ArcGIS Pro

Module 4 - Data Analysis

Marcel Fortin, University of Toronto Map and Data Library, mdl@library.utoronto.ca

Last Updated: September 29, 2022

Marcel Fortin
University of Toronto
Map and Data Library
mdl@library.utoronto.ca

Goals

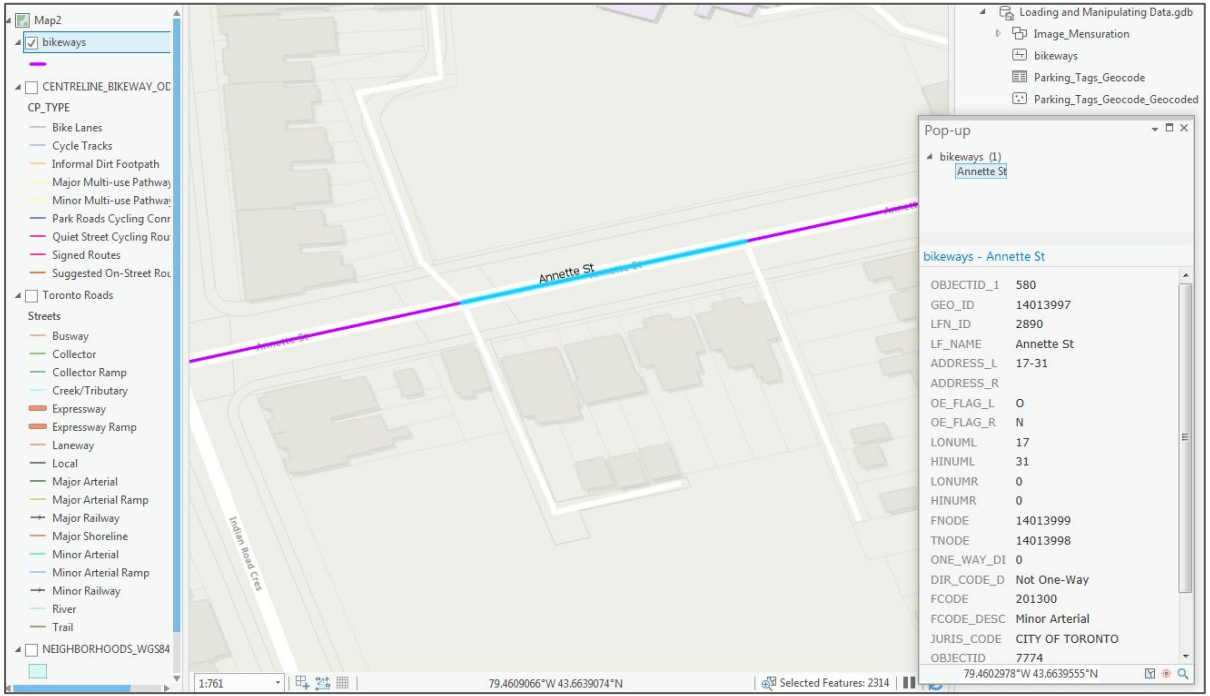
- Learn some basic geospatial analysis through the use of various tools
- Solve a basic spatial problem using GIS

Calculating Distances

1 - Load your bikeways dataset

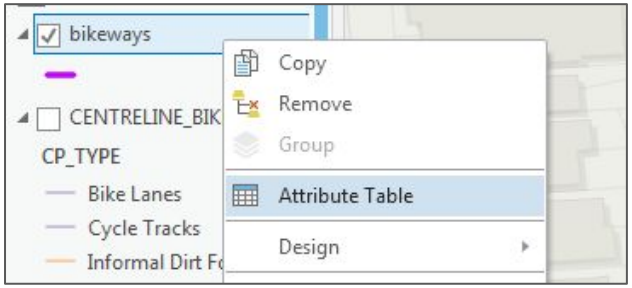
2 - click on various segments of lines from the “bikeways” dataset

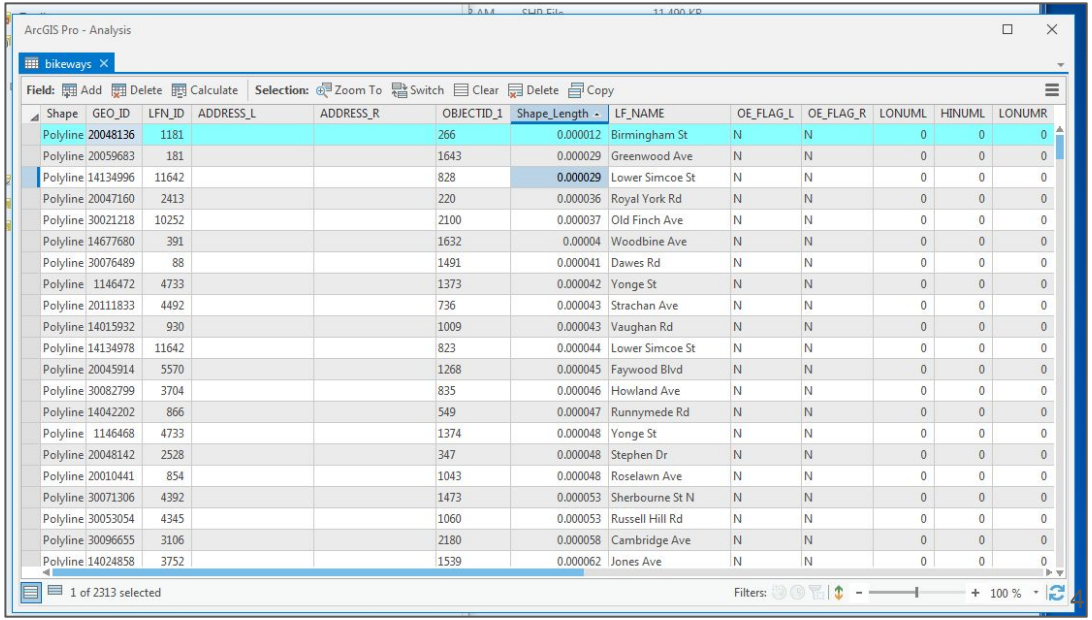
3 - notice the various attributes



Calculating Distances

1 - right-click on the “bikeways” layer in the “Contents” pane and select “Attribute Table”





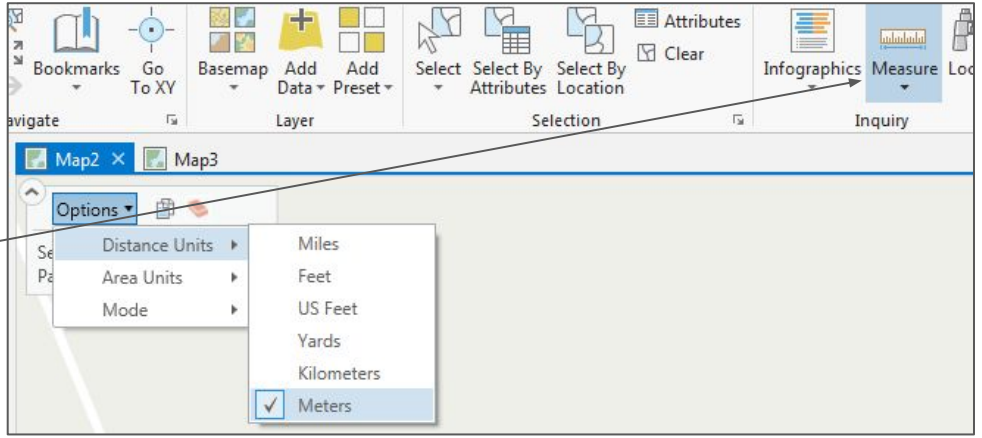
Shape	GEO_ID	LFN_ID	ADDRESS_L	ADDRESS_R	OBJECTID_1	Shape_Length	LF_NAME	OE_FLAG_L	OE_FLAG_R	LONUML	HINUML	LONUMR
Polyline	20048136	1181			266	0.000012	Birmingham St	N	N	0	0	0
Polyline	20059683	181			1643	0.000029	Greenwood Ave	N	N	0	0	0
Polyline	14134996	11642			828	0.000029	Lower Simcoe St	N	N	0	0	0
Polyline	20047160	2413			220	0.000036	Royal York Rd	N	N	0	0	0
Polyline	30021218	10252			2100	0.000037	Old Finch Ave	N	N	0	0	0
Polyline	14677680	391			1632	0.00004	Woodbine Ave	N	N	0	0	0
Polyline	30076489	88			1491	0.000041	Dawes Rd	N	N	0	0	0
Polyline	1146472	4733			1373	0.000042	Yonge St	N	N	0	0	0
Polyline	20111833	4492			736	0.000043	Strachan Ave	N	N	0	0	0
Polyline	14015932	930			1009	0.000043	Vaughan Rd	N	N	0	0	0
Polyline	14134978	11642			823	0.000044	Lower Simcoe St	N	N	0	0	0
Polyline	20045914	5570			1268	0.000045	Faywood Blvd	N	N	0	0	0
Polyline	30082799	3704			835	0.000046	Howland Ave	N	N	0	0	0
Polyline	14042202	866			549	0.000047	Runnymede Rd	N	N	0	0	0
Polyline	1146468	4733			1374	0.000048	Yonge St	N	N	0	0	0
Polyline	20048142	2528			347	0.000048	Stephen Dr	N	N	0	0	0
Polyline	20010441	854			1043	0.000048	Roselawn Ave	N	N	0	0	0
Polyline	30071306	4392			1473	0.000053	Sherbourne St N	N	N	0	0	0
Polyline	30053054	4345			1060	0.000053	Russell Hill Rd	N	N	0	0	0
Polyline	30096655	3106			2180	0.000058	Cambridge Ave	N	N	0	0	0
Polyline	14024858	3752			1539	0.000062	Jones Ave	N	N	0	0	0

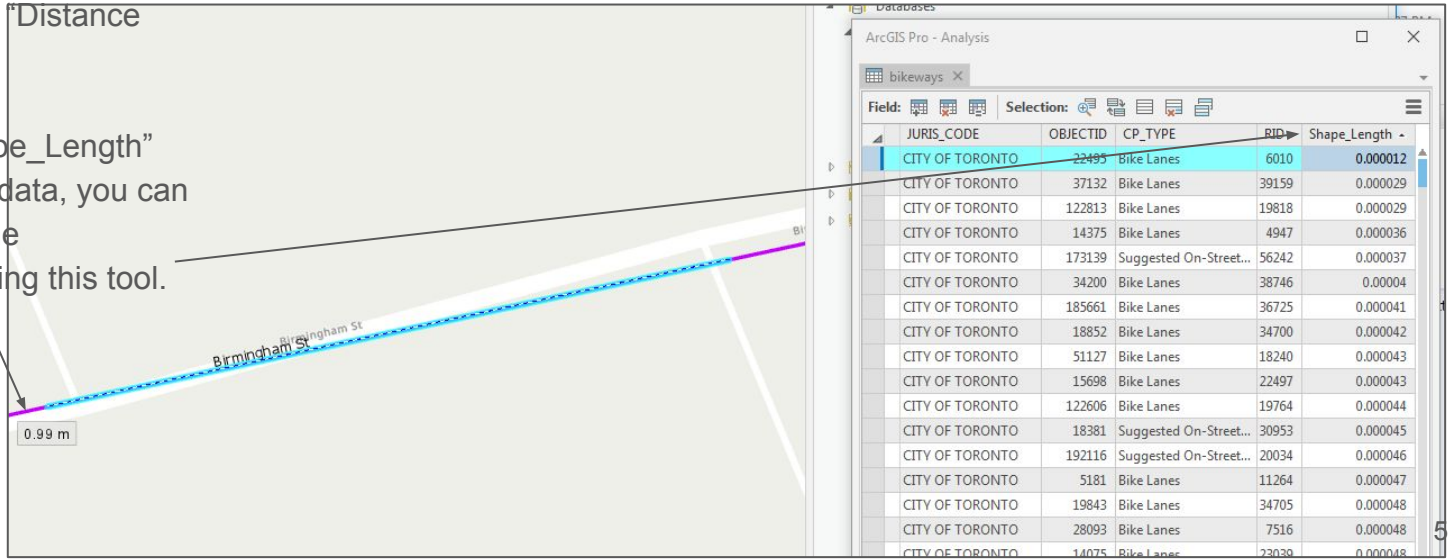
Calculating Distances

1 - Under the “Map” menu, select the “Measure” tool

2 - under the measure “Options” drop-down menu, select “Distance Units” → “Meters”

3 - while there is a “Shape_Length” column in your attribute data, you can see that it is not the same measurement we get using this tool.



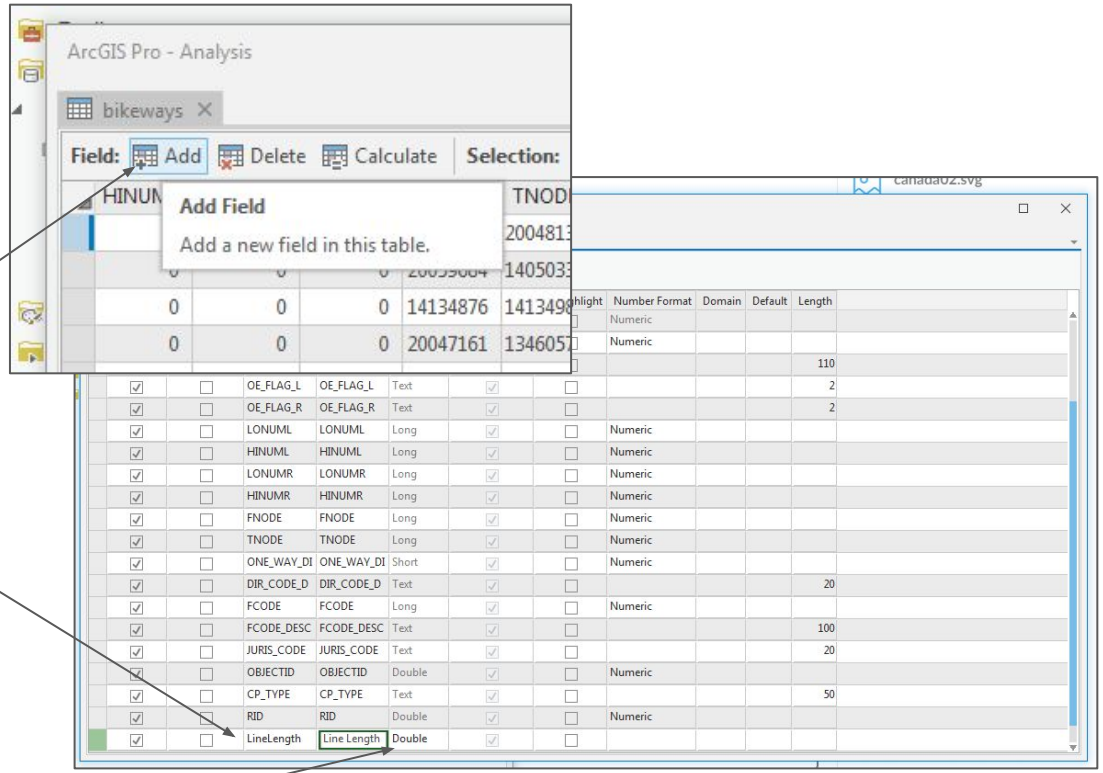


Calculating Distance

1 - with the “bikeways” attribute table open, click on the “Add” button to create a new field

3 - name the field “LineLength” with the alias “Line Length”

4 - select “Double” as the “Data Type”



Calculating Distances

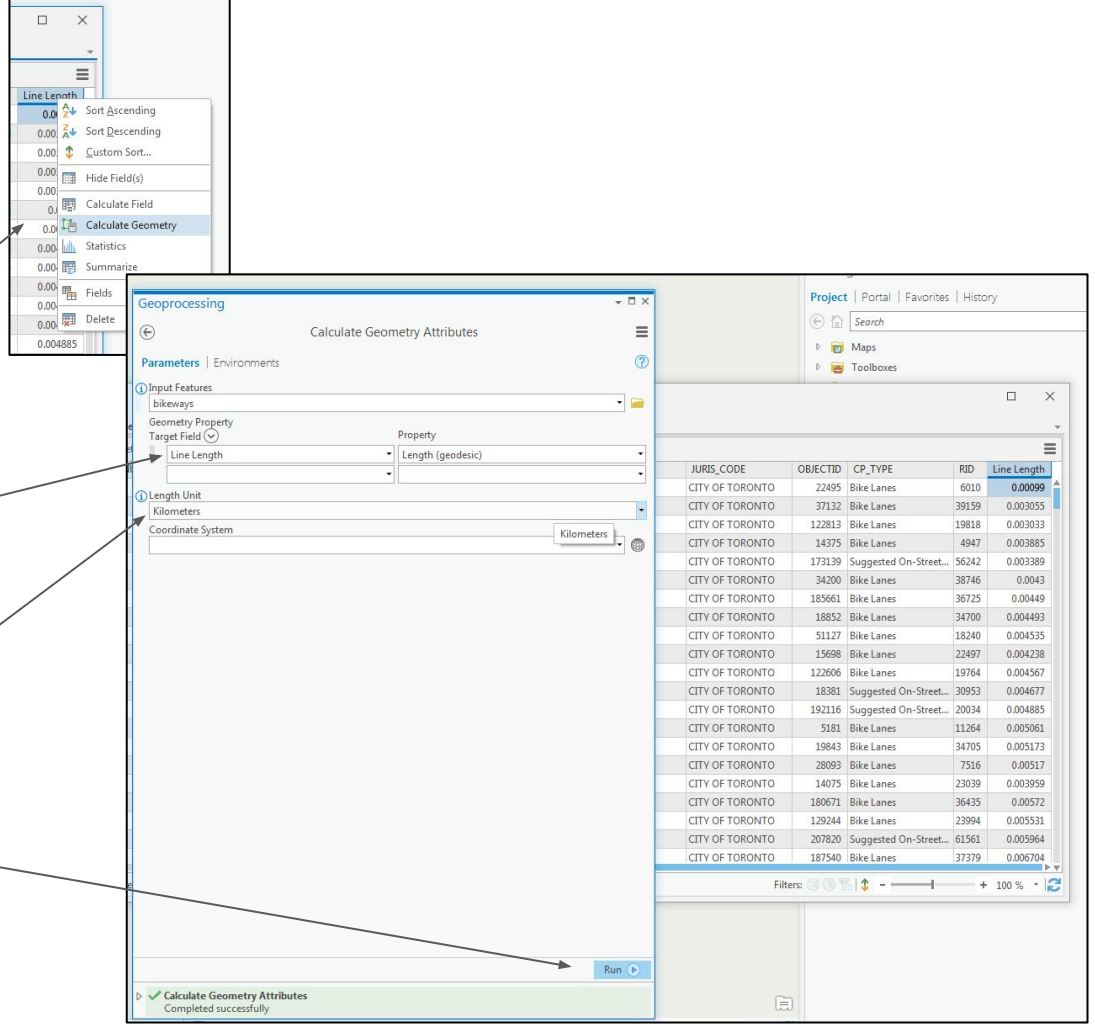
1 - right-click on the “Line Length” column and select “Calculate Geometry”

2 - in the “Geoprocessing” popup select “Length (geodesic)” in the “Property” column

3 - select “Kilometers” as the “Length Unit”

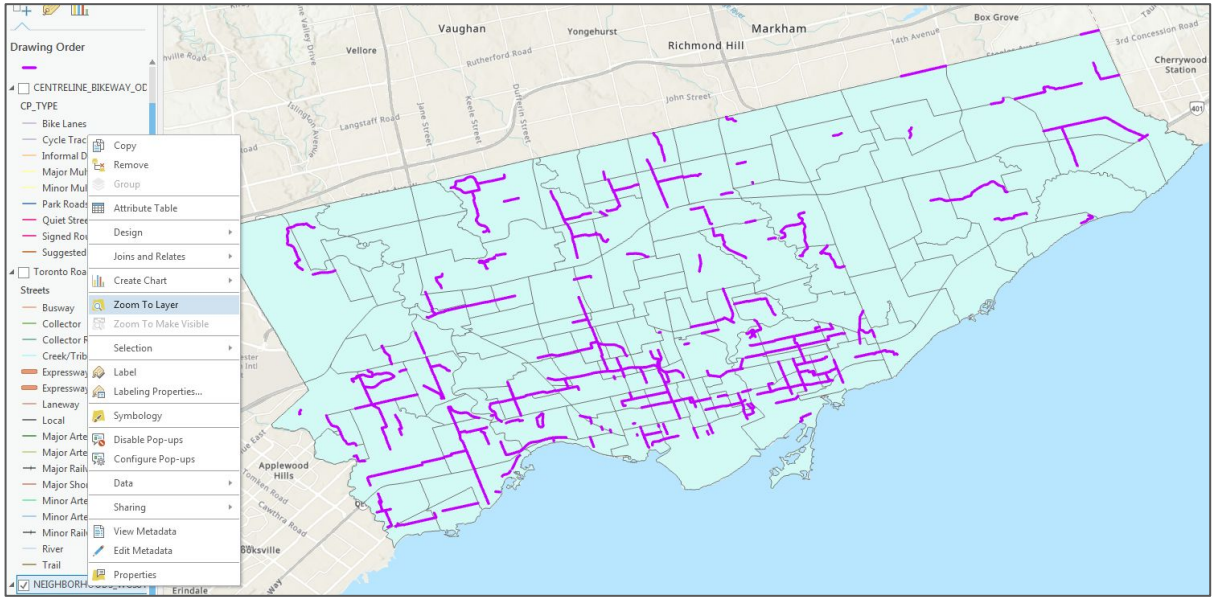
4 - click on “Run”

The column should now be populated with the line length of every segment of the bikeways layer.



