# Introduction to QGIS and Geoprocessing using the Census of Canada

Rebecca Bartlett, GIS & Digital Resources Librarian, October 2017

1. Download Census Tracts shapefile and add to map document

Census Tracts (CTs) are geographic areas in Census Metropolitan Areas (CMAs) that have a population of 2500-8000. The data is downloaded as a shapefile which can then be viewed in QGIS.

1) In an internet browser, go to statcan.gc.ca.
2) After you've selected English, click **Browse by key resource**
3) Scroll down and click on **Maps and geography**
4) Click **Geographic products**, then **Boundary files**, then **2016**
5) In Boundary files options, select **English** and **ArcGIS (.shp)** format
6) Click the radio button for **Census tracts** and **Cartographic Boundary File** then the **Continue** button
7) Click the **lct_000b16a_e.zip (ZIP Version, 8247.0 kb)** link
8) Once it has downloaded, copy the ZIP folder into the project folder for today's class. We suggest a hierarchy like this, but we're particular about this stuff:
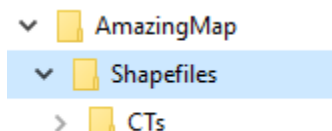
9) Unzip the file. If you don't, you may get an error when you try to import it into QGIS.
10) Once unzipped, your folder should look something like this:

Name	Date modified	Type	Size
92-160-g2016002-eng.pdf	8/23/2017 8:28 AM	Adobe Acrobat D...	350 KB
census_tract.html	3/1/2017 11:20 AM	Chrome HTML Do...	171 KB
lct_000b16a_e.dbf	9/20/2016 4:45 PM	DBF File	1,280 KB
lct_000b16a_e.prj	9/20/2016 4:45 PM	PRJ File	1 KB
lct_000b16a_e.shp	9/20/2016 4:45 PM	DWG TrueView Sh...	15,791 KB
lct_000b16a_e.shx	9/20/2016 4:45 PM	DWG TrueView Co...	45 KB
lct_000b16a_e.zip	10/17/2017 3:04 PM	Compressed (zipp...	8,247 KB

This is as good a time as any to note that shapefiles look like multiple different files when you view them in Windows Explorer. If you ever copy and paste them somewhere, you need to have all the parts (e.g.: all the files that have the same name before the file extension – there may be up to 8 of them).

11) You're done the download!
12) Open QGIS. Once QGIS opens, add the dataset by clicking on **Layer > Add Layer > Add Vector Layer…** Browse to the project's folder and find the shapefile. Click on the .shp and click **Ok** to add the shapefile to QGIS.

    a. You could also have the project folder open in Windows Explorer and just drag and drop the .shp file in. Either way works!

13) You'll notice that much of the country is missing – this is normal! Remember that census tracts are only in cities with a population of over 100,000.

14) Click **View >Panels** and make sure **Browser** and **Layers** are both selected.

15) Because we will want our map to look reasonably nice, let's follow steps 1-13 except download the **Provinces/territories Cartographic Boundary File** and unzip it and add it to QGIS. Your map should now look something like this:

16) If you're wondering why the census tracts seem to have disappeared, it's because that layer is below the provinces/territories layer in the Layers Panel:
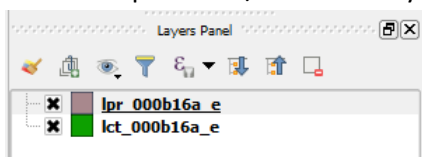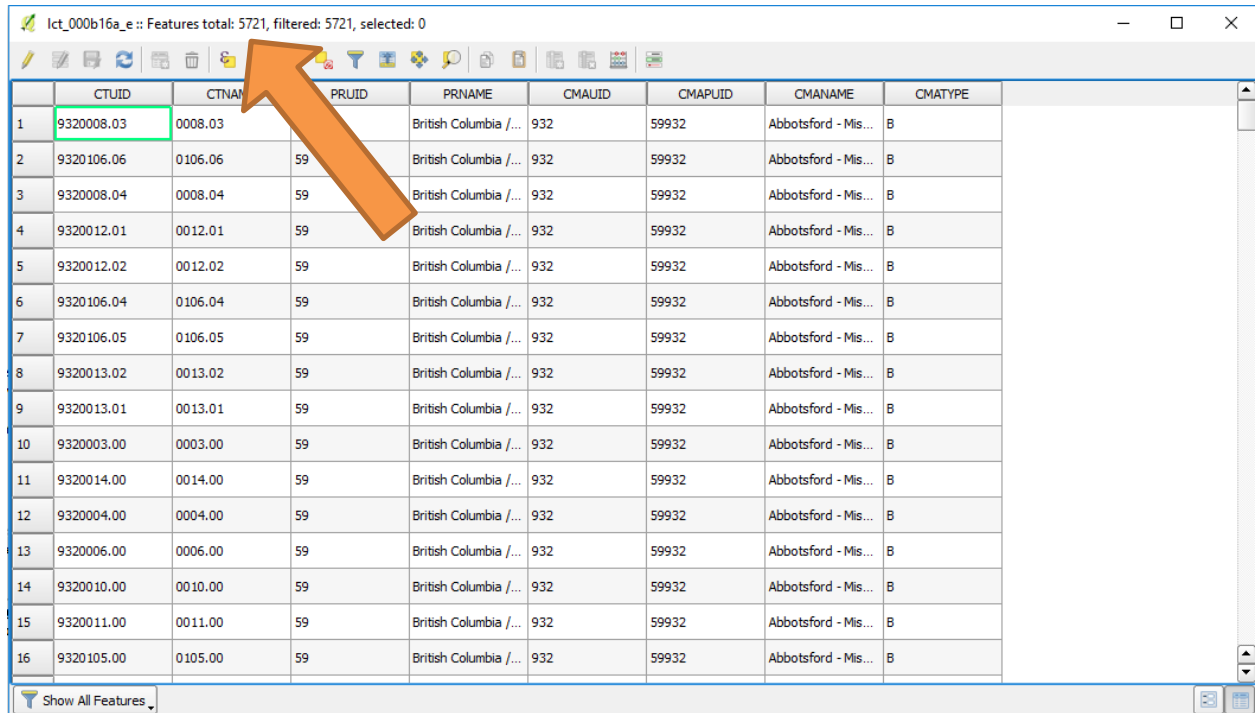
17) Click and drag **lct_000b16a_e** above the provinces/territories layer to view it on top. The layers always show top to bottom. You can also turn the layers on and off with the checkbox.

