Working with Survey of Construction Microdata in a Microsoft Excel PivotTable

Introduction

What is a Microdata file?

Generally speaking, a microdata file is a file that contains a subset of actual records from a given survey (in this case, the Survey of Construction). These files are made available so that the public can create their own custom tabulations of the data, allowing a far greater level of detail than our published tables. By law (Title 13, USC), all identifying information must be removed from these files to protect the respondent's privacy.

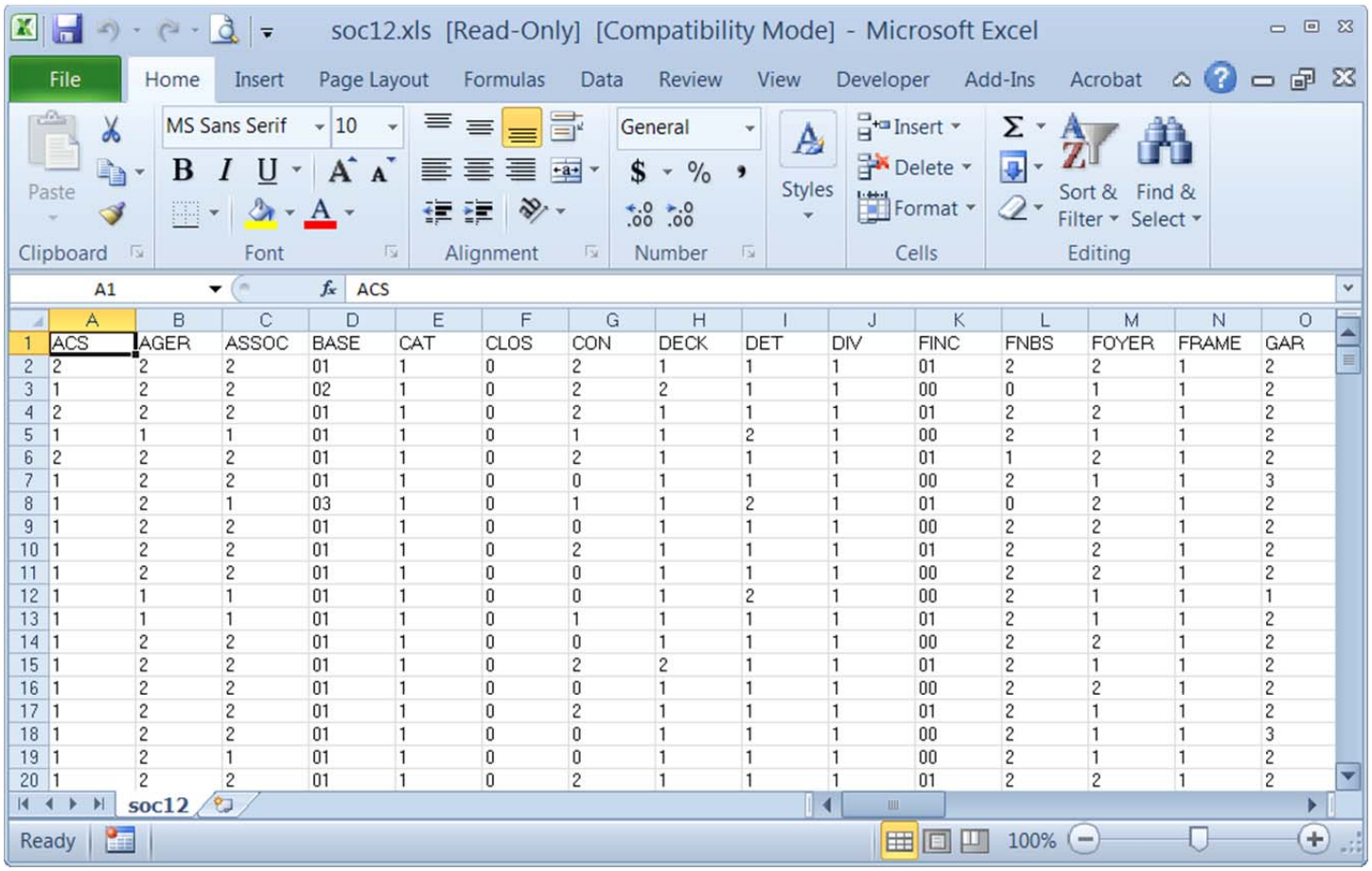
What is a PivotTable?

By far, the easiest and most efficient way to use the microdata file is by utilizing an Excel feature called a "PivotTable". This is an interactive feature that allows a person to summarize the data on their own, and create custom tables with the variables they want. A PivotTable also allows you to easily 'pivot' to change your summary to a different variable, or change the way in which the data is displayed. PivotTables are particularly useful in summarizing large amounts of data (such as this microdata file).

Note that this guide is specifically for Microsoft Excel PivotTables. Other spreadsheet software may or may not have a similar feature. Also be advised that this guide only scratches the surface of PivotTable capabilities. The examples contained in this guide were made to be relatively simple and easy to follow. It is possible to create tables that are much more complex than the ones contained in this document.

Note: For the purposes of this guide, the 2012 file was used. If you are using the examples contained this guide as a walkthrough, it is advised you use the 2012 file so that your results match the illustrations.

