie
View
Developer
Tableau
Tell me what you want to do
Clear
Reapply
Advanced
Filter
Sort & Filter
Text to Columns
Data Tools
What-If Analysis
Forecast Sheet
Forecast
Group
Ungroup
Subtotal
Outline
Data Anal
Solver
Analyze
N
O
P
Q
R
al Project Cost
Public Access Date
Year
Month
Day
117,000.00
28/07/2017
=year(O2)
875,000.00
28/07/2017


ACOA_ProjectInformation_20170729134518.csv - Excel
David McKie
View
Developer
Tableau
Tell me what you want to do
Clear
Reapply
Advanced
Filter
Sort & Filter
Text to Columns
Data Tools
What-If Analysis
Forecast Sheet
Forecast
Group
Ungroup
Subtotal
Outline
Data Analys
Solver
Analyze
N
O
P
Q
R
tal Project Cost
Public Access Date
Year
Month
Day
117,000.00
28/07/2017
09/07/1905
875,000.00
28/07/2017
114,000.00
28/07/2017
98,716.00
27/07/2017

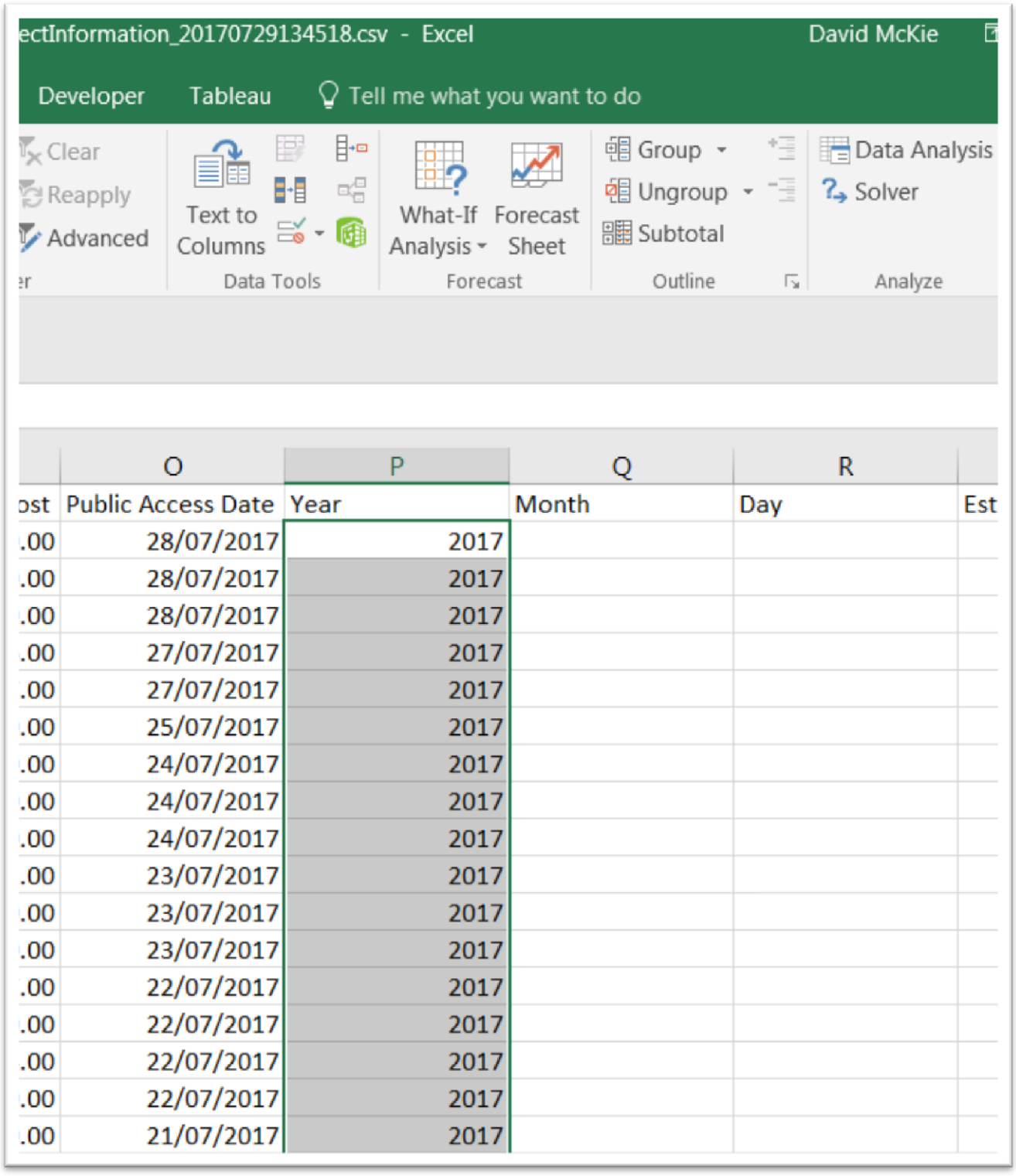
Move your cursor back up to P2, reformat the value as general, or a number with no decimal points, and then copy the result to the bottom of the column.



Select the OK tab.



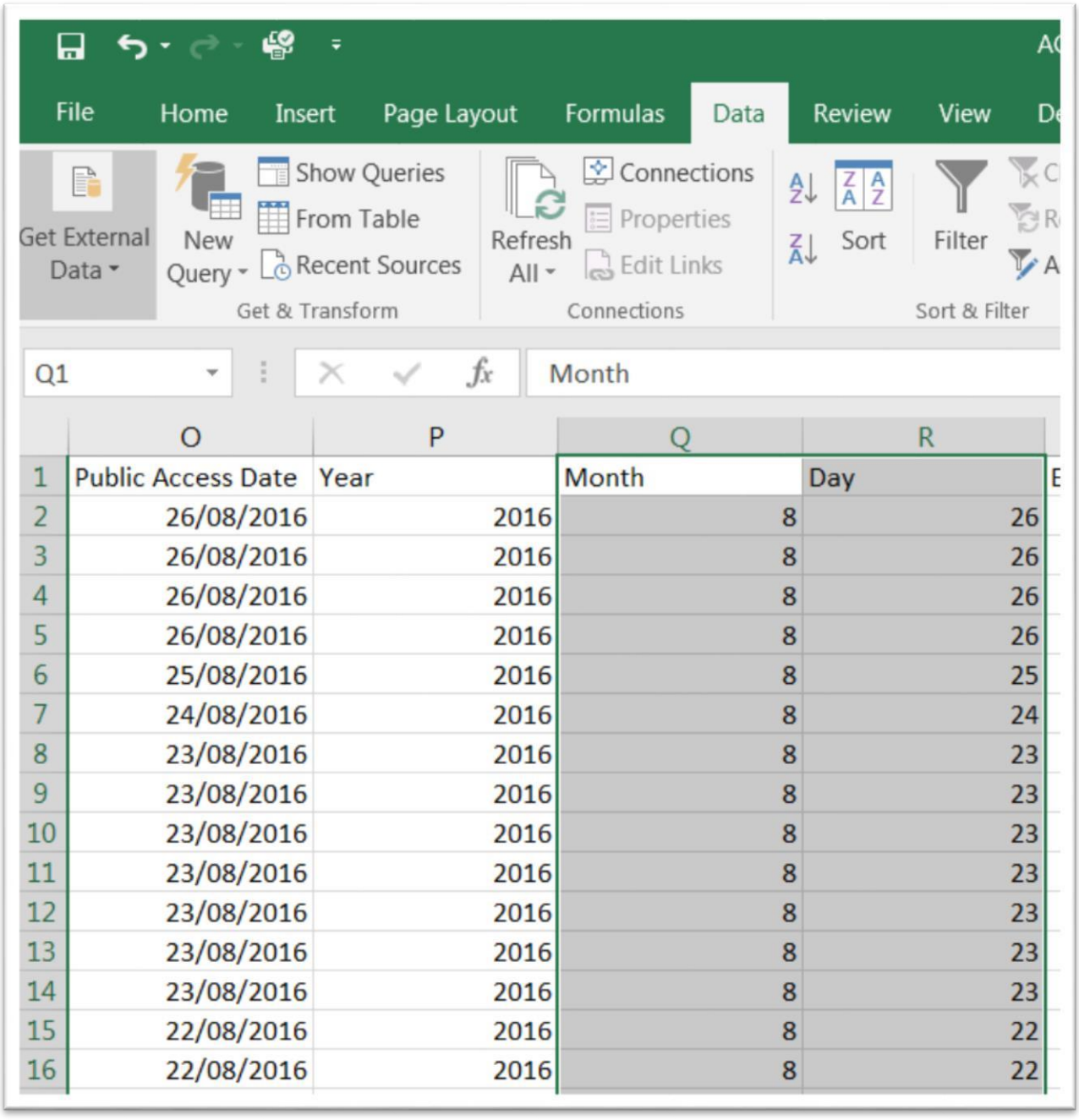
Copy the new value in P2 to the bottom of the column.



	O	P	Q	R	
ost	Public Access Date	Year	Month	Day	Est
.00	28/07/2017	2017			
.00	28/07/2017	2017			
.00	28/07/2017	2017			
.00	27/07/2017	2017			
.00	27/07/2017	2017			
.00	25/07/2017	2017			
.00	24/07/2017	2017			
.00	24/07/2017	2017			
.00	24/07/2017	2017			
.00	23/07/2017	2017			
.00	23/07/2017	2017			
.00	23/07/2017	2017			
.00	22/07/2017	2017			
.00	22/07/2017	2017			
.00	22/07/2017	2017			
.00	22/07/2017	2017			
.00	21/07/2017	2017			

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.



