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 is the data geographically located?*). There are two main types of spatial data: **vector** and **raster**.

## Vector data

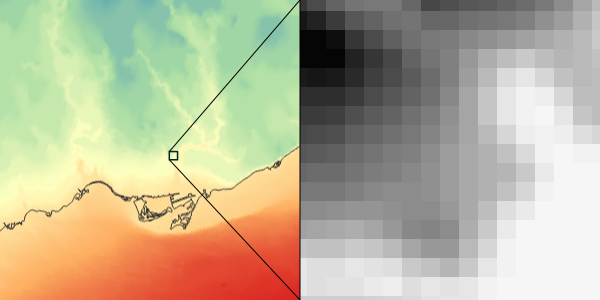
Vector data uses geographic coordinates to create **points**, **lines**, and **polygons** representing real-world features. For example, in the map below (a screenshot of [OpenStreetMap](https://www.openstreetmap.org/)) lines are used to represent roads and rail, points for retail, polygons for parks 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. Open QGIS from 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.

A screenshot of a computer

Description automatically generated

1. 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 only default base map, but you can look up how to add other ones with different designs. Zoom into Toronto so you can see the data you’ll be mapping. Your screen should look something like this:

A map of a city

Description automatically generated

# Download spatial census data and create new variable

In this lab, we’ll be mapping **vector** data. Let’s use the 2021 census data you downloaded in Lab #2: census21\_data.csv. Since this is only attribute data, it contains 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 percent of employees age 15+ who commuted to work via public transit. This is the variable we are going to include in our map.

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 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). Add this shapefile as a layer to the map in QGIS. 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 screenshot of a computer

Description automatically generated

## 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:
   1. To download the census tract boundaries shapefile from 2021, go to the [Statistics Canada website](https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21). 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.***
   2. Under the “*Statistical boundaries*” section, select “*Census tracts*”. Scroll down to the “*Format*” section and select “*Shapefile (.shp)*”. Download the zipped file.
2. 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.
   1. There will be a pop-up window asking you to select the coordinate reference system. *A*[*Coordinate Reference System (CRS)*](https://en.wikipedia.org/wiki/Spatial_reference_system)*is a schema for referencing where features are on the earth's surface. Each CRS has specified units (e.g. degrees, metres, 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 metres 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!
   2. Your screen should now look something like this:

A map of a city

Description automatically generated

1. Add census21\_data.csv, which you created in Lab #2, to the map by dragging the file to the “Layers” panel:

A screenshot of a computer

Description automatically generated

1. 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):

A screenshot of a computer

Description automatically generated

1. 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.
   1. When joining datasets, make sure that the variable(s) you’re joining on are formatted the same. 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.

A white wall with three white lines

Description automatically generated with medium confidence

1. 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’.

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

1. 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).

A screenshot of a computer

Description automatically generated

1. 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.
   1. 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.

A screenshot of a computer

Description automatically generated

1. Click ‘OK’ and exit out of the window. Open the attribute table to make sure the new variable was created.

# Create a choropleth map

Now it’s time to 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. Start by right-clicking on the shapefile layer in the ‘Layers’ panel and selecting ‘Properties’.
   1. In the top drop-down bar, select ‘Graduated’.
   2. Customize the legend format and color palette.
   3. Click ‘Classify’ on the bottom left and the numbers shown below should show up. Then hit ‘Apply’ and ‘OK’.

A screenshot of a computer

Description automatically generated

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

A map of a city

Description automatically generated

# Add point and line layers to map

[transit stops and lines]

# Add legend and scale bar

[FROM [JEFF’S TUTORIAL](https://schoolofcities.github.io/mapping-workshops-2023/mapping-census-data): The legend and scale bar were added by going to Project, selecting New Print Layout, and then adding in a map view, legend, and scale bar to the blank page.]

# Other resources