Basic File Structure



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ACS	AGER	ASSOC	BASE	CAT	CLOS	CON	DECK	DET	DIV	FINC	FNBS	FOYER	FRAME	GAR
2	2	2	2	01	1	0	2	1	1	1	01	2	2	1	2
3	1	2	2	02	1	0	2	2	1	1	00	0	1	1	2
4	2	2	2	01	1	0	2	1	1	1	01	2	2	1	2
5	1	1	1	01	1	0	1	1	2	1	00	2	1	1	2
6	2	2	2	01	1	0	2	1	1	1	01	1	2	1	2
7	1	2	2	01	1	0	0	1	1	1	00	2	1	1	3
8	1	2	1	03	1	0	1	1	2	1	01	0	2	1	2
9	1	2	2	01	1	0	0	1	1	1	00	2	2	1	2
10	1	2	2	01	1	0	2	1	1	1	01	2	2	1	2
11	1	2	2	01	1	0	0	1	1	1	00	2	2	1	2
12	1	1	1	01	1	0	0	1	2	1	00	2	1	1	1
13	1	1	1	01	1	0	1	1	1	1	01	2	1	1	2
14	1	2	2	01	1	0	0	1	1	1	00	2	2	1	2
15	1	2	2	01	1	0	2	2	1	1	01	2	1	1	2
16	1	2	2	01	1	0	0	1	1	1	00	2	2	1	2
17	1	2	2	01	1	0	2	1	1	1	01	2	1	1	2
18	1	2	2	01	1	0	0	1	1	1	00	2	1	1	3
19	1	2	1	01	1	0	0	1	1	1	00	2	1	1	2
20	1	2	2	01	1	0	2	1	1	1	01	2	2	1	2

The **header row** (row 1) shows all the characteristics we collect for each house in the SOC sample. Each variable is described in our microdata guide, available at http://www.census.gov/construction/chars/pdf/socmicro_info.pdf.

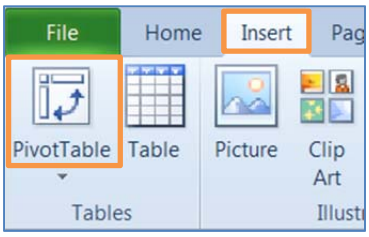
All **other rows** represent actual houses in the SOC sample for the given year, and selected characteristics of these houses.

All identifying information has been removed from these houses to protect the respondent's privacy. For this reason, the only geographic information we provide is the Census Division. Lower geographic levels (state, county, city, zip code, etc.) could possibly compromise a respondent's privacy, therefore we cannot release this information to the public.

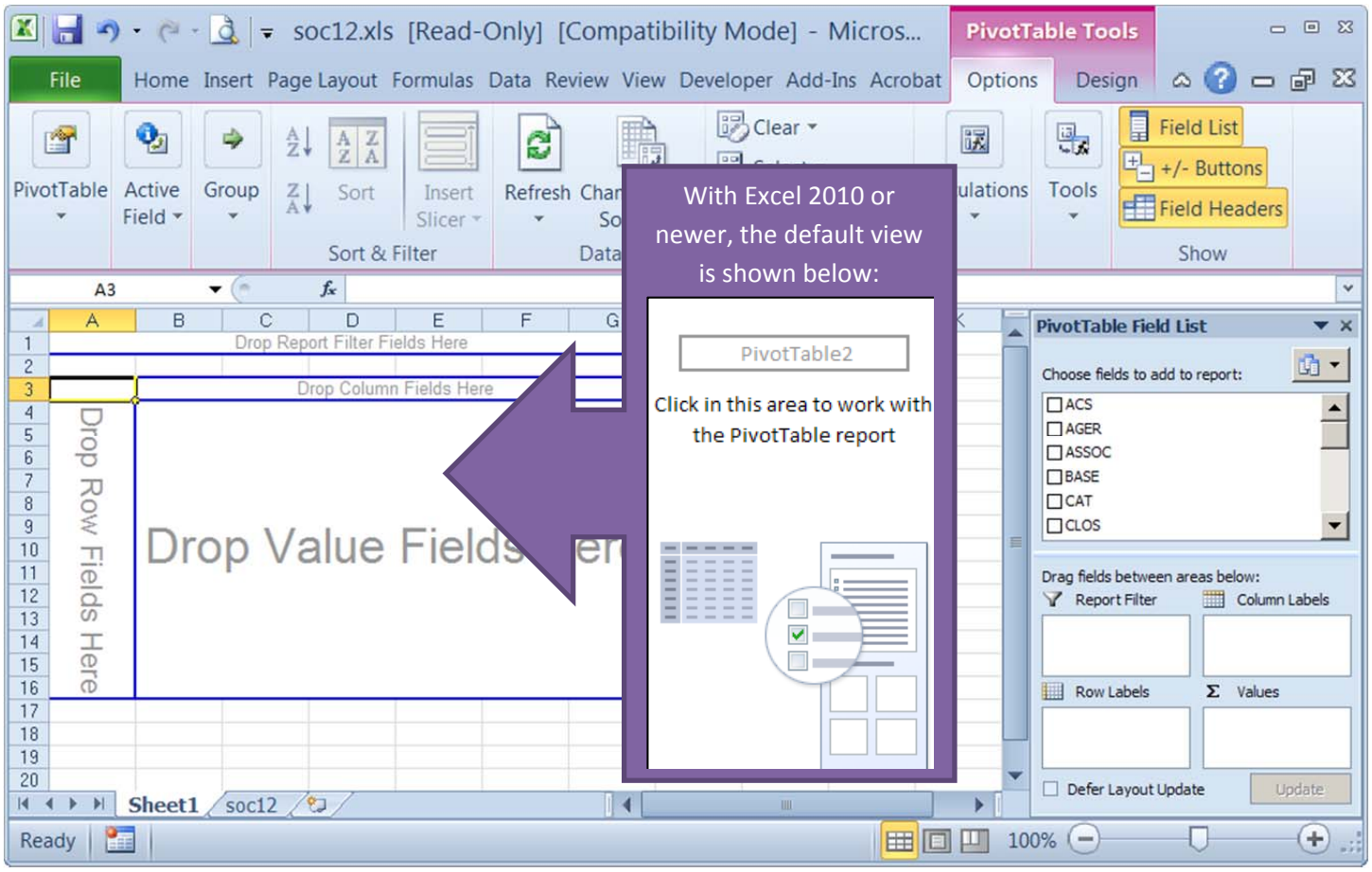
Note: Do not make changes to the original data. This is the source data for your PivotTable, and changing to it will be reflected in your PivotTable calculations. (If you want to make custom columns, it is advised you do it in the columns to the right of the existing data.)

