

Introduction to QGIS (version 3.16) and Geoprocessing using COVID-19 data

Rebecca Bartlett, GIS Librarian, November 2021

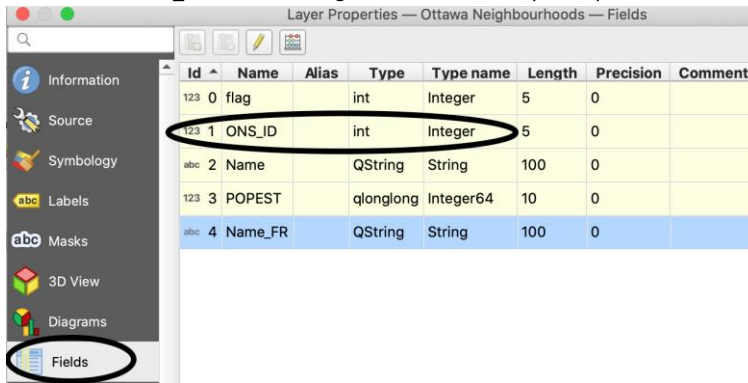
Table of Contents

- 5. PREPARING THE DATA FOR JOINS.....1
- 6. JOINING ONS_TRANSPOSED_CSV TO THE NEIGHBOURHOODS SHAPEFILE2
- 7. MAKING SURE THE 5-10 FIELDS YOU CHOSE EARLIER ARE NUMERIC3
- 8. EXPORTING THE NEIGHBOURHOOD SHAPEFILE WITH YOUR NEW DATA ATTACHED5
- 9. JOINING THE ONSMAPCOVID_EN CSV TO THE NEIGHBOURHOOD SHAPEFILE6
- 10. CONVERTING THE FIELDS WITH SUPPRESSED/<5 ENTRIES TO NUMERIC FIELDS10
- 11. SYMBOLOGY12

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




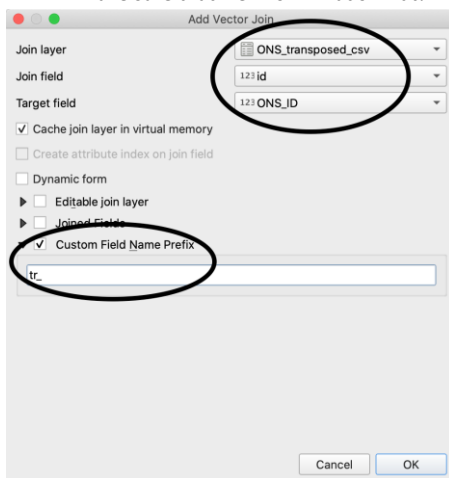
- b.
- 3) Repeat this process for **ONS_transposed_csv**
 - c. The **id** field is also an integer, excellent!
- 4) Also check the fields for **Data_tables_for_ONS_Neighbourhood_COVID-19_Maps**
 - d. The **_ONS_ID** field is an integer, hooray!
- 5) Also check the fields for **Data_tables_for_ONS_Neighbourhood_COVID-19_Vaccination_Maps_**

- e. It also has an **ONS_ID** field that's an integer. We're set!

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 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.

- 6) Right-click on the **Ottawa Neighbourhoods** shapefile and select **Properties..**
- 7) Select **Joins**
- 8) Click the **Add new join** button (the green plus sign at the bottom) 
- 9) 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
- 10) 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
- 11) Open the attribute table for the Ottawa Neighbourhoods shapefile

- 12) 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!

	Name_FR	r_pop2016_fromDE	tr_area	tr_pop_density	tr_walkScore
1	Cimetière Notre-Dame	426	0.2	na	81.3
2	Orléans Industriel	132	0.5	323.6	53.0
3	NULL	3943	0.5	10479.0	57.8
4	Cimetière Beechwood	560	0.7	na	66.2
5	Basse-ville	6211	0.8	7826.8	89.4
6	Université Carleton	5	0.9	na	48.4
7	Développement Lebreton	615	0.9	678.1	66.2
8	NULL	6508	1.2	5736.1	67.6
9	Marché By	5997	1.2	4111.9	89.7
10	Vanier Sud	6783	1.3	4913.1	82.9

d.

