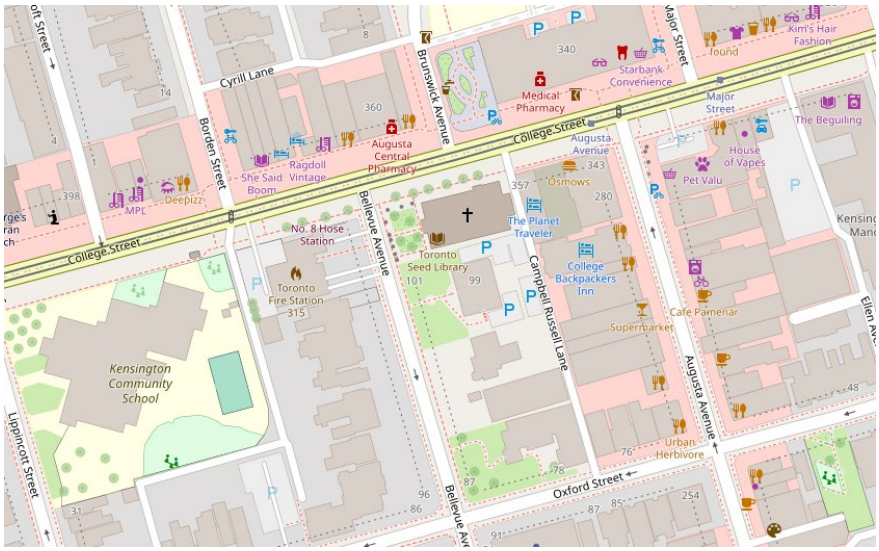
Lab 4: Introduction to Mapping with QGIS

Introduction to spatial data

To create maps using data, you'll need a spatial dataset. Spatial datasets combine attribute data (*what variables do you want to map?)* and location data (*where are the data geographically located?*). There are two main types of spatial data: **vector** and **raster**.

Vector data

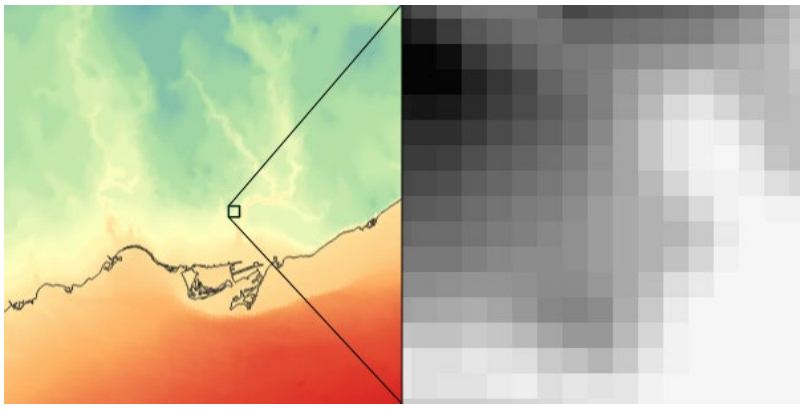
Vector data uses geographic coordinates to create **points**, **lines**, and **polygons** representing real-world features. For example, in the map below (a screenshot of OpenStreetMap) lines are used to represent roads and rail, points for retail, polygons for parks and buildings, etc.



Raster data

Raster data represents space as a continuous grid with equal cell sizes. Each cell contains a value pertaining to the type of feature it represents. These values can be quantitative (e.g. elevation) or categorical (e.g. type of land use). Common examples of raster data include digital elevation models (DEMs), satellite imagery, and scanned images (e.g. historical maps).

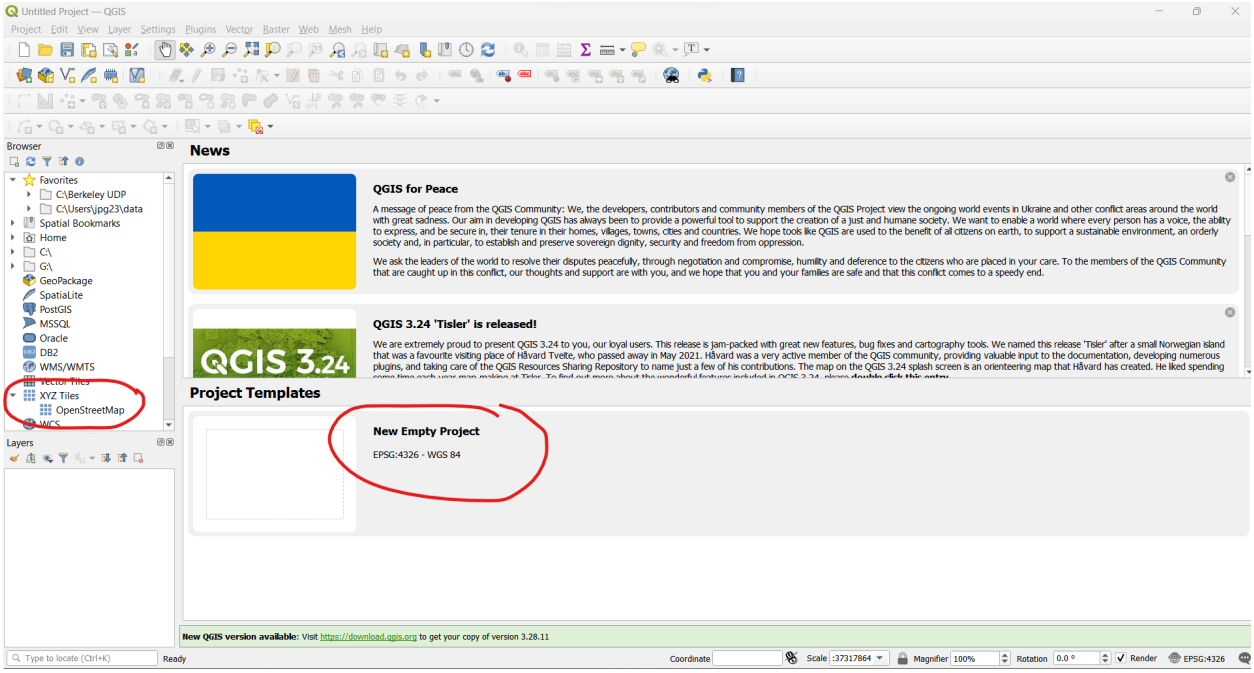
The map below shows a DEM for Toronto at two different scales.



Let's get started!