2. Querying and Extracting

When confronted with a large dataset you may want to use the querying capabilities of QGIS to extract one or more subsets.

1) Right click on the **lct_00616a_e** and select **Open Attribute Table**.
2) You will note that there are 5721 features in the dataset.

3) In order to query the data, click on the **Select features using an expression** button on the attribute table toolbar.
4) The **Select by expression** dialog box will appear. It allows you to create a query statement that will select features that conform to the parameters you set. In this example, we will search for only **New Brunswick** census tracts by using the **PRNAME** field.
   a. Start off by making the **Select by expression** dialog box wider. It will make sense shortly.
   b. In the middle window, **click the + sign** beside **Fields and Values** to see all the fields in the attribute table.
   c. Double-click **PRNAME**. It will show up in the Expression window.
   d. Above the Expression window, **type in LIKE** so it shows up after PRNAME.
   e. On the right, beside **Load values** click the **all unique** button. The result will be all the province names. **Double-click 'New Brunswick / Nouveau-Brunswick'**.
   f. Your Select by expression dialog box should now look like this:
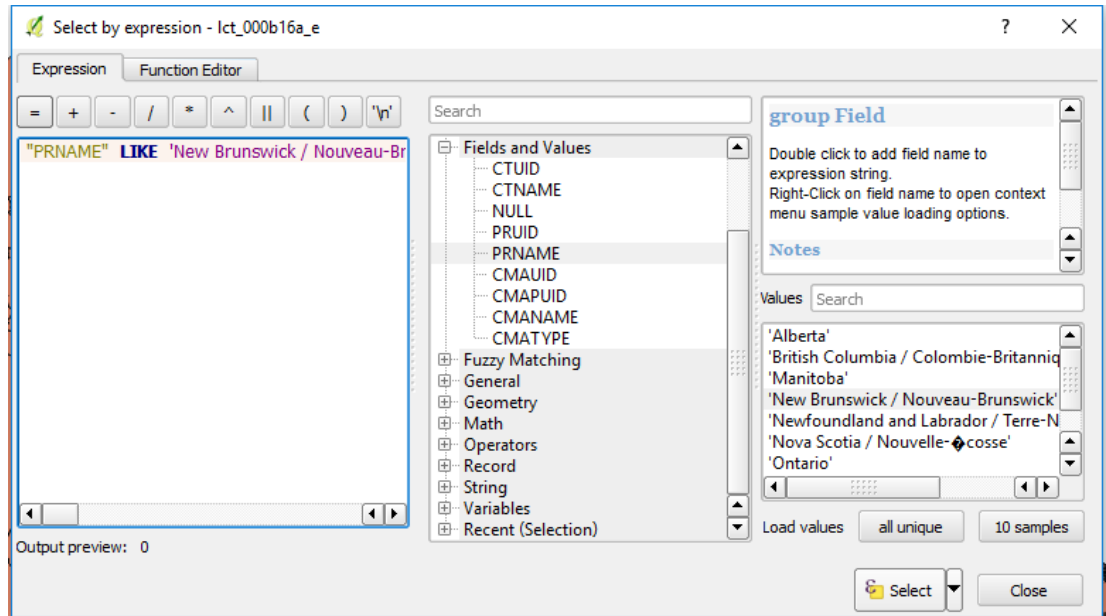
g. Click the **Select** button. Note that the CTs in New Brunswick are now yellow.

h. Click **Close**.

i. In the bottom left of the attribute table, you can click the Show All Features button and select Show Selected Features instead.
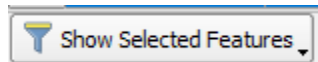
j. Close the attribute table.

5) The selected features are only temporary and we want to deal ONLY with the New Brunswick CTs, so we'll export the currently-selected census tracts to a new shapefile.

