


Chapter 7

Joining Maps to Other Datasets in QGIS



Skills you will learn: How to join a map layer to a non-map layer in preparation for analysis, based on a common joining field shared by the two tables.

If you are unfamiliar with the basic functionality of QGIS, such as how to add map layers and other data tables to the map document, please review the tutorial **A Quick Tour of QGIS Desktop**, which you can access by clicking [here](#).

Getting started

Add the map layer and the non-geographic layer to the data frame. For the purposes of this illustration, we are using a shapefile of census tracts in Winnipeg, Manitoba, Canada and a dataset of median household income from the census. This is what the attribute table looks like for the map layer.

| | CTUID | CMAUID | PRUID |
|----|------------|--------|-------|
| 0 | 6020001.00 | 602 | 46 |
| 1 | 6020002.00 | 602 | 46 |
| 2 | 6020003.00 | 602 | 46 |
| 3 | 6020004.01 | 602 | 46 |
| 4 | 6020004.02 | 602 | 46 |
| 5 | 6020005.00 | 602 | 46 |
| 6 | 6020006.00 | 602 | 46 |
| 7 | 6020007.00 | 602 | 46 |
| 8 | 6020008.00 | 602 | 46 |
| 9 | 6020009.00 | 602 | 46 |
| 10 | 6020010.00 | 602 | 46 |
| 11 | 6020011.00 | 602 | 46 |
| 12 | 6020012.00 | 602 | 46 |
| 13 | 6020013.00 | 602 | 46 |
| 14 | 6020014.00 | 602 | 46 |
| 15 | 6020015.00 | 602 | 46 |
| 16 | 6020016.00 | 602 | 46 |
| 17 | 6020017.00 | 602 | 46 |
| 18 | 6020018.00 | 602 | 46 |
| 19 | 6020019.00 | 602 | 46 |

And this is what the csv data table looks like in the data table when we open it in Excel:

| Geography | Tract | Total_Private_household | Median_2005_household_income | Median_2005_aftertax_household_income | Average_2005_household_income | Standard_error_of_average |
|---------------------------|------------|-------------------------|------------------------------|---------------------------------------|-------------------------------|---------------------------|
| 0001.00 (602000100) 01000 | 6020001 | 1850 | 53436 | 45133 | 63800 | |
| 0002.00 (602000200) 00000 | 6020002 | 2400 | 41184 | 36694 | 49562 | |
| 0003.00 (602000300) 01010 | 6020003 | 2630 | 39250 | 34345 | 45973 | |
| 0004.01 (602000401) 00000 | 6020004.01 | 2305 | 45270 | 39247 | 52837 | |
| 0004.02 (602000402) 01000 | 6020004.02 | 1700 | 36937 | 33134 | 41922 | |
| 0005.00 (602000500) 00000 | 6020005 | 2335 | 75475 | 60342 | 90117 | |
| 0006.00 (602000600) 01000 | 6020006 | 2630 | 43616 | 37830 | 51066 | |
| 0007.00 (602000700) 00010 | 6020007 | 1595 | 52045 | 43546 | 61740 | |
| 0008.00 (602000800) 00000 | 6020008 | 1140 | 91369 | 70850 | 122958 | |

Making the join

Eventually, we will join “Tract”, column B, with the first column in the census tract attribute table that we’ve opened in QGIS, as displayed in the first screen grab.

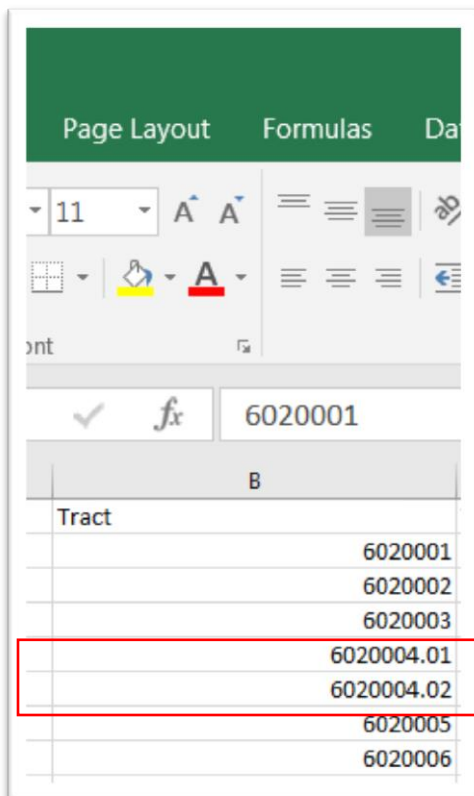
However, we must add a decimal and two zeros to the census tract ID numbers in column B above. Failure to do so, means that tracts without decimal points will not be joined to their corresponding tracts in the census tract table.

And unlike ArcGIS, QGIS will import the values in column B as numbers. Under normal circumstances this would be fine. However, you’ll notice that the census tract field in QGIS, CTUID, is left-justified, meaning that it’s text. In order for the join to happen, the corresponding field in the csv file must be the same datatype.

So we’ll have to carry out two steps: run two functions to add two decimal points to the census tract numbers in the csv file, and then convert the column to text; then we must create what’s called a csvt file that essentially instructs QGIS to import the values in that column as text.

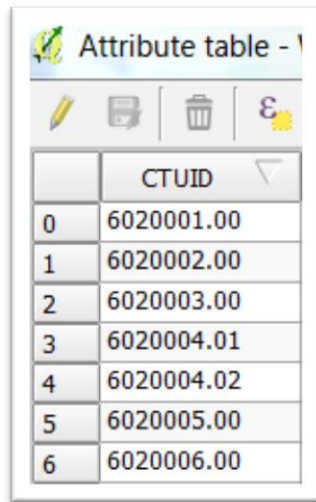
Adding the decimal point and two zeros

Take a closer look at the Tract column in our csv file.



| B | |
|-------|------------|
| Tract | |
| | 6020001 |
| | 6020002 |
| | 6020003 |
| | 6020004.01 |
| | 6020004.02 |
| | 6020005 |
| | 6020006 |

The values in the first three rows contain no decimal places. The values in the fourth and fifth rows do. Now let's look at the corresponding values in the census tract file that we've opened in QGIS's attribute table.



The screenshot shows the 'Attribute table' window in QGIS. The table has a header row with the column name 'CTUID'. Below the header, there are seven rows of data. The first three rows have values ending in '.00', the fourth and fifth rows have values ending in '.01' and '.02' respectively, and the last two rows have values ending in '.00'. The values are left-justified in the cells.

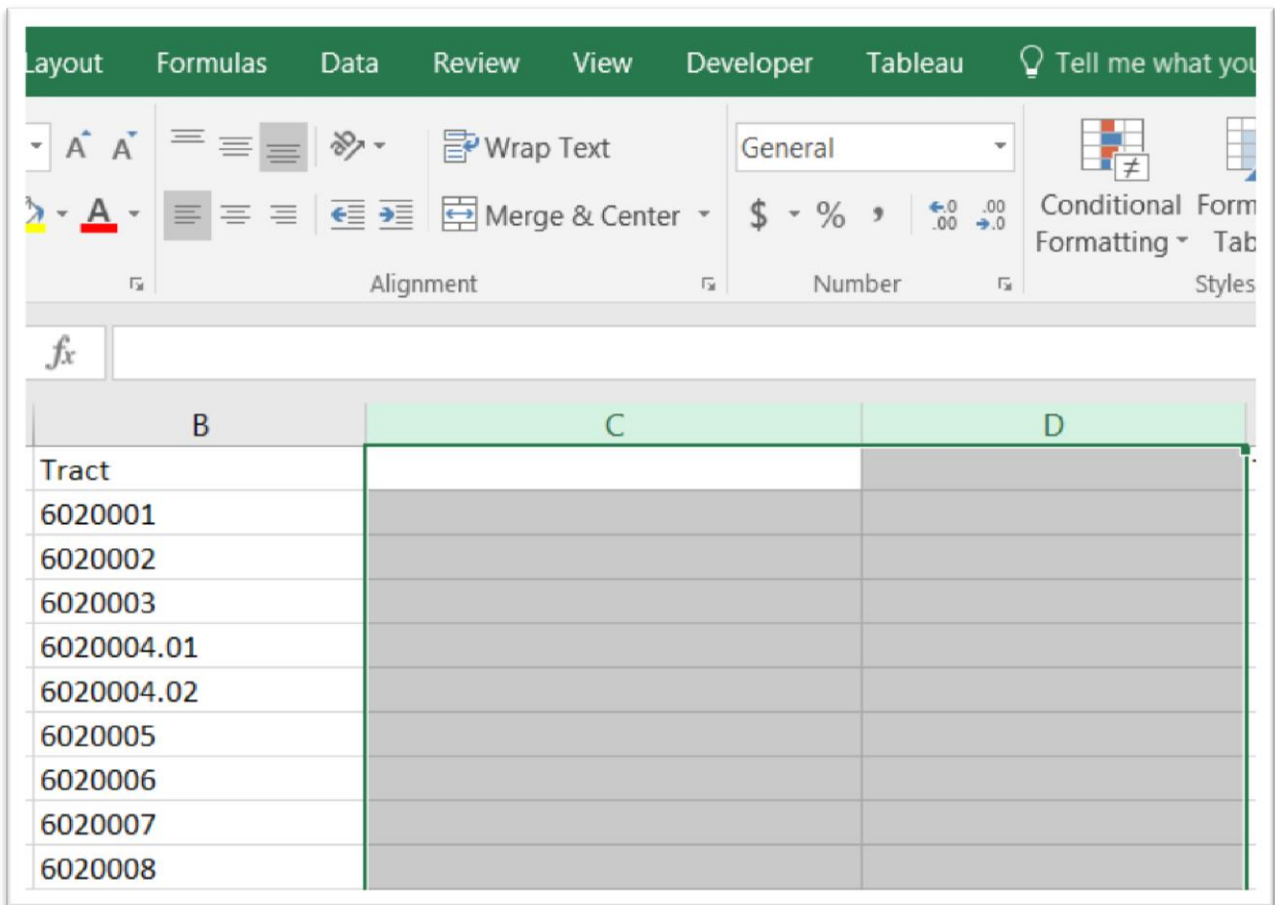
| | CTUID |
|---|------------|
| 0 | 6020001.00 |
| 1 | 6020002.00 |
| 2 | 6020003.00 |
| 3 | 6020004.01 |
| 4 | 6020004.02 |
| 5 | 6020005.00 |
| 6 | 6020006.00 |

Two problems. The first three values have a decimal place, followed by two zeros. And, as we mentioned previously, QGIS is interpreting the numbers as text because they are left-justified.