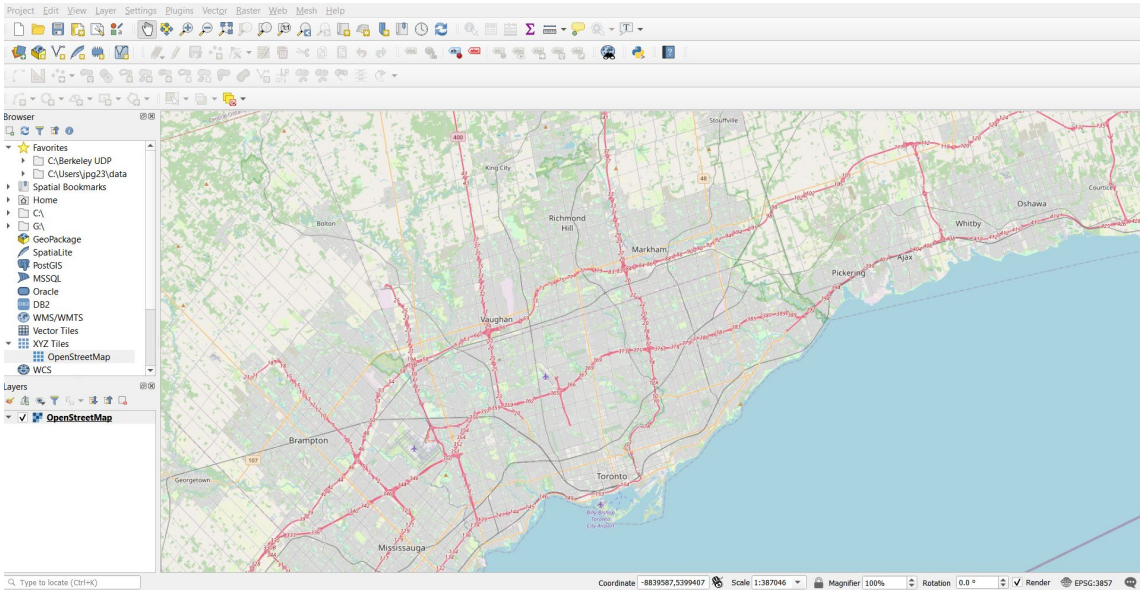
Open QGIS

1. [Download QGIS] and open it on your desktop. The "Browser" on the left allows for navigating and loading datasets, the "Layers" panel will populate with each dataset that is added to the project, and the big blank square is where your map data will be visualized.



2. Click on 'New Empty Project'. You should see a blank white rectangle – this is because you haven't selected a base map. A base map is a background layer that

shows geographic information like highways and rivers. Double-click on the "OpenStreetMap" tiles (see red circle above). In QGIS, OpenStreetMap is the default base map, but you can look up how to add other ones with different designs (more on that later). Zoom into Toronto so you can see the data you'll be mapping. Your screen should look something like this:



Download spatial census data and create new variable

In this lab, we'll be mapping **vector** data to answer the question: *"which parts of Toronto have the highest proportion of workers commuting via public transit?"* Let's use the 2021 census data you downloaded in Lab #2: `census21_data.csv`. This file contains only attribute data – it has the variables we are interested in mapping, but not the spatial information associated with each census tract.

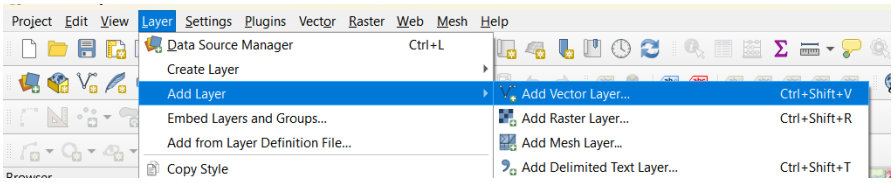
We also want to create a new variable, `transit_pc`, which equals the percentage of workers age 15+ who commute to work via public transit. Dividing the number of workers who commute via transit by the total number of workers helps to standardize the data—we wouldn't want to compare the raw number of workers across census tracts because tracts have different populations and land areas.

Here are two different ways we can get the spatial information for this data and create the new variable:

Option 1: Use cancensus package in R

One option is to download the shapefile version of the file you downloaded in Lab #2 directly using `cancensus`. A **shapefile** is a *geospatial file format that stores vector data*. The `cancensus` package gives you the option of downloading either only attribute data, *or* the attribute and location data together in a single shapefile. You can then create the new variable using the `tidyverse` package directly in R.

See `lab04/lab04_download_shapefile.r` for the code to create the shapefile (`census21_data.shp`) with the new variable. Add this shapefile as a layer to the map in QGIS. You can do this either by dragging the file to the 'Layers' panel or by going to 'Layer' in the toolbar → selecting 'Add Layer' → then 'Add Vector Layer…' → and navigating to where the file is saved on your computer.



Option 2: Download shapefile separately and use QGIS

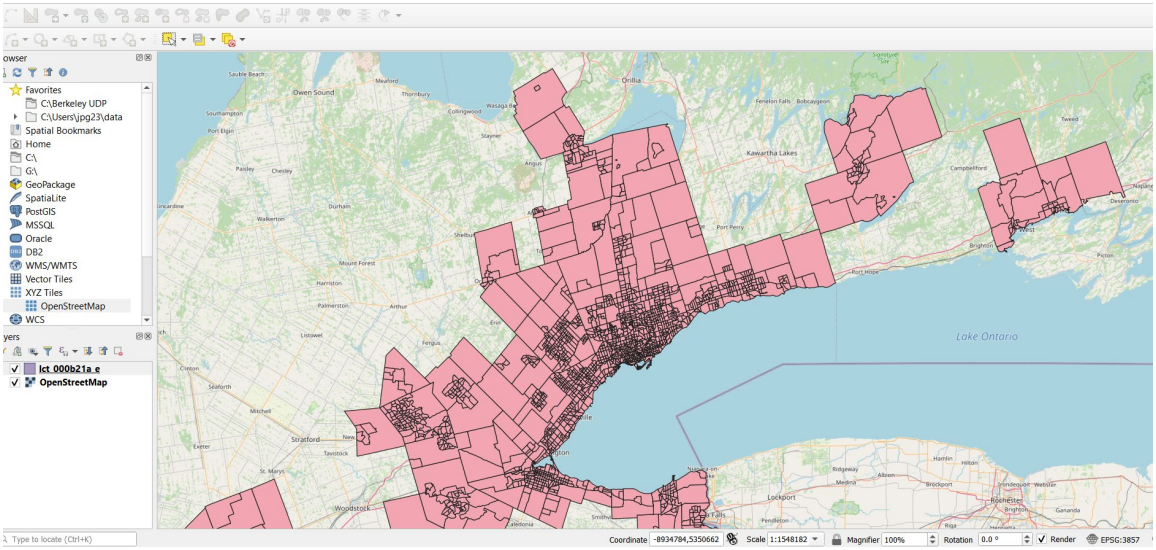
Another option, which is much more tedious but does not require using the `cancensus` API, is to download a shapefile of census tract geographic boundaries and "join" it with the attribute data:

1. The tract boundaries shapefile has already been downloaded for you – see `lab04/lab04_data_lct_000b21a_e.shp`. If you want to practice downloading it yourself, follow these steps:
    a. To download the census tract boundaries shapefile from 2021, go to the Statistics Canada website. This file contains all census tracts in Canada, not just Toronto. ***Keep in mind that census boundaries can change over time, so make sure to always join attribute data and boundary shapefiles from the same year.***
    b. Under the "*Statistical boundaries*" section, select "*Census tracts*". Scroll down to the "*Format*" section and select "*Shapefile (.shp)*". Download the zipped file.
3. In QGIS, add the tract boundaries shapefile to the map. Reminder: you can either do this by dragging the file to the 'Layers' panel or by going to 'Layer' in the toolbar →

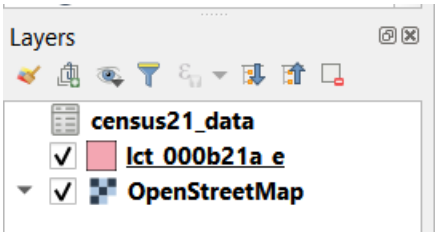
selecting ‘Add Layer’ → then ‘Add Vector Layer…’ → navigating to where the file is saved on your computer.

a. There will be a pop-up window asking you to select the coordinate reference system. *A Coordinate Reference System (CRS) is a schema for referencing where features are on the earth's surface. Each CRS has specified units (e.g. degrees, meters, etc.).* When working with urban data, we often want to have data in a CRS that does not distort local areas or distances, and has intuitive units (e.g. working with distances in meters rather than in degrees). You can select the default CRS in this case, but when working with multiple shapefiles (especially when doing something like a spatial join), make sure they all have the same CRS!

