Micro-module A: Online Urban Data Gathering

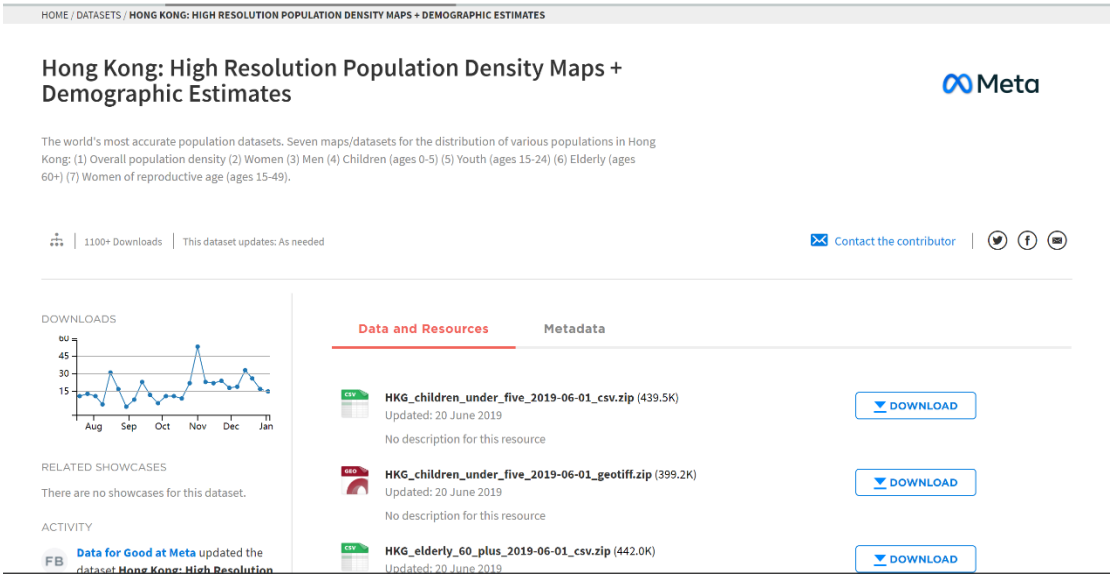
A5- Geo-tagged Socioeconomic Data Processing

In this micro module, we will take census data in Hong Kong as a case study to learn how to obtain and process socioeconomic metrics, which could associate with urban morphology analytics.
In the first part, this tutorial will introduce how to process high-resolution GeoTIFF to show the population density of large scale, for example, overall Hong Kong.
In the second part, this tutorial will introduce how to join small street-level census data to a spatial vector layer, that could apply to the local-scale analysis.

1. Processing population density in geo-tiff of overall Hong Kong

1.1 To obtain and input geo-tiff in QGIS

Firstly, we could download the population density of tiff and geo-tagged csv from the followed website.

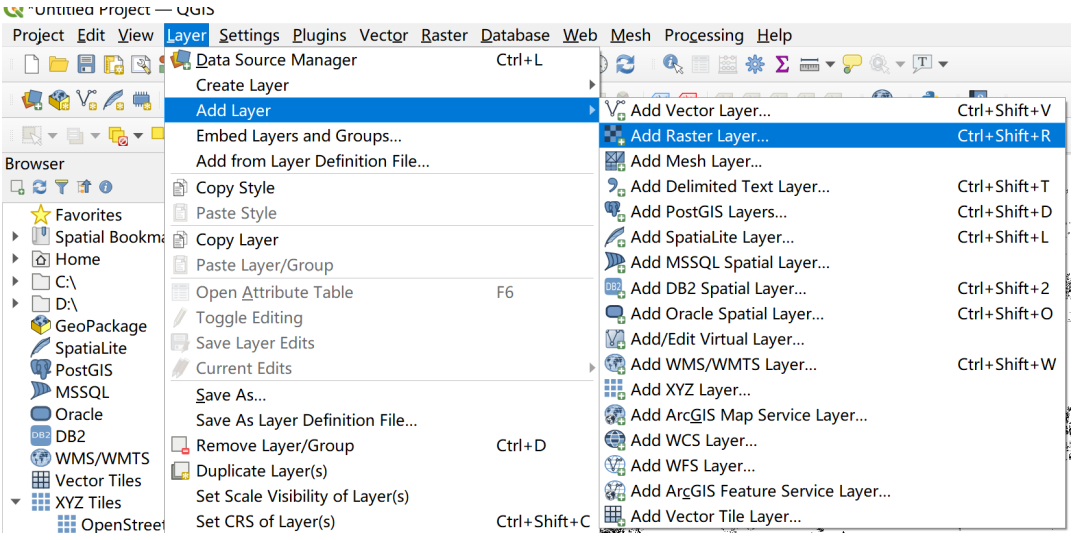


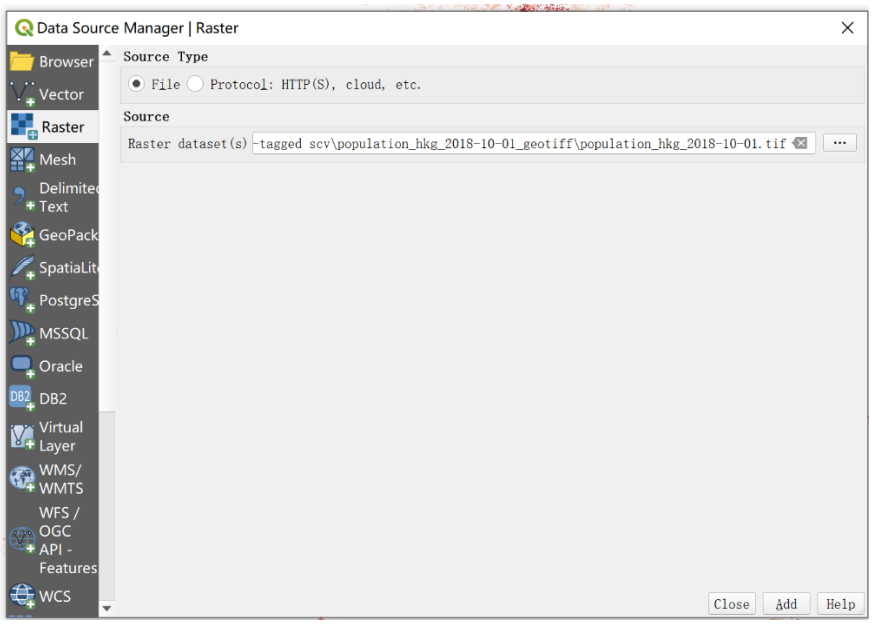
https://data.humdata.org/dataset/worldpop-population-density-for-china-hong-kong-special-administrative-region

