Census Data: Access, Mapping and Visualization

Trent University is a member of the Data Liberation Initiative (DLI), an agreement that provides academic institutions with access to otherwise restricted data files from Statistics Canada. These include spatial boundary files and statistical data at finer levels of geography, such as dissemination areas (DA), enumeration areas (EA), forward sortation areas (FSA), and census tracts (CT). Registered members of participating post-secondary institutions are permitted access to the collection under the restriction that data is used for research and teaching purposes only.

A number of these data sets can be downloaded directly through Trent Library's subscription to several Internet data systems, including the University of Toronto's CHASS Census Analyser and Scholars Portal's ODESI.

Links to Statistics Canada's census Web sites and to a variety of data providers can be found on MaDGIC's site. Please note that if you are attempting to access licensed databases from a computer that is not on the Trent network, you will be prompted to enter your Trent user ID and password at some point in the process.

The steps below outline the process of downloading and mapping a subset of DLI data from the *CHASS Census Analyser.* For an in-person demonstration, or for any request related to statistical data, please contact us at madgichelp@trentu.ca.

Step 1   Downloading census data with CHASS Census Analyser

Visit the Bata Library home page and browse to *Locations > MaDGIC.* On the MaDGIC home page, select *Statistical Data* from the left menu and then select *CHASS Census* from the results. A short description of the CHASS tools will be provided. Use the active link to *download data from the Canadian Census Analyser*.

Note: *If you are accessing the CHASS tool from a computer that is connected to the Trent network, you will be provided seamless access to the database. However, if you are working off-campus then you will be asked to identify yourself by signing in with your Novell credentials.*

The CHASS Welcome page allows you to select various starting points to discover data, such as browsing Census Profile Tables by geography level or census year, as well as searching microdata and postal code conversion files.

1. **Census Profiles Tables**
   - by Census Geography:
     - Canada, Provinces and Territories
     - Federal Electoral District
     - Census Division
     - Census Subdivision
     - Census Tract
     - Enumeration area / Dissemination area
   - by Census Year
     - 2011 NHS

2. **Public use microdata files (PUMF)**
   - Census PUMF files
   - NHS PUMF files **-new 2014-07-31-**

The first step in the process is to select your starting point for locating statistical files. In this example, we have chosen the option to browse profile tables by census tract. The figure below displays the results of this selection. Select a profile table by clicking its title to proceed to the next step.

2011 NHS (cumulative)
2011 (cumulative)

2006 (cumulative)

- Age and Sex
- Marital status, Common-law status, Families, Dwellings and households
- Language, immigration, citizenship, mobility and migration
- Aboriginals people
- Labour market activity, industry, occupation, education, language of work, place of work and mode of transportation
- Ethnic origin and visible minorities
- Income and earnings and housing and shelter costs
- Additional immigration and place of birth

2001 (cumulative)

The subsequent steps in the process require that you specify the level of census geography, census variables, and the output format for your results. The CHASS web pages break these three steps down in order.

1. **Specify Census Geography for retrieval**

   Users are provided with the option to filter by CMA or CA name, or by province. In this example, we have checked the letter 'P' in the list provided, and then selected the Peterborough census agglomeration. Note that you are also provided with the option to filter by province, or to select all geography for the entire Canadian census.

⦿ Locate census geography (CMAs/CAs with CTs) in the database using the following filtering criteria then select the items you would like to include in your search.

| by Name | by Province |

-check all- | -uncheck all-
☐A ☐B ☐C ☐D ☐E ☐F ☐G
☐H ☐K ☐L ☐M ☐N ☐O ☑P
☐Q ☐R ☐S ☐T ☐V ☐W

-check all- | -uncheck all-   -remove all- | -order-

2 items found: Selected items:
☑ Peterborough (Ont.)   ☑ Peterborough (Ont.)
☐ Prince George (B.C.)

◎ Or select all geography

2. **Specify Census Profile variables for retrieval**

   Variables are categorized under three tabs: Age and sex; Marital, families & households; and Language. Click through each tab and select variables of interest by placing a check mark beside the variables that you wish to include in your results. In this example, we have selected the variables of Population, 2011 and Population, 2006, as well as the Population percent change for this time period.