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

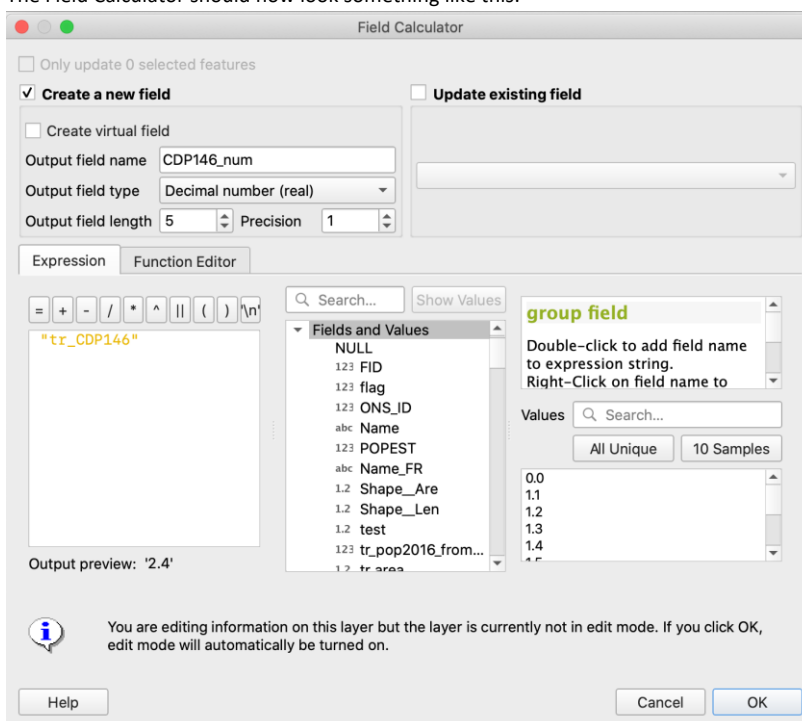
The fields were all numbers in the CSV file, but are the fields a numeric data type or are they now text/string instead?

- 13) Right-click on the Ottawa Neighbourhoods shapefile and select Properties...
- 14) 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)
- 15) **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...
- 16) 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**.
- 17) Open the Ottawa Neighbourhoods attribute table



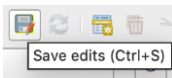
- 18) Click the **Open Field Calculator** button at the top of the attribute table
- 19) **What we need to do is to create a new numeric field for any of the fields that were imported as text, and copy the data into the new field.** Fill out the Field Calculator as follows:
 - a. Check off **Create a 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:



20) Click OK

21) In the attribute table, **Save edits**



a.

22) Repeat for the other fields you want to convert to numeric.

a. Don't forget to save the edits!

23) When you've finished, the fields should be at the very end (far right) of the attribute table

	nority23	tr_minority24	tr_minority25	tr_minority26	CDP146_num	CDP54b_num	CDP199_num
1		na	na	na	NULL	NULL	NULL
2		0.0	0.0	88.2	0	10.0	NULL
3		0.2	0.4	47.3	3.8	1350.0	39499.00
4		na	na	na	NULL	NULL	NULL
5		0.2	0.0	65.5	3.2	960.0	34677.00
6		na	na	na	NULL	NULL	NULL
7		0.0	0.0	78.8	2.5	60.0	73543.00
8		0.0	0.3	51.2	3.8	1010.0	45765.00
9		0.3	0.4	79.4	2.4	720.0	54686.00

- a.
- b. A *NULL* value means it probably had a 'na' designation or other letters in the original spreadsheet entry, which is often due to low population or unavailable information.

