

Analyzing Census Data with Excel – Course Guide

Learn how to use Excel to summarize, analyze, and visualize Census data.

Developed by Adam Hecktman, Microsoft Director of Technology and Civic Innovation for Chicago

Module 1: [Basic Census Data Access and Table Formatting](#)

Module 2: [Quick Census Data Analysis in Excel](#)

Module 3: [Advanced Census Data Access and Hierarchical Charts in Excel](#)

Module 4: [Advanced Census Data Analysis in Excel](#)

Extra Modules: [Mapping Census Data in Excel](#)

Introduction

The U.S. Census Bureau's mission is to serve as the leading source of quality data about the nation's people and economy. Whether you are looking for the most current economic indicators or for demographic and socioeconomic characteristics about your community and how they compare statistically to other areas, data from the U.S. Census Bureau are available online through a variety of tools.

Every year, the Census Bureau publishes population, socioeconomic, housing and business statistics for all communities in the country. These data can be sorted by several characteristics such as age, sex and race, as well as different levels of geographies from the nation to areas below cities and towns. They can be charted, compared with other data (from inside or outside Census), and can be used to derive insights about the nation's characteristics.

Excel is a tool that many people are familiar with, and many people use every day. Excel can be leveraged to unlock the value of open data of all kinds, and it is particularly well-suited to transforming, analyzing, and visualizing Census data. This course will show how to use Excel to access, manipulate, and visualize Census data. It will also cover mapping that data, and tools for doing advanced statistical analysis.

MODULE 1: Basic Census Data Access and Table Formatting

One very easy way to find basic data is by browsing the Community Facts feature found in the American Fact Finder (AFF). American Fact Finder itself is an online, self-service tool designed to search a variety of population, economic, geographic, housing, education, and other types of data. It is a very easy way to look for the data you may need.

The Community Facts section shows you popular tables. You can use the Guided Search feature to narrow down what you are looking for and have AFF suggest datasets for you. You can use the Advanced Search feature (which we will explore later) to build a query for your search. In all cases, the data can be modified prior to being downloaded. We will start with the Community Facts to find our data.

Let's say we simply want to see the population growth (or decline) of every county in the US over the last 6 years.

Open <http://census.gov> in your browser.

On the menu bar at the top, click on Data, and then go to Data Tools & Apps

Select American Fact Finder, and click on Community Facts on the menu bar

In AFF, look at the options you have on the left to find the data you are looking for

We want population data, so choose population.

Note that at the top, shaded in blue, you can see very common queries presented as a number for the geography you may be looking for. In the case below, it is the Census 2010 Total Population. You can select other common demographic data from the drop down. But in this case, we want several years of population estimates.

Select Annual Population Estimates

The screenshot shows the American Fact Finder website interface. At the top, there is a navigation bar with 'United States Census Bureau' and 'FactFinder' logos. Below the navigation bar, there is a search bar and a 'GO' button. The main content area is titled 'Community Facts - Find popular facts (population, income, etc.) and frequently requested data about your community.' On the left side, there is a sidebar with a list of categories: Population, Age, Business and Industry, Education, Governments, Housing, Income, Origins and Language, Poverty, Race and Hispanic Origin, Veterans, and Show All. The 'Population' category is selected and highlighted with a red circle. In the main content area, the 'United States' section shows the 'Population' dropdown menu set to 'Census 2010 Total Population' with the value '308,745,538'. Below this, there is a section titled 'Popular tables for this geography:' which lists several tables. The '2016 Population Estimates Program' section is highlighted with a red arrow, and the 'Annual Population Estimates' table is selected.

This will bring up a view of the table, and this is the table that we will open in Excel. One of the nice features of this view is that you can modify the table before you even download it into Excel. These are modifications that you can do in Excel after you download it as well, but sometimes it is nice to have less data to download. We will make some modifications in this view, and the rest of the transformation we can do in Excel.

Another way to search for data is to use the Advanced Search feature.

Data Transformation

One of the many benefits of using Excel is that it can be used to easily transform your data. In this case, we are going to start modifying our table using the Census tools, and then we will put the finishing touches on the structure using Excel. You can do all of this in Excel, but this example shows you useful methods for using both.

You should now be seeing the table with April 2010 census data, and yearly population estimates. Let's say that we only want to download the yearly population estimates.

Click on the Modify Table button

Geography	April 1, 2010		Population Estimate (as of July 1)						
	Census	Estimates Base	2010	2011	2012	2013	2014	2015	2016
United States	308,745,538	308,758,105	309,348,193	311,663,358	313,998,379	316,204,908	318,563,456	320,896,618	323,127,513

Uncheck the April 1, 2010 Census and the Estimate base check boxes to modify the table.

We will do the rest of the transformations in Excel, so click on Download.

Choose "Use" and keep the Merge Annotations and Include Descriptive Elements checked.

This will prepare the file for download to a zip file.

Click Download

Unzip the file called PEP_2016_PEPANNRES_with_ann.csv

It is a best practice to get in the habit of saving your CSV files as an Excel Workbook right away. This will avoid you wondering later why all the features of an Excel workbook are not available to you.

Click on File/Save As, and choose the Excel Workbook (usually the top option).

Name this file "Population"

If you expand the columns in this spreadsheet, you will see right away that there are some columns and at least two rows we will not be using. For example, the first two columns are geographic IDs. We

won't be using those, so it makes sense to delete them. There is a top row with field names that we won't be using, so it makes sense to delete that.

Delete columns A and B with the geography IDs

Delete first row with the Field Names.

Delete the row with the total United States population information

A	B	C	D	E	F	G	H	I	J
GEO.id	GEO.id2	GEO.display-label	respop72010	respop72011	respop72012	respop72010	respop72011	respop72012	respop72010
Id	Id2	Geography	Population Estimate (as of July 1) - 2010	Population Estimate (as of July 1) - 2011	Population Estimate (as of July 1) - 2012	Population	Population	Population	Population Estim
0100000U\$		United States	309348193	311663358	313998379	3.16E+08	3.19E+08	3.21E+08	3.23E+08
0500000U\$	1001	Autauga County, Alabama	54742	55255	55027	54792	54977	55035	55416
0500000U\$	1003	Baldwin County, Alabama	183199	186653	190403	195147	199745	203690	208563
0500000U\$	1005	Barbour County, Alabama	27348	27326	27132	26938	26763	26270	25965
0500000U\$	1007	Bibb County, Alabama	22861	22736	22645	22501	22511	22561	22643
0500000U\$	1009	Blount County, Alabama	57376	57707	57772	57746	57621	57676	57704
0500000U\$	1011	Bullock County, Alabama	10892	10722	10654	10576	10712	10455	10362
0500000U\$	1013	Butler County, Alabama	20938	20848	20665	20330	20283	20126	19998
0500000U\$	1015	Calhoun County, Alabama	118468	117736	117208	116475	115837	115285	114611
0500000U\$	1017	Chambers County, Alabama	34101	34006	34084	34123	33996	34043	33843
0500000U\$	1019	Cherokee County, Alabama	25977	26073	26017	26074	25944	25726	25725
0500000U\$	1021	Chilton County, Alabama	43675	43758	43693	43769	43874	43809	43941
0500000U\$	1023	Choctaw County, Alabama	13840	13610	13549	13384	13310	13201	12993
0500000U\$	1025	Clarke County, Alabama	25762	25549	25148	25105	24862	24726	24392
0500000U\$	1027	Clay County, Alabama	13882	13676	13461	13464	13503	13495	13492
0500000U\$	1029	Cleburne County, Alabama	14975	14976	14928	15031	15076	14996	14924

Your table will now look like this:

A	B	C	D	E	F	G	H
Geography	Population Estimate (as of July 1) - 2010	Population Estimate (as of July 1) - 2011	Population Estimate (as of July 1) - 2012	Population	Population	Population	Population Estim
Autauga County, Alabama	54742	55255	55027	54792	54977	55035	55416
Baldwin County, Alabama	183199	186653	190403	195147	199745	203690	208563
Barbour County, Alabama	27348	27326	27132	26938	26763	26270	25965
Bibb County, Alabama	22861	22736	22645	22501	22511	22561	22643
Blount County, Alabama	57376	57707	57772	57746	57621	57676	57704
Bullock County, Alabama	10892	10722	10654	10576	10712	10455	10362
Butler County, Alabama	20938	20848	20665	20330	20283	20126	19998
Calhoun County, Alabama	118468	117736	117208	116475	115837	115285	114611
Chambers County, Alabama	34101	34006	34084	34123	33996	34043	33843
Cherokee County, Alabama	25977	26073	26017	26074	25944	25726	25725
Chilton County, Alabama	43675	43758	43693	43769	43874	43809	43941
Choctaw County, Alabama	13840	13610	13549	13384	13310	13201	12993
Clarke County, Alabama	25762	25549	25148	25105	24862	24726	24392
Clay County, Alabama	13882	13676	13461	13464	13503	13495	13492
Cleburne County, Alabama	14975	14976	14928	15031	15076	14996	14924
Coffee County, Alabama	50183	50484	51180	50729	50757	51063	51226
Colbert County, Alabama	54521	54460	54493	54456	54427	54295	54216
Conecuh County, Alabama	13209	13193	13007	12866	12621	12597	12395
Coosa County, Alabama	11758	11369	11201	11058	10794	10687	10581

Another best practice is to format your data as a table as soon as it is convenient. Doing so is easy, it makes it more readable, and it gives the data some analysis features right off the bat.

Click inside your data

On the Ribbon, in the Styles section, click on Format as Table

Pick a look that you like.

It will confirm your table range, and keep the My Table Has Headers checked.

You will notice that it gives arrows next to the headers. Click on the arrows and you will see options for sorting and filtering your data. This all comes with being formatted as a table! This can be very useful in finding specific information. For example, let's say you want to find population data for Cook County.

Click on the arrow next to Geography

Under Text Filters, start type "Cook" and hit OK

But wait...There are several counties with Cook in the name, and you just wanted the one for Cook County, Illinois. It would be better if the state were in its own column, separate from the county. That way you could filter the state to Illinois and the County to Cook. Let's do a very common transformation with Census data. Let's break up a cell into multiple cells.

Clear the filter by clicking on the arrow next to Geography again and choosing Clear Filter.

Insert a row next to Geography (select Row B, right click and hit insert)

Geography	Population	Population estimate for July 1, 2011	Population estimate for July 1, 2011
Autauga County, Alabama	54762	55255	55027
Barbour County, Alabama	181199	186451	190818
Bibb County, Alabama	27868	27236	27232
Blount County, Alabama	22861	22726	22645
Bolivar County, Alabama	57379	57707	57772
Bullock County, Alabama	12802	12722	12654
Butler County, Alabama	29938	29848	29605
Calhoun County, Alabama	114468	117726	117208
Chambers County, Alabama	94131	96906	96384
Cherokee County, Alabama	23077	26873	26037
Chilton County, Alabama	46375	41758	46698
Chocoma County, Alabama	13862	13810	13549
Clarke County, Alabama	23762	25149	25148
Clay County, Alabama	13882	13676	13461
Clayton County, Alabama	14974	14976	14638
Coffee County, Alabama	93181	90484	91180
Colbert County, Alabama	14521	14440	14403
Covadonga County, Alabama	13208	11193	11807
Craig County, Alabama	11758	11349	11203
Crenshaw County, Alabama	17021	28564	17642
Cross County, Alabama	13857	13903	13920
Cullman County, Alabama	89376	80457	83951
Dale County, Alabama	50920	50119	50296
Dallas County, Alabama	43820	43207	42748
DeKalb County, Alabama	71143	71149	70903
DeWitt County, Alabama	79578	80010	80291

Select the Geography column

On the Ribbon, choose the Data Tab

Click on Text to Column and note the preview, this is very useful

You have some options for how you want to split up your column. You can use a delimiter, meaning that Excel should split up the column based on a character (a comma in this case). Or you can use a fixed width, which you sometimes need for data coming off of legacy systems like mainframes. In this case, we are delimited by the comma that separates the county from the state.

