02.221 – Lab 2: Creating and using spatial data

The data needed for this lab can be found on the course website or downloaded from Dropbox: https://dl.dropboxusercontent.com/u/4837647/02221-Lab2.zip. Extract the data from the zip file to an appropriate location within your documents.

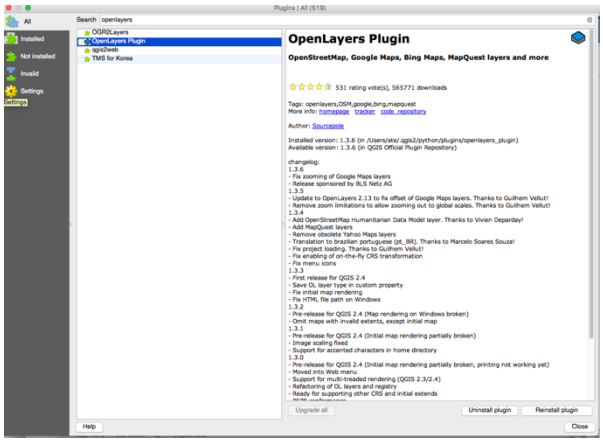
Goals

The primary goal of this lab to learn about the different ways to create, load and save spatial data within a GIS environment. In doing so, you will also learn how to use QGIS' plugin system and create your first thematic map of population per planning subzone in 2015.

Creating spatial data from scratch

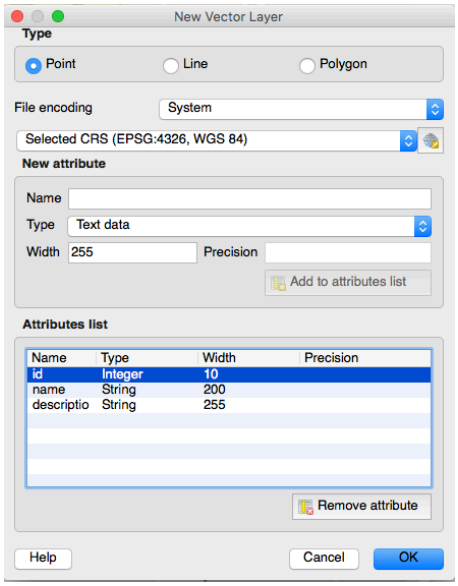
Let's assume you want to create a map that shows your five favorite places to eat in Singapore. To do so, you would need to create a completely new layer of spatial information so that you can add a data point for each restaurant – analogous to creating a new Word or Excel document.

Luckily, this is easy to do in QGIS. Open the Singapore base map project that you worked on in the last lab. First of all, despite having all kinds of information already present, depending on your map design, it might be difficult to pinpoint the location of your restaurants. To aid you in finding them on the map, we can add an online map service as reference layer. QGIS has a plug-in structure that allows you to extend the functionality of the base program by installing additional plugins. Go to Plugins | Manage and Install Plugins. You should see a long list of possible plugins now. Feel free to browse around a bit and then search for the Openlayers plugin and install it.



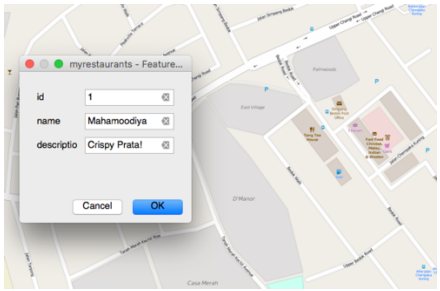
Once installed, you can navigate to Web | Openlayers Plugin and pick any of the map layers available there. If you don't have a preference, go for Bing Road or OpenStreetMap.

To subsequently create a new layer to hold our restaurants, go to Layer | Create Layer | New Shapefile Layer. This will create a new, empty file and add it to our project as a Layer. Apart from the points themselves, we also want to make sure we give each point additional attributes, consisting of at least a name and a short description.



After you have defined your attributes, hit OK, give your file an appropriate name and hit Save. We now have an empty layer, ready to receive our new points. To start adding points, we need to enter 'Edit mode'. In the edit toolbar, click the little pencil  to 'Toggle Editing' and subsequently click  to start adding features.