Insert a PivotTable

From the Insert tab, click the PivotTable menu icon (Range and Worksheet automatically selected), then click "Ok"



This will open a new tab named "Sheet1" that looks like this:



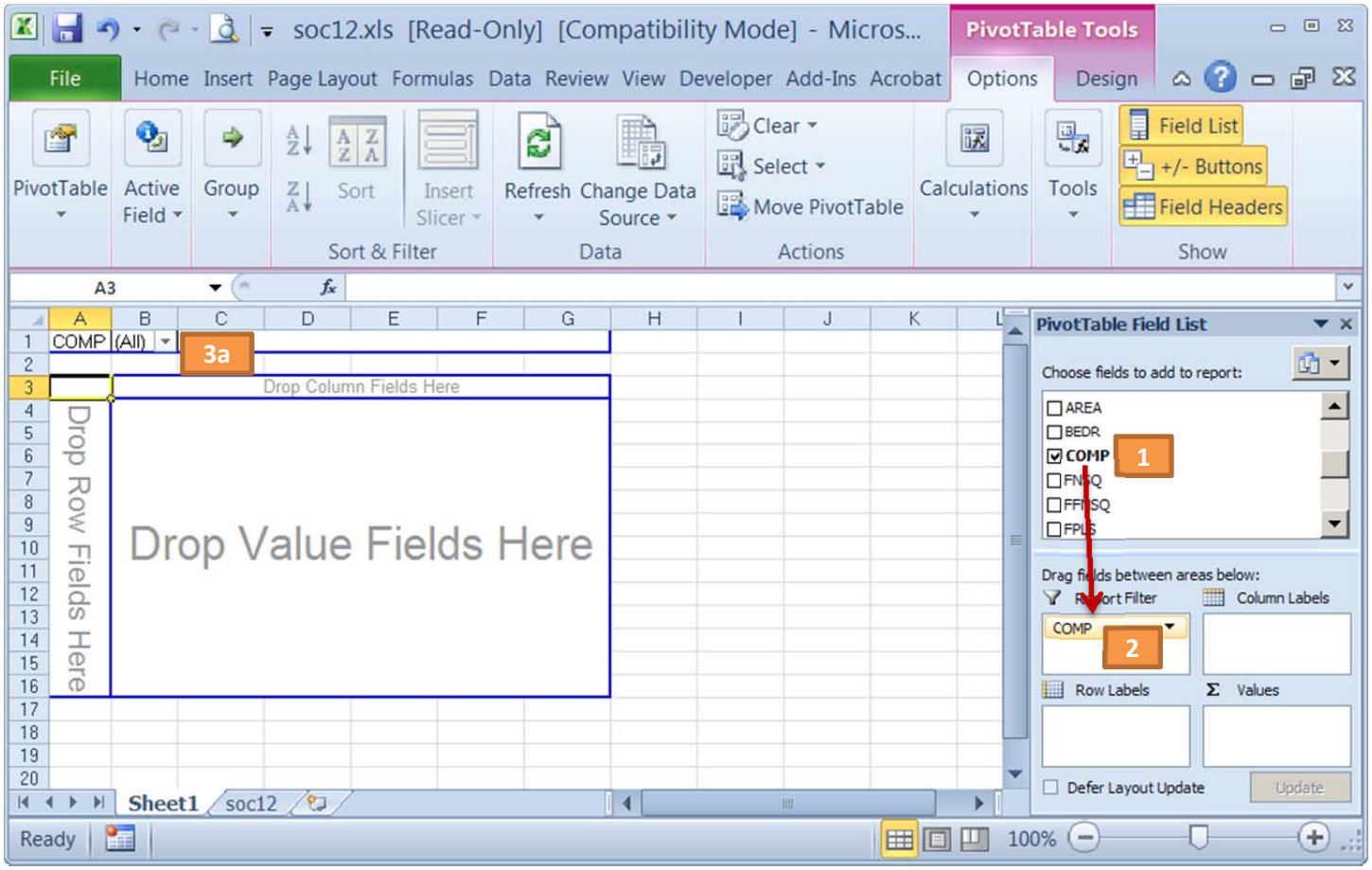
Notice that your screen is now split into 2 sections. The PivotTable is on the left side, and the PivotTable Field List is on the right side. The "classic" PivotTable view is shown above.

> The **PivotTable** (left side) is where the actual calculations will be viewed. It is also interactive. You can drag and drop fields to any of the sections that are outlined in blue.
>
> The **PivotTable Field List** (right side) shows a list of all the available data items (in the order they appear in the source table). Dragging an item to the four boxes below has the same effect as dragging to the corresponding box in the PivotTable itself.

Note: The field list is not visible if you click on a cell that is not part of the PivotTable. When you click back in the PivotTable area, the field list will re-appear.

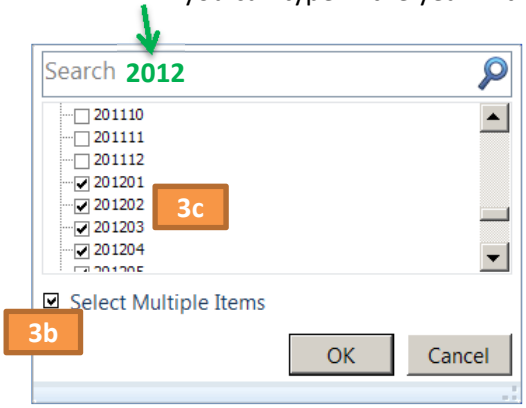
Set up your date filter

1. In the PivotTable Field List, scroll down until you find COMP (completions)
2. Drag it to the "Report Filter" box



3. Choose only the dates for this year
    a. In the PivotTable, select the dropdown arrow next to COMP
    b. Check the "Select Multiple Items" box
    c. Check only the dates for this year (or type in the year in the search box), click Ok

    TIP: you can type in the year in the search box and hit enter instead of selecting each month manually



Date Fields (yyyymm format):

AUTH = authorization date (permit issued)
COMP = date of completion (final flooring installed)
START = start date (groundbreaking for foundation)
SALE = sale date (contract signing)

Note: You can also do this with any date field (AUTH, COMP, START, or SALE). It is also possible to filter by multiple variables by dragging additional items to the "Report Filter" box.