Select the Census Profile variables (-see definitions-) to include in your search.

Age, sex	Marital, families & households	Language

Population and dwelling counts | Age characteristics Total | Age characteristics Males | Age characteristics Females

-check all- | -uncheck all-

Population and dwelling counts
- ☑ Population, 2011 (v1)
- ☑ Population, 2006 (v2)
- ☑ Population percentage change, 2006 to 2011 (v3)
- ☐ Total private dwellings (v4)
- ☐ Private dwellings occupied by usual residents (v5)
- ☐ Population density per square kilometre (v6)
- ☐ Land area in square kilometres (v7)

-remove all-

Selected items:
- ☑ Population, 2011 (v1)
- ☑ Population, 2006 (v2)
- ☑ Population percentage change, 2006 to 2011 (v3)

3. Specify the output details and submit query

The final step in the process involves specifying output details, such as options related to how you would like to see the data, and the output format you would like to use. These options will vary depending on how you plan to use the data. Note that it is crucial to include the unique identifier that relates to your selected level of geography in order to facilitate joining the census data with spatial data in subsequent steps. In this example, since we are interested in mapping profile information by census tract, we choose to include the Census Tract name in our results. If you are planning to visualize the data in a GIS environment, then it is a good idea to select either CSV or MS Excel formats.

Optionally include in the result:

- ☐ Province code   ☐ Province name
- ☑ CMA code   ☐ CMA name
- ☑ Census Tract name

Optionally filter CT names (the format of CT name is xxxx.yy):
(e.g: 1.01-4.00 7-9 59 302)
[   ]

Census variables to be listed as: (apply only to Screen output format)

columns: ⦿ rows: ○

Optionally provide an email address for larger downloads notifications: (apply only to Download to a file output format)

[demo@trentu.ca]

Optionally enable zip file compression: (apply only to Download to a file output format)

none: ○ zip: ⦿

Select the output format:

```
Screen output
   Text
   HTML
   Comma-Separated Values (CSV) for spreadsheet
   MS Excel ready
   SAS
   SPSS
Download to a file
   Comma-Separated Values (CSV) file for spreadsheet
   dBase (DBF) file
```

Submit Query. This will open a new tab or window in the browser.

[Submit Query]

The *Screen output* option will generate HTML results directly in your browser. For example, if you select a screen output of *Comma-Separated Values (CSV) for spreadsheet*, you will see your results formatted as shown below.

```
2011 Census Profiles Files / Profile of Census Tracts

"COL0","GEO UID"
"COL1","CMA code"
"COL2","Census Tract name"
"COL3","Population and dwelling counts / Population, 2011"
"COL4","Population and dwelling counts / Population, 2006"
"COL5","Population and dwelling counts / Population percentage change, 2006 to 2011"

"COL0","COL1","COL2","COL3","COL4","COL5"

5290000.00,529,0000.00,118975.00,116570.00,2.10
5290001.01,529,0001.01,3281.00,3006.00,9.10
5290001.02,529,0001.02,4648.00,4733.00,-1.80
5290002.00,529,0002.00,3853.00,3802.00,1.30
5290003.00,529,0003.00,6162.00,5744.00,7.30
5290004.00,529,0004.00,4386.00,4237.00,3.50
5290005.00,529,0005.00,2242.00,2307.00,-2.80
5290006.00,529,0006.00,5641.00,5610.00,0.60
5290007.00,529,0007.00,4441.00,4715.00,-5.80
5290008.00,529,0008.00,4027.00,3969.00,1.50
5290009.01,529,0009.01,3058.00,2938.00,4.10
5290009.02,529,0009.02,5502.00,5571.00,-1.20
5290009.03,529,0009.03,6278.00,4977.00,26.10
5290009.04,529,0009.04,5475.00,5732.00,-4.50
5290010.00,529,0010.00,2384.00,2406.00,-0.90
5290011.00,529,0011.00,2664.00,2515.00,5.90
5290012.00,529,0012.00,6463.00,5849.00,10.50
5290013.00,529,0013.00,5296.00,4879.00,8.50
5290014.00,529,0014.00,2513.00,2030.00,23.80
5290100.00,529,0100.00,3551.00,3608.00,-1.60
5290101.00,529,0101.00,3254.00,3346.00,-2.70
5290102.01,529,0102.01,6660.00,6675.00,-0.20
5290102.02,529,0102.02,3732.00,3832.00,-2.60
5290103.00,529,0103.00,4924.00,4954.00,-0.60
5290104.00,529,0104.00,2917.00,3012.00,-3.20
5290105.00,529,0105.00,5650.00,5848.00,-3.40
5290106.00,529,0106.00,1372.00,1447.00,-5.20
5290200.00,529,0200.00,3893.00,3996.00,-2.60
5290201.00,529,0201.00,4708.00,4832.00,-2.60
```