So, in the csv file, we must add a decimal place and two zeros to the numbers that don't have them.

Save the csv file as an Excel workbook. This is necessary because the csv file will not retain the functions that you'll be writing to add the zeros, as well as the new column.

Create two columns to the right of column B.



Reformat column B as text.

And in column C, type "FIND_DEC" (find decimal) in C1.

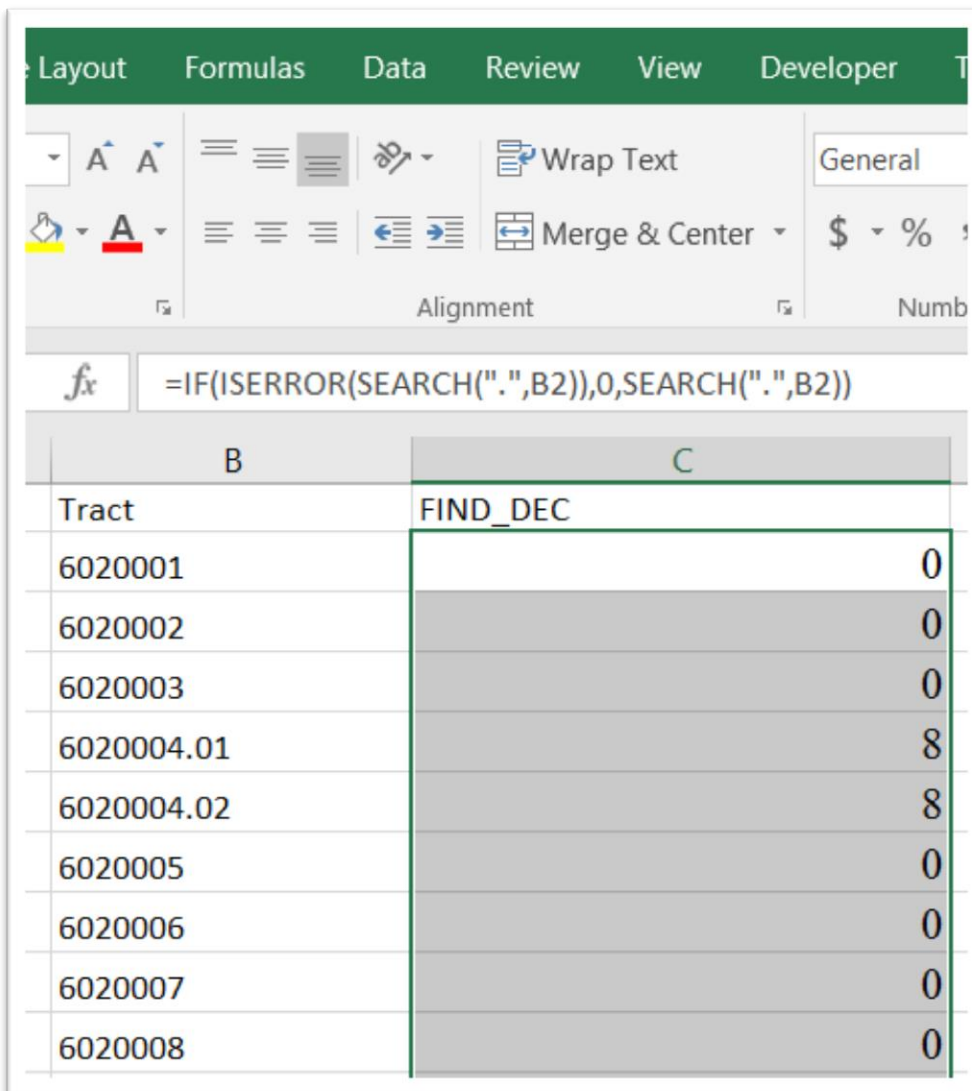
Then type the following formula in C2.

“=IF(ISERROR(SEARCH(".",B2)),0,SEARCH(".",B2))”

The screenshot displays the Microsoft Excel interface. The ribbon at the top includes 'Layout', 'Formulas', 'Data', 'Review', 'View', and 'Developer'. The 'Formulas' ribbon is active, showing options for 'Wrap Text', 'Merge & Center', and 'Alignment'. The formula bar shows the formula: `=IF(ISERROR(SEARCH(".",B2)),0,SEARCH(".",B2))`. Below the formula bar, a spreadsheet grid is visible with columns B and C. Column B contains the following values: 'Tract', '6020001', '6020002', '6020003', '6020004.01', '6020004.02', '6020005', '6020006', '6020007', and '6020008'. Column C has a header 'FIND_DEC'. Cell C2 is selected and contains the value '0'.

| B | C |
|------------|----------|
| Tract | FIND_DEC |
| 6020001 | 0 |
| 6020002 | |
| 6020003 | |
| 6020004.01 | |
| 6020004.02 | |
| 6020005 | |
| 6020006 | |
| 6020007 | |
| 6020008 | |

Copy the formula to the bottom.



| B | C |
|------------|----------|
| Tract | FIND_DEC |
| 6020001 | 0 |
| 6020002 | 0 |
| 6020003 | 0 |
| 6020004.01 | 8 |
| 6020004.02 | 8 |
| 6020005 | 0 |
| 6020006 | 0 |
| 6020007 | 0 |
| 6020008 | 0 |

The formula does the following: if there is a decimal in the numbers in the 'Tract' column, it will return the number that represents the number of characters from the left. If the decimal does not appear, it returns 0. You should have a column full of 8s, and 0s.

Copy column C and use the "paste special" option to plug the values into column D. If you can't remember how to do this, please refer to Chapter 4's [paste-special tutorial](#). Once you've pasted the new numbers, delete column C, and give our new column the same name. Looking at the value in the formula bar, you'll notice that the paste special has eliminated the function and just retained the value.

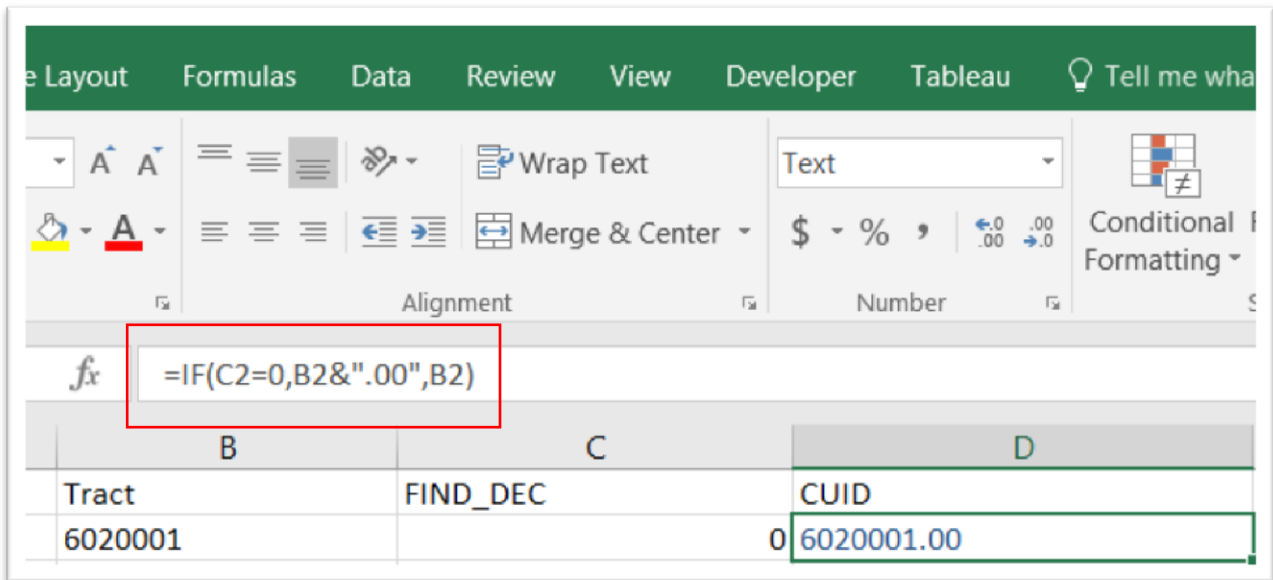
Layout Formulas Data Review View Devel

Alignment

fx 0

| B | C |
|------------|----------|
| Tract | FIND_DEC |
| 6020001 | 0 |
| 6020002 | 0 |
| 6020003 | 0 |
| 6020004.01 | 8 |
| 6020004.02 | 8 |
| 6020005 | 0 |
| 6020006 | 0 |
| 6020007 | 0 |
| 6020008 | 0 |

Now you should still have an empty column to the left of column C. Give column D the label CUID, and type this function. “=IF(C2=0,B2&".00",B2)”.



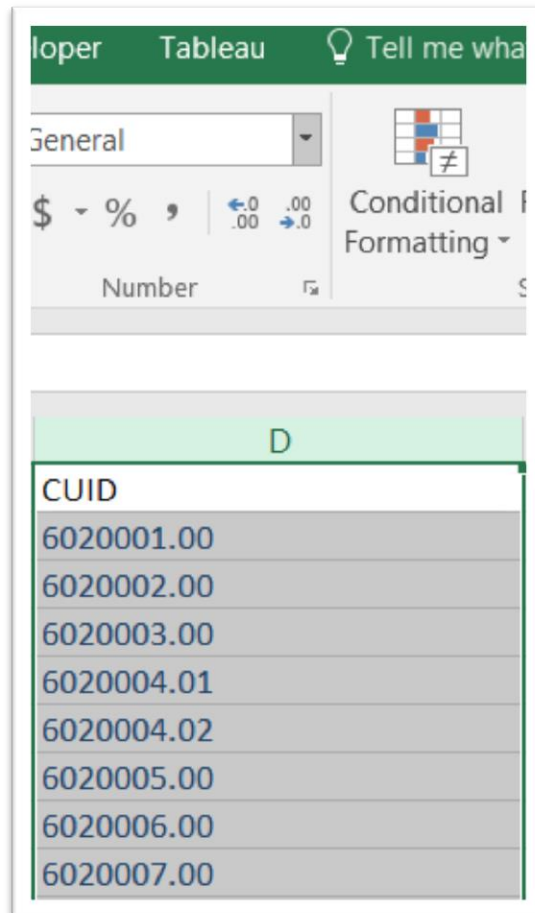
Pages 73-74 in Chapter 4 discusses IF statements, as does the accompanying tutorial, “[For working with specialized functions in Excel tutorial](#)”. Translated, the function in the formula bar says if the value in C2 equals zero, then use the concatenation operator (&) to add a decimal and two zeros (.00) to the value in B2. The decimal and two zeros constitute a condition, and as such must be bracketed by quotation marks. If C2 does not equal zero, then simply re-produce the value in the corresponding cell in column B.

