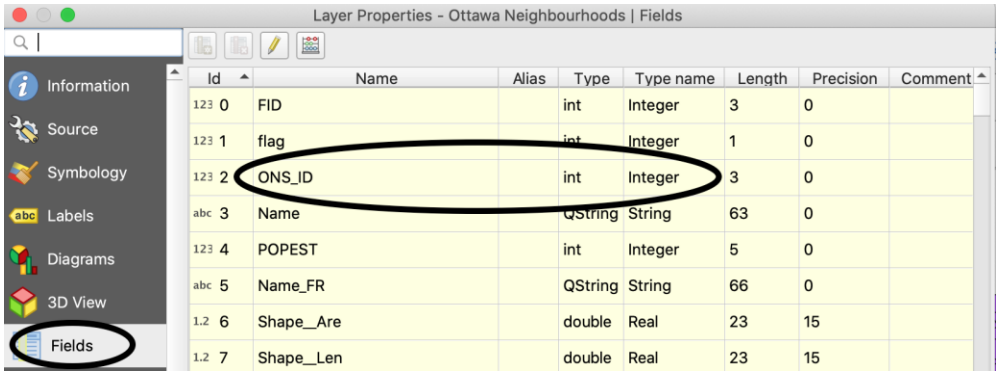
Introduction to QGIS (version 3.10) and Geoprocessing using COVID-19 data

Rebecca Bartlett, GIS & Digital Resources Librarian, November 2020

Table of Contents

5. Preparing the data for joins

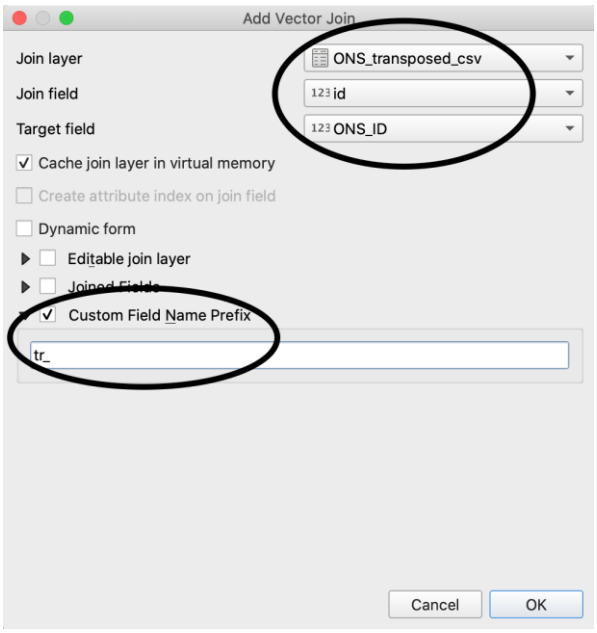
In order to join the **neighbourhoods** shapefile and the **CSVs**, ***we need to have a field that includes identical data***. For example, it's not possible to join a text (string) field to a number field. Specifically, in this case, the neighbourhood identification numbers are in both the shapefile and the CSV files. **We need to make sure those fields are all numeric**.

1) Right-click on the **Ottawa Neighbourhoods shapefile** and select Properties…
2) In the pop-up, select the **Fields** option
    a. Note that the ONS_ID field is an integer, then Close the Layer Properties window
    b. 

3) Repeat this process for **ONS_transposed_csv**
    c. The **id** field is also an integer, excellent!
4) Finally, check the fields for **ONSMapCOVID_En**
    d. There's a problem! There is no neighbourhood ID field
    e. ***This means we'll have to join the ONSMapCOVID_En CSV to the shapefile using the names of the neighbourhoods***, which means we need to hope that the neighbourhood names are spelled exactly the same (no extra spaces, no issues with dashes -/–, etc.). Fingers crossed!

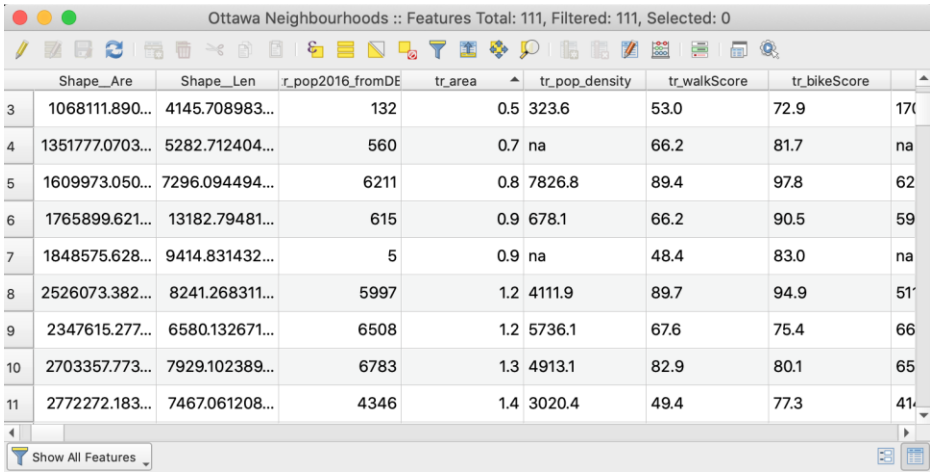
6. Joining ONS_transposed_csv to the neighbourhoods shapefile

We're going to "stick" the data from **ONS_transposed_csv** onto the neighbourhoods shapefile by joining the two together based on a common identifier. The reason we join the CSV data to the shapefile is so the spatial components (e.g. the shapes on the map) are preserved. If we stuck the shapefile to the CSV, only the non-spatial components would remain and we wouldn't be able to map it.

5) Right-click on the **Ottawa Neighbourhoods** shapefile and select **Properties..**
6) Select **Joins**
7) Click the **Add new join** button (the green plus sign at the bottom) 
8) Fill in as follows:
    a. Join layer: **ONS_transposed_csv**
    b. Join field (field in the CSV): **id**
    c. Target field (field in the shapefile): **ONS_ID**
    d. **Check off Custom Field Name Prefix**
    e. Make the Custom Field Name Prefix something short, like **tr_**
        i. The field names from the CSV will be added *after* the Custom Field Name Prefix, and the field names are only about 16 characters long, and we want to make sure that we know what's what!
    f. 

    g. Click OK
9) Click OK again
    a. You'll probably have to wait a few seconds for the join to happen, and again when you open the attribute table next
10) Open the attribute table for the Ottawa Neighbourhoods shapefile
11) We're going to check to see if the join worked well, and we'll do this by sorting to see if there are any NULL values.

a. Scroll to the right in the attribute table and find a field with the **tr_** prefix, such as **tr_area**
b. Click on the column header to sort
c. There shouldn't be any NULL values, which means our join worked!
d.