    a. Right-click on *lct_000b16a_e* and select **Save As…** from the menu

    b. **Before you do anything else**, select the *Save only selected features* option
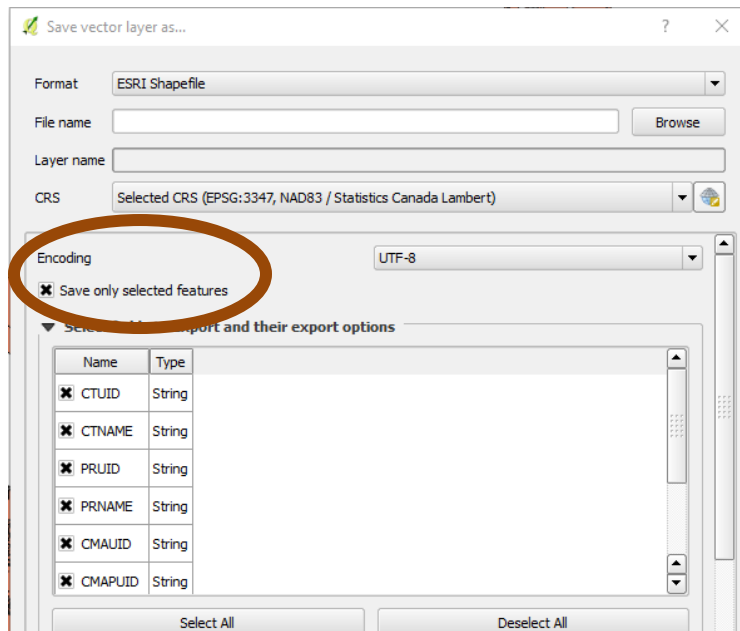
c. Browse to where you would like to save your file. We recommend naming it something helpful, such as ***CTs_NewBrunswick.shp***. Ensure that you save as type **ESRI Shapefile**
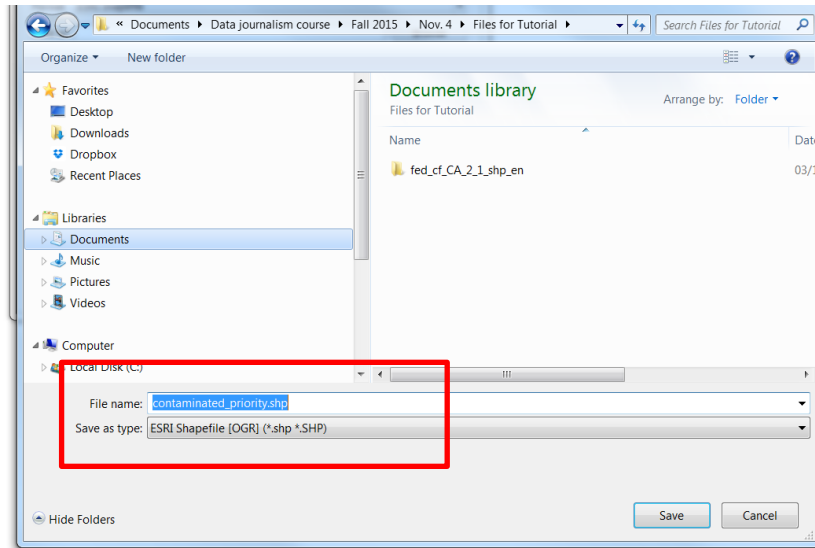
d. Click ***Save*** and then ***OK in the "Save vector layer as" dialog box***.
e. Wait a moment for it to export and add the file.
f. Right-click on ***lct_000b16a_e*** and select ***Remove***.
g. Right-click on CTs_NewBrunswick and select ***Zoom to layer***.

3. Downloading Census Data

We now want to look at the census data for the CTs in New Brunswick to perhaps find a pattern with relation to income.

1) Go to www.statcan.gc.ca
2) Scroll down and click on **Census Program**
3) Scroll down to **Information and Services**, then select **Data Products**
4) Click on **Data Tables**
5) Select **Income**
6) Scroll down to table 12, **Household Income Statistics (3) and Household Type Including Census Family Structure (11) for Private Households**. Click on the table number.
7) Click on the **View/Download this Product** button
8) While you can view details for Census Metropolitan Areas or Census tracts within them, we want to download the entire table. Click on the **Download** tab.
9) Select the CSV for the entire table.

10) Copy it to your project folder and unzip it.

Shapefiles		10/17/2017 3:13 PM	File folder	
98-400-X2016100_ENG_CSV.ZIP		10/16/2017 9:21 AM	Compressed (zipp...	949 KB
98-400-X2016100_English_CSV_data.csv		9/19/2017 12:35 PM	Microsoft Excel C...	7,091 KB
98-400-X2016100_English_meta.txt		9/19/2017 12:35 PM	Text Document	22 KB
Geo_starting_row_CSV.csv		9/19/2017 12:35 PM	Microsoft Excel C...	161 KB
README_meta.txt		9/19/2017 12:37 PM	Text Document	2 KB

11) In QGIS, go to **Layer** > **Add Layer** > **Add Delimited Text Layer…**
12) Fill in the following information then click OK
    a. File format: **CSV**
    b. Geometry definition: **No geometry (attribute only table)**

13) Right-click on **98-400-X2016100_English_CSV_data** and open the attribute table.
14) You'll note that in the **GEO_CODE (POR)** field, the New Brunswick codes start with a 3. We will extract all the New Brunswick features.
    a. Click the Select by expression button
    b. Your final expression should be **"GEO_CODE (POR)" > 3000000 AND "GEO_CODE (POR)" < 4000000**
    c. Close the dialog box, note that 1199 features are selected.
    d. Export the selected features as CSV and call it something like
       CTdata_NewBrunswick.csv
    e. The export may take a moment or two
15) To further whittle down the data, when you open the CTdata_NewBrunswick CSV you may notice that there are multiple household types. We want to look at **Total Household Income** so we'll extract those values.
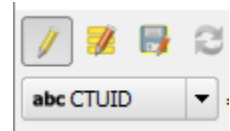   a. Click the Select by expression button
    b. Your final expression should be **"DIM: Household type including census family structure (11)" LIKE 'Total - Household type including census family structure'**
    c. Close the dialog box, note that 109 features are selected.

d. Export the selected features as CSV and call it something like **CTdata_NewBrunswick_Totalincome.csv**

16) Open the attribute table for the new CT CSV file. Note that the **GEO_CODE (POR)** attributes have decimals, even .00. Close the attribute table.

17) The decimals are key and we need to make sure that the **CTs_NewBrunswick** file has a matching field. Open the attribute table for the shapefile. The **CTUID** field has the same kind of data. We should be set!
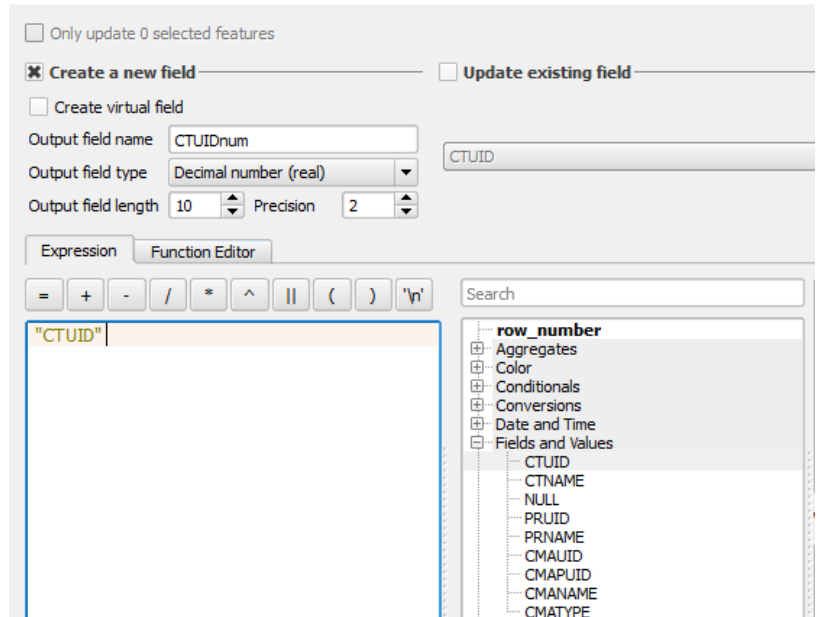
4. Ensuring the data can be joined