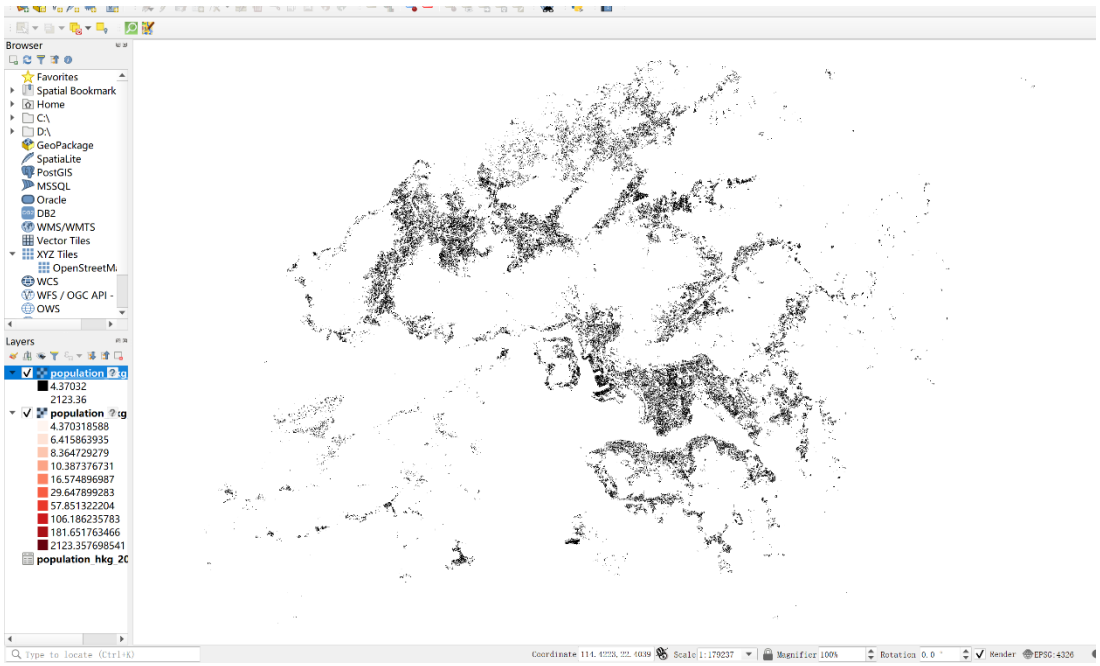
The csv has latitude and longitude values, that could be imported into QGIS with geo-information.

To import the geo-tiff, click ‘Layer’- ‘Add Raster Layer’, and find the source of the tiff file.

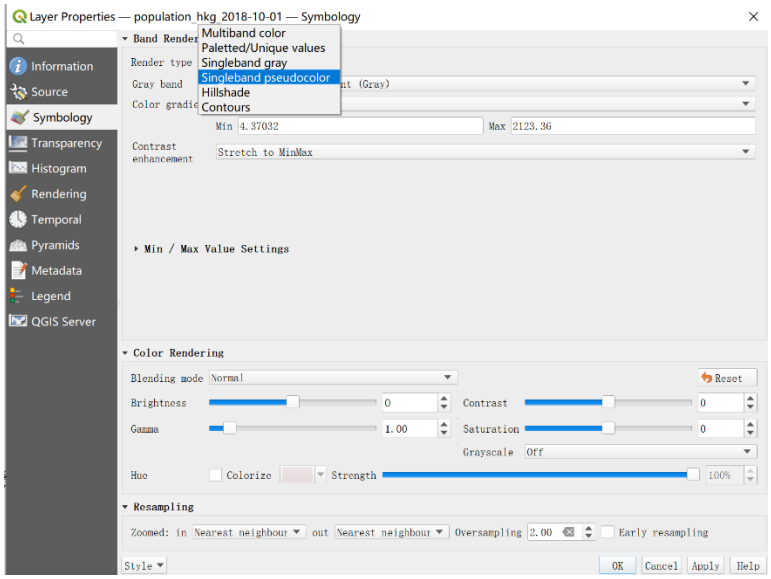
	A	B	C	D	E
1	latitude	longitude	population_2015	population_2020	
2	22.16014	113.899	9.337328933	9.659726936	
3	22.16347	114.2532	17.16392845	17.44693997	
4	22.16375	114.2543	17.16392845	17.44693997	
5	22.16431	114.2513	17.16392845	17.44693997	
6	22.16486	113.9101	9.337328933	9.659726936	
7	22.16486	113.9121	9.337328933	9.659726936	
8	22.16486	114.2521	17.16392845	17.44693997	
9	22.16486	114.2543	17.16392845	17.44693997	
10	22.16514	114.2529	17.16392845	17.44693997	
11	22.16514	114.2535	17.16392845	17.44693997	
12	22.16514	114.2546	17.16392845	17.44693997	
13	22.16542	113.9118	9.337328933	9.659726936	
14	22.16542	114.254	17.16392845	17.44693997	
15	22.16597	113.9113	9.337328933	9.659726936	