Copy the formula to the bottom of column D and format the numbers as text.

The screenshot shows the Microsoft Excel interface. The ribbon includes 'Layout', 'Formulas', 'Data', 'Review', 'View', 'Developer', 'Tableau', and 'Tell me what'. The 'Layout' ribbon is active, showing options for 'Alignment' (Wrap Text, Merge & Center), 'Number' (Text, \$, %, ,), and 'Conditional Formatting'. The formula bar contains the formula: `=IF(C2=0,B2&".00",B2)`. Below the formula bar is a table with three columns: B, C, and D.

| B | C | D |
|------------|----------|--------------|
| Tract | FIND_DEC | CUID |
| 6020001 | | 0 6020001.00 |
| 6020002 | | 0 6020002.00 |
| 6020003 | | 0 6020003.00 |
| 6020004.01 | | 8 6020004.01 |
| 6020004.02 | | 8 6020004.02 |
| 6020005 | | 0 6020005.00 |
| 6020006 | | 0 6020006.00 |
| 6020007 | | 0 6020007.00 |
| 6020008 | | 0 6020008.00 |
| 6020009 | | 0 6020009.00 |
| 6020010 | | 0 6020010.00 |
| 6020011 | | 0 6020011.00 |
| 6020012 | | 0 6020012.00 |
| 6020013 | | 0 6020013.00 |
| 6020014 | | 0 6020014.00 |
| 6020015 | | 0 6020015.00 |
| 6020016 | | 0 6020016.00 |

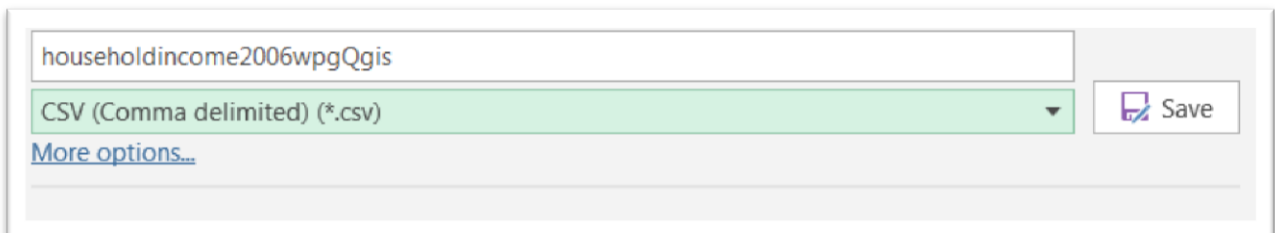
Now let's use the paste special to get rid of the formula in column D by copying the column.



Be sure to save the values as text, meaning that they are left-justified. Now you can delete all the columns to the left of D.

| | A | B | C | D | E |
|----|------------|---------|------------|-----------|------|
| 1 | CUID | Tract | Total_Priv | Median_20 | Medi |
| 2 | 6020001.00 | 6020001 | 1850 | 53436 | 4 |
| 3 | 6020002.00 | 6020002 | 2400 | 41184 | 3 |
| 4 | 6020003.00 | 6020003 | 2630 | 39250 | 3 |
| 5 | 6020004.01 | 6020004 | 2305 | 45270 | 3 |
| 6 | 6020004.02 | 6020004 | 1700 | 36937 | 3 |
| 7 | 6020005.00 | 6020005 | 2335 | 75475 | 6 |
| 8 | 6020006.00 | 6020006 | 2630 | 43616 | 3 |
| 9 | 6020007.00 | 6020007 | 1595 | 52045 | 4 |
| 10 | 6020008.00 | 6020008 | 1140 | 91369 | 7 |

If we were importing this file into ArcGIS, we could simply save the file in Excel format. However, QGIS deals with csv files.

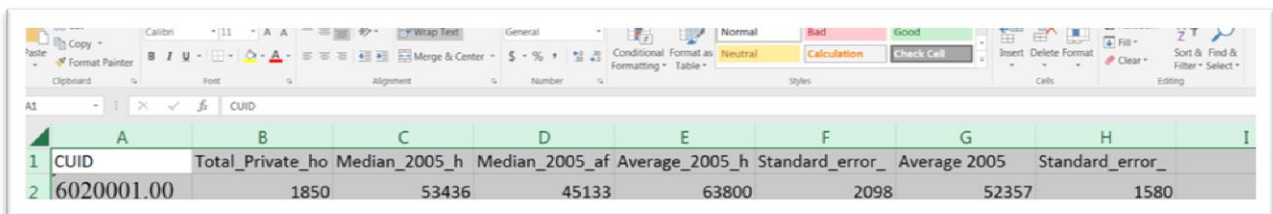
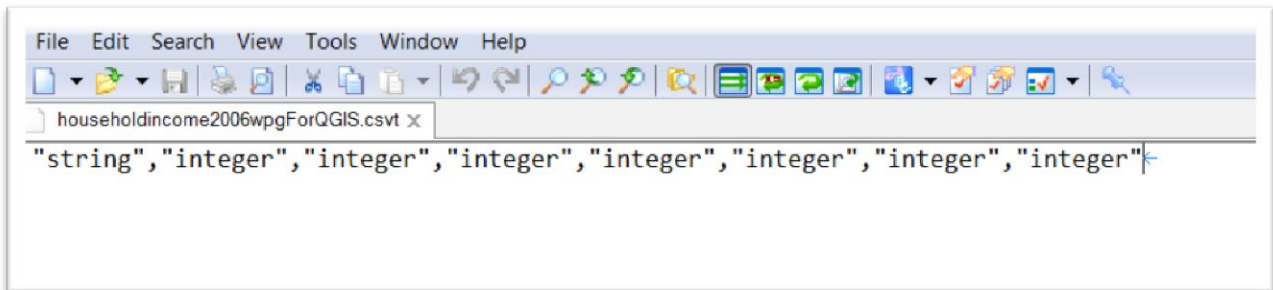


There remains one more step before saving the Excel file as a csv file, and then importing the table into QGIS.

QGIS will import column A as a number format. So we have to use a text file with a [csvt](#) extension that will, in essence, force QGIS to recognize our column as text, not

numbers. The csvt file only contains one row which specifies the datatypes for each column.

And this is crucial: the csvt file MUST have the SAME name as the Excel file that we will save in csv format, and go in the SAME directory as that csv file. You can create the csvt file in a notepad, or one of the many open-sourced text editors discussed in the Appendix A. Our csvt file looks like this, which is created in the text editor, [EmEditor](#). String defines text; integer, a number.



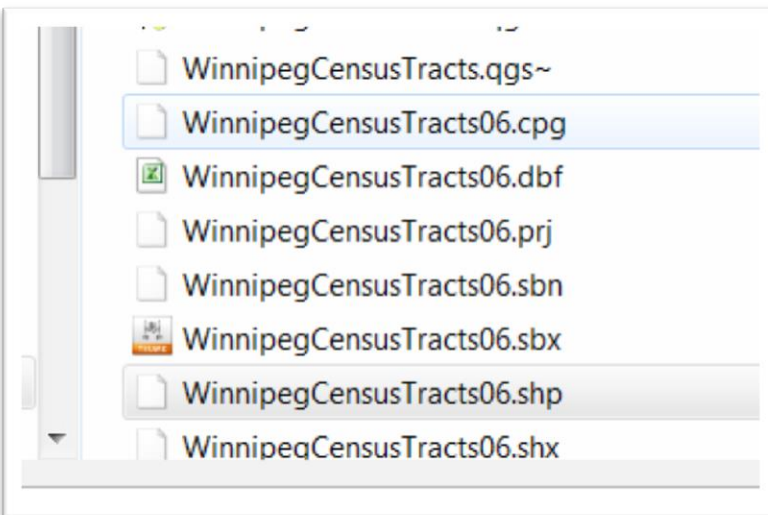
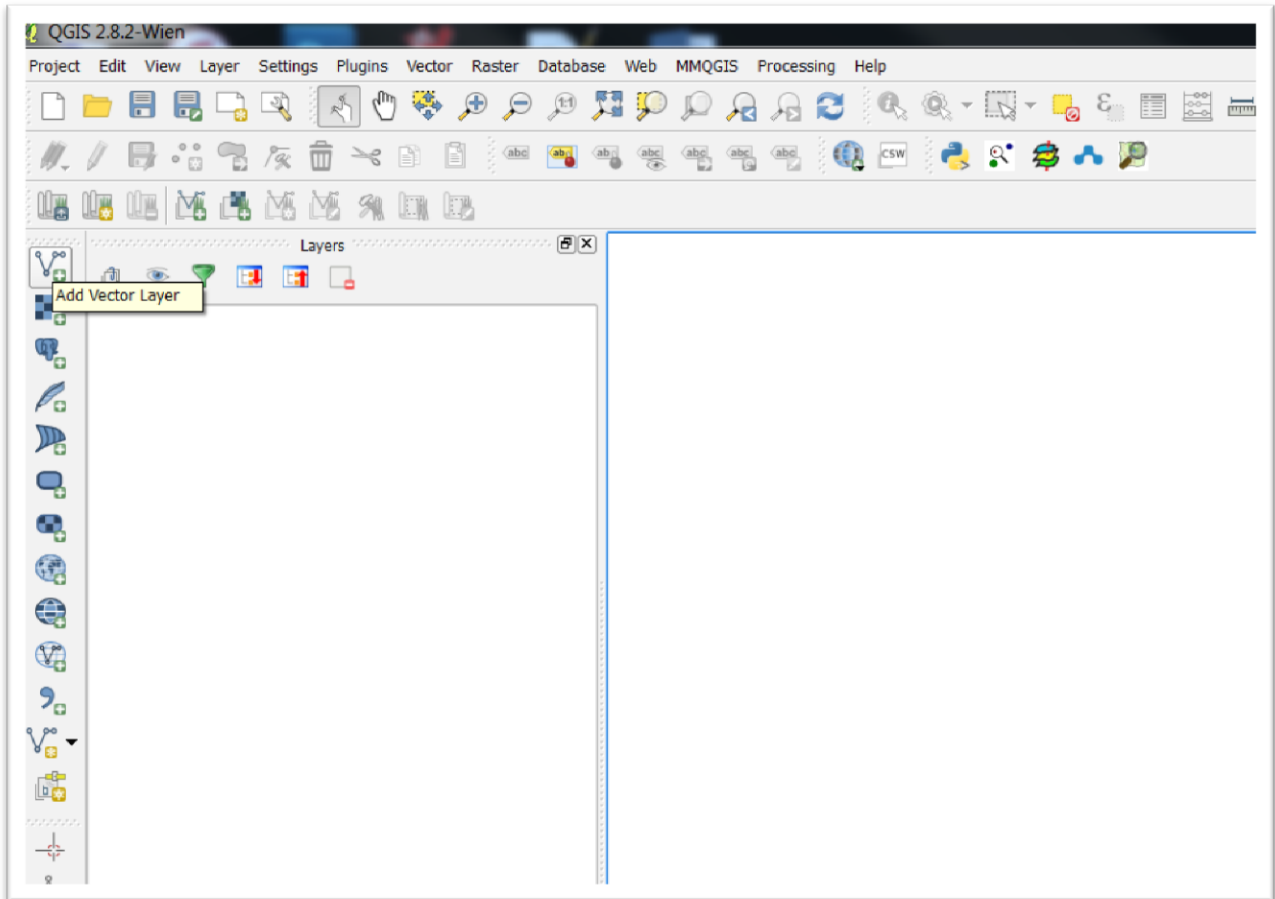
| | A | B | C | D | E | F | G | H | I |
|---|------------|------------------|---------------|----------------|----------------|-----------------|--------------|-----------------|---|
| 1 | CUID | Total_Private_ho | Median_2005_h | Median_2005_af | Average_2005_h | Standard_error_ | Average 2005 | Standard_error_ | |
| 2 | 6020001.00 | 1850 | 53436 | 45133 | 63800 | 2098 | 52357 | 1580 | |