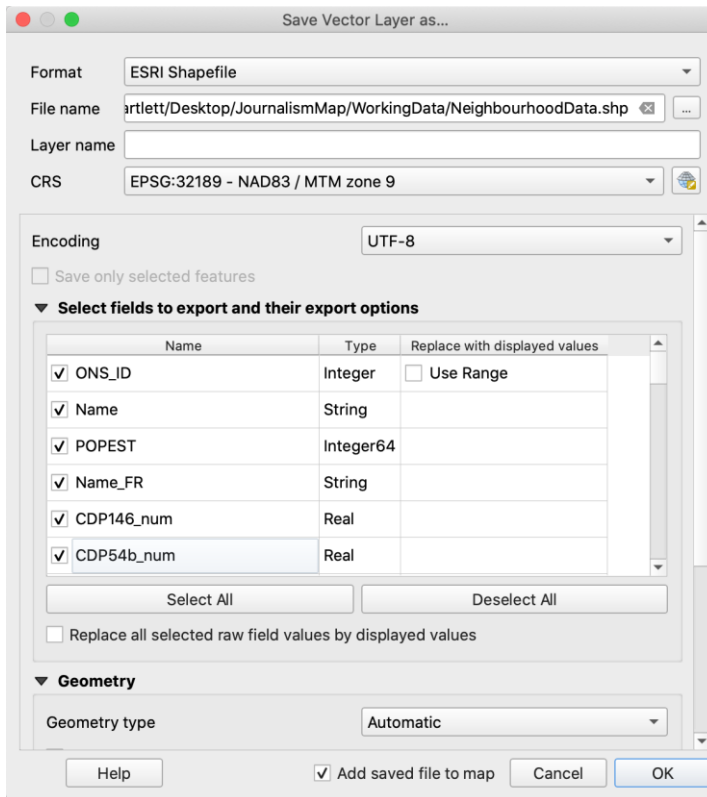
24) **Save your edits!**

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.

25) 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.

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

9. Joining the Data_tables_for_ONS_Neighbourhood_COVID-19_Vaccination_Maps_CSV to the Neighbourhood shapefile

27) Right-click on Data_tables_for_ONS_Neighbourhood_COVID-19_Vaccination_Maps_ and open the attribute table

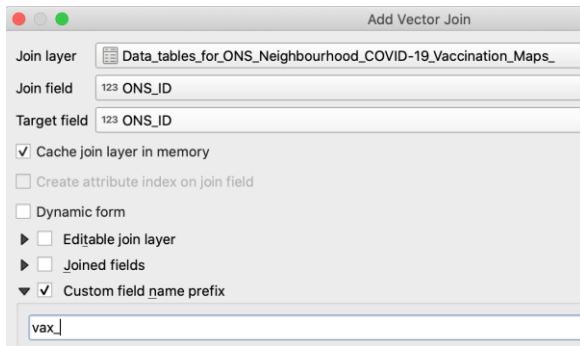
28) As we found out earlier, there is an ONS_ID field. Great news!

29) Right-click on Ottawa Neighbourhoods and select 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:



b.

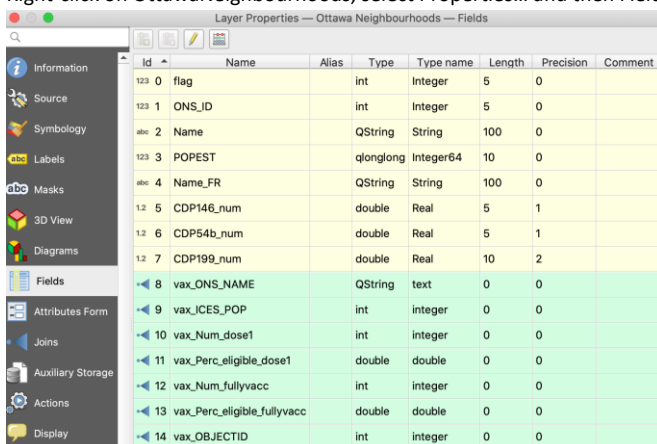
32) Click OK

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. A few will have NULL values

34) Let's also have a look at the fields to see if they're numeric

a. Right-click on OttawaNeighbourhoods, select Properties... and then Fields. It looks good!



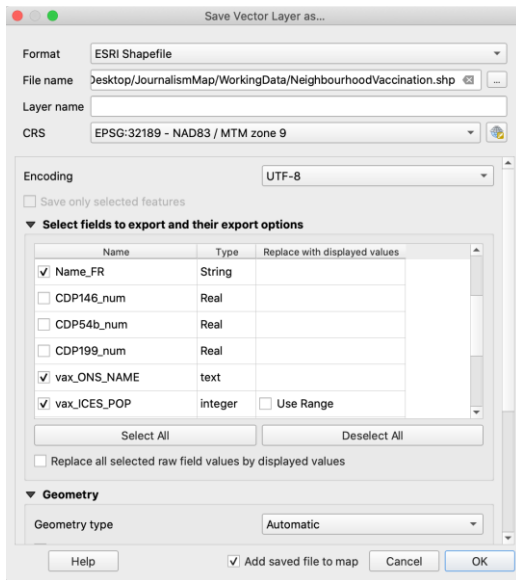
b.