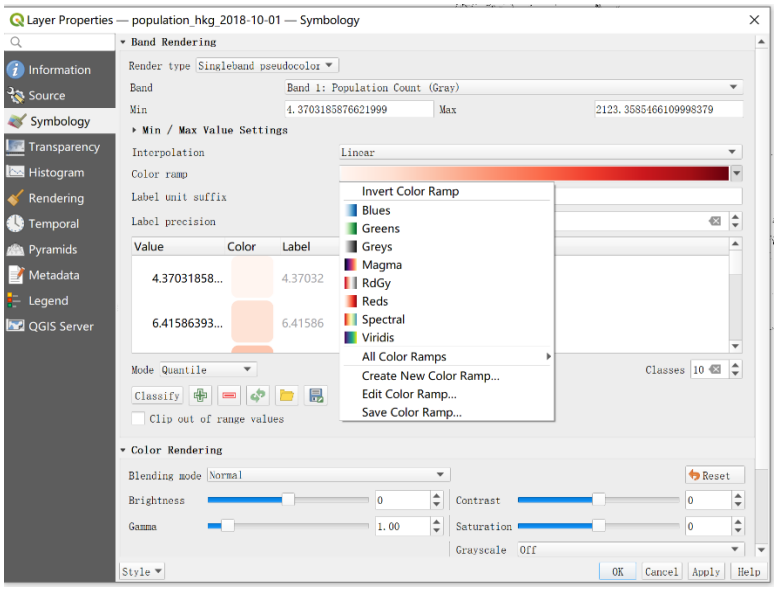




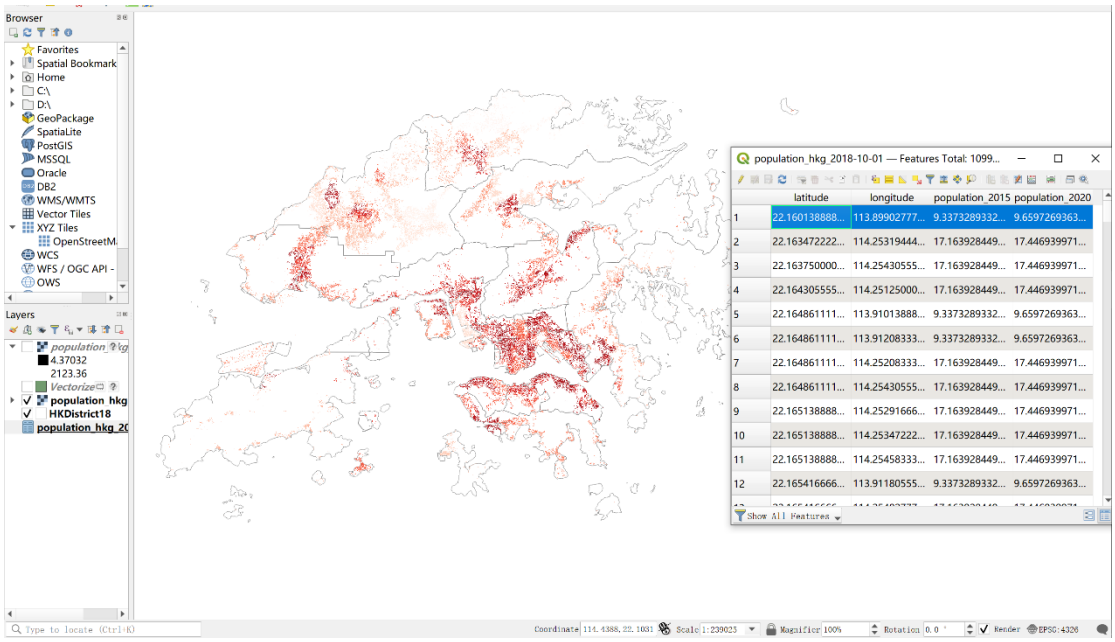


To change the Symbology of tiff layer, click 'layer properties'- 'Symbology', change the render type from Singleband gray to Singleband pseudocolour, change 'interpolation' to 'linear', we could select a specific colour ramp, we could identify the numbers of categories, and click 'Classify'.



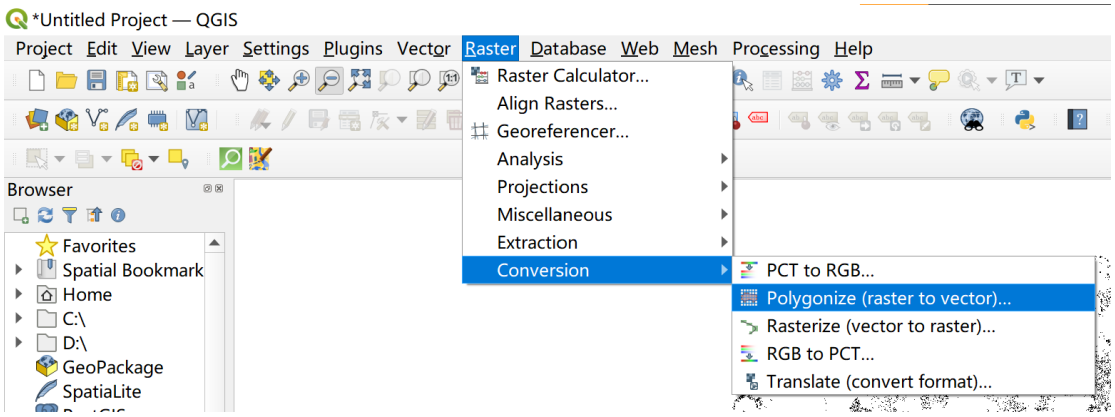


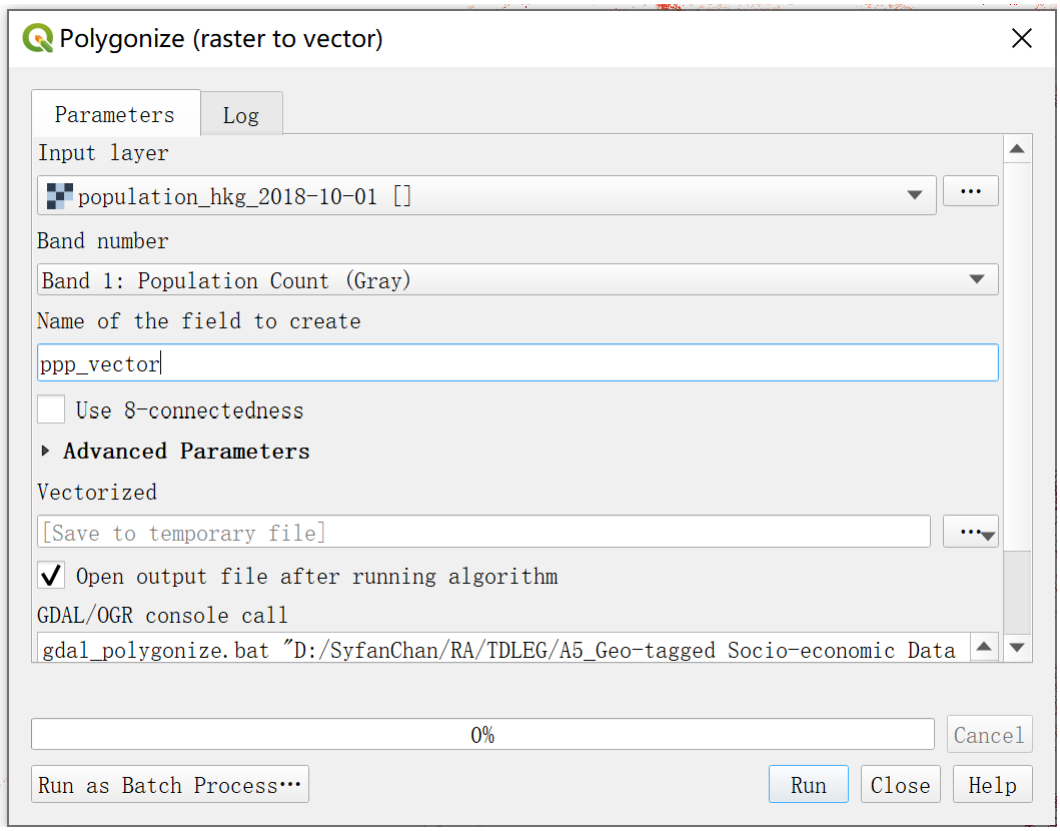
Finally, we could get the density map with attribute table and population metrics.



