

Using and Formatting Pivot Tables

Creating custom estimates with ACS PUMS in Excel

Pivot Table Basics

A pivot table is a tool that helps you quickly analyze “long” formatted data based on various fields, or categories, in the data. “Long” data are those where each observation gets its own row and each category of the data gets its own column. By arranging your data in this format first, Excel is able to quickly filter the observations you want to analyze, and then perform some computation on that subset of data. The power of pivot tables is that they can filter and summarize data over as many categories as you want simultaneously, displaying the results in a tidy summary table. Moreover, should you want to split your data along a different category, you can quickly swap in new filters.

Long Versus Wide Data

At first, it can be difficult to spot whether your data is indeed in long format, but understanding that a pivot table wants to be able to filter your data helps to build the intuition around how to structure your input table. A basic guiding rule is that your table will almost always be taller (longer) than it is wide if your data is set up correctly. This naturally occurs when your data has a column for each measure and a row for each observation, as we tend to have more observations than measures. Unfortunately, many data come in wide format because this is more compact. For example, the data below from American FactFinder table S1701 are still in wide format. You can tell because each county has its own set of columns, rather than its own single row.

Subject	Allegan County, Michigan					Bay County, Michigan				
	Total	Estimate	Margin of	Below poverty level	Percent below poverty	Total	Estimate	Margin of	Below poverty level	Percent below poverty
Population for whom poverty status is	113,608	+/-463	12,901	+/-2,781	11.4%	103,844	+/-845	15,196	+/-2,259	14.6%
AGE										
Under 18 years	28,445	+/-686	4,408	+/-1,702	15.5%	21,218	+/-740	4,270	+/-1,094	20.1%
Under 5 years	6,382	+/-320	1,181	+/-693	18.5%	5,066	+/-419	1,000	+/-577	19.7%
5 to 17 years	22,063	+/-670	3,227	+/-1,216	14.6%	16,152	+/-816	3,270	+/-944	20.2%
18 to 64 years	67,803	+/-716	7,076	+/-1,302	10.4%	63,638	+/-633	9,404	+/-1,581	14.8%
18 to 34 years	22,148	+/-666	2,673	+/-865	12.1%	21,172	+/-741	3,908	+/-1,226	18.5%
35 to 64 years	45,655	+/-551	4,403	+/-958	9.6%	42,466	+/-711	5,496	+/-1,071	12.9%
60 years and over	24,916	+/-936	2,306	+/-721	9.3%	27,286	+/-1,040	2,185	+/-608	8.0%
65 years and over	17,360	+/-469	1,417	+/-466	8.2%	18,988	+/-537	1,522	+/-470	8.0%

In long format, this table would look like this:

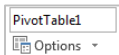
Age	State	County	Poverty Measure	Estimate	Margin of Error
Under 18 years	Michigan	Allegan	Total	28,445	+/-686
Under 5 years	Michigan	Allegan	Total	6,382	+/-320
5 to 17 years	Michigan	Allegan	Total	22,063	+/-670
18 to 64 years	Michigan	Allegan	Total	67,803	+/-716
18 to 34 years	Michigan	Allegan	Total	22,148	+/-666
35 to 64 years	Michigan	Allegan	Total	45,655	+/-551
60 years and over	Michigan	Allegan	Total	24,916	+/-936
65 years and over	Michigan	Allegan	Total	17,360	+/-469
Under 18 years	Michigan	Allegan	Below poverty	4,408	+/-1,702
Under 5 years	Michigan	Allegan	Below poverty	1,181	+/-693
5 to 17 years	Michigan	Allegan	Below poverty	3,227	+/-1,216
18 to 64 years	Michigan	Allegan	Below poverty	7,076	+/-1,302
18 to 34 years	Michigan	Allegan	Below poverty	2,673	+/-865
35 to 64 years	Michigan	Allegan	Below poverty	4,403	+/-958
60 years and over	Michigan	Allegan	Below poverty	2,306	+/-721
65 years and over	Michigan	Allegan	Below poverty	1,417	+/-466
Under 18 years	Michigan	Allegan	Percent below	16	+/-6.0
Under 5 years	Michigan	Allegan	Percent below	19	+/-10.7
5 to 17 years	Michigan	Allegan	Percent below	15	+/-5.5
18 to 64 years	Michigan	Allegan	Percent below	10	+/-1.9
18 to 34 years	Michigan	Allegan	Percent below	12	+/-3.8
35 to 64 years	Michigan	Allegan	Percent below	10	+/-2.1
60 years and over	Michigan	Allegan	Percent below	9	+/-2.8
65 years and over	Michigan	Allegan	Percent below	8	+/-2.7