Click on the location of your first restaurant. A pop-up window will allow you to enter the attributes for that feature.

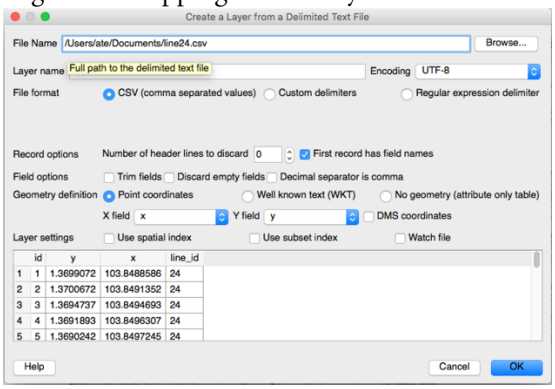


Finish the other four restaurants in the same way. Once you are finished entering restaurants, hit  'Save Layer Edits' and exit edit mode by clicking on  'Toggle Editing' again. This will save all your edits to the underlying file. Hang on to the file for now, we will use it later in this lab assignment.

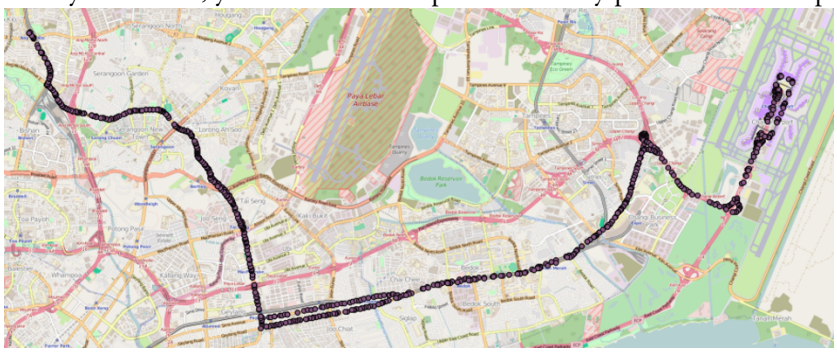
Converting one type of spatial data to another

Often the data we need for our projects is not yet in the format we ultimately need it to be in. For example, information from GPS traces (think bus trackers, cell phone traces, social media data) is often stored in text files. Open the file line24.csv (it's in the lab assignment data zip file) in Microsoft Excel to inspect its contents. Based on the contents of the file, what do you think the data represents?

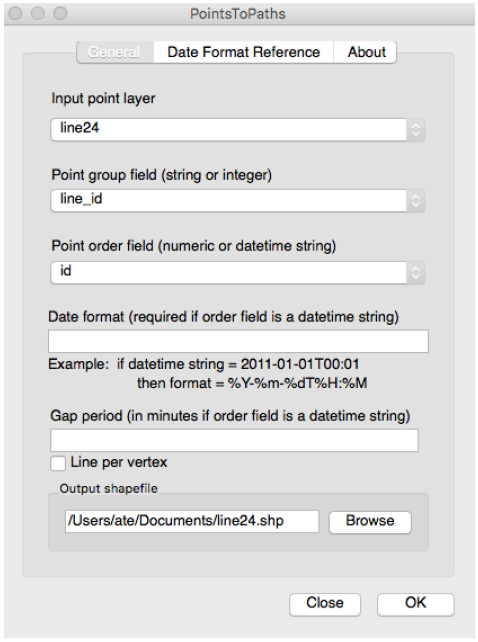
Although the file is just a flat table, two of the columns do contain geographic information. With this info, we can instruct QGIS to make our data explicitly spatial and plot each point on the map. Go to Layer | Add Layer | Add Delimited Text Layer. Select the csv file and pay attention to the Geometry definition section of the window. You will notice QGIS is pretty smart in discovering our spatial columns! This is not always the case so you do have to pay close attention to this part. You can imagine that flipping the x and y column will create intelligible results.