Spatial Joins

1 - right-click on the “Neighborhoods” layer and select “Zoom to Layer”



Spatial Joins

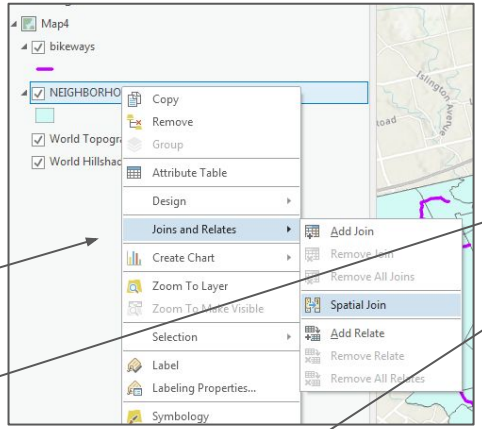
1 - right-click on the “Neighbourhoods” layer and select “Joins and Relates” → “Spatial Join”

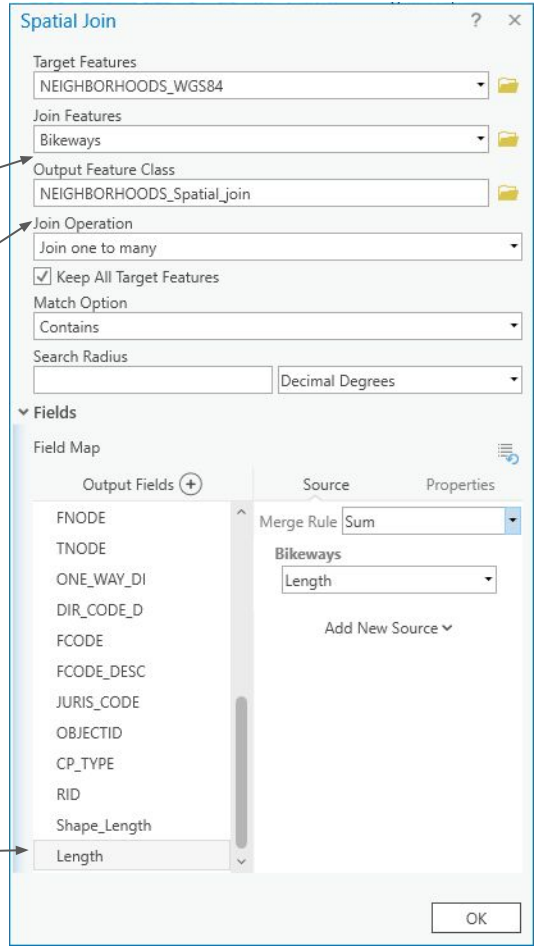
2 - select “bikeways” in the “Join Features” option

3 - give the name “NEIGHBORHOODS_spatial_join” to the resulting joined layer

4 - select “Join one to many” and “Contains” as the “Match option”

5 - highlight the “LineLength” field and select “Sum” as the “Merge Rule”





Spatial Joins

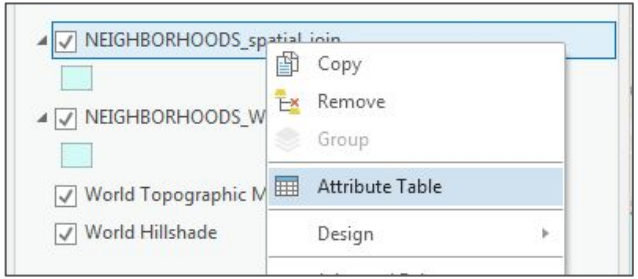
1 - right-click on the resulting “NEIGHBORHOODS_spatial_Join” layer in the “Contents” pane and select “Attribute Table”

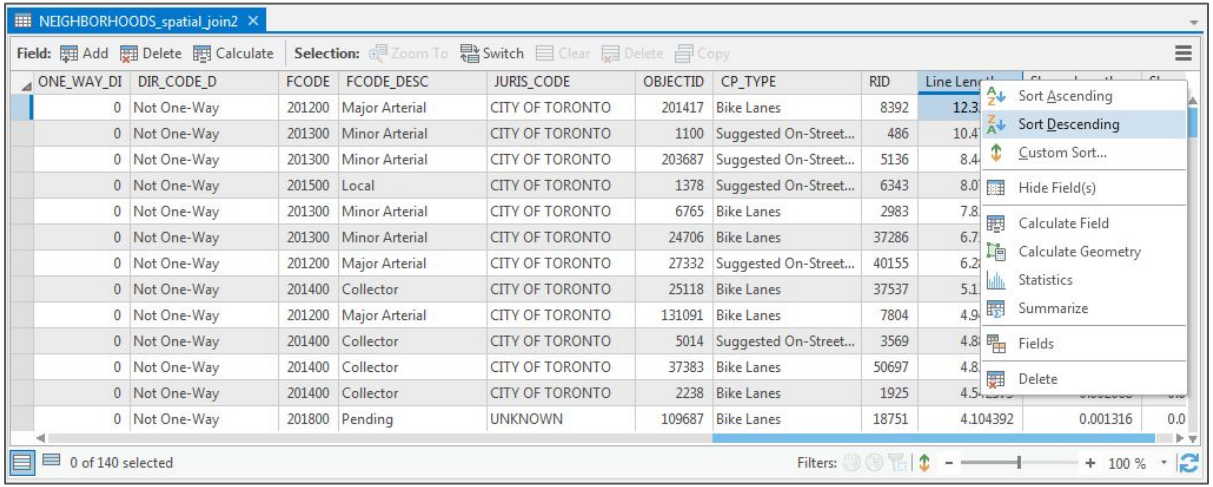
2 - examine the “Join_Count” column (this is the number of bikeway segments are in each neighbourhood”

3 - slide along the table until you find the “Line Length” field

4 - right-click on the column and select “Sort Descending”

Which Toronto Neighbourhood has the longest total KMs of bikeways?





Spatial Joins method II

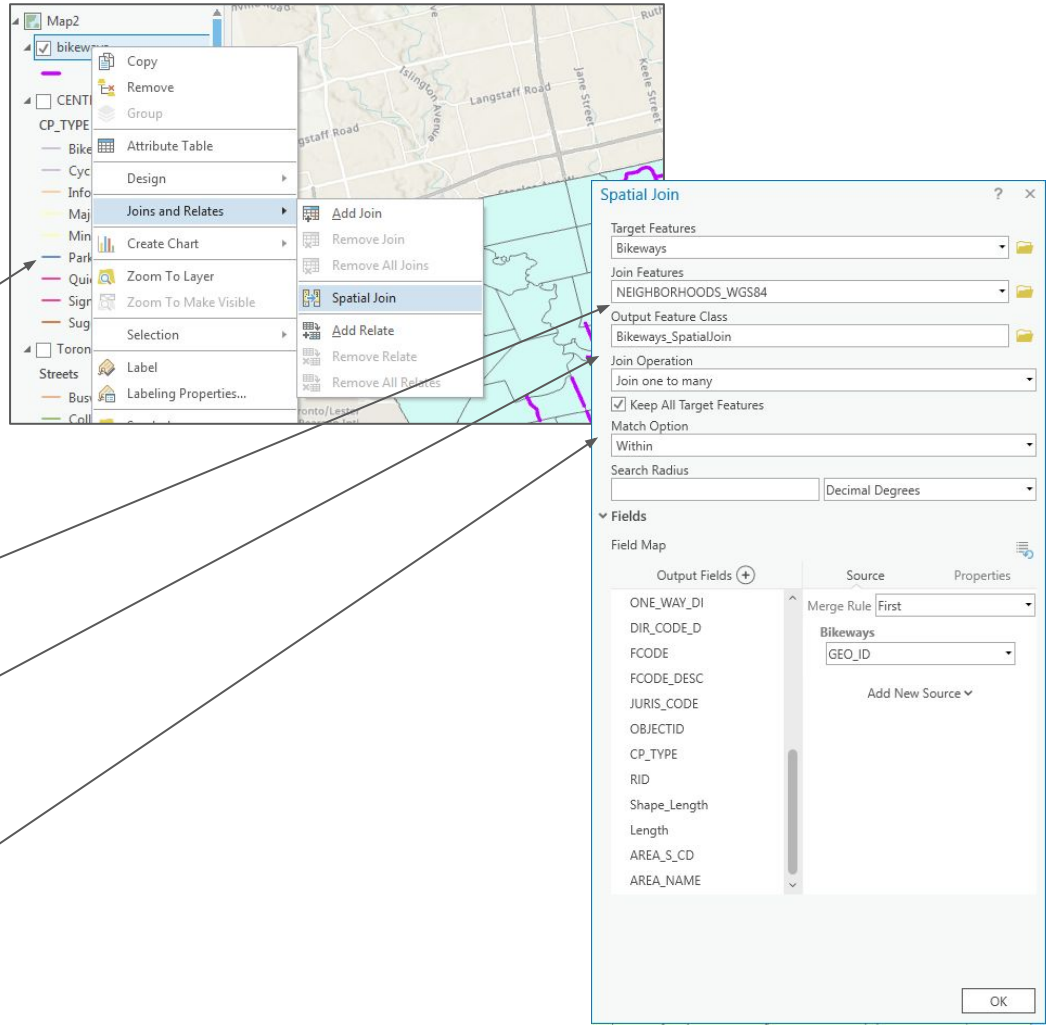
1 - right-click on the “bikeways” layer and select “Joins and Relates” → “Spatial Join”

2 - in the “Geoprocessing” popup, select “NEIGHBORHOODS_WGS84” in the “Join Features” option

3 - give the name “bikeways_SpatiaJoin” to the resulting joined layer

4 - Join Operation to “Join one to many” and Match option to “Within”

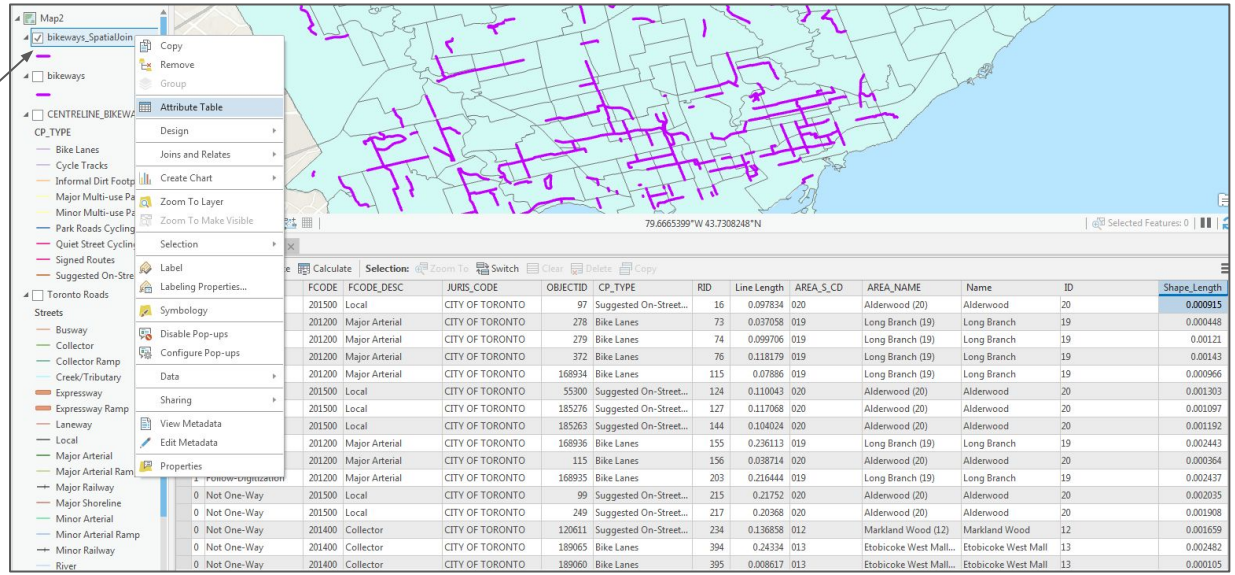
5 - click “Ok”



Spatial Joins method II

1 - right-click on the new joined layer “bikeways_SpatialJoin” and select “Attribute Table”

Examine the attribute data. You will notice that the city wards data are now attached to each bike lane



Summarizing statistics

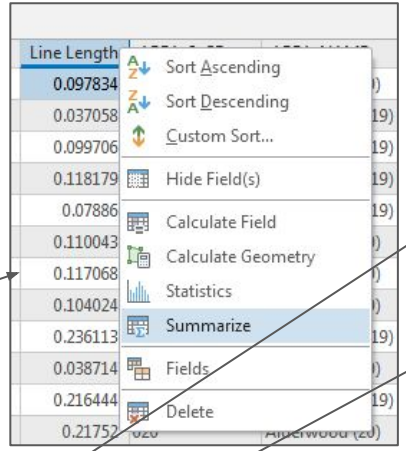
1 - in the new “bikeways_SpatialJoin” attribute table, right-click on the “Line Length” column and select “Summarize”

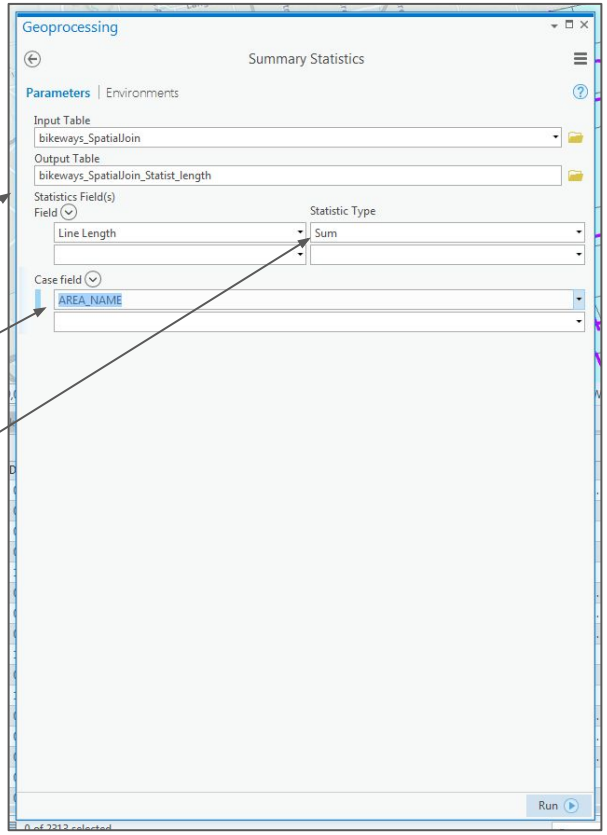
2 - in the Geoprocessing popup, name your summary table “bikeways_SpatialJoin_Statist_length”

3 - in the “Case field” option, select “AREA_NAME”

4 - in the “Statistic Type”, select “Sum”

5 - note the name of the “Output Table”

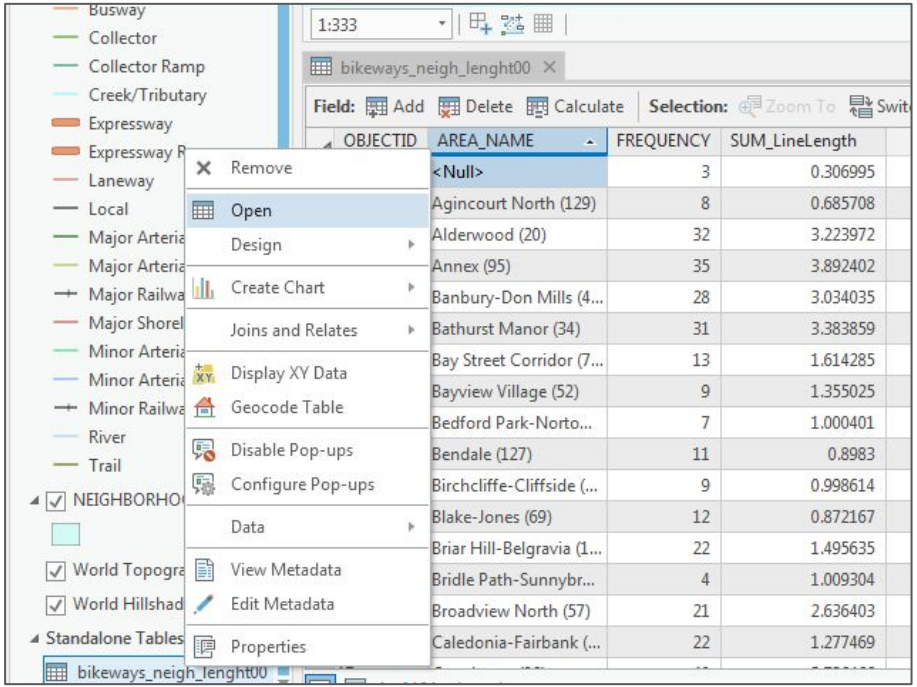




Summarizing statistics

1 - right-click on the resulting table in the “Contents” pane and select “Open”

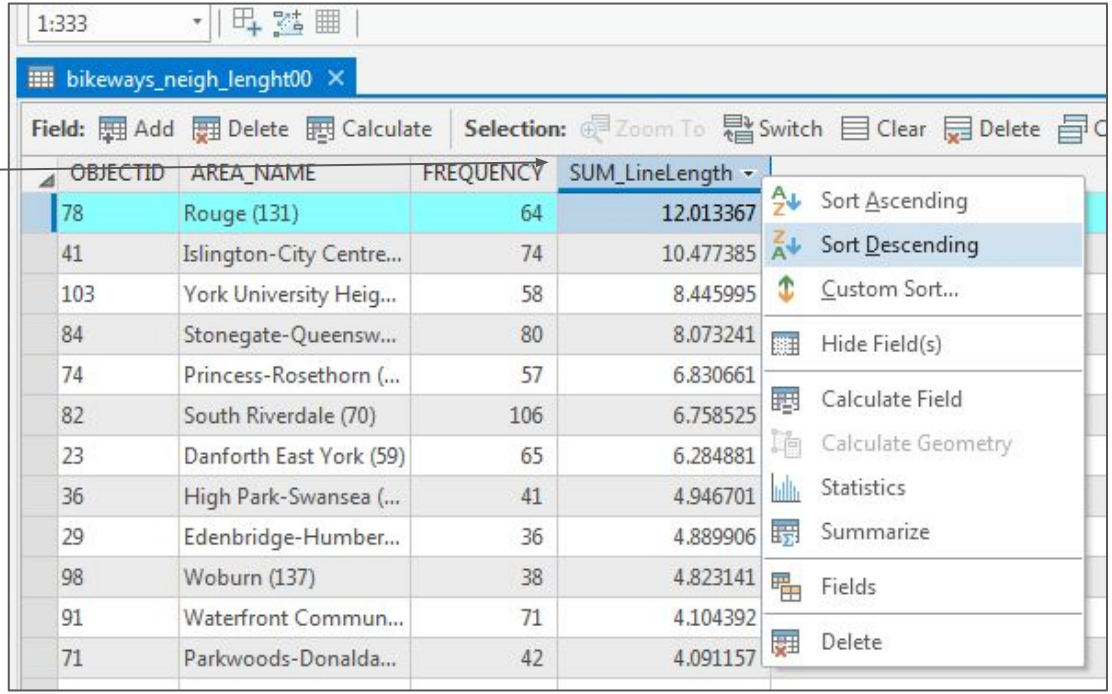
This table has one entry for each Toronto Neighbourhood. Each entry has the number of “bikeways” segments (FREQUENCY) within the Neighbourhood and a “SUM_LineLength”, which is the total amount of Kilometres of “bikeways” are in that neighbourhood.



Summarizing statistics

1 - right-click on the “SUM_LineLength” column and select “Sort Descending”

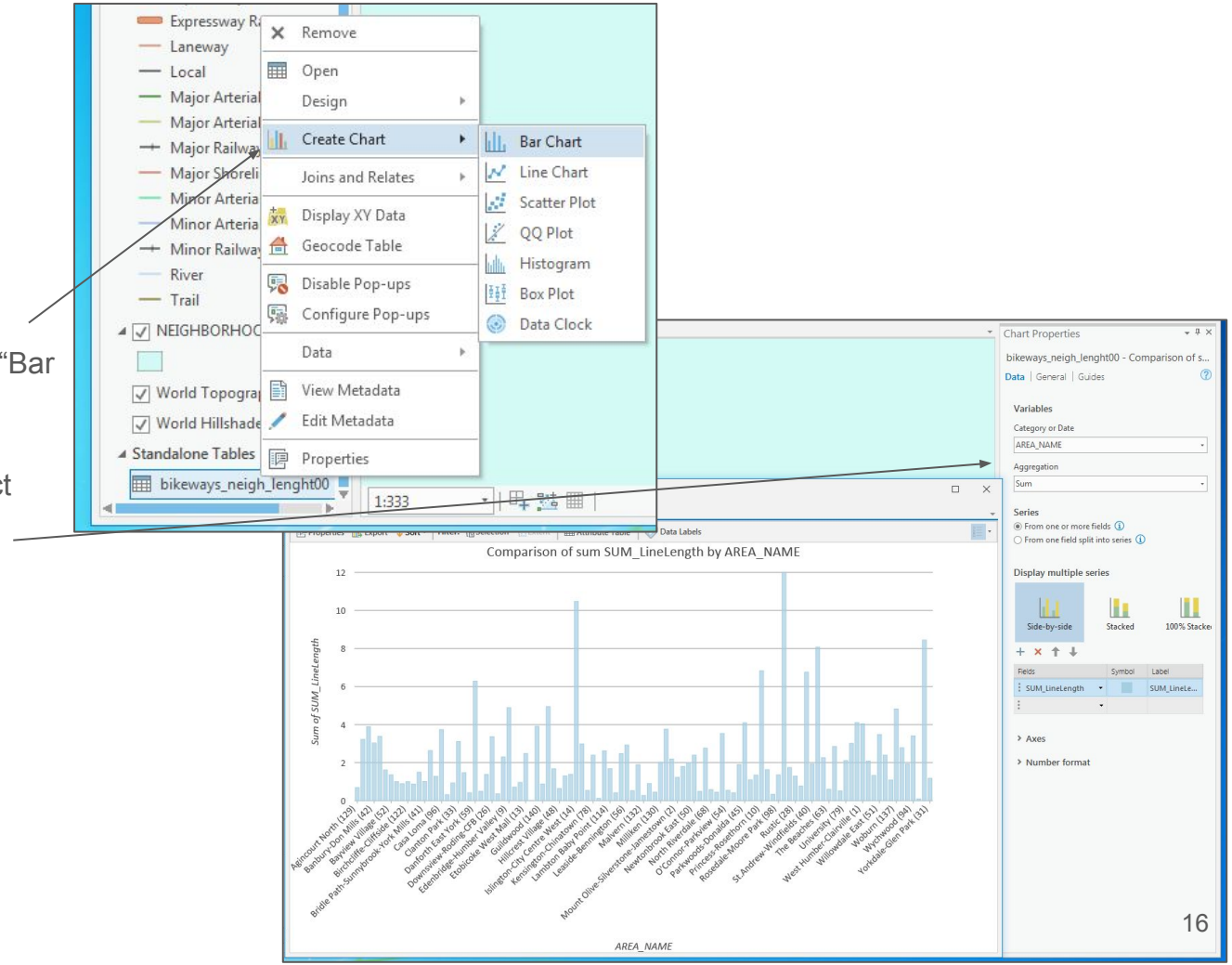
You can see that the “Rouge (131)” neighbourhood has the most kilometres of bikeways in Toronto



Charts

1 - right-click on the new summary table once again in the “Contents” pane and select “Create Chart” → “Bar Chart”