1.2 To convert geo-tiff (raster) to shapefile (vector data)

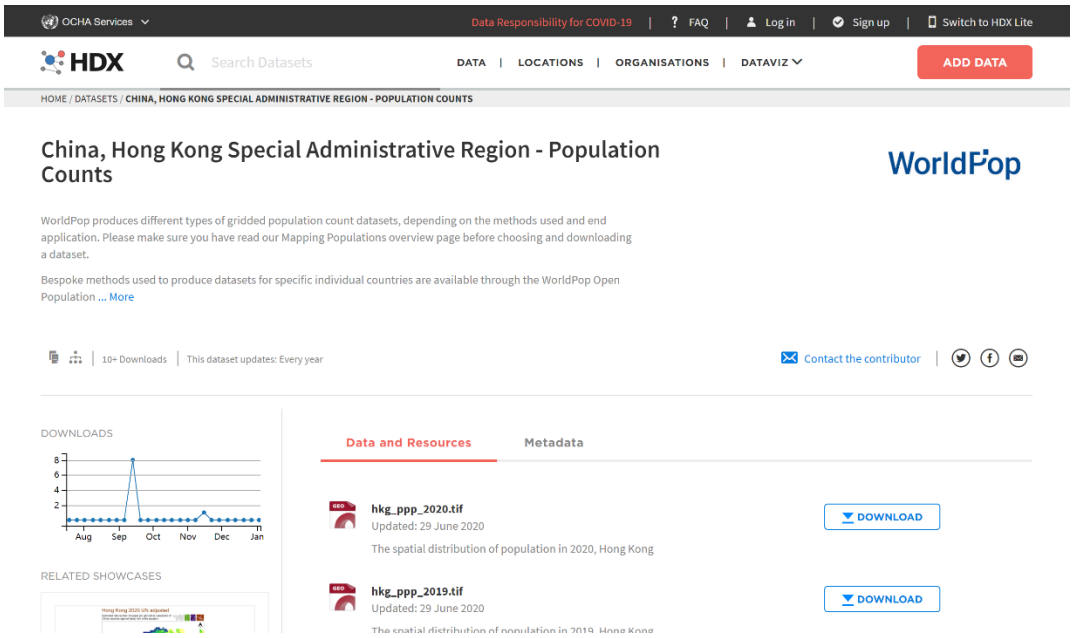
We could also convert the rater data to an editable shapefile, click 'Raster'- 'Conversion'- 'Polygonize'.





1.3 To compare population density in different years

By imputing population density of tiff in different years through followed URL, we could compare the difference and change of population density through the urbanism process.



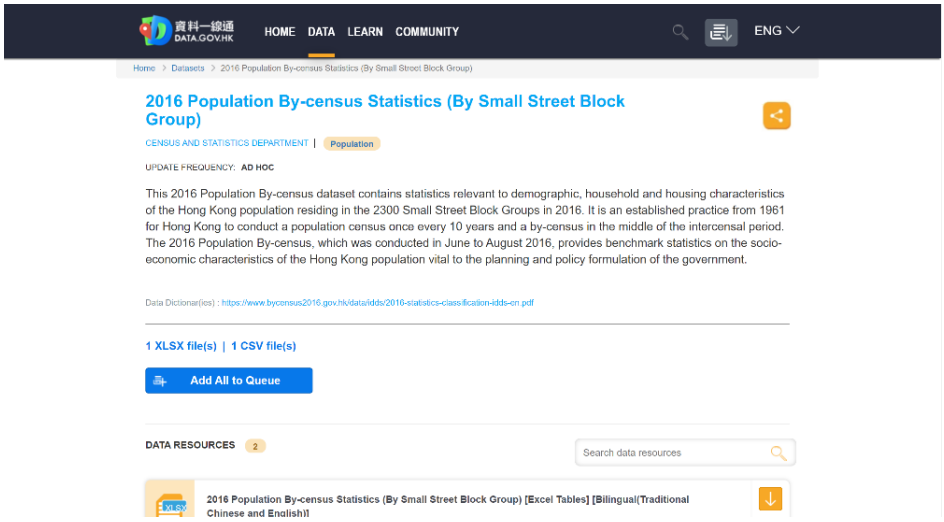
https://data.humdata.org/dataset/worldpop-population-counts-for-china-hong-kong-special-administrative-region#

2. Processing small street-level census data in QGIS

2.1 To obtain census data and the small street-level outline

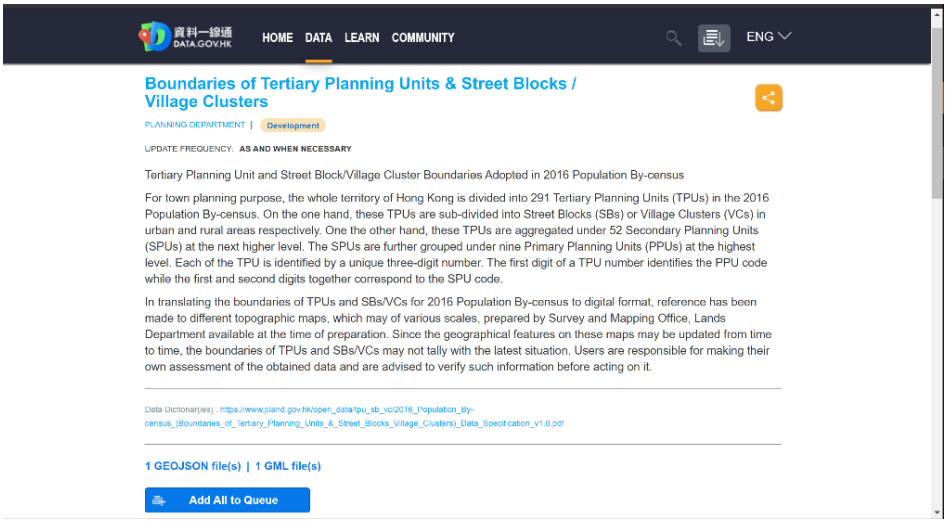
When we need more accurate census data to analyse smaller scale urban areas, we could

download population by-census statistics by small street block group.