Frequently Asked Questions about the date fields

The four date fields for SOC are as follows:

- **AUTH** – when the building permit was issued (authorized)
- **STRT** – excavation for the foundation
- **COMP** – when final flooring is put in, or the house is occupied
- **SALE** – when a sales contract is signed or a deposit is given (can be any time during construction)

Why do I have to filter the dates? If I use the soc12.xls file, shouldn't the dates already be for 2012 only? No they shouldn't. Remember that this file is a 'snapshot' of all the houses that are included in the survey for the year. We keep a house in sample from the time the permit is issued, until the house is completed or sold - whichever comes last. A typical house takes about 1 month from permit to start, and another 5-7 months from start to completion. Owner-built houses typically take around 10 months to complete. (Current figures can be found on our "Length of Time" page at http://www.census.gov/construction/nrc/lengthoftime.html.) If a house is authorized, under construction, or for sale *at any point in the survey year*, it will be included in this file.

Ok, but I see dates that are older than last year. Why is that? Older dates usually indicate a delay in construction. We still continue to follow these houses as long as the permit retains "active" status (meaning it is not cancelled, revoked, or expired) by the issuing jurisdiction. Another reason for old dates could indicate a stoppage after construction has been started. This can happen for many reasons, including zoning problems, legal issues, financing problems, etc. In addition, houses that are for sale can have old dates simply because the seller is having trouble finding a buyer. These houses will remain in the survey until they are sold, or until the sale category changes (they are rented out for example).

So theoretically, you can follow a house for 10 or 20 years? No. The Survey of Construction will only keep a house in sample for 7 years (84 months). If a house is not completed and sold by then, it is dropped from the sample. The reason for this is to save costs. These handful of old houses are not worth the cost or effort of following up every month. Due to the small number of cases, these dropped houses have little to no effect on the published estimates.

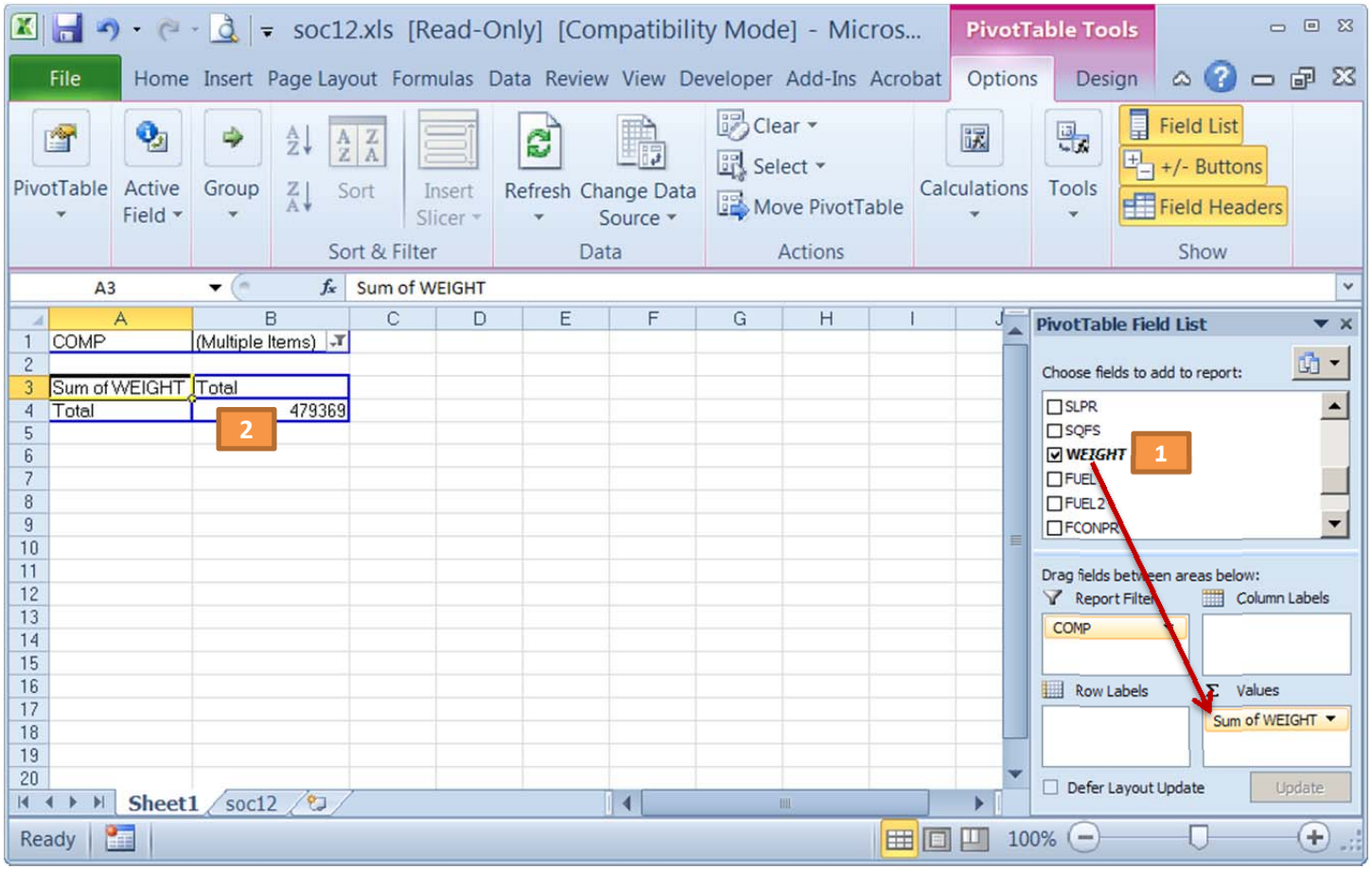
Why do you include 4 months of data past the end of the survey year? We include the first 4 months of the next year to ensure that the latest revisions and corrections to the data are included for the current year. This improves the accuracy of the data.