The first section of the file provides information related to the data in each column of your results. The second section displays the results in CSV format. The results must be formatted properly for use in a GIS, which means that the column heading information must be separated from the data file. Simply copy the data (beginning with "COL0", "COL1",…) and paste it into a new file in a text editor. Delete the line break between the column headings and the data, so that your final data file is structured exactly like the example below.
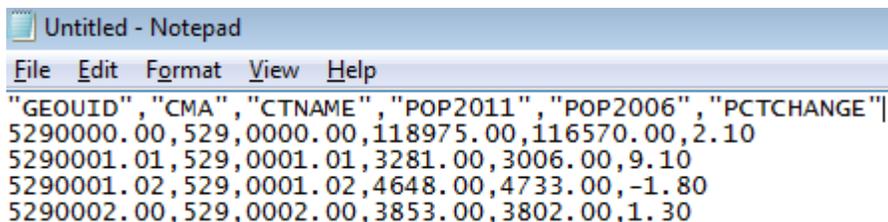

```
"GEOUID","CMA","CTNAME","POP2011","POP2006","PCTCHANGE"|
5290000.00,529,0000.00,118975.00,116570.00,2.10
5290001.01,529,0001.01,3281.00,3006.00,9.10
5290001.02,529,0001.02,4648.00,4733.00,-1.80
5290002.00,529,0002.00,3853.00,3802.00,1.30
```

At this point, it is up to you if you'd like to rename your columns according to the heading information provided in the screen output, as shown in this example. If you choose to leave the defaults, then it is a good idea to save the column heading information to a separate text file so that you can reference it when needed.

Once you have formatted your data file as required, save the file in CSV format to your working directory.
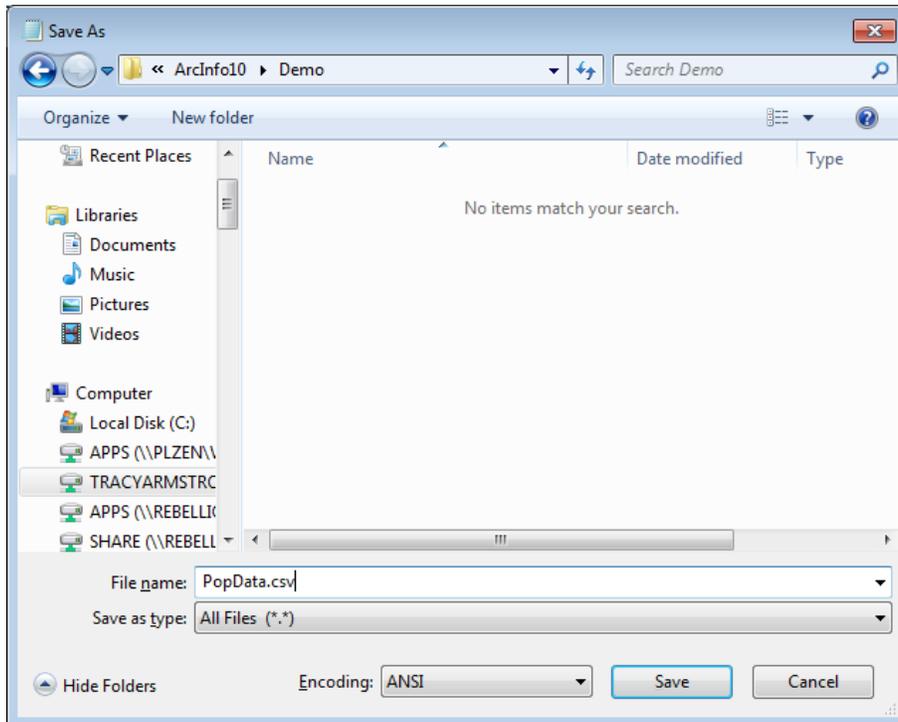
Step 2   Downloading census boundary files