Table continued...

Under 18 years	Michigan	Bay	Total	21,218	+/-740
Under 5 years	Michigan	Bay	Total	5,066	+/-419
5 to 17 years	Michigan	Bay	Total	16,152	+/-816
18 to 64 years	Michigan	Bay	Total	63,638	+/-633
18 to 34 years	Michigan	Bay	Total	21,172	+/-741
35 to 64 years	Michigan	Bay	Total	42,466	+/-711
60 years and over	Michigan	Bay	Total	27,286	+/-1,040
65 years and over	Michigan	Bay	Total	18,988	+/-537
Under 18 years	Michigan	Bay	Below poverty	4,270	+/-1,094
Under 5 years	Michigan	Bay	Below poverty	1,000	+/-577
5 to 17 years	Michigan	Bay	Below poverty	3,270	+/-944
18 to 64 years	Michigan	Bay	Below poverty	9,404	+/-1,581
18 to 34 years	Michigan	Bay	Below poverty	3,908	+/-1,226
35 to 64 years	Michigan	Bay	Below poverty	5,496	+/-1,071
60 years and over	Michigan	Bay	Below poverty	2,185	+/-608
65 years and over	Michigan	Bay	Below poverty	1,522	+/-470
Under 18 years	Michigan	Bay	Percent below	20	+/-5.0
Under 5 years	Michigan	Bay	Percent below	20	+/-10.5
5 to 17 years	Michigan	Bay	Percent below	20	+/-5.5
18 to 64 years	Michigan	Bay	Percent below	15	+/-2.5
18 to 34 years	Michigan	Bay	Percent below	19	+/-5.8
35 to 64 years	Michigan	Bay	Percent below	13	+/-2.6
60 years and over	Michigan	Bay	Percent below	8	+/-2.2
65 years and over	Michigan	Bay	Percent below	8	+/-2.5

Other Pivot Table Tips

While properly formatting your input table is the most important step in working with pivot tables, here are some other tips to keep in mind:

- Data should all fit within a rectangle, with no blank columns or rows.
 - Blank columns will mark the end of the table.
 - Blank rows can cause calculation errors.
- The first row of the data should have the column or field names. Columns without a name will cause an error.
- Unhide rows and columns to make sure the above two rules are not violated.
- Remember, you can apply sum, average, count, max, min, and other functions to your data.
- You can add custom calculations as a new field if needed.
- If referencing cells in the pivot table, manually typing in their cell address allows for easier copying and pasting.
- Pivot table fields can be renamed by editing them in the formula bar.
- Explore the pivot table options drop down menu under the “ANALYZE” tab:  .