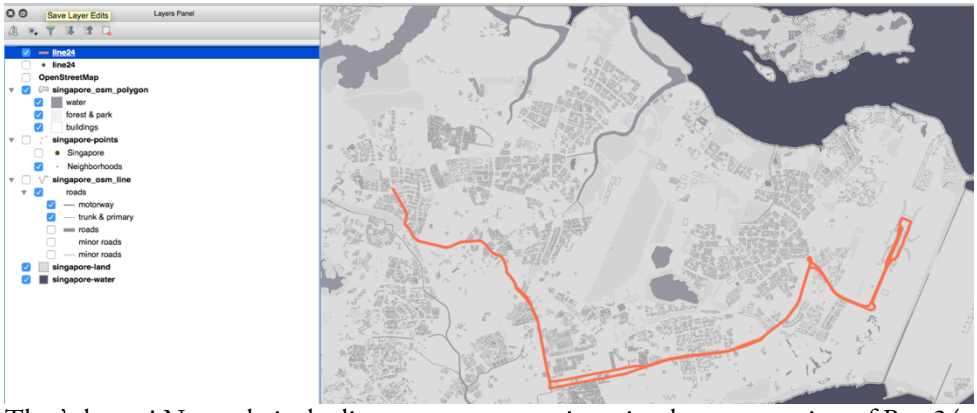
After you hit OK, you should see the points correctly plotted in the Map window.



Since these points actually represent a bus line, it would make sense to convert the individual points to a line. We can make use again of QGIS' plugins. Find the plugin called 'PointsToPaths' (look in the previous section if you forgot how) and install it. Once installed, go to the appropriate menu item to activate the tool.



The 'point group field' can be used to group points together into a single line. In this case, we only have one bus line but this would be very helpful if we had all the bus lines in Singapore in a single file. By setting the grouping based on the line_id, each bus would get its own line. Similarly, the 'point order field' is used to determine the order in which points will be connected. Often this is a timestamp field, but in this case we use the sequential 'id' integer to make things a bit easier. Give your output file an appropriate name and hit OK.



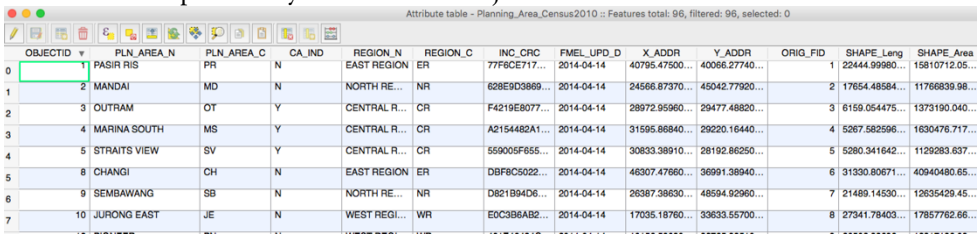
That's better! Not only is the line a more appropriate visual representation of Bus 24, it also helps us analytically. For example, if you wanted to find out what the total length of the bus line is, you can now do so. Try to think through (or Google!) the best way of calculating the total length of the bus route in QGIS. *Write down both the correct answer as well as the way you arrived at it.*

Joining non-spatial data

In the next section, we will create a thematic population map of Singapore. To start that process, open the file population_2015_pa.csv in Excel. You will see that it contains the population in each planning area in Singapore. Of course, there is some geographic information contained in this file but it is not explicit. I.e. in the previous section we had the exact coordinates for each data point on the bus line, but here we have just the name of each planning area. This is a common occurrence as secondary data often exists in these flat tables instead of explicitly spatial file formats. What we need to do to be able to visualize this population data is to find a spatial file that does have that information. We can then *join* our population data to the spatial file based on some *common attribute*.

	A	B	C	D
1	planning_area	population_2015	planning_area_caps	
2	Ang Mo Kio	174770	ANG MO KIO	
3	Bedok	289750	BEDOK	
4	Bishan	90700	BISHAN	
5	Boon Lay	30	BOON LAY	
6	Bukit Batok	139270	BUKIT BATOK	
7	Bukit Merah	155840	BUKIT MERAH	
8	Bukit Panjang	139030	BUKIT PANJANG	
9	Bukit Timah	74470	BUKIT TIMAH	
10	Central Water Catchment	10	CENTRAL WATER CATCHMENT	
11	Changi	2530	CHANGI	
12	Changi Bay	-	CHANGI BAY	
13	Choa Chu Kang	174330	CHOA CHU KANG	
14	Clementi	91630	CLEMENTI	
15	Downtown Core	3720	DOWNTOWN CORE	