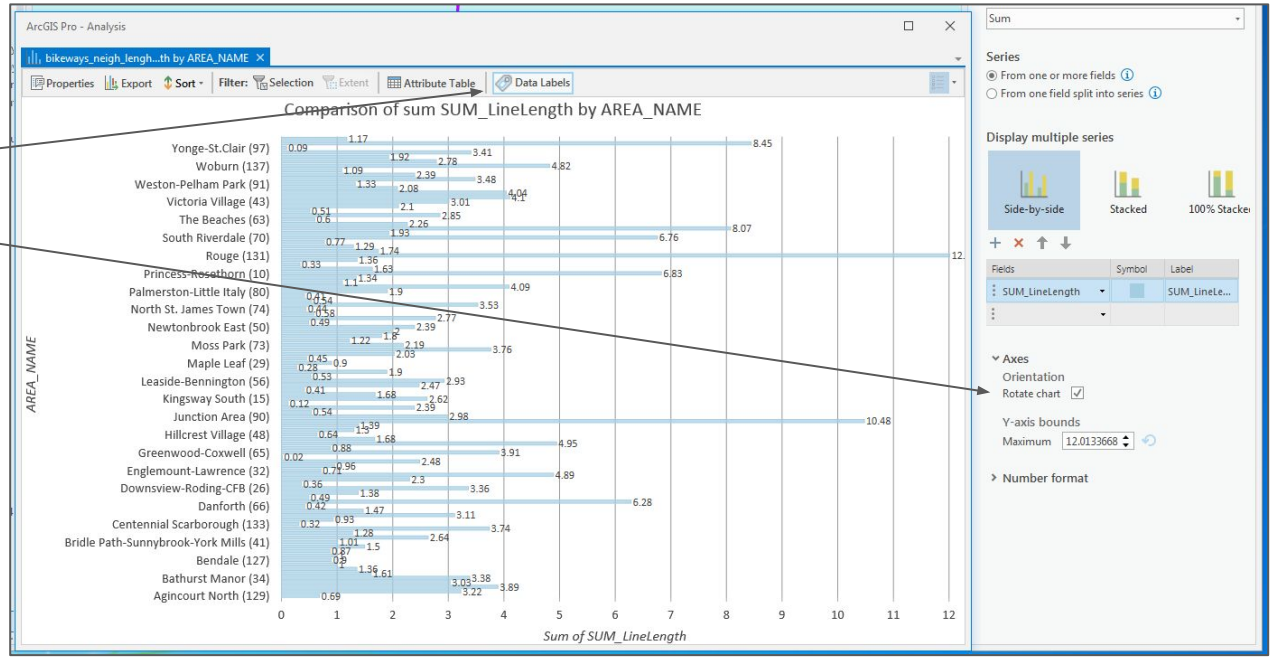
2 - under “Category or Date”, select “AREA_NAME” and under “Aggregation”, select “Sum”



Charts

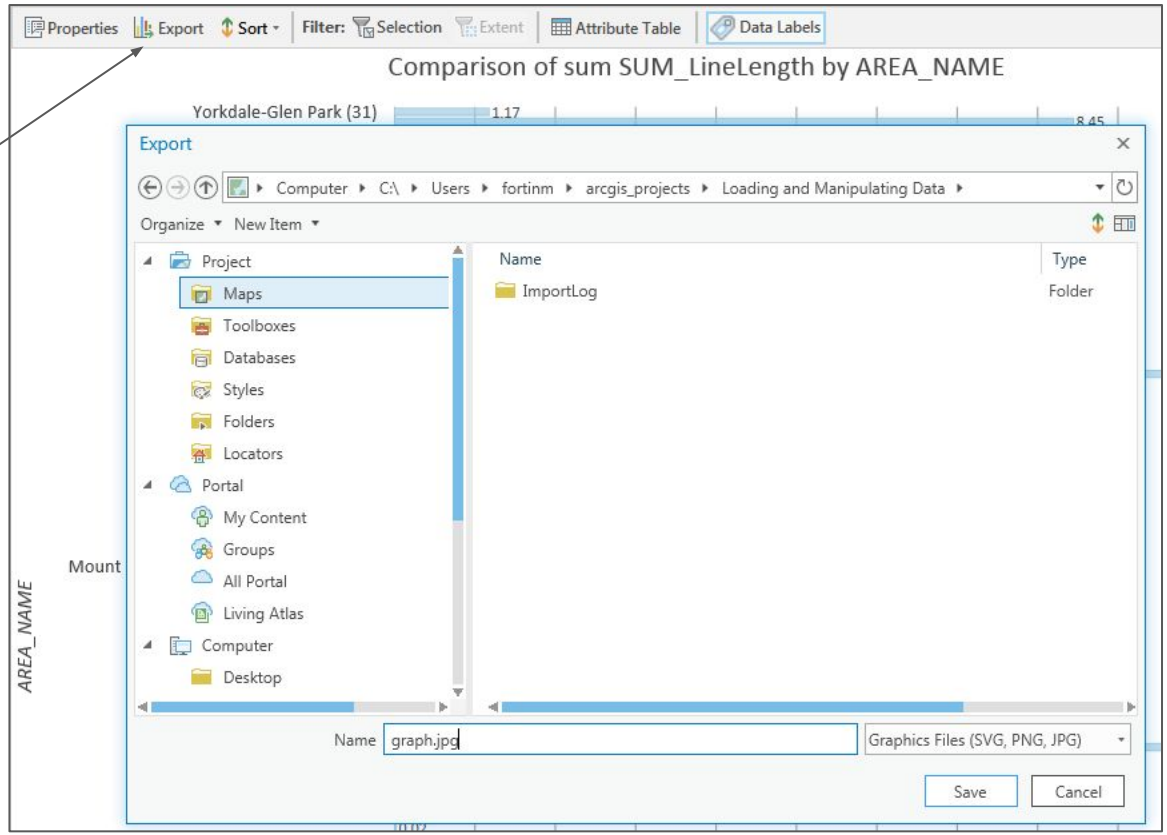
1 - click on “Data Labels”

2 - in the newly created chart, click on the “Axes” option and select “Rotate chart”



Charts

1 - click on “Export” and save your chart



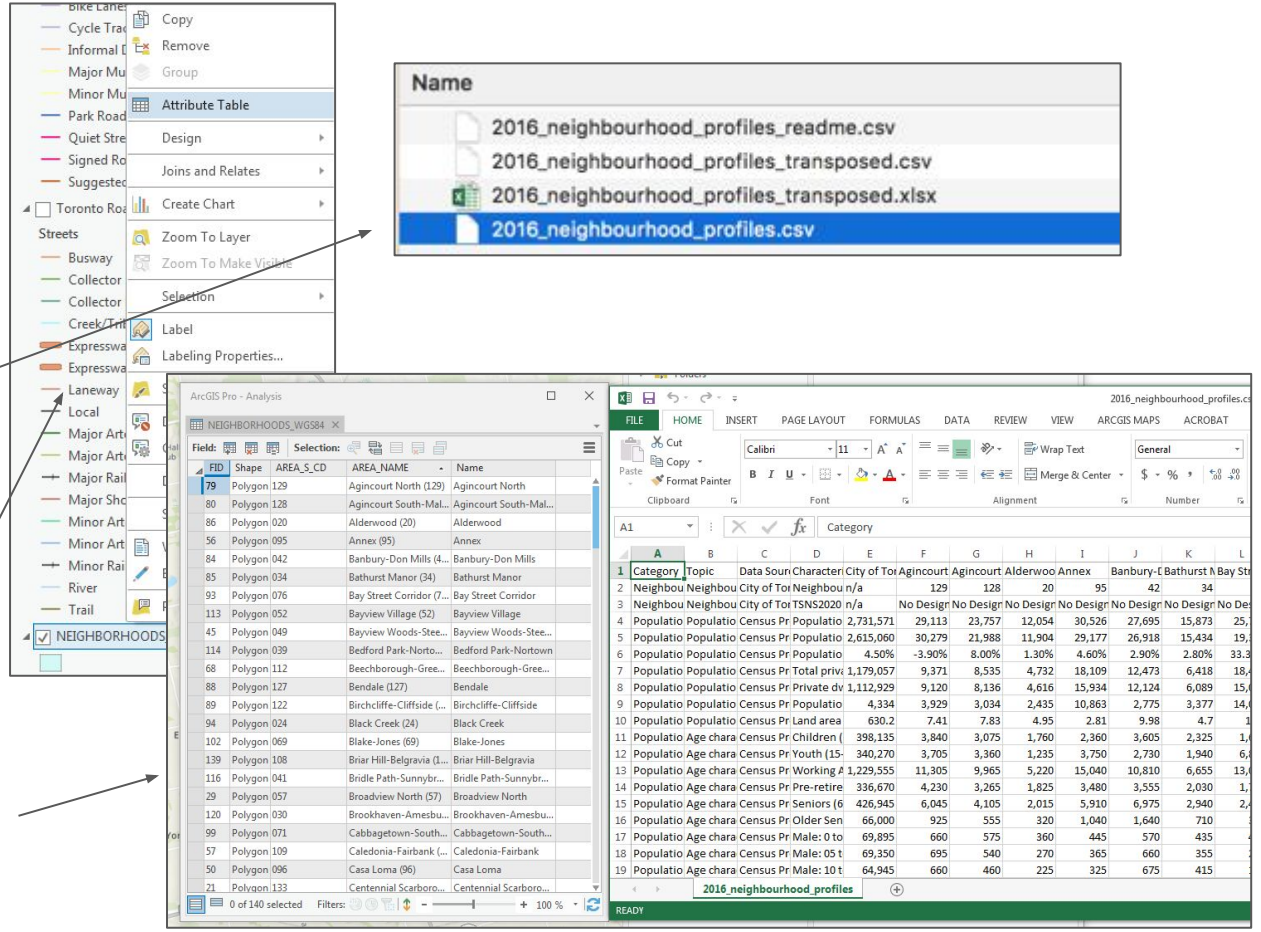
Attribute Joins

1 - in Windows, navigate to the “ArcGISWorkshop” → “data” → “Neighbourhood_Profiles” folder and double-click on the file “2016_neighbourhood_profiles.csv”

2 - right-click on the “NEIGHBORHOODS_WGS84” layer in the “Contents” pane and select “Attribute Table”

3 - compare the two. What might be some of the issues here?*

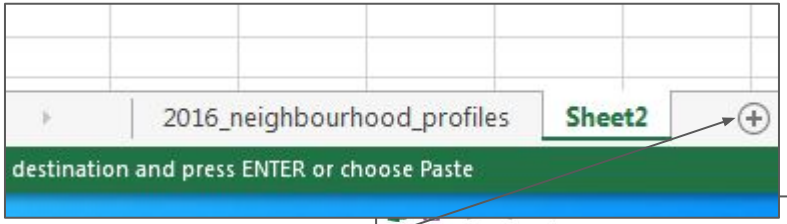
*Hint - the spreadsheet file has the neighbourhoods listed in individual columns. The GIS file has the neighbourhoods all in one column

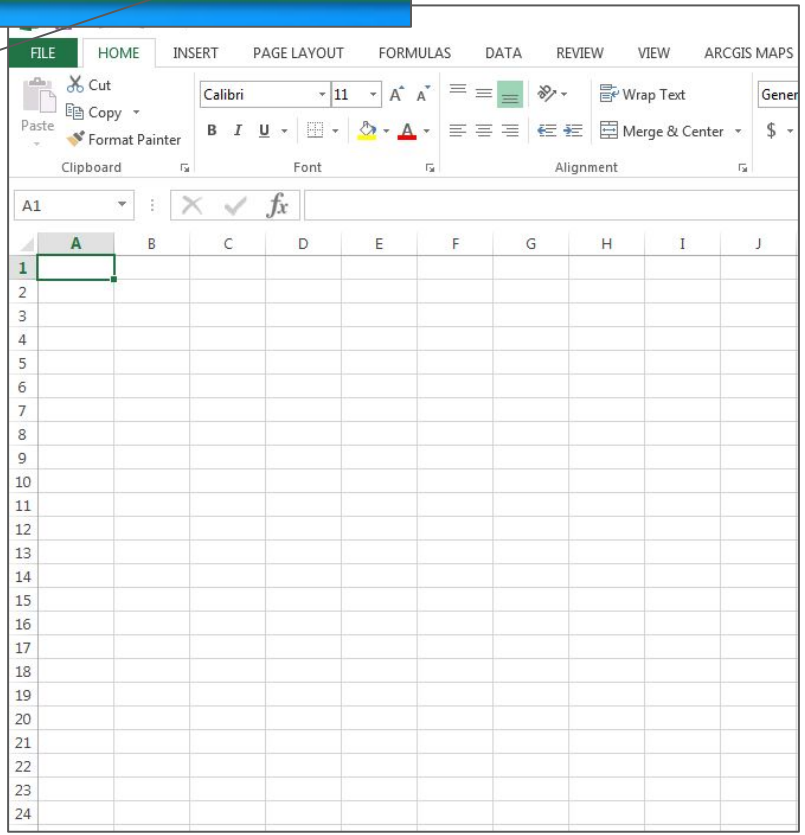


Attribute Joins

1 - to begin formatting the spreadsheet in a matching fashion, add a new tab to your spreadsheet by clicking on the “+” icon

2 - return to the “2016_neighbourhood_profiles” tab



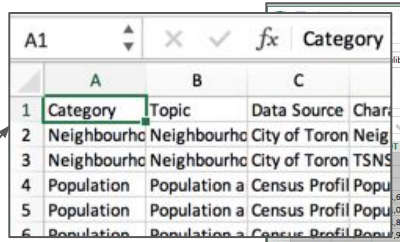


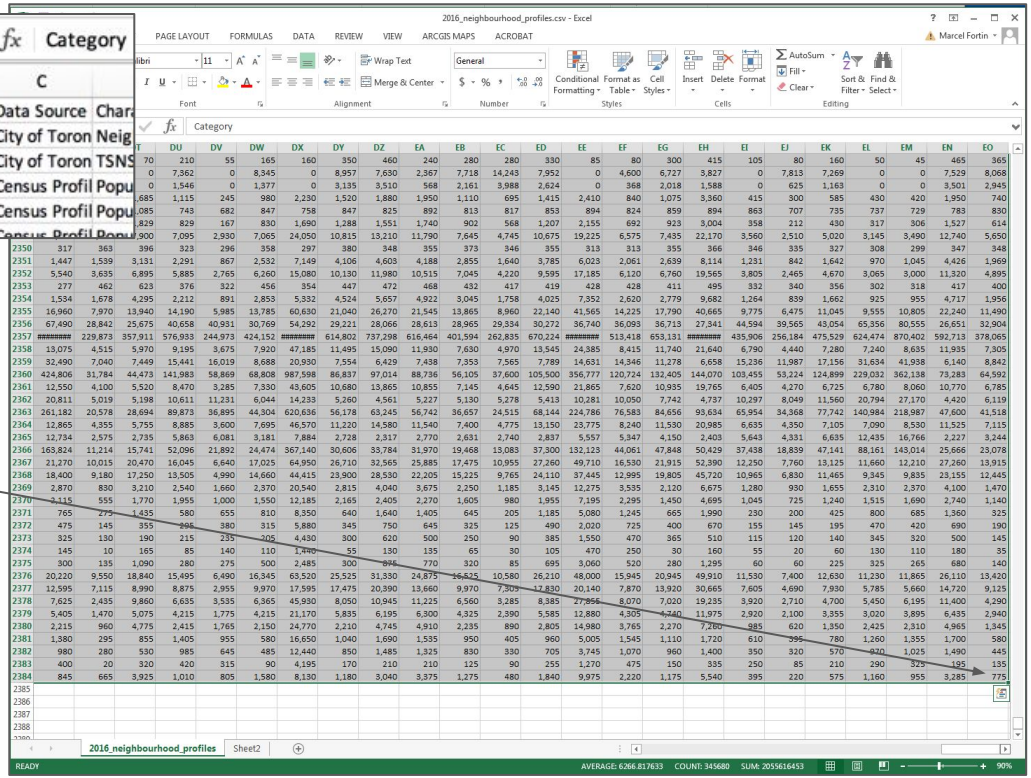
Attribute Joins

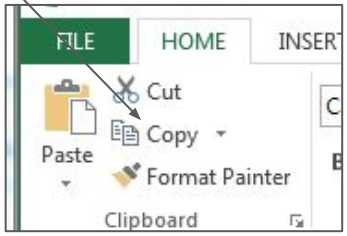
1 - click the first cell (A1) at the top-left of the table

2 - while holding down the shift key, click the last cell (EO2384)

3 - click on “Copy”





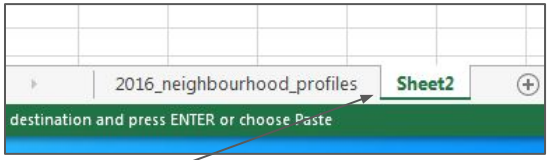


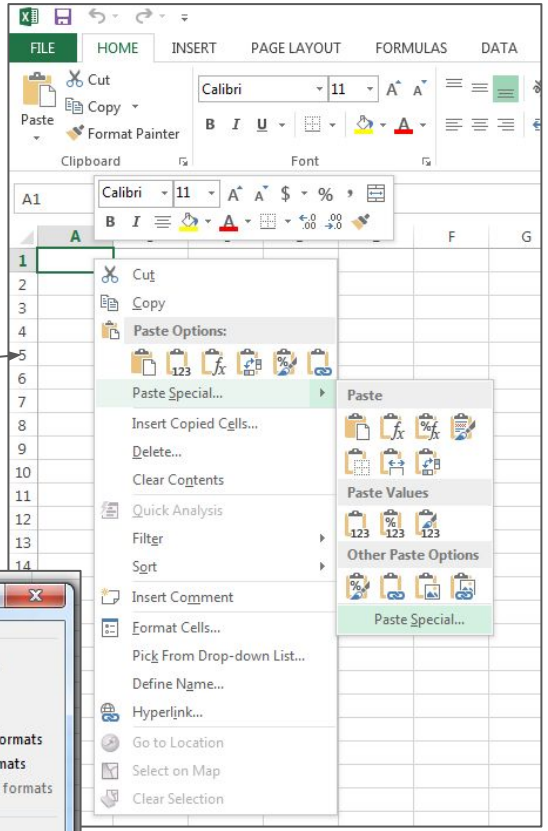
Attribute Joins

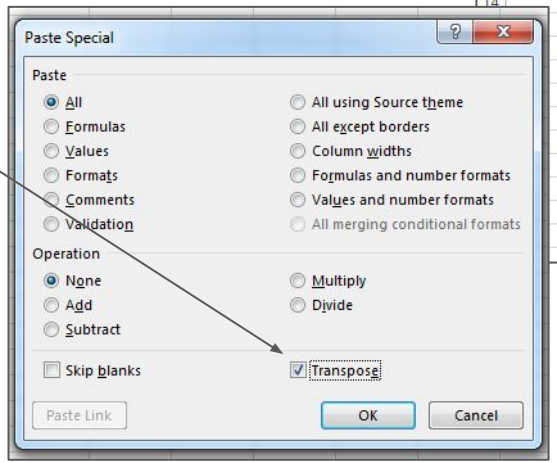
1 - click on the “Sheet2” tab

2 - right-click in the top-left (A1) cell, select “Paste Special…” → “Paste Special…”

3 - in the pop-up click to turn on the “Transpose” option and click “OK”





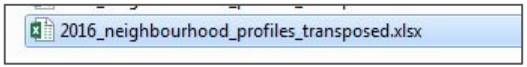


Attribute Joins

1 - delete the first three lines, and the fifth line of the spreadsheet

2 - you will notice that our spreadsheet now has the neighbourhood identifiers in columns

3 - close the spreadsheet (you can save it if you wish, but a saved version has already been created for you in the same folder called “2016_neighbourhood_profiles_transposed.xlsx”



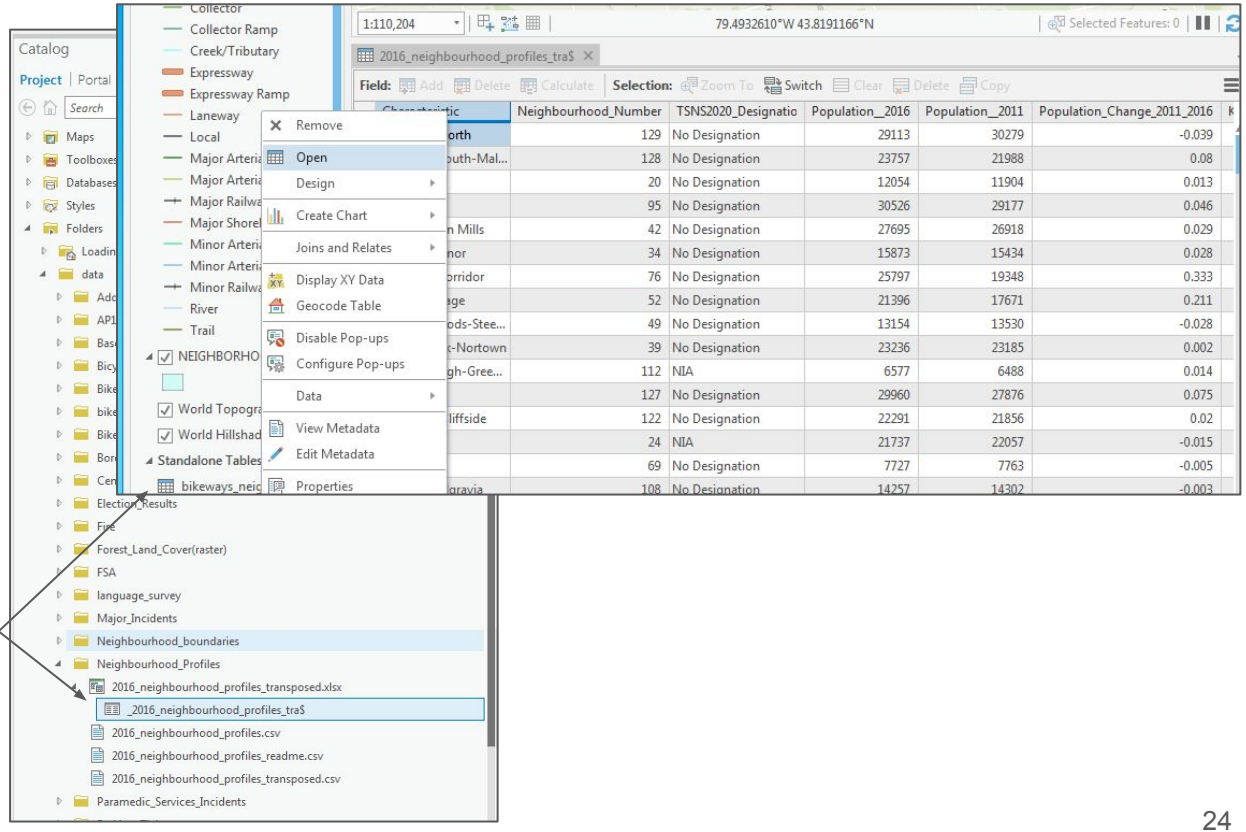
Attribute Joins

1 - In ArcGIS Pro, turn on the “Catalog” pane under the “View” menu

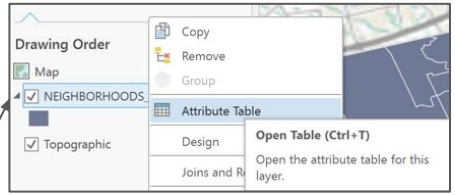
2 - in the “Catalog” pane, navigate to the “Neighbourhood_Profiles” folder