What does "0" mean in a date field? A zero in the COMP field simply means that the house hasn't been completed yet. The same applies to the STRT, and SALE fields. The only date field that cannot have zeros is the AUTH field. This is because there must at least be an authorization (building permit) for the house to be selected for our sample. In the small number of jurisdictions in the country that do not require a building permit, the AUTH field is filled with the date in which the house was started (AUTH and STRT will be the same).

Why use "YYYYMM" format? Why not use the built-in date format for Excel? The data in this file was not created in Excel. We just convert it to Excel to make it more user-friendly. Further, the SOC survey pre-dates Excel by over 30 years. We keep the old format for consistency, so that we can make comparisons with older files and data. It also helps keep the file sizes down, and lets us use the data in different programs and platforms.

Set up your weights

1. In the PivotTable field list, find WEIGHT and drag it down to the box titled “Values”
2. Your “Total” will appear in the PivotTable showing the weighted number of houses completed in 2012



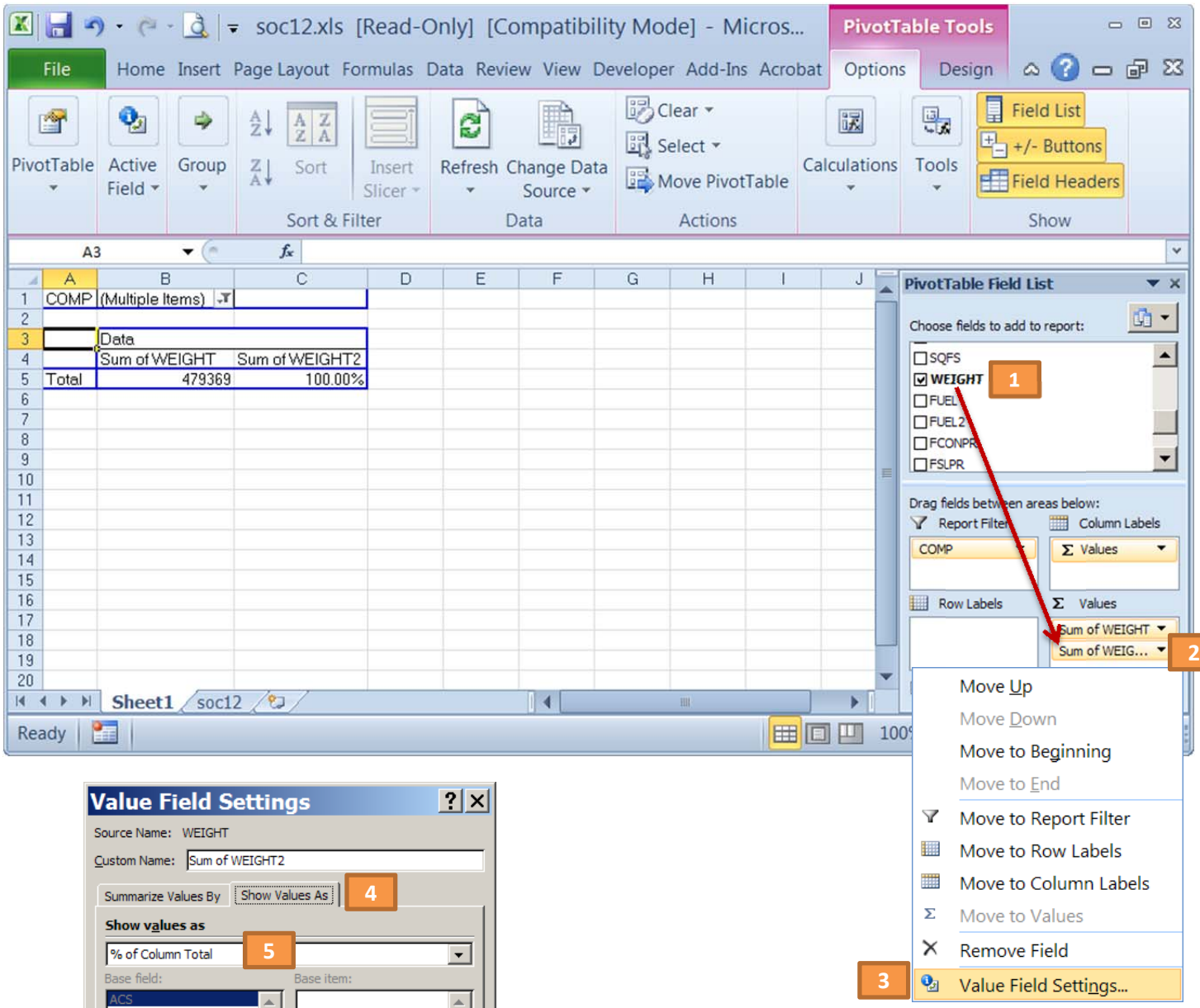
Why are the cases weighted? This survey (Survey of Construction) uses a sample of new residential buildings authorized in roughly 900 permit-issuing places and a sample of land areas where building permits are not required. To estimate houses for the entire United States, each case is weighted to represent other houses in similar geographic areas that were not selected for the survey. On average, each case represents about 50 houses. For a more detailed explanation, see our “How the Data are Collected” page on our website at: http://www.census.gov/construction/chars/how_the_data_are_collected/.

Be advised that your “Total” will be close (but **NOT** an exact match) to the number of houses completed from the published characteristics tables. This is primarily because the Microdata and the Annual Characteristics are produced at different times and the published estimates are adjusted to account for expected late reports.

Also, the Microdata is more of a “snapshot” of the houses that are in sample at a particular time, not a final tabulation. Some of the data may still change as new information is collected and revisions are made. Remember that for the Survey of Construction, a house is in sample from the time a building permit is issued until the time it is completed/sold - on average around 7 months, but can be much longer (particularly for owner-built houses). It should also be noted that more complete data is available at the time of completion (final flooring is installed) than at the time the house is authorized (a permit is issued) or started (excavation is begun for the foundation).