In order to join the **CTs_NewBrunswick** shapefile and the **CTdata_NewBrunswick_Totalincome.csv** together, *we need to have a field that includes identical data*. For example, it's not possible to join a text (string) field to a number field. Specifically in this case, the CT identification numbers have decimals (.00, .01, etc.) and those are critical. *We need to make sure those fields are both numeric and have 2 decimal places for each entry*.

1) Open the attribute table for **CTs_NewBrunswick** shapefile

2) Click the **Toggle editing mode** button

   a. Note that the CTUID field has a little abc icon beside it. This means it's a text field and we want it to be a numeric field.

3) Click the **Open field calculator** button

   a. Make sure **Create a new field** is checked

   b. In **Output field name**, enter **CTUIDnum**

   c. **Output field type** needs to be Decimal number (real)

   d. **Output field length** is 10, **Precision** is 2

      i. Length is how many digits are in the number, and precision is how many digits appear after the decimal; we want .00 to show up for whole numbers so we'll force that precision by giving it a value of 2

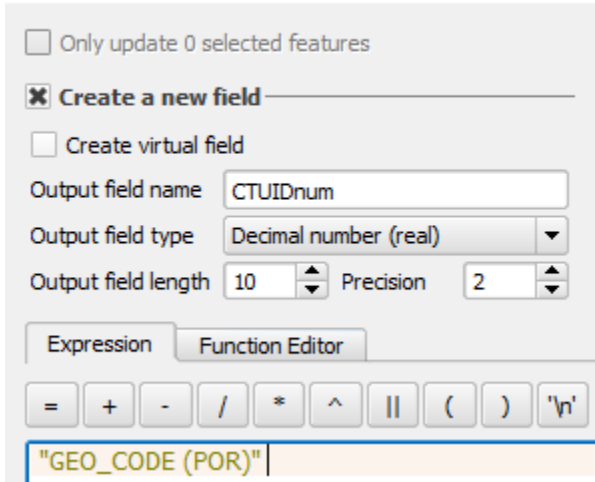   e. In the Expression area, click **Fields and Values** and then double-click **CTUID**

   f. Click OK

4) In your attribute table, you should see a new field on the far right called CTUIDnum

5) Click the **Save edits** button

6) Click **Toggle editing mode** to stop editing

7) Now, we need to do the same thing for **CTdata_NewBrunswick_Totalincome.csv**.

8) Open the attribute table for **CTdata_NewBrunswick_Totalincome.csv**

9) Repeat steps 2-6 except you want to copy the values in the **GEO_CODE (POR)** field, so the result should look like this.
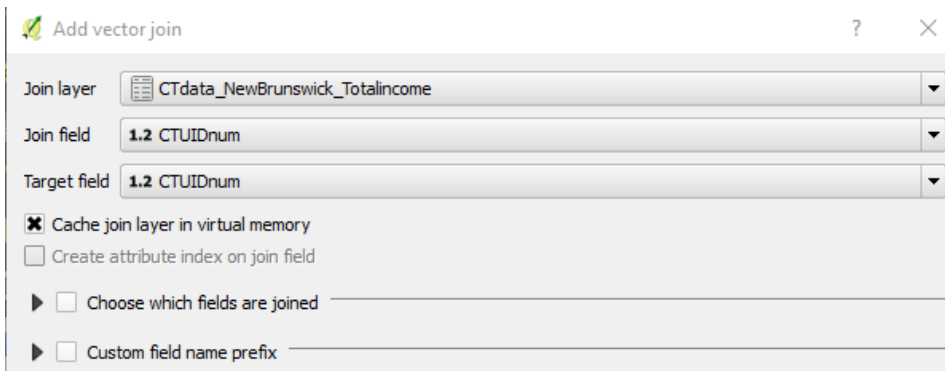

10) Save your edits, close the editing mode, and close the attribute tables.

5. Joining Census Data

We're at the most critical point! Attention to detail is key, as you've undoubtedly noticed.

1) Right click on the **CTs_NewBrunswick** polygon file and select **Properties**.
2) Click Joins, then click the green plus sign
3) Ensure that join of the table data to the polygon is based on a common field. In this case, it's **CTUIDnum** and **CTUIDnum** which are the unique census tract identification numbers that we just made sure are numeric and have all the required decimal places.

4) Click **OK**. Then **Apply**. Then **Close**.
5) Right-click on **CTs_NewBrunswick** to check that the join worked. You should see data added to the attribute table:

	CTUID	CTNAME	PRUID	PRNAME	CMAUID	CMAPUID	CMANAME	CMATYPE	nswick_Totalincome	unswick_Totalincom
1	3100026.00	0026.00	13	New Brunswick / ...	310	13310	Saint John	B	2016	2
2	3100028.00	0028.00	13	New Brunswick / ...	310	13310	Saint John	B	2016	2
3	3100029.00	0029.00	13	New Brunswick / ...	310	13310	Saint John	B	2016	2
4	3050003.01	0003.01	13	New Brunswick / ...	305	13305	Moncton	B	2016	2
5	3050010.01	0010.01	13	New Brunswick / ...	305	13305	Moncton	B	2016	2
6	3100100.04	0100.04	13	New Brunswick / ...	310	13310	Saint John	B	2016	2
7	3200018.00	0018.00	13	New Brunswick / ...	320	13320	Fredericton	K	2016	2
8	3200019.00	0019.00	13	New Brunswick / ...	320	13320	Fredericton	K	2016	2
9	3050003.04	0003.04	13	New Brunswick / ...	305	13305	Moncton	B	2016	2
10	3200020.00	0020.00	13	New Brunswick / ...	320	13320	Fredericton	K	2016	2
11	3200021.00	0021.00	13	New Brunswick / ...	320	13320	Fredericton	K	2016	2
12	3050004.00	0004.00	13	New Brunswick / ...	305	13305	Moncton	B	2016	2
13	3200008.00	0008.00	13	New Brunswick / ...	320	13320	Fredericton	K	2016	2
14	3200004.00	0004.00	13	New Brunswick / ...	320	13320	Fredericton	K	2016	2
15	3100130.03	0130.03	13	New Brunswick / ...	310	13310	Saint John	B	2016	2
16	3050014.03	0014.03	13	New Brunswick / ...	305	13305	Moncton	B	2016	2

CTs_NewBrunswick :: Features total: 111, filtered: 111, selected: 1

Show All Features

6) Right click on **CTs_NewBrunswick** > **Save as…** and save the file as **CTs_NewBrunswick_IncomeData.** The join is now complete.
7) **Remove all the CSV files and the CTs_NewBrunswick shapefile** so we don't get confused.

6. Checking feature properties for symbolization

In order to properly symbolize, we need to make sure that the numbers you see in the attribute table are actually numbers and not text.

1) Open the properties for **CTs_NewBrunswick_IncomeData**
2) Click the **Fields** option from the menu
3) Under the **Type** name field, all of the fields are String. This means we'll have to change some to numbers.
4) The **CTdata_N10 field is Median total income of households in dollars**, which you can find out by looking at the original CSV downloaded from Statistics Canada. We'll convert that to a numeric field.
5) Close the Properties.
6) Open the attribute table.

7) Click the **Toggle editing mode** button

8) Click the **Open field calculator** button

a. Make sure **Create a new field** is checked

b. In **Output field name**, enter MedianInc

c. **Output field type** can remain a Whole number (integer)

d. In the Expression area, click **Fields and Values** and then double-click **CTdata_N10**

e. Click OK

9) Scroll to the right to see the new MedianInc field

10) Click the **Save edits** button

11) Click Toggle editing mode again
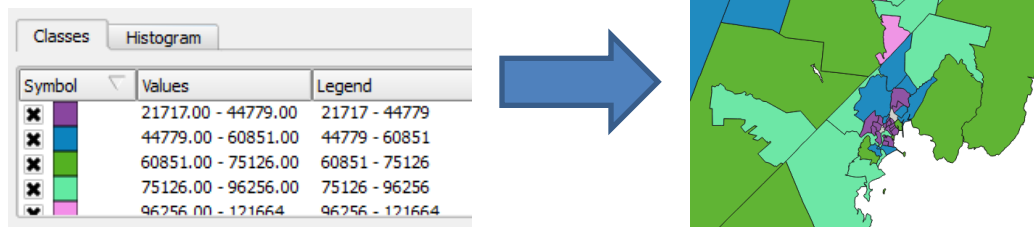
12) Now we can symbolize!

7. Symbology

1) To symbolize the **CTs_NewBrunswick_IncomeData** shapefile based on the MedianInc field, close the table and in the table of contents right-click **CTs_NewBrunswick_IncomeData** and select **Properties.**

2) Click the **Style** tab in the Layer Properties dialog box.

3) Click **Graduated** in the drop-down box at the top

4) For Column, select **MedianInc**

5) Select a colour ramp, then click the **Classify** button

6) It should be filled in as you see below:
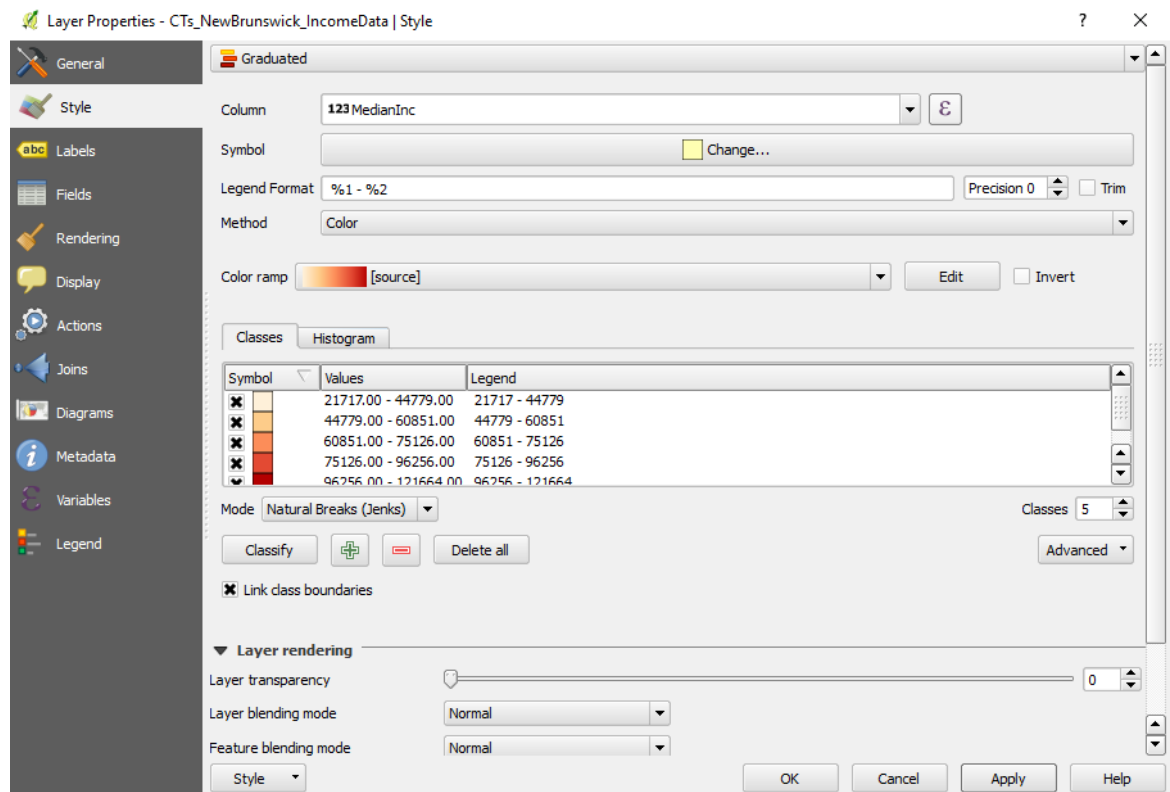
(NOTE: depending on your version of QGIS, your dialog box may look slightly different.)
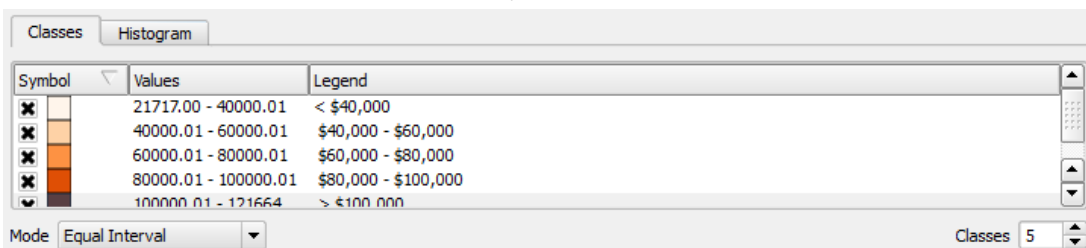
a. Column should be **MedianInc**