Mapping census profile information requires two data components; the statistical data obtained in the previous steps, and a spatial file (called a *census boundary file*) that is used to link the statistical results to a specific geographical region for mapping and visualization.

Census boundary files are available for download via Scholars GeoPortal. Use the active links from MaDGIC's web page to access the site. The portal offers the option to search, preview, and download datasets licensed by Ontario university libraries, including census boundaries.

***Note:** If you are accessing the site from an off-campus location or a non-Trent computer, then it will be necessary to sign in with your Novell credentials before downloading data.*

There are a number of options for searching data via Scholars GeoPortal. You can use the pre-defined categories to locate census boundaries under the *'Census and administrative boundaries'* heading, or you can refine your search by entering a more specific term to the search box. In this example, we have searched the portal for the term 'census tract' to narrow down the results.

Census Tracts - Cartographic Boundary File (CT-CBF), 2006 Census
Add Details
Producer: **Statistics Canada**
Date published: **2007-03-01 (publication)**
Type of data layer: **Vector**

Census Tracts - Cartographic Boundary File (CT-CBF), 2011 Census
Add Details
Producer: **Statistics Canada**
Date published: **2011-11-01 (publication)**
Type of data layer: **Vector**

Census Tracts - Digital Boundary File (CT-DBF), 2011 Census
Add Details
Producer: **Statistics Canada**
Date published: **2011-11-01 (publication)**
Type of data layer: **Vector**

Census Tracts - Digital Boundary File (CT-DBF), 2006 Census
Add Details
Producer: **Statistics Canada**
Date published: **2007-03-01 (publication)**
Type of data layer: **Vector**

Statistics Canada provides two types of spatial boundary files; cartographic and digital. Cartographic files show geographic areas with only the major land mass of Canada and its coastal islands, while digital files show the full extent of the areas, including coastal waters. For this example, we have chosen to download the digital boundary file from the 2011 census.

To download data, click the blue *Add* button under the dataset to add it to the map. Switch your view to the 'Download' tab at the top of the screen, and select your option to either download by area of interest, or download the entire dataset. Should you choose to download by area of interest, use the tools provided to either select a predefined area, or draw an area on the map. Once you are satisfied with your options, click the blue 'Download' button. The files will be added to 'My download list' – click each file to initiate transfer.

Extract the files to your working directory. You now have everything you need to visualize your statistical data in a GIS environment.

Step 3 Visualizing statistical data with ArcGIS 10.2.x

Open a new, blank project in ArcMap (*Programs > ArcGIS > ArcMap*) and click the 'Add Data'   button.
Browse to your working directory and add the spatial and statistical files to the map.

The first crucial step in visualizing census data is to join the spatial boundary with the statistical data. This step combines the statistical data with its associated census tract, and allows you to then symbolize your map based on the census results. In order to join these two separate data sources, it is essential that you have a common field in each table with matching records. Have a look at the statistical file (right-click and select *Open*), and then compare it with the attribute table of the spatial file (right-click > *Open Attribute Table*). The field GEOUID in the statistical file contains records that match those of the CTUID field in the spatial file. These are the common fields we will use to join the tables.

Unfortunately, ArcGIS software automatically sets the data types for the columns in CSV and Excel files. If you right-click on the GEOUID field and select *Properties*, you will see that the data type for this column is *Double*, while the CTUID column in the spatial boundary layer is *String*. Since it is not possible to join fields with two different data types, it is necessary to reformat one of the fields to match the other. This is most easily accomplished by adding a new field to the spatial file using the steps below.