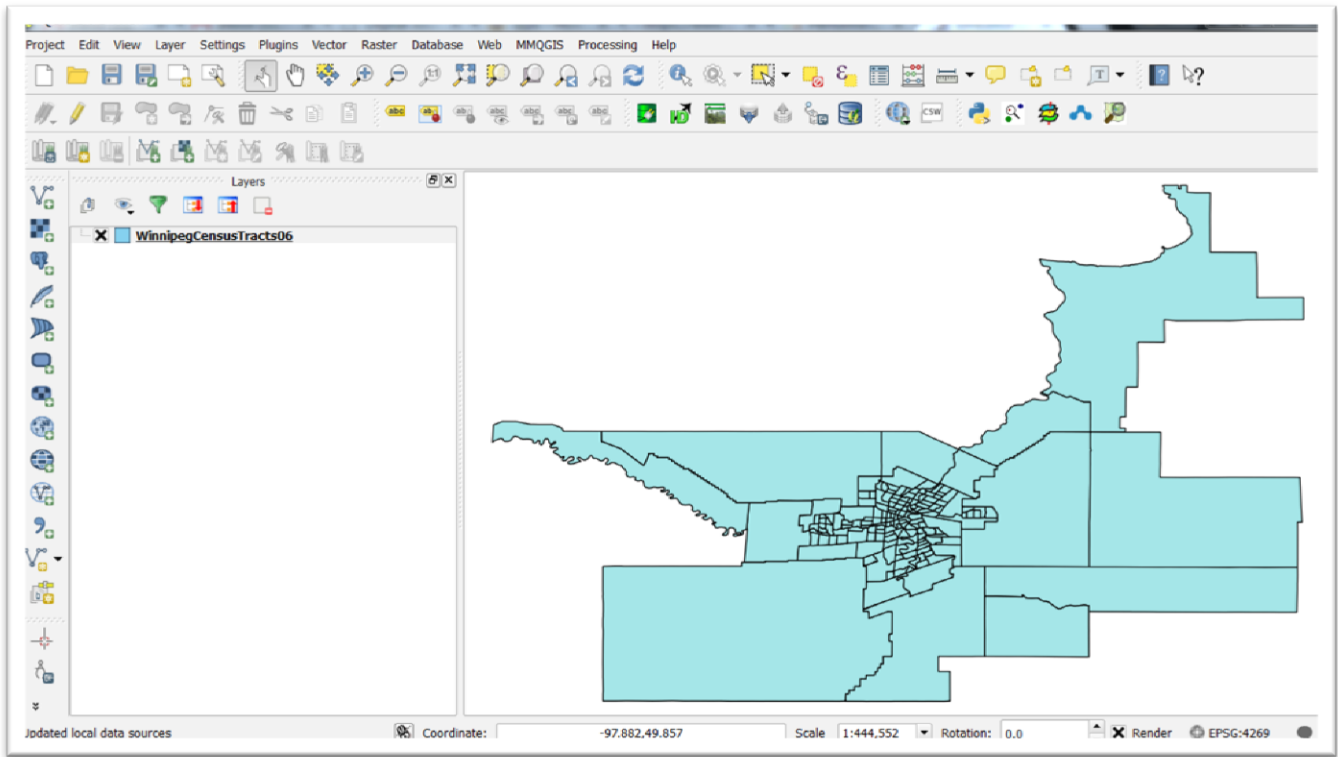
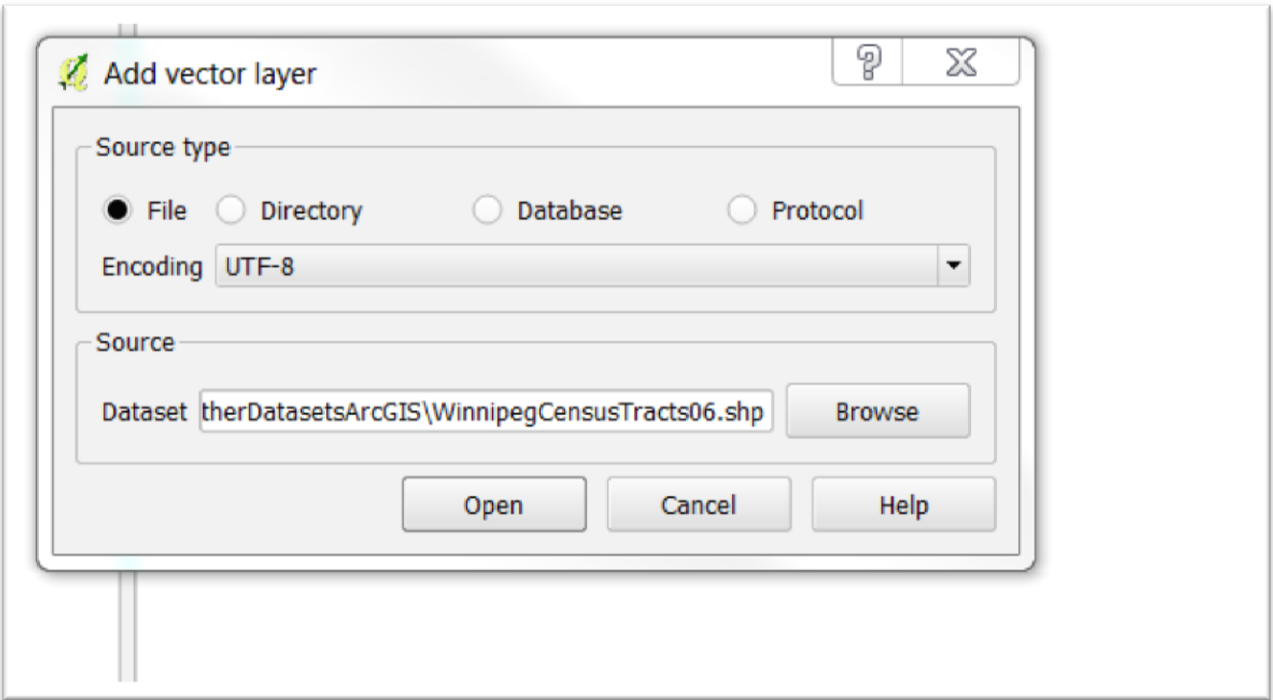
The csvt file defines each datatype in the csv file. The key is the first column. The “string” tells QGIS to import the CUID column as text.

Once you’ve created the csvt file, save the Excel worksheet as a csv file. Remember, the csv and csvt files MUST have the SAME name, and be in the SAME folder.

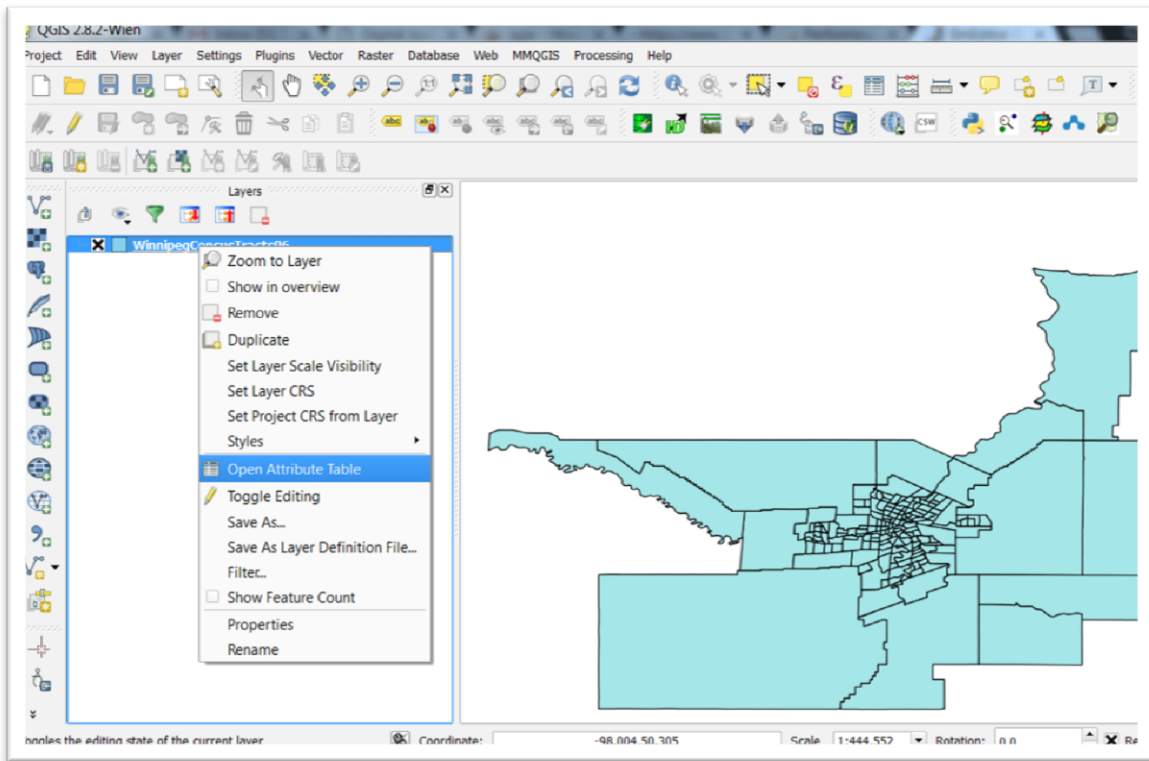
Importing the files into QGIS

Open QGIS and use the “Add Vector Layer” icon to browse for, and then import the Winnipeg census tract shape file.