Choose delimited and click next

Check the Comma as your delimiter (you can uncheck tab if you like) and click next

Look at the preview and verify it will look the way you want

Click Finish and accept the warning if it comes up

You now have two columns, one for the county name and one for the state name.

Rename the new column State

Now you can do things like analyze population data changes for a single state. For example, let's say we just want to see the data for the state of Hawaii.

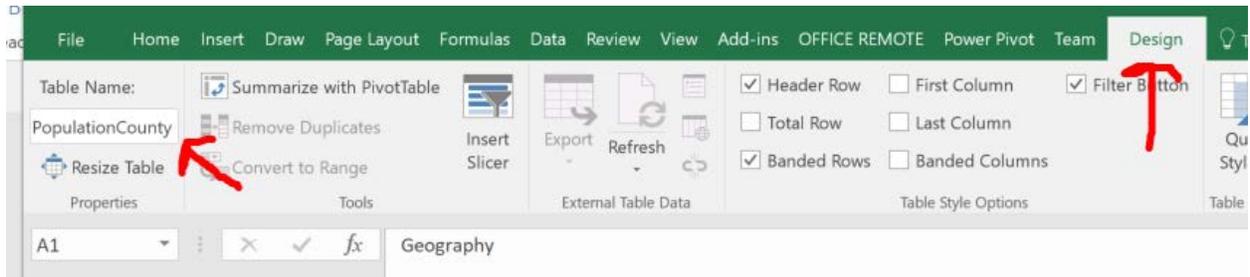
Click on the drop down on the column header

Under Text Filter, start typing Haw and you will see that only Hawaii is selected

Now you are seeing the population data corresponding to the four counties in Hawaii

To go back to the full table, simply click on the drop down again and click on clear filter.

You now have a fully functioning table. Yet another best practice is to name your table. This will make it obvious which data you are referring to when you do more advanced things like working with formulae and maps. To name this table, simply click on the Design tab on the ribbon, and change the Table Name (in the Properties section of the ribbon) from Table1 to something more meaningful like PopulationCounty.



MODULE 2: Quick Census Data Analysis in Excel

There are some very simple ways to create quick and effective analyses in Excel. Later we will talk about more advanced analysis, but this section will focus on easy ways to derive insights from Census data.

With a table like this, with year over year data, it is useful to see trends. You can certainly quickly create a line chart that shows the trend for each state, but with so many counties, it would be difficult to see the trend for any particular county, as it would get lost in the chart. This is where Sparklines come in handy.

Sparklines are like little charts that live inside a single cells that can be used to visually represent and show a trend in your data. Sparklines can draw attention to important items, in this case population changes. We can also use them to highlight the maximum and minimum values in the row.

Create another column header in the column to the right of the population estimate for 2016 Name it "Trend". Notice that Excel automatically makes this part of the table for you, complete with formatting.

I	J	K	L
Population Estimate (as of July 1) - 201	Trend		
55416			
208563			
25965			
22643			
57704			
10362			
19998			
114611			
33843			
25725			
43941			

Now, we will use this cell to create a Sparkline to see the population trend for that county.

Click in the first cell under Trend

On the ribbon, click on the Insert tab. In the Sparklines section, you will see some options.

Start with the Line Sparkline. For Data Range, select the row of population estimates (C2:K2)

That sparkline shows the trend, year over year, of the population for that county!

Now, you can copy the sparkline for every county.

*Using a shortcut, you can select every cell in that column in that table by holding **ctrl+shift+end***

*Now, from the home tab on the ribbon, in the editing section, choose **fill/down**.*

I	J
Population Estimate (as of July 1) - 2016	Trend
55416	
208563	
25965	
22643	
57704	
10362	
19998	
114611	
33843	
25725	
43941	
12993	
24392	
13492	
14924	
51226	
54216	
12395	
10581	
37458	
13913	
82471	
49226	
40008	
70000	

Now you can quickly see the trends for each county! That is a lot of great insight in a single cell. But it would be also very useful if you could see where the highs and lows are.

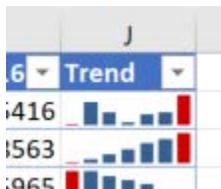
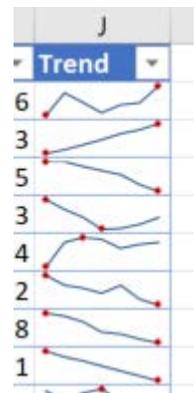
Click on the Sparkline Tools Design tab on the ribbon.

Note in the Show section that you have a variety of things you can show in each cell.

Select the high point and low point

You can even change the colors of what is the high point and what is the low point.

Experiment with the style options



Line charts are not your only option. You can also use a column chart or a win/loss chart.

You can experiment with these, but the trend is easiest viewed with the line chart in this scenario.

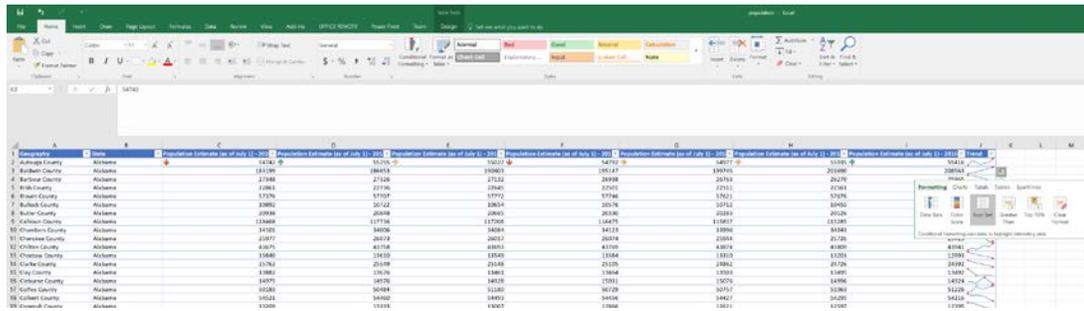
There is a quick menu of tools called the Quick Analysis Tools that provides a trove of insights that you can add to your Census data. You can instantly create different types of charts, including line and column charts, or sparklines. You can also apply a table style, create PivotTables, quickly insert totals, and apply conditional formatting. Here, we will add icons, another way to do quick visual analysis right from within the cell. To experiment with these, select the cells in the first row of data (the population estimates for Autauga County, Alabama)

Hit **Ctrl+Q**, which is the keyboard shortcut for Quick Analysis Tools

In the mini toolbar, you can see you have a variety of options for doing analysis of your data.

Hover over some of these to get a preview of what they do.

Then click on **Icon Set**



The icon set puts icons directly in the cell based on conditions. In this example, it tells you if the population went significantly up, down or stayed relatively flat.

Next, select the data on one of the other rows.

Hit **Ctrl+Q** again to bring up the quick analysis tool.

This time choose **Data Bar**. This shades the cell based on the data in that cell.

Experiment with some of the other quick analysis options.

These tools are actually examples of conditional formatting. As you saw, conditional formatting rules to help you do things the built-in formatting rules can't do. To see what they look like, click on the **Conditional Formatting** in the **Style Section** of the **Home** tab. Here you can not only see the wide variety of icons, data bars, and color schemes you can use, at the bottom (**Manage Rules**), you can manage how these visualizations are applied.

MODULE 3: [Advanced Census Data Access and Hierarchical Charts in Excel](#)

In this section, we will explore two topics. First, we will look at how to use the Advanced Search feature of the Census American Fact Finder. Then we will modify and download a dataset, and use that to create a hierarchical chart (a chart that highlights data in multiple categories).

Let's say you want to understand the income gap between males and females with varying levels of education. You can see that we are looking at three separate variables. Income, male and female, and education level. Let's search the American Fact Finder using those variables to find exactly what we are looking for.

Go back to American Fact Finder (<http://factfinder.census.gov>).

Click on Advanced Search on the menu.

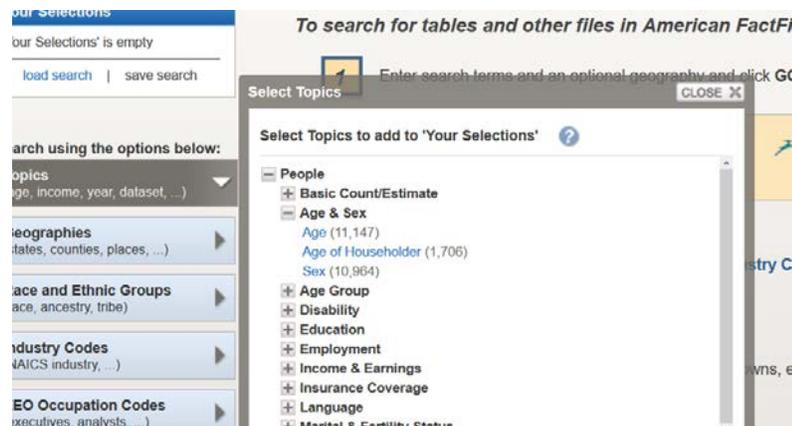
On the left, click on Topics under the search options.

Here you will choose the three topics mapping to the three variables.

Start with male vs. female. These are people, so click on People.

Click on Age & Sex.

Click on Sex.



Notice that when you do that, it has added Sex to your search criteria in the box on the left marked "Your Selections".

Now we want to find Income. Choose Income and Earnings.

Choose Income/Earnings (Individuals). It is added to your search criteria.

Last, we want this data by levels of education. Click Education

Choose Educational Attainment

Close the search box

For this example we will look at national level data. You will not need to select a geography in AFF since the US is the default geography. However, if you are interested in conducting the same analysis for a different geography, you just need to use the Geography search option right below topics to select it.

The second report gives you median earnings by sex by educational attainment. Open this.

		United States	
		Estimate	Margin of Error
18 of 18	Total:	36,860	+/-44
	Less than high school graduate	21,320	+/-62
	High school graduate (includes equivalency)	29,004	+/-110
	Some college or associate's degree	34,377	+/-178
	Bachelor's degree	50,930	+/-73
	Graduate or professional degree	67,286	+/-141
	Male:	42,378	+/-53
	Less than high school graduate	25,159	+/-109
	High school graduate (includes equivalency)	34,472	+/-272
	Some college or associate's degree	41,582	+/-79
	Bachelor's degree	61,780	+/-115
	Graduate or professional degree	85,202	+/-357
	Female:	31,072	+/-42
	Less than high school graduate	16,151	+/-105
	High school graduate (includes equivalency)	22,461	+/-70
	Some college or associate's degree	28,643	+/-136
	Bachelor's degree	41,952	+/-91
	Graduate or professional degree	56,726	+/-168

5 American Community Survey 1-Year Estimates

Quite often, you will want to analyze or visualize census data across several different categories. You may be already familiar with using Excel charts to map single categories. But there are several charts in Excel that are particularly good for hierarchal data. An example of such data may be the average earnings (the data) for people with varying levels of education (one category), broken down by gender (a second category). With a bit of cleaning, we can do that with this table.

First, click Modify Table.

Uncheck the Margin of Error column.

Uncheck the education levels under Total, as we won't be using those.

We can structure the data further in Excel.