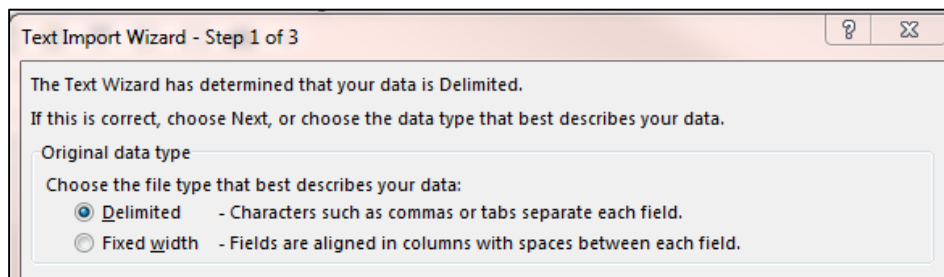
Importing Data from DataFerrett

Warning: ACS PUMS files can be very large, and analyzing them in Excel can cause your PC to run slowly or freeze. You might wish to save all other work you are doing before proceeding and save frequently during this tutorial.

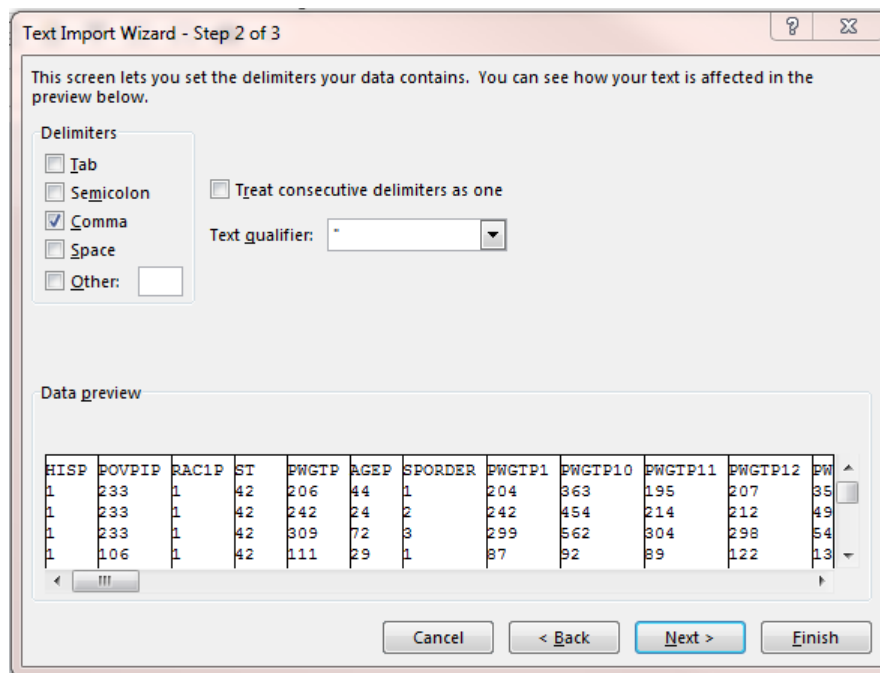
The starting point for this tutorial assumes that you have used DataFerrett to extract several variables from the 1-year 2015 ACS PUMS population file for the state of Pennsylvania. DataFerrett’s download feature will allow you to save the microdata as a comma separated values file, though the file extension is .acs by default. To open these data in Excel:

- Unzip your file if needed.
- Launch Excel. From the main window, click the open file icon (📁) to launch a dialog.
- Select the DataFerrett file, which should be in .acs format, and click open.

- The Text Import Wizard should launch automatically. In “Step 1,” select “Delimited” and click Next.



- In “Step 2,” select the box next to “Comma” and deselect any other box that might be checked. Click Finish to complete the import process.



Organize Data and Recode Variables

ACS PUMS data comes largely ready to use out of the box. If your data is not already in long format, the fastest way to transform it is often to go back to the source and see if the data can be output into a long format. If that is not an option, using Excel VBA is often the next best approach. Both of those are outside the scope of this tutorial.

Organizing Data

The extract we are working from should import 107 variables, in the same order as they were added in DataFerrett. To start off on the same foot, we will move our important columns towards the front of the sheet, then hide the rest of the columns as we recode the data. Reorder your data so that the following columns appear in these positions.