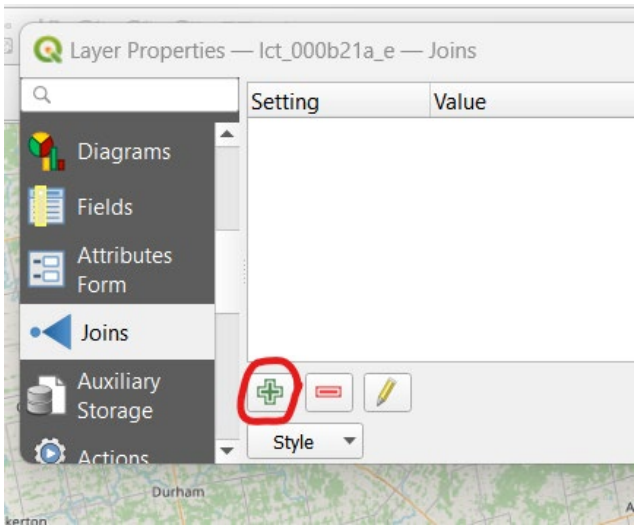
b. Your screen should now look something like this:



4. Add `census21_data.csv`, which you created in Lab #2, to the map by dragging the file to the “Layers” panel:

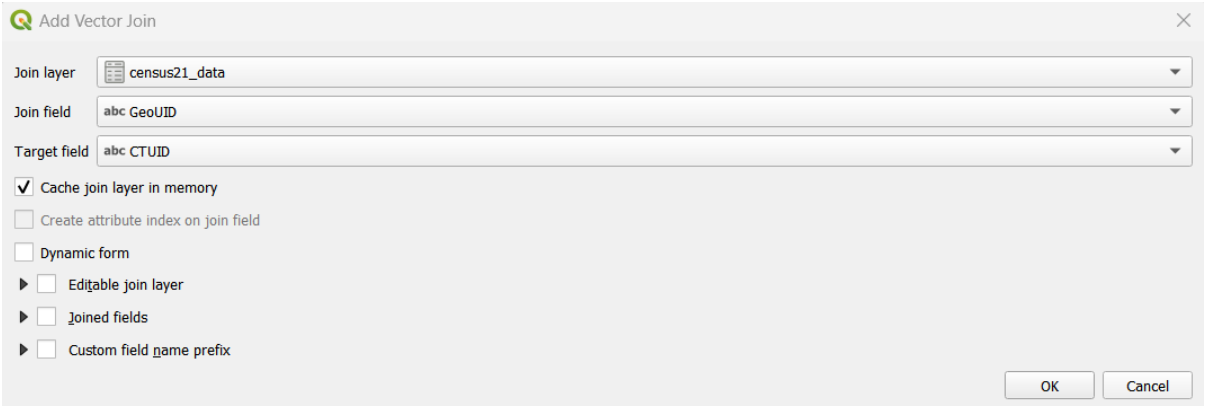


5. Next, join the census attribute data to the shapefile so that you have variables and spatial information all in one layer. Do this by right-clicking on the `lct_000b21a_e` layer, selecting ‘Properties’, and scrolling down to ‘Joins’. Click on the green plus sign (circled in red below):

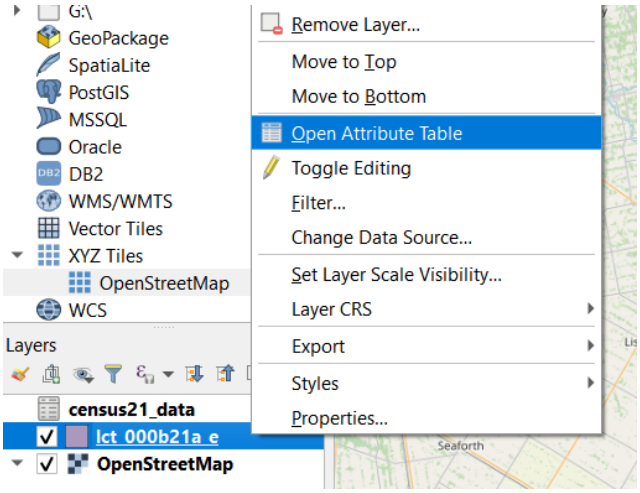


6. Select `census21_data` as the join layer, `GeoUID` as the join field, and `CTUID` as the target field. The join field is the census tract ID variable from `census21_data`, and the target field is the census tract ID variable from the shapefile.

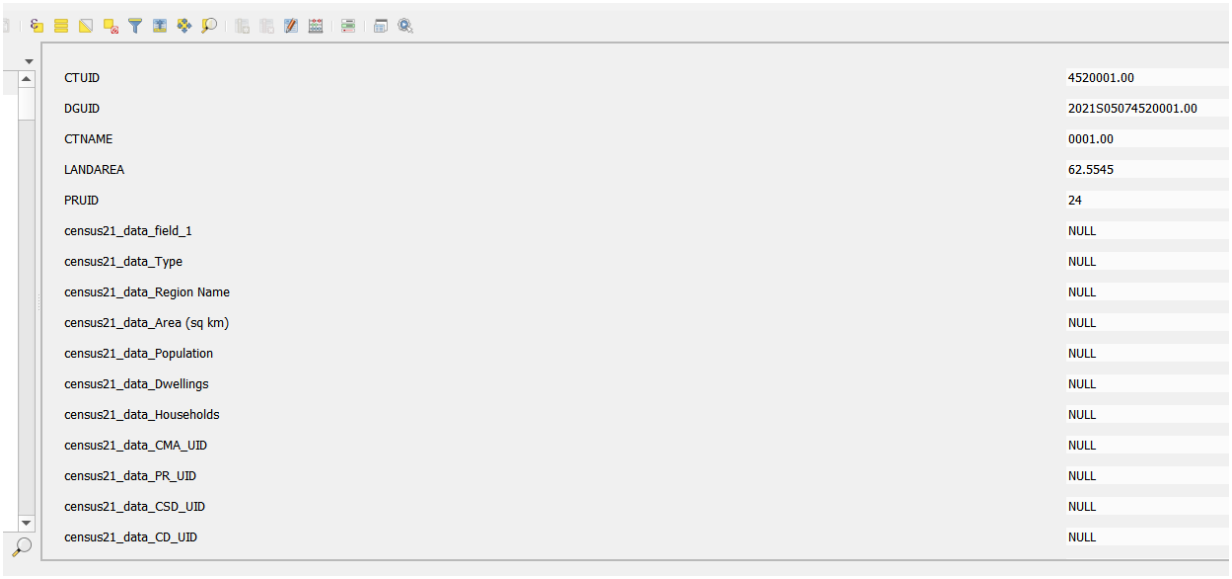
    a. When joining datasets, make sure that the variable(s) you're joining on are formatted the same way. For example, these two variables are both formatted as strings (the 'abc' symbol next to the field names indicates the field type). If one field was formatted as an integer and the other as a string, the join wouldn't work correctly.



7. Click 'OK', 'Apply', then 'Cancel'. To check that the join worked, make sure that the census variables show up in the shapefile's attribute table. You can open the attribute table by right-clicking on the layer and selecting 'Open Attribute Table'.