1. Open the spatial attribute table (right-click > *Open Attribute Table*) and select the *Table Options*   menu from the top left corner of the window. Select the option *Add Field.*
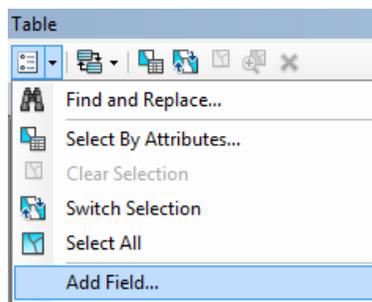

2. In the Add Field dialog, name the new field GEOUID and select a data type of DOUBLE, then click OK. A new field is added to the spatial attribute table.
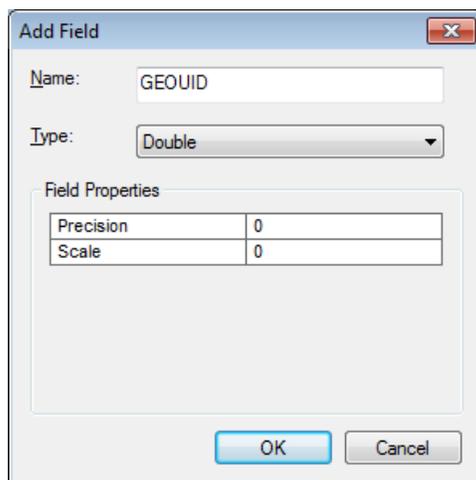
3. Start an editing session. If you do not see the *Editor* toolbar on your screen, browse to *Customize > Toolbars* and place a check mark beside *Editor* in the list. Then, click the *Editor* dropdown to open the menu and select *Start Editing*. If you are asked which layer you would like to edit, select the spatial file from the list and click *OK*.
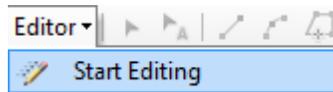

Right-click the heading of your new field and select *Field Calculator*. Place your cursor into the text box and then double-click CTUID from the field list to enter it to the expression. Click *OK*. The new GEOUID field will be populated with the values from CTUID.
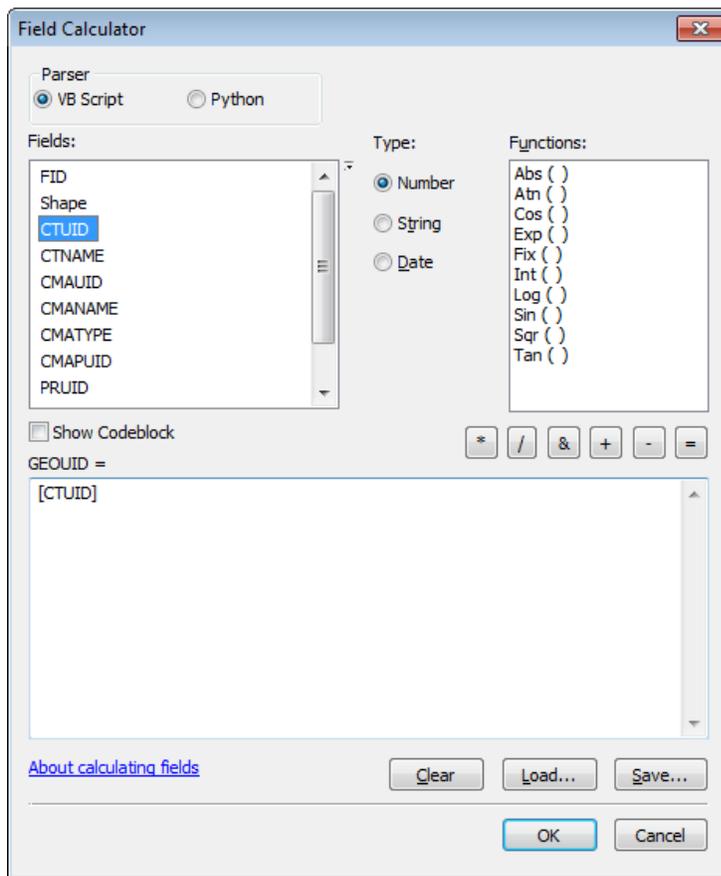