35) Export the Ottawa Neighbourhoods file to a new shapefile

c. Right-click > Export > Save features as...

d. All you really need to do is browse to your WorkingData folder and give the file a helpful name, like **NeighbourhoodVaccination**

e. If you feel like it, uncheck the fields that we created with the transposed CSV (e.g. CDP54bnum, etc.)



f.

10. Joining the Data tables_for_ONS_Neighbourhood_COVID-19_Maps CSV to the Neighbourhood shapefile

36) Right-click on Data_tables_for_ONS_Neighbourhood_COVID-19_Maps and open the attribute table

37) Like the Vaccine CSV there is an ONS_ID field. Woohoo!

38) However, 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.

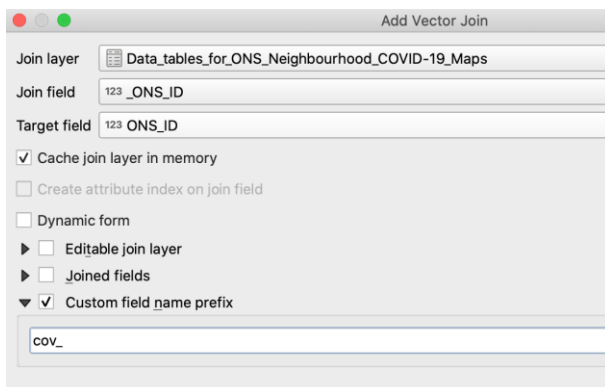
g. SIGH. But we will fix it later with our new best friend, the Field Calculator.

ONS Neighbourhood	Cases	Deaths	Recovered
Bayshore - Belltown	734	68	215
Beacon Hill South - Cardinal Heights	454	33	193
Beaverbrook	153	16	Suppressed
Beechwood Cemetery	Suppressed	<5	Suppressed
Bells Corners East	153	7	Suppressed
Bells Corners West	373	15	283
Billings Bridge - Alta Vista	358	45	170
Blackburn Hamlet	225	18	75
Borden Farm - Fisher Glen	112	8	Suppressed

h.

39) First though, the join. Right-click on Ottawa Neighbourhoods and open Properties.

- 40) Go to the Joins section and click the Add Join button.
 - i. You should remove the previous join just to keep things tidy.
- 41) Fill out the Join window as below:



- j.
- 42) Click OK
- 43) 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.
 - k. Yay, it worked!
 - l. ...but we still have that pesky Suppressed/<5 issue to deal with ☹️
- 44) Let's export the Ottawa Neighbourhoods file to a new shapefile
 - m. Right-click > Export > Save features as...
 - n. All you really need to do is browse to your WorkingData folder and give the file a helpful name, like **NeighbourhoodCOVID**
 - o. If you feel like it, uncheck the fields that we created with the transposed CSV (e.g. CDP54bnum, etc.)



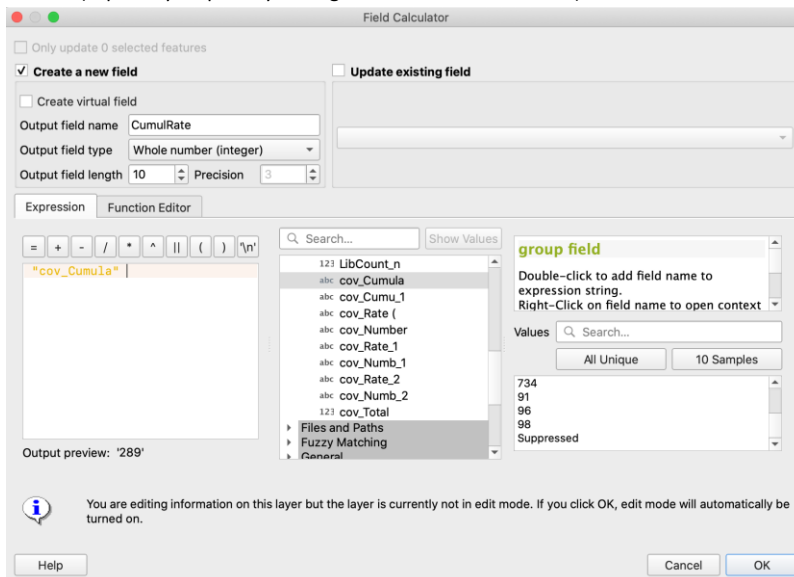
p.