Set up a weighted percent field (optional but recommended)

1. Drag WEIGHT to the 'Values' box a second time (place it beneath the first WEIGHT)
2. Click on the arrow next to the second WEIGHT
3. Select "Value Field Settings" from the menu
4. Click on the "Show Values As" tab
5. Show values as "% of Column Total" (should display 100% in the table)

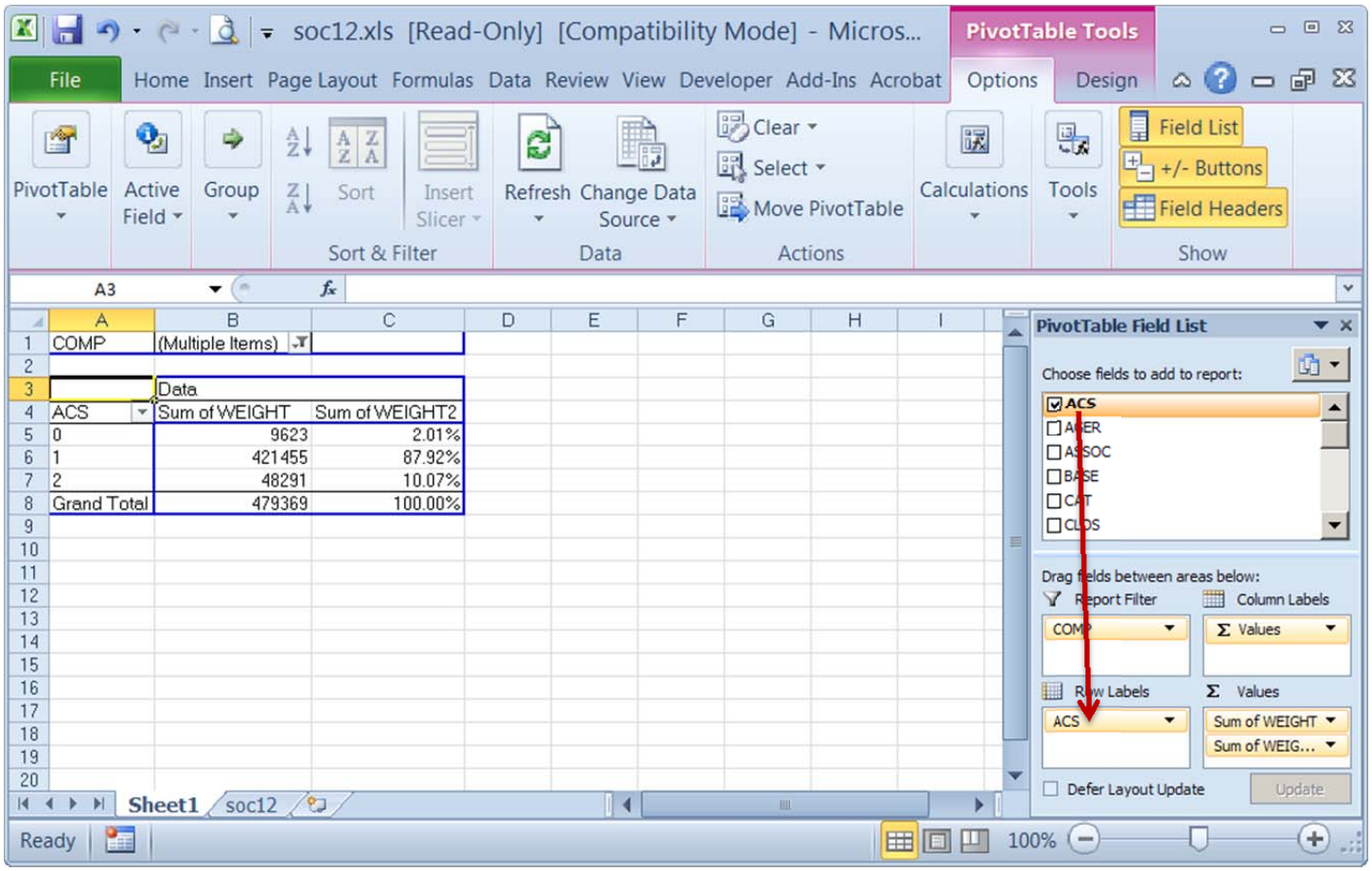


Now you are ready to make basic tables of the data (in a vertical format).

Example 1: Air-Conditioning

Now let's try a simple example with a single variable.

If you want to see the breakdown of ACS (air-conditioning) for houses completed in 2012, just click on the checkbox next to the item, or drag ACS to the "Row Labels" box. Your screen should look like this:



By checking and unchecking items, you can get basic tabulations of the data, and the corresponding percentages. Use the microdata guide to see what each category represents. It can be downloaded from: http://www.census.gov/construction/chars/pdf/socmicro_info.pdf.

Using the above table, we can see that approximately 88% of houses completed have air-conditioning, 10% do not, and 2% did not report that item (meaning the data was unavailable).