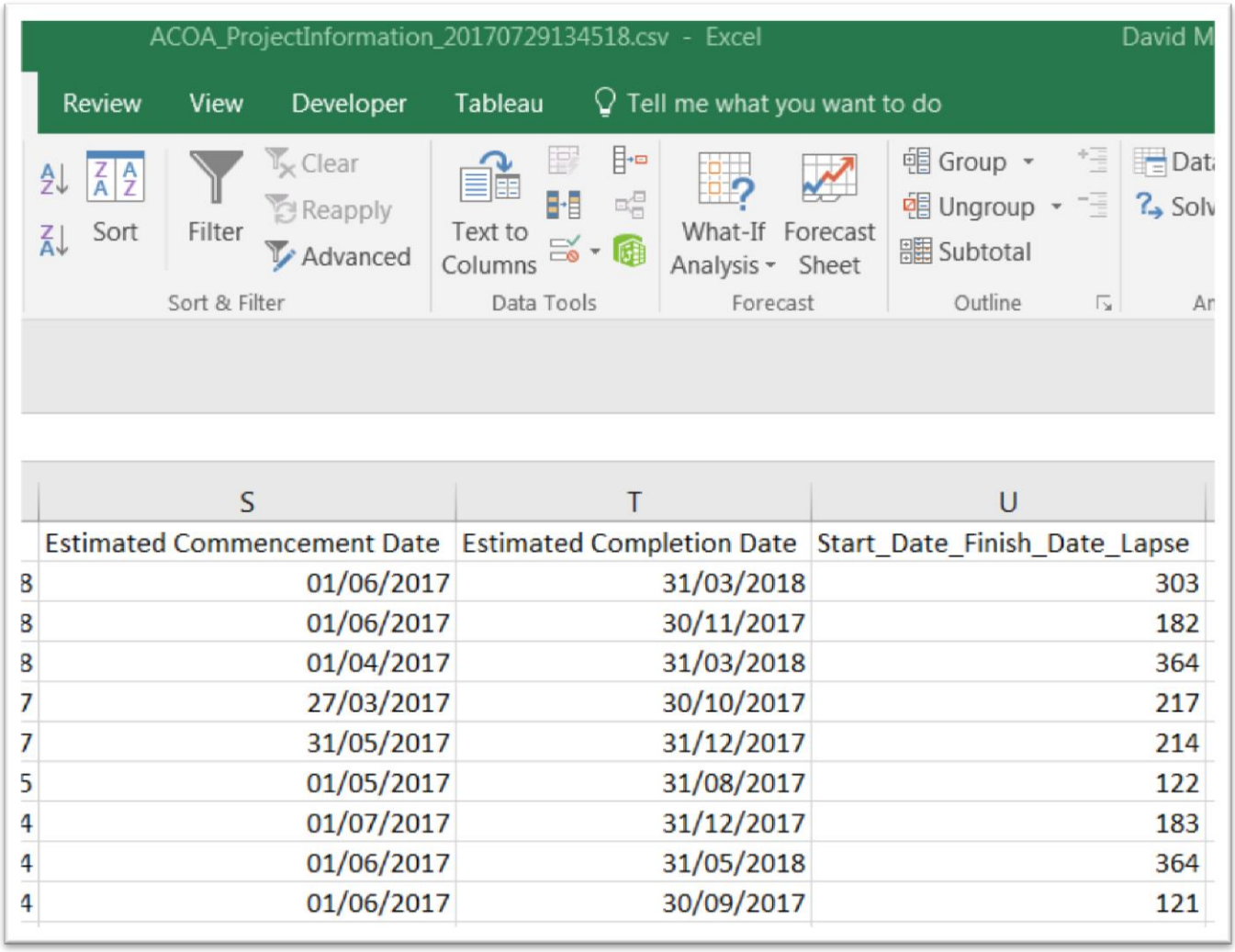
	O	P	Q	R
1	Public Access Date	Year	Month	Day
2	26/08/2016	2016	8	26
3	26/08/2016	2016	8	26
4	26/08/2016	2016	8	26
5	26/08/2016	2016	8	26
6	25/08/2016	2016	8	25
7	24/08/2016	2016	8	24
8	23/08/2016	2016	8	23
9	23/08/2016	2016	8	23
10	23/08/2016	2016	8	23
11	23/08/2016	2016	8	23
12	23/08/2016	2016	8	23
13	23/08/2016	2016	8	23
14	23/08/2016	2016	8	23
15	22/08/2016	2016	8	22
16	22/08/2016	2016	8	22

As we learned in this table's data dictionary, 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.

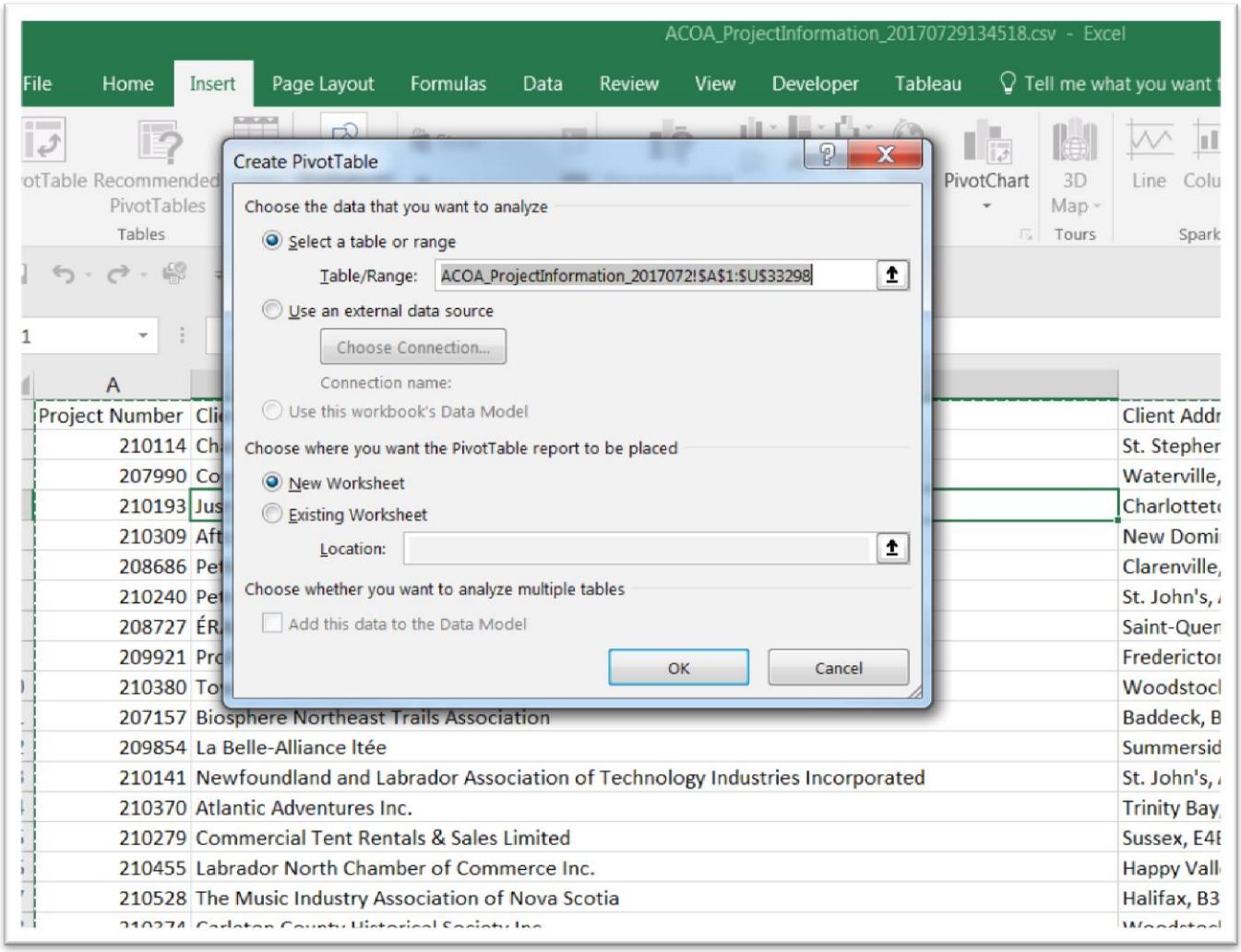
This comes in handy when describing length of time, which might indicate delays, negligence or corruption.



	S	T	U
	Estimated Commencement Date	Estimated Completion Date	Start_Date_Finish_Date_Lapse
8	01/06/2017	31/03/2018	303
8	01/06/2017	30/11/2017	182
8	01/04/2017	31/03/2018	364
7	27/03/2017	30/10/2017	217
7	31/05/2017	31/12/2017	214
5	01/05/2017	31/08/2017	122
4	01/07/2017	31/12/2017	183
4	01/06/2017	31/05/2018	364
4	01/06/2017	30/09/2017	121

Task three: How to create a pivot table

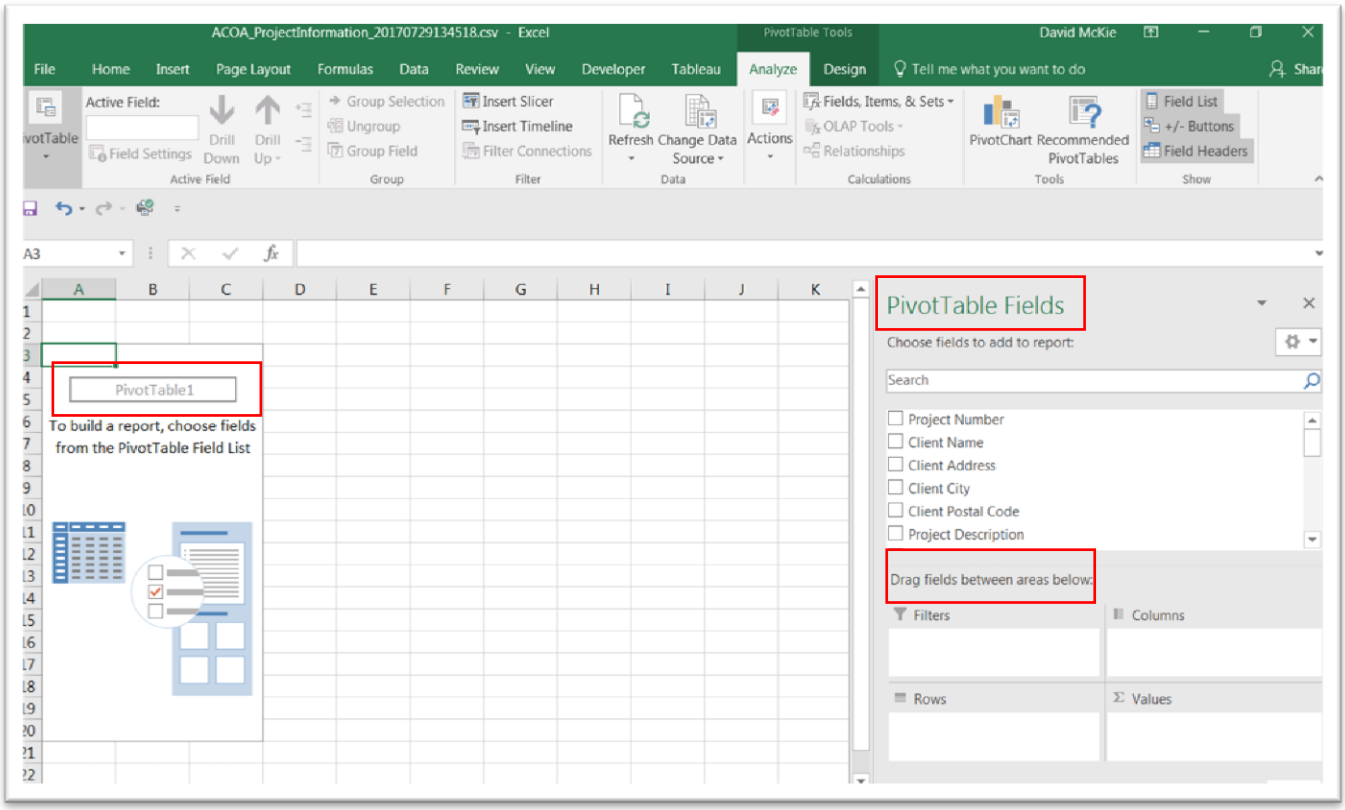
First, let's create the pivot table. Place your cursor anywhere within the table. Go the "Insert" menu of the ribbon, select the "Pivot-Table" option, which produces a dialogue box that identifies the range of cells to be included. Pay attention to the range to ensure that all the cells are covered. 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: <$A1$:U33298>. 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.