Prepared by Rebecca Bartlett, Carleton University Library, October 2017

b. Under the **Classes** tab, Mode should be **Natural Breaks (Jenks)** with 5 classes
   i. You can play around with this if you'd like
c. Pick your favourite colour ramp. As you are showing intervals, go with something that's the same hue but with varying intensity. If you go with a many-coloured ramp your map becomes meaningless, like so:

Classes	Histogram	
Symbol	Values	Legend
✖	21717.00 - 44779.00	21717 - 44779
✖	44779.00 - 60851.00	44779 - 60851
✖	60851.00 - 75126.00	60851 - 75126
✖	75126.00 - 96256.00	75126 - 96256
⌄	96256.00 - 121664	96256 - 121664

7) Your on-screen map should now look something like this, with the darker hues showing us which CTs have higher and lower median household incomes.
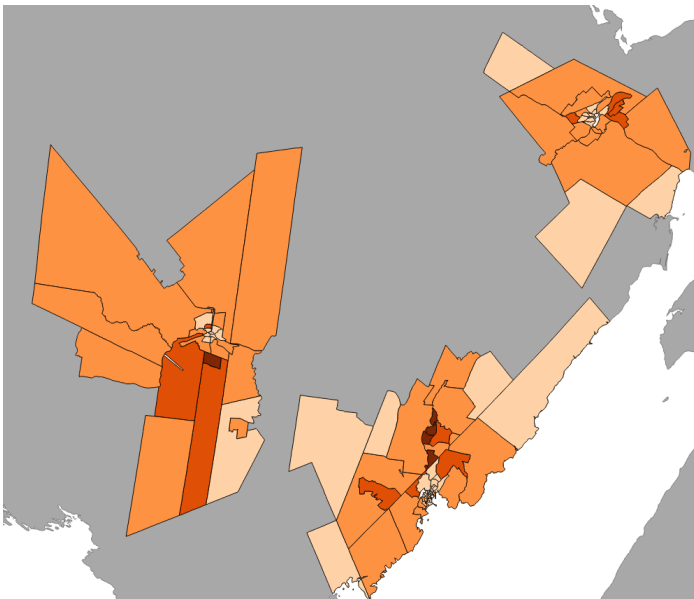
8) You can also manually adjust the intervals. For example, you can set the classes to 5 and then edit the values to reflect rounder numbers, like so:

Classes	Histogram	
Symbol	Values	Legend
✖	21717.00 - 40000.01	< $40,000
✖	40000.01 - 60000.01	$40,000 - $60,000
✖	60000.01 - 80000.01	$60,000 - $80,000
✖	80000.01 - 100000.01	$80,000 - $100,000
⌄	100000.01 - 121664	> $100,000

Mode Equal Interval ▼   Classes 5 ⇕

You can also edit the legend text, as you can see above, to include dollar signs.

8. Creating a Map Layout

To create a map layout in QGIS, you need to make a Composer. This will enable you to place and arrange map elements such scale bar, a legend, and a map title.