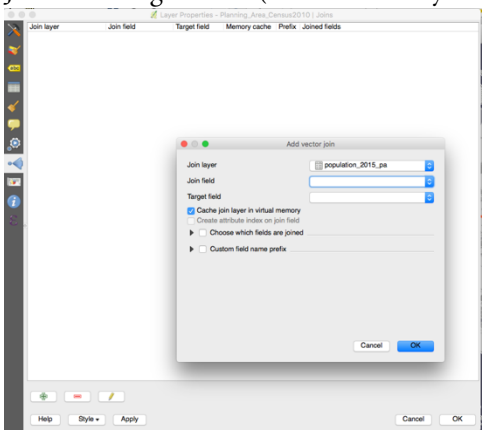
Go to https://data.gov.sg and search for 'planning area 2010'. This is the spatial file with planning areas that we are looking for. Download the SHP version of this data, unzip and add it to QGIS by going to Layer | Add Layer | Add Vector Layer (or by dragging it in). After you've added the data, open the attribute table to inspect which fields we could potentially use for our data join.



To do a successful join between our population data and the spatial data we just downloaded, you need to find a field that both data sets have in common. The field needs to match exactly otherwise the join will fail! Which fields do you think satisfy this criterium?

To start the joining process, we first need to add the population csv file to QGIS as well. Go to Layer | Add Layer | From Delimited Text File. Remember, the file *does not contain any geometry info* so you will have to make sure to set that option correctly in the dialog window.

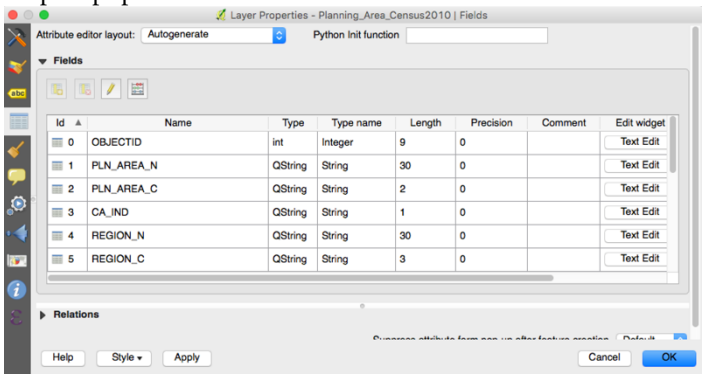
Once the population layer is added, we can create our join. Go to the properties of the Planning_Area_Census2010 layer and go to the 'Joins' tab. Create a new join ('+') and select the population_2015 layer as the 'Join Layer'. Select the appropriate Join and Target fields (remember: they need to match exactly) and click OK.



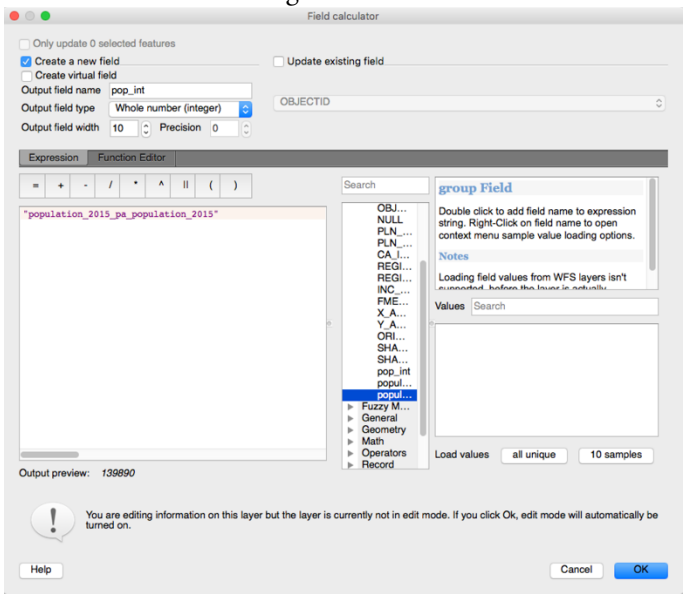
Open the Attribute Table again to make sure the join worked correctly.