	Shape__Are	Shape__Len	r_pop2016_fromDE	tr_area	tr_pop_density	tr_walkScore	tr_bikeScore	
3	1068111.890...	4145.708983...	132	0.5	323.6	53.0	72.9	17(
4	1351777.0703...	5282.712404...	560	0.7	na	66.2	81.7	na
5	1609973.050...	7296.094494...	6211	0.8	7826.8	89.4	97.8	62
6	1765899.621...	13182.79481...	615	0.9	678.1	66.2	90.5	59
7	1848575.628...	9414.831432...	5	0.9	na	48.4	83.0	na
8	2526073.382...	8241.268311...	5997	1.2	4111.9	89.7	94.9	51
9	2347615.277...	6580.132671...	6508	1.2	5736.1	67.6	75.4	66
10	2703357.773...	7929.102389...	6783	1.3	4913.1	82.9	80.1	65
11	2772272.183...	7467.061208...	4346	1.4	3020.4	49.4	77.3	41

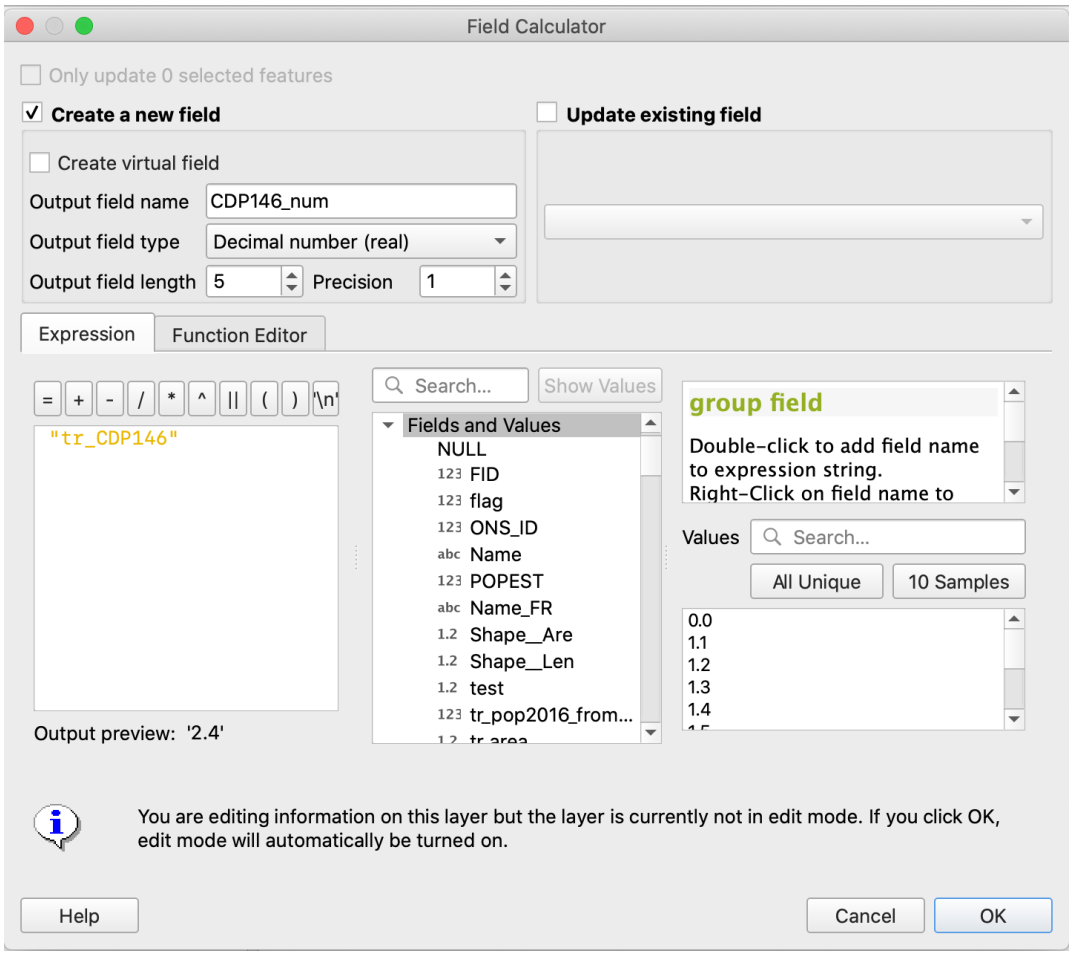
7. Making sure the 5-10 fields you chose earlier are numeric

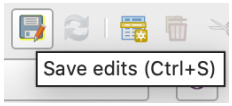
The fields were all numbers in the CSV file, but are the fields a numeric data type or are they now text/string instead?

12) Right-click on the Ottawa Neighbourhoods shapefile and select Properties…
13) Select **Fields** in Layer Properties and scroll down to check on the field types
    a. Problem: many of the fields from the CSV (with the prefix tr_) are text fields and not numeric (integer or double)
14) **Go back to your notes on which fields you're interested in**. Check to see if they are text or integer/double. If they're all integer/double, great! If not…
15) Changing the fields from text to numeric involves the Field Calculator
    a. Refer back to your notes with the field names, and which ones are integer (no decimal places) and which are double (includes decimal places), such as tr_CDP146.
16) Open the Ottawa Neighbourhoods attribute table
17) Click the **Open Field Calculator** button at the top of the attribute table 
18) **What we need to do is to *create a new numeric field* for any of the fields that were imported as text, and copy the data into the new field.** Fill out the Field Calculator as follows:
    a. Check off **Create a new field**
    b. Write an **output field name** (needs to be less than 16 characters!), e.g. CDP146_num
    c. Output field type: **Whole number (integer) or Decimal number (real)**
        i. If you select Decimal number (real): *Precision = number of decimal places* so if you want 2 decimal places, set it to 2, and so on.
    d. In the lower middle section of the Field Calculator, find the **Fields and Values** item and click on the arrow to view all the different fields in the dataset