		United States	
		Estimate	Margin of Error
<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Total:	<input checked="" type="checkbox"/>	36,860	+/-44
<input checked="" type="checkbox"/> Less than high school graduate	<input checked="" type="checkbox"/>	21,320	+/-62
<input checked="" type="checkbox"/> High school graduate (includes equivalency)	<input checked="" type="checkbox"/>	29,004	+/-110
<input checked="" type="checkbox"/> Some college or associate's degree	<input checked="" type="checkbox"/>	34,377	+/-178
<input checked="" type="checkbox"/> Bachelor's degree	<input checked="" type="checkbox"/>	50,930	+/-73
<input checked="" type="checkbox"/> Graduate or professional degree	<input checked="" type="checkbox"/>	67,286	+/-141
<input checked="" type="checkbox"/> Male:	<input checked="" type="checkbox"/>	42,378	+/-53
<input checked="" type="checkbox"/> Less than high school graduate	<input checked="" type="checkbox"/>	25,159	+/-109
<input checked="" type="checkbox"/> High school graduate (includes equivalency)	<input checked="" type="checkbox"/>	34,472	+/-272
<input checked="" type="checkbox"/> Some college or associate's degree	<input checked="" type="checkbox"/>	41,582	+/-79
<input checked="" type="checkbox"/> Bachelor's degree	<input checked="" type="checkbox"/>	61,780	+/-115
<input checked="" type="checkbox"/> Graduate or professional degree	<input checked="" type="checkbox"/>	85,202	+/-357
<input checked="" type="checkbox"/> Female:	<input checked="" type="checkbox"/>	31,072	+/-42
<input checked="" type="checkbox"/> Less than high school graduate	<input checked="" type="checkbox"/>	16,151	+/-105
<input checked="" type="checkbox"/> High school graduate (includes equivalency)	<input checked="" type="checkbox"/>	22,461	+/-70
<input checked="" type="checkbox"/> Some college or associate's degree	<input checked="" type="checkbox"/>	28,643	+/-136
<input checked="" type="checkbox"/> Bachelor's degree	<input checked="" type="checkbox"/>	41,952	+/-91
<input checked="" type="checkbox"/> Graduate or professional degree	<input checked="" type="checkbox"/>	56,726	+/-168

015 American Community Survey 1-Year Estimates

The view should look like this:

		United States
		Estimate
<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Total:	<input checked="" type="checkbox"/>	36,860
<input checked="" type="checkbox"/> Male:	<input checked="" type="checkbox"/>	42,378
<input checked="" type="checkbox"/> Less than high school graduate	<input checked="" type="checkbox"/>	25,159
<input checked="" type="checkbox"/> High school graduate (includes equivalency)	<input checked="" type="checkbox"/>	34,472
<input checked="" type="checkbox"/> Some college or associate's degree	<input checked="" type="checkbox"/>	41,582
<input checked="" type="checkbox"/> Bachelor's degree	<input checked="" type="checkbox"/>	61,780
<input checked="" type="checkbox"/> Graduate or professional degree	<input checked="" type="checkbox"/>	85,202
<input checked="" type="checkbox"/> Female:	<input checked="" type="checkbox"/>	31,072
<input checked="" type="checkbox"/> Less than high school graduate	<input checked="" type="checkbox"/>	16,151
<input checked="" type="checkbox"/> High school graduate (includes equivalency)	<input checked="" type="checkbox"/>	22,461
<input checked="" type="checkbox"/> Some college or associate's degree	<input checked="" type="checkbox"/>	28,643
<input checked="" type="checkbox"/> Bachelor's degree	<input checked="" type="checkbox"/>	41,952
<input checked="" type="checkbox"/> Graduate or professional degree	<input checked="" type="checkbox"/>	56,726

American Community Survey 1-Year Estimates

Now, we can take this opportunity to learn another way to download data from Census.

Click on Download, and instead of choosing Use (which downloads in CSV), choose View (which downloads in PDF or Excel).

Select Excel as the presentation format.

Hit Download.

Open the file.

Your data is now exactly as you structured it, with a few extra columns and rows. There is some verbiage which (once you read it), you can delete.

Delete the first 8 rows above your table, and then all the rows of text below it (should start at row 16).

	United
Total:	36,860
Male:	42,378
Female:	31,072
Less than high school graduate	25,159
High school graduate (includes)	34,472
Some college or associate's degree	41,582
Bachelor's degree	61,780
Graduate or professional degree	85,202
Less than high school graduate	16,151
High school graduate (includes)	22,461
Some college or associate's degree	28,643
Bachelor's degree	41,952
Graduate or professional degree	56,726

Now, recall our best practice of saving these files as an Excel Workbook right away? That applies even here where it is an Excel file already. The reason is that this is actually an Excel file in compatibility mode. This means it is meant to be used by someone with Excel 97 – 2003.

There are many...loads...of features that are available in more recent versions that you will want to take advantage of. Saving this is a two-step process:

File/Save As/type is Excel Workbook

Close and re-open the file so that you are no longer in compatibility mode.

Now let's simplify this table a bit.

Your table will look like this.

	A	B	C	E	F
1				United	
2				Estimate	
3	Total:			36,860	
4	Male:			42,378	
5	Less than high school graduate			25,159	
6	High school graduate (includes			34,472	
7	Some college or associate's degree			41,582	
8	Bachelor's degree			61,780	
9	Graduate or professional degree			85,202	
10	Female:			31,072	
11	Less than high school graduate			16,151	
12	High school graduate (includes			22,461	
13	Some college or associate's degree			28,643	
14	Bachelor's degree			41,952	
15	Graduate or professional degree			56,726	

At first glance, it looks fine, but actually there are some extra columns that are a little hidden. Start by deleting the first row.

Delete the first row.

Next, delete the row that says total

Delete Column C

Delete Column B

At this point, you should have this:

	A	B	C	D
1		Estimate		
2	Male:	42,378		
3	Less than high school	25,159		
4	High school graduate	34,472		
5	Some college or associate's	41,582		
6	Bachelor's degree	61,780		
7	Graduate or professional	85,202		
8	Female:	31,072		
9	Less than high school	16,151		
10	High school graduate	22,461		
11	Some college or associate's	28,643		
12	Bachelor's degree	41,952		

Now, again, a hierarchical chart has several columns of criteria. We want to separate Male and Female into their own columns.

Add a column to the right of A

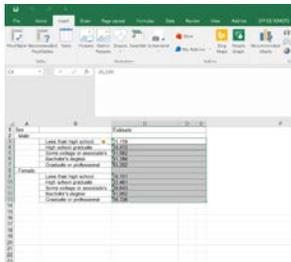
Move the labels "Male" and "Female" into those columns, and delete the estimates for each.

Label the Column "Sex"

We are very close. You should have:

	A	B	C	D
1	Sex		Estimate	
2	Male:			
3		Less than high school	25,159	
4		High school graduate	34,472	
5		Some college or associate's	41,582	
6		Bachelor's degree	61,780	
7		Graduate or professional	85,202	
8	Female:			
9		Less than high school	16,151	
10		High school graduate	22,461	
11		Some college or associate's	28,643	
12		Bachelor's degree	41,952	
13		Graduate or professional	56,726	

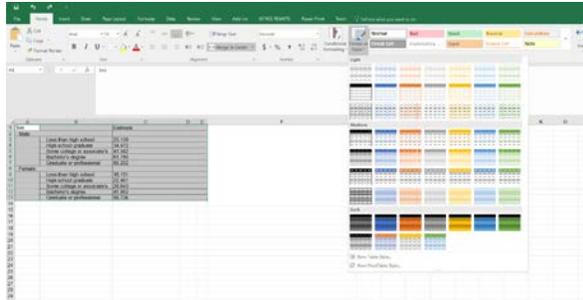
But what is with those green little triangles in the corner of the cells in column C? That is Excel telling you it has found something that could save you hours of frustration later on. Select those cells (or even just one of those cells) and you will see a little warning icon in the upper left of the selection.



Click on it and you will see that it is saying that these could be numbers or these could be text!
Choose convert to numbers.

Now we are ready to make this look nice.

Select your table and choose Format as Table



Get rid of those two extraneous row

Give Column B the title "Education Level"

Now you have a hierarchical table.

	A	B	C
1	Sex	Education Level	Estimate
2	Male:		
3		Less than high school	25,159
4		High school graduate	34,472
5		Some college or associate's	41,582
6		Bachelor's degree	61,780
7		Graduate or professional	85,202
8	Female:		
9		Less than high school	16,151
10		High school graduate	22,461
11		Some college or associate's	28,643
12		Bachelor's degree	41,952
13		Graduate or professional	56,726

There are charts in Excel 2016 that are particularly well suited to give you visual insights into data like this. A treemap chart provides a hierarchical view of your data and makes it easy to spot patterns. The tree branches are represented by rectangles and each sub-branch is shown as a smaller rectangle. The treemap chart displays categories by color and can easily show lots of data which would be difficult with other chart types.

Select your table

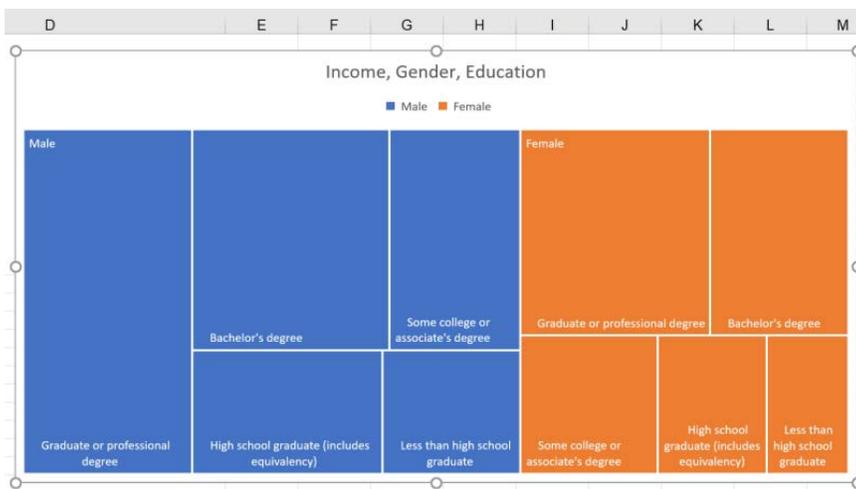
Click on Recommended Charts on the Insert tab of the ribbon

Scroll down and look at the previews

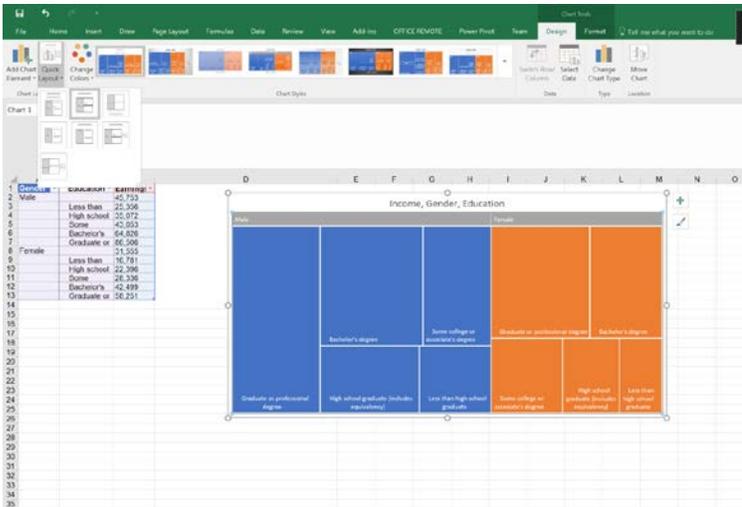
Choose the Treemap

Sex	Education Level	Estimate
Male:		
	Less than high school	25,159
	High school graduate	34,472
	Some college or associate's	41,582
	Bachelor's degree	61,780
	Graduate or professional	85,202
Female:		
	Less than high school	16,151
	High school graduate	22,461
	Some college or associate's	28,643
	Bachelor's degree	41,952
	Graduate or professional	56,726

When you expand the chart, you will see many of the same options you see in other charts. For example, you can click on Chart Title and give it a meaningful name.