Column A	Column B	Column C	Column D	Column E	Column F
SERIALNO	SPORDER	RELP	AGE	WKW	WKHP

Trick: To move a column quickly:

- Click on the column's header to select the whole column at once.
- Position the mouse's pointer on the column's green outline and then left-click and hold to move the column.
- While still holding the mouse button, press and hold the shift key and then drag your column to where you want to insert it by hovering over the column divider between two columns.
- When the cursor changes to a long green insertion icon, release the mouse button to complete the insertion. Repeat as necessary.

Recoding Work Experience

Analysts will likely want to add a few variables of their own before creating a pivot table. These new variables will simply be recodes of existing variables, but will result in easier use and display of data in pivot tables.

One common measure that analysts are interested in is the work experience of a person. Census often describes work experience as “full-time, year round,” “part-time,” or “did not work.” We will create this measure by recoding several variables—WKW and WKHP—into a new variable called “WorkExperience.” First, though, we will create a variable called “WorkingAge” based off the AGEP variable.

WorkingAge

- In the first empty column, create a new variable name by entering “WorkingAge” in the top row.
- In the cell below that, enter the following formula: `=AND(D2>=25,D2<=64)`, where D is the column that contains the variable AGEP.
- With that cell still highlighted, either drag the formula to the bottom of the data table or double-click on the cell handle (the green box in the lower right corner of the selected cell) to do this automatically.

X ✓ <i>fx</i> <code>=AND(D2>=25,D2<=64)</code>			
D	E	F	DD
AGEP	WKW	WKHP	WorkingAge
44	4	35	TRUE
24	0	0	FALSE

The result will be boolean value of TRUE or FALSE, depending on the age of the person. Booleans are dichotomous (a.k.a. binary, indicator, or dummy) variables that take on the values 0 and 1. When Excel performs a comparison like we did in cell DD2, the result is shown as TRUE or FALSE, but these represent the implicit values of 1 and 0, respectively.

WorkExperience

- In the next column, enter the variable name “WorkingExperience” in the topmost cell, which should be DE1.
- In the cell below that, enter the following formula:
=IF(AND(E2=1,F2>=35),"FullTime",IF(E2>0,"PartTime","NotWorking")), where E is the column that contains the variable WKW and F is the column that contains the variable WKHP.
- Copy the formula down the column.

DE2				✕ ✓ fx		=IF(AND(E2=1,F2>=35),"FullTime",IF(E2>0,"PartTime","NotWorking"))					
	A	B	C	D	E	F	DD	DE	DF	DG	DH
1	SERIALNO	SPORDER	REL	AGEP	WKW	WKHP	WorkingAge	WorkExperience			
2	31	1	0	44	4	35	TRUE	PartTime			
3	31	2	2	24	0	0	FALSE	NotWorking			

Organizing Data Again


Now that you’ve added your two new variables, move their columns along with several other variables so that they are in the order listed below. As we will see shortly, this makes finding these fields easier in the pivot table window. As you do this, unhide all columns and ensure there are no blank columns in the data.

Column G	Column H	Column I	Column J	Column K	Column L	Column M
WorkingAge	WorkExperience	DIS	RECODE1	RECODE2	RECODE3	PWGTP

Making the Pivot Table

Now that your data is organized, you are ready to select the data, insert the pivot table, and populate the fields.

Selecting the data

- Select all of the data: Move to the top-left corner of the data and then select all by typing Ctrl + A.
Note: Using this keyboard shortcut selects all contiguous data. If you have an empty row or column only part of the data will be selected. Now is a good time to check that there are no empty rows or columns and all of your data is selected.
- With your data selected, click on the INSERT tab in the ribbon.
- Click the PivotTable icon () in the left-hand side of the ribbon.
- From the popup dialog, click on the OK button to accept the default choices.

Parts of a PivotTable

Excel will open a new sheet and place an empty PivotTable in the top-left corner. On the right side of the screen, Excel opens the PivotTable's Field List, which consists of the columns—or variables as we have them organized—in the order that they appear in our data. Each of our fields can be used as one of the four main parts of a pivot table: filters, columns, rows, or values.