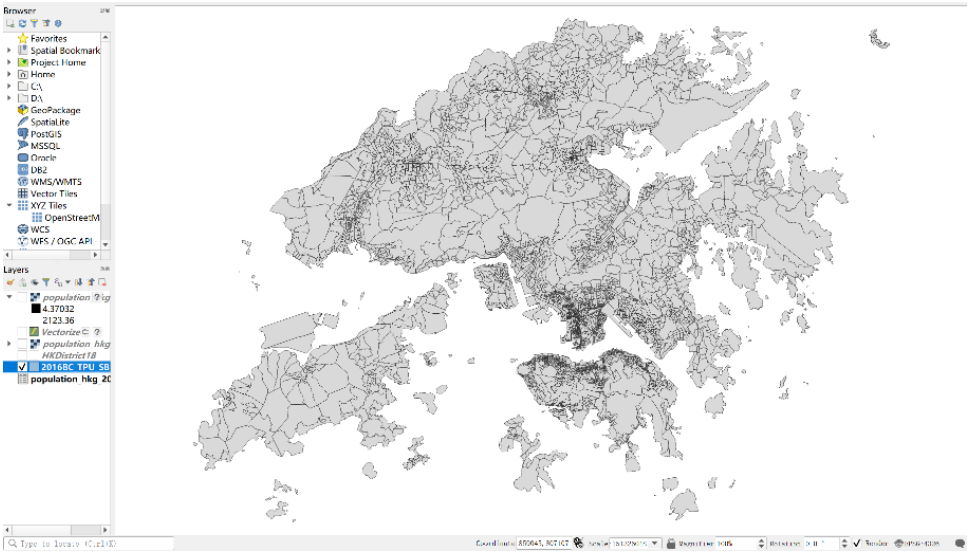
The csv data includes the number of small street block groups, total population density of domestic households and occupied quarters. However, there is no geo reference value in the census data, to link it to urban spatial data, we could download the vector data of small street block group outline. Through the match of the attribute table of two data, we could add census value to the vector layer.

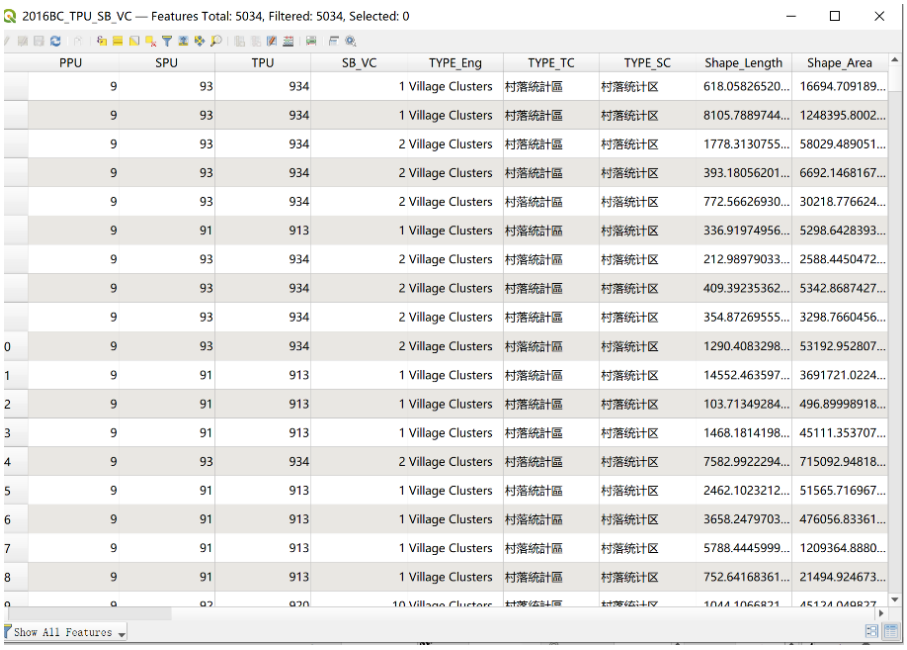
	A	B	C	D	E	F	G
1	街段_小合併組	街段_小合併組（英文名稱）	街段_小合併組（中文名稱）	總人口	家庭住戶數目	有人居住的屋宇單位數目	
2	Small Street Block Group	Name of Small Street Block Group (in English)	Name of Small Street Block Group (in Chinese)	Total population	Domestic households	Occupied quarters	
3	ssbg	ssbg_eng	ssbg_chi	t_pop	dh	oq	
4	11101S	111/01-02 and 111/37-38	111/01-02 及 111/37-38	1766	418	433	
5	11103	111/03	111/03	1539	670	670	
6	11104	111/04	111/04	2269	818	810	
7	11105	111/05	111/05	1210	428	428	
8	11106	111/06	111/06	1032	414	424	
9	11107	111/07	111/07	2189	721	742	
10	11108	111/08	111/08	649	280	280	
11	11109	111/09	111/09	879	383	391	
12	11110	111/10	111/10	1041	399	409	
13	11111	111/11	111/11	2102	868	868	
14	11112	111/12	111/12	2888	1204	1185	
15	11113S	111/13 and 111/15	111/13 及 111/15	1058	402	402	
16	11114S	111/14 and 111/40-42	111/14 及 111/40-42	1637	660	671	
17	11116S	111/16-17	111/16-17	8962	2544	2578	
18	11118S	111/18-20	111/18-20	2910	1069	1069	
19	11121	111/21	111/21	1660	729	740	
20	11122	111/22	111/22	5175	2053	2042	
21	11123	111/23	111/23	909	382	382	
22	11124	111/24	111/24	1265	470	482	
23	11125	111/25	111/25	1830	737	725	
24	11126	111/26	111/26	8674	3205	3520	
25	11127	111/27	111/27	805	302	322	
26	11128	111/28	111/28	858	353	353	
27	11129S	111/29-30	111/29-30	1092	393	393	
28	11131S	111/31-32	111/31-32	713	313	313	
29	11135S	111/35 and 111/43	111/35 及 111/43	1668	595	736	
30	11139	111/39	111/39	2149	678	685	



2.2 To pre-processing census data in csv

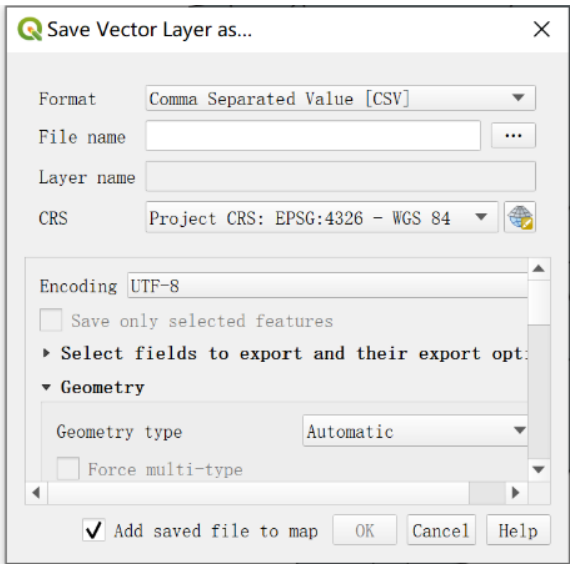
Firstly, to open the small street-level block group outline, drag the GML into QGIS, open the attribute table, the values named ‘TPU’ and ‘SB_VC’ is consistent with ‘Name of Small Street Block Group’ in census .csv.