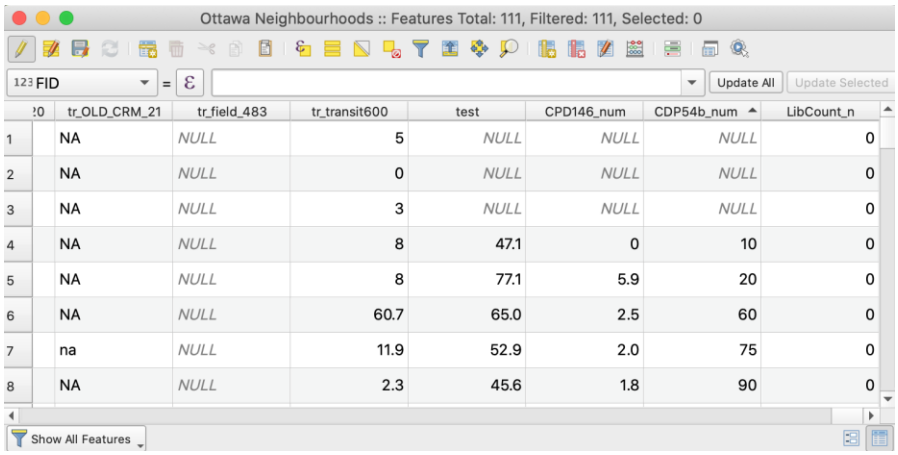
e. Select the field you want to convert to numeric (e.g. CDP146) and **double-click** on it, which will pop that field name into the Expression box.
    i. On the right, you can click **All Unique** to view all the values in that field (especially helpful if you forgot if it has decimals or not)

f. The Field Calculator should now look something like this:



19) Click OK
20) In the attribute table, **Save edits**
    a. 

21) Repeat for the other fields you want to convert to numeric.
    a. Don't forget to save the edits!
22) When you've finished, the fields should be at the very end (far right) of the attribute table

a. 

	tr_OLD_CRM_21	tr_field_483	tr_transit600	test	CPD146_num	CDP54b_num	LibCount_n
1	NA	NULL	5	NULL	NULL	NULL	0
2	NA	NULL	0	NULL	NULL	NULL	0
3	NA	NULL	3	NULL	NULL	NULL	0
4	NA	NULL	8	47.1	0	10	0
5	NA	NULL	8	77.1	5.9	20	0
6	NA	NULL	60.7	65.0	2.5	60	0
7	na	NULL	11.9	52.9	2.0	75	0
8	NA	NULL	2.3	45.6	1.8	90	0

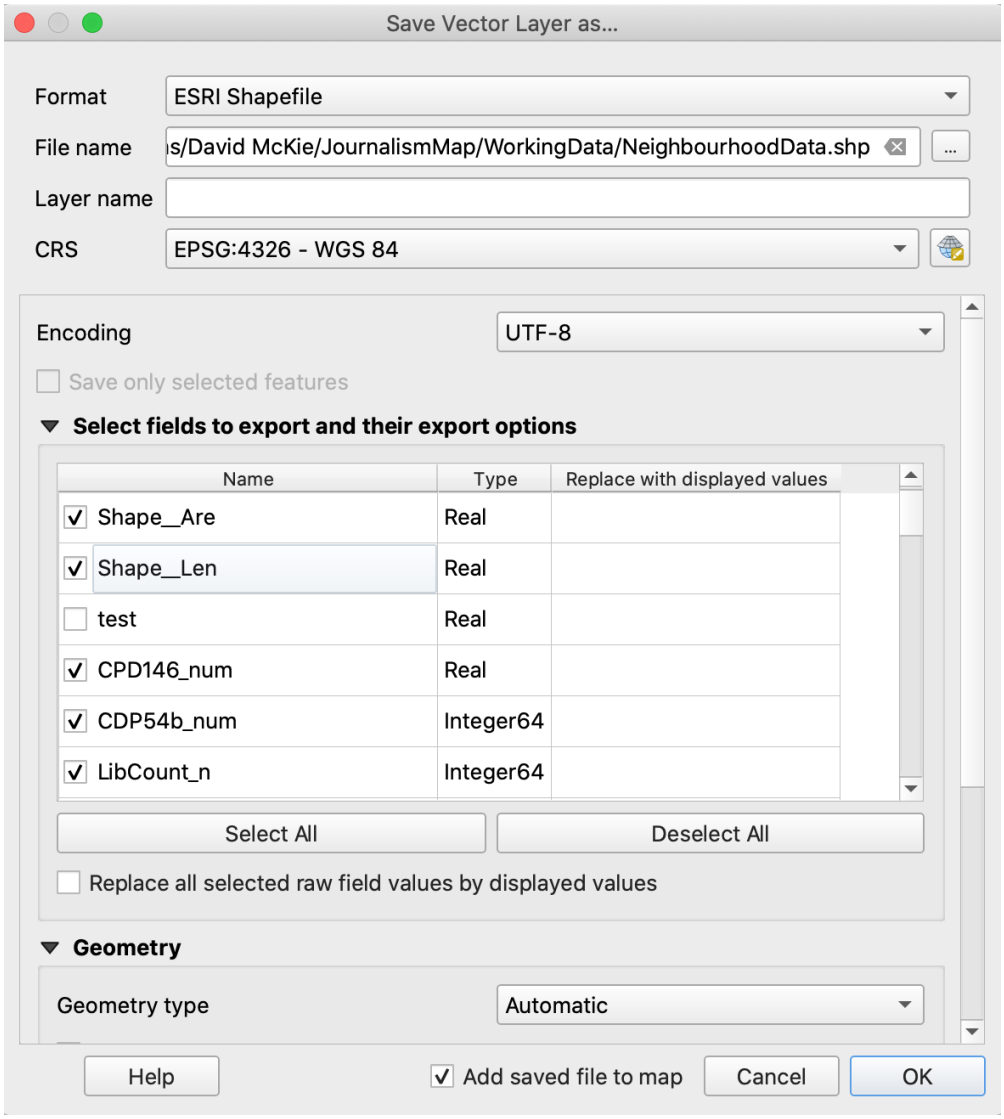
b. A *NULL* value means it probably had a 'na' designation or other letters in the original spreadsheet entry

23) **Save your edits!**

8. Exporting the neighbourhood shapefile with your new data attached

In order to make the join permanent, we need to save the shapefile. Otherwise the join is just temporary.