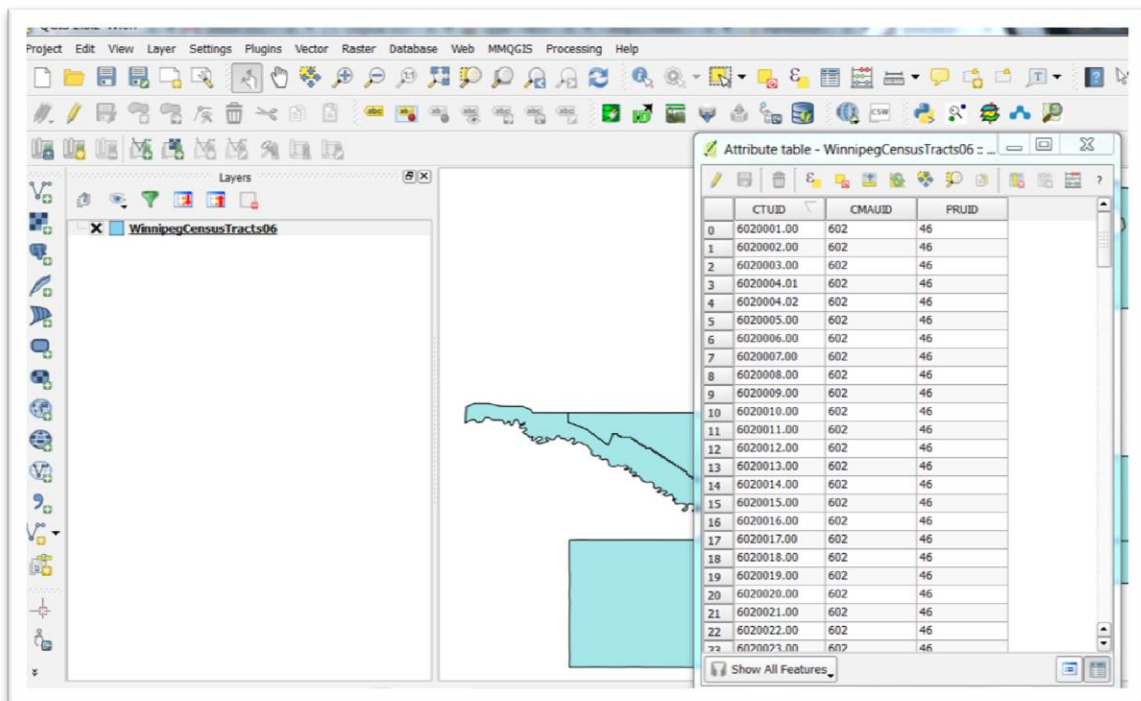


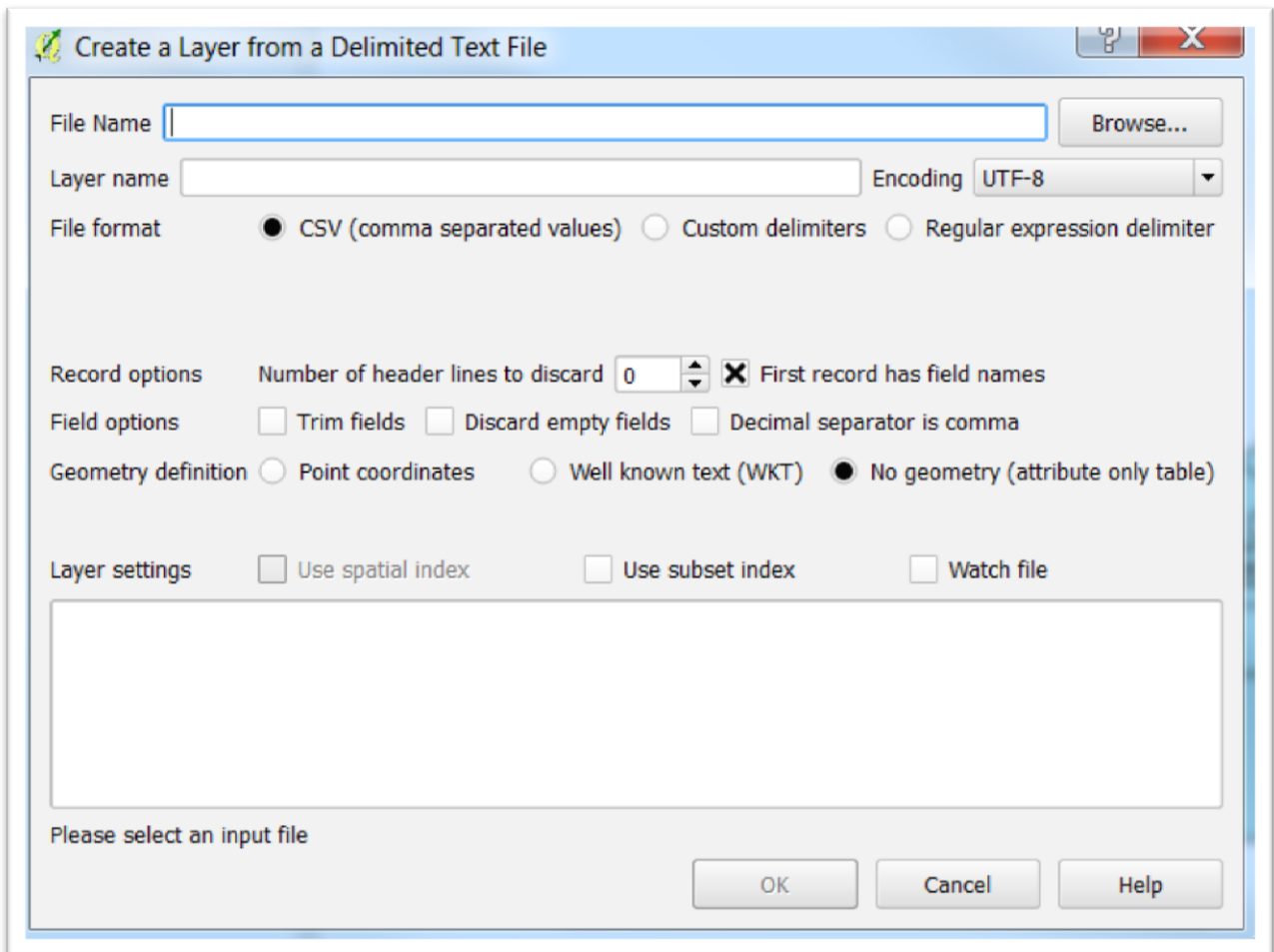


Right-click on the layer in the menu to the left to obtain your attribute table.

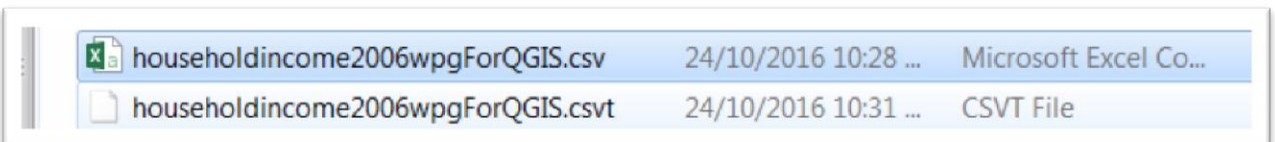


Selecting the “Open Attribute Table” option produces a dialog table which contains the geographic information QGIS – like ArcGIS – uses to map the census boundaries.





Now browse for the csv file, which should be right with the csvt file that we have created.



Next, we get a “Create a Layer from a Delimited Text File” dialogue box.

The dialog box is titled "Create a Layer from a Delimited Text File". It contains the following fields and options:

- File Name: `pter seven/JoiningMapstoOtherDatasetsArcGIS/householdincome2006wpgForQGIS.csv` (with a "Browse..." button)
- Layer name: `householdincome2006wpgForQGIS`
- Encoding: `UTF-8`
- File format: CSV (comma separated values), Custom delimiters, Regular expression delimiter
- Record options: Number of header lines to discard: `0`; First record has field names
- Field options: Trim fields, Discard empty fields, Decimal separator is comma
- Geometry definition: Point coordinates, Well known text (WKT), No geometry (attribute only table)
- X field: [empty], Y field: [empty], DMS coordinates
- Layer settings: Use spatial index, Use subset index, Watch file

A preview table is shown below the settings:

| | Tract | Total_Private_households | Median_2005_household_income | Median_2005_after |
|---|------------|--------------------------|------------------------------|-------------------|
| 1 | 6020001.00 | 1850 | 53436 | 45133 |
| 2 | 6020002.00 | 2400 | 41184 | 36694 |
| 3 | 6020003.00 | 2630 | 39250 | 34345 |

At the bottom, there is a message: "X and Y field names must be selected". Buttons for "OK", "Cancel", and "Help" are located at the bottom right.

QGIS has rightly guessed that it's a csv file. Since it doesn't have any X and Y geographic coordinates which come into play when performing spatial joins, click the box that specifies that there are “No geometry” coordinates.