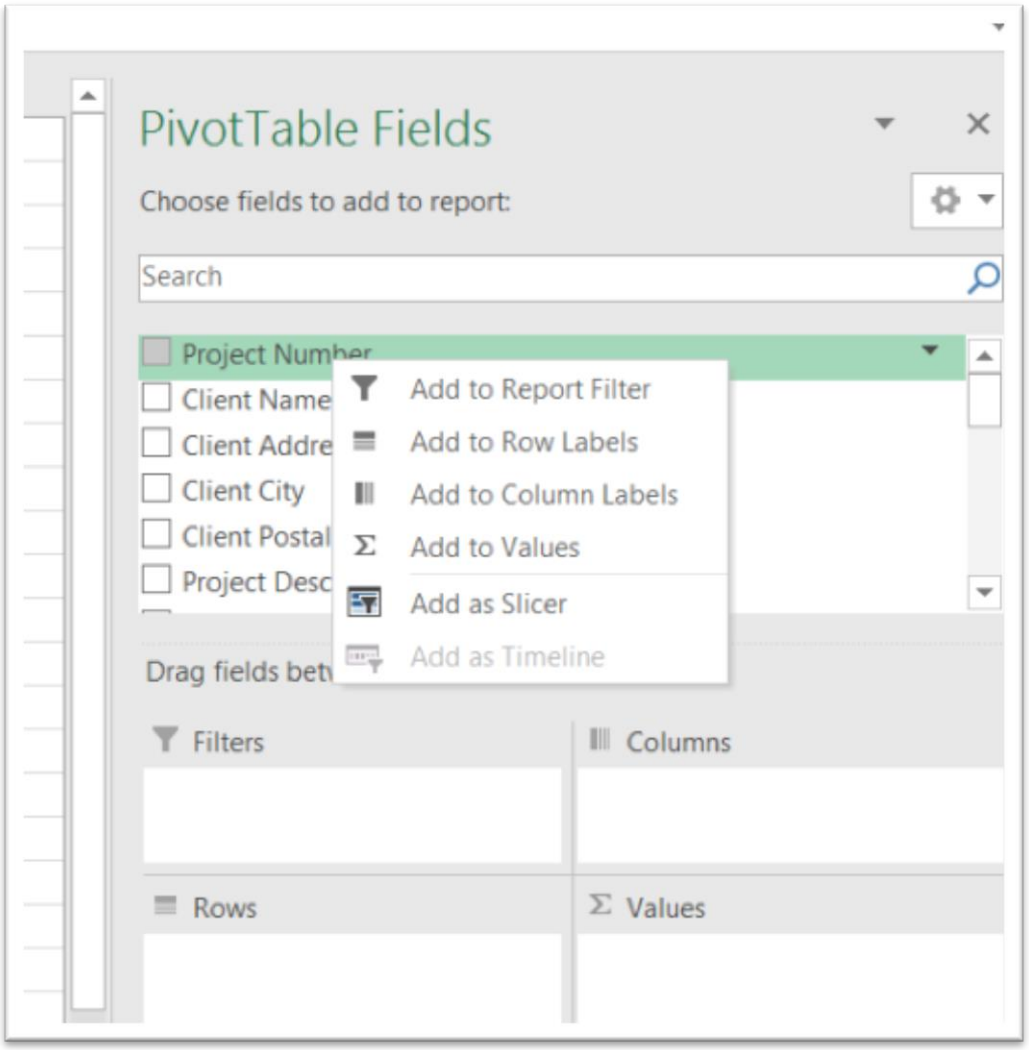
The worksheet is divided into three sections: the "Pivot Table Fields" list; the box containing Filters, Columns, Rows and Values; and the interface to the left.

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

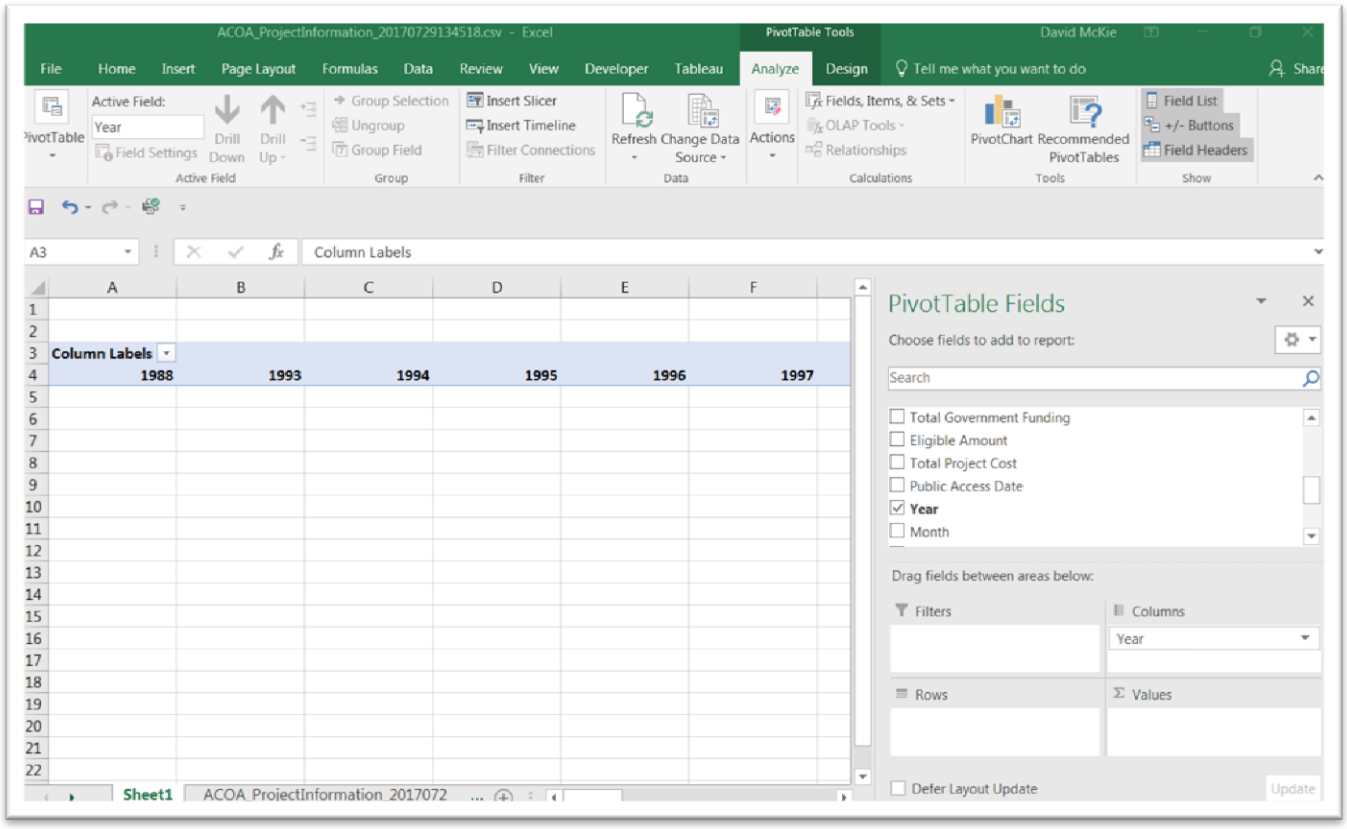
If you click outside the pivot table, the "Pivot Table Fields" list, or the "Pivot Table Builder" disappears. To get it back, click inside the pivot table.

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.

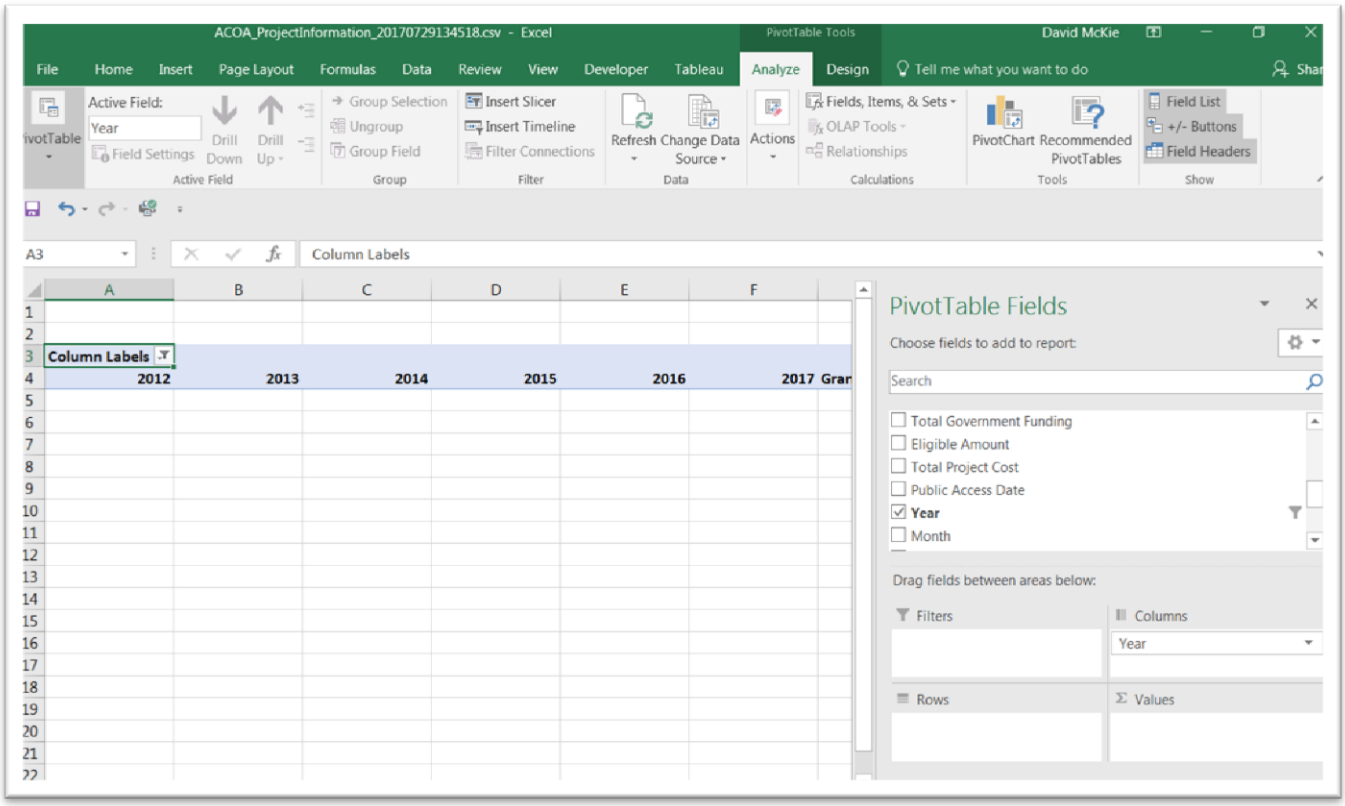


Add "Year" to the "Columns" section, using any of the methods described above.