From the ribbon, you can also experiment with styles and layouts by hovering over the options. That way, you do not commit to the look you like until you select it:



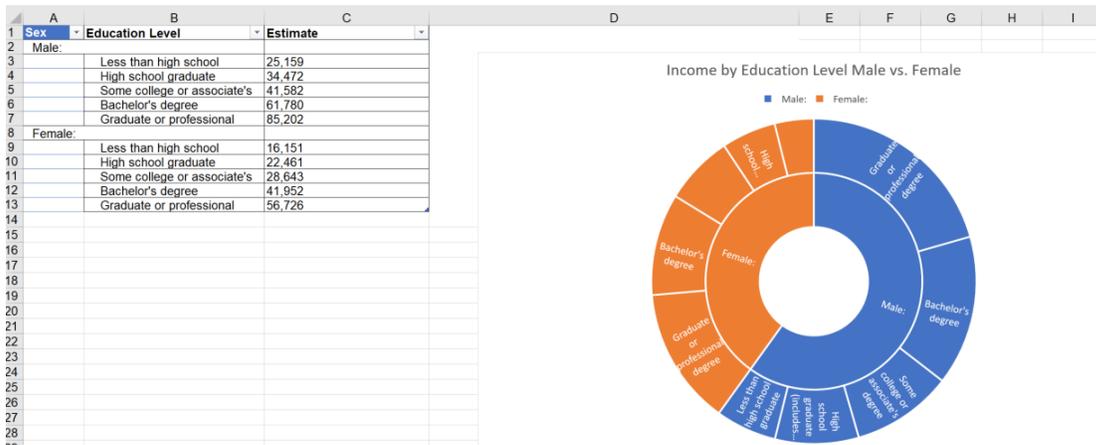
You can also Treemap charts are good for comparing proportions within the hierarchy, however, treemap charts aren't great at showing hierarchical levels between the largest categories and each data point. A sunburst chart is a much better visual chart for showing that. To see how this same chart looks as a sunburst chart:

Click on the treemap chart

From the Design tab on the Ribbon, choose Change Chart Type

Select Sunburst

The sunburst chart is actually quite similar to the donut chart. It is as if you had a donut chart inside a donut chart, each level reflecting a category of your information



If you want to move around the categories (for example, instead of earnings by education by gender, you want earnings by gender by level of education), a much more flexible and dynamic way

of looking at this data is through a pivot table. That is beyond the scope of this particular course, but it opens up new worlds for manipulating your data.

MODULE 4: Advanced Census Data Analysis in Excel

The vast majority of time, Excel as-is it provides you with what you will need to access, combine, query, massage, transform, visualize, and otherwise play with your data using the visualizations, formulas, and functions. There are, however, advanced tools for statistical analysis that you can turn on as an add-in. You do not need to download anything, you just need to turn it on. Doing so will give you tools for

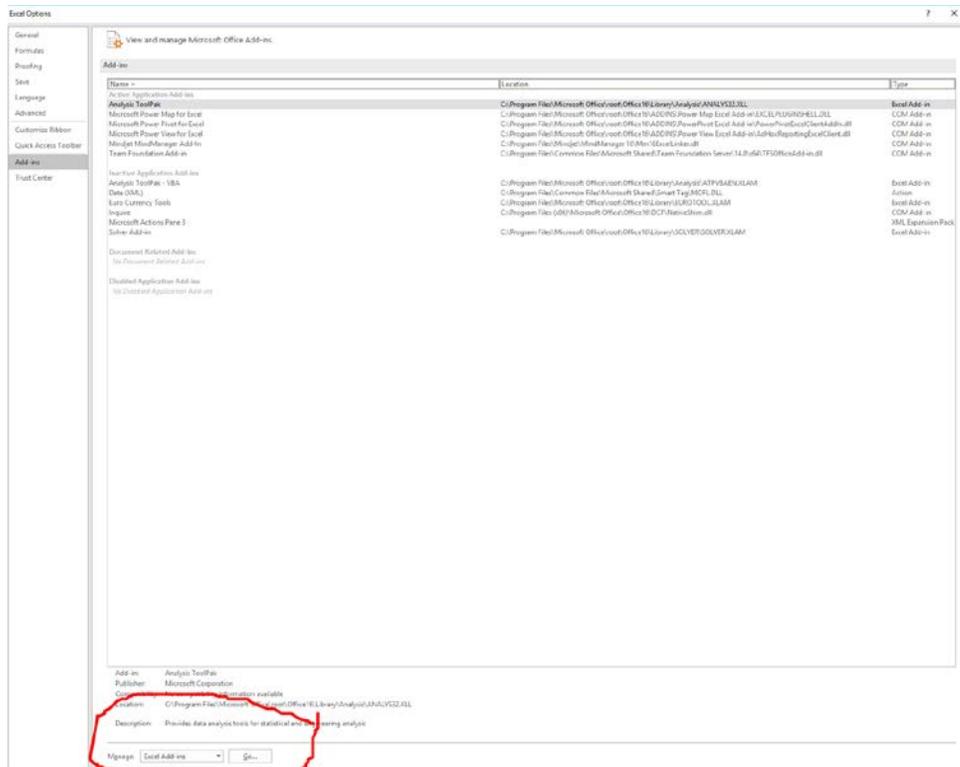
- Regression testing
- Histograms
- Anova, correlation, and covariance
- Exponential smoothing
- f-tests, z-tests, and t-tests
- Fourier analysis
- Moving averages, rank and percentile summaries
- Sampling and a more sophisticated random number generator than the standard function

Enabling the Add-in

Going through all of these statistical functions and methods is beyond the scope of this training, so I will just give you a basic idea of how it can be used with a few examples. We will enable the add on, and do some basic analysis on the Income spreadsheet.

To enable the Add-in, click File, Options, and choose Add-ins.

If you do not see the Analysis Tool Pack already in the list of add-ins, choose Excel-Add-ins from the Manage drop down in the bottom and hit Go



You will see a list of available add-ins. Check Analysis Tool Pack and click OK.
Now, when you click on the Data tab, you will see the Analysis Tool Pack.

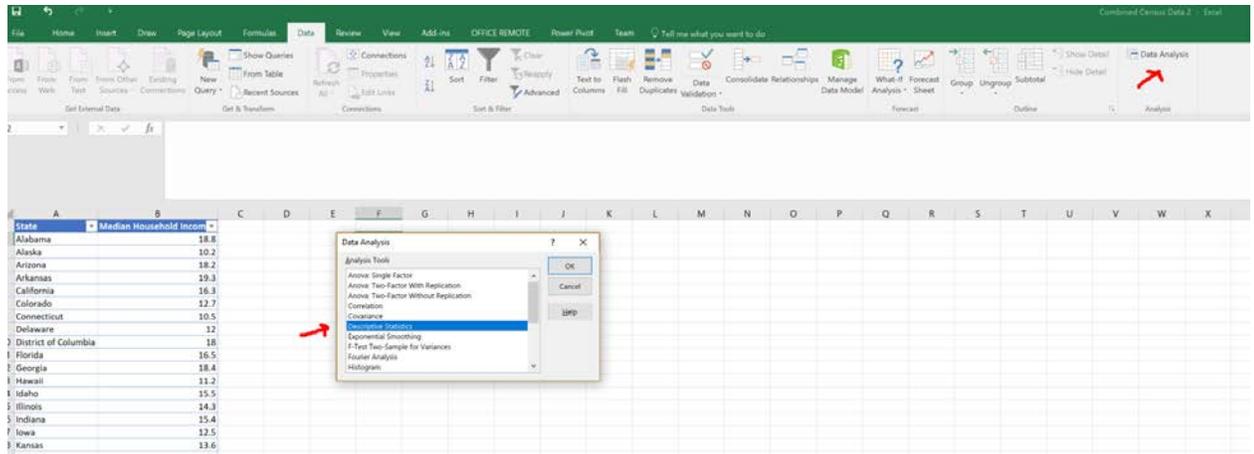
Descriptive Statistics

Descriptive statistics describe the main features of a Data set. They are the basic statistics you would generally want to start with, even if you are doing more sophisticated analysis of data. Descriptive statistics provide, for example, information about the central tendency, variability, and overall shape of your data in a few clicks.

Click in the Income table

Choose Data Analysis Tools from the Data tab

Choose Descriptive Statistics.



Give it the input range (which should be the median household income, including the column heading at the top, and Puerto Rico at the bottom).

For this example, have it group the data by column

Output the data to a New Worksheet Ply name it Descriptive Stats

Check Summary Statistics

Hit "OK"

You will now have a new tab called Descriptive stats. Again, these are useful for your first glance at the data to initiate your statistical analysis.

	A	B
1	Median Household Income	
2		
3	Mean	15.43461538
4	Standard Error	0.732574772
5	Median	15.3
6	Mode	11.5
7	Standard Deviation	5.282671807
8	Sample Variance	27.90662142
9	Kurtosis	20.2805969
10	Skewness	3.681717088
11	Range	36.6
12	Minimum	8.9
13	Maximum	45.5
14	Sum	802.6
15	Count	52
16		

Histogram

A histogram is a chart that tells you the number of occurrences of a value or range in a data set. You may recall from school that sometimes letter-grades were distributed in odd boundaries. Chances

are good that your teacher used a histogram to determine those letter-grade boundaries and the number of scores that fell in the range.

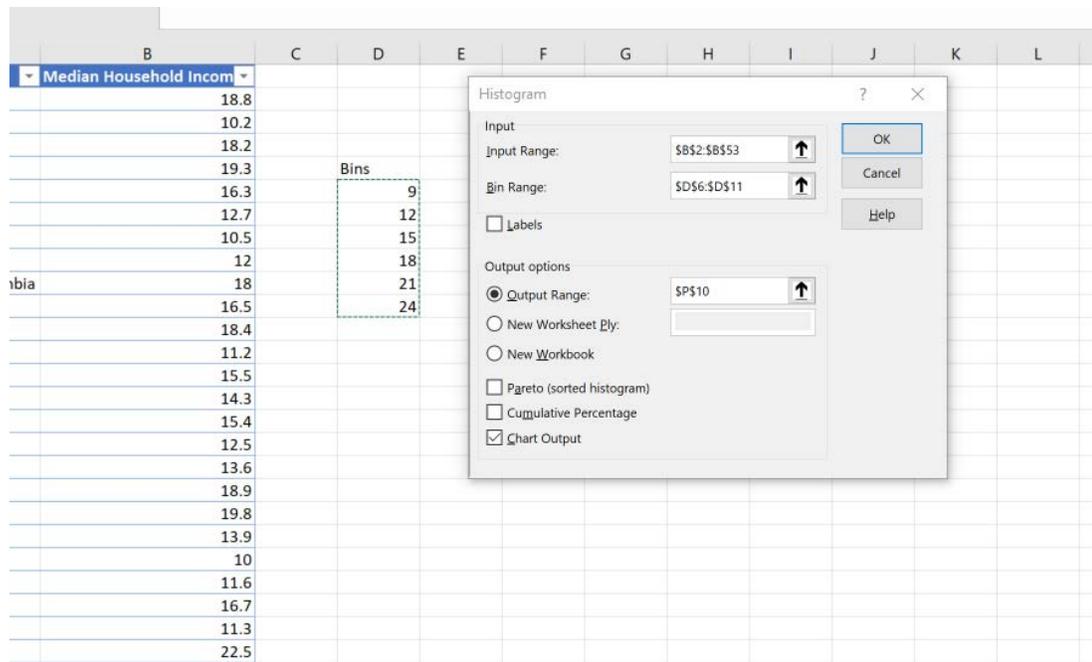
In this case, we are going to look at how the median income data is distributed. We will create a column of “bins” where we want the data to fit. For example, we want a bin representing median income of \$9,000 and under, another between \$9,001 and \$12,000, another between \$12,001 and \$15,000, etc.

Somewhere in the spreadsheet, enter in separate rows on a column the bin numbers 9, 12, 15, 18, 21, 24. I picked these ranges arbitrarily.

B	C	D	E
Median Household Income			
18.8			
10.2			
18.2			
19.3		Bins	
16.3		9	
12.7		12	
10.5		15	
12		18	
18		21	
16.5		24	
18.4			
11.2			
15.5			
14.3			
15.4			
12.5			
13.6			

Now, go to the data tab, and click on Data Analysis Tools.