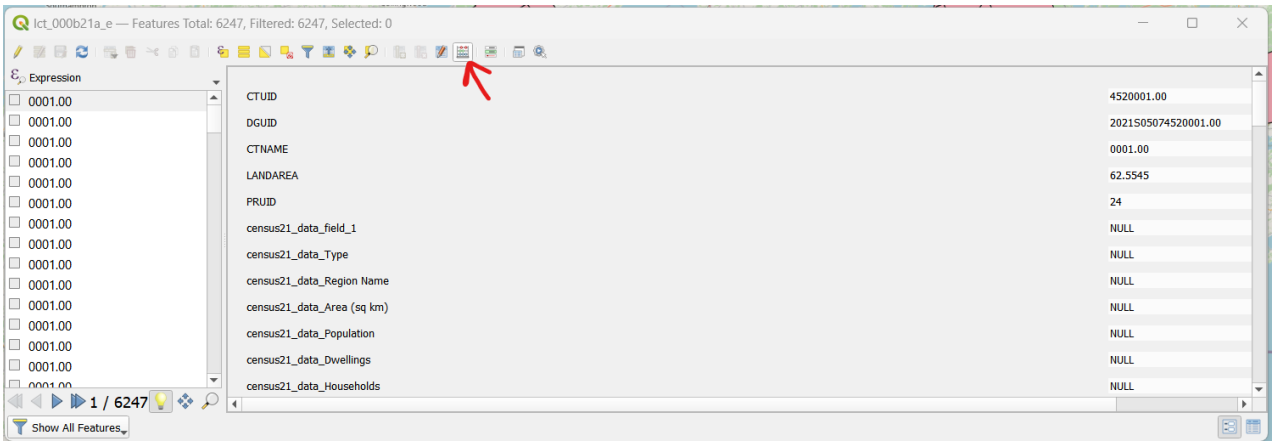


8. You should see a window that looks like this:



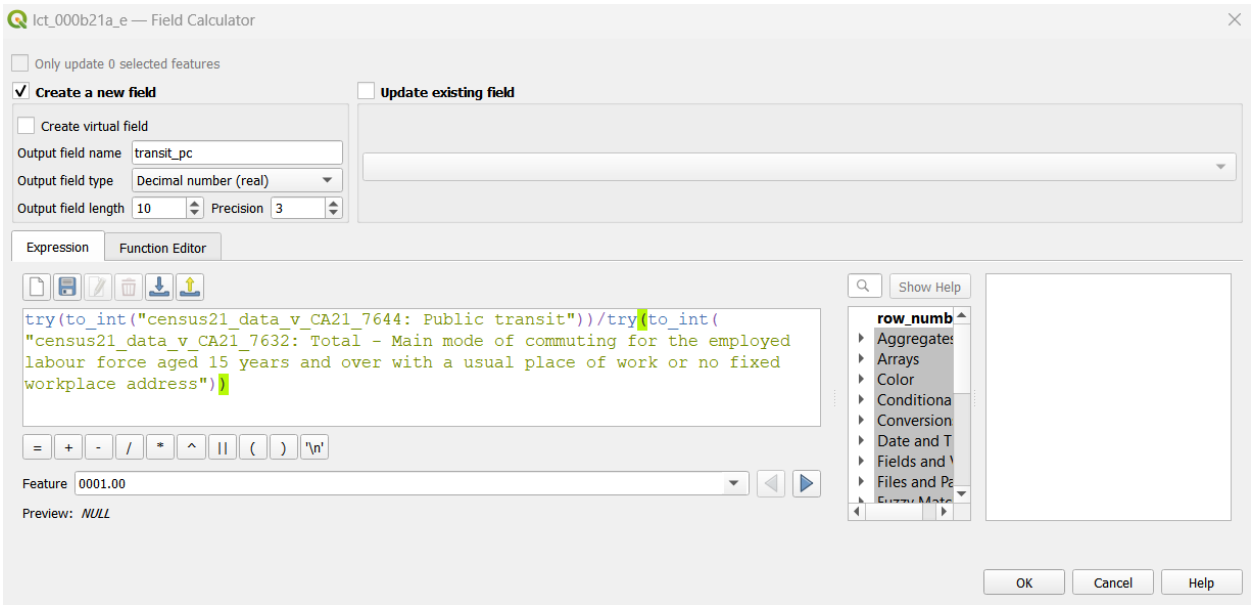
CTUID	4520001.00
DGUID	2021S05074520001.00
CTNAME	0001.00
LANDAREA	62.5545
PRUID	24
census21_data_field_1	NULL
census21_data_Type	NULL
census21_data_Region Name	NULL
census21_data_Area (sq km)	NULL
census21_data_Population	NULL
census21_data_Dwellings	NULL
census21_data_Households	NULL
census21_data_CMA_UID	NULL
census21_data_PR_UID	NULL
census21_data_CSD_UID	NULL
census21_data_CD_UID	NULL

9. Next, create the new variable we want to show on the map. Open the attribute table again and click on the ‘Open field calculator’ button (see red arrow below).



10. To create the new variable, make sure that 'Create a new field' is checked. Name the new field `transit_pc` and make sure it is a real number, not an integer, since it will include decimals. In the 'Expression' box, type what you see in the image below. Then hit 'OK' and exit the window.

    a. Since both the numerator and denominator are strings, we must use the `as_int()` function to convert them to integers before dividing them. Because QGIS doesn't recognize string 'NA' values as NULL and therefore has trouble converting these values to integers, the `try()` function is used to return the output of `to_int()` if it succeeds, and return NULL if it results in an error.