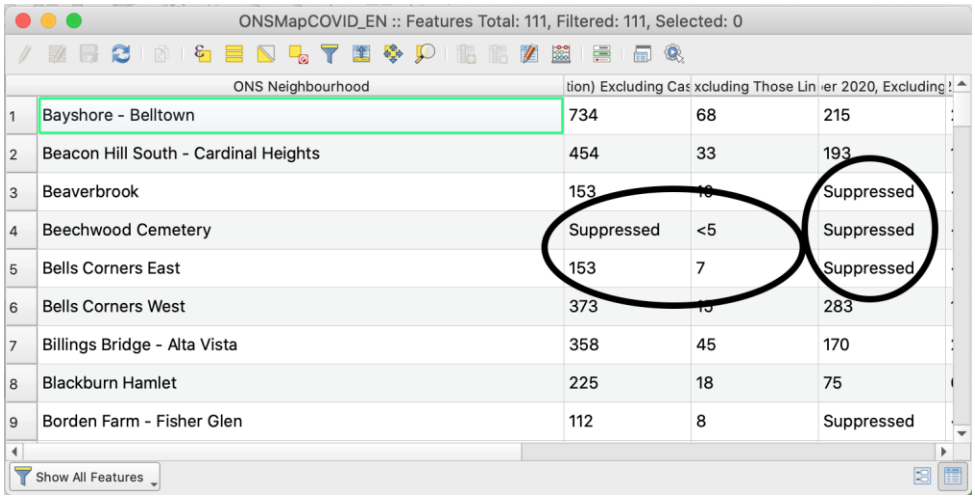
24) Right-click on Ottawa Neighbourhoods and select **Export > Save features as…**
   a. Format: **Esri shapefile**
   b. Filename: browse to your WorkingData folder and give the new shapefile an informative name
   c. Click **Select fields to export and their export options**
      i. **Deselect All**
         1. If you select all, you'll get an error. Converting all those fields to numeric not only saves us from that error, but also allows you to symbolize the data by the numeric value later.
      ii. Select the first several fields (neighbourhood name, ONS_ID, shape length, etc.) and your new numeric fields which should helpfully be near the top of the list.
      iii. If you select other fields from the CSV for the hell of it, you may get an error and have to re-do the save.
   d. Click OK when you're done

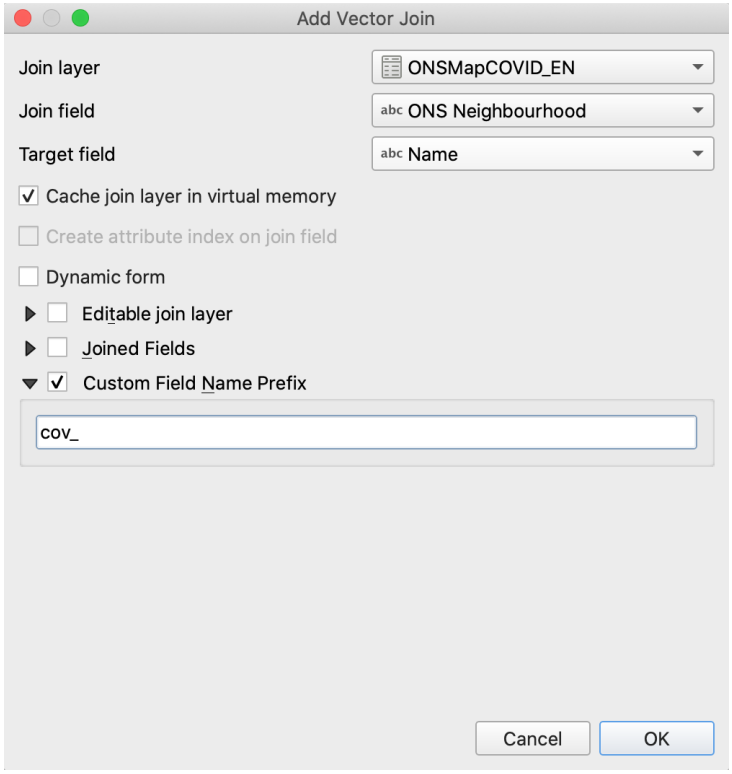
e. 

25) The new shapefile should appear in your map and Layers pane.

9. Joining the ONSMapCOVID_EN CSV to the Neighbourhood shapefile

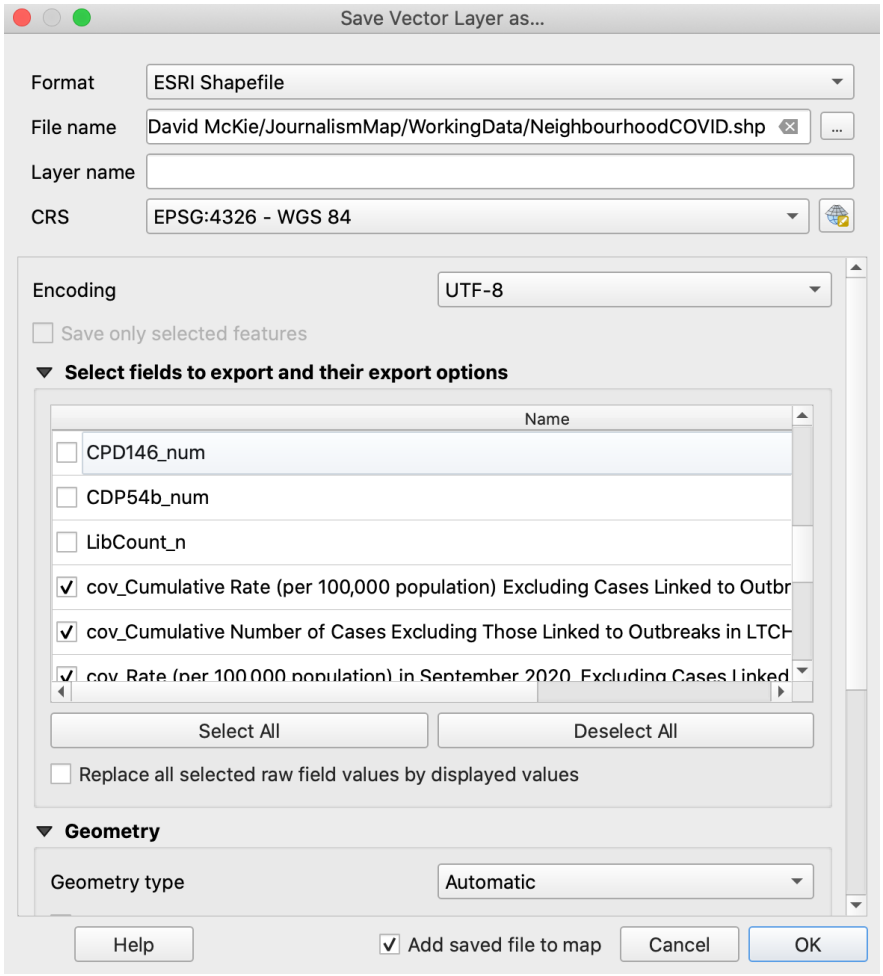
26) Right-click on ONSMapCOVID_EN and open the attribute table
27) Note that there is no neighbourhood ID number, so we'll have to rely on the neighbourhood names matching.
    a. Spoiler: it works out fine, but we have other fish to fry…
28) We can see some issues with the data right off the top, namely that the dataset has both text and numbers in the same fields. This presents a problem since the data will be joined as a text field and not a numeric field if there's any text.
    a. SIGH. But we will fix it later with our new best friend, the Field Calculator.

b. 