Click Histogram and the following dialog box will appear:



Select the income data for the input range (no need to include the label, unless you want it)

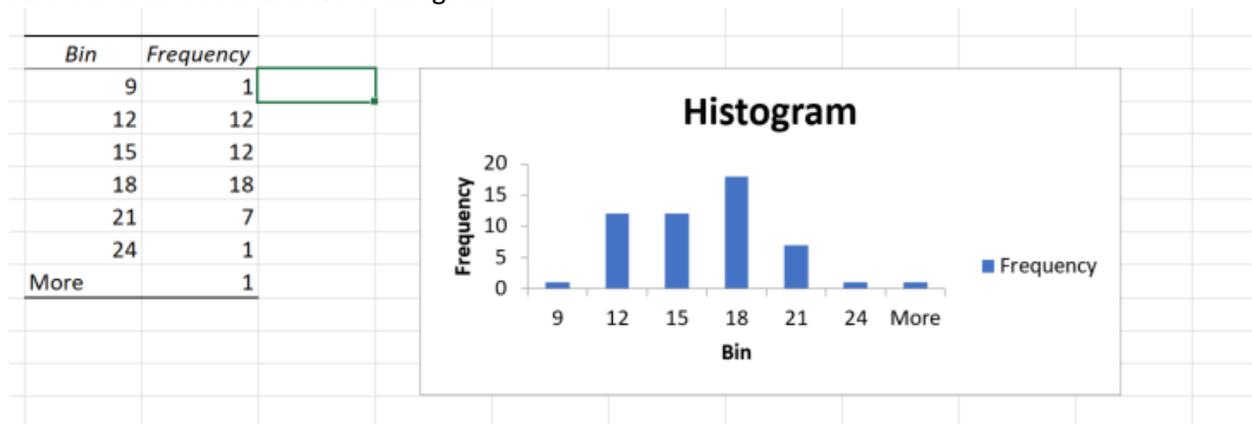
Select your Bins for the Bin

Select some empty space on your spreadsheet for the output range

Click the Chart Output box

Click OK

And that is it. You now have a histogram



Sampling

The Data Analysis tools make it super simple to get samples from a population. Although this is a very small amount of data (51 data points), you can imagine if you were doing research on a data set with hundreds, thousands, tens of thousands, etc. of data points. Getting a random sample

from such a large dataset is not that easy. So even though this is a trivial example, it is good to know how it is done.

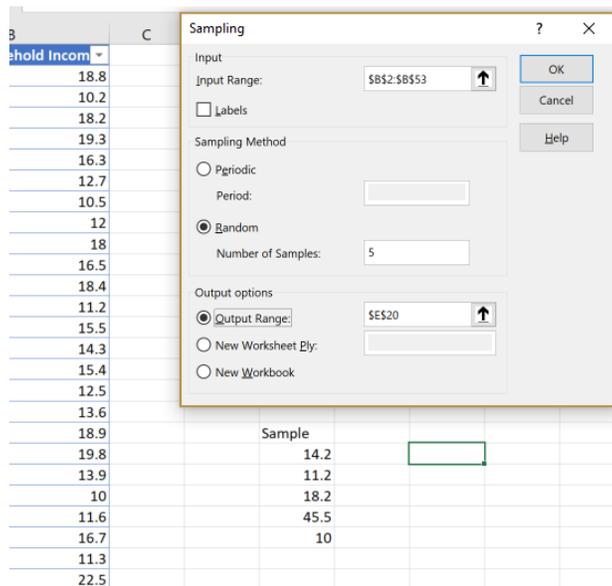
Choose Data Analysis Tools from the Data tab

Choose Sampling.

Give it the input range of state data

Tell it how big your sample size should be in Number of Samples. Try 5

Tell it where you want it to show the output



Now you have your sample, with no complex formula, no macros, and no repetitive operations required.

EXTRA MODULE: Mapping Census Data in Excel

(Not available on video)

Let's say we want to plot on a map the median household income for each state, along with the number of people in each state that have completed high school education.

We can leverage the Community Facts in the American Fact Finder again to find the data that we need. Let's start with a very simple table, a comparison of median household income per state found in the Income section.

Click on the Community Facts from the top menu on American Fact Finder

Click on Income on the left

Choose Compare States for Median Household Income

From the table page, choose Download

Select Use Data

Click Download

Unzip to your computer and open the CSV file



Again, apply your best practice of saving this CSV file as an Excel Workbook file right away.

File/Save As/Excel Workbook with a meaningful name

Next, we can delete columns we won't be using.

Delete the columns containing GEO.id, GEO.id2, GCT_STUB.target-geo-id, GCT_STUB.target-geo-id2), and the margin of error column, HC02.

Also delete that first data row for the overall United States (row 2).

Let's give our headers useful names. Name the header of Column A "State" and Column B, "Median Household Income".

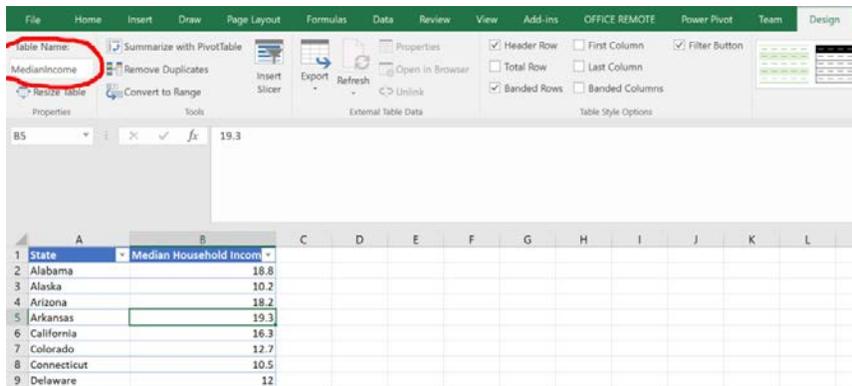
State	Median Household Income
Alabama	18.8
Alaska	10.2
Arizona	18.2
Arkansas	19.3
California	16.3
Colorado	12.7
Connecticut	10.5
Delaware	12
District of Columbia	18
Florida	16.5
Georgia	18.4
Hawaii	11.2
Idaho	15.5
Illinois	14.3
Indiana	15.4
Iowa	12.5
Kansas	13.6
Kentucky	18.9

Next, let's turn this into a table to access all the table features.

Click on Format As Table

Pick your favorite style, and make sure the "My Table Has Headers" box is checked.

Give your table a meaningful name by changing the "Table1" in the Table Name Box (in the Table Tools Design tab) to something meaningful:



Let's find the high school education data and put it on another tab in this workbook

Click on Community Facts

Select Education

Click on Compare States for People Who have Completed High School

United States
Bureau

AMERICAN
FactFinder

MAIN COMMUNITY FACTS GUIDED SEARCH ADVANCED SEARCH DOWNLOAD CENTER

Community Facts - Find popular facts (population, income, etc.) and frequently requested data a

Enter a state, county, city, town, or zip code: GO

Population
Age
Business and Industry
Education
Governments
Housing
Income
Origins and Language

United States

Educational Attainment: Percent high school graduate or higher
86.7% Source: 2011-2015 American Community Survey 5-Year Estimates

Popular tables for this geography:

2015 American Community Survey

- Educational Attainment (High School, Bachelor's, Advanced Degree, ...)
- School Enrollment (Preschool, Kindergarten, High School, College, ...)
- Median Earnings by Educational Attainment
- Compare States for People Who Have Completed High School
- Compare Cities and Towns for People Who Have Completed High School

Census 2000

Again, if your data transformations are simple, you can do them before you even download the data.

With the table up, click on the Modify Table button.

2011-2015 American Community Survey 5-Year Estimates

Table View

Actions: [Modify Table](#) [Bookmark/Save](#) [Print](#) [Download](#)

Tell us what you think. [Provide feedback to help make American Community Survey data more i](#)

Although the American Community Survey (ACS) produces population, demographic and housing

Geography:

Versions of this table are available for the following years:
2015
2014
2013
2012
2011
2010
2009

Geographic Area	Percent	Margin of Error
1 United States	86.7	+/-0.1
53 Alabama	84.3	+/-0.2
53 Alaska	92.1	+/-0.3
53 Arizona	86.0	+/-0.1
Arkansas	84.8	+/-0.2
California	81.8	+/-0.1
Colorado	90.7	+/-0.1
Connecticut	89.9	+/-0.1
Delaware	88.4	+/-0.3
District of Columbia	89.3	+/-0.3
Florida	86.9	+/-0.1
Georgia	85.4	+/-0.1
Hawaii	91.0	+/-0.2
Idaho	89.5	+/-0.2
Illinois	87.9	+/-0.1
Indiana	87.8	+/-0.1
Iowa	91.5	+/-0.1
Kansas	90.2	+/-0.2

Again, you have the ability to modify the table by deleting elements, moving elements, even transposing rows and columns.

We are simply going to uncheck the Margin of Error column.

ns: Hide Table Tools | Bookmark/Save | Print | Download | Create a Map

Table Tools: Reset Table | Show Hidden Rows/Columns | Transpose Rows/Columns ?

Click 'Back to Search' to select other tables or geographies

Legend: show/hide rows and columns | collapse/expand data categories | rearrange columns | rearrange rows | sort ascending/descending | filter rows

What you think. [Provide feedback to help make American Community Survey data more useful for you.](#)

When the American Community Survey (ACS) produces population, demographic and housing unit estimates, it is the Census Bureau's Population Estimates

Geography: United States

53 of 53

Geographic Area	Percent	Margin of Error
United States	86.7	+/-0.1
Alabama	84.3	+/-0.2
Alaska	92.1	+/-0.3
Arizona	86.0	+/-0.1
Arkansas	84.8	+/-0.2
California	81.8	+/-0.1
Colorado	90.7	+/-0.1
Connecticut	89.9	+/-0.1
Delaware	88.4	+/-0.3
District of Columbia	89.3	+/-0.3
Florida	86.9	+/-0.1
Georgia	85.4	+/-0.1

Download as before, and extract the file into Excel as a CSV file.

Remember to save this as an Excel Workbook right away.

Delete columns A through F, as before.

Delete the row showing the total for the United States

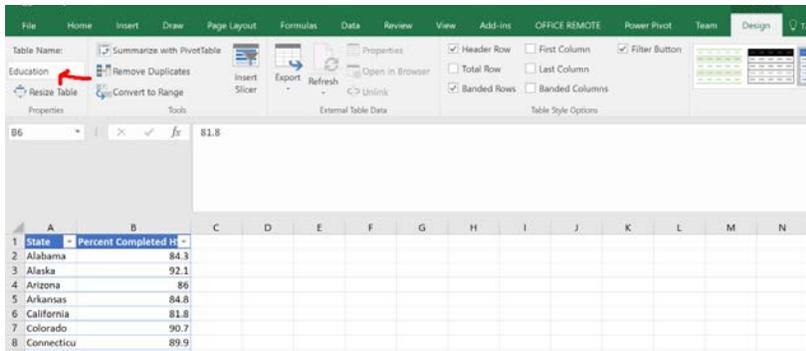
Names your headers State (or StateEducation), and Percent Completed HS.

A	B	C
State	Percent Completed HS	
Alabama	84.3	
Alaska	92.1	
Arizona	86	
Arkansas	84.8	
California	81.8	
Colorado	90.7	
Connecticut	89.9	
Delaware	88.4	
District of C	89.3	
Florida	86.9	
Georgia	85.4	
Hawaii	91	
Idaho	89.5	
Illinois	87.9	
Indiana	87.8	
Iowa	91.5	
Kansas	90.2	
Kentucky	84.2	
Louisiana	83.4	
Maine	91.6	
Maryland	89.4	
Massachus	89.8	

Click on Format as Table from the Styles section of the Home tab on the ribbon.

Ensure that "My table has headers" is checked, and hit ok.

Give this table a meaningful name by changing Table Name in the Table Tools Design tab of the ribbon from Table1 to something you like.