4. Click the *Editor* dropdown and select *Stop Editing.* Click *Yes* to save your edits when prompted.

You are now ready to join the GEOUID fields. Right-click the spatial file in the ArcMap table of contents and select *Joins and Relates > Join.*
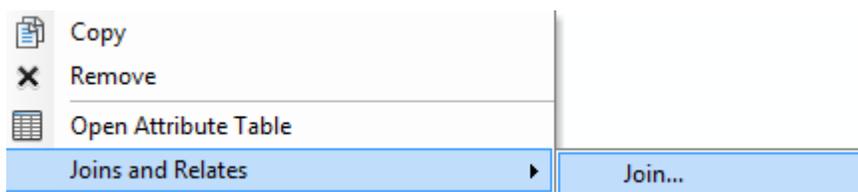
In the *Join Data* dialog, specify GEOUID as the field in the layer that will be used in the join. Select the table you wish to join to the layer, and specify the table's GEOUID field as the second join field. Select the option to Keep only matching records – your spatial file holds census tract boundaries for all of Canada, but for this example we are only interested in those that are situated within the Peterborough census area. Click OK to perform the operation.
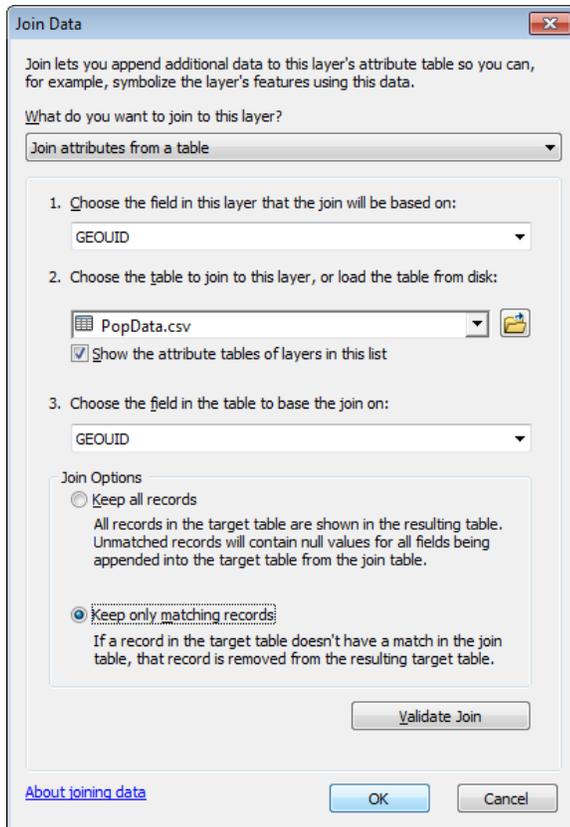
Right-click the layer in the table of contents and select *Open Attribute Table*. You should see the original attribute fields along with the joined table appended at the right. The sample below shows the two GEOUID fields and the joined statistical data fields highlighted in blue. This join has made it possible to visualize the population data on your map document.

CTNAME	CMAUID	CMANAME	CMATYPE	CMAPUID	PRUID	PRNAME	GEOUID	GEOUID	CMA	CTNAME	POP2011	POP2006	PCTCHANGE
0103.00	529	Peterborough	B	35529	35	Ontario	5290103	5290103	529	103	4924	4954	-0.6
0009.03	529	Peterborough	B	35529	35	Ontario	5290009.03	5290009.03	529	9.03	6278	4977	26.1
0006.00	529	Peterborough	B	35529	35	Ontario	5290006	5290006	529	6	5641	5610	0.6
0014.00	529	Peterborough	B	35529	35	Ontario	5290014	5290014	529	14	2513	2030	23.8
0005.00	529	Peterborough	B	35529	35	Ontario	5290005	5290005	529	5	2242	2307	-2.8
0200.00	529	Peterborough	B	35529	35	Ontario	5290200	5290200	529	200	3893	3996	-2.6

Close the attribute table, then right-click the layer name in the table of contents and select *Properties.* Click the *Symbology* tab to view the symbology menu.