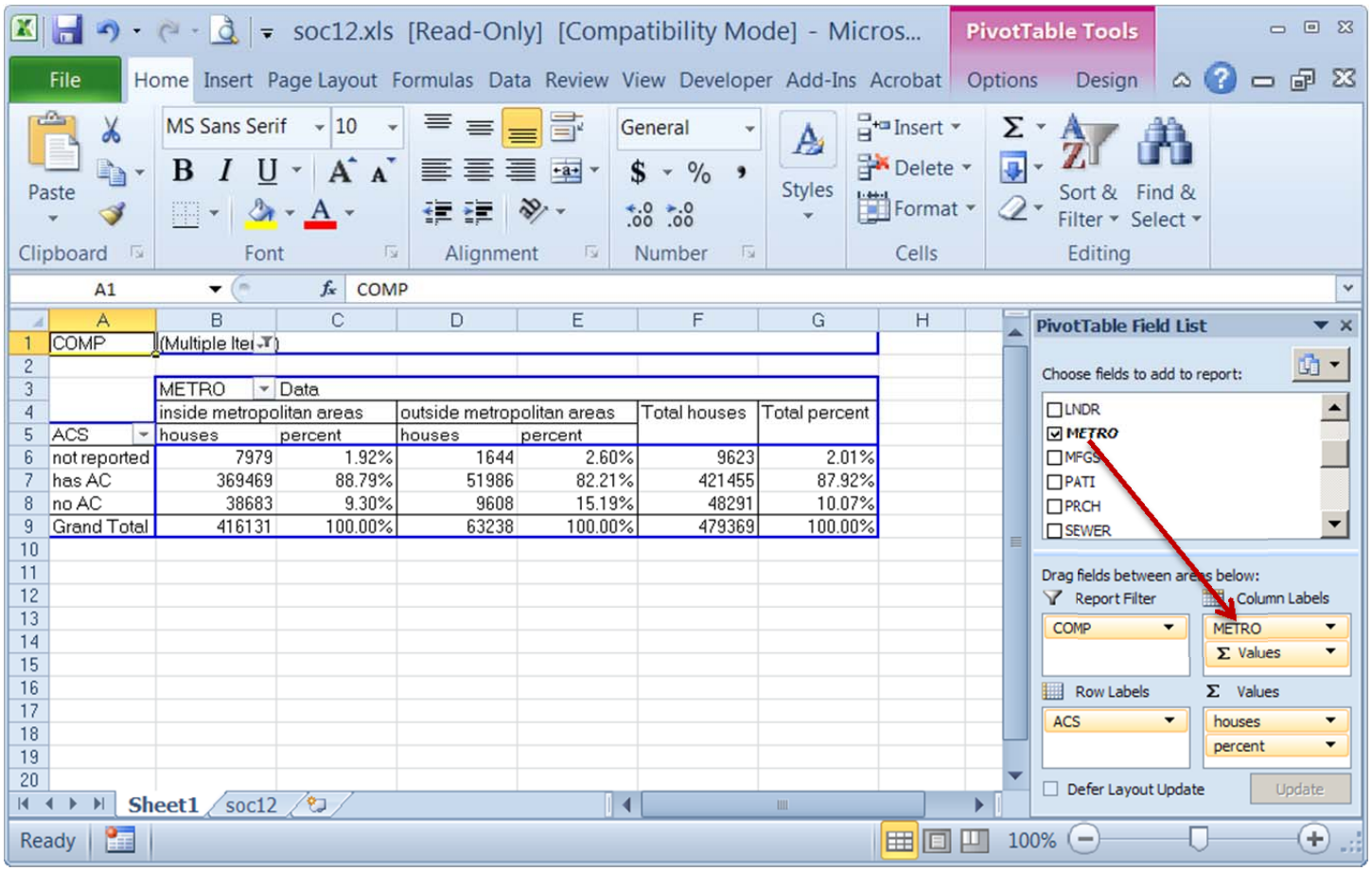
Note: You can edit the row and column headings (but not the data) in the PivotTable by typing over the titles. Excel will remember the new headings, but it will not affect the original data table.

	A	B	C
1	completion	(Multiple Items)	
2			
3		Data	
4	air-conditioning	number of houses	percent
5	not reported	9623	2.01%
6	has a/c	421455	87.92%
7	no a/c	48291	10.07%
8	Grand Total	479369	100.00%
9			

Example 2: Air-Conditioning by Metro

Using Example 1 above, let's try something a little more advanced and make a "crosstab" (table of X by Y) of air-conditioning by metropolitan area.

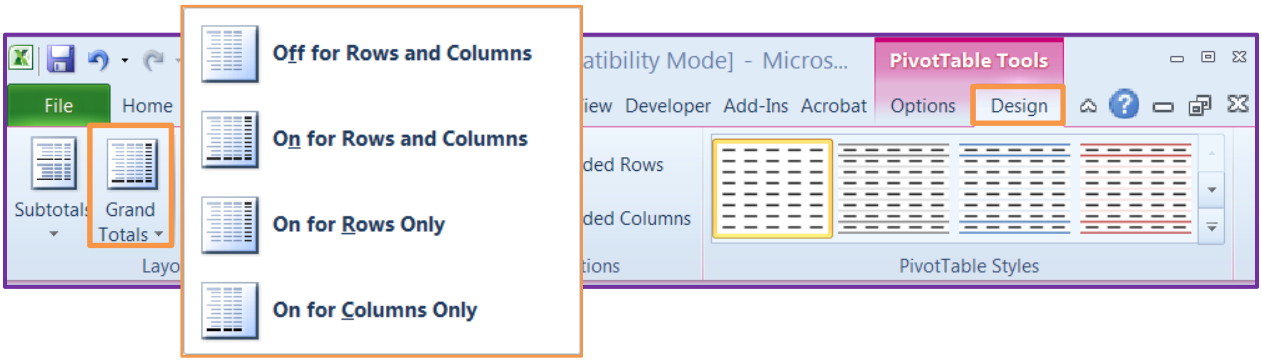
In the "Field list", find the variable METRO and drag it to the "Column Labels" box and place it above the "Values" item.



Notice that in the above example, I replaced the row headings with text to indicate inside/outside metro area. (This is optional, but helps keep track of what we are tabulating.)

According to the table above, we can see that 89% of houses within metropolitan areas have air-conditioning, compared to 82% of houses that are not located in metropolitan areas.