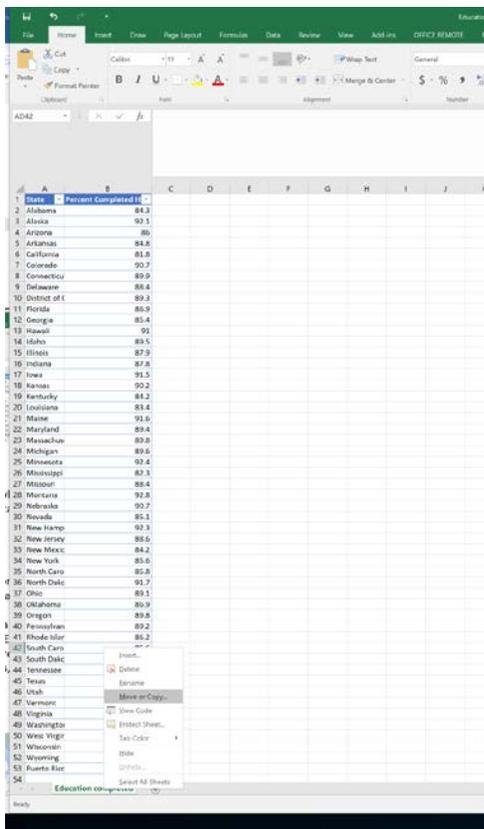


To minimize jumping in and out of multiple workbooks, move this spreadsheet into the workbook with the Income data.

Right click on the tab at the bottom with the name of your current spreadsheet.

I've named mine "Education"

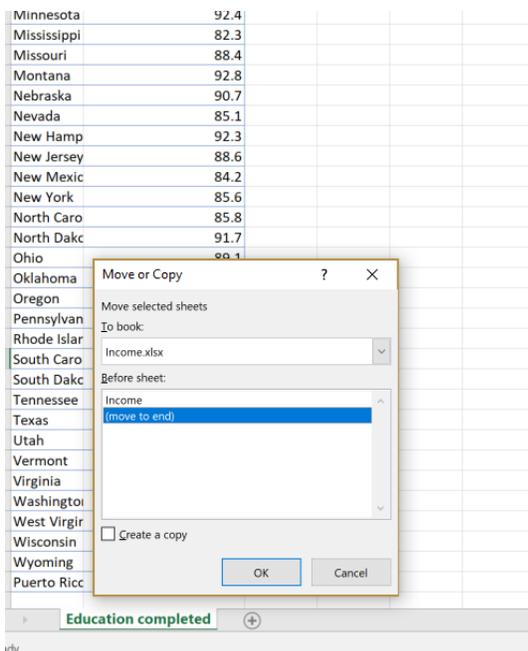
Choose Move or Copy...



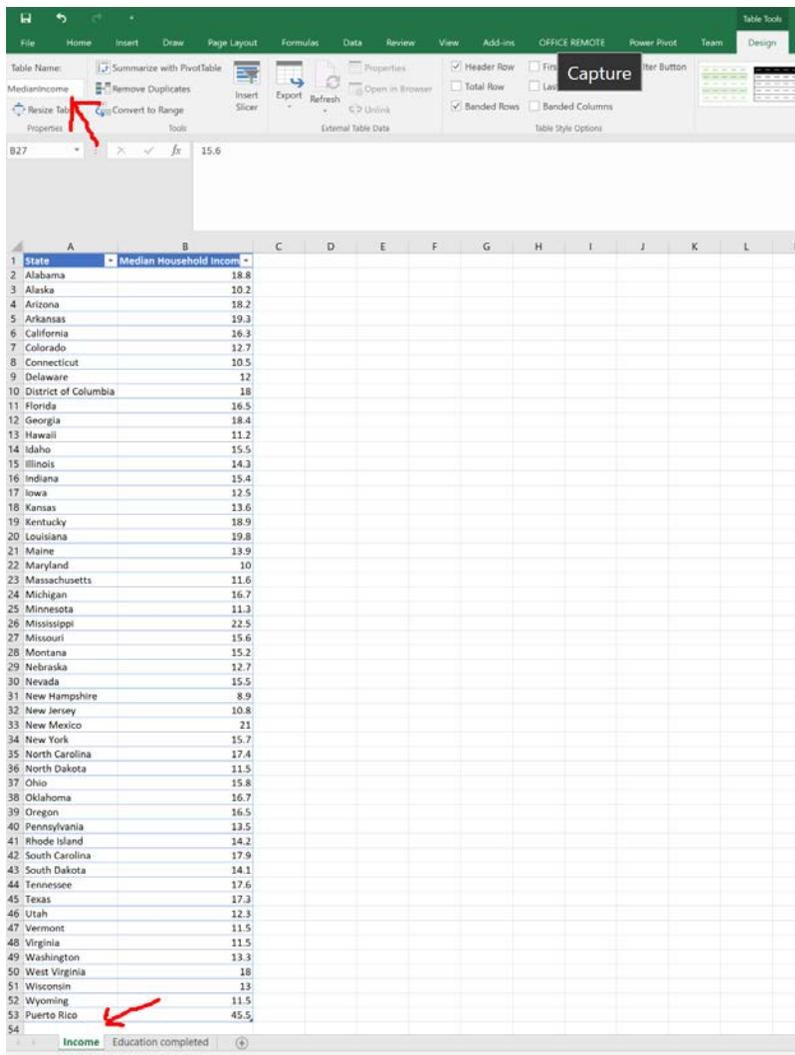
You will see that you can change the workbook you can copy or move this to.

Click on the drop down and select the first spreadsheet you had been working on.

I named mine "Income".



Now you will have one workbook with two tables. You can verify that the tables have been named the way you want by choosing the tab with the data, clicking in the table, choosing the Design tab on the ribbon and looking at the table name:



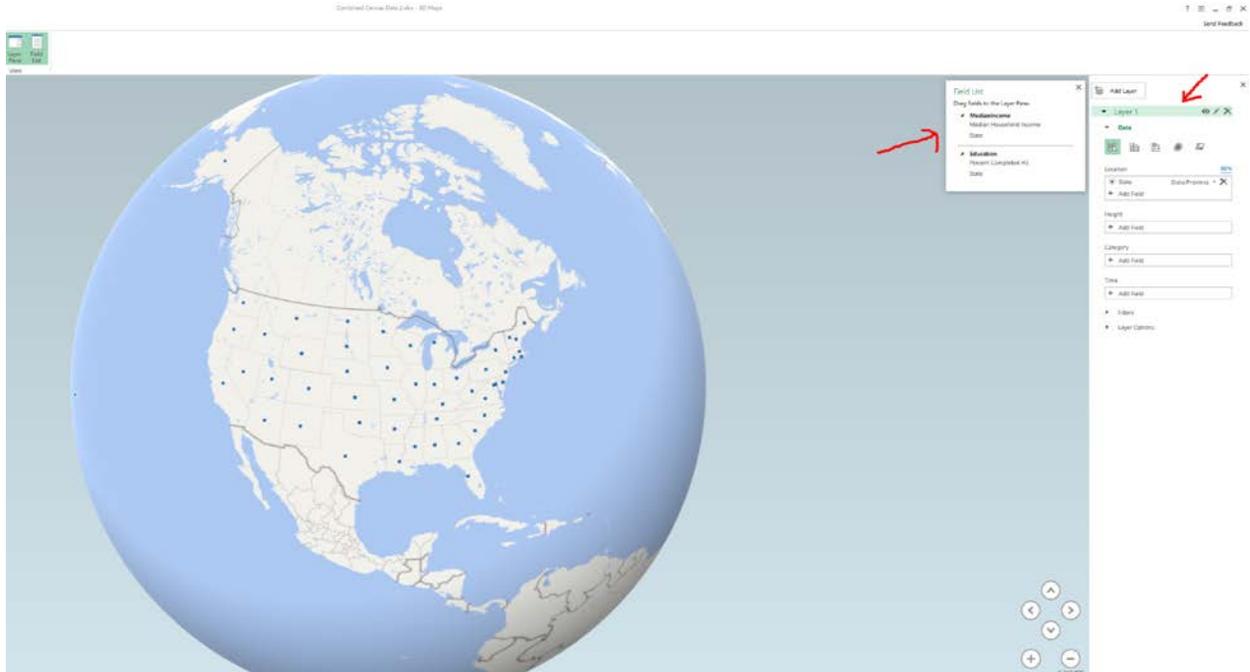
Now we are going to create a 3D Map to visualize both the income and the education levels across the states. 3D Maps allow you to take geographical (and time) data and turn them into maps, and even create videos to share and post those maps.

Click 3D Map from the Insert tab

First time you do this with your data, it will say on the bottom "loading model" and this may take a while for very large data sets, but it only happens once. When it says loading model, it is referring to a data model.

A Data Model is an approach for integrating data from multiple tables. Effectively, what you are doing is building a relational database inside an Excel workbook, without having to have database expertise. Within Excel, Data Models are used pretty much without you having to know anything about them. They are used in Maps, PivotTables, Pivot Charts, and Power View reports, etc. But today, we are just going to build a map.

You will have a screen that looks like this:



Let's talk about some terminology. On the upper right, you will see the fields in your tables. You should now be able to appreciate the fact that you named your tables something useful!

Layers

To the right of that are the options for the *layer*. You create layers that show specific data on your map (like the number of graffiti requests). You can combine layers on top of layers, and then build *scenes* from those layers. When you have a time dimension (such as something that happens every day, week, month, etc), map scenes are like scenes in a movie, with transition effects and all. Then, you can combine scenes into a tour. Here we only have spatial data.

The first decision you need to make is the type of map we want to create. We can choose a column chart, a stacked column chart, a bubble chart, a heat map, a shaded region, even add a custom map type. Let's start with a stacked column chart so that we can have a column whose height represents the median income for the state.

The second decision you need to make is which fields in your table you want to use to denote location. Excel Maps make it easy to do, but it is an important question to consider. It relates to the question of what you want your analysis to answer. In our case, want to review data at the state level. But Census is rich with data for cities, counties, metropolitan statistical areas, census tracts, etc. Some datasets may even contain latitude/longitude coordinates. Excel's 3D maps can handle all of them.

In this case, we are going to use the state data for median income. Excel has already predicted that you want to use State, and that the field named State refers to a US state. Go ahead and delete that by hitting the x next to it so you can see how to do this manually.

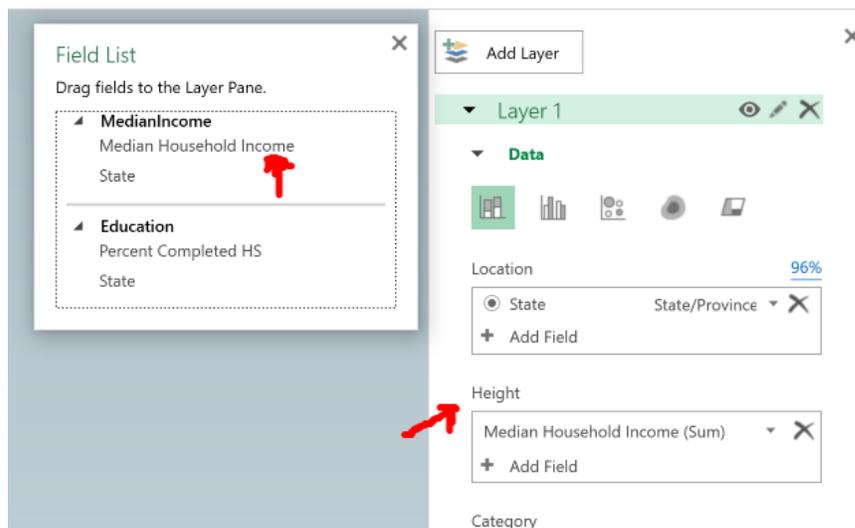


You could add the geography field this way:

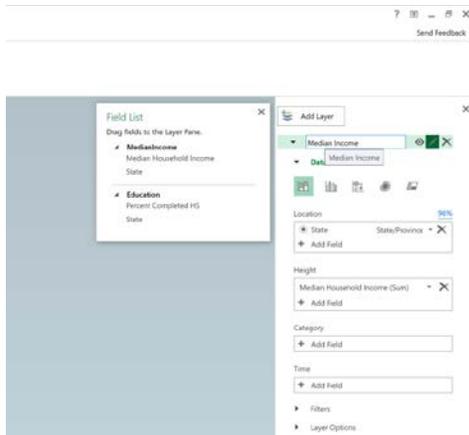
In the Location box of the Layer options, choose +Add Field, and add State from the Median Income table

Now, let's add data to represent the height of that column.