There are a number of options to symbolize attribute data for both discrete and continuous values. In this example, we wish to display the percent change of the population by Peterborough census tract between 2006

and 2011. We select *Quantities* from the options on the left, and then set the *Value* field to PCTCHANGE – the field that holds the values we wish to symbolize. We then select a colour ramp and apply the symbolization. The software will attempt to use the Natural Breaks method to symbolize the data using five classes. You can manually set the number of classes and break values using the tools provided under the *Classification* section.

Feel free to explore the various options for symbolization. The methods and options will change depending on the nature of your data and the message you are trying to convey. When you are satisfied with your symbolization, click *OK*. The new symbology will be reflected on the map.
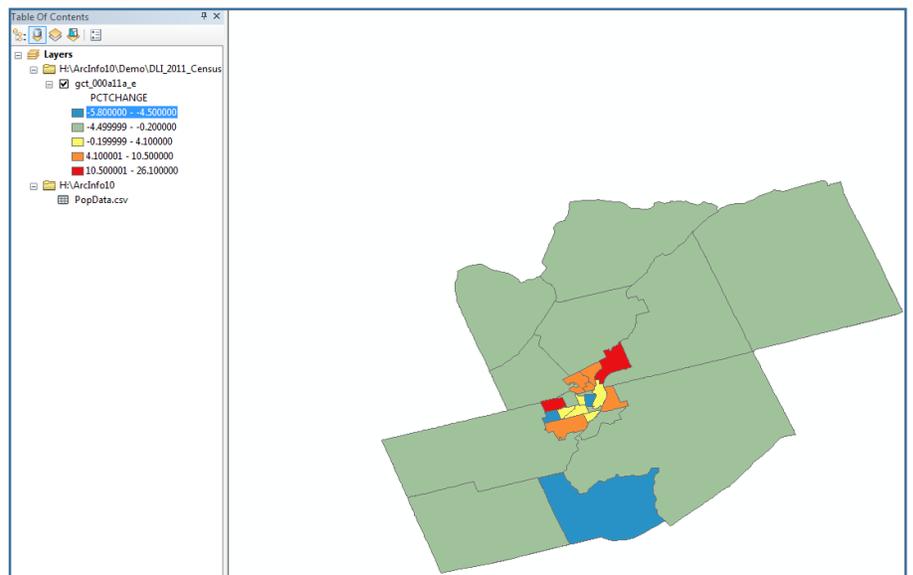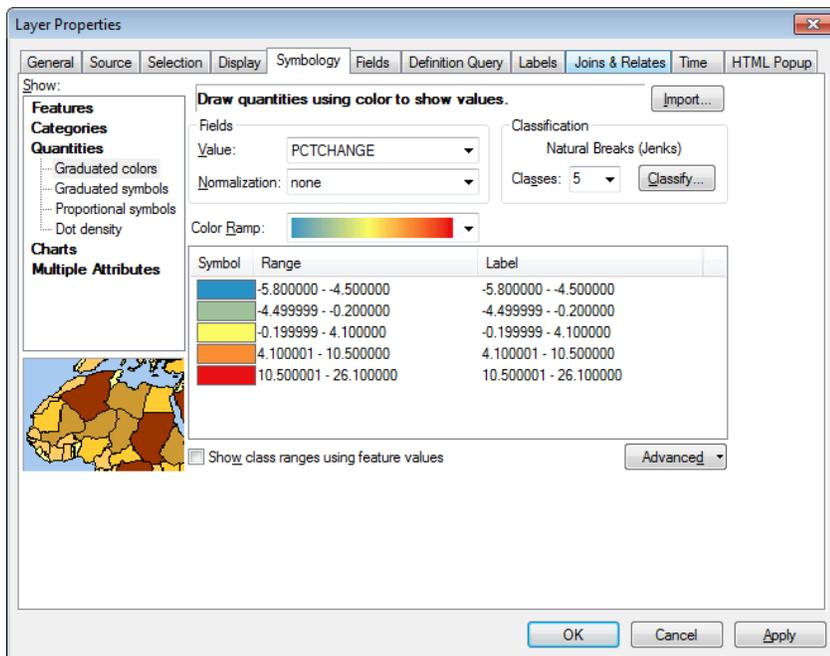
For further assistance with this guide or other topics related to geospatial and/or statistical data and software, please contact us at madgichelp@trentu.ca.