Create a Layer from a Delimited Text File

File Name

Layer name Encoding

File format CSV (comma separated values) Custom delimiters Regular expression delimiter

Record options Number of header lines to discard First record has field names

Field options Trim fields Discard empty fields Decimal separator is comma

Geometry definition Point coordinates Well known text (WKT) No geometry (attribute only table)

Layer settings Use spatial index Use subset index Watch file

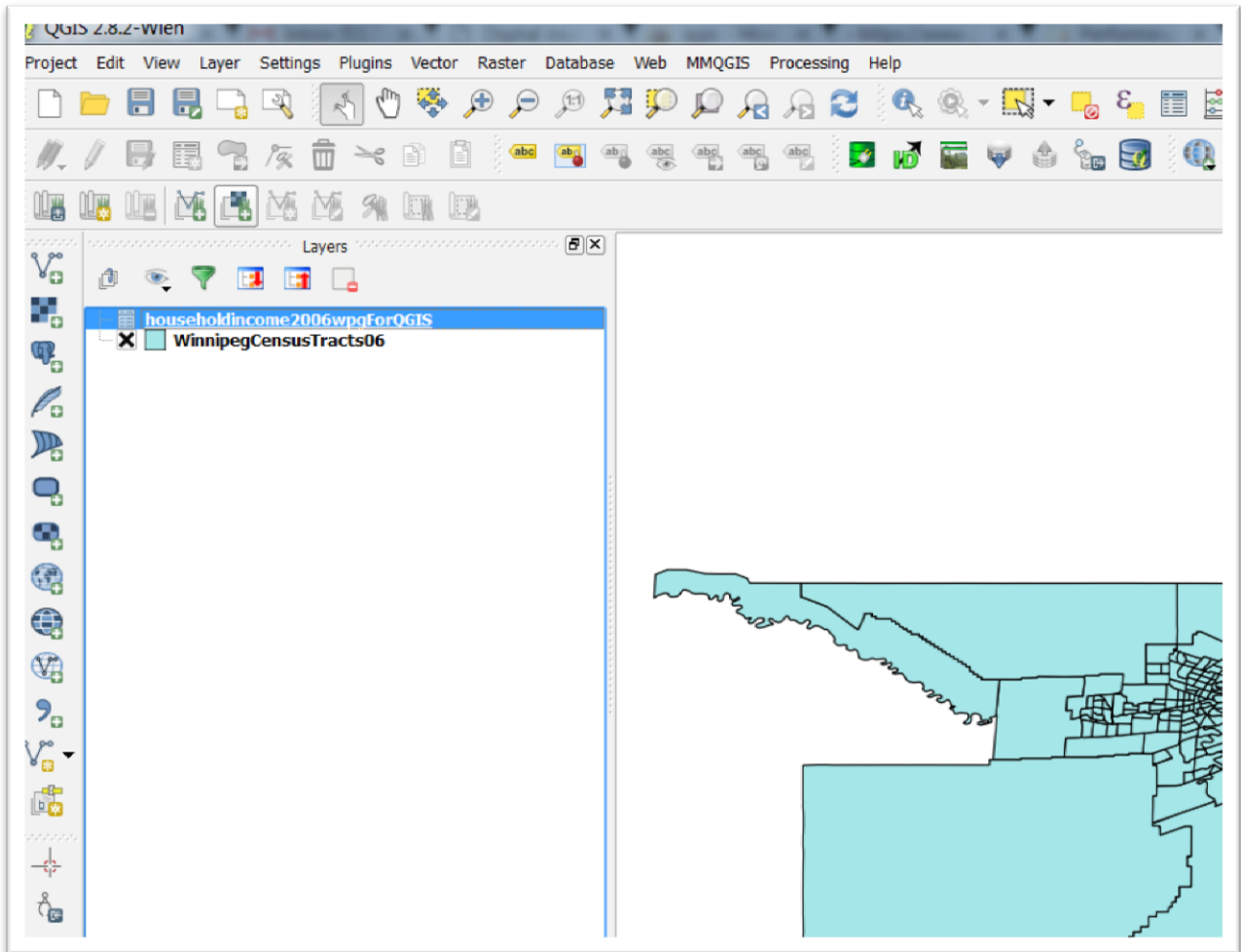
| | Tract | Total_Private_households | Median_2005_household_income | Median_2005_aft |
|---|------------|--------------------------|------------------------------|-----------------|
| 1 | 6020001.00 | 1850 | 53436 | 45133 |
| 2 | 6020002.00 | 2400 | 41184 | 36694 |
| 3 | 6020003.00 | 2630 | 39250 | 34345 |

OK

Cancel

Help

Select the “OK” tab.



Now we have a second file in our layer menu. Just as we did with the census tract file, open the attribute table to see what's there.

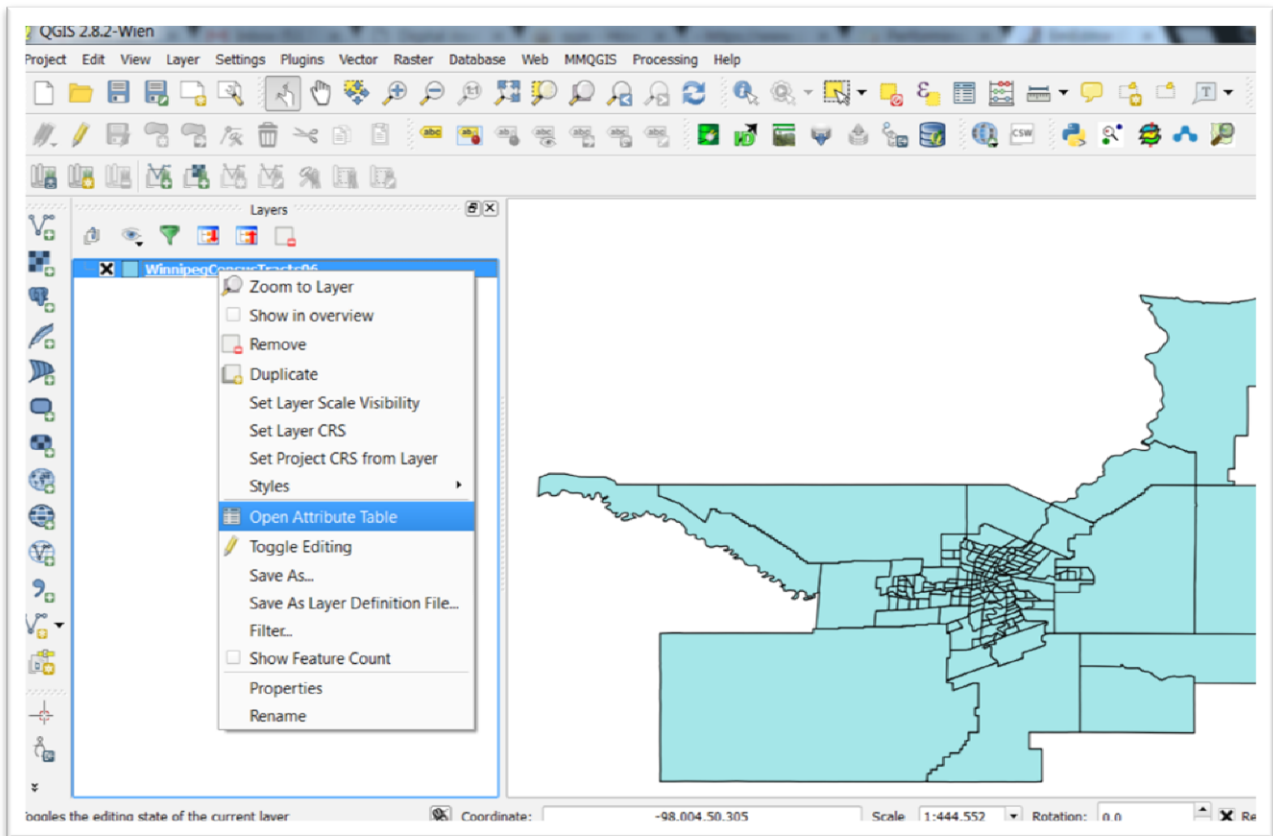
Attribute table - householdincome2006wpgForQGIS :: Features total: 167, filtered: 167, selected: 0

| | CUID | _Private_household | 2005_household | _aftertax_household | 2005_household | _of_average_household | after-tax household | average_aftertax |
|----|------------|--------------------|----------------|---------------------|----------------|-----------------------|---------------------|------------------|
| 0 | 6020001.00 | 1850 | 53436 | 45133 | 63800 | 2098 | 52357 | 1580 |
| 1 | 6020002.00 | 2400 | 41184 | 36694 | 49562 | 2438 | 41537 | 1538 |
| 2 | 6020003.00 | 2630 | 39250 | 34345 | 45973 | 1417 | 38835 | 1068 |
| 3 | 6020004.01 | 2305 | 45270 | 39247 | 52837 | 1530 | 44227 | 1157 |
| 4 | 6020004.02 | 1700 | 36937 | 33134 | 41922 | 1313 | 36099 | 1025 |
| 5 | 6020005.00 | 2335 | 75475 | 60342 | 90117 | 3634 | 70788 | 2386 |
| 6 | 6020006.00 | 2630 | 43616 | 37830 | 51066 | 1496 | 42711 | 1114 |
| 7 | 6020007.00 | 1595 | 52045 | 43546 | 61740 | 2291 | 50829 | 1719 |
| 8 | 6020008.00 | 1140 | 91369 | 70850 | 122958 | 8591 | 90696 | 5103 |
| 9 | 6020009.00 | 1225 | 90350 | 71015 | 106568 | 5024 | 81378 | 2966 |
| 10 | 6020010.00 | 2410 | 62693 | 51968 | 98276 | 5363 | 74343 | 3530 |
| 11 | 6020011.00 | 4025 | 47899 | 39792 | 62896 | 2944 | 49839 | 1754 |
| 12 | 6020012.00 | 2975 | 28865 | 25643 | 34264 | 1025 | 29543 | 775 |
| 13 | 6020013.00 | 915 | 19108 | 17393 | 26686 | 1598 | 23288 | 1266 |
| 14 | 6020014.00 | 3855 | 30029 | 27163 | 36575 | 1043 | 31176 | 797 |
| 15 | 6020015.00 | 3395 | 20486 | 18965 | 28371 | 1001 | 25096 | 784 |
| 16 | 6020016.00 | 1070 | 32861 | 27868 | 47162 | 3436 | 39182 | 2428 |
| 17 | 6020017.00 | 1450 | 52400 | 44766 | 62636 | 2551 | 51178 | 1908 |
| 18 | 6020018.00 | 1260 | 46977 | 41392 | 57152 | 2622 | 47689 | 1911 |
| 19 | 6020019.00 | 1265 | 47457 | 40948 | 52776 | 1808 | 44766 | 1405 |
| 20 | 6020020.00 | 1020 | 41881 | 37465 | 49440 | 1895 | 42561 | 1499 |
| 21 | 6020021.00 | 2350 | 32446 | 29331 | 39619 | 1315 | 35140 | 1119 |
| 22 | 6020022.00 | 1845 | 20374 | 19670 | 26888 | 1122 | 24745 | 953 |