1) Start off by right-clicking the layer you want to create a layout for, and select **Zoom to layer**.
2) In the menu, select *Project > New Print Composer*.
3) In the Composer title dialog box, give your new composition a name. Click **OK**.
4) A new Composer window will appear. To add your map of NB CTs, click *Layout > Add Map*, then draw a box that is just inside the perimeter of the page.
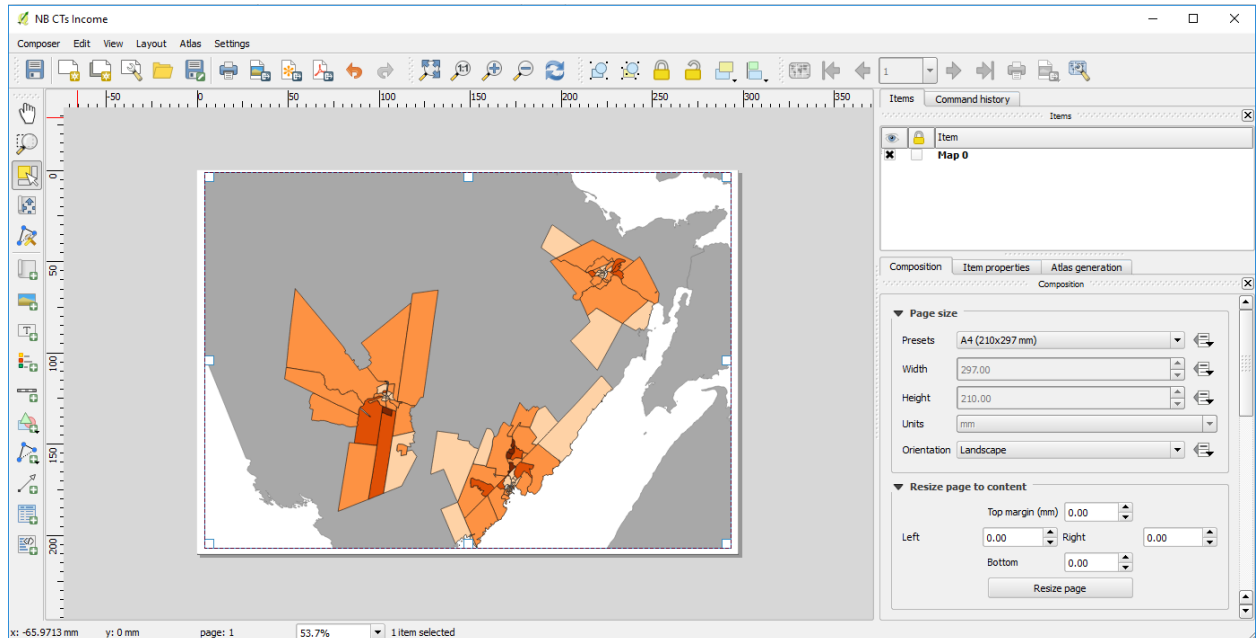
5) Under Item Properties tab, under Extents, click the **Set to map canvas extent** button to centre the layer.
6) If the CTs are either too large or too small for your liking, fiddle with the **Scale** under **Item properties** on the right.
   a. A larger number will make the CTs smaller and a smaller number will make the CTs larger.
7) The QGIS composer is remarkably finicky. Maneuver (including deleting and re-adding) the map around until you're reasonably satisfied.

Adding Other Map Elements
Adding a Legend

1. In the menu, select *Layout > Add Legend*.
2. Draw a box where you would like your legend to appear.
3. By default, all the layers on the map will appear as legend items in the legend.

4. To remove a legend item, you need to **uncheck the Auto update checkbox** and then manually remove the layers you don't want in your legend using the red minus button.

5. To change the name of a legend item as it shows up in the legend, click the legend item and then the **Edit** button to change its title in the legend.

6. If you would like to change the title of your legend (or prefer that it doesn't have a title at all), change it in the Title textbox under **Main properties**.

7. There are a lot of options in the Legend panel, from font selections to spacing to background colour.

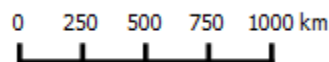8. You can click and drag the legend into place on your map.

Adding a North Arrow

1. In the menu, select *Layout > Add Image*.
2. Draw a box where you would like to see your north arrow.
3. In the **Item properties** panel on the right, click on the **Search Directories** submenu, and under **Image search paths** find the one ending in **/arrows**.
4. Select an arrow that meets your needs. You could always design your own if you feel creative, or just go for a Star Trek vibe (never a bad idea):

Adding a Scale Bar

1. In the menu, select *Layout > Add Scalebar*.
2. Draw a box where you would like to see the scale bar.
3. In the Item properties tab, you can select one of six styles under the **Style** dropdown.
4. Choose the units for the scale bar under **Units**.
5. Adjust the segments in the **Segments** area.
   a. E.g.: left 0 and right 4 on a Line Ticks Up style will look something like this:

   0   250   500   750   1000 km

6. Click and drag the scale bar into place on your map.

Adding a Title to a Map

1. In the menu, select *Layout > Add Label*.
2. Draw a box where you'd like your title to be (you can move it later).
3. Type your title into the **Main properties** text box.
4. Adjust the font as you see fit and move it around until you resign yourself to the lack of design elements included in QGIS.

Household Income in New Brunswick Census Tracts, 2016
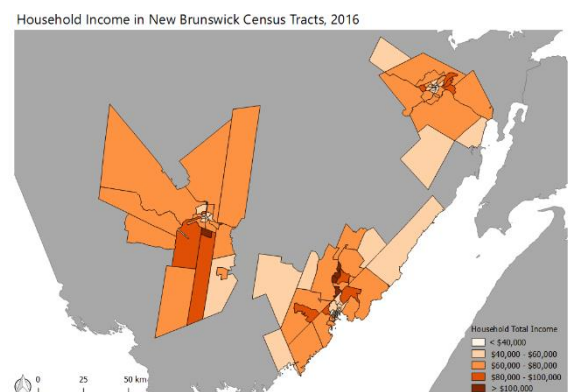
This is not a very pretty map. You can make a nicer one.

9. Geocoding (2016)

We've looked at a lot of files today but most were already in a mapping format. The following section will describes the steps to geocode a file that only street addresses. We'll take a look at inspection violation file, from which we've extracted Tim Hortons stores. We'll explore two methods of geocoding as well as the format and cleaning-up process that's required.