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

We need the fields to be numeric in order to symbolize the data – and therefore interpret it – in a meaningful way.

- 45) Okay, now let's deal with those Suppressed/<5 entries in NeighbourhoodCOVID
 - q. **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**
- 46) Open the attribute table for NeighbourhoodCOVID and fire up the Field Calculator
- 47) **For all fields with the cov_ prefix that you're interested in, do the following:**
 - r. **Make a note somewhere of whether that field has Suppressed or <5 values**
 - i. Spoiler: it's basically all of them ☹
 - ii. You'll also need this info later on for when you label your symbology
 - s. Check off **Create a new field**
 - t. Write an **output field name** (needs to be less than 16 characters!), e.g. CumulRate
 - u. 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.
- v. 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
- w. Select the field you want to convert to numeric (e.g. cov_Cumula) 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)



- x.
- y. When it matches the example above, click OK

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

	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

Z.

49) **Save your layer edits!**

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

12. Symbology

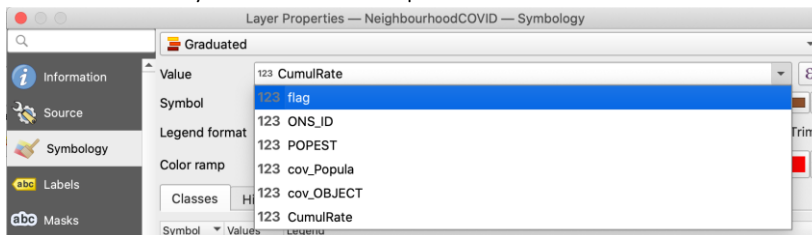
50) To symbolize right-click on the **NeighbourhoodCOVID** and select **Properties**.

51) Click the **Symbology** tab in the Layer Properties dialog box.

52) Click **Graduated** in the drop-down box at the very top

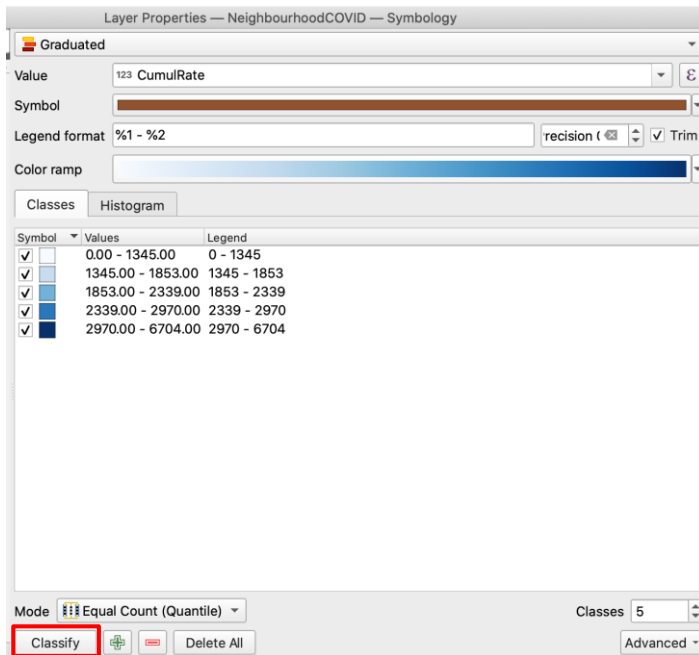
53) For **Value**, select **CumulRate**

a. You'll notice that only numeric fields show up and there are no text fields.



b.