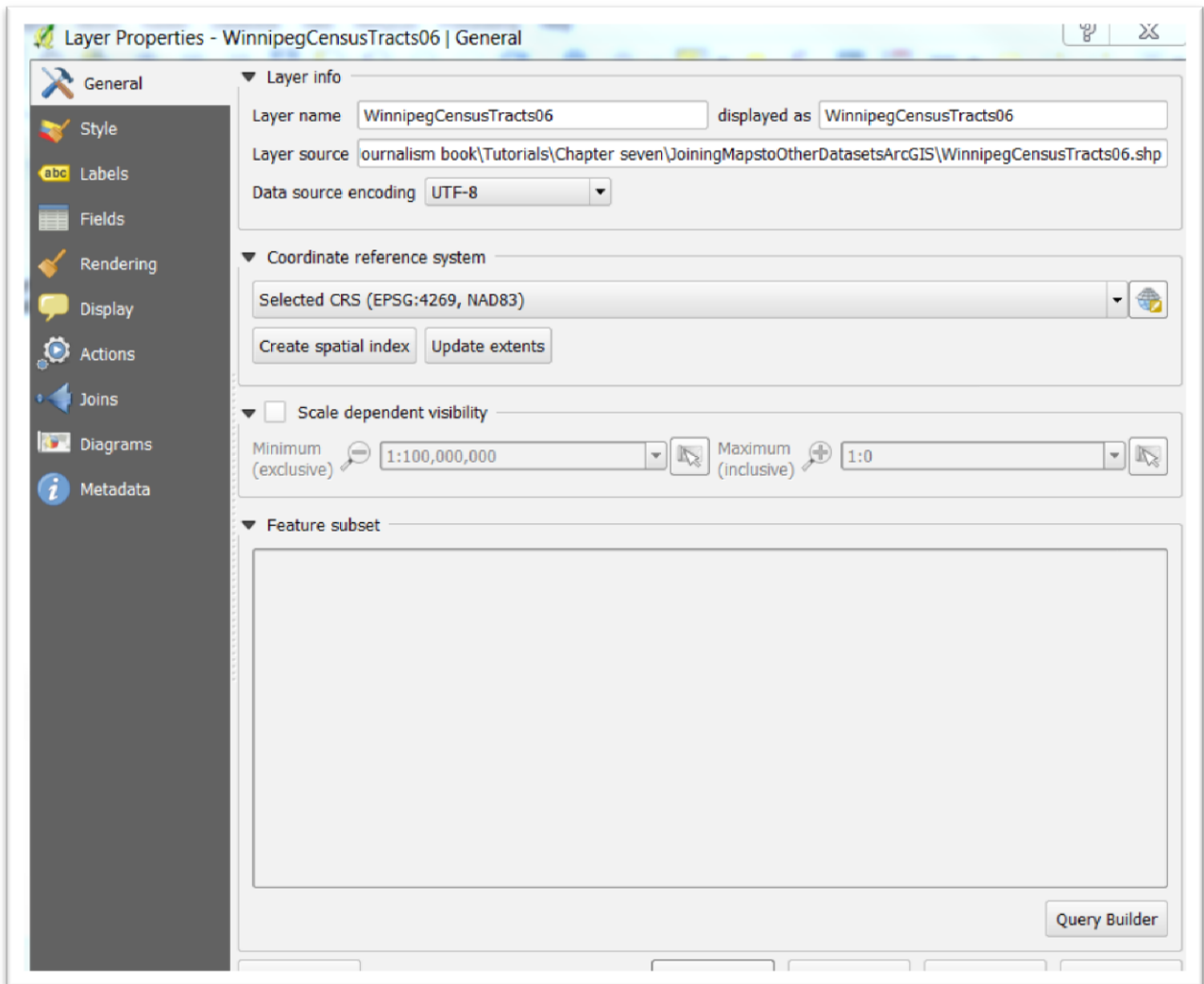
Show All Features

Thanks to the csvt file, QGIS imported the values in the CUID column as and “string”, or text, and the rest of the values as “integers” or numbers. The latter is also important because QGIS, like ArcGIS, (Or Excel or MySQL, for that matter) can only do math on numbers.

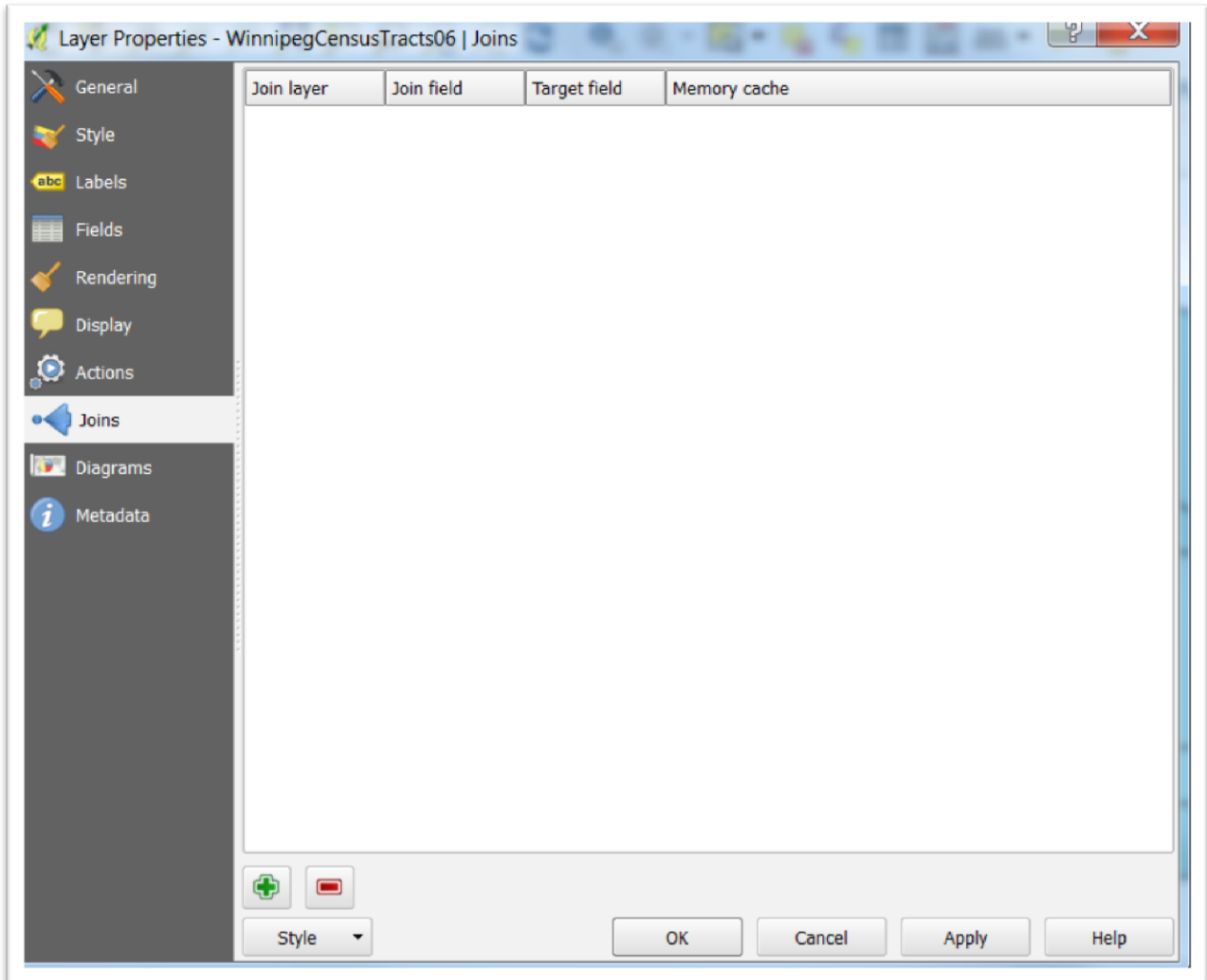
Close the attribute table. And right click on the Winnipeg census tract layer to obtain our short cut menu.



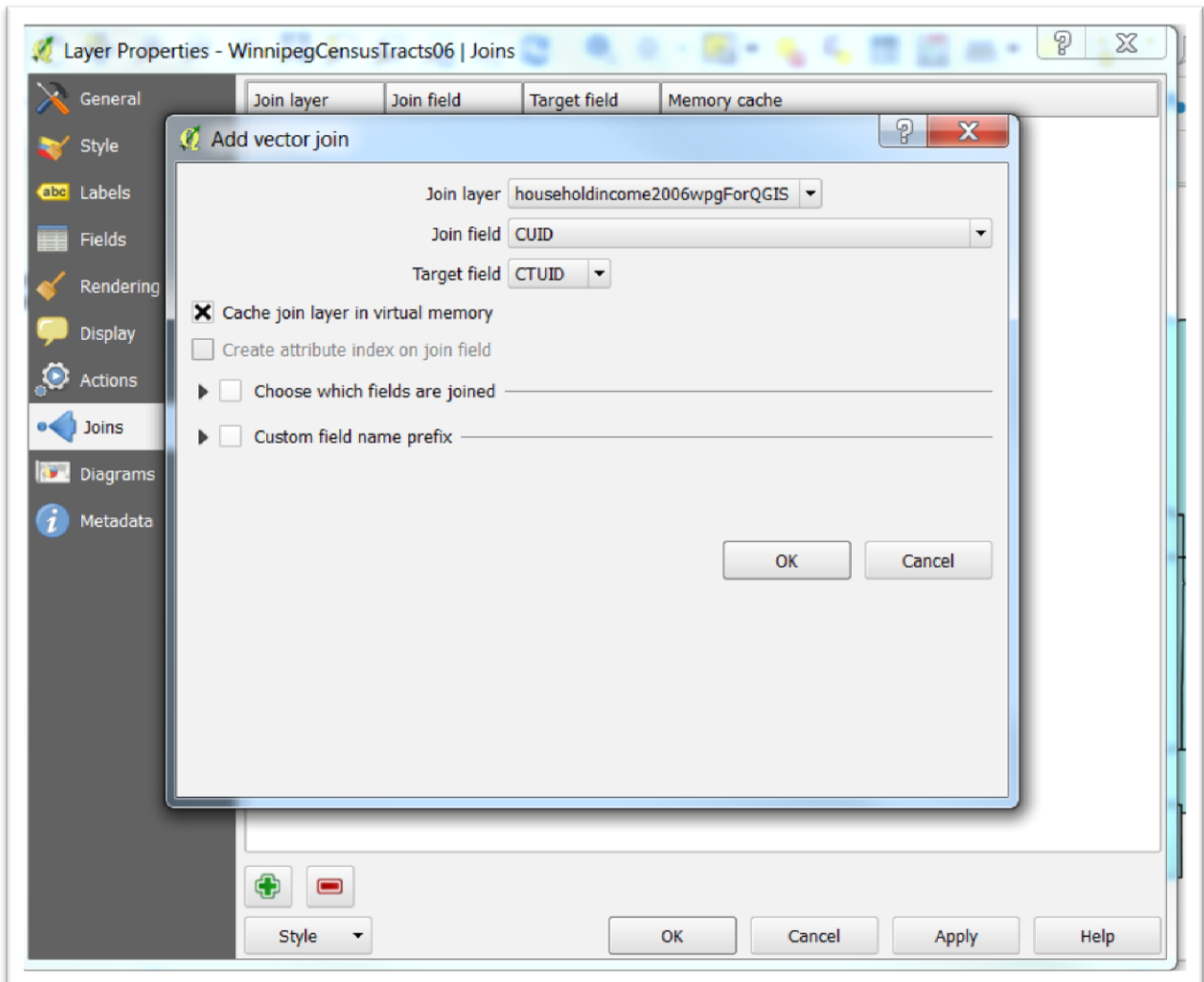
Select “Properties”, which should open to the general tab which contains information we’ll explain in the spatial join tutorial.