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.

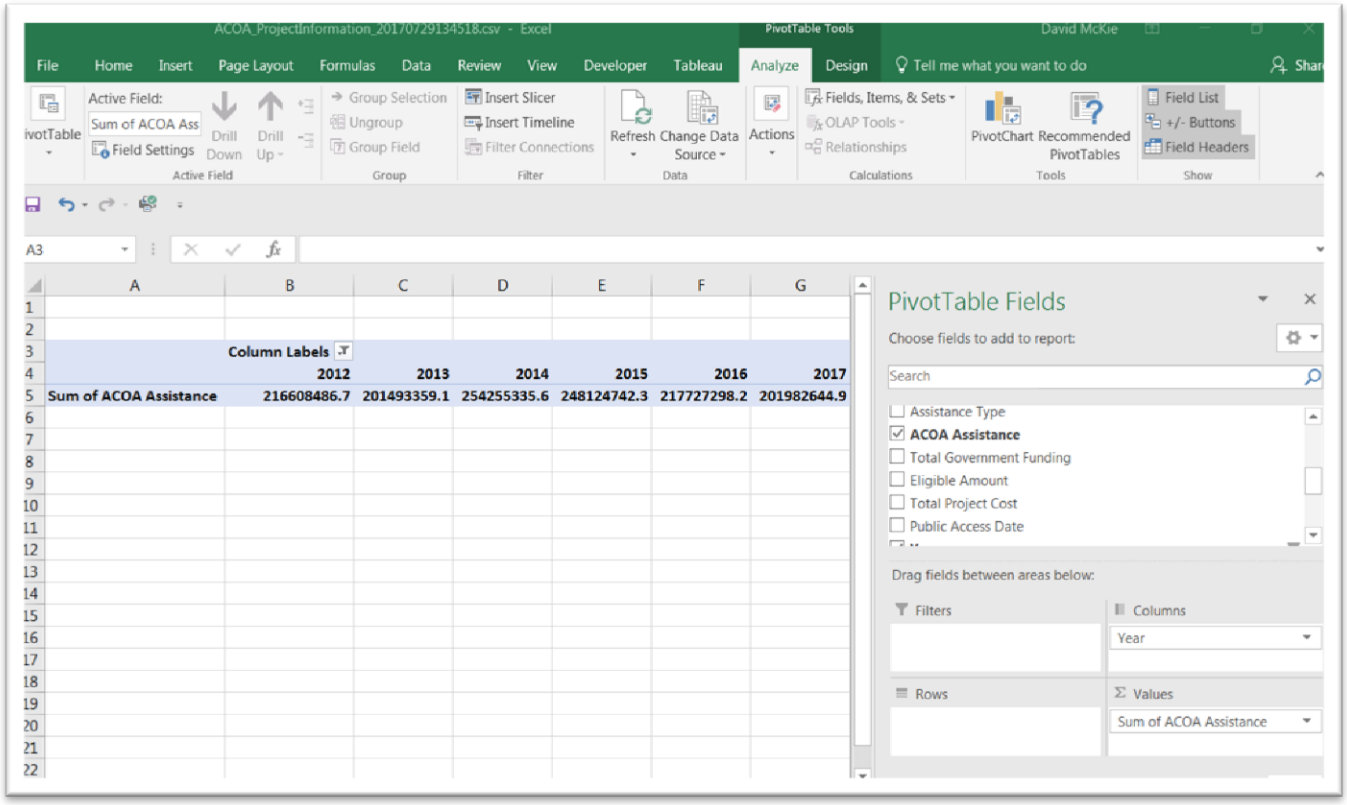
De-select all the years before 2012 from the filter, leaving 2017 for now. To de-select the dates, we'll need to get the short cut menu by selecting the arrow to the right of "Column Labels". From the short cut menu, de-select the unwanted dates.


ACOA_ProjectInformation_20170729134518.csv - Excel
PivotTable Tools
David McKie
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Tableau
Analyze
Design
Tell me what you want to do
Active Field:
Year
Field Settings
Drill Down
Drill Up
Group Selection
Ungroup
Group Field
Insert Slicer
Insert Timeline
Filter Connections
Refresh
Change Data Source
Actions
Fields, Items, & Sets
OLAP Tools
Relationships
PivotChart
Recommended PivotTables
Field List
+/- Buttons
Field Headers
Active Field
Group
Filter
Data
Calculations
Tools
Show
A3
Column Labels
Column Labels
2012
2013
2014
2015
2016
2017
PivotTable Fields
Choose fields to add to report:
Search
Total Government Funding
Eligible Amount
Total Project Cost
Public Access Date
Year
Month
Drag fields between areas below:
Filters
Columns
Year
Rows
Values

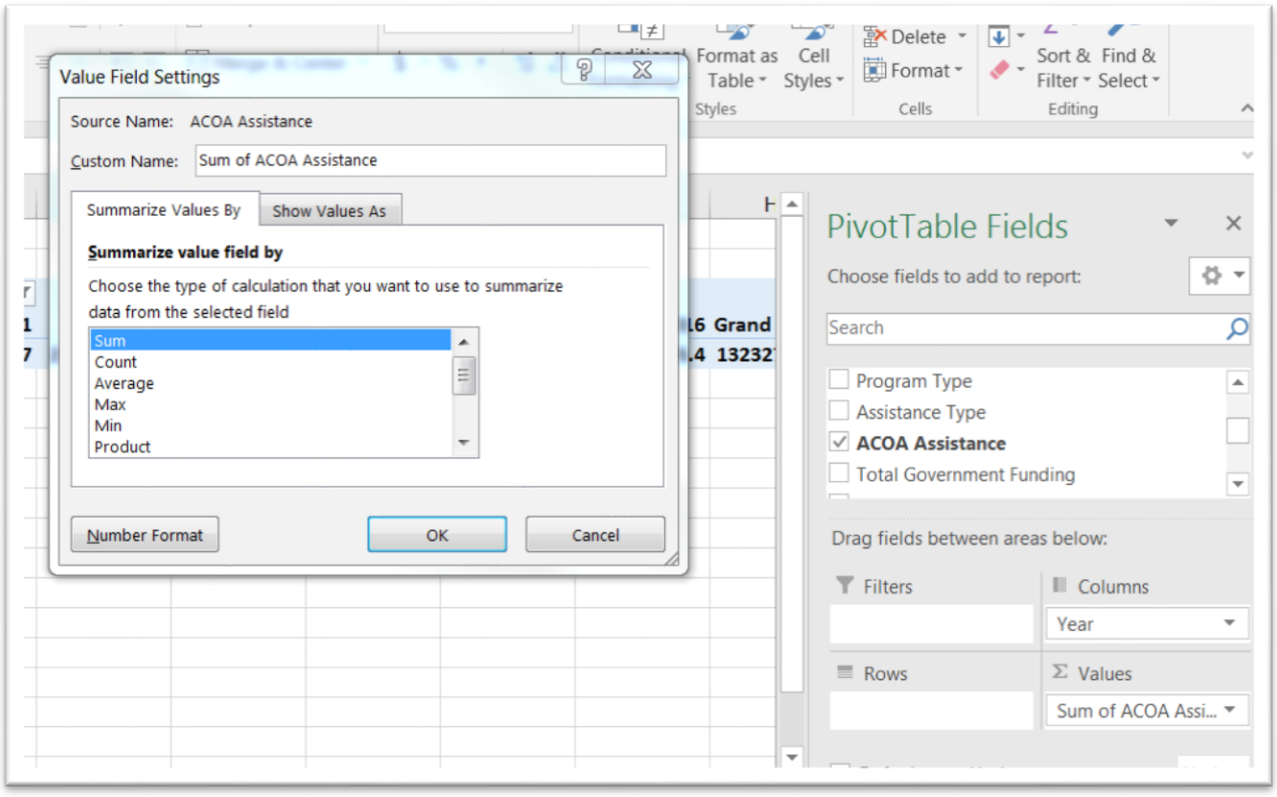
Task four: How to SUM, COUNT and group by year

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.

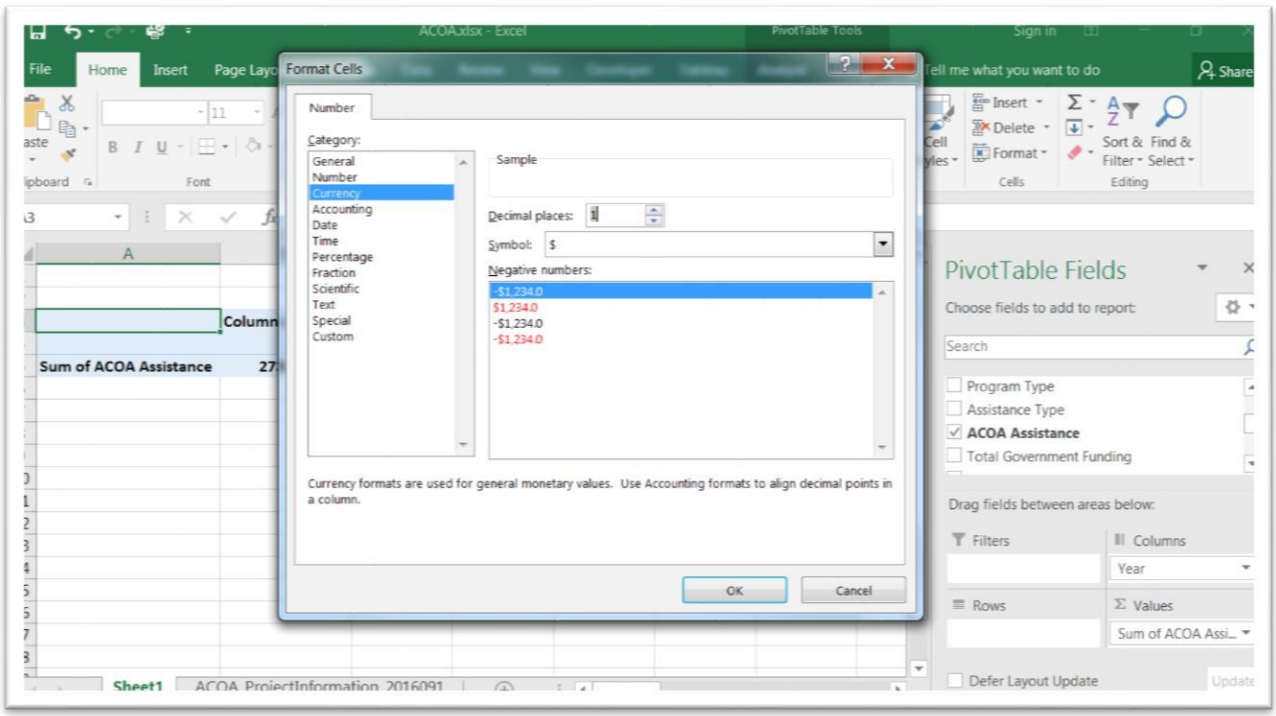
Drag the "ACOA Assistance" row into the "Values" section.