- Filters are used to determine which data is used in your analysis. For example, you might filter out all persons with AGEP greater than 17 to focus solely on children.
- Columns here are not like columns in the raw data, but are fields (variables) that you wish to break out horizontally.
- Rows, like columns, are again fields (variables) that you want to break out, but display vertically.
- Values, finally, are the numbers which you want to summarize by the fields you have chosen in the three other parts of the pivot table.

Designing a PivotTable of the Working Poor

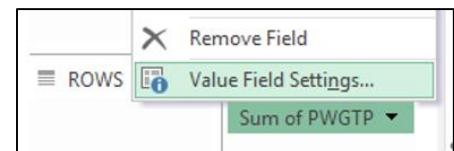
We will build our pivot table piece by piece, starting off with the values. Typically, one might start elsewhere, but by placing the values in first we can watch our pivot table update as we select other fields, giving you a sense of how powerful and useful pivot tables can be. In the Field List on the right-hand side:

- First, drag the field “PWGTP” to the VALUES area. Upon releasing, the pivot table should update and the only value should be the total sum of the variable PWGTP. This is the person weight variable, which is the number of people each observation represents in the full population, and by summing it, we are calculating the total population of the Pennsylvania.

	A
1	
2	
3	Sum of PWGTP
4	12802503

- *Trick:* By default, pivot tables will sum up the fields in the values area. This can be changed in several ways, but the quickest way is to:

- Click on the drop down area next to the field in the VALUES area.
- Then click on the option Value Field Settings.
- Then, select one of the options from the list.



- Next, click and drag the field WorkingAge to the FILTERS area. In the pivot table, a new filter automatically shows up in the top left-hand corner of the table. Note that value(s) remain the same in the pivot table because we have simply chosen a field upon which to filter, but we have not yet applied the filter itself.
- Apply the filter by clicking on the dropdown arrow next to (All) and select the value “TRUE”. Watch how the total drops, with the result being the total number of working age adults in Pennsylvania.

- Add another filter using the field DIS and apply a filter so that only the value 2 is showing. (2 is the value for “Not Disabled” in the ACS PUMS data.) Watch how the total drops once again.

The pivot table now shows the total count of non-disabled working adults.

Next, we’ll add the work experience and poverty status of this sub-population.

	A	B
1	WorkingAge	TRUE
2	DIS	2
3		
4	Sum of PWGTP	
5	5861727	

- Drag the field WorkExperience to the ROWS area of the Field List window.
- Drag the field RECODE1 to the COLUMNS area of the Field List window. RECODE1 is a recode of the POVPIP variable down in DataFerrett. Those with a POVPIP value between 0 and 99 (i.e., below the poverty level) were recoded to the value 1, and those with a POVPIP value between 100 and 501 were recoded to the value 2. There’s a third category of POVPIP values possible, which is the value of -1 in DataFerrett, which stands for missing. These observations do not have an income-to-poverty-level ratio because they are outside of the poverty universe.

The result of our pivot table design is a table that shows the work experience and poverty status of working age adults without a disability in the state of Pennsylvania. Before moving on to formatting, we’ll add one more element of data, the poverty rate.

- As done previously, drag and drop the field PWGTP to the VALUES window for a second time. As one my might expect, this will duplicate the values that already exist because PWGTP is repeated in the values area.
- To keep track of this repeated field, Excel adds a “Values” box to the COLUMN area in the Field List window. Move this box from the COLUMNS area to the ROWS area by dragging and dropping it. Watch how the pivot table quickly reorganizes itself.

Our final step is to change how this extra field is calculated.

- Click on drop down area of the second PWGTP field in the VALUES area.
- From the pop-up menu, click “Value Field Settings.”
- In the pop-up window, click on the “Show Values As” tab.
- In the drop down list, select the option “% of Row Total.” Click OK.