3 - expand the “2016_neighbourhood_profiles_transposed.xlsx” spreadsheet and select and drag into the map, the tab named “_2016_neighbourhood_profiles_tra$”

4 - right-click on “_2016_neighbourhood_profiles_tra$” in the “Contents” pane and click on “Open”

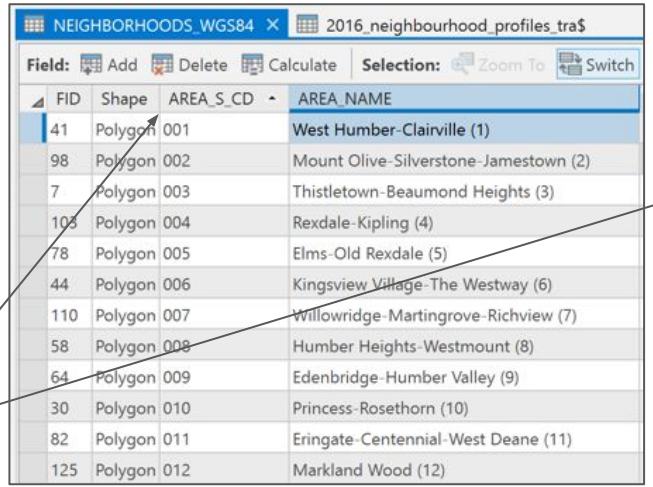


Attribute Joins

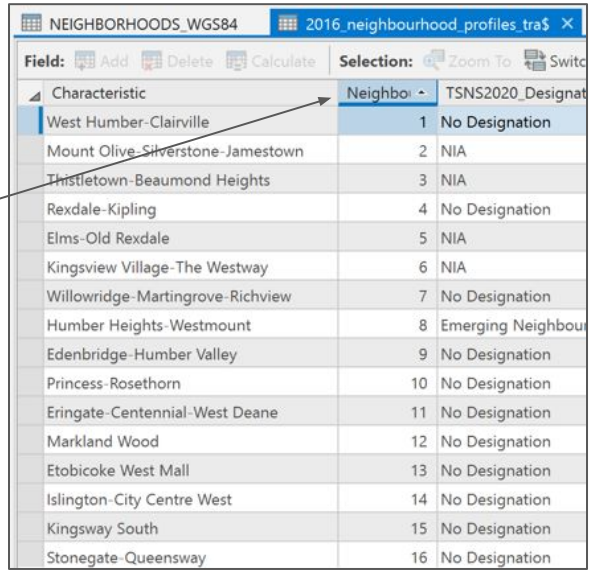


1 - right-click on the "NEIGHBORHOODS_WGS84" layer and select "Attribute Table" and compare the geography columns between this table and the spreadsheet

2 - notice that there are similar fields, but no common fields. "AREA_S_CD" in the boundary file is a text field. "Neighbourghood_Number" in the spreadsheet is a number field. ArcGIS Pro will not match these together despite having the same number. The column with the names of the neighbourhoods would not work either because they are not written the same way



FID	Shape	AREA_S_CD	AREA_NAME
41	Polygon	001	West Humber-Clairville (1)
98	Polygon	002	Mount Olive-Silverstone-Jamestown (2)
7	Polygon	003	Thistletown-Beaumond Heights (3)
103	Polygon	004	Rexdale-Kipling (4)
78	Polygon	005	Elms-Old Rexdale (5)
44	Polygon	006	Kingsview Village-The Westway (6)
110	Polygon	007	Willowridge-Martingrove-Richview (7)
58	Polygon	008	Humber Heights-Westmount (8)
64	Polygon	009	Edenbridge-Humber Valley (9)
30	Polygon	010	Princess-Rosethorn (10)
82	Polygon	011	Eringate-Centennial-West Deane (11)
125	Polygon	012	Markland Wood (12)



Characteristic	Neighbo	TSNS2020_Designat
West Humber-Clairville	1	No Designation
Mount Olive-Silverstone-Jamestown	2	NIA
Thistletown-Beaumond Heights	3	NIA
Rexdale-Kipling	4	No Designation
Elms-Old Rexdale	5	NIA
Kingsview Village-The Westway	6	NIA
Willowridge-Martingrove-Richview	7	No Designation
Humber Heights-Westmount	8	Emerging Neighbou
Edenbridge-Humber Valley	9	No Designation
Princess-Rosethorn	10	No Designation
Eringate-Centennial-West Deane	11	No Designation
Markland Wood	12	No Designation
Etobicoke West Mall	13	No Designation
Islington-City Centre West	14	No Designation
Kingsway South	15	No Designation
Stonegate-Queensway	16	No Designation

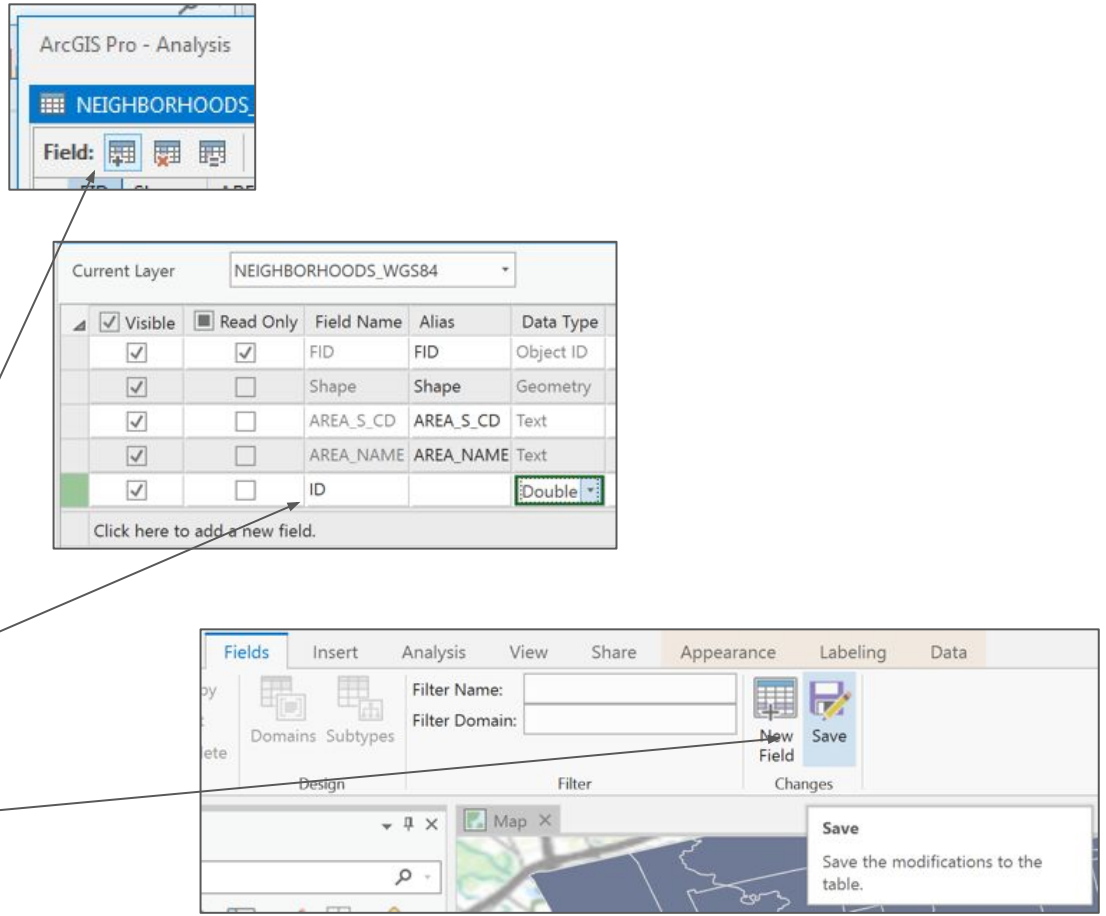
Attribute Joins

There are many ways to tackle this problem, but in this case we will create a new numeric field for our neirghbourhood ID in our shapefile

1 - in the “NEIGHBORHOODS_WGS84” attribute table, click on the “Add Field” icon

2 - name your field “ID” and select “Double” as the “Data Type”

3 - under the “Fields” menu, click on “Save”

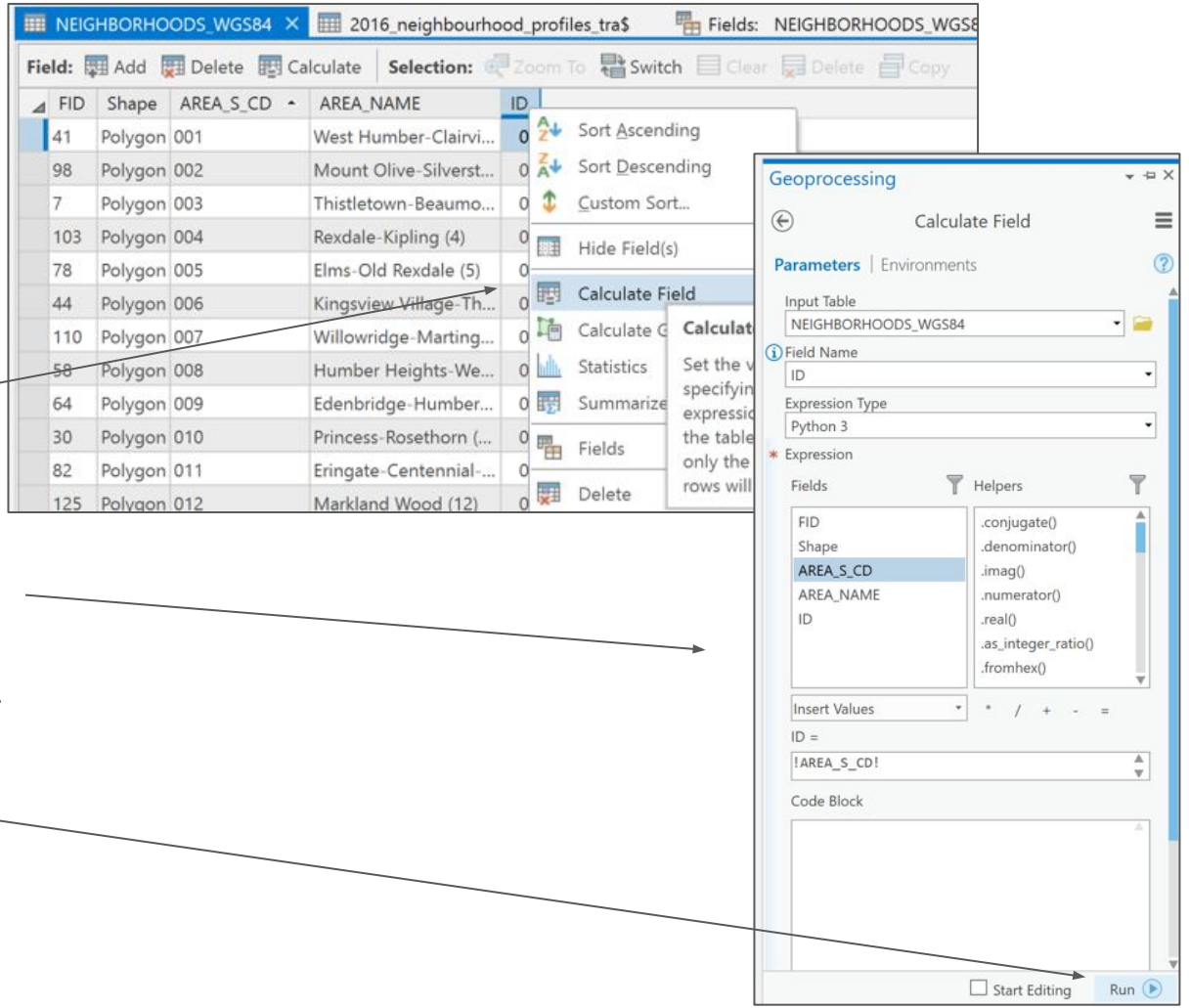


Attribute Joins

1 - go back to the “NEIGHBORHOODS_WGS84” attribute table and right-click on the new “ID” column and select “Calculate Field”

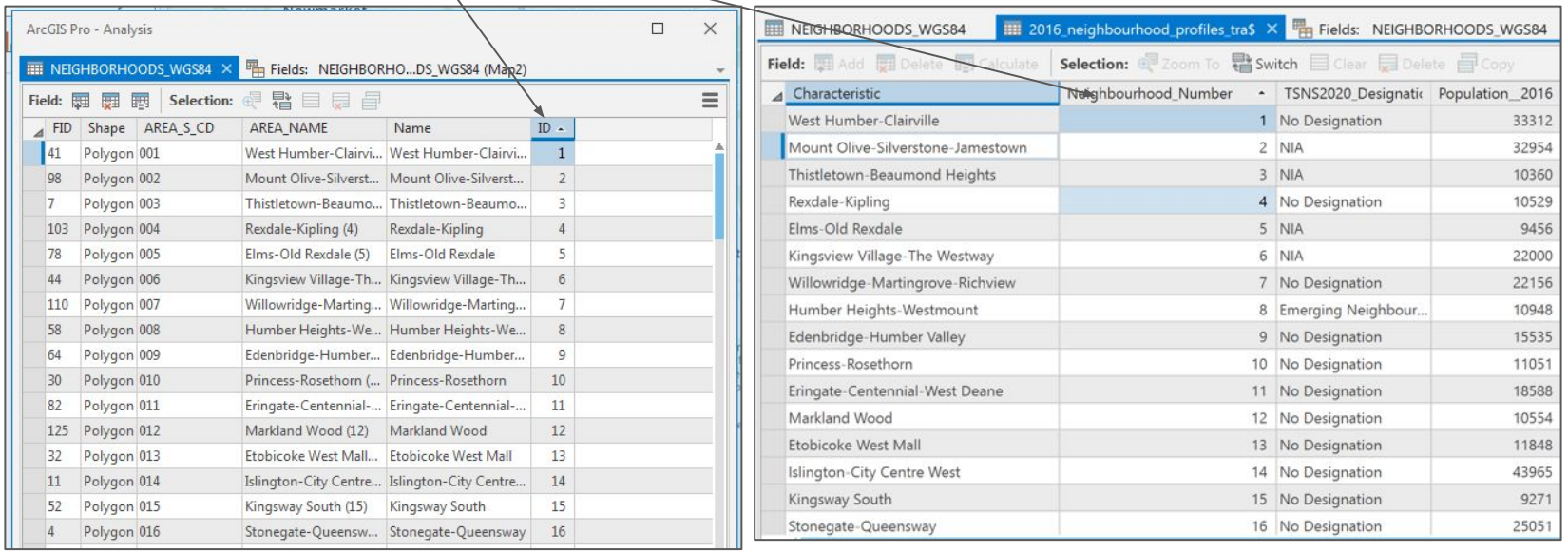
2 - double-click on the “AREA_S_CD” field to populate the “ID =” box to read “!AREA_S_CD!” This operation will convert our text field to a numeric field.

3 - click on Run



Attribute Joins

Our two tables now contain matching columns and we can now join the two layers.

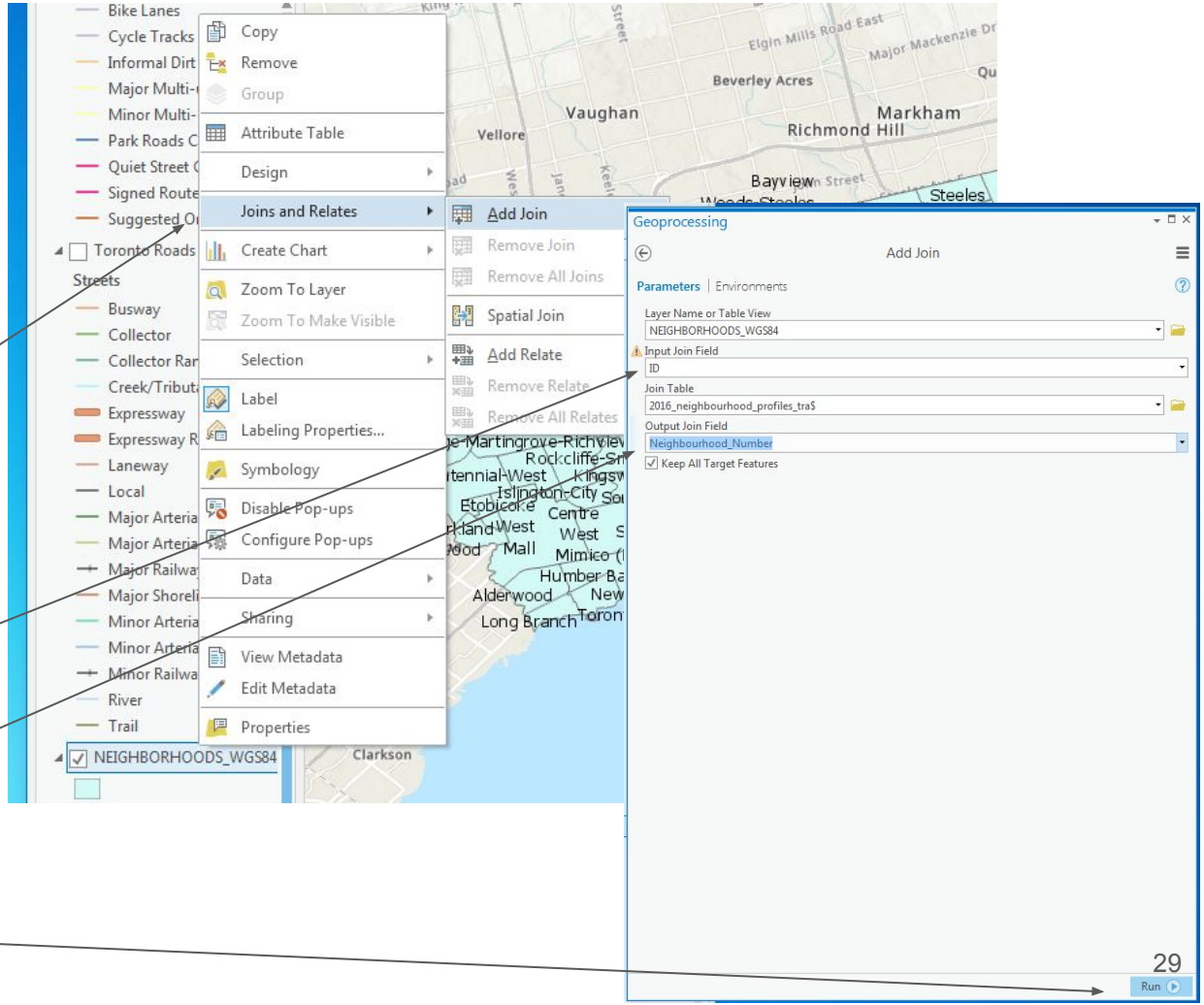


Attribute Joins

1 - right-click on the “NEIGHBORHOODS_WGS84” layer in the “Contents” pane and select “Joins and Relates” → “Add Join”

2 - select “ID” field as the “Input Join Field” from the “NEIGHBORHOODS_WGS84” table, and select the “Neighbourhood_Number” from the “Join Table”, “2016_neighbourhood_profiles_tra$”

3 - Click on “Run”

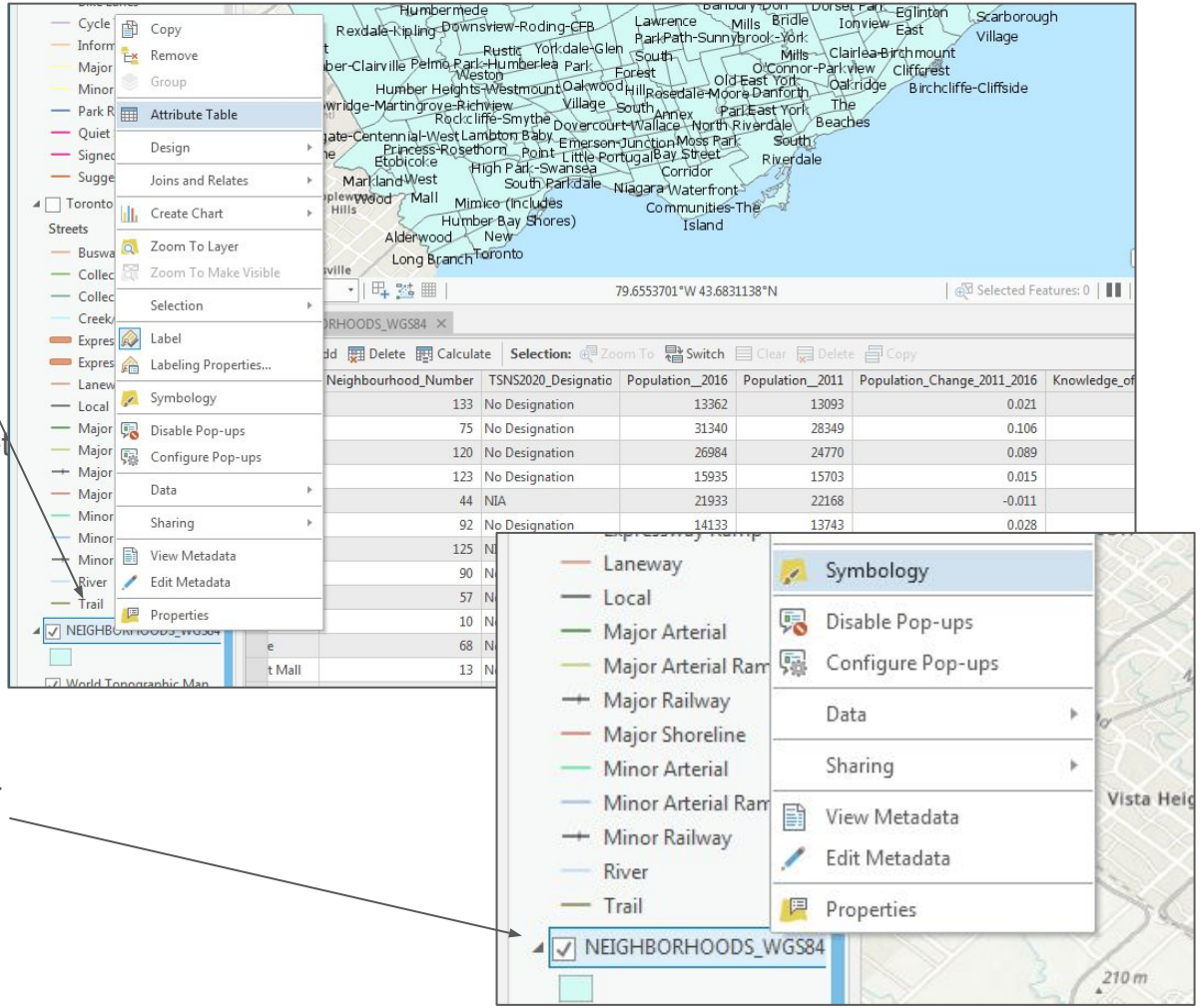


Attribute Joins

1 - open the “NEIGHBORHOODS_WGS84” “Attribute Table”. You will notice that now all the data from the spreadsheet are now available for each of the Toronto neighbourhoods.