29) Okay, so the join. Right-click on Ottawa Neighbourhoods and open Properties.
30) Go to the Joins section and click the Add Join button.
    a. You should remove the previous ONS_transposed_csv join to get rid of all those dozens of fields, especially since you saved the new shapefile with all that data.
31) Fill out the Join window as below:
    a. 

32) Click OK
33) Open the Ottawa Neighbourhoods attribute table to take a look and see if everything joined up okay and there aren't any neighbourhoods with no joined data.
    a. Yay, it worked!
    b. …but we still have that pesky Suppressed/<5 issue to deal with ☹

34) Let's export the Ottawa Neighbourhoods file to a new shapefile
    a. Right-click > Export > Save features as…
    b. All you really need to do is browse to your WorkingData folder and give the file a helpful name, like **NeighbourhoodCOVID**
    c. If you feel like it, uncheck the fields that we created with the transposed CSV (e.g. CDP54bnum, etc.)
    d.



10. Converting the fields with Suppressed/<5 entries to numeric fields

As with the previous join, we need the fields to be numeric in order to symbolize the data – and therefore interpret it – in a meaningful way.

35) Okay, now let's deal with those Suppressed/<5 entries
    a. **You may be wondering how to change Suppressed and <5 to actual numbers. You can do it by using numeric entries that will function as codes, but you need to note this as you make these changes.**

i. **Converting those fields to numeric will change both Suppressed and <5 values to NULL**

36) Open the attribute table for NeighbourhoodCOVID, then fire up the Field Calculator

37) **For all fields with the cov_ prefix, do the following:**

a. **Make a note somewhere of whether that field has Suppressed or <5 values**
   i. You'll need this later on for when you label your symbology

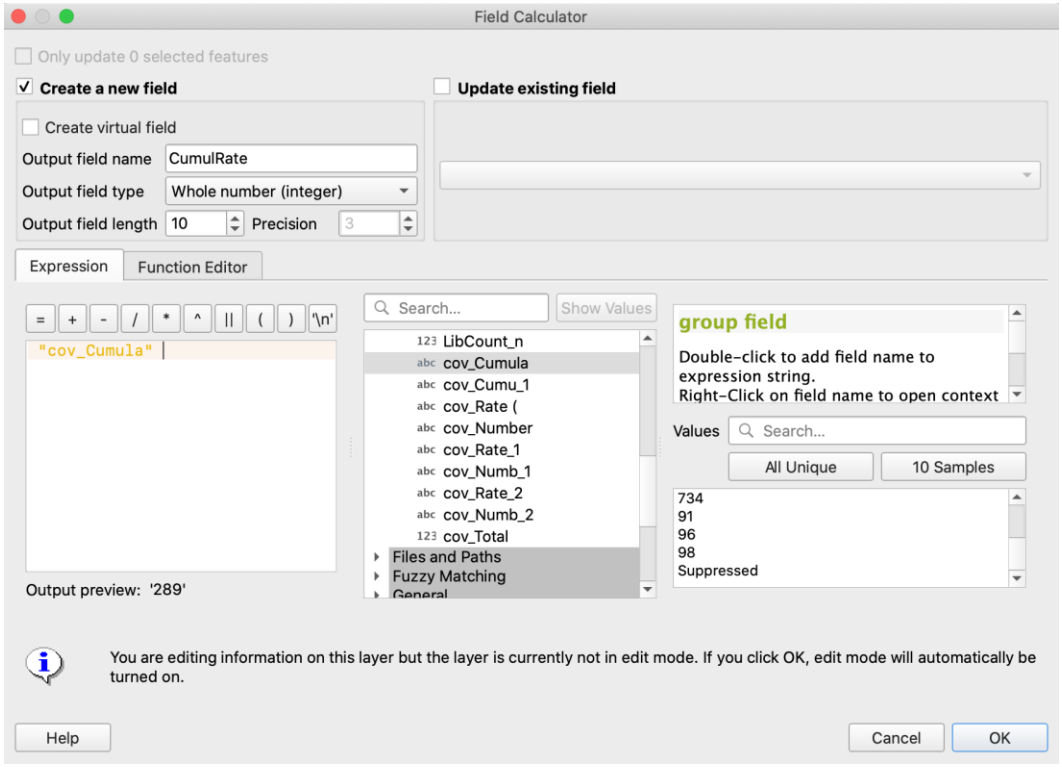
b. Check off **Create a new field**

c. Write an **output field name** (needs to be less than 16 characters!), e.g. CumulRate

d. Output field type: **Whole number (integer) or Decimal number (real)**
   i. If you select Decimal number (real): *Precision = number of decimal places* so if you want 2 decimal places, set it to 2, and so on.

e. In the lower middle section of the Field Calculator, find the **Fields and Values** item and click on the arrow to view all the different fields in the dataset

f. Select the field you want to convert to numeric (e.g. cov_Cumul) and **double-click** on it, which will pop that field name into the Expression box.
   i. On the right, you can click **All Unique** to view all the values in that field (especially helpful if you forgot if it has decimals or not)