OBJECTID	PLN_AREA_N	PLN_AREA_C	CA_IND	REGION_N	REGION_C	INC_CRC	FMEL_UPD_D	X_ADDR	Y_ADDR	ORIG_FID	SHAPE_Leng	SHAPE_Area	ı_2015_pa_plan	2015_pa_popul
1	PASIR RIS	PR	N	EAST REGION	ER	77F6CE717…	2014-04-14	40795.47500…	40066.27740…	1	22444.99980…	15810712.05…	Pasir Ris	139890
2	MANDAI	MD	N	NORTH RE…	NR	628E9D3869…	2014-04-14	24566.87370…	45042.77920…	2	17654.48584…	11766839.98…	Mandai	2120
3	OUTRAM	OT	Y	CENTRAL R…	CR	F4219E8077…	2014-04-14	28972.95960…	29477.48820…	3	6159.054475…	1373190.040…	Outram	22080

You can inspect the properties of each field by going to the properties of the layer and going to the 'Fields' tab. You will notice that the newly joined fields are coded as String. This makes sense for the planning area name but less so for the population. After all, population is a *ratio* variable and it would make more sense to encode as *integer* instead. In fact, we need to encode it this way if we are to create a thematic map of population in the different areas.

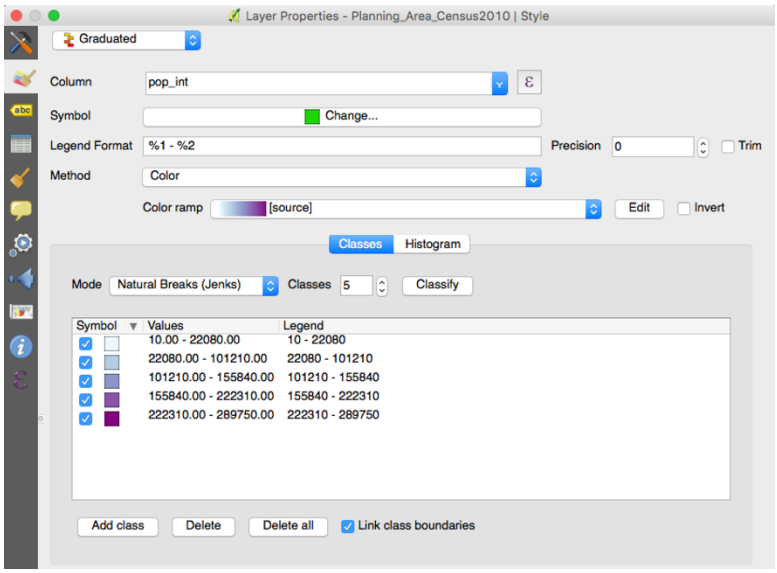


Go to the Attribute Table of the layer and select the Field Calculator  . Now we need to create a new integer field and fill it with the value of the joined String field:



Check the attribute table to make sure the conversion to integer was successful. Finally, we have everything lined up to visualize our population data!

Go to Properties again and select the Style tab. Choose 'Graduated', select the correct column and set the Mode to 'Natural Breaks' with 5 classes. Hit 'Classify' and then OK.



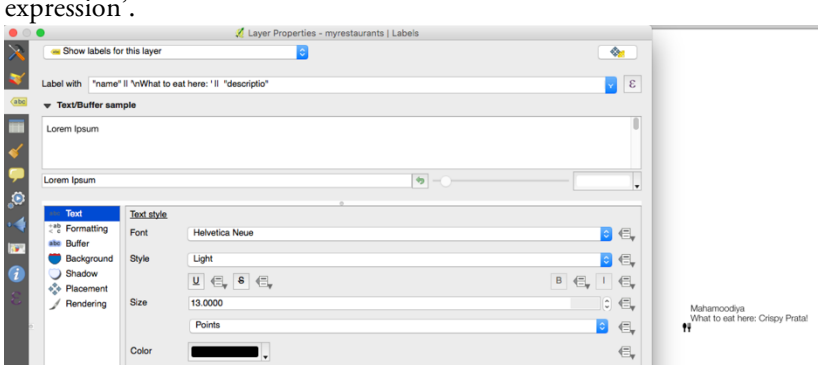
Do not worry too much about the different choices and options here. We will discuss data classification and color theory in depth during Week 5. The most important thing is: you have made your first thematic map!