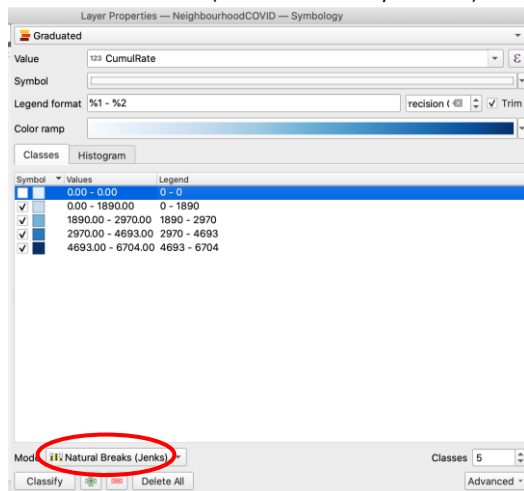
54) Select a colour ramp, then click the **Classify** button near the bottom. Only after you click Classify will you see classes.



c.

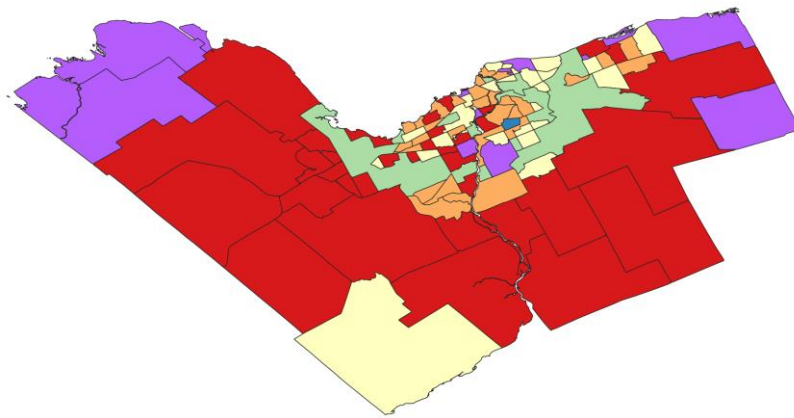
55) It should be filled in as you see below, then click OK (NOTE: depending on your version of QGIS, your dialog box may look slightly different.)

a. Column should be **CumulRate** (or whichever one you chose).

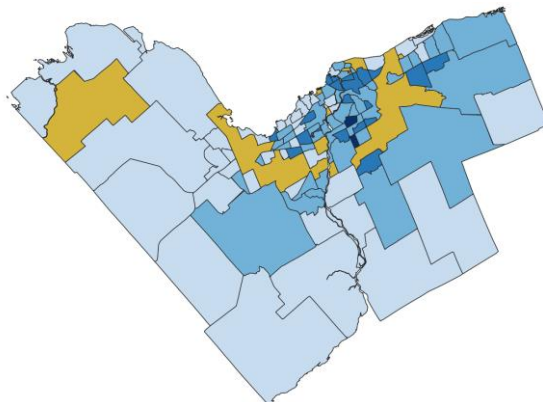


i.

- ii. See that weird 0.00-0.00 field at the top? Uncheck it for now and we'll deal with it shortly.
- 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:



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



- d.
- e. **The neighbourhoods with NULL values in this field are invisible.** The yellowish areas are there because the Ottawa Neighbourhoods shapefile is yellow and shows up below it.

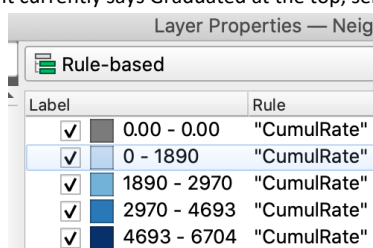
Commented [DM1]: The references to CTs and population change are confusing. Are you referring to census tracts when it should be neighbourhoods?