In the Location box of the Layer options, choose +Add Field, and add State from the Median Income table



Just as with tables, it is always a best practice to name your layers. Whatever you name your layer will also show up in a legend that gets created automatically for you. In the upper part of the box, click on the pencil to the right of the label Layer 1 and give it a meaningful name. We can call it Median Income:



To make it obvious which state is which, use Map Labels:
Click on Map Labels from the ribbon

Take a few minutes to explore this map.

Zoom in using the mouse wheel or using the + and - buttons

Tilt up and down using the arrows

Generally explore the map by moving around

When you hover over one of the data points, you see text from that data point in the form of what is called a custom card. You can create custom cards that show all kinds of data. You can also add your own annotations. Let's annotate Mississippi and point out that it has the highest median household income.

Right click on the column in Mississippi and choose Add annotation

In Title, type Highest Median Income

In description, note that you can type anything you want, add a field, or insert an image.

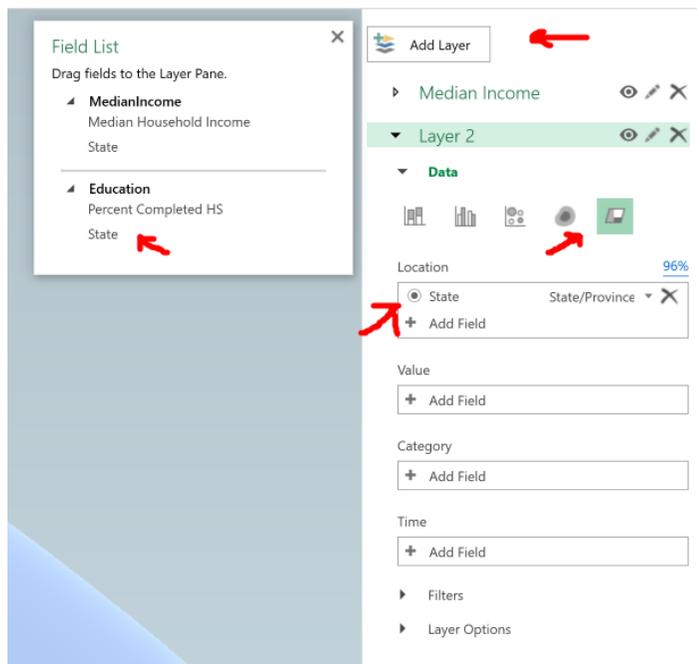
Here we will add a field, and select both data points.

Feel free to add other annotation. Let's add a second layer with a different chart type, this one a shaded region chart.

Click on the Add Layer button to get your next layer. Note that the first layer will still be visible (unless you click on the little eye icon next to the layer name).

In the Location box of the Layer options, choose +Add Field, and add State from the Education table

Change the chart type to the Shaded Region chart



Now, let's add data to represent the shading within the region.

In the Value box of the Layer options, drag Percent Completed HS into the value field

You now have a visualization that shows you which states have a relatively higher percentage of people completing high school. But you can make it even more clear.

Click on Layer Options and change the color scale to 20%

Rename the layer from "Layer 2" to something more relevant like "Percent Completed High School"

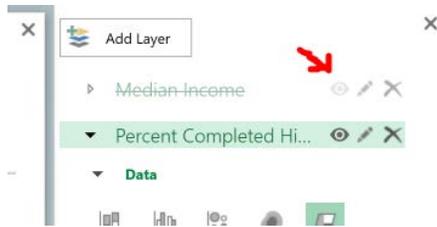
The image shows a software interface for map layer management. On the left, a 'Field List' panel allows users to drag fields to the Layer Pane. It lists two main categories: 'MedianIncome' (with sub-items 'Median Household Income' and 'State') and 'Education' (with sub-items 'Percent Completed HS' and 'State'). On the right, the 'Layer Pane' shows two layers: 'Median Income' and 'Percent Completed Hi...'. The 'Percent Completed Hi...' layer is selected and expanded to show various configuration options: Location (set to 'State' with a '96%' indicator), Value (set to 'Percent Completed HS (Sum)'), Category, Time, Filters, Layer Options (Color Scale at 19%, Opacity at 100%, Color set to 'Percent Completed HS (Sum)'), Show Values (Zeroes, Negatives, Nulls), and Data Card (Customize).

On the lower left, you will see the legends clearly.
Move them it so that it is in the upper left (in the sky).

Scenes

Scenes allow you to visualize your map like a movie. You can even add a soundtrack! In this case, we are going to build two scenes: one that shows just the shaded region map that denotes the percent completed high school, and then a second scene that adds the median income. We will combine the scenes into what is called a “tour”, and each scene will have different visual effects to allow you to see it better.

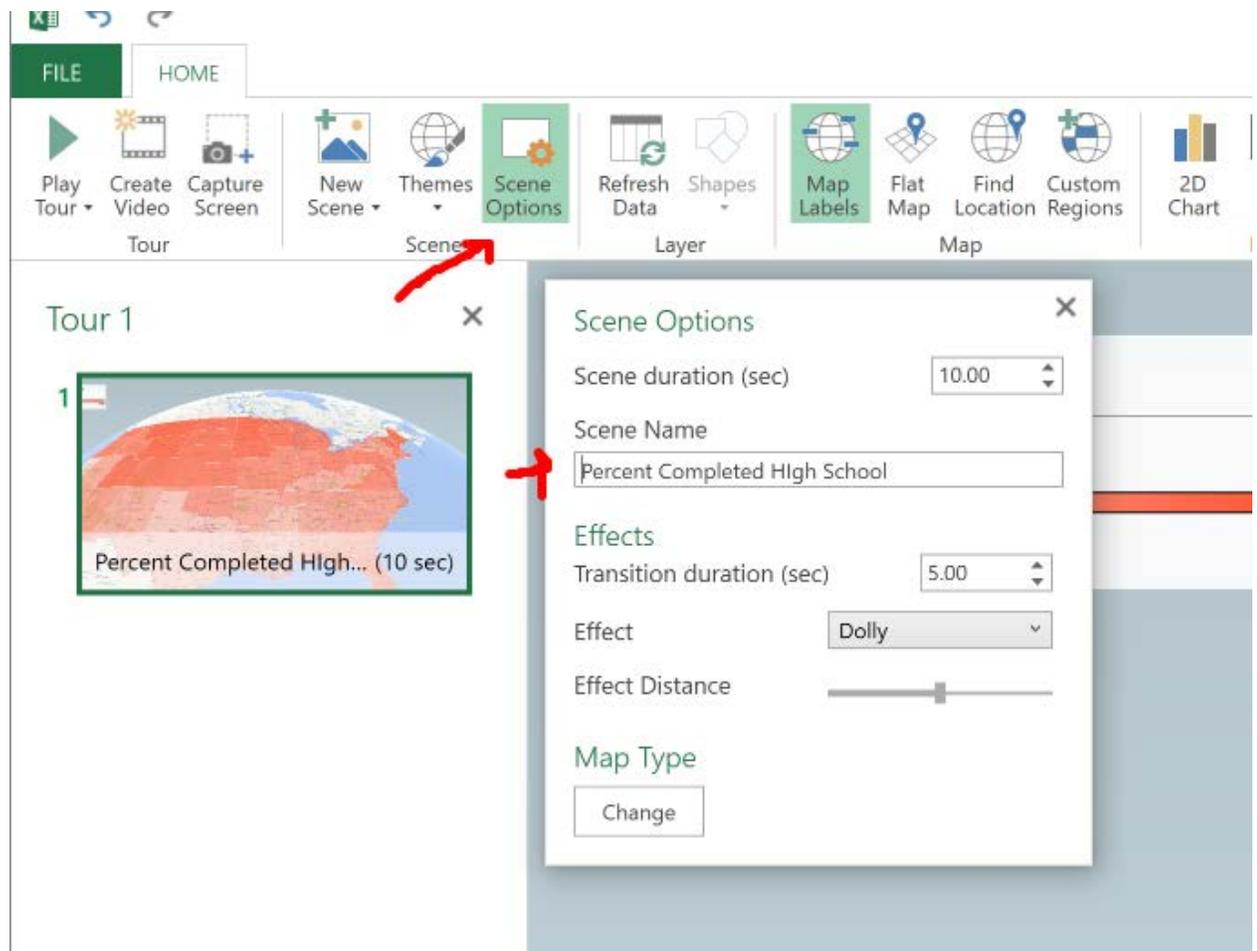
In the Layer Pane, click on the eye icon to hide the Median Income layer.



On the left side of the screen, you will see your tour pane with one scene.

Click on Scene Options in the ribbon

Give the scene the name Percent Completed High School



Change the transition duration to 5 seconds. This is how much time the scene spends on this particular view of your map.

Assign an effect (I suggest Dolly).

You can test your work by clicking on the Play Tour button on the ribbon towards the left side.

Now let's create our second scene. We will add to this shaded region the column chart of median income.

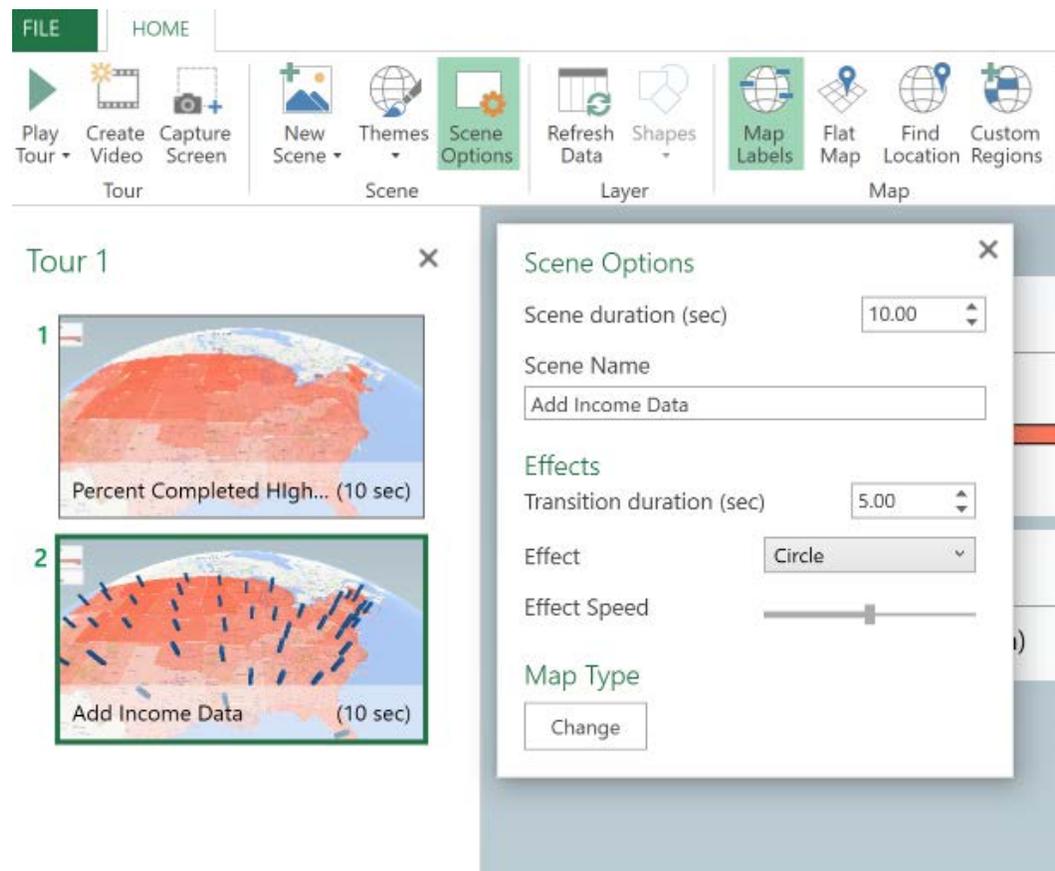
Choose New Scene from the ribbon (or right click on the first scene)

Unhide the Median Income Layer by clicking on the eye icon

Click on Scene Options and name this scene "Add Income Data"

Change the transition duration to 5 seconds

Assign a different effect (I suggest Circle)



Play the tour

You can see that when the income data points come in, the movement of the map changes.