You will notice that the larger planning areas generally have a higher population and densely populated but smaller areas in the center of the city actually have relatively few people. This makes sense, the larger the area the higher the chance of people living within it. This is why it is almost never a good idea to visualize absolute numbers when making a thematic choropleth map like this one. The larger areas will always have more of whatever absolute variable you are interested in visualizing. Think about how you could visualize the same data in a more appropriate way. *Write your answer down.*

Assignment

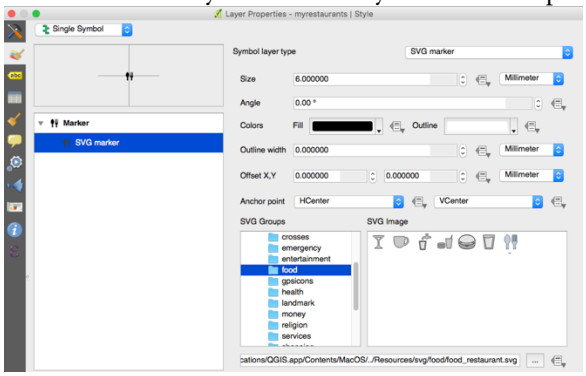
On the class website, you will find the assignment for this lab. It consists of the questions above where you were asked to write down your answer. In addition, you need to prepare two additional maps:

1. Go back to the data on your 5 favorite restaurants. Finish the map with appropriate labeling. Since there are only 5 features to label, you have bit more leeway in what you put on the labels.
    a. In this case, you need to put both the name and the description of each place on the label. To do so, you need to use the Expression dialog on the Labels tab of the Properties window. If you get stuck, try Googling 'multi-line labels QGIS' or 'QGIS label expression'.



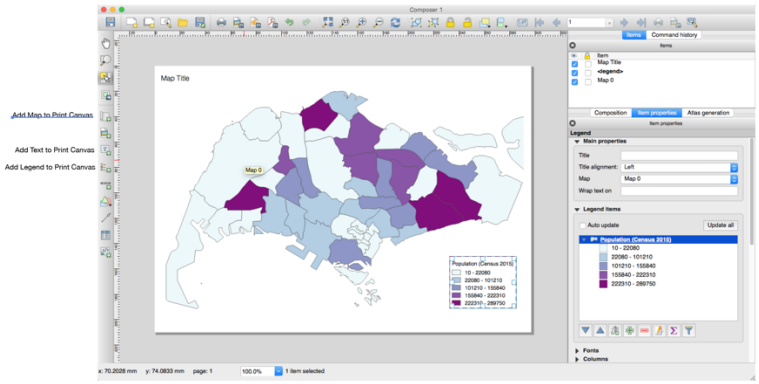
    b. The default small circles are a little boring. It would be better use an appropriate *SVG marker*. You can do this on the Style tab of the Layer Properties. You can either use some of QGIS built-in SVG symbols or go to https://thenounproject.com/ for an awesome

selection of icons you can use in your future maps.



c. Make sure you use an appropriate basemap to provide some geographic context to the map reader and export the map as PNG.

2. Finish the thematic map of population:
   a. Go to http://colorbrewer2.org/ and select an appropriate sequential color scheme for your data.
   b. Add an appropriate map title and a legend to your map. This is especially important with thematic maps as we need to tell the map reader what each color represents. In QGIS you can do so by creating a Print Composer (Project | New Print Composer). Make sure you de-clutter the legend so it only displays the essential info. When you are finished, export to PNG by clicking  .



You will quickly notice how dealing with the Print Composer is a bit clunky. This is why, in practice, the last few polishing steps of many visualization projects take place in a graphic vector design software like Adobe Illustrator. You will have to make do with QGIS for now but you will learn how to use Illustrator in a few weeks.

The assignment needs to be submitted as Word or PDF file. Please also upload the zipped Shapefile of your favorite restaurants. Please make sure you submit the assignment by **February 8.**