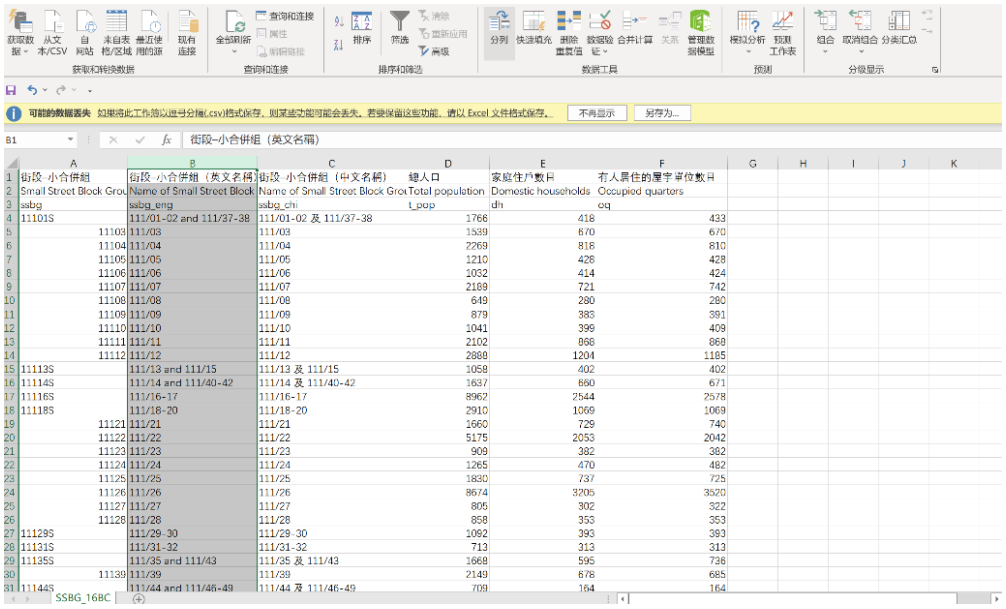
	PPU	SPU	TPU	SB_VC	TYPE_Eng	TYPE_TC	TYPE_SC	Shape_Length	Shape_Area
	9	93	934	1	Village Clusters	村落統計區	村落统计区	618.05826520...	16694.709189...
	9	93	934	1	Village Clusters	村落統計區	村落统计区	8105.7889744...	1248395.8002...
	9	93	934	2	Village Clusters	村落統計區	村落统计区	1778.3130755...	58029.489051...
	9	93	934	2	Village Clusters	村落統計區	村落统计区	393.18056201...	6692.1468167...
	9	93	934	2	Village Clusters	村落統計區	村落统计区	772.56626930...	30218.776624...
	9	91	913	1	Village Clusters	村落統計區	村落统计区	336.91974956...	5298.6428393...
	9	93	934	2	Village Clusters	村落統計區	村落统计区	212.98979033...	2588.4450472...
	9	93	934	2	Village Clusters	村落統計區	村落统计区	409.39235362...	5342.8687427...
	9	93	934	2	Village Clusters	村落統計區	村落统计区	354.87269555...	3298.7660456...
0	9	93	934	2	Village Clusters	村落統計區	村落统计区	1290.4083298...	53192.952807...
1	9	91	913	1	Village Clusters	村落統計區	村落统计区	14552.463597...	3691721.0224...
2	9	91	913	1	Village Clusters	村落統計區	村落统计区	103.71349284...	496.89998918...
3	9	91	913	1	Village Clusters	村落統計區	村落统计区	1468.1814198...	45111.353707...
4	9	93	934	2	Village Clusters	村落統計區	村落统计区	7582.9922294...	715092.94818...
5	9	91	913	1	Village Clusters	村落統計區	村落统计区	2462.1023212...	51565.716967...
6	9	91	913	1	Village Clusters	村落統計區	村落统计区	3658.2479703...	476056.83361...
7	9	91	913	1	Village Clusters	村落統計區	村落统计区	5788.4445999...	1209364.8880...
8	9	91	913	1	Village Clusters	村落統計區	村落统计区	752.64168361...	21494.924673...

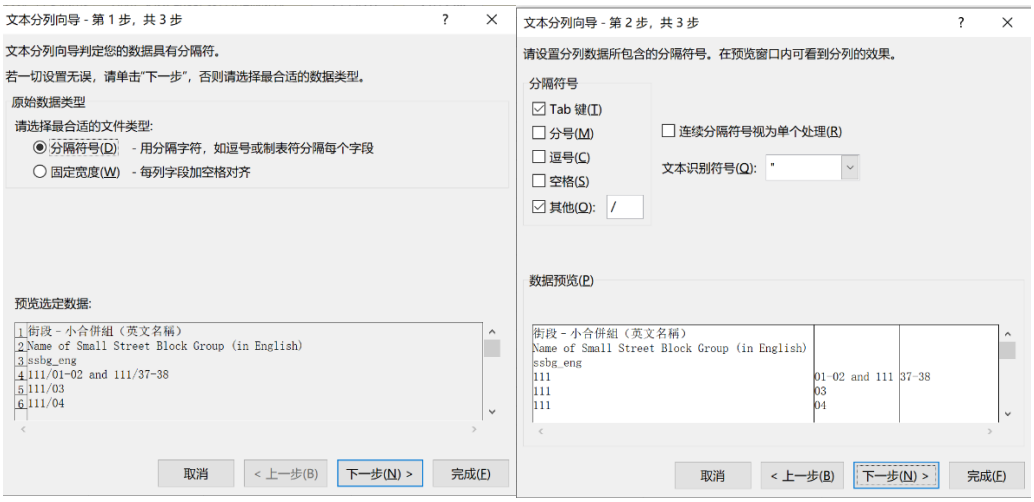
Save vector layers as csv, we could get the files below, each block has a unique value named ‘OBJECTID’, for joining the census data, we need to add the ‘OBJECTID’ to the census csv.