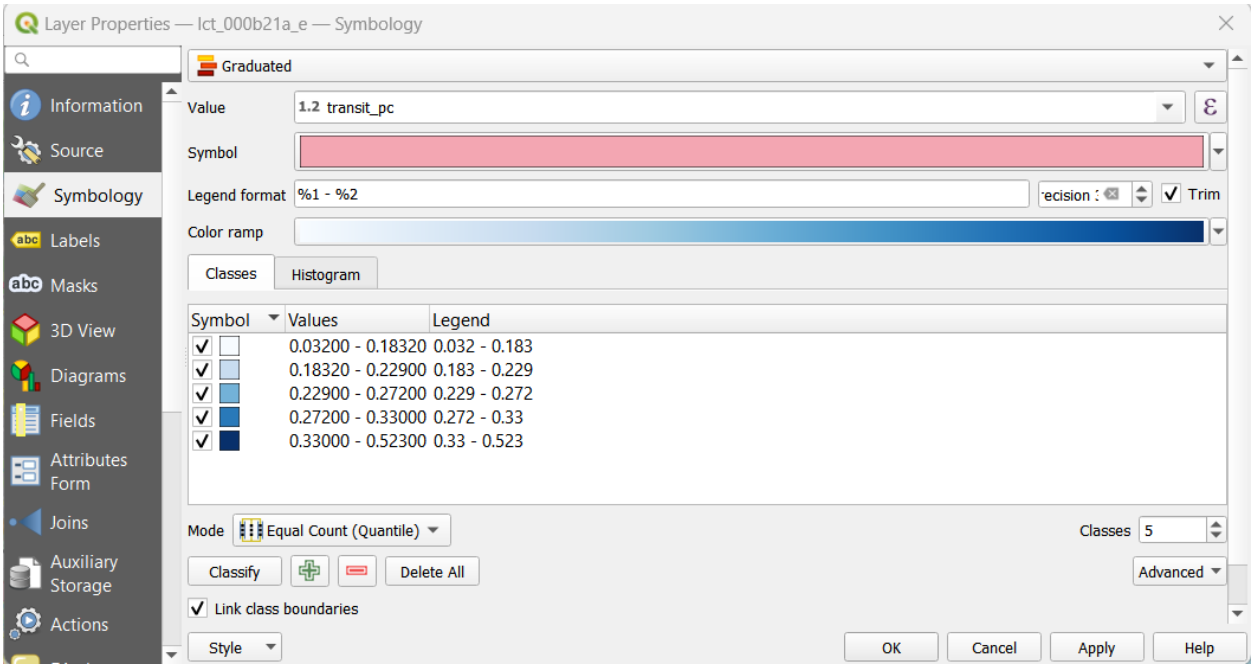


11. Click 'OK' and exit out of the window. Open the attribute table to make sure the new variable was created.

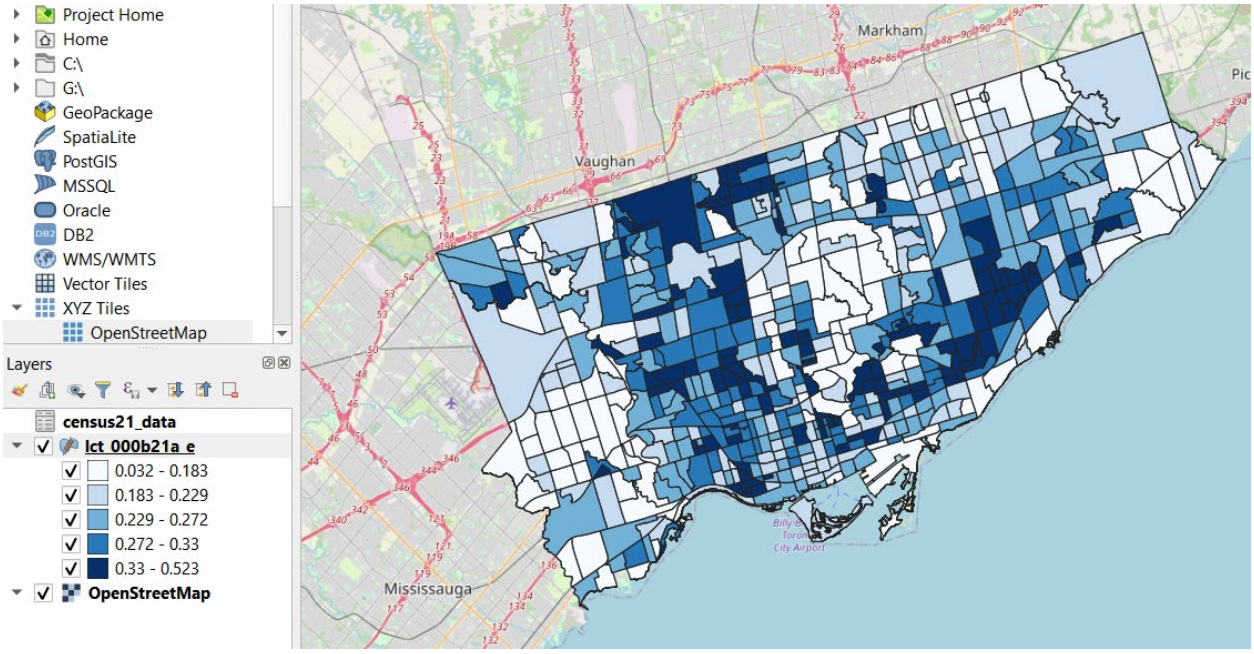
Create a choropleth map

Next, make a choropleth map that shows the percentage of the workforce in each census tract that commutes to work via public transit. A choropleth map is a type of map in which geographical areas are colored to correspond to attribute values.

1. In QGIS, start by right-clicking on the shapefile layer in the 'Layers' panel and selecting 'Properties'. Navigate to the 'Symbology' tab.
    a. In the top drop-down bar, select 'Graduated'.
    b. In the 'Value' field, select 'transit_pc' – this is the variable whose values we want colored on the map.
    c. Customize the legend format and color palette.
    d. Click 'Classify' on the bottom left and the numbers shown below should show up. Then hit 'Apply' and 'OK'.

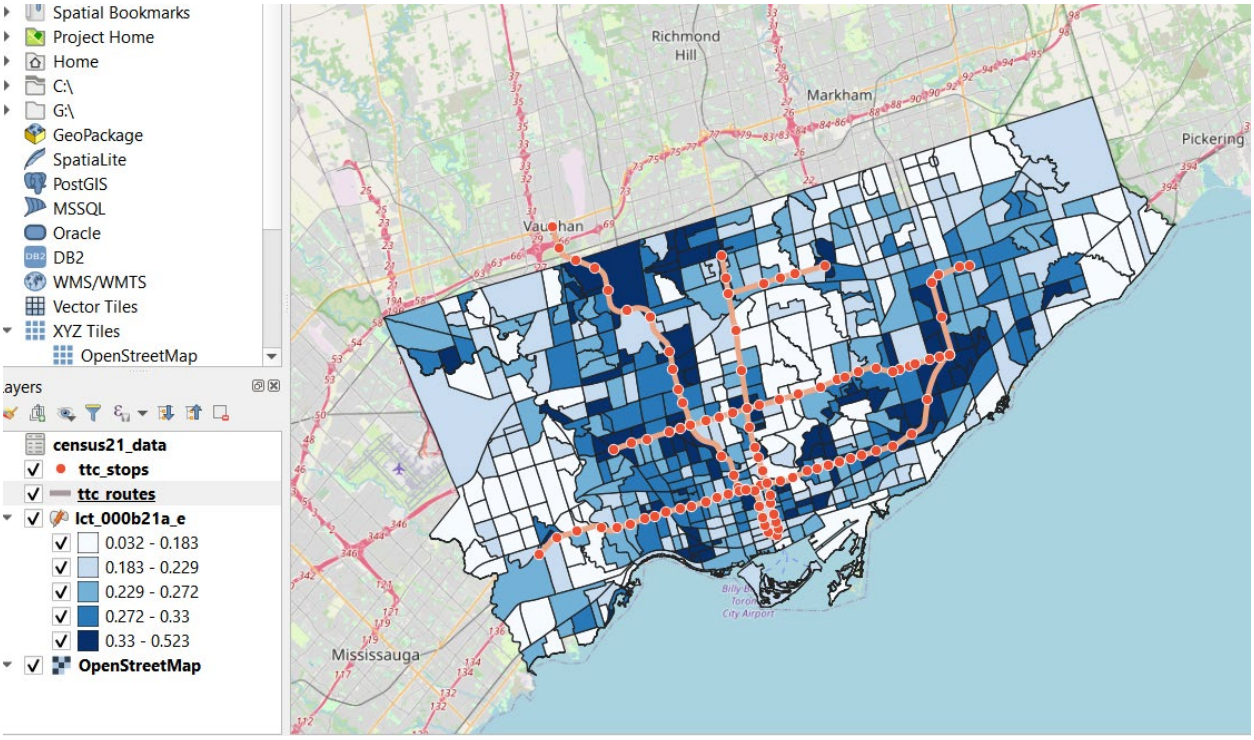


2. Your screen should look like the image below. In this map, darker blue means there is a higher proportion of workers in the census tract who commute via public transit.