Now let’s map some of these variables.

2 - right-click on the “NEIGHBORHOODS_WGS84” layer and select “Symbology”



Choropleth Mapping

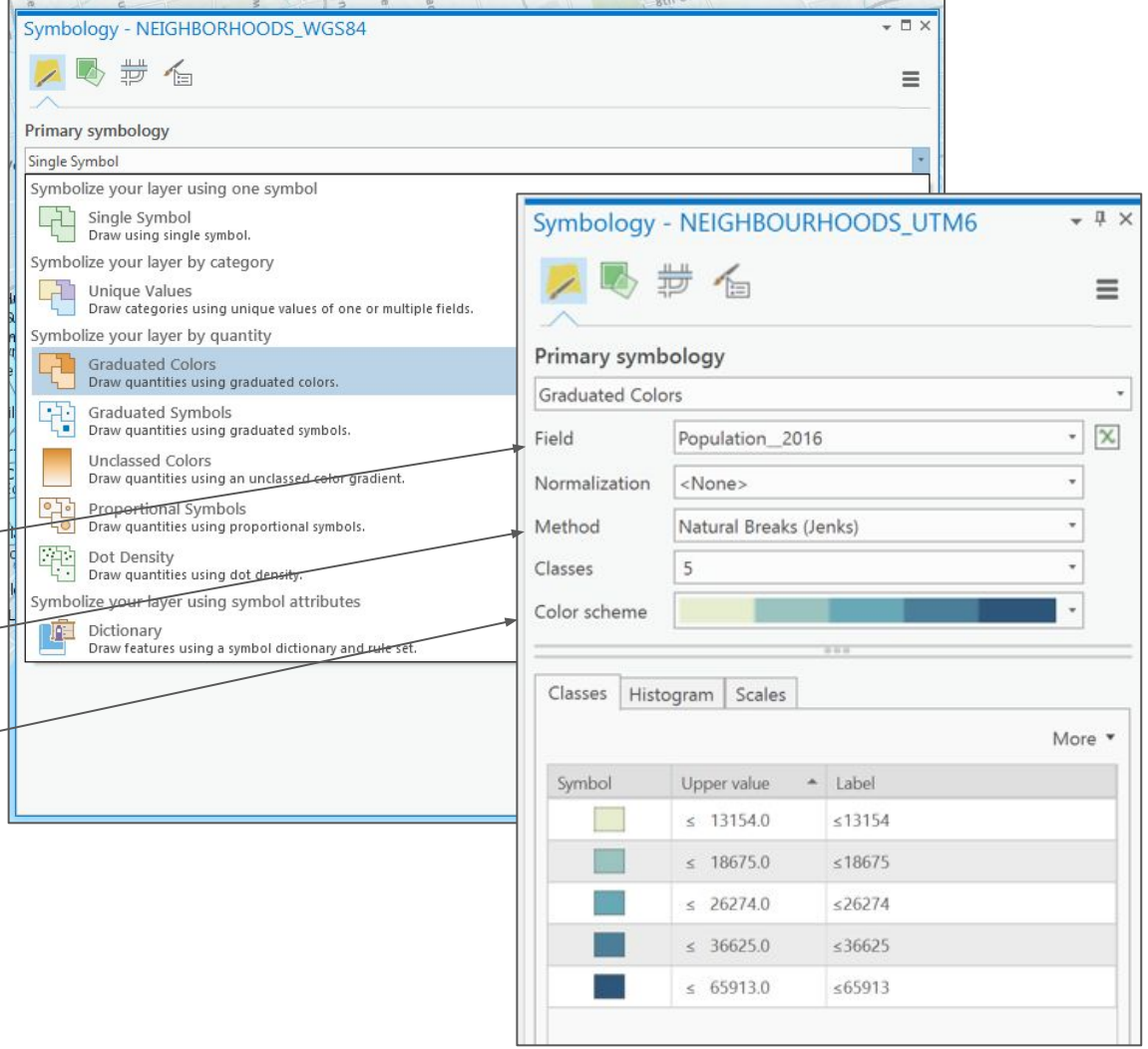
1 - in the “Symbology” popup window, select “Graduated Colors” under “Primary symbology”

2 - select “Population_2016” as the “Field”

3 - select “Natural Breaks (Jenks)” as “Method”

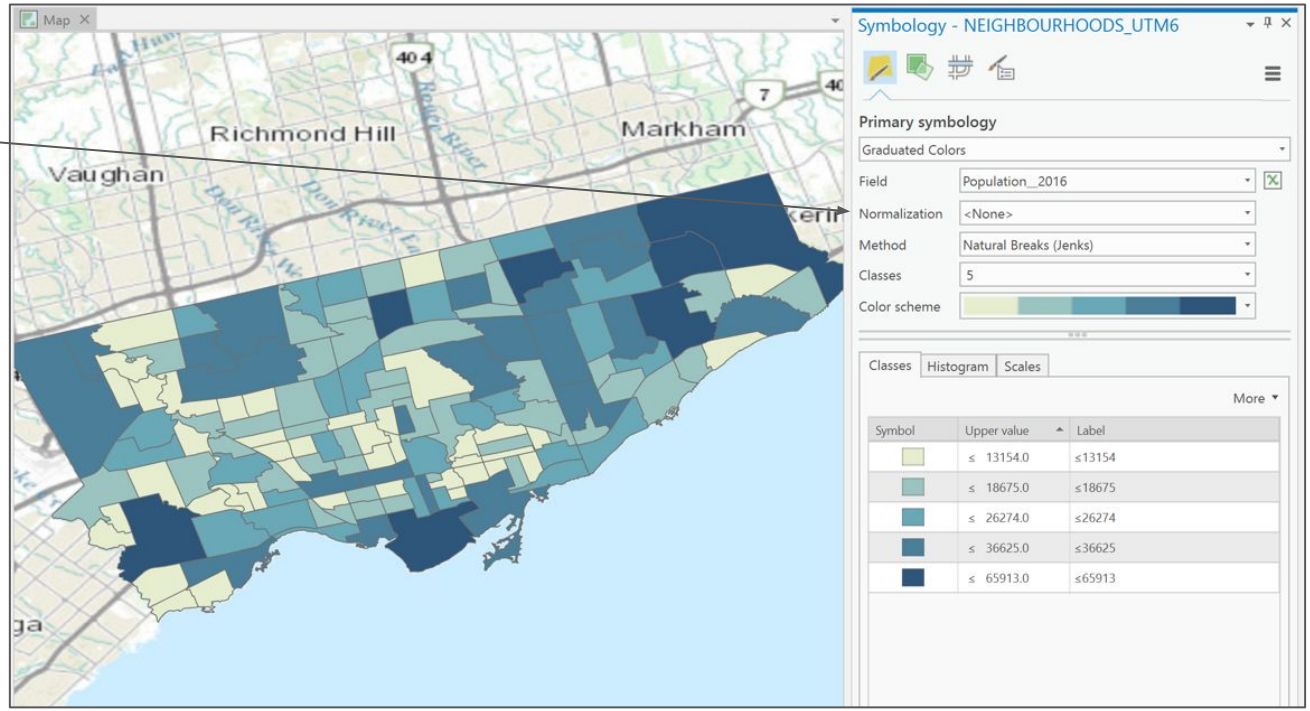
4 - choose a “Color scheme”

5 - Examine the resulting map



Choropleth Mapping

Since we have not “normalized” our data, what does is this map really expressing since we are comparing just population numbers?

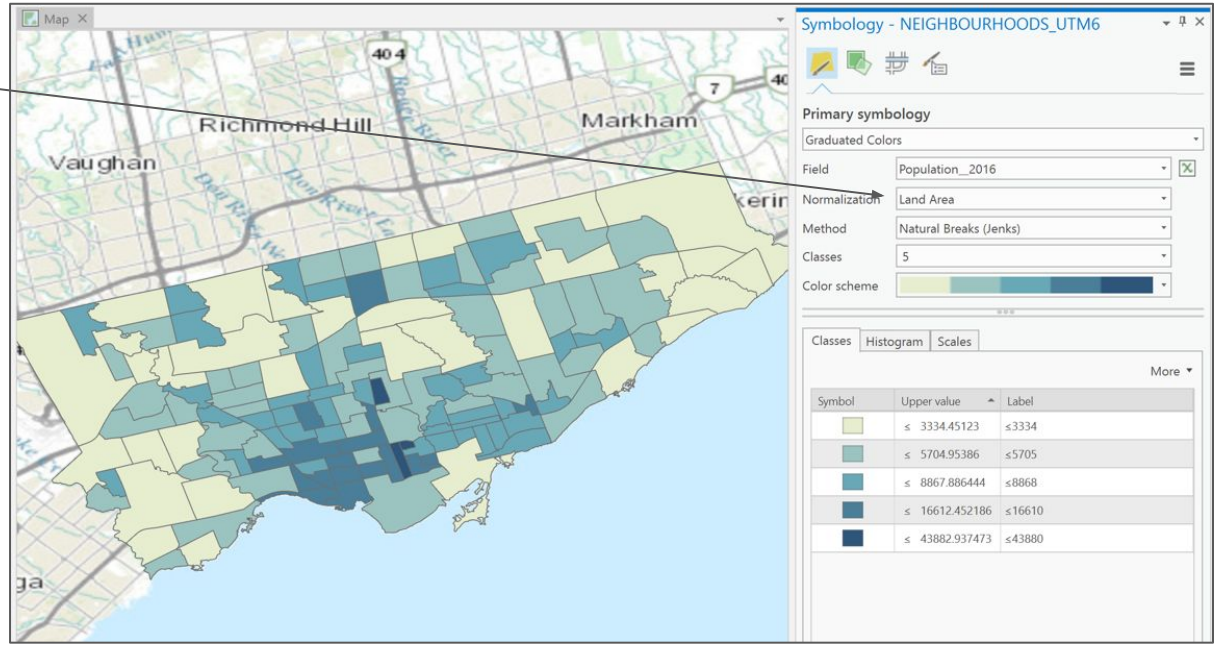


Choropleth Mapping

1 - Normalize your data using the field “Land Area”

2 - This now gives us population density

3 - “Natural Breaks” group data into natural groups within the data range. Using this method, features are divided into classes where there are relatively large differences in values. Since the least populated neighbourhood in Toronto is about 10% of the most populated one, this is a good starting point to classify our data.

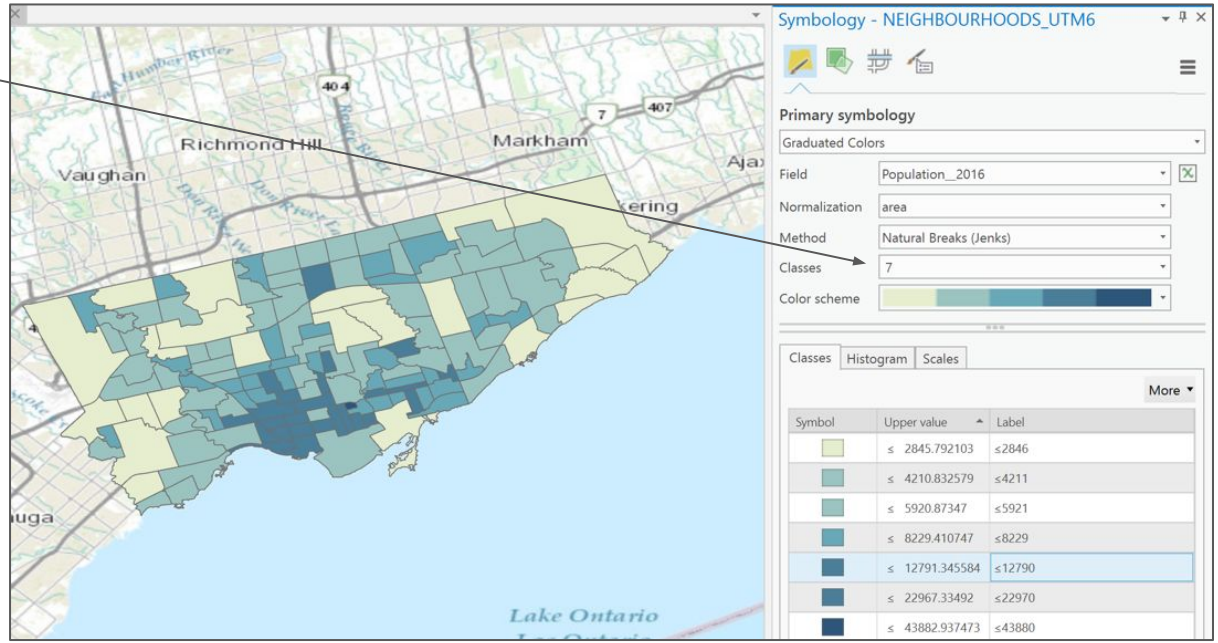


Choropleth Mapping

1 - change the number of “Classes” to “7”

You can see that this provides a little more precision to demonstrate where the highest densities are located.

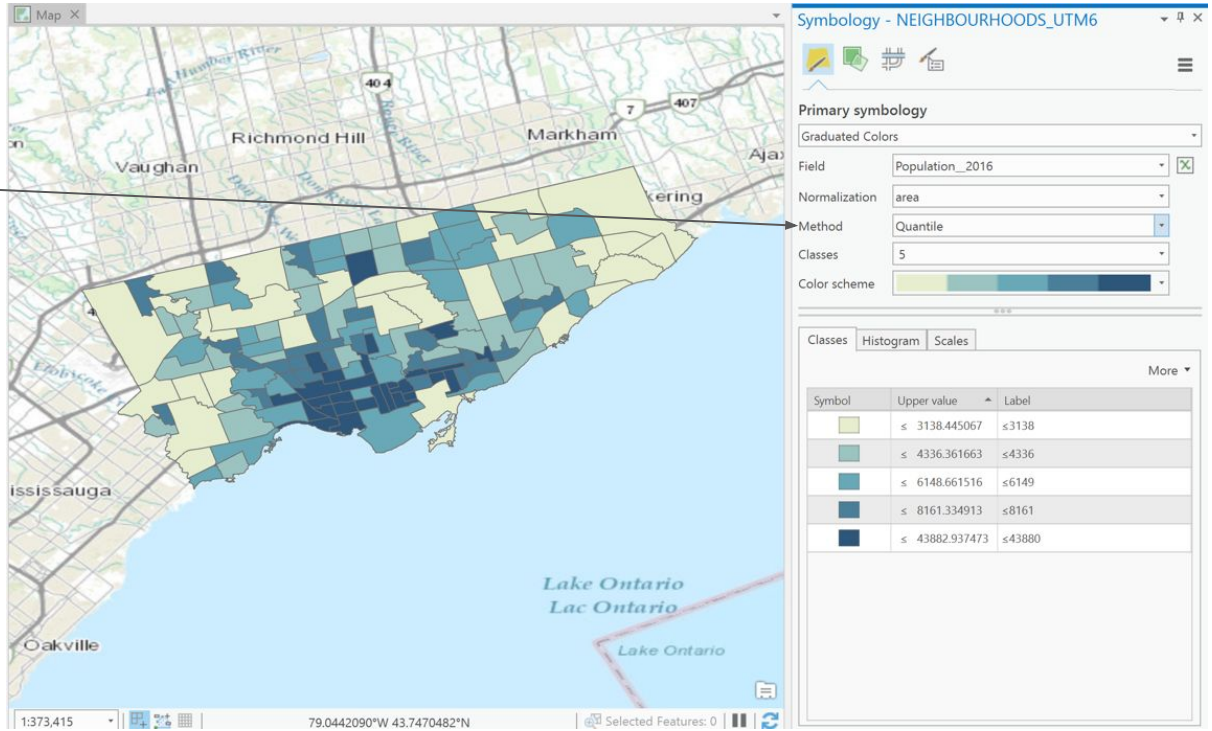
2 - Return the “Classes” to “5”



Choropleth Mapping

1 - Change the “Method” for classification to “Quantile”

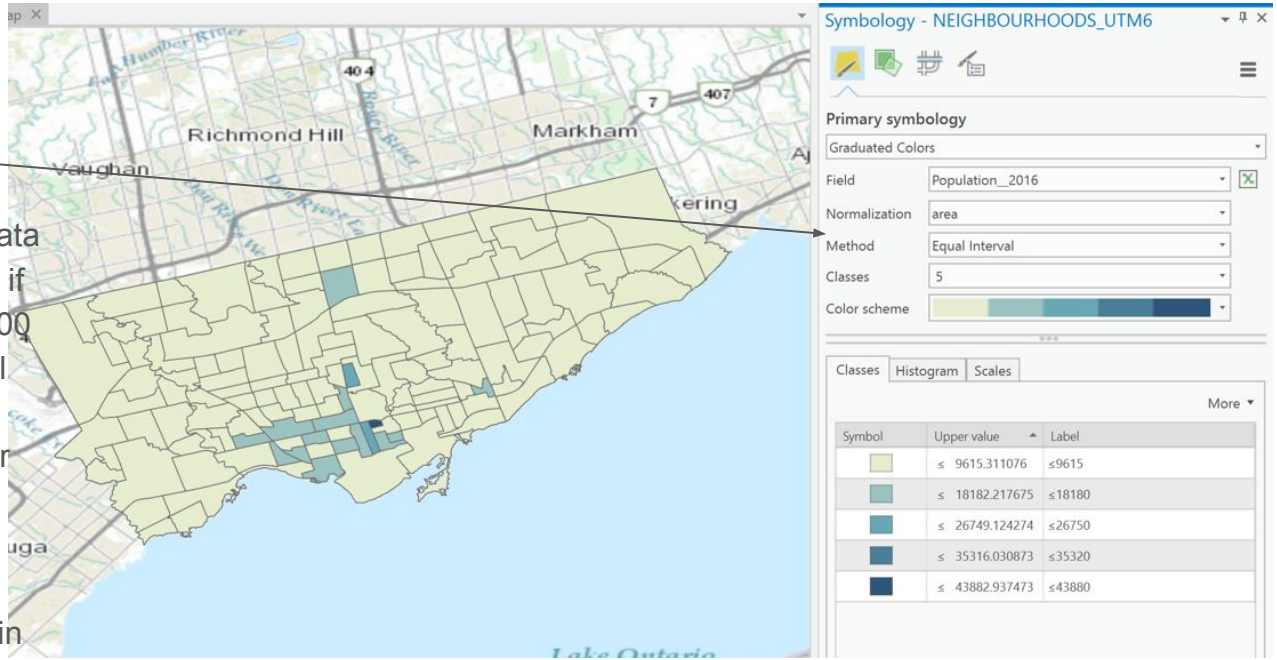
Quantile distributes evenly the number of features (or entries in the attribute table) within the number of classes regardless of the jumps in values. Best suited for linear data, this is not a particularly good choice for Toronto’s big differences in population and density across all of its geography.



Choropleth Mapping

1 - Change the “Method” for classification to “Equal Interval”

Equal Interval divides the attribute data into equal sub-ranges. For example, if you have values ranging from 0 to 100 and you specify 5 ranges, values will vary from 0-25, 26-50, 51-75 and 76-100. Unlike quantiles, the number of features within a class will differ. Because the density of Toronto’s population is mostly in the southern core, most values end up displayed in few classification colours.