Custom Maps and Local Census Data

This section brings together many of the skills you have learned so far, and adds two more. Census data is not just about national statistics. There is an abundance of data at the state,

county, city and town level. You can also access data for areas below these levels such as zip codes, census tracts, block groups, and blocks. But how would you know what a census tract looks like, and how it relates to other geographic areas? This is where custom maps come in.

Let's say we want to map both the 10 year population change and the 10 year housing change for all of the census tracts in the state of Illinois. We need two things. First, we need the data. Second, we need a way to tell Excel where those census tracts are on a map. Census has both of these.

First, let's get the data and bring it into Excel. We are looking for 2000 to 2010 census tract population changes. You can use the search options in American Fact Finder. But for simplicity, the spreadsheet can be loaded here:

<http://www2.census.gov/programs-surveys/decennial/tables/time-series/tract-change-00-10/censustract-00-10.xlsx>

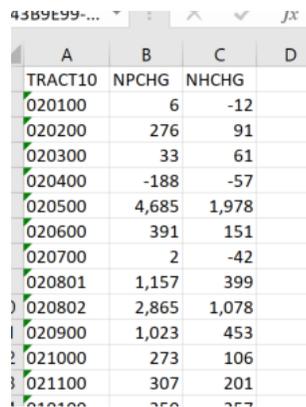
Open this spreadsheet and we will simplify it.

Delete all columns but TRACT10, NPCHG, and NHCHG.

These are the census tract number, the population change over 10 years, and the housing change over 10 years.

Format it as a table in whatever style you like

Save it as an Excel Workbook somewhere on your computer



TRACT10	NPCHG	NHCHG
020100	6	-12
020200	276	91
020300	33	61
020400	-188	-57
020500	4,685	1,978
020600	391	151
020700	2	-42
020801	1,157	399
020802	2,865	1,078
020900	1,023	453
021000	273	106
021100	307	201

Now, we need something that tells us what all those tracts actually look like on a map of the state of Illinois. Census maintains files called KML or SHP files that can be used for this purpose.

From Census.gov, choose Geography from the menu bar at the top of the page

Choose Maps & Data

Choose TIGER products

Click on KML-Cartographic Boundary Files.

These are the files that you can use not only in Excel, but a variety of mapping tools.

Click on Census Tract under State-based Files.

Select Illinois

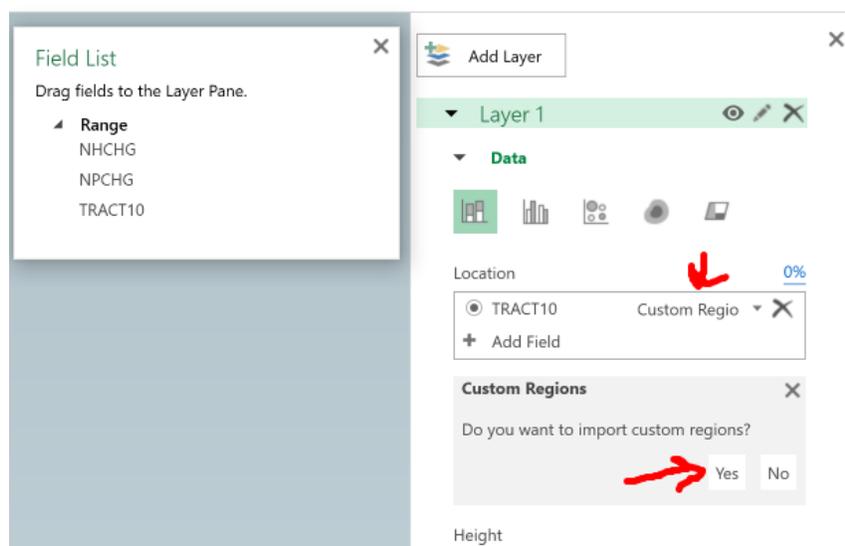
This will give you a zip file that you should extract to your computer. The one we are most interested in is the KML file. We will import that into Excel.

Now, to bring it all together!

In the spreadsheet you just created, choose 3D Map from the Insert tab and add a map.

For location, choose TRACT10. But you need to tell Excel what a tract looks like, right?

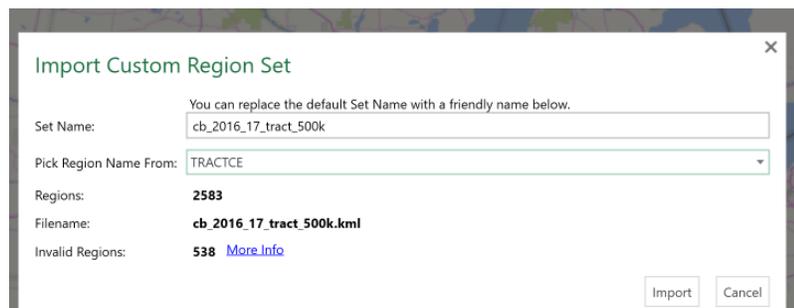
Set the location type to be a custom region and you will be prompted to import the region.



Click Yes and point it to the KML file that you downloaded and unzipped.

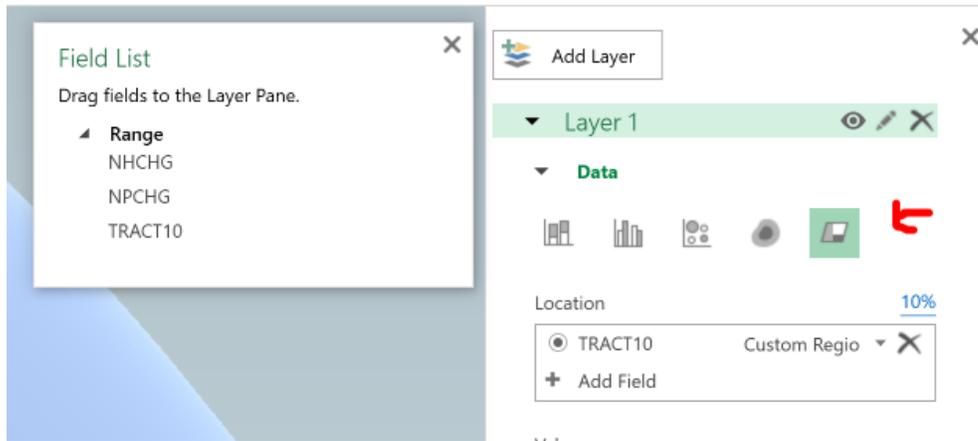
Give it a friendly name, if you wish

Now, the region name in this KML file is TRACTTCE, so select that



Now, let's give it a more professional look.

In the map type, choose the shaded region type.

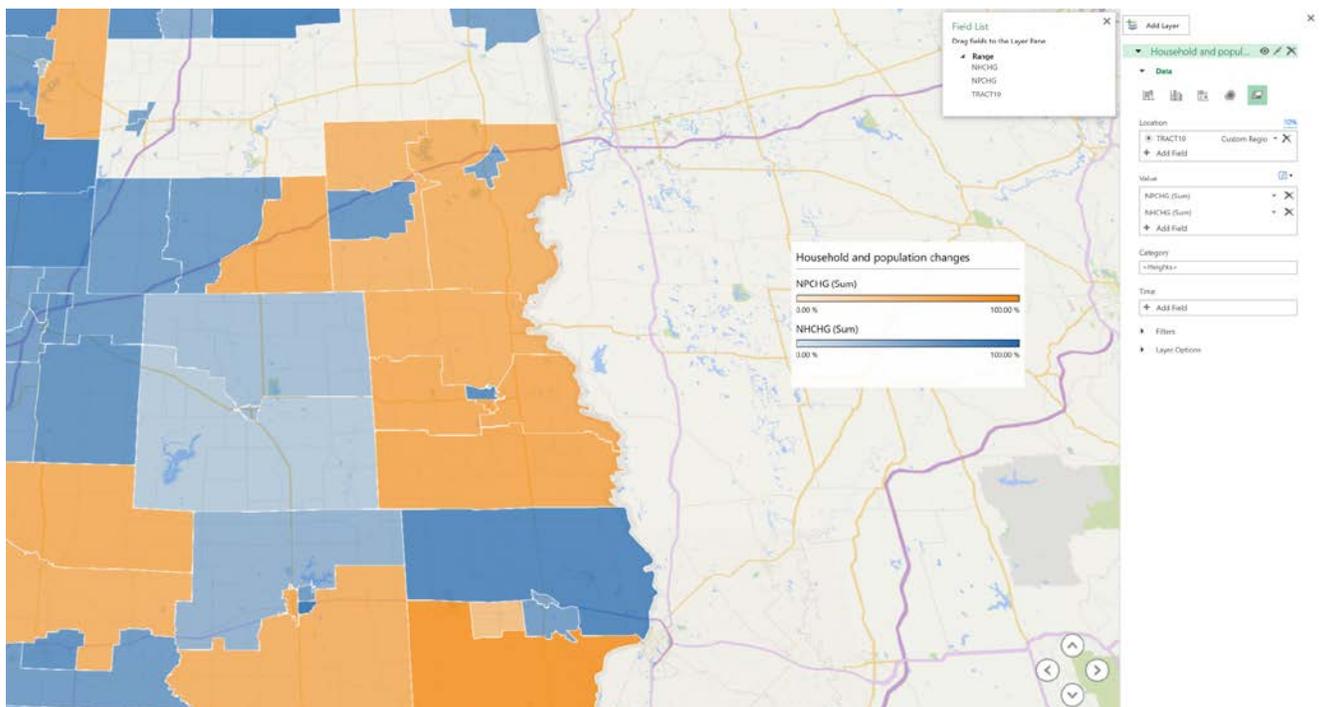


You are starting to see the census tracts come alive.

For the final flare, drag NHCHG (number household change) and NPCHG (number population change) to the value field.

Name the Layer something useful

You now have a custom map with data by census tract



Where to go from here

This course has barely scratched the surface both in terms of the useful data provided by US Census and the tools in Excel that you can use to work with that data. We explored just a few data sets and a few ways to use it. However, the use cases are almost limitless. Here are just a few examples of how data from the Census Bureau are used:

- Decision making at all levels of government.
- Forecasting future transportation needs for all segments of the population.
- Planning for hospitals, nursing homes, clinics and the locations of other health services as well as for distributing medical research.
- Designing public safety strategies and creating maps to speed emergency services to households in need of assistance.
- Planning health and educational services for people with disabilities.
- Attracting new businesses to state and local areas.
- Planning for school construction and other projects.

Further, Census provides a myriad of tools for finding the data that you need, accessing it with a variety of applications and devices, and incorporating it into research and other data-heavy activities. Go to <http://census.gov> to discover more.

Excel provides numerous easy to use tools for taking that data and combining it with other data, manipulating the data through formulae and functions, visualizing the data, pivoting it to understand the data in new ways, and sharing the data with a large number of other data sources and stores. It provides ways to help users not only gain insights from the data, but also to collaborate on it. I hope that this course has inspired you to explore further on your own.