	A	B	C	D
1	WorkingAge	TRUE		
2	DIS	2		
3				
4	Column Labels			
5	Row Labels	1	2	Grand Total
6	FullTime			
7	Sum of PWGTP	3460704	194236	3654940
8	Sum of PWGTP2	94.69%	5.31%	100.00%
9	NotWorking			
10	Sum of PWGTP	823306	68692	891998
11	Sum of PWGTP2	92.30%	7.70%	100.00%
12	PartTime			
13	Sum of PWGTP	1234270	80519	1314789
14	Sum of PWGTP2	93.88%	6.12%	100.00%
15	Total Sum of PWGTP	5518280	343447	5861727
16	Total Sum of PWGTP2	94.14%	5.86%	100.00%

Formatting a Pivot Table

Once you have decided that your pivot table analysis is complete, you might want to improve its look by changing the formatting of its elements or alter the labeling so that the table reads more intuitively. Here are a few improvements to this table that demonstrate common formatting tasks.

Values Formatting

To show the number of people using Excel's comma formatting:

- Return to the PivotTable Fields window and click on the drop down arrow next to the first PWGTP field in the VALUES area.
- From the pop-up window, click "Value Field Settings."
- Click on the Number Format button in the lower right-hand corner in the pop-up window.
- In the new pop-up window, change the number Category by clicking on the "Number" option in the list.
- In the context window, next to the words "Decimal places:" click on the down arrow until the window shows 0.
- In the context window, select the box next to "Use 1000 Separator(,)."
- Click OK to exit the Format Cells window.
- Click OK to exit the Value Field Settings window.

Table Labels

Almost all labels in the pivot table can be changed manually by clicking on the cell and then typing the new label into the Formula Bar. For example:

- Click on cell A2 and type in the word "Disabled"
- Click on cell A5 and type in the words "Work Experience"
- Click on cell B4 and type in the words "Poverty Status"
- Click on cell D5 and type in the word "Total"
- Click on cell B5 and type in the words "Below Poverty"
- Click on cell C5 and type in the words "Above Poverty"
- To change the labels in cells A7, A10, and A13 simultaneously:
 - Return to the PivotTable Fields window and click on the drop down arrow next to the first PWGTP field in the VALUES area.
 - From the pop-up menu, click "Value Field Settings."
 - Next to the field "Custom Name:" type Count, and then click OK.
- To change the labels in cells A8, A11, and A14 simultaneously:

- Return to the PivotTable Fields window and click on the drop down arrow next to the second PWGTP field in the VALUES area.
- From the pop-up window, click “Value Field Settings.”
- Next to the field “Custom Name:” type Poverty Rate, and then click OK.

	A	B	C	D
1	WorkingAge	TRUE		
2	Disabled	2		
3				
4	Poverty Status			
5	Work Experience	Below Poverty	Above Poverty	Total
6	FullTime			
7	Count	3,460,704	194,236	3,654,940
8	Poverty Rate	94.69%	5.31%	100.00%
9	NotWorking			
10	Count	823,306	68,692	891,998
11	Poverty Rate	92.30%	7.70%	100.00%
12	PartTime			
13	Count	1,234,270	80,519	1,314,789
14	Poverty Rate	93.88%	6.12%	100.00%
15	Total Count	5,518,280	343,447	5,861,727
16	Total Poverty Rate	94.14%	5.86%	100.00%

Pivot Table Options

Beyond appearance and formatting, pivot tables have other options that can be modified by accessing the Options menu in the bottom left-hand side of the Ribbon in the Pivot Tables Design tab. These options are worth exploring and control things such as how often data is refreshed, whether formatting is reset when the table is modified, and the ability to use a custom list to sort the results.

Vincent Palacios

Center on Budget and Policy Priorities
 palacios@cbpp.org
 202.408.1080
 www.cbpp.org