Select Joins.

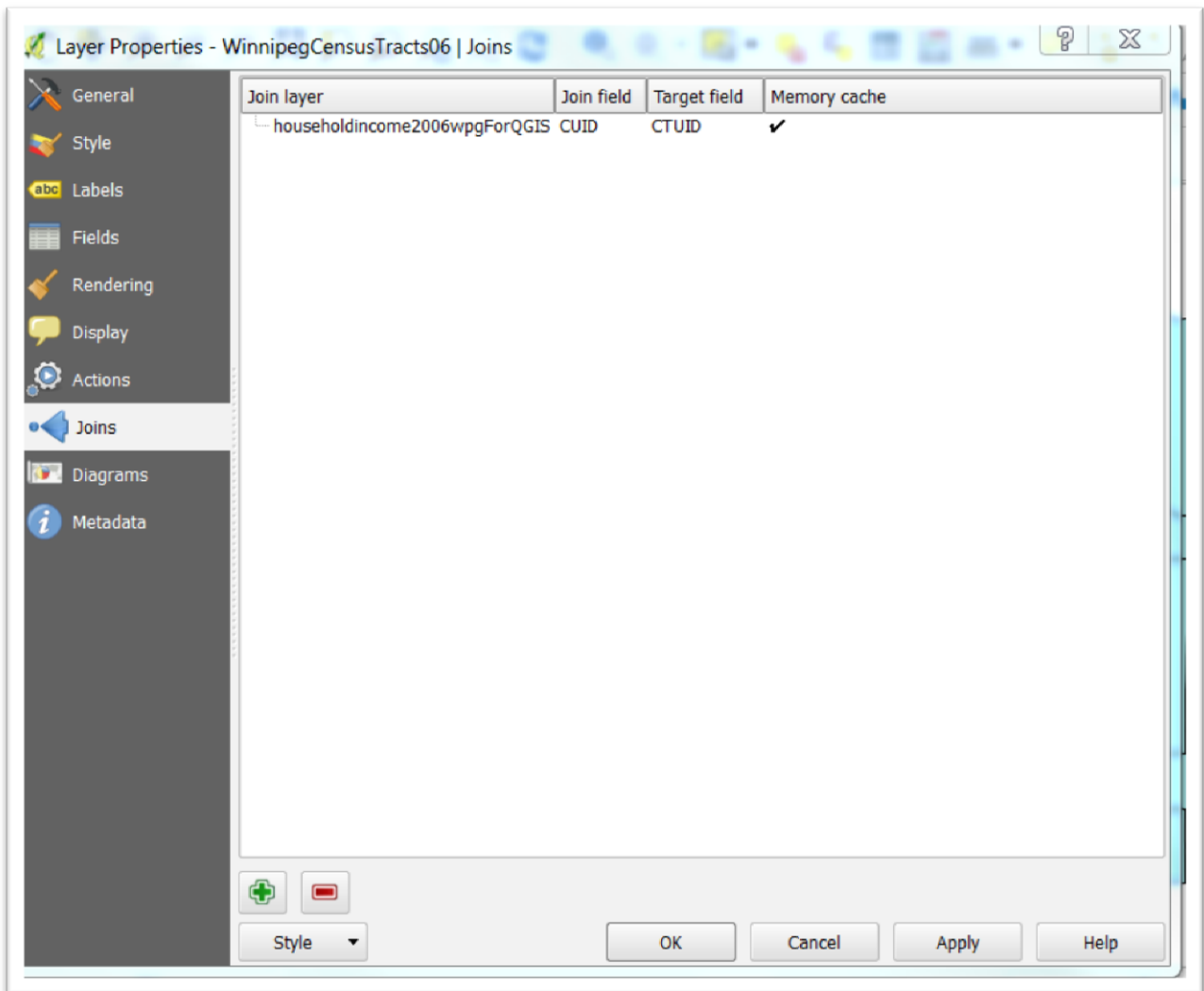


For specific values to appear on the map, we must join it to the Map. And the green plus sign at the bottom left.



Because we've already selected the Winnipeg census tract layer, our "Join layer" is the csv file. The "Join field" is "CUID", the new one we created and renamed earlier in this tutorial. The "Target field" in the Winnipeg census tract file is the CTUID field.

Select the OK tab.



At the top of the dialogue box, we can see that QGIS has informed us that the join has been executed. So select the “Apply” tab, and then “OK.”

For further evidence that we have successfully joined the csv file to the census tract shape file, right click on the latter to obtain the attribute table, and expand the width in order to see all the columns.

Attribute table - WinnipegCensusTracts06 -: Features total: 168, filtered: 168, selected: 0

| | CTUID | CMAUID | PRUID | wppForQGIS_To | ForQGIS_Mediar | GIS_Median_20 | ForQGIS_Averag | 3_Standard_ern | GIS_Average 20 | andard_error_of |
|----|------------|--------|-------|---------------|----------------|---------------|----------------|----------------|----------------|-----------------|
| 0 | 6020001.00 | 602 | 46 | 1850 | 53436 | 45133 | 63800 | 2098 | 52357 | 1580 |
| 1 | 6020002.00 | 602 | 46 | 2400 | 41184 | 36694 | 49562 | 2438 | 41537 | 1538 |
| 2 | 6020003.00 | 602 | 46 | 2630 | 39250 | 34345 | 45973 | 1417 | 38835 | 1068 |
| 3 | 6020004.01 | 602 | 46 | 2305 | 45270 | 39247 | 52837 | 1530 | 44227 | 1157 |
| 4 | 6020004.02 | 602 | 46 | 1700 | 36937 | 33134 | 41922 | 1313 | 36099 | 1025 |
| 5 | 6020005.00 | 602 | 46 | 2335 | 75475 | 60342 | 90117 | 3634 | 70788 | 2386 |
| 6 | 6020006.00 | 602 | 46 | 2630 | 43616 | 37830 | 51066 | 1496 | 42711 | 1114 |
| 7 | 6020007.00 | 602 | 46 | 1595 | 52045 | 43546 | 61740 | 2291 | 50829 | 1719 |
| 8 | 6020008.00 | 602 | 46 | 1140 | 91369 | 70850 | 122958 | 8591 | 90696 | 5103 |
| 9 | 6020009.00 | 602 | 46 | 1225 | 90350 | 71015 | 106568 | 5024 | 81378 | 2966 |
| 10 | 6020010.00 | 602 | 46 | 2410 | 62693 | 51968 | 98276 | 5363 | 74343 | 3530 |
| 11 | 6020011.00 | 602 | 46 | 4025 | 47899 | 39792 | 62896 | 2944 | 49839 | 1754 |
| 12 | 6020012.00 | 602 | 46 | 2975 | 28865 | 25643 | 34264 | 1025 | 29543 | 775 |
| 13 | 6020013.00 | 602 | 46 | 915 | 19108 | 17393 | 26686 | 1598 | 23288 | 1266 |
| 14 | 6020014.00 | 602 | 46 | 3855 | 30029 | 27163 | 36575 | 1043 | 31176 | 797 |
| 15 | 6020015.00 | 602 | 46 | 3395 | 20486 | 18965 | 28371 | 1001 | 25096 | 784 |
| 16 | 6020016.00 | 602 | 46 | 1070 | 32861 | 27868 | 47162 | 3436 | 39182 | 2428 |
| 17 | 6020017.00 | 602 | 46 | 1450 | 52400 | 44766 | 62636 | 2551 | 51178 | 1908 |
| 18 | 6020018.00 | 602 | 46 | 1260 | 46977 | 41392 | 57152 | 2622 | 47689 | 1911 |
| 19 | 6020019.00 | 602 | 46 | 1265 | 47457 | 40948 | 52776 | 1808 | 44766 | 1405 |
| 20 | 6020020.00 | 602 | 46 | 1020 | 41881 | 37465 | 49440 | 1895 | 42561 | 1499 |
| 21 | 6020021.00 | 602 | 46 | 2350 | 32446 | 29331 | 39619 | 1315 | 35140 | 1119 |
| 22 | 6020022.00 | 602 | 46 | 1845 | 20374 | 19670 | 26888 | 1122 | 24745 | 953 |
| 23 | 6020023.00 | 602 | 46 | 2665 | 21520 | 19748 | 27351 | 858 | 24463 | 702 |
| 24 | 6020024.00 | 602 | 46 | 200 | 53509 | 51461 | 74120 | 11129 | 58100 | 7854 |

Show All Features

The first three columns belong to our shape file; the rest, to the csv file.

Close the attribute table.

We will learn how to colour code the results in the “7_16_MakingChoroplethinQgis” tutorial.

Now we have successfully mapped the cities income levels and have an idea what areas have neighborhoods where we might want to visit.