Method 1 – Using Google Maps / OpenStreetMap

This method uses the Google Maps API or OpenStreetMap API to geocode the location of the points. This is optimal if you have a thousand points or less to geocode.
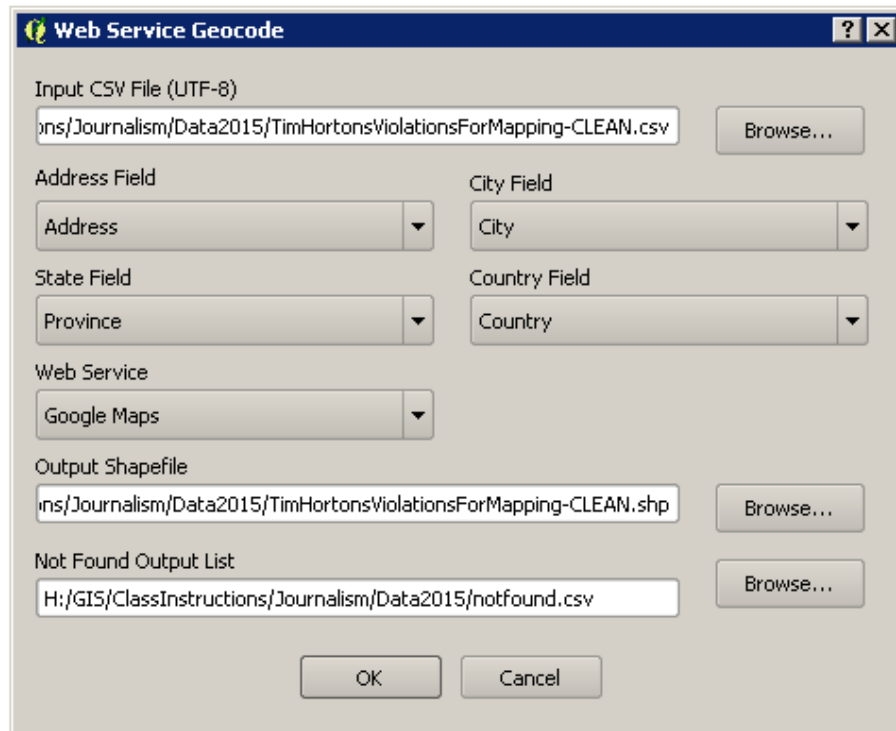
Cleaning-up the data:

1) In Excel, open the Tim Hortons file - **TimHortonsViolationsForMapping.csv**.
2) Clean-up the file to have the address elements in separate columns.
   a. Move the content of **row C** to **row G**
   b. Highlight **Row B**, from **Find and Replace**, select **Replace**. Do the following Find & Replace
      i. Replace "**S** " (S with two spaces) with a S comma
      ii. Replace " , "(space comma space) with a comma
      iii. Replace ", " (comma space) with a comma
      iv. Replace " " (two spaces) with one space
   c. Highlight **Row B**, click on **Text to Columns**
      i. Select Delimited, space as the delimiter.
   d. Scroll through your data to ensure that all data was changed properly. Line 87 has to be changed.
   e. Change the column headers to **Store, Address, Province, Country**
   f. Save as a new Excel file - **TimHortonsViolationsForMappingCLEAN.csv**

Geocoding in QGIS

1) In **QGIS**, go to **Plug-ins** > **Manage and Install Plug-ins** and ensure that MMQGIS is selected.
2) In the menu, click on **MMQGIS** > **Geocode** > **Geocode CSV with Google / Open StreetMap**
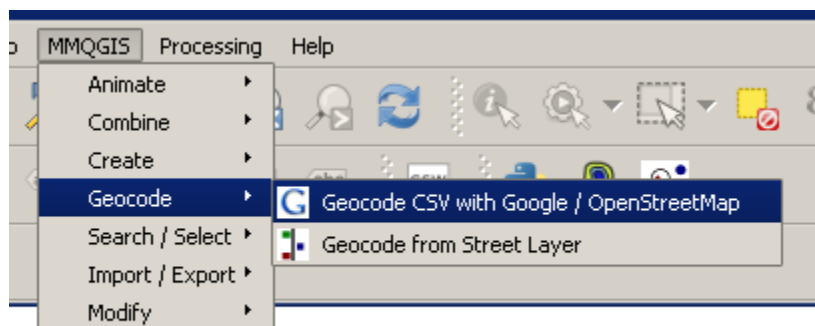
   a. Ensure that the following is selected for the geocoding. You'll have to match the proper fields to the geocoding field. (ie. Address Field needs to correspond to the filed that has an address – 123 colonel by drive).

   b. Click **Ok.** Geocoding will run.

Method 2 – Using Street Layer

This method uses an existing street layer file to geocode your points. In this case, we've downloaded the street file from Ottawa Open data.

Cleaning-up the data:

   1) In Excel, open the Tim Hortons file - **TimHortonsViolationsForMappingCLEAN.csv**.
   2) Clean-up the file to have the address elements in separate columns.
      i. Use **Find&Replace** and **TextToColumns** to separate address number and streetname into separate columns.
      ii. Use the **Find&Replace** function to change the street type suffix to the acronym. (ie.Road to RD, Street or ST, etc.)
      iii. Save the file as **TimHortonsViolationsForMapping-CLEAN-STREET.csv**

Geocoding in QGIS

   1) In QGIS, add the **Road_Segments.shp** to QGIS (this was downloaded from Ottawa OpenData)
   2) Go to **MMQGIS** > **Geocode** > **Geocode from Street Layer**
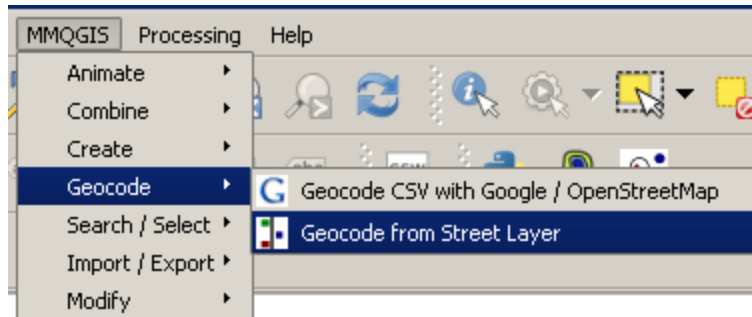
3) Fill in the following information in the dialog box. Ensure that you select **ROAD_NAME_** and not **ROAD_NAME**

4) Click **Ok** to run the geocoding.
5) Once complete, you'll notice that it didn't geocode all of the files. Take a look at the notfound.csv file to see which records weren't geocoded. Can you tell us why?