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

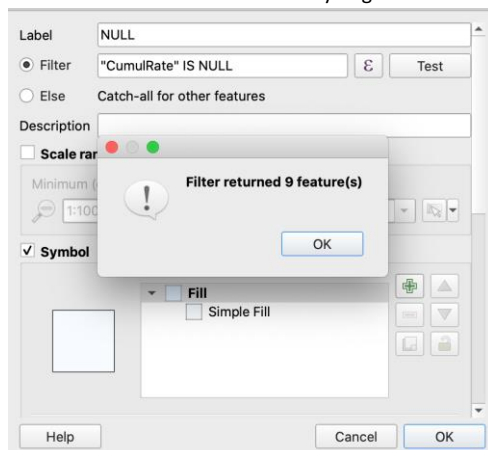
- f. Remember that weird 0.00-0.00 field at the top of the list? We're going to deal with that now. Check it off.
 - i. If you're using a different QGIS version and don't have that field, in the Symbology window in Properties, click the **Add class** button



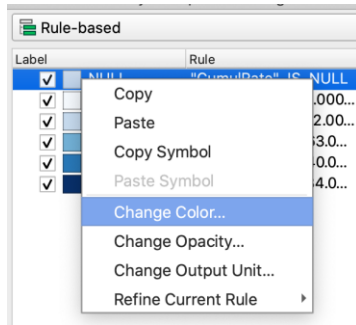
- 1.
 - ii. It should automatically add a 0.00 class at the top of the list
- g. Where it currently says Graduated at the top, select Rule-based from the drop-down:



- ii.
- h. **Double-click on the first entry under the Rule column**
 - iii. In the pop-up, change the label to NULL
 - iv. In the Filter, delete all but the field name (e.g. "CumulRate") and then type **IS NULL** in all caps
 - v. Click the Test button to make sure you get some features returned:



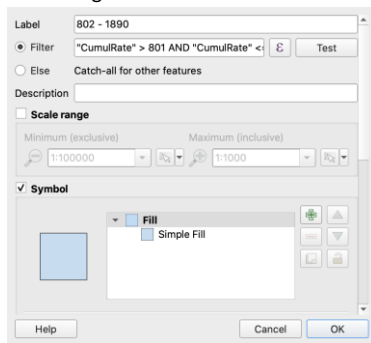
- vi.
 - vii. Click OK, then OK again
- i. **Change the colour**
 - viii. Right-click on the new NULL category and select Change Color...



1.
 - ix. Choose a medium grey (or something different than the colour scale you've chosen)
 - x. Click OK, then OK again
- j. Your first field may have something that looks like 0-1890. However, if we look at the attribute table, there are no 0 values – just NULL and then 802.

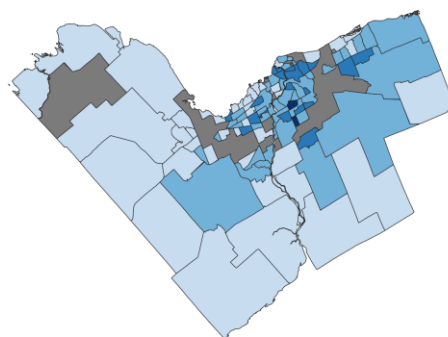
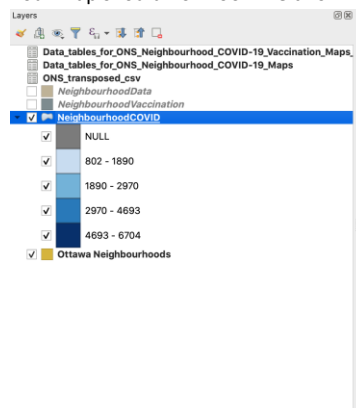
	_Rate12	cov_Numb12	cov_Rate13	cov_Numb13	cov_Popula	cov_OBJECT	CumulRate
7	essed	<5	Suppressed	<5	1041	53	NULL
8	essed	<5	Suppressed	<5	1426	60	NULL
9	essed	<5	Suppressed	<5	1990	46	NULL
10	essed	<5	Suppressed	<5	4722	32	802
11	essed	<5	Suppressed	<5	7787	20	804
12	essed	<5	Suppressed	<5	21164	58	967
13	essed	<5	Suppressed	<5	11317	61	1034

- i.
- k. Since it's not entirely accurate to say the range starts at 0, we'll change the label in the legend.
 - i. Go back into NeighbourhoodCOVID Properties and then Symbology
 - ii. Change the label to 802 – 1890
 - iii. We can also change the filter to "CumulRate" > 801 (instead of > 0)



1.
 - iv. Click OK and OK again

I. Your map should now look like this:



m.

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