Add point and line layers to map

Next, add public transit routes (`ttc_routes.geojson`) and stops (`ttc_stops.geojson`) to the map, and adjust the symbology (colors, size, etc.) so they're visible against the choropleth map in the background. Unsurprisingly, there appears to be a spatial relationship between the location of transit stops/routes and the proportion of workers commuting via transit:



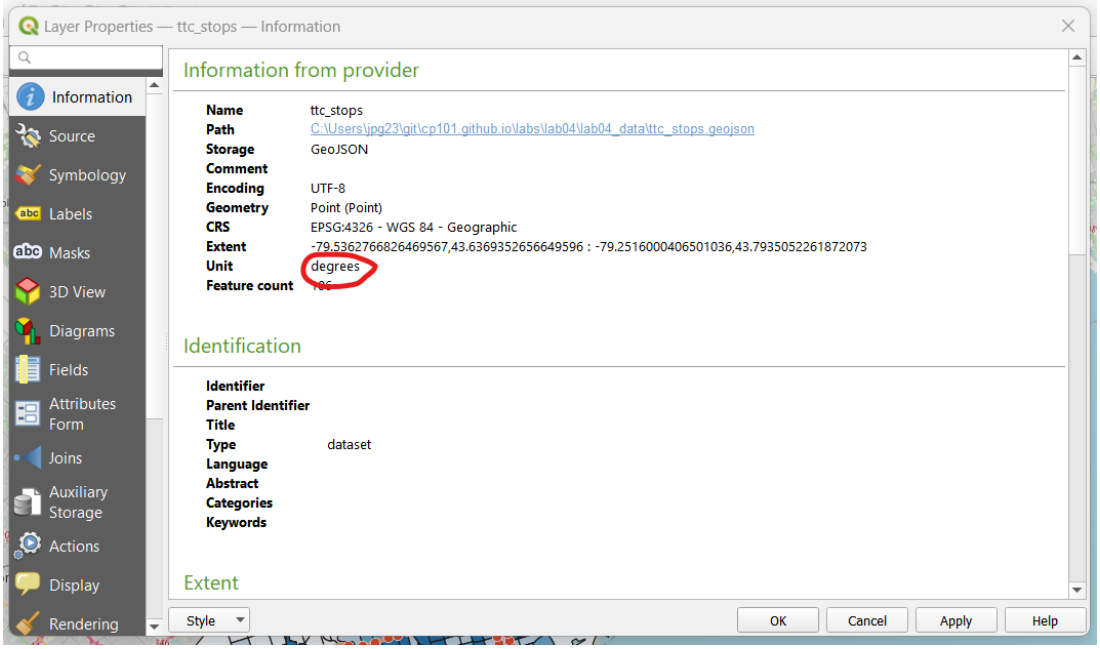
Create buffers around transit stops

Let's see whether census tracts that overlap with a 'transit walkshed' (a radius of about half a mile around a transit stop) appear[1] to have higher proportions of transit commuters. To do this, let's create a buffer around each transit stop.

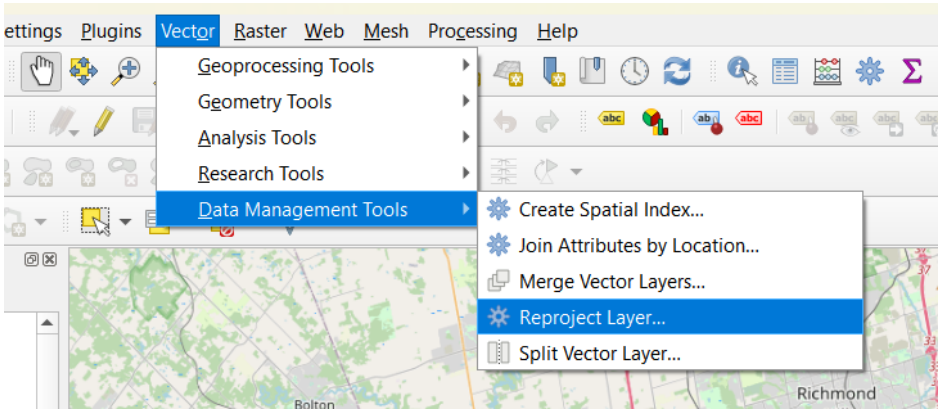
Reproject data to a different Coordinate Reference System

1. First, let's check the Coordinate Reference System of the `ttc_stops` layer we want to create the buffer for. As a reminder, a **Coordinate Reference System (CRS)** is a schema for referencing where features are on the earth's surface. Each CRS has specified units (e.g. degrees, meters, etc.). To check the CRS and units in QGIS, right-click on the `ttc_stops` layer, select 'Properties' → go to the 'Information' tab:

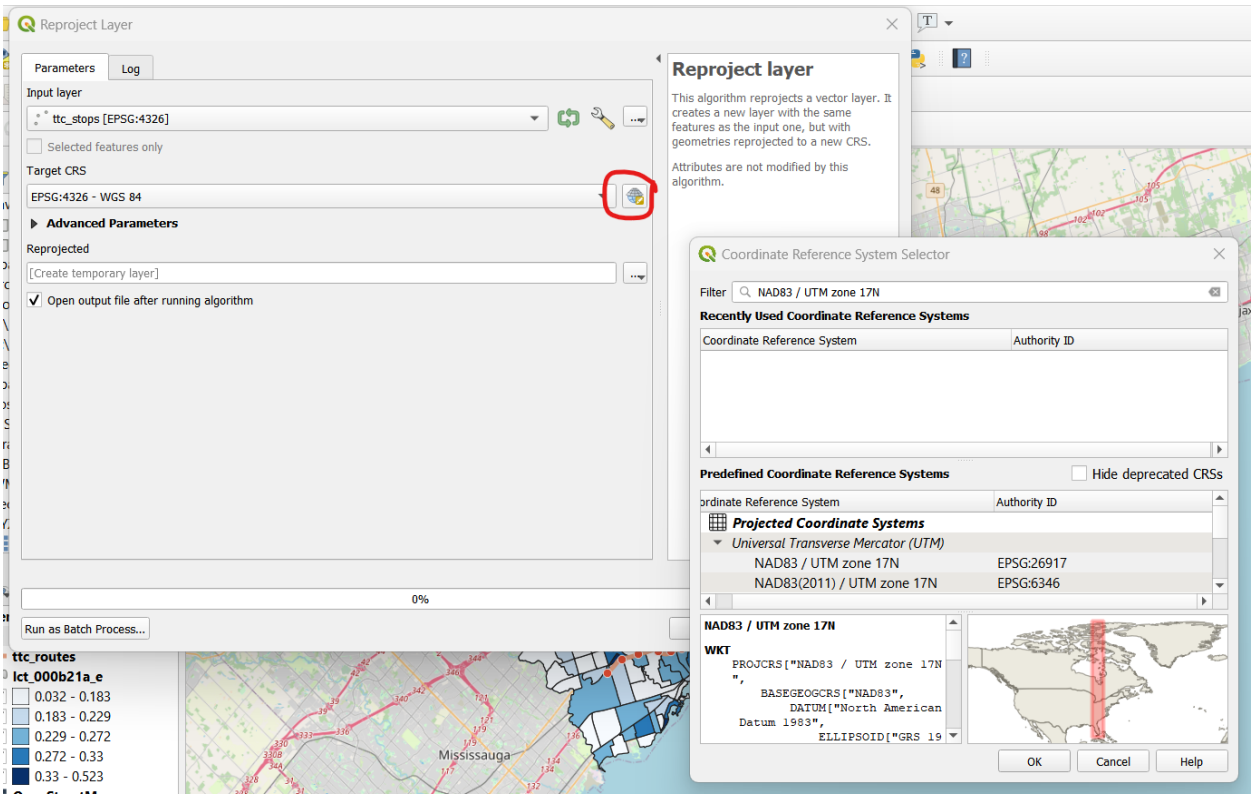
[1] I say 'appear' because we are not doing any statistical analysis to actually evaluate this relationship.