	A	B	C	D	E	F	G	H	I	J
1	OBJECTID	PPU	SPU	TPU	SB_VC	TYPE_Eng	TYPE_TC	TYPE_SC	Shape_Leng	Shape_Area
2	1469	2	22	227	1	Street Bloc	琛楁	缇庡	7326.231463	
3	1475	2	22	227	2	Street Bloc	琛楁	缇庡	6357.760481	
4	1480	2	23	232	4	Street Bloc	琛楁	缇庡	43967.89517	
5	1512	2	22	227	3	Street Bloc	琛楁	缇庡	6660.845249	
6	1517	2	22	227	4	Street Bloc	琛楁	缇庡	6424.370144	
7	1558	2	22	227	5	Street Bloc	琛楁	缇庡	6793.069837	
8	1564	2	23	235	3	Street Bloc	琛楁	缇庡	4262.862473	
9	1570	2	22	222	15	Street Bloc	琛楁	缇庡	11044.89187	
10	1576	2	22	222	11	Street Bloc	琛楁	缇庡	6706.736233	
11	1650	2	23	235	1	Street Bloc	琛楁	缇庡	17811.53334	
12	1655	2	23	235	2	Street Bloc	琛楁	缇庡	11191.20356	
13	1664	2	23	232	1	Street Bloc	琛楁	缇庡	87558.7058	
14	1703	2	22	222	30	Street Bloc	琛楁	缇庡	5990.609449	
15	1777	2	23	235	4	Street Bloc	琛楁	缇庡	24330.17528	
16	1819	2	22	222	26	Street Bloc	琛楁	缇庡	3840.17097	
17	1824	2	23	232	10	Street Bloc	琛楁	缇庡	18027.18127	
18	1946	2	22	222	20	Street Bloc	琛楁	缇庡	7639.480697	
19	1984	2	22	222	21	Street Bloc	琛楁	缇庡	7982.482446	
20	1988	2	22	222	22	Street Bloc	琛楁	缇庡	7788.985414	

Before adding the new value, we need to pre-processing the original census csv data, select followed area, click ‘Statistics’- ‘Branch’, split corresponding TPU and SB_VC into two columns, after that, we could join ‘‘OBJECTID’ value by searching same values. Taking Mong Kok as an example, the processed data shows as follow.





	A	B	C	D	E
1	OBJECTID	TPU	SB_VC	T_pop	
2	1576	222	11	843	
3	1430	222	12	1663	
4	1570	222	15	2062	
5	1620	222	17	1921	
6	1942	222	19	510	
7	1946	222	20	1179	
8	1984	222	21	1936	
9	1988	222	22	2072	
10	1774	222	24	708	
11	1700	222	29	1370	
12	1819	222	26	696	
13	1703	222	30	1006	
14	1469	227	1	590	
15	1475	227	2	595	
16	1512	227	3	1387	
17	1517	227	4	438	
18	1558	227	5	787	
19	1563	227	6	871	
20	2059	227	8	644	

2.3 To join census data to geo-tagged shapefile (join attribute table)

We could import this csv file without any further action and it would be imported. But, the default type of each column would be a String (text).

To tell QGIS to import the field as a number, we need to create a sidecar file with a .csvt extension. This file will have only 1-row specifying data types for each column. Save this file as ca_tracts_pop.csvt in the same directory as the original .csv file.

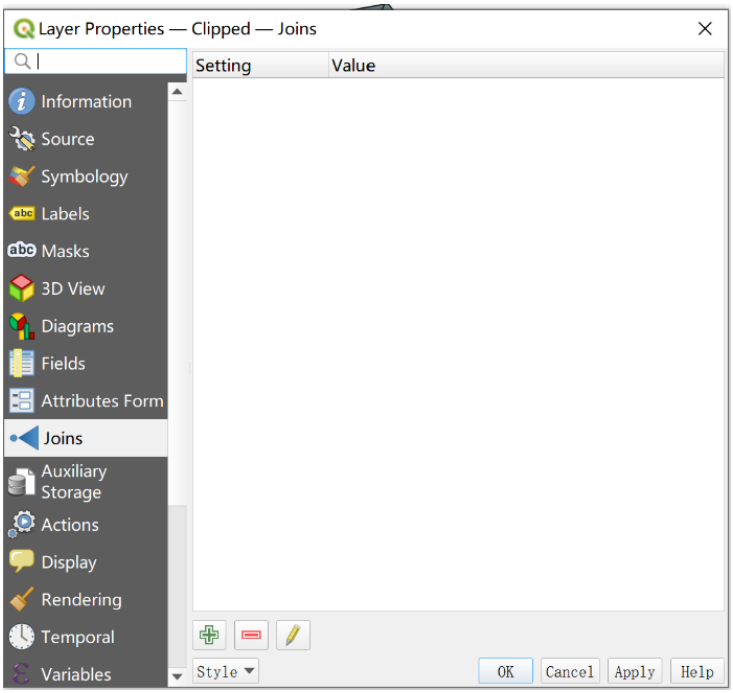
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

"Integer","Real","Integer","Integer"

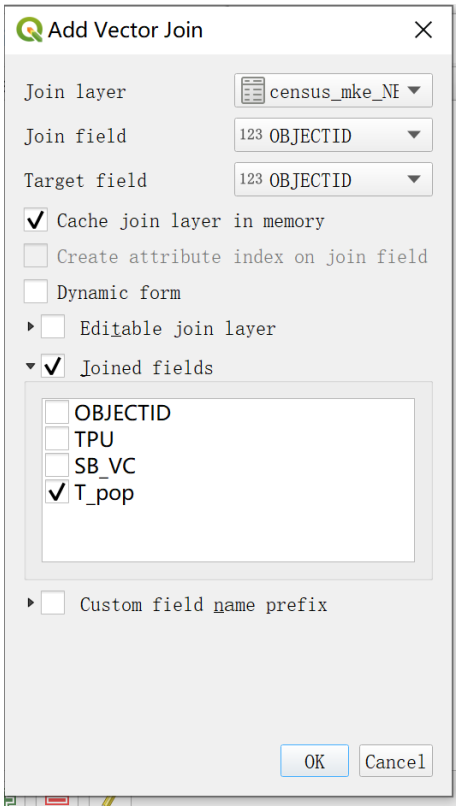
Now we are ready to import the CSV file to QGIS. Go to Layer ‣ Add Delimited Text Layer. Browse to the folder containing the CSV file and select it. Make sure you have selected File format as CSV (comma-separated values). Since we are importing this as a table, we must specify that our file contains no geometry. Select the No geometry (attribute only

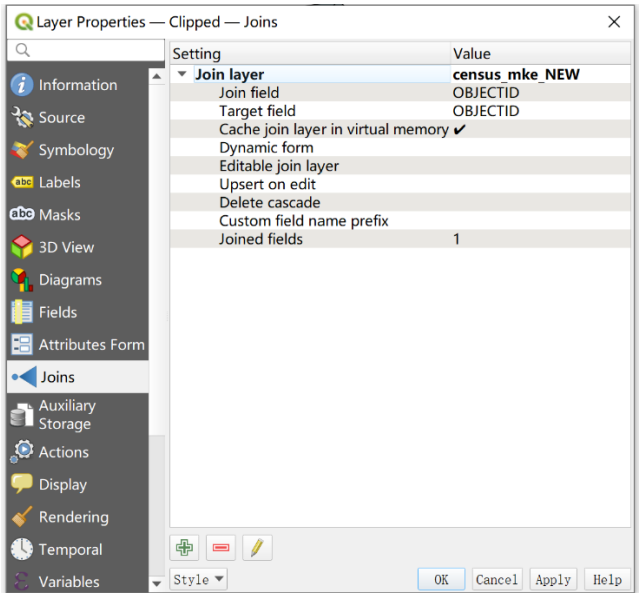
table) option. Click OK.