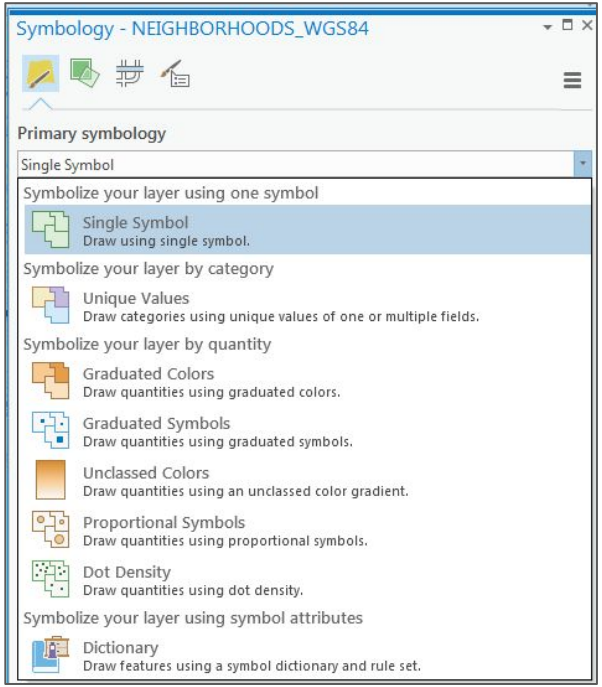


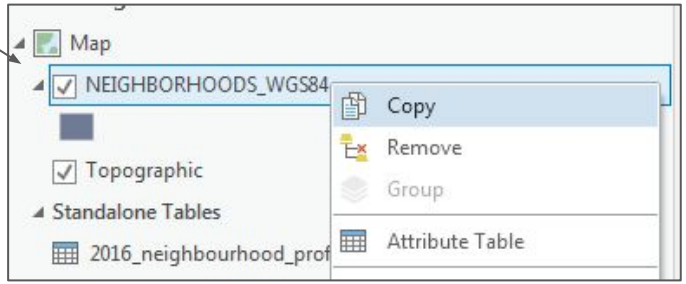
Graduated Symbols

1 - return your map to “Single Symbol” classification under the “Primary symbology” option

2 - right-click on the “NEIGHBORHOODS_WGS84” layer in the “Contents” pane and select “Copy”

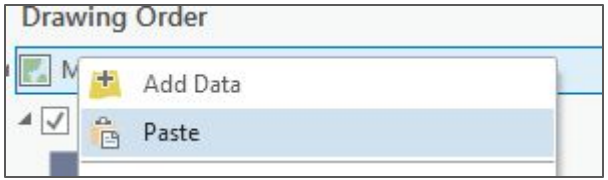
Graduated symbols show quantitative differences between mapped features by varying the size of symbols. Data are classified into ranges that are each assigned a symbol size to represent the range. See https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/graduated-symbols.htm



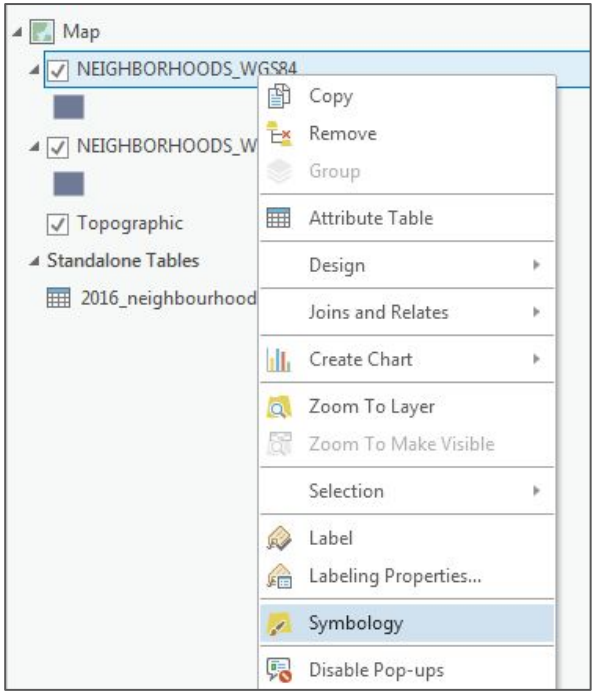


Graduated Symbols

1 - right-click on the “Map” in the “Contents” pane and select “Paste”



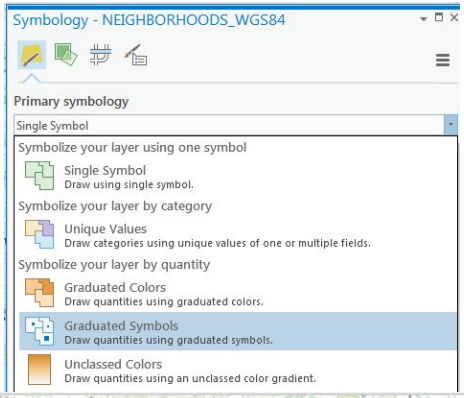
2 - right-click on the top-resulting “NEIGHBORHOODS_WGS84” layer in the “Contents” pane and select “Symbology”

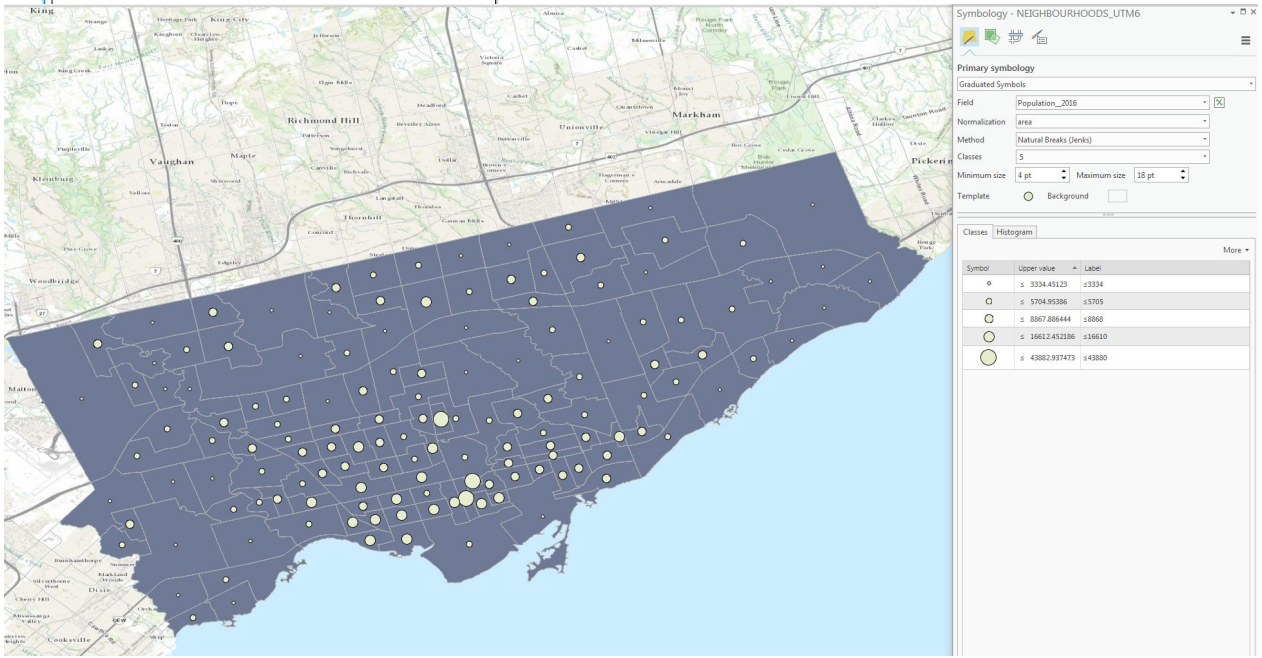


Graduated Symbols

1 - in the “Symbology” popup window, select “Graduated Symbols”. Graduated Symbols show quantitative differences between features by varying the size of the symbols.

2 - use “Population_2016 for the “Field” and normalize by “area” and play around with different classification “Methods”; “Natural Breaks”, “Quantile” and “Equal Interval” and with different “Classes” 5, 6 or 7.

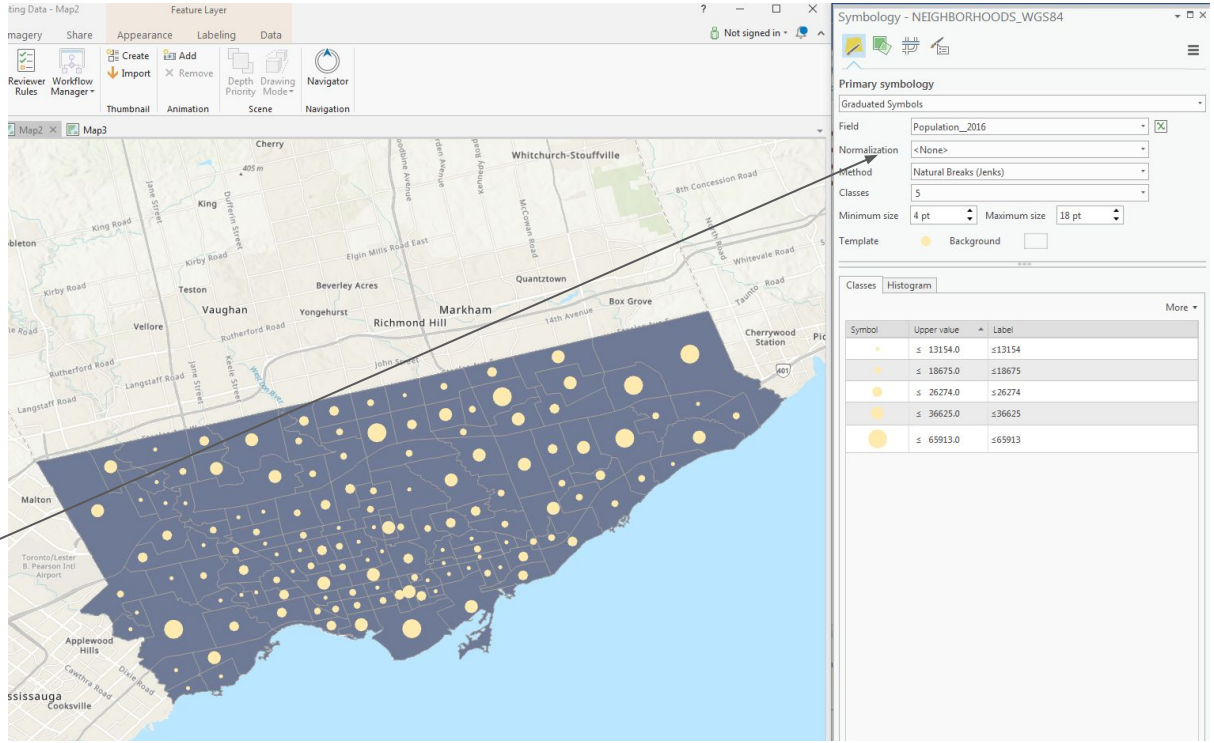




Graduated Symbols

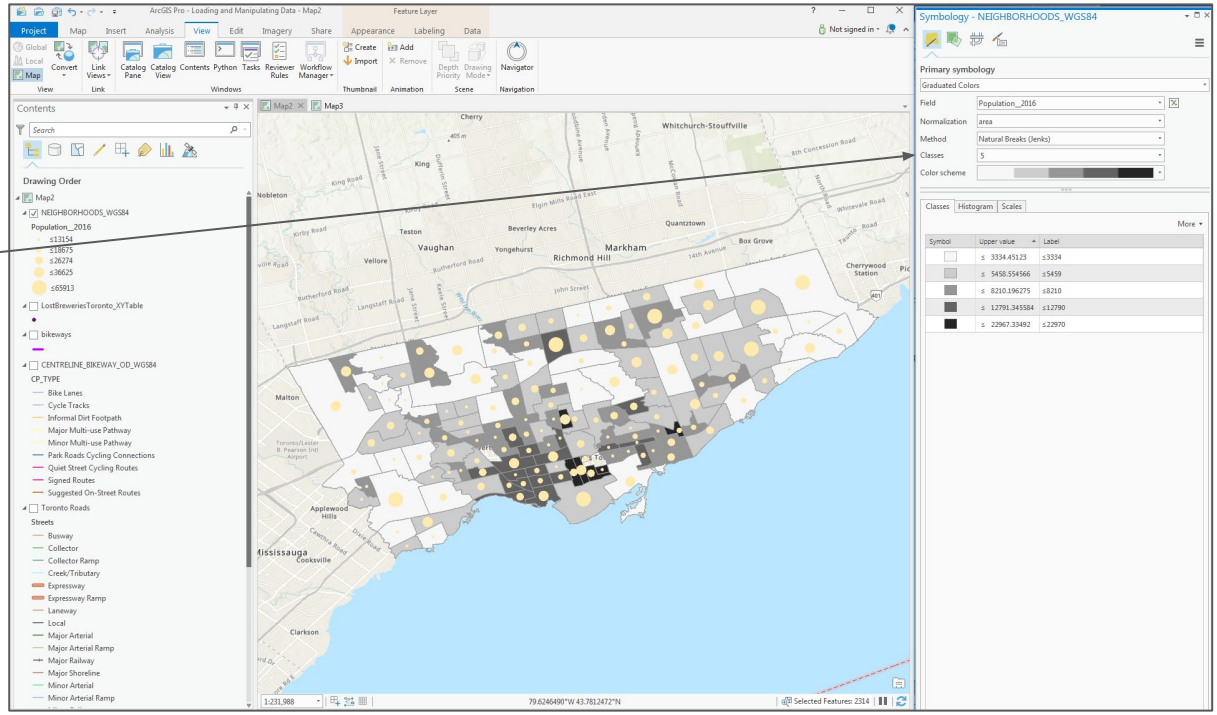
Graduate colors do not work well with raw population numbers. However, graduated symbols do work well with raw populations.

1 - change the “Normalization” to your symbology for your “Graduated Symbols” to “<None>”



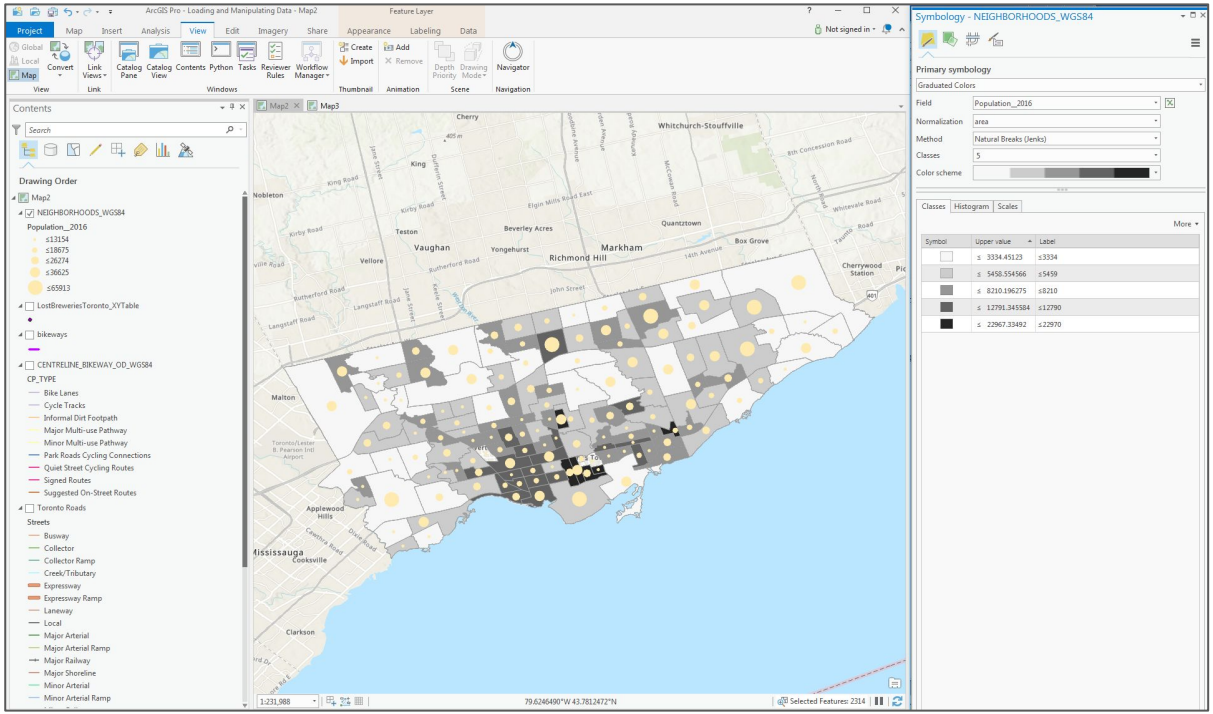
Graduated Symbols

1 - change the “Symbology” of the second copy of “NEIGHBORHOODS_WGS84” to “Graduated Colors” with the 2016 population normalized by area using the “Natural Breaks (Jenks)” “Method”



Our resulting map is interesting because it allows for displaying both full population numbers and the population density at the same time.

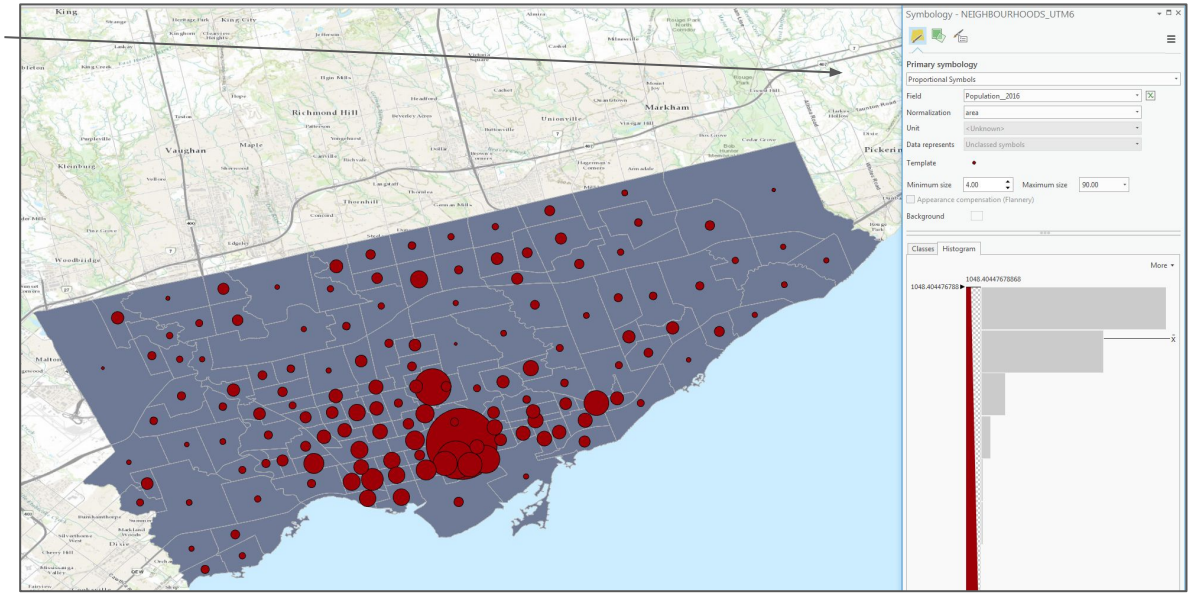
Graduated Symbols



Proportional Symbols

1 - Return to “Single Symbol” for the second neighbourhood layer and change the “Primary symbology” to “Proportional Symbols” on the other

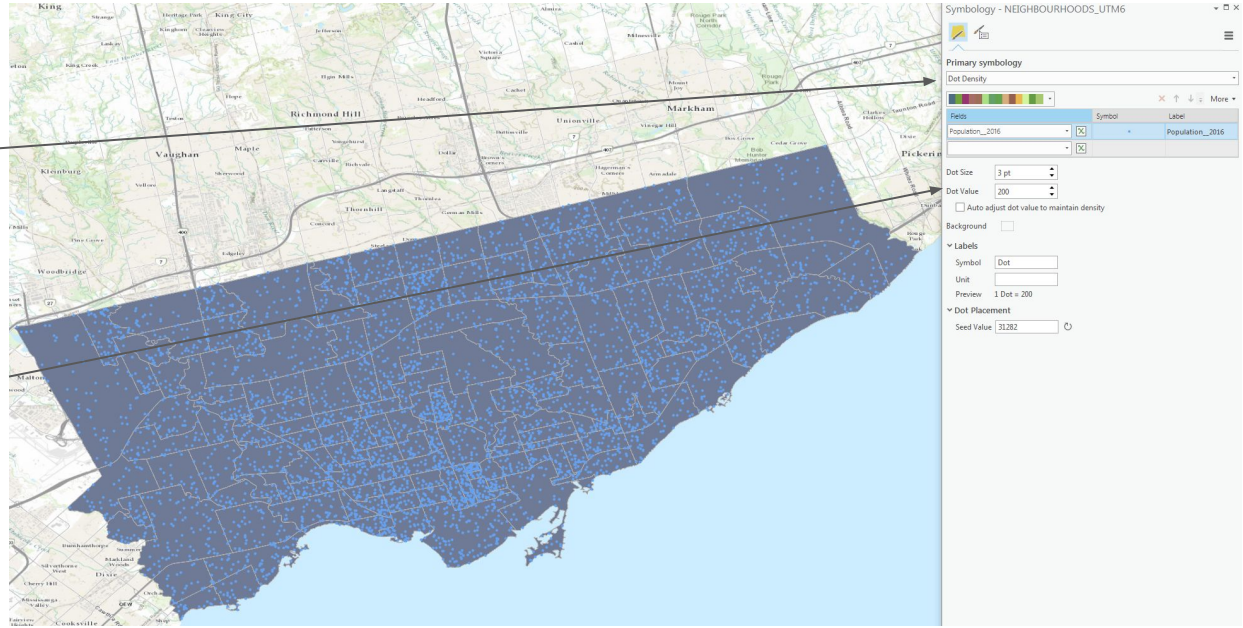
Proportional Symbols are used to show relative differences in quantities among features. Unlike graduated symbols, proportional symbols represent quantitative values as a series of unclassed symbols, sized according to each specific value.



Dot Density

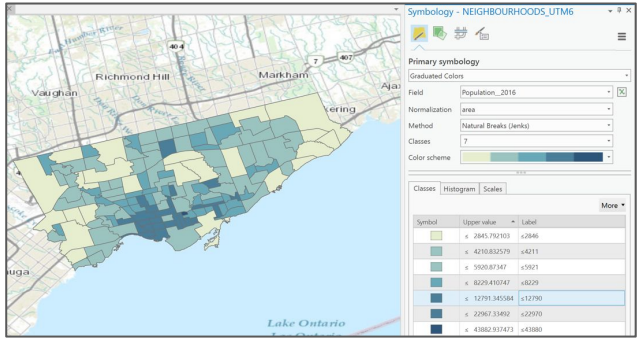
1 - change the “Primary Symbology” to “Dot Density” and select “Population_2016” as the Field to represent

2 - Change around the values for the “Dot Size” and the “Dot Value”. “Dot Value” in this case means the number of people the dots represent.

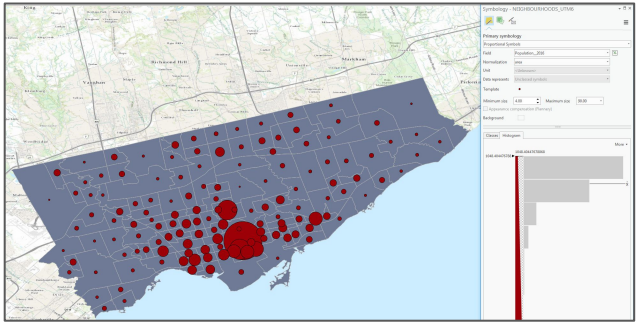


Discussion

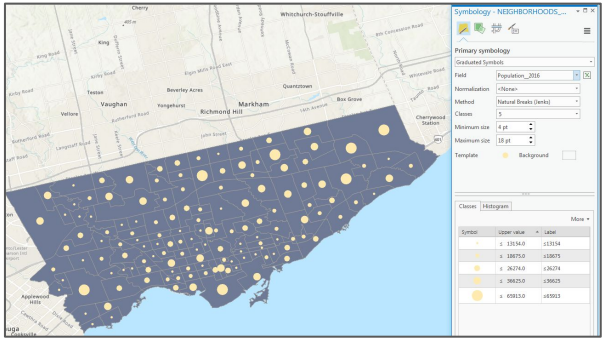
In your estimation, which better represents the population density of the City of Toronto? Graduated Colours, Graduated Symbols, Proportional Symbols or Dot Density?