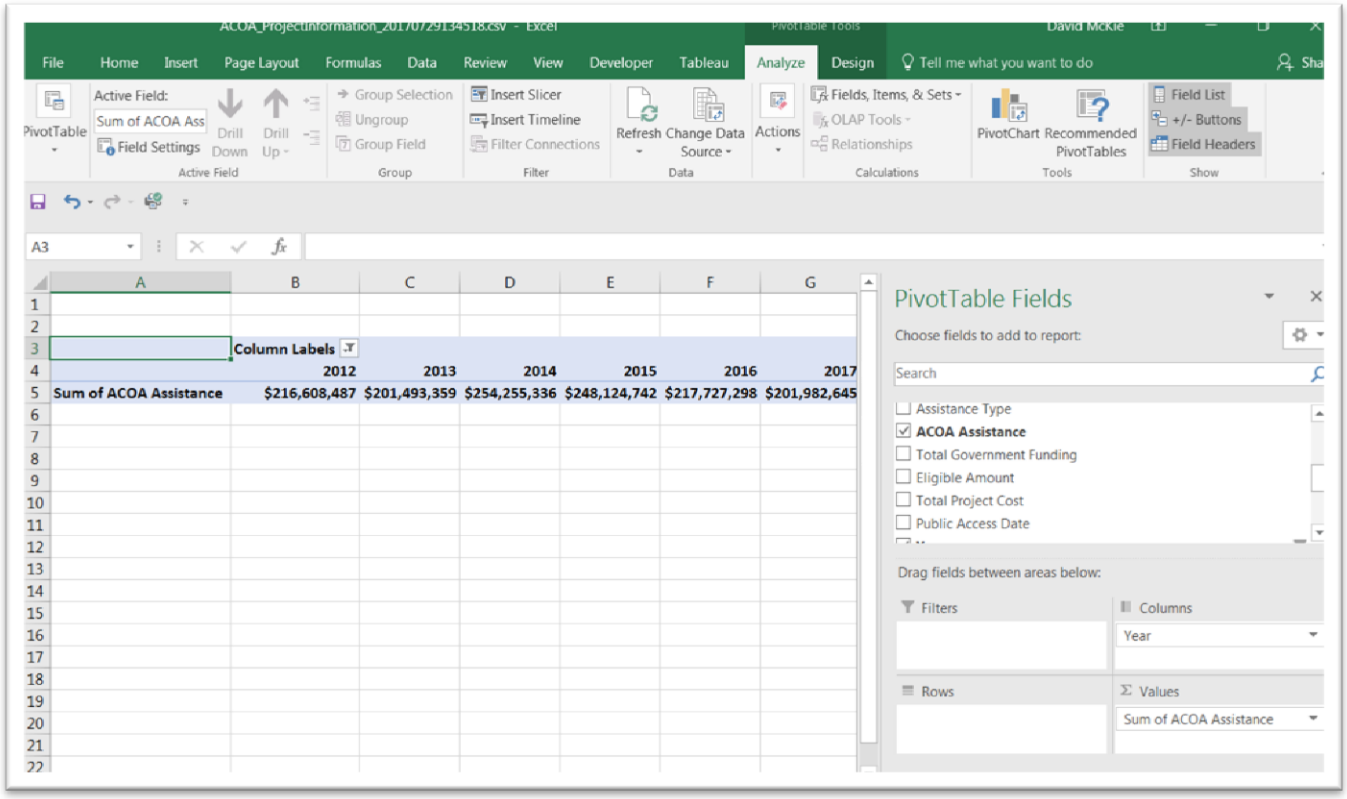
Format the numbers as currency by selecting the arrow to the right of the SUM label for PC users -- or the icon for Mac users that resembles the letter <i> -- to produce a value field setting dialogue box that looks like this.



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.

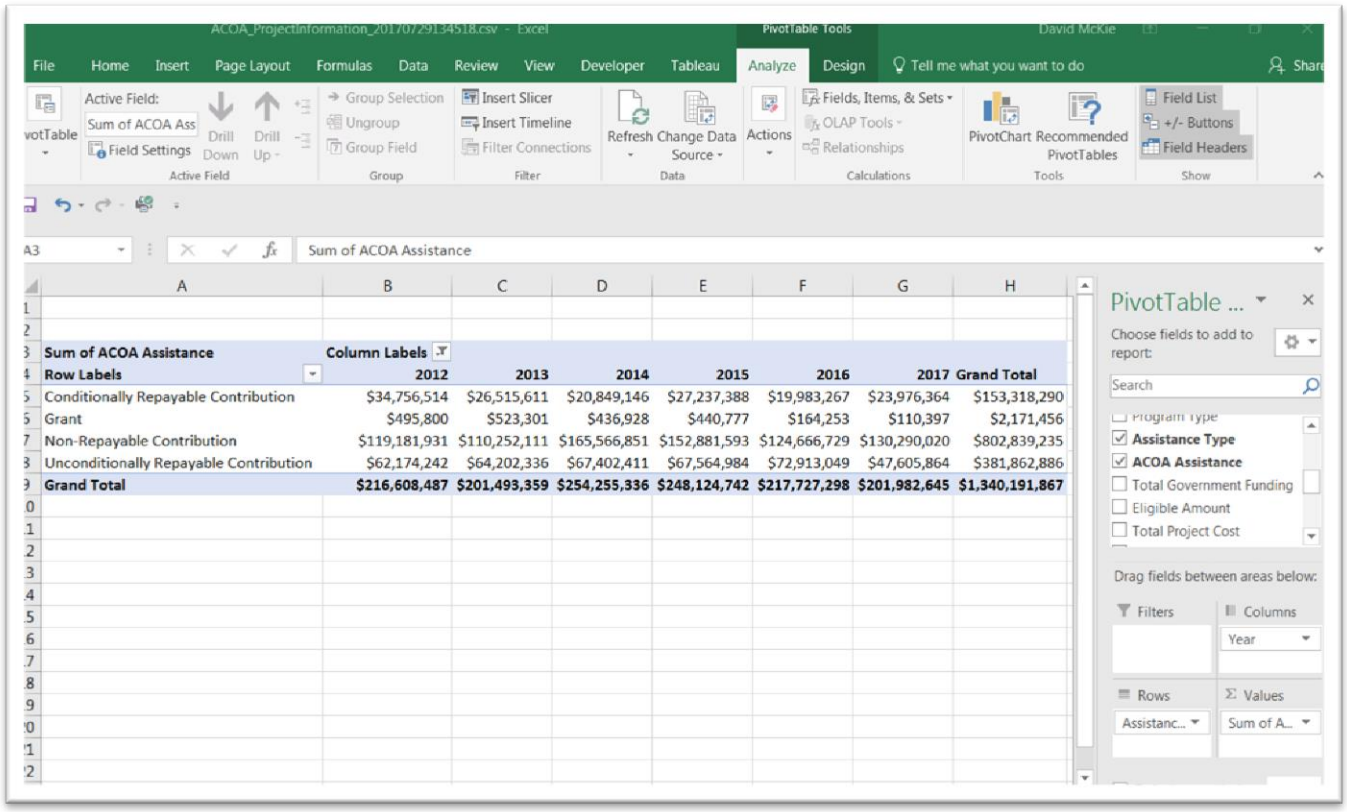


If this was all we were interested in, we could simply copy and use the paste-special option that we learned in the similarly named tutorial that goes with this chapter, and calculate the percent changes. However, from simply eyeballing the numbers, there doesn't seem to be anything worth pursuing.

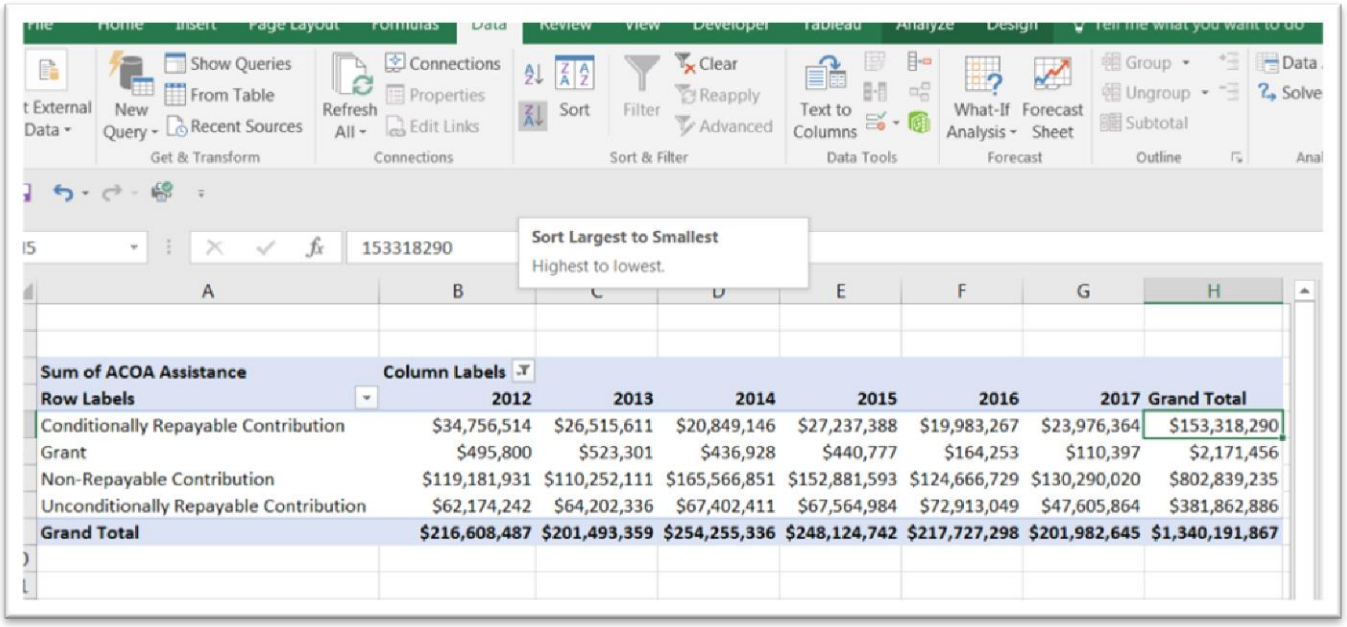
Let's keep building the pivot table in order to take a deeper dive into the data.