In the Layer Properties dialog, select the Joins tab. Click on the + button at the bottom to create a new table join.

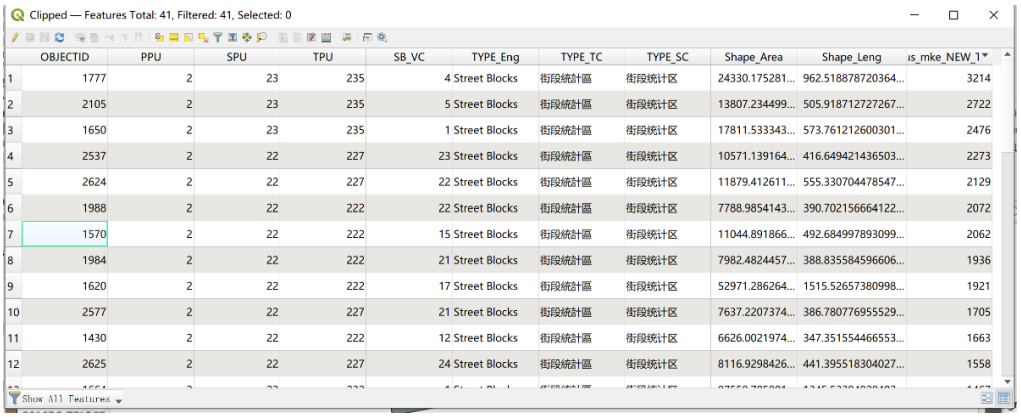


In the Add vector join dialog, select census_mke_NEW as the Join layer. Next, we have to select the field with unique ids in both the shapefile and the CSV. Select T_pop as the Join field and Target field respectively. Click OK.



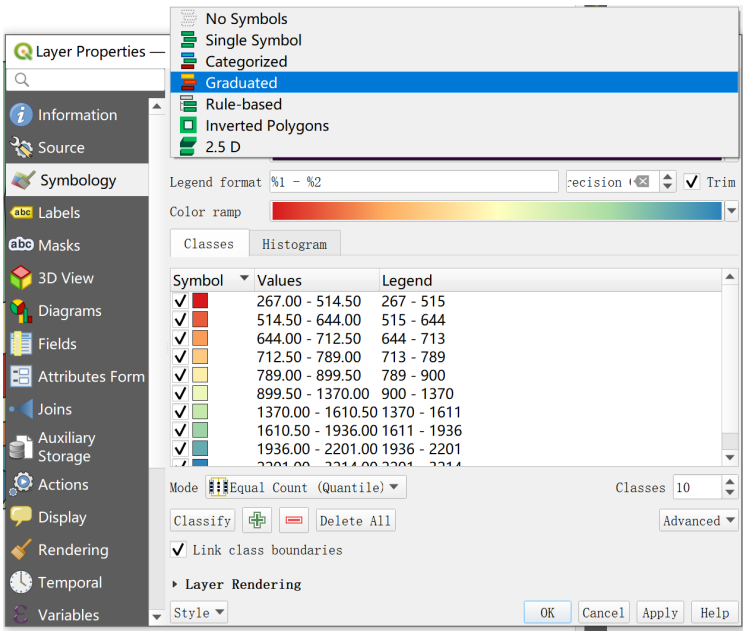


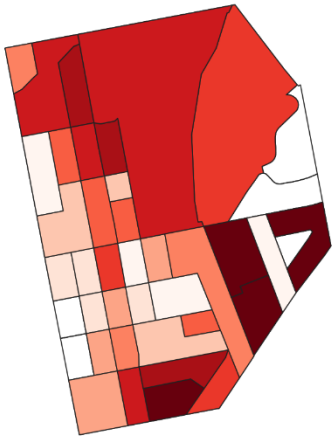
Open the Attribute table, we could find the census value is already added.

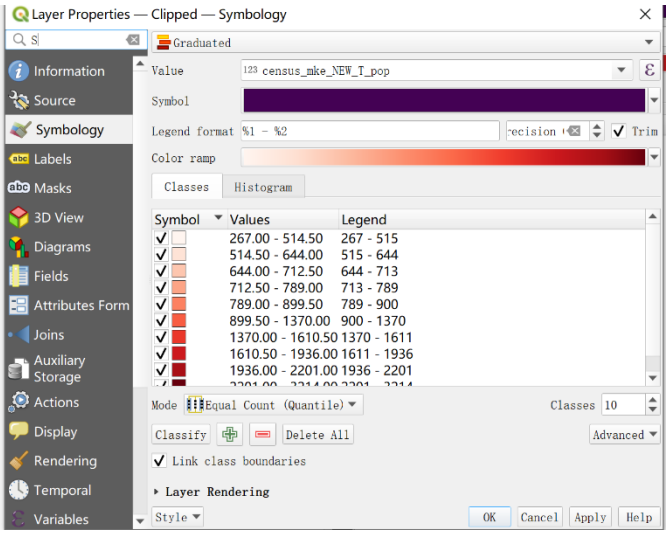


2.4 To edit the symbology

To visualize the population, open the layer properties, change the mode to ‘Graduated’, select the ‘colour ramp’, identify classes number, click ‘Classify’, and click OK.






Layer Properties — Clipped — Symbology
Graduated
Information
Source
Symbology
Labels
Masks
3D View
Diagrams
Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Temporal
Variables
Value
123 census_mke_NEW_T_pop
Symbol
Legend format
%1 – %2
Trim
Color ramp
Classes
Histogram
Symbol
Values
Legend
267.00 - 514.50 267 - 515
514.50 - 644.00 515 - 644
644.00 - 712.50 644 - 713
712.50 - 789.00 713 - 789
789.00 - 899.50 789 - 900
899.50 - 1370.00 900 - 1370
1370.00 - 1610.50 1370 - 1611
1610.50 - 1936.00 1611 - 1936
1936.00 - 2201.00 1936 - 2201
Mode
Equal Count (Quantile)
Classes
10
Classify
Delete All
Advanced
Link class boundaries
Layer Rendering
Style
OK
Cancel
Apply
Help