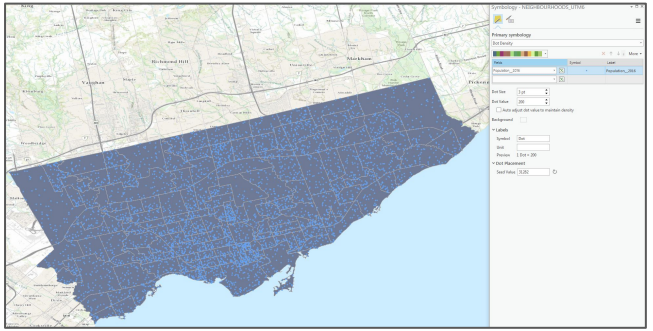
Graduated Colours



Proportional Symbols



Graduated Symbols

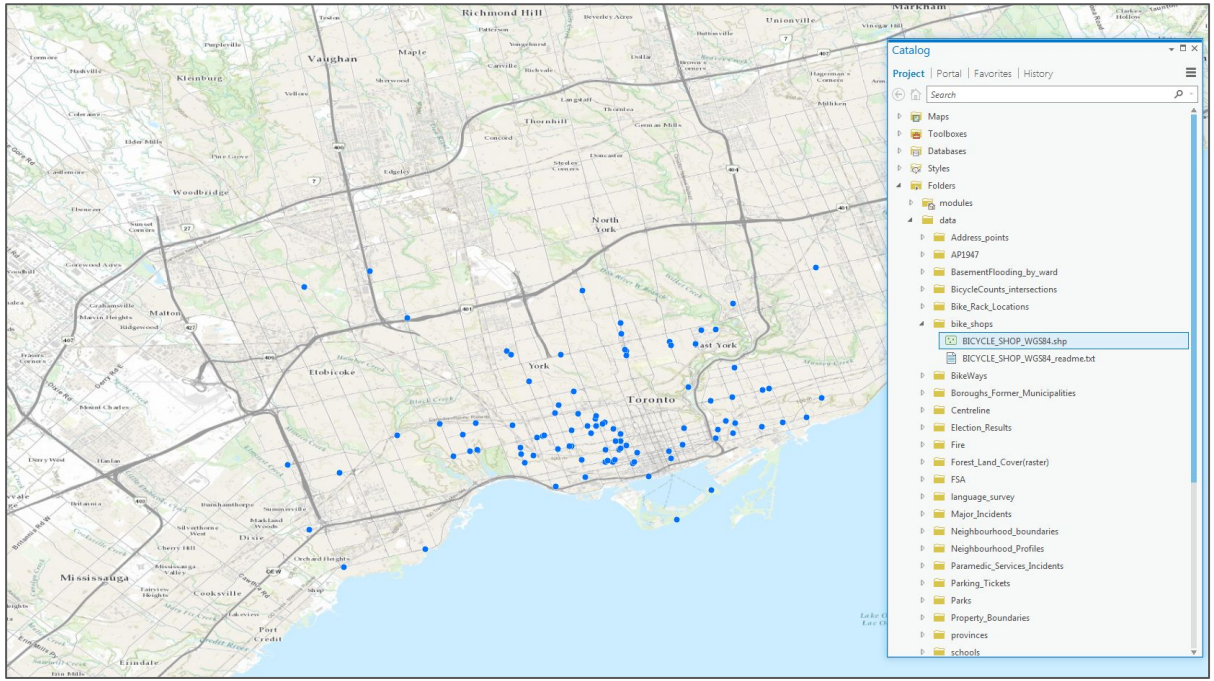


Dot Density

Solving Problems with GIS

Problem: how do we find an ideal location to set up a new bicycle shop in Toronto where there are no established stores, but within a short distance of bicycle traffic in Toronto?

1 - In a new map, start by loading the “BICYCLE_SHOP_WGS84.shp” shapefile from the “bike_shops” folder



Buffering

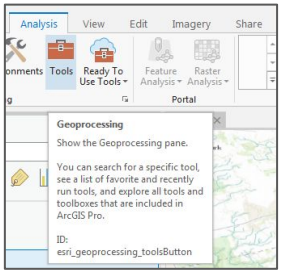
1 - using the “Analysis” menu, click on “Tools”

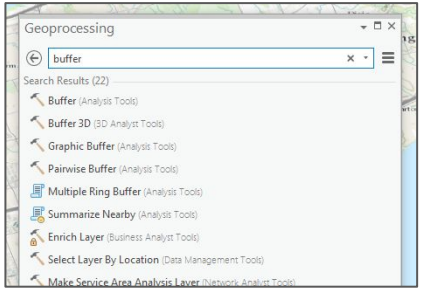
2 - in the “Geoprocessing” popup, type in “buffer” to find the “Buffer” tools

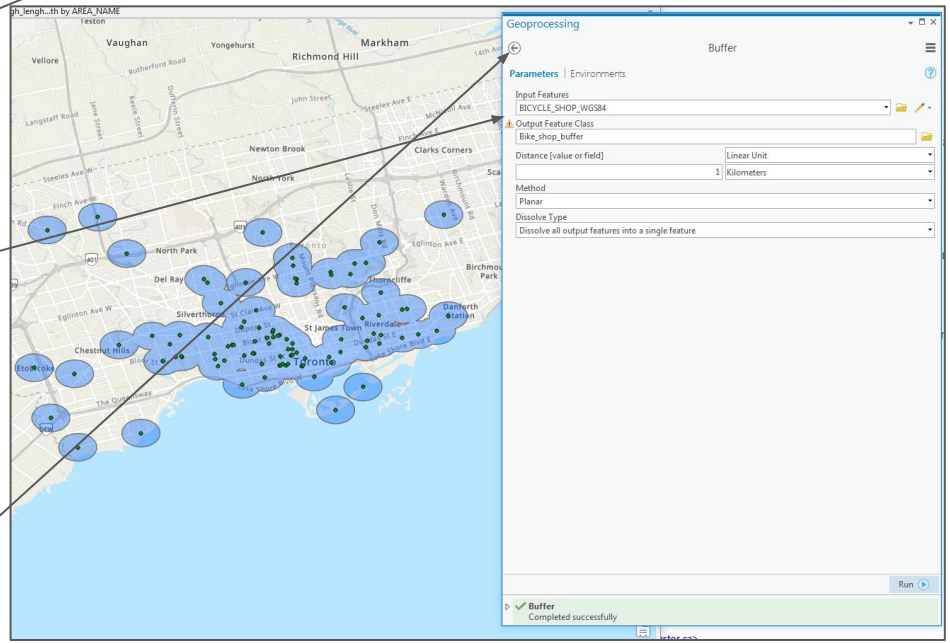
3 - click on “Buffer (Analysis Tools)”

4 - provide a name to the resulting buffer layer “Bike_shop_buffer” select a distance of 1 KM buffer and select “Dissolve all output features into a single feature” option for the “Dissolve Type”

5 - click on “Run”and return to the main “Geoprocessing” window by clicking on the back arrow





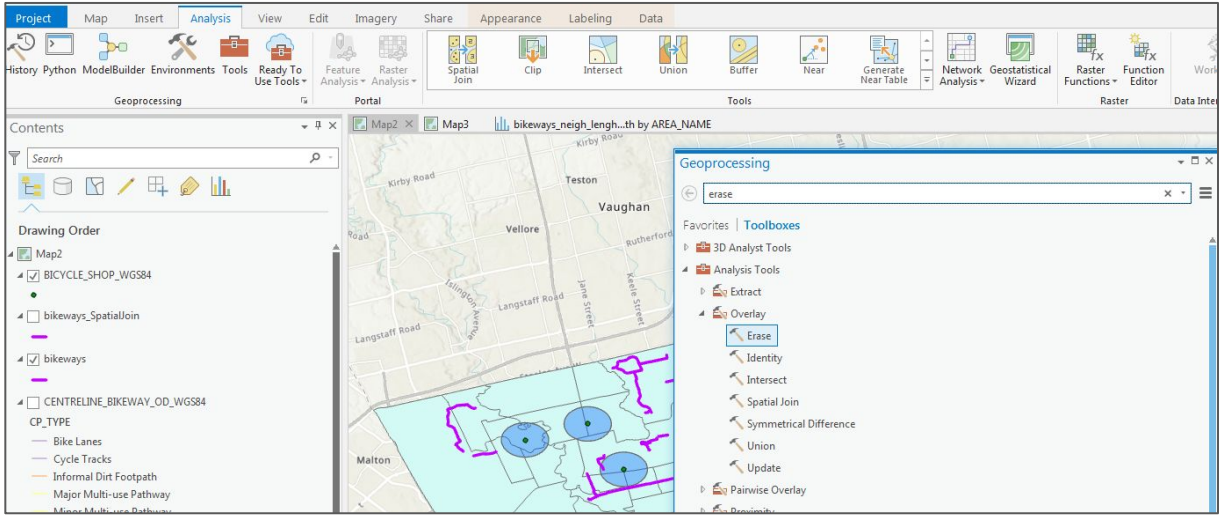


Erase

Our buffer represents areas we do not want to build a new bike shop in since the areas are already serviced. We need to “erase” these from our possible locations.

1 - in the “Geoprocessing” window search box, type “erase”

2 - click on the “Erase” toolbox

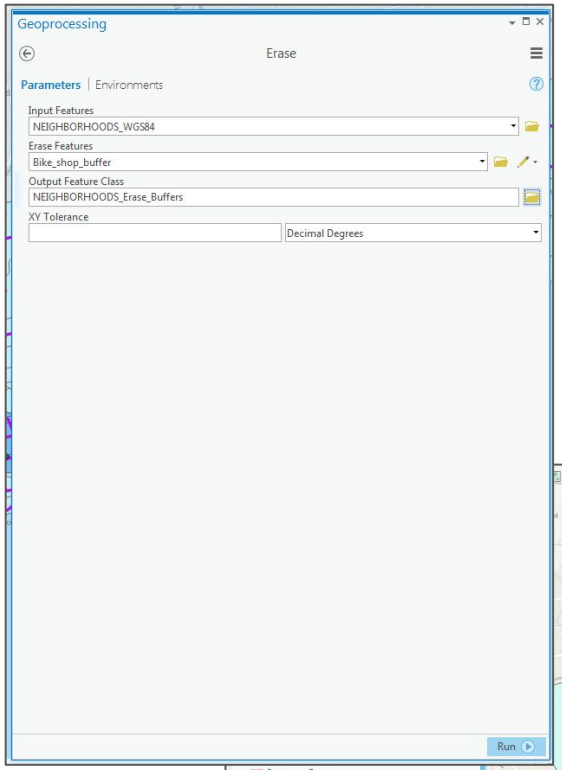


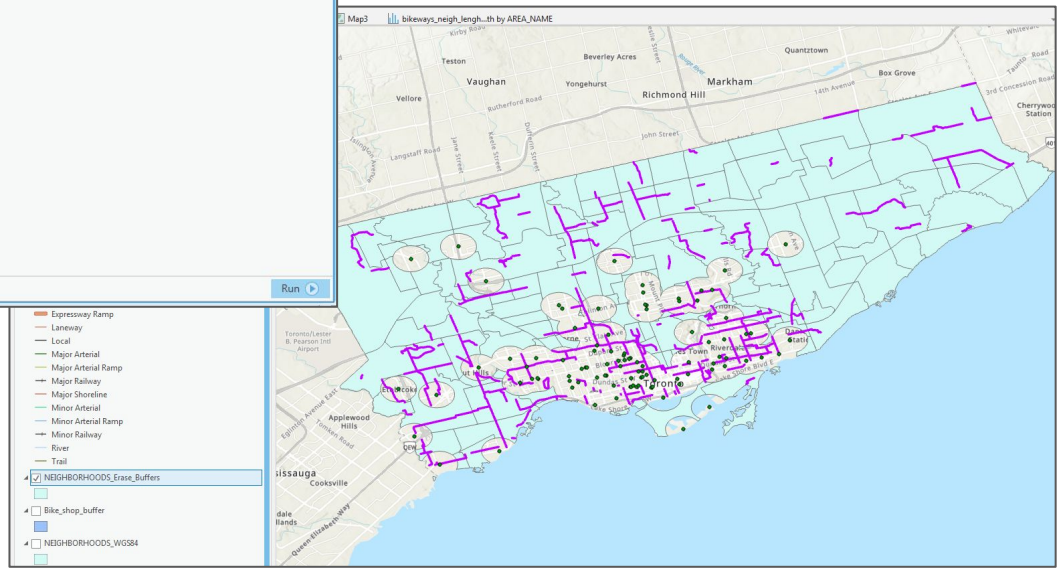
Erase

1 - your “Input Features” are the “NEIGHBOURHOODS_WGS84” layer from which we want to erase the buffer we just created “Bike_shop_buffer”

2 - provide a name from the resulting neighbourhoods “Output Feature Class” “NEIGHBOURHOODS_Erase_Buffers”

3 - click on “Run”



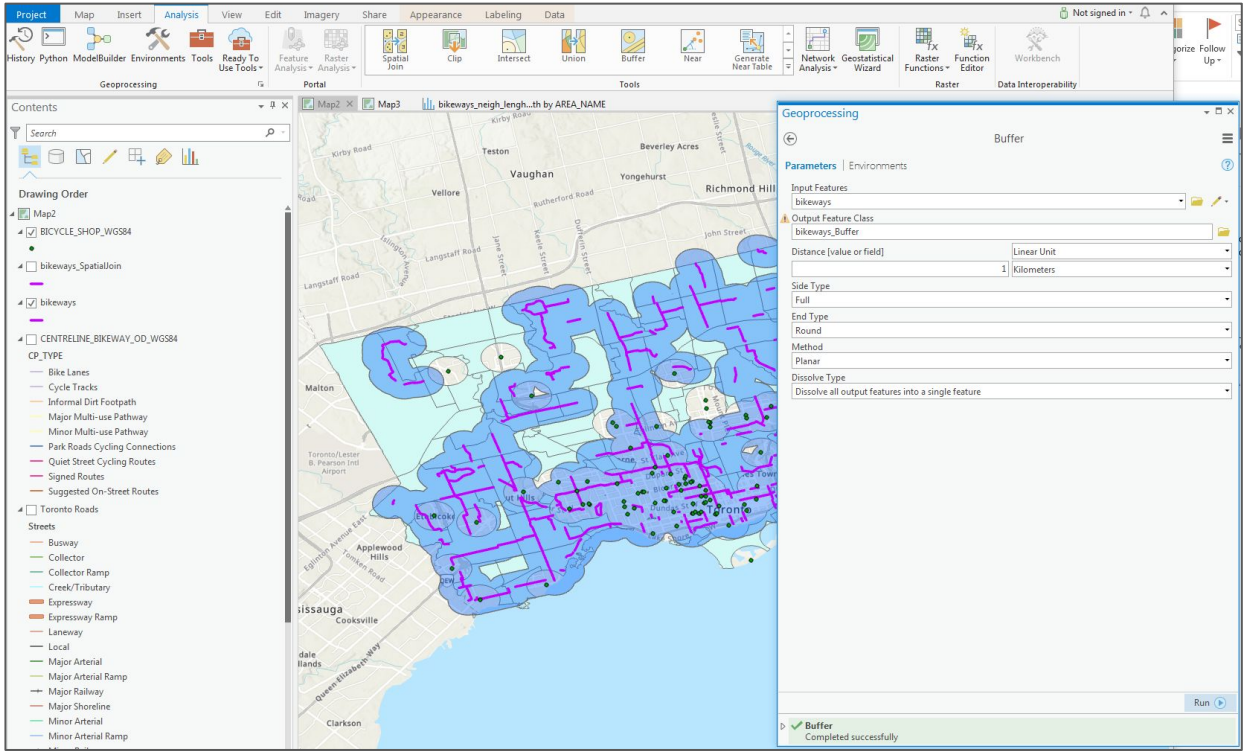


Buffering Lines

We now want to find areas that are within 1KM of our “bikeways” features created earlier. To accomplish this, we will create a new buffer, but we will not delete from this one, we will instead “clip” out our city layer from our resulting buffer.

1 - in the “Geoprocessing” popup, search for the “buffer” tool again, but this time build a buffer of 1 km around all bikeways features

2 - name your “Output Feature Class”, “bikeways_Buffer”



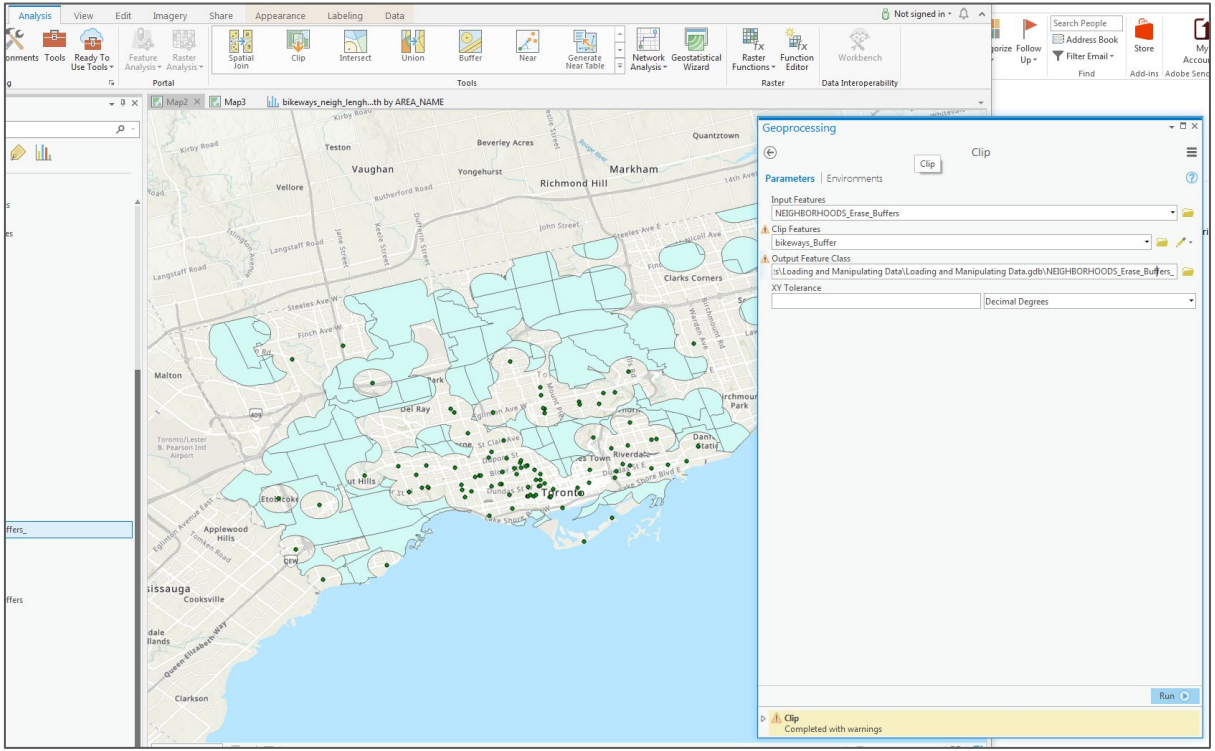
Clipping

Our new buffer represents areas that are of good potential because they are close to bikeways. But we need to combine these results with what we erased earlier so we can get closer precision to finding potential sites.

1 - search for the “Clip” tool

2 - clip out of the “NEIGHBOURHOOD_Erase_Buffers” layer, the new “bikeways_Buffer”

The results now show areas that do not have a bike shop within 1km but that are within 1km of a bikeway feature.

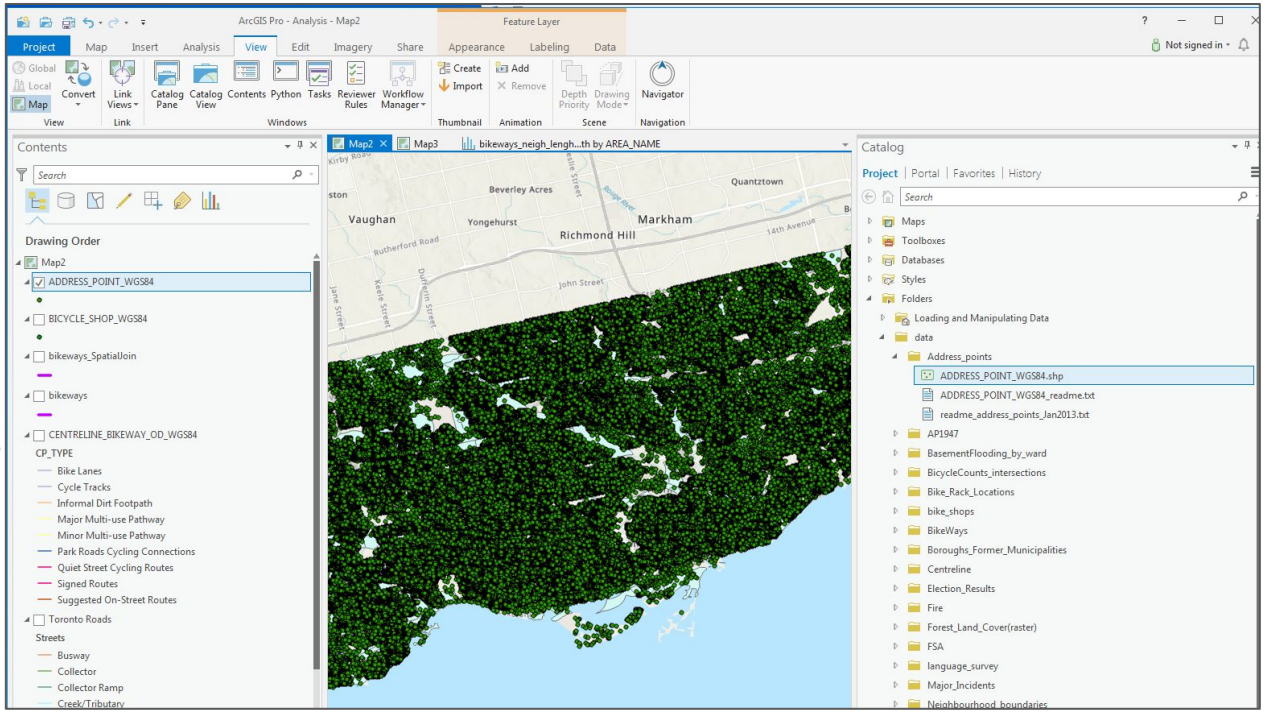


Select by Location or Intersect

1 - add the “ADDRESS_POINT_WGS84.shp” shapefile to your map

This layer represents all address locations in the city. The addresses contain “land use” categories* for each location. Let’s search for commercial locations that fall within our ideal areas identified.