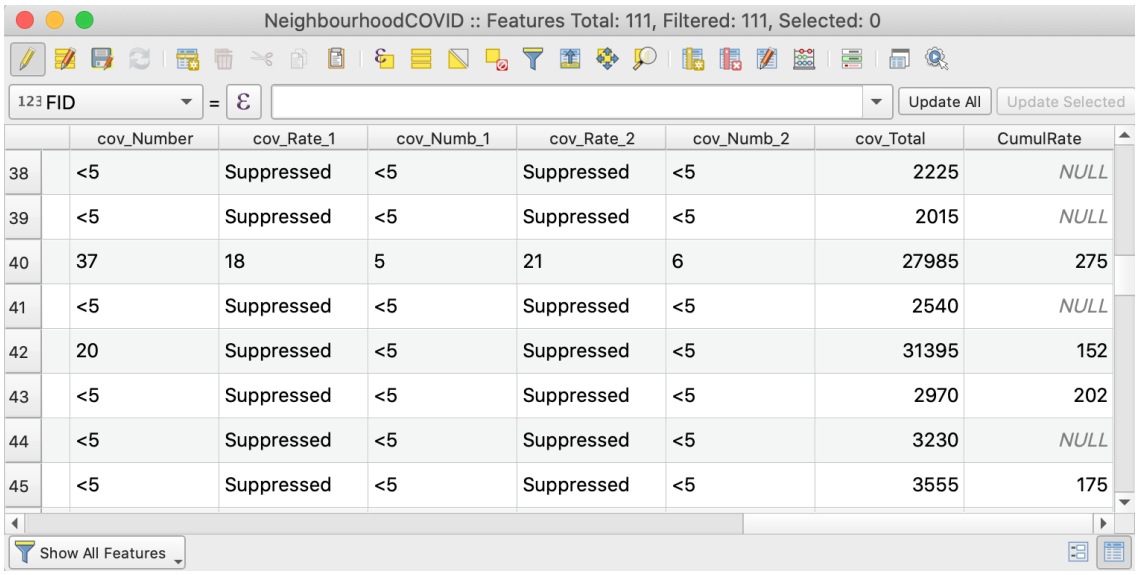


g.

h. When it matches the example above, click OK

38) This results in a new numeric field with NULL values for the Suppressed or <5 values.

a.



	cov_Number	cov_Rate_1	cov_Numb_1	cov_Rate_2	cov_Numb_2	cov_Total	CumulRate
38	<5	Suppressed	<5	Suppressed	<5	2225	NULL
39	<5	Suppressed	<5	Suppressed	<5	2015	NULL
40	37	18	5	21	6	27985	275
41	<5	Suppressed	<5	Suppressed	<5	2540	NULL
42	20	Suppressed	<5	Suppressed	<5	31395	152
43	<5	Suppressed	<5	Suppressed	<5	2970	202
44	<5	Suppressed	<5	Suppressed	<5	3230	NULL
45	<5	Suppressed	<5	Suppressed	<5	3555	175

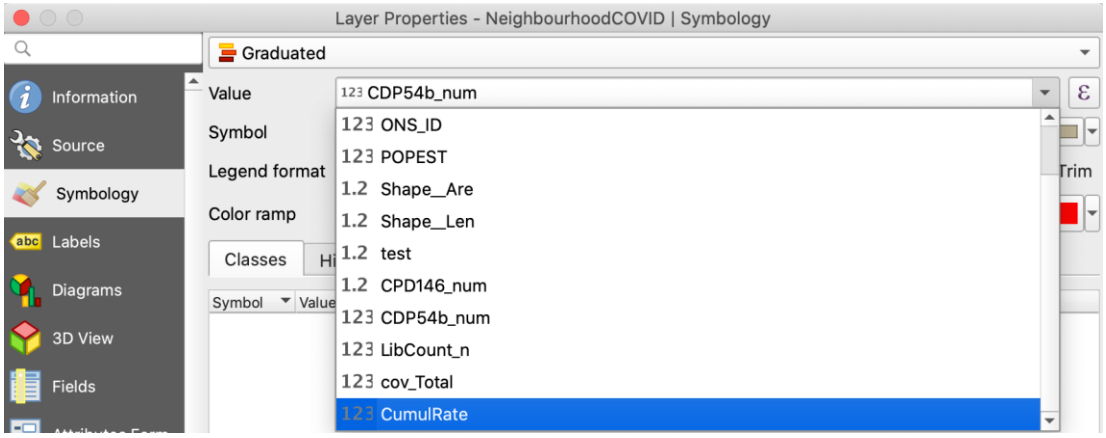
39) **Save your layer edits!**

Once you have fields converted to numeric, you can symbolize, FINALLY!

11. Symbology

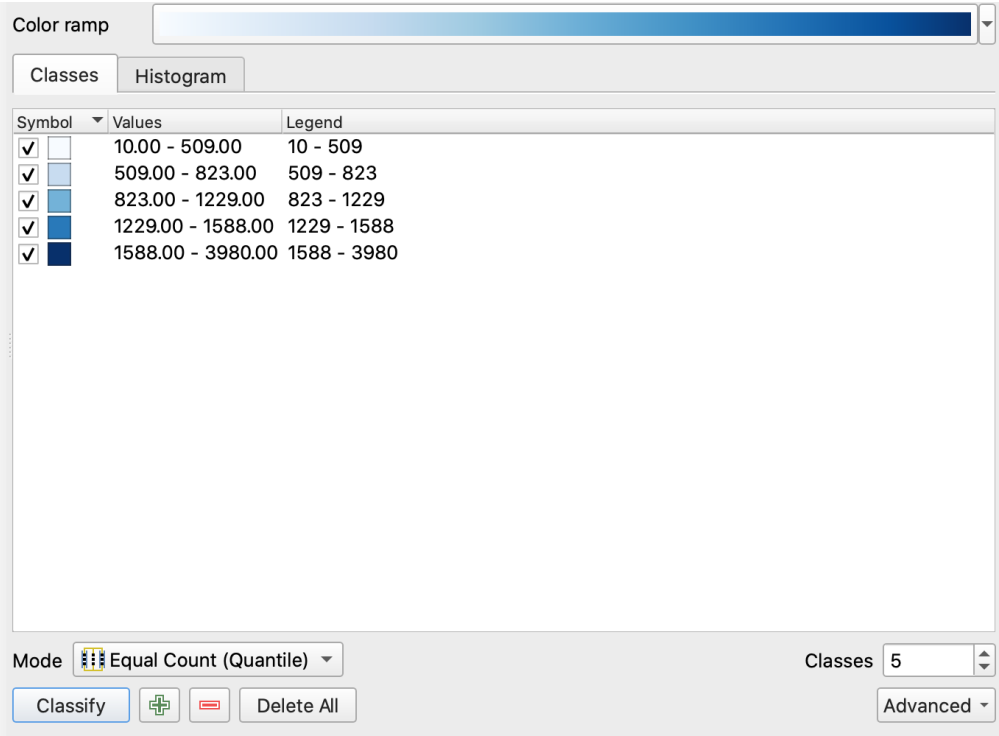
40) To symbolize right-click on the **NeighbourhoodCOVID** and select **Properties.**
41) Click the **Symbology** tab in the Layer Properties dialog box.
42) Click **Graduated** in the drop-down box at the very top
43) For **Value**, select **CumulRate**
    a. You'll notice that only numeric fields show up and there are no text fields.

b.