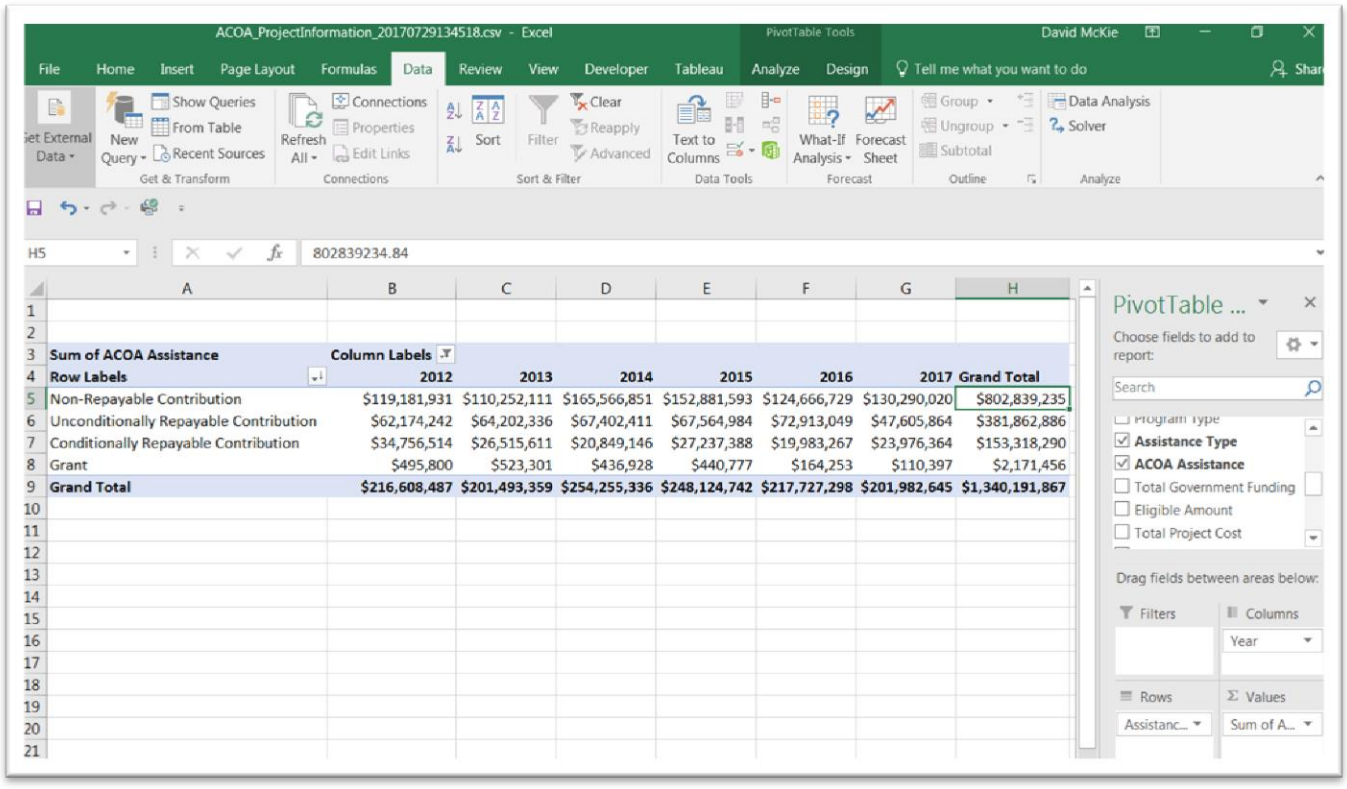
The "Row Labels" section should contain the column on which you want to do your analysis. Put another way, the column you care the most about.

Select "Assistance Type".

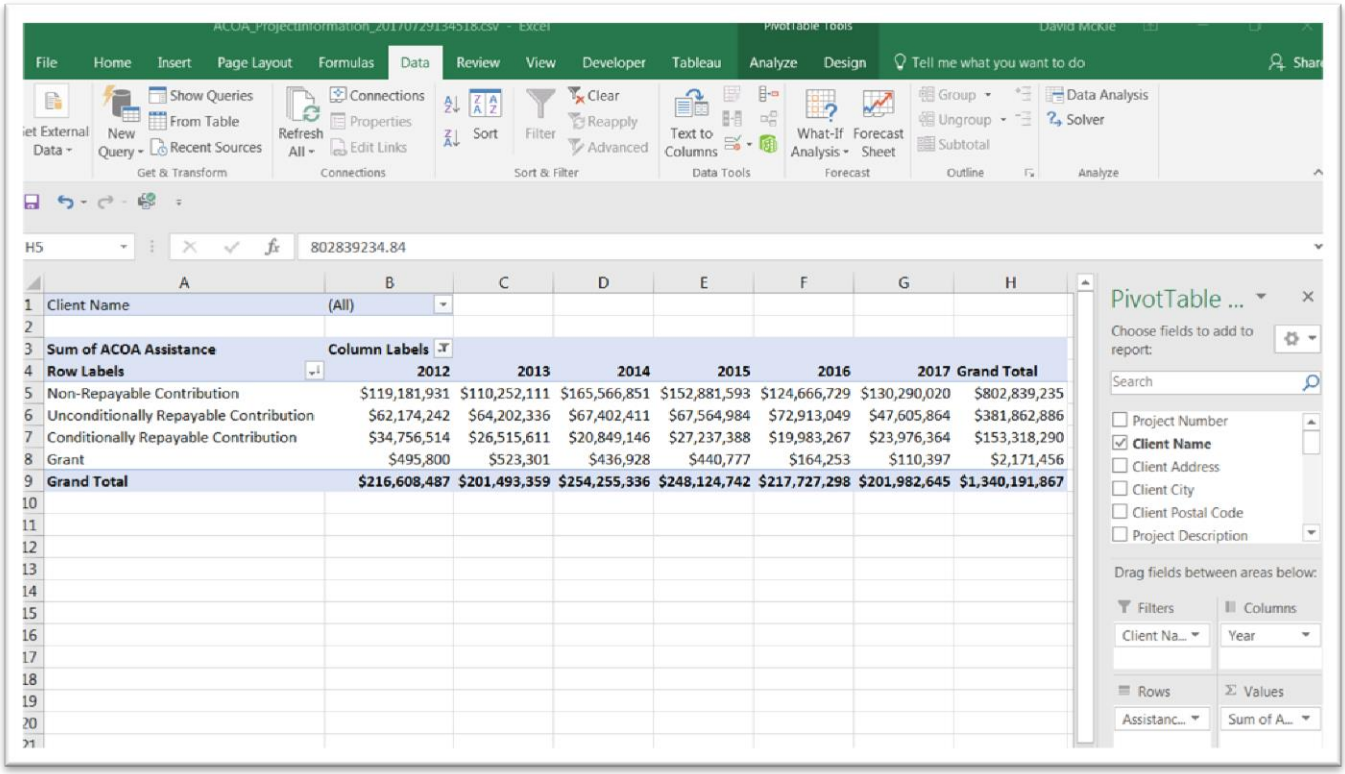


Now we can see the categories of financial assistance. Because as journalists we like seeing who got the most, sort the "Grand Total" section in descending order.





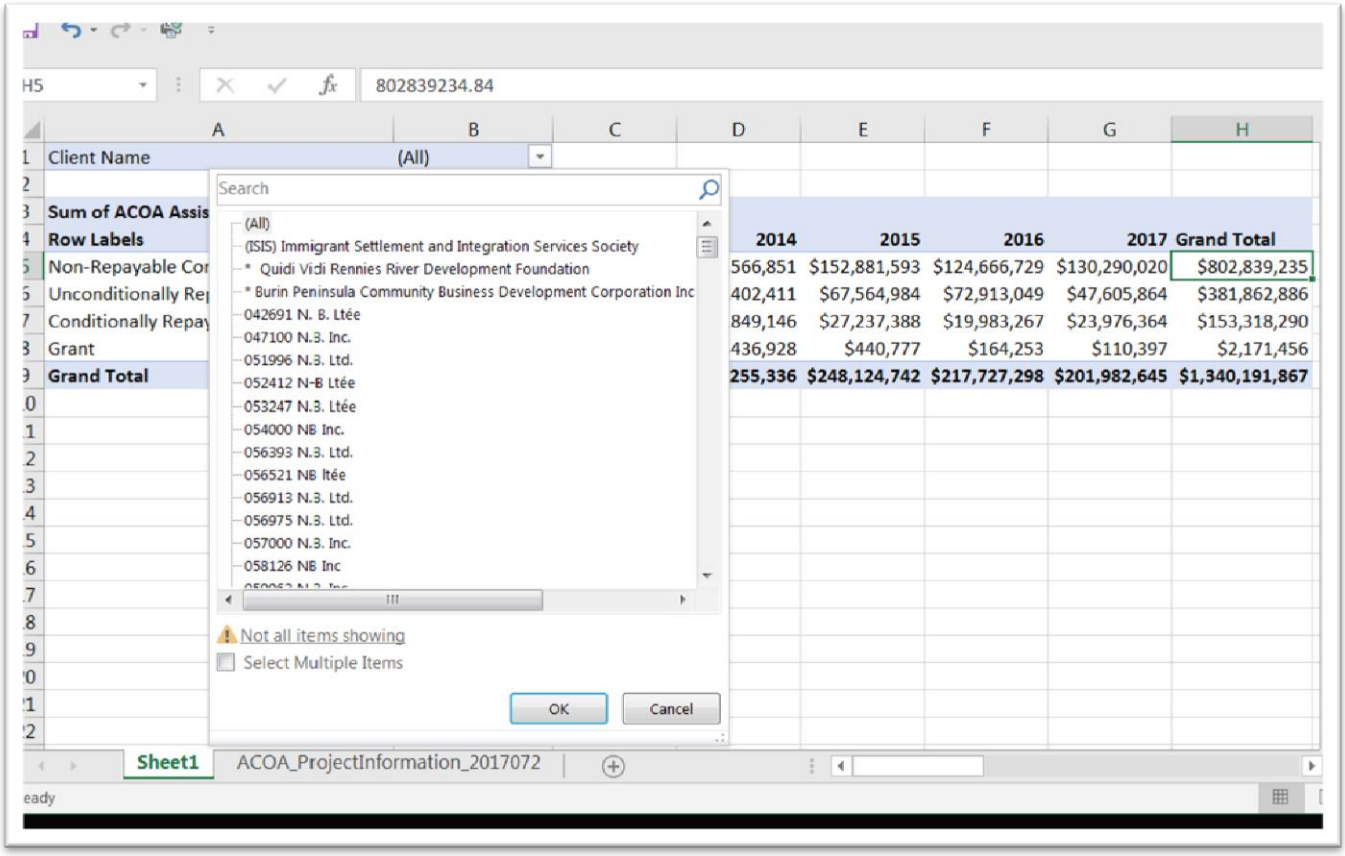
The "Non-Repayable Contribution" category takes top spot. We can also see how the other types behave year after year. Be sure to consult the data dictionary for a definition of each category.