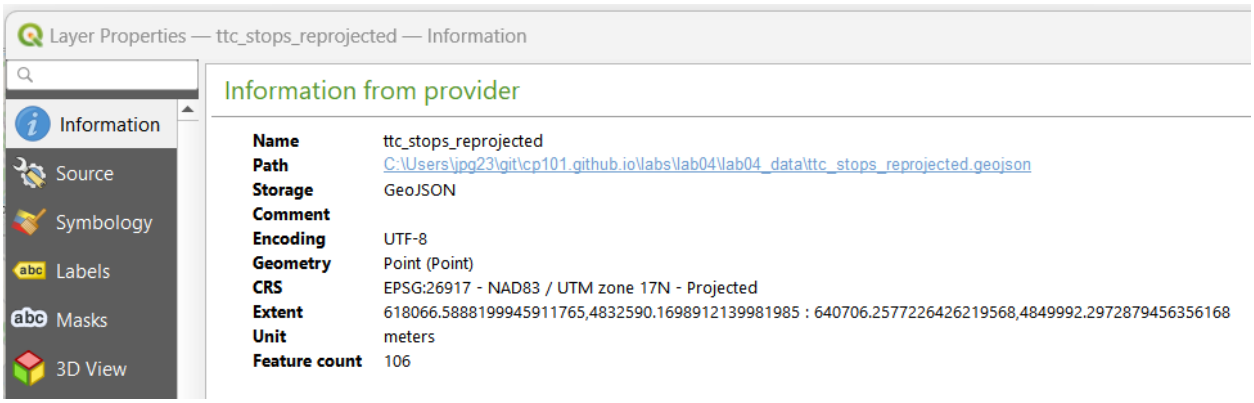
2. The CRS, `EPSG:4326 - WGS 84 - Geographic`, has units in degrees. Since it's more intuitive to work with meters than degrees, let's **reproject** the data to a different CRS whose units are in meters. You'll typically have to do some research on your own to figure out which CRS makes sense for the geographic area you're working with, but for now, let's use `NAD83 / UTM zone 17N - EPSG:26917`, a CRS that's commonly used for Toronto with units in meters. Click on 'Vector' in the toolbar → select 'Data Management Tools' → 'Reproject Layer':



3. Select `ttc_stops` (see CRS listed next to layer name) from the 'Input layer' drop-down menu. Next, click on the button to the right of the 'Target CRS' drop-down and search for the new CRS in the search bar. Select the one that says `EPSG:26917`, click 'OK', and in the 'Reproject Layer' box under 'Reprojected' (where it currently says '[Create temporary layer]') use the '[…]' button to select where on your computer to save the newly projected layer *as a new file*. Save the file with a name like `ttc_stops_reprojected`. Hit 'Run', then close the window.