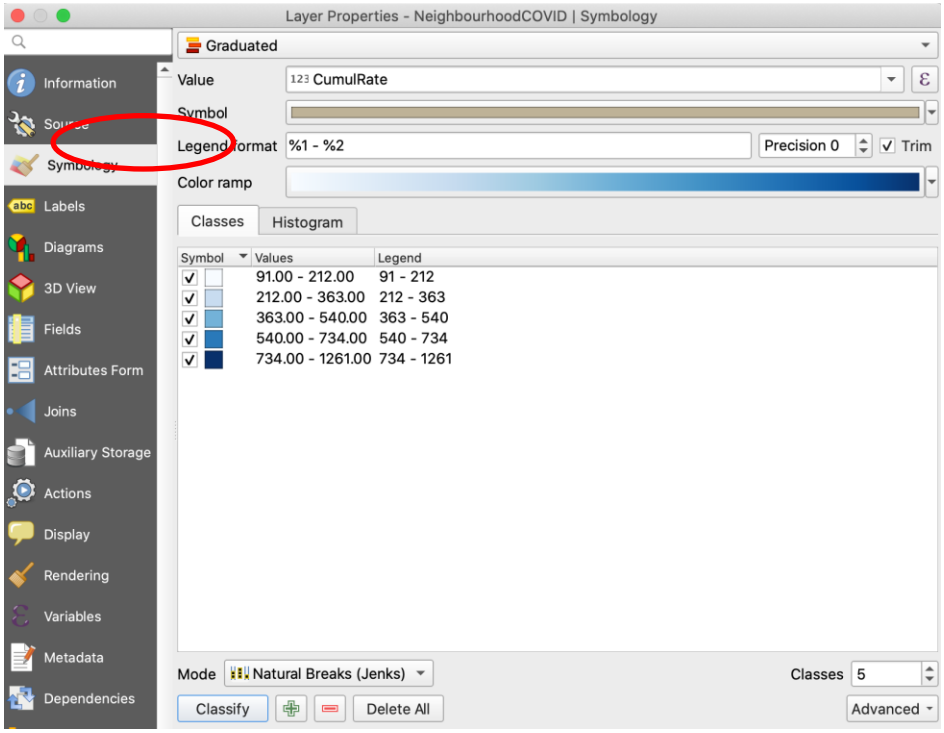


44) Select a colour ramp, then click the **Classify** button near the bottom. Only after you click Classify will you see classes.

a.

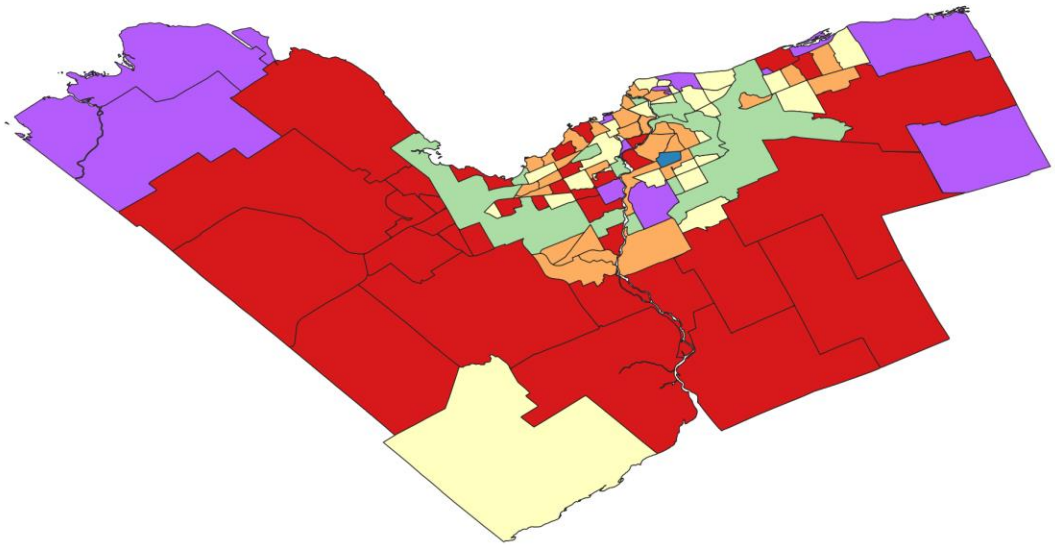


45) It should be filled in as you see below, then click OK (NOTE: depending on your version of QGIS, your dialog box may look slightly different.)
    a. Column should be **CumulRate** (or whichever one you chose).
    i.

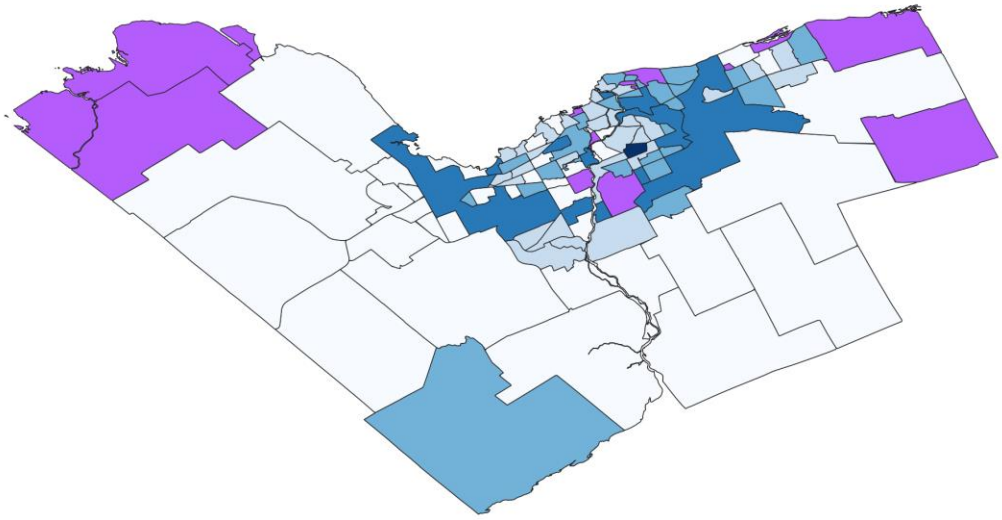


    b. Under the **Classes** tab, Mode should be **Natural Breaks (Jenks)** with 5 classes
        i. You can play around with this if you'd like

c. Pick your favourite colour ramp. As you are showing intervals, go with something that's the same hue, but with varying intensity. If you go with a many-coloured ramp your map becomes meaningless, like so:



46) Your on-screen map should now look something like this, with the darker hues showing us which CTs have higher and lower population changes.



a.