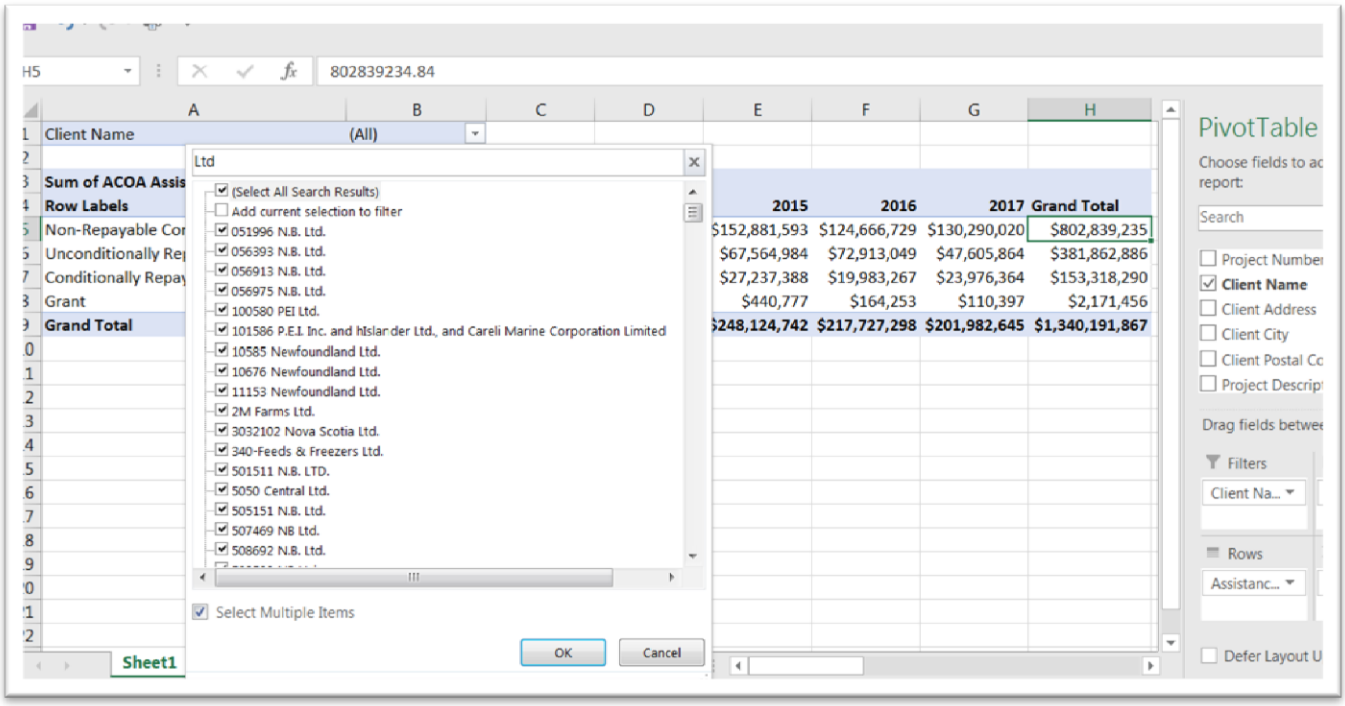
Sum of ACOA Assistance	Column Labels						
Row Labels	2012	2013	2014	2015	2016	2017	Grand Total
Non-Repayable Contribution	$119,181,931	$110,252,111	$165,566,851	$152,881,593	$124,666,729	$130,290,020	$802,839,235
Unconditionally Repayable Contribution	$62,174,242	$64,202,336	$67,402,411	$67,564,984	$72,913,049	$47,605,864	$381,862,886
Conditionally Repayable Contribution	$34,756,514	$26,515,611	$20,849,146	$27,237,388	$19,983,267	$23,976,364	$153,318,290
Grant	$495,800	$523,301	$436,928	$440,777	$164,253	$110,397	$2,171,456
Grand Total	**$216,608,487**	**$201,493,359**	**$254,255,336**	**$248,124,742**	**$217,727,298**	**$201,982,645**	**$1,340,191,867**

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.

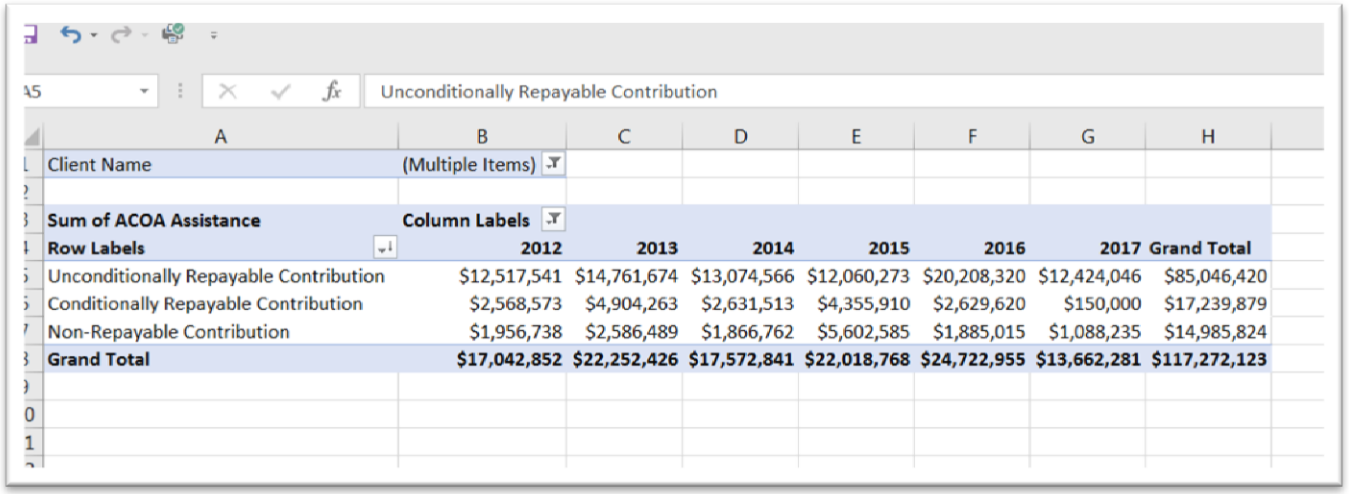
Click on the arrow to obtain the drop-down menu. From perusing the categories, it's obvious that many businesses receive assistance. A number of the business names end with "Ltd."



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.



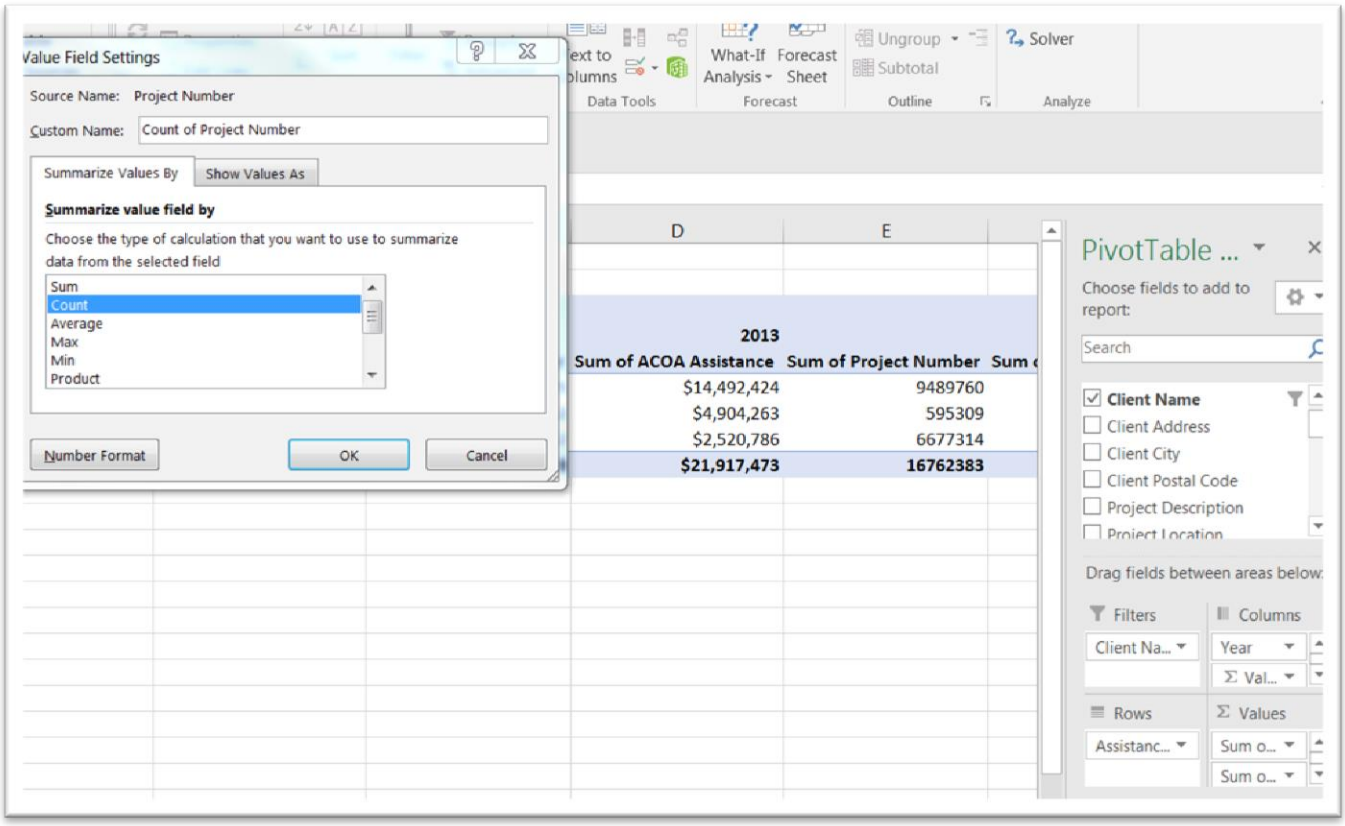
Select "Ltd.", the acronym for the designation limited. Once you've made the selection, in this case the companies with the "Ltd." suffix, select the OK tab.