*Note that as of the summer of 2021, this data has been “washed” out of the dataset



Select by location or Intersect

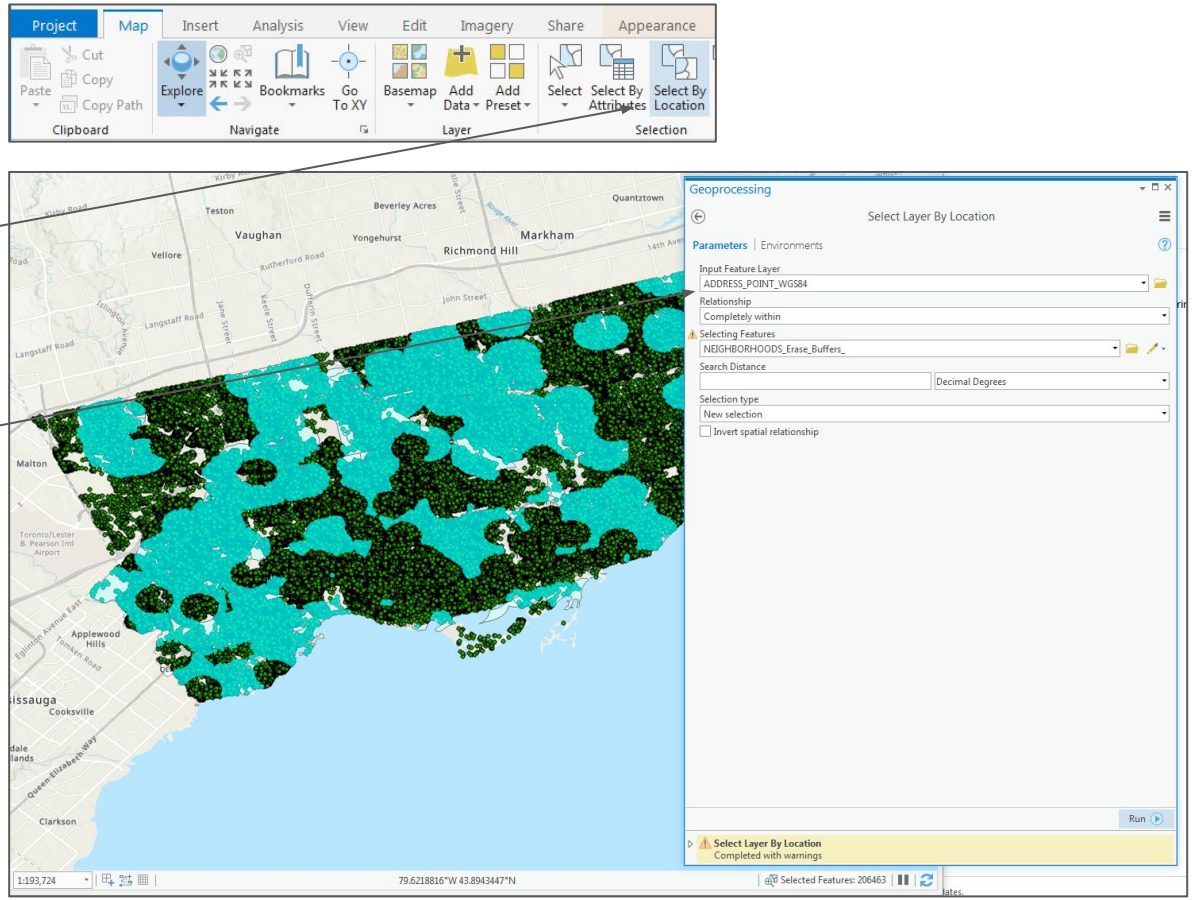
1 - under the “Map” menu, click “Select by Location”

2 - in the “Geoprocessing” popup, make sure your “Input Feature Layer” is set to “ADDRESS_POINT_WGS84”

3 - select “Completely within” under the “Relationship”

3 - under “Selecting Features” select “NEIGHBORHOODS_Erase_Buffers

This query will be looking for addresses that fall complete within our buffered, erased, and clipped layer “NEIGHBORHOODS_Erase_Buffers_”

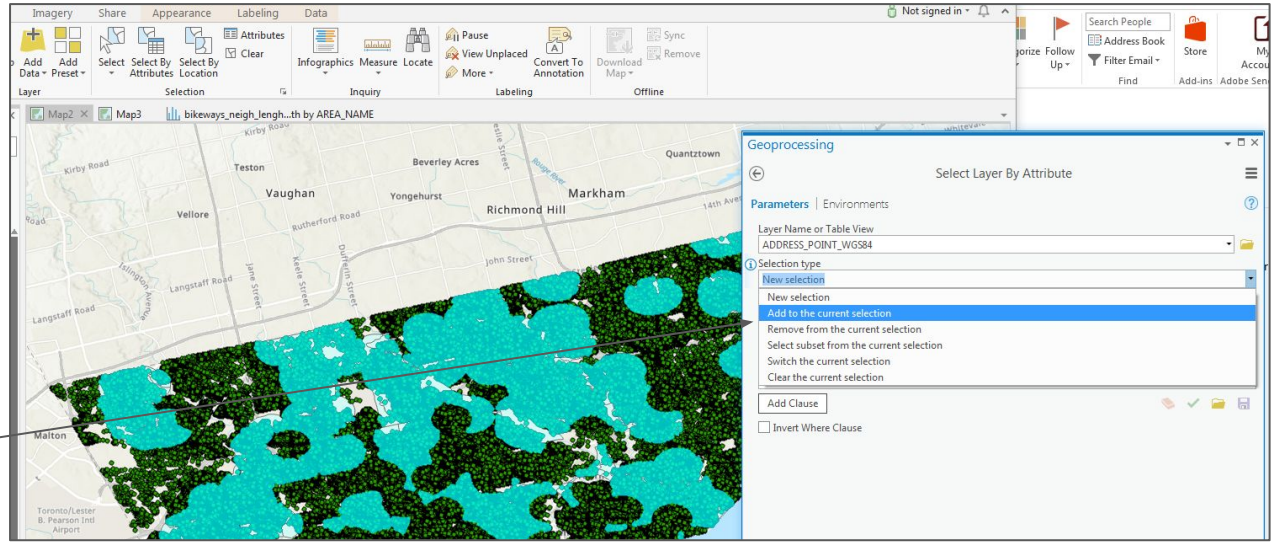


Select by Attribute

Now we want to find only “Commercial” locations within the addresses found within our ideal area.

1 - Under the “Map” menu, click on “Select by Attributes”

2 - in the “Geoprocessing” popup, make sure you are querying the “ADDRESS_POINT_WGS84” layer and that your “Selection Type” is “Add to the current selection”



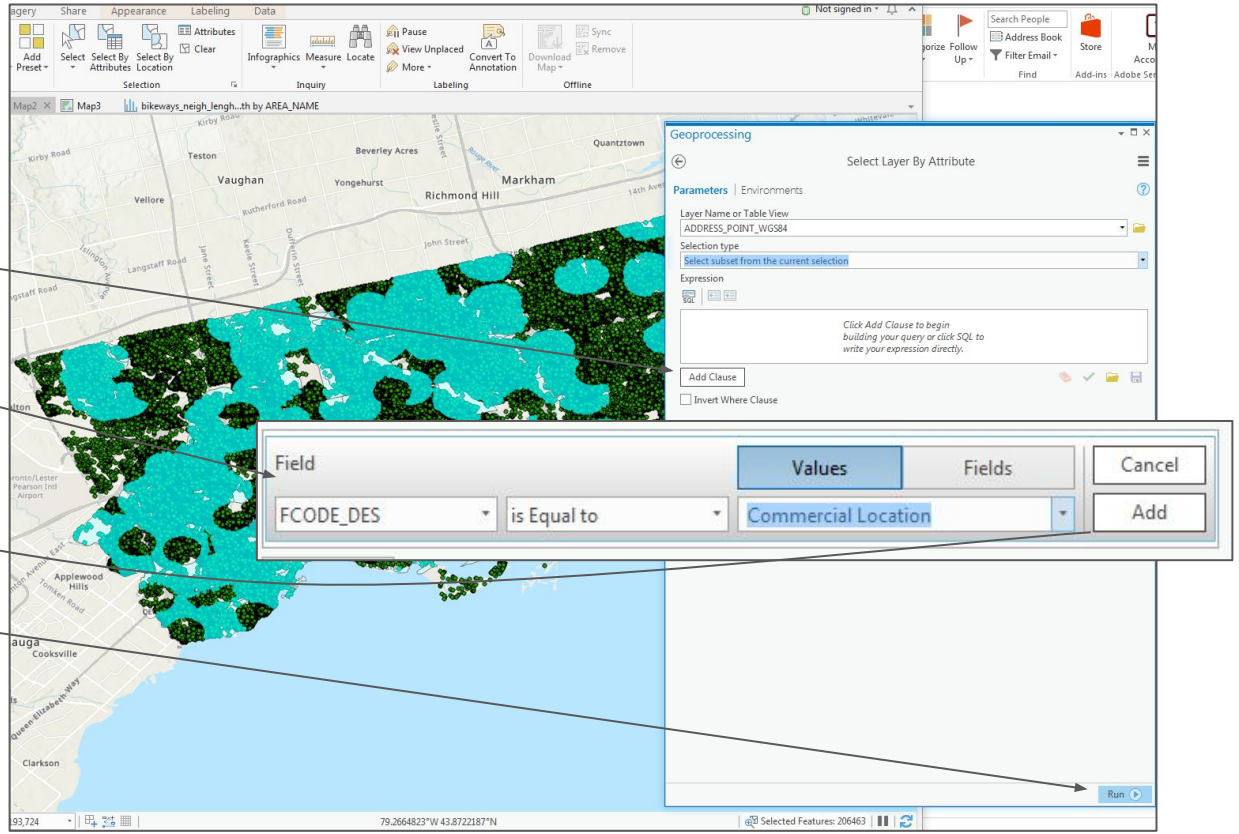
Select by attributes

1 - click on “Add Clause”

2 - in the resulting clause query tool, select the “Field” “FCODE_DES” “is Equal to” “Commercial Location”

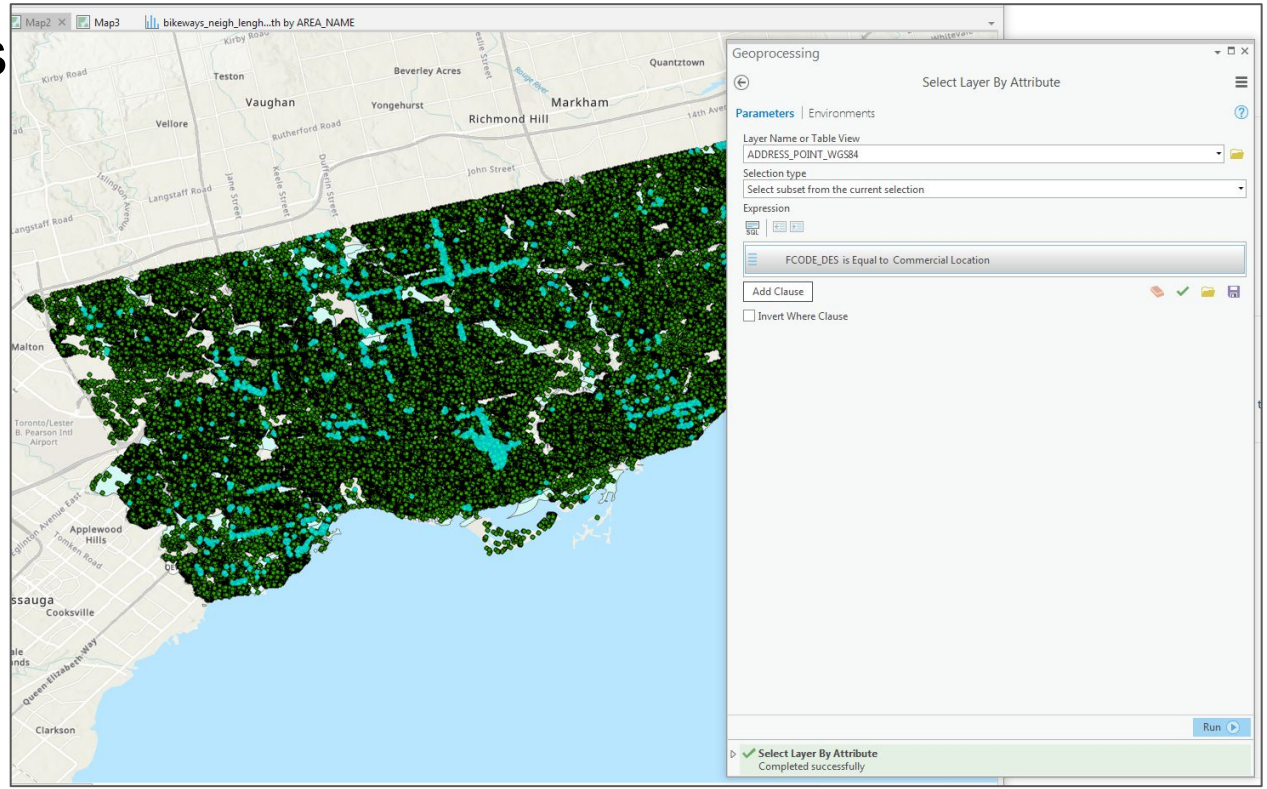
3 - click on “Add”

4 - click on “Run”



Select by Attributes

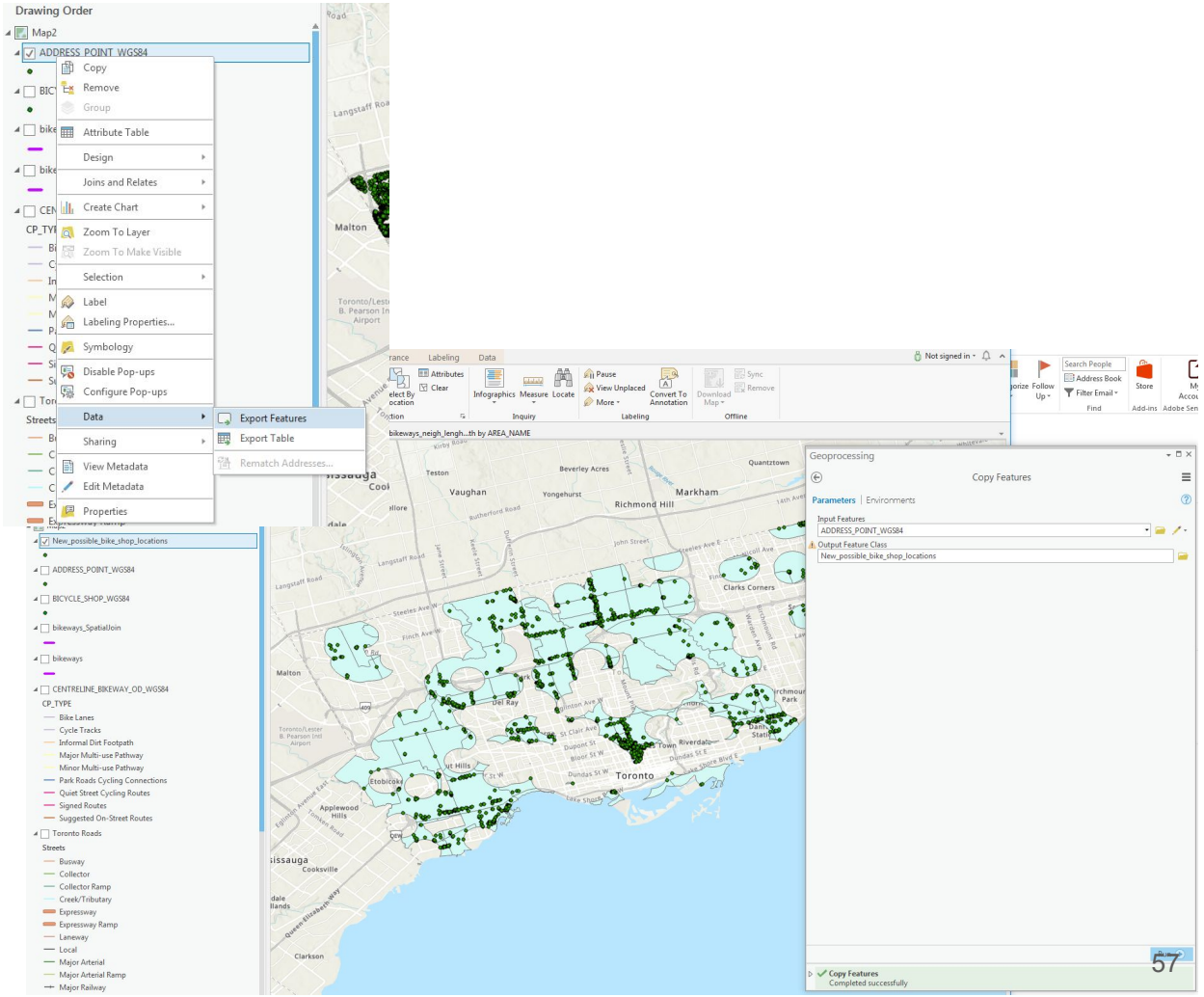
Let’s create a new layer from these results.



Site Selection

1 - right-click on the “ADDRESS_POINT_WGS84” layer, select “Data” → “Export Features”

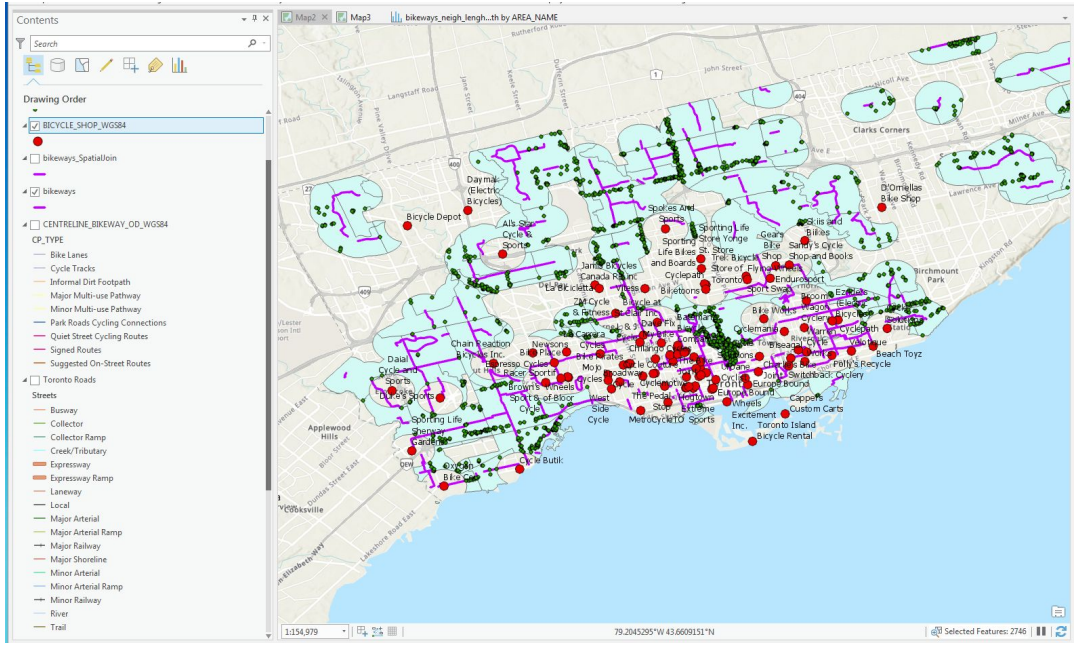
2 - provide a name for the resulting export feature



Site locations

1 - label your current bicycle shops by right-clicking the layer in the “Contents” pane and selecting “Label”

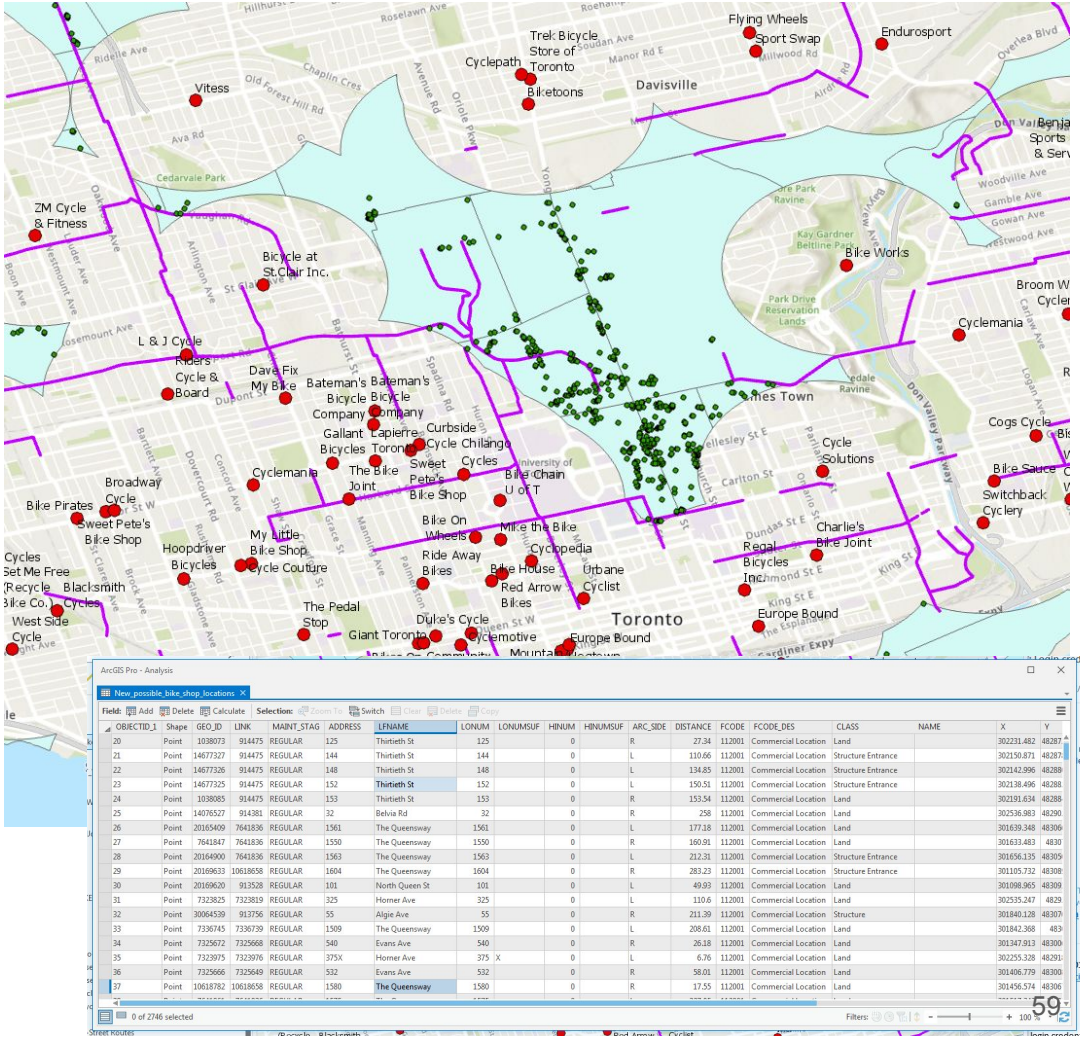
2 - zoom into your map



Site selection

Now it's just a matter of refining your searches by finding out which of these locations are for rent or for sale!

One could of course add new queries into the equation by looking at population numbers and income ranges, bike rack locations (available from the city) etc.



End of Module 4