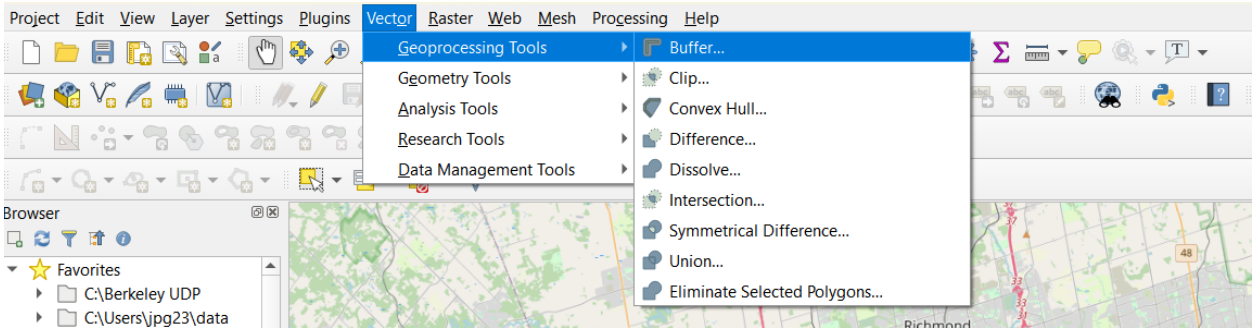


4. Now if you look at 'Information' for the new layer, it should show the new CRS:

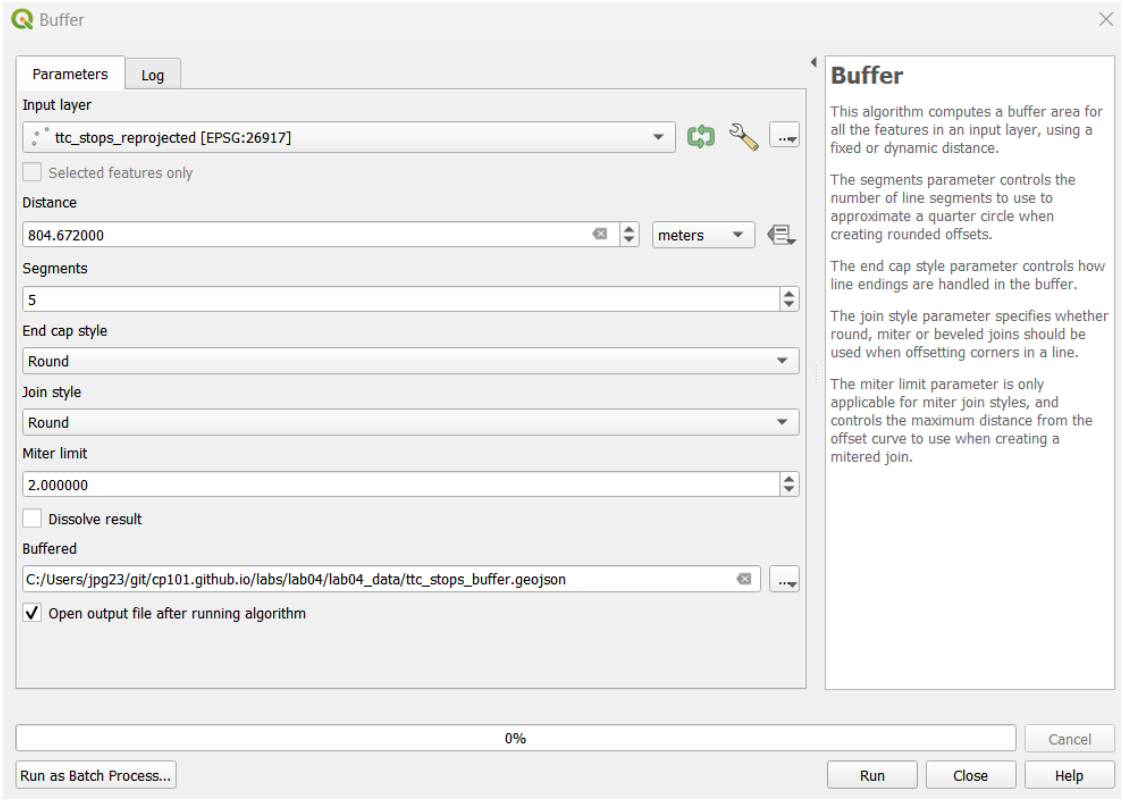


Create the buffer layer

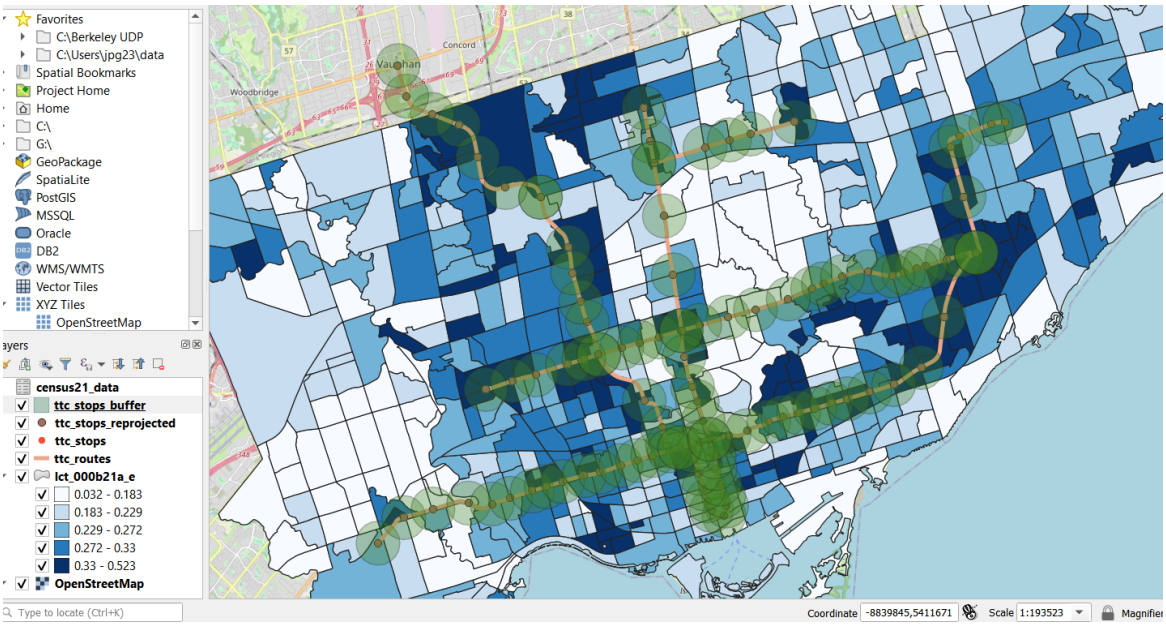
1. Click on 'Vector' in the toolbar at the top → 'Geoprocessing Tools' → 'Buffer'.



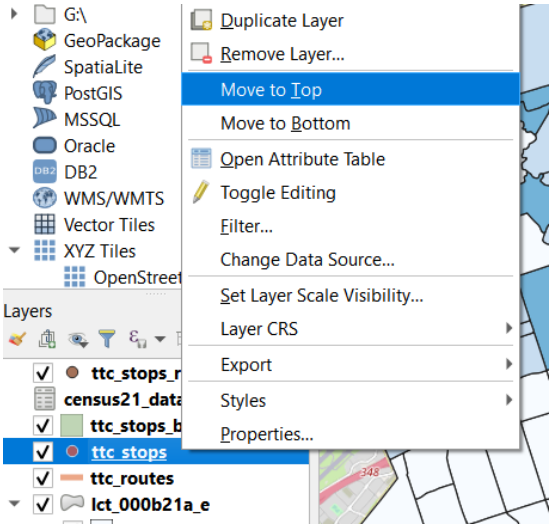
2. Select `ttc_stops_reprojected` as the input layer, set the distance to 804.672 (make sure the unit is meters; this is equivalent to half a mile), select a location and name for the new file, click 'Run' and then exit the window.



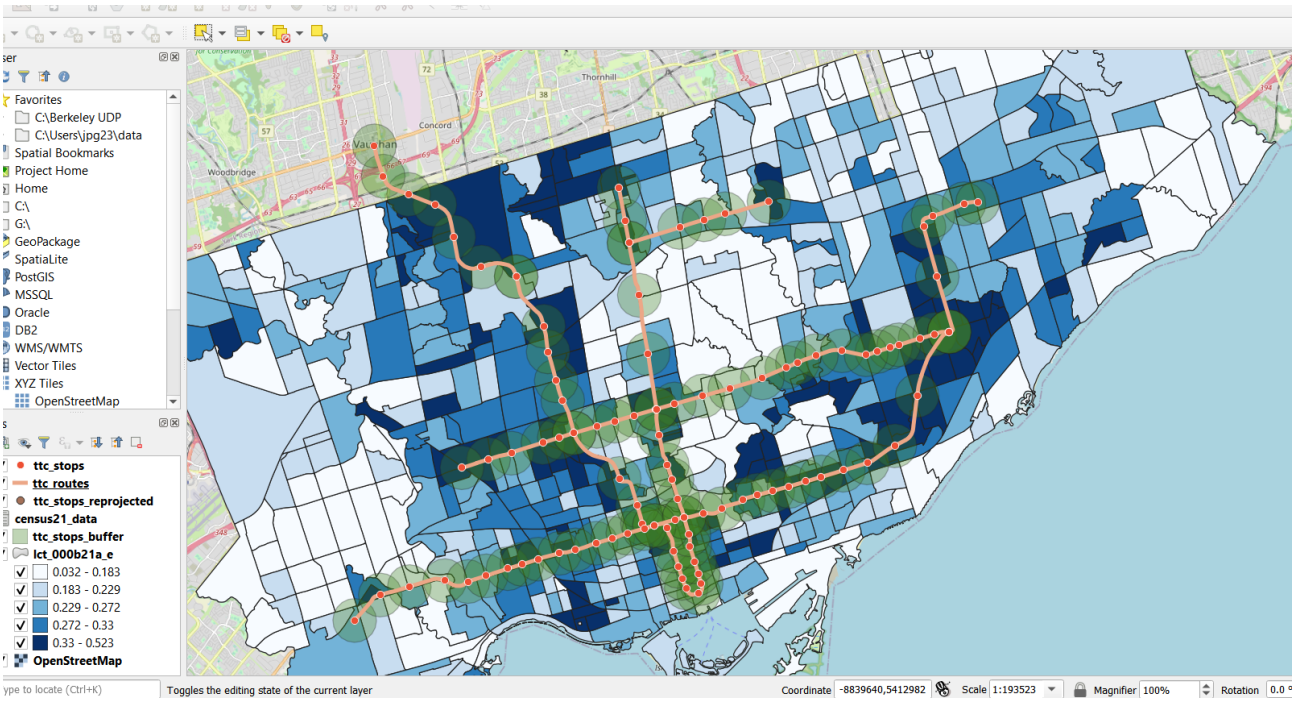
3. The new layer should show up on the map. Lower the opacity and change the color so you can see the layers beneath the circles:



4. It's still difficult to see the transit stops and routes underneath the buffer circles. To fix this, click on the `ttc_stops` and `ttc_routes` layers and select 'Move to Top'.



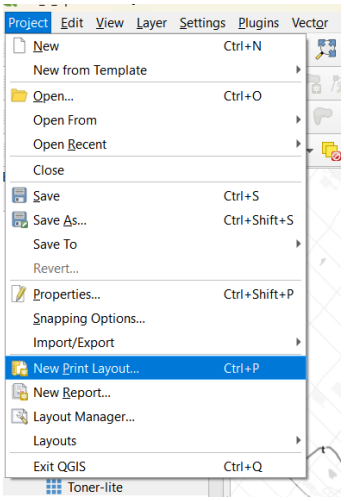
5. Now your map should look something like this:



Format the map

Finally, it's time to make the map look good!

1. Click on 'Project' in the toolbar, then 'New Print Layout':



2. Here, you can add a static image of your map ('Add Item' → 'Add Map') and then customize it to make it look publishable. See the QGIS documentation for more information. In the map I created below, I added a legend, title, scale bar, and north arrow. I also changed the base map to make it look cleaner (see details here – this is

done to the underlying map, not in the ‘New Print Layout’ window). When you make updates to the underlying map, you can refresh the static image in this window so that you don’t have to keep opening a new window each time. You can also export the final image as a PDF.

Voila!