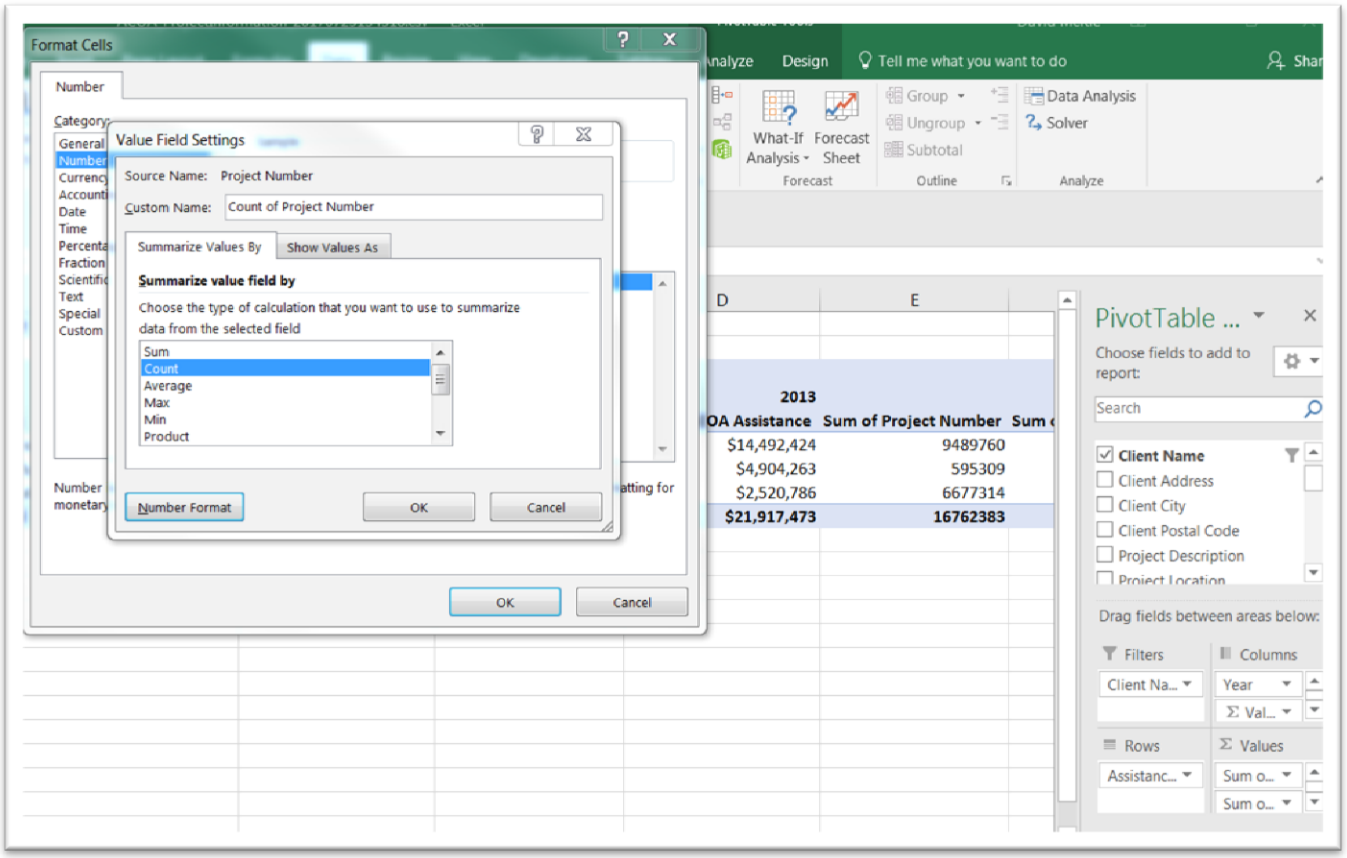


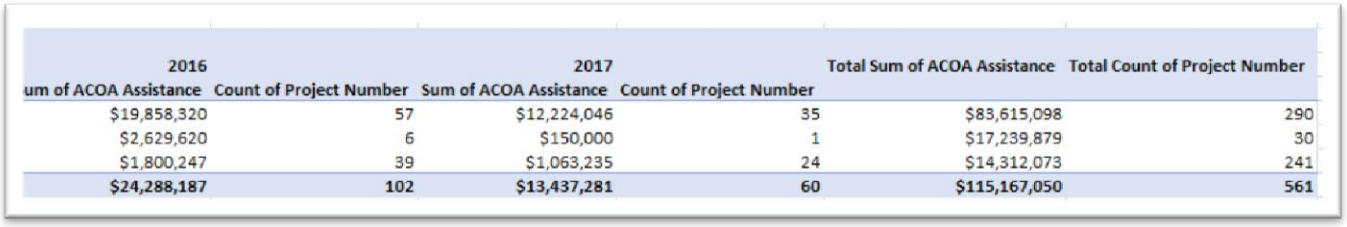
ACOA hands out most of its money to businesses with the "Ltd" suffix as in the form of unconditionally repayable contributions. If you're happy with this result, you can give the pivot table a specific name, and return to your original table to build another one. Or if you wanted to perform some math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







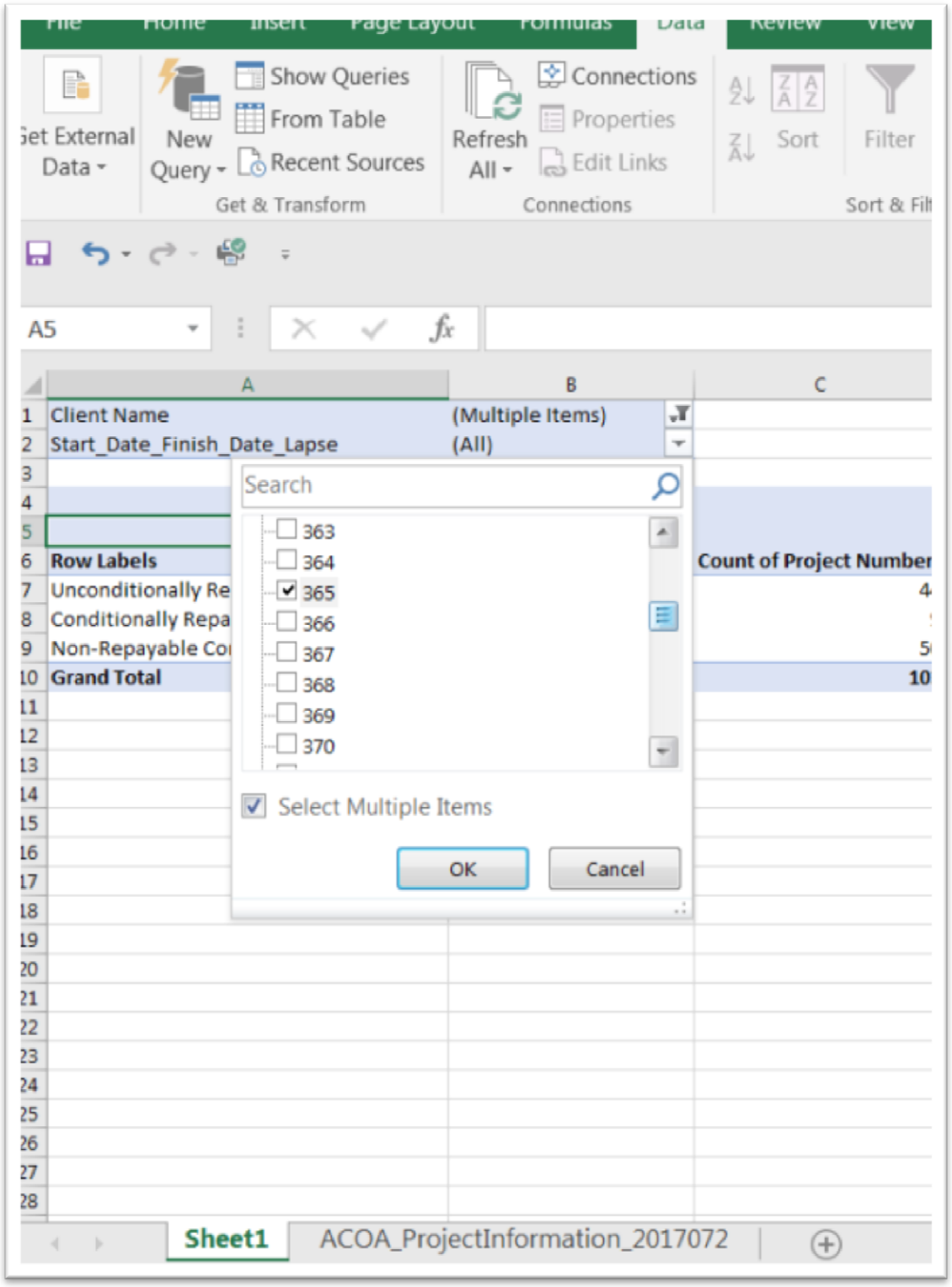
2016		2017		Total Sum of ACOA Assistance	Total Count of Project Number
um of ACOA Assistance	Count of Project Number	Sum of ACOA Assistance	Count of Project Number		
$19,858,320	57	$12,224,046	35	$83,615,098	290
$2,629,620	6	$150,000	1	$17,239,879	30
$1,800,247	39	$1,063,235	24	$14,312,073	241
$24,288,187	**102**	**$13,437,281**	**60**	**$115,167,050**	**561**

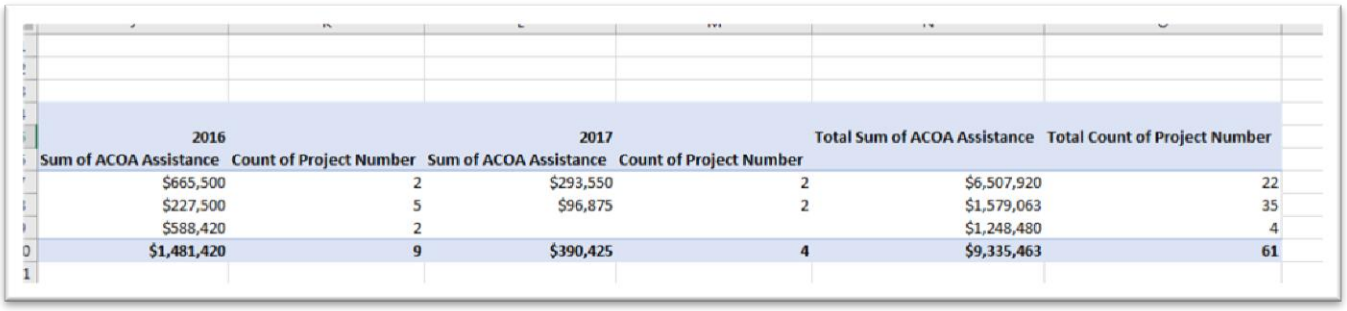
To the right of each year, and the grand total, the pivot table has added a column that COUNTS the number of recipients. In 2016, 57 corporations received $19.8 million in unconditional repayable contributions.

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.


File
Home
Insert
Page Layout
Formulas
Data
Review
View
Get External Data
New Query
Show Queries
From Table
Recent Sources
Get & Transform
Refresh All
Connections
Properties
Edit Links
Connections
Sort
Filter
Sort & Filt
A5
A
B
C
1 Client Name (Multiple Items)
2 Start_Date_Finish_Date_Lapse (All)
3
4
5
6 Row Labels Count of Project Number
7 Unconditionally Re 4
8 Conditionally Repa
9 Non-Repayable Co 5
10 Grand Total 10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
Search
363
364
365
366
367
368
369
370
Select Multiple Items
OK
Cancel
Sheet1
ACOA_ProjectInformation_2017072



2016		2017		Total Sum of ACOA Assistance	Total Count of Project Number
Sum of ACOA Assistance	Count of Project Number	Sum of ACOA Assistance	Count of Project Number		
$665,500	2	$293,550	2	$6,507,920	22
$227,500	5	$96,875	2	$1,579,063	35
$588,420	2			$1,248,480	4
$1,481,420	**9**	**$390,425**	**4**	**$9,335,463**	**61**

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 worthwhile 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.