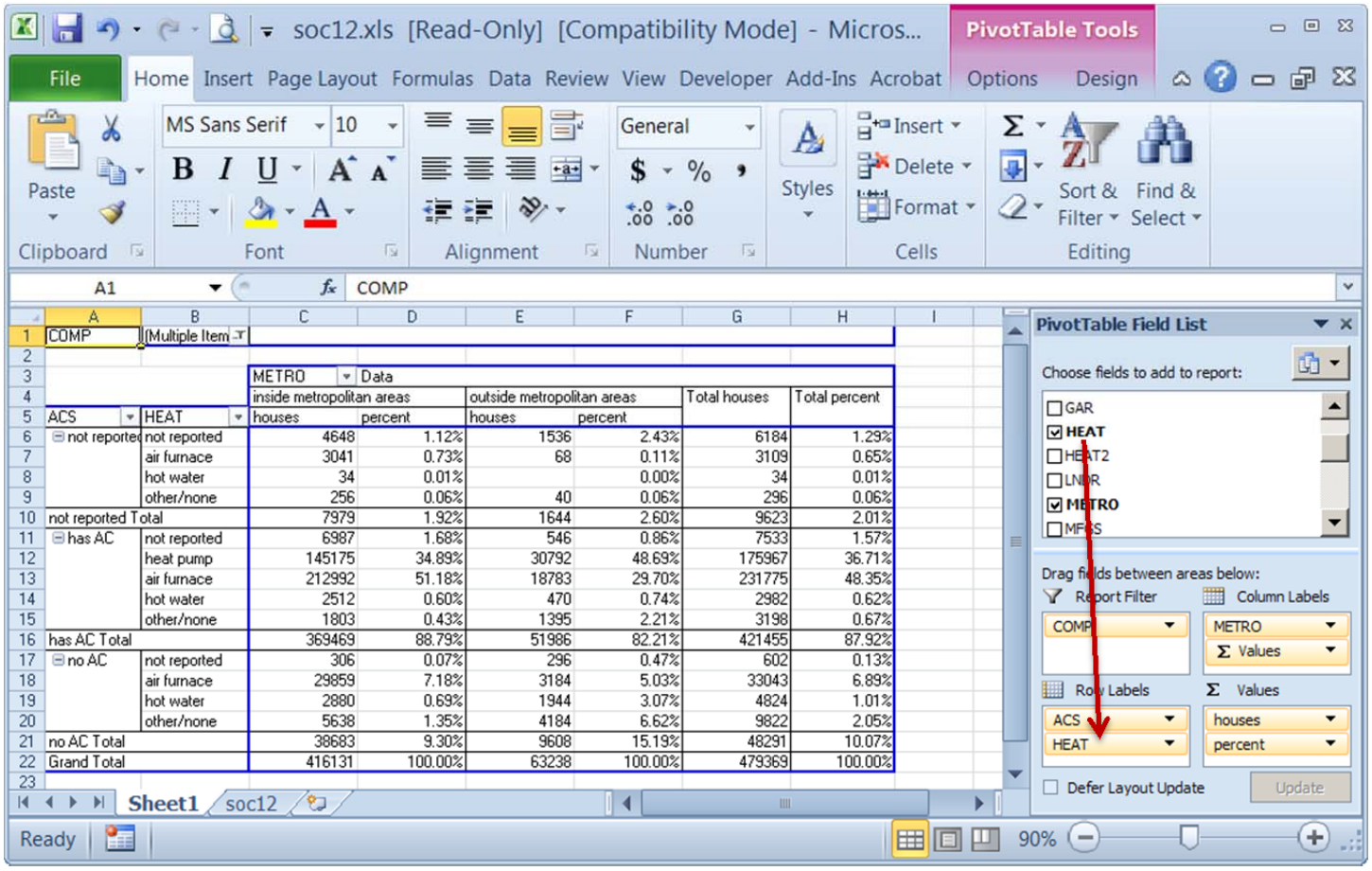
Note: You can turn off row totals, column totals, or both by clicking the "Design" tab and selecting "Grand Totals".



Example 3: Air-Conditioning, Metro, and Heating System

Now let's add a third variable to our example.

This time, find the variable HEAT (type of heating system) and drag it to the "Row Labels" below ACS.



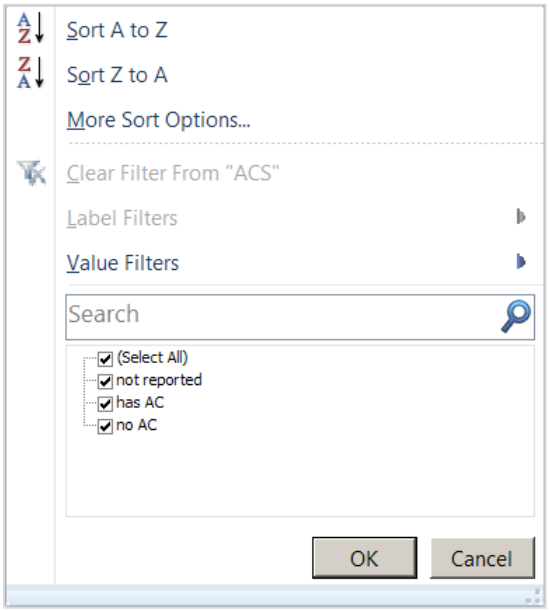
According to the above table, we can make a statement such as "Of houses within metropolitan areas, 51.2% have air-conditioning AND a furnace, versus 29.7% of homes outside of metropolitan areas".

Note: When dragging items from the Field List, the order you place them in matters. For example, if you placed HEAT above ACS in the "Row Labels" box, HEAT would become your primary variable instead of ACS. (ACS by HEAT is **not** the same as HEAT by ACS.)

Helpful Hints

Below are some helpful tips for working with the SOC microdata, and PivotTables in general.

- **Source Data** – Don't modify or change the source data table. Changing the original data will obviously change your results when you create a PivotTable.
- **Keep it simple** – PivotTables are a powerful tool, but for the majority of the time, you want to keep it simple. Too many variables in a table will make it difficult to read and analyze.
- **Use the Microdata guide** – download SOC Microdata Guide from: http://www.census.gov/construction/chars/pdf/socmicro_info.pdf.
- **Filtering** – clicking the arrows next to the row or column headings let you filter your results. This is helpful if you want to exclude a certain category from your tabulations. The menu also gives you the capability to sort data or search for an item.



- **Expand/Collapse** – if you have multiple items in the 'Row Label' box, the PivotTable results will display small "+" or "-" signs in front of the data item. These let you expand or collapse the data columns.
- **Formatting** – you can change the format of your PivotTable by clicking the "Design" tab from the main Excel menu. This gives you various color and display options to apply to your PivotTable.
- **PivotTable Help** – help is available by clicking the Excel Help icon (blue question mark in the top right corner), then go to Analyzing Data > PivotTable reports. You can also search online at: http://office.microsoft.com/en-us/support/?CTT=97.