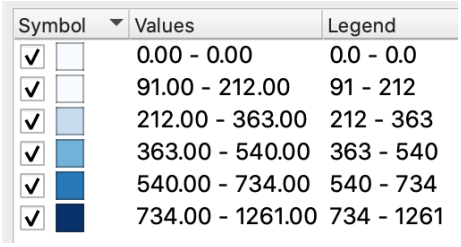
b. ***The neighbourhoods with NULL values in this field are invisible***. The purple areas are there because the Ottawa Neighbourhoods shapefile is purple and shows up below it.

47) One way to address the invisible NULL neighbourhoods is using **Rule-based Classification** as follows:

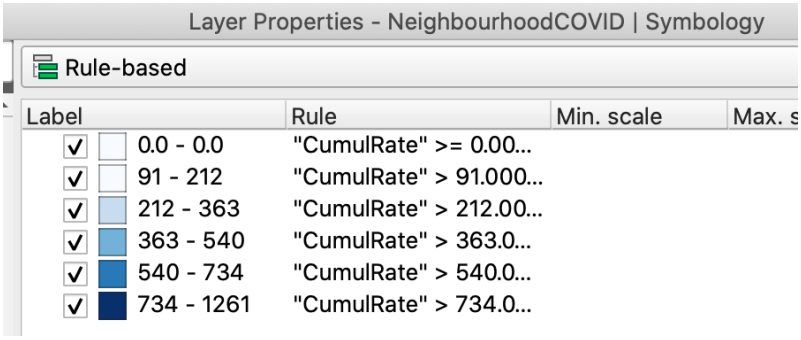
a. In the Symbology window in Properties, click the **Add class** button

i. 

b. It should automatically add a 0.00 class at the top of the list:

i. 

c. Where it currently says Graduated at the top, select Rule-based from the drop-down:

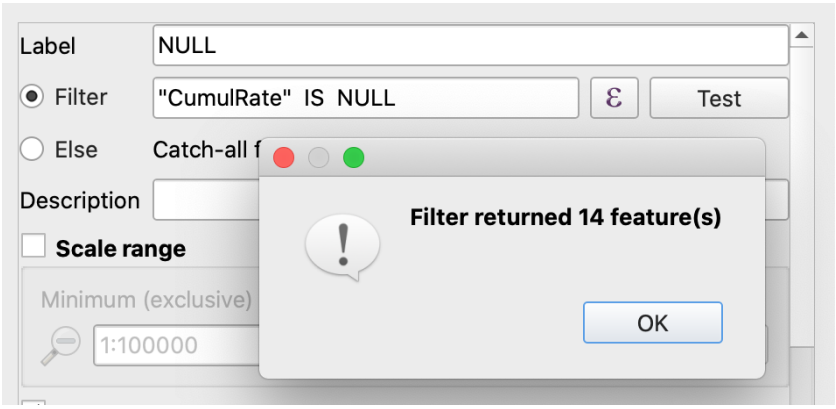
i. 

d. **Double-click on the first entry under the Rule column**

i. In the pop-up, change the label to NULL

ii. In the Filter, delete all but the field name (e.g. “CumulRate”) and then type **IS NULL** in all caps

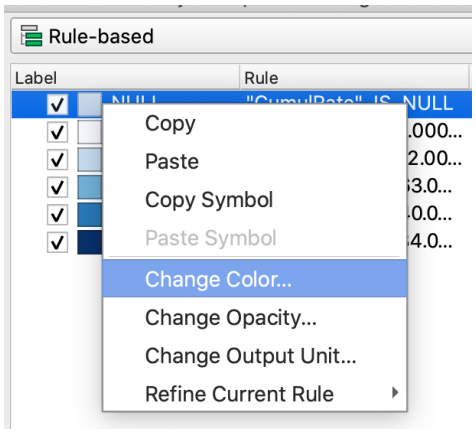
iii. Click the Test button to make sure you get some features returned:

iv. 

v. Click OK, then OK again

e. **Change the colour**

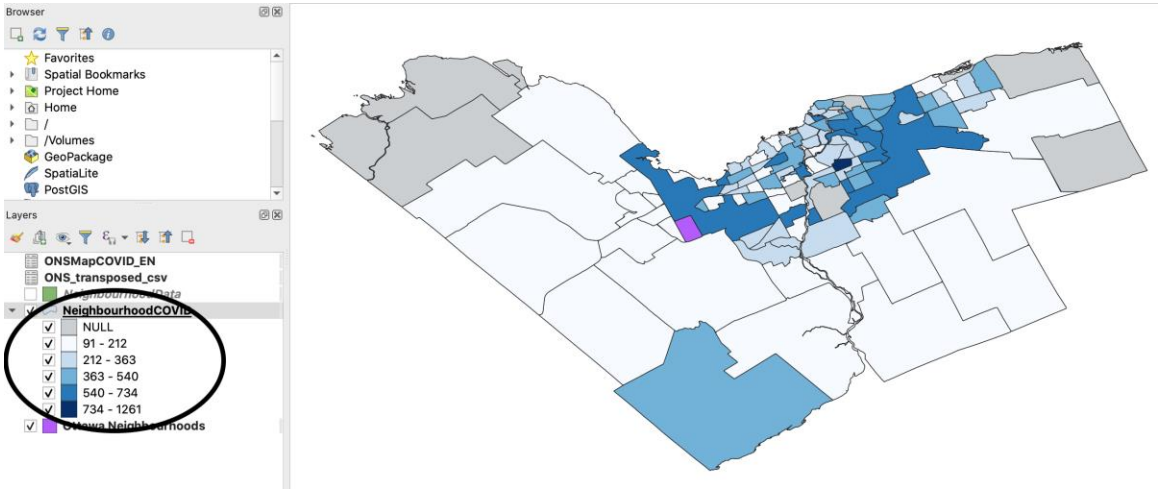
i. Right-click on the new NULL category and select Change Color…

1. 

ii. Choose a medium grey (or something different than the colour scale you've chosen)

iii. Click OK, then OK again

f. Your map should now look like this:

g. 

You did it!! Have a treat of some kind (screen break, nap, stretch, favourite snack, something else you enjoy) to celebrate!