THE GEO PROJECT
GEOSPATIAL EDUCATION & OUTREACH

GEO TUTORIAL

UNLOCK A HIDDEN POTENTIAL OF ONE-LINE EXPRESSIONS IN QGIS

Kate Grala
Krzysztof Raczynski
John Cartwright

Geosystems Research
Institute Mississippi
State University

MARCH 2024

UNLOCK A HIDDEN POTENTIAL OF ONE-LINE EXPRESSIONS IN QGIS

Kate Grala[1] (kgrala@ngi.msstate.edu)

Krzysztof Raczynski[2] (chrisr@gri.msstate.edu)

John Cartwright[2] (johnc@gri.msstate.edu)

[1] Northern Gulf Institute

[2] Geosystems Research Institute, Mississippi State University

REQUIRED RESOURCES

- QGIS (tutorial developed using QGIS 3.28 LTR; any 3.x version is acceptable)

FEATURED DATA SOURCES

- https://geo-trainingctr.maps.arcgis.com/home/item.html?id=0d6cb35c61b04e059bf631cb87a61fce

OVERVIEW

This tutorial aims to teach you how to compare values from multiple columns in QGIS. The main idea is to illustrate the integration of a comparison statement into a Field Calculator expression. You'll discover how to assess similarities between individual row values across several attribute columns. The presented workflow involves creating a new field and a comparison expression to populate the attribute cells with integer values representing true (1) or false (0) conditions. If the values across the evaluated columns are the same, the returned attribute value is 1; otherwise, it is 0. You don't need any programming background to complete this tutorial, just a basic familiarity with QGIS.

To spark your interest, imagine yourself in the following situation: The Mississippi Wetland Association has contracted your GIS consulting company to map wetland changes in coastal Mississippi counties. From an extensive request list, you have decided to first tackle a change detection assessment. This tutorial will guide you through the required steps to get this task promptly off your list.

STEP 1. DOWNLOAD, OPEN, AND REVIEW THE DATA

In this step, you will download the tutorial data, open the shapefile, and review its attribute table.

A. Download the GEO_Tutorial data using the link listed in the Feature Data Sources section. Save the zipped file to a location on your local drive you can easily find and unzip. Remember, there should be no spaces or special characters in the name of the folder where you save the tutorial data.

B. Create a new QGIS project and add your favorite basemap layer to provide a visual context. For example, you can use the Browser panel and select the OpenStreetMap available under the XYZ tiles.
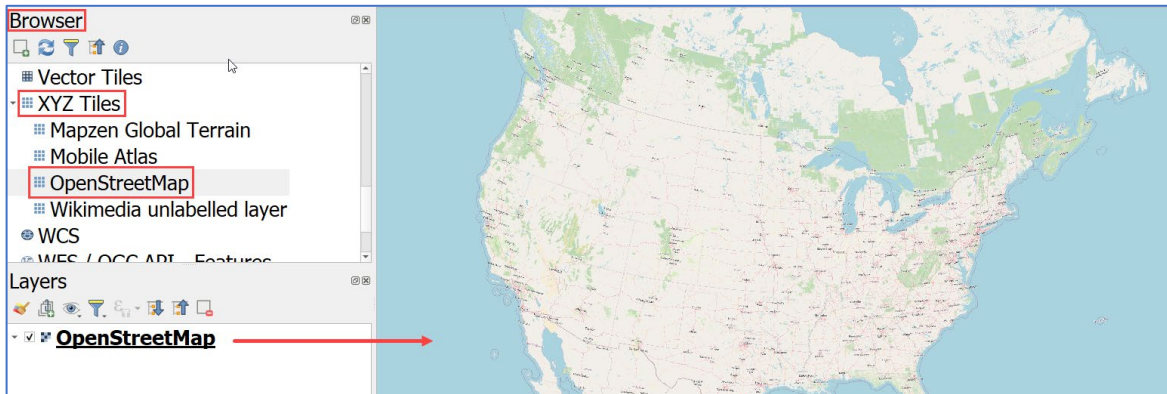
C. Insert the Woody_Wetlands_2001 shapefile from the GEO_Tutorial folder you unzipped in step 1A.
D. Zoom in to the point layer and ensure its correct position. The locations shown on the map cover the extent of coastal zone in Mississippi and represent the Woody Wetlands land cover class extracted from the 2001 NLCD Land Cover dataset.
E. Right-click on the Woody_Wetlands_2001 layer. Select the option to Open Attribute Table. Review the existing columns and check the number of features. Confirm that the table has three NLCD-related attribute fields and 2,000 records. Note that all attributes under the NLCD_2001 column represent a single land cover class: Woody Wetlands. Most attributes stored in the other two columns similarly represent the Woody Wetlands class. However, occasionally, they correspond to other land classes. For example, as shown in the figure below, some represent Emergent Herbaceous Wetlands.
F. Scroll down the table to evaluate the attributes stored in the NLCD_2011 and NLCD_2021 columns. As expected, a visual assessment of 2,000 records for multiple columns is ineffective, and you need to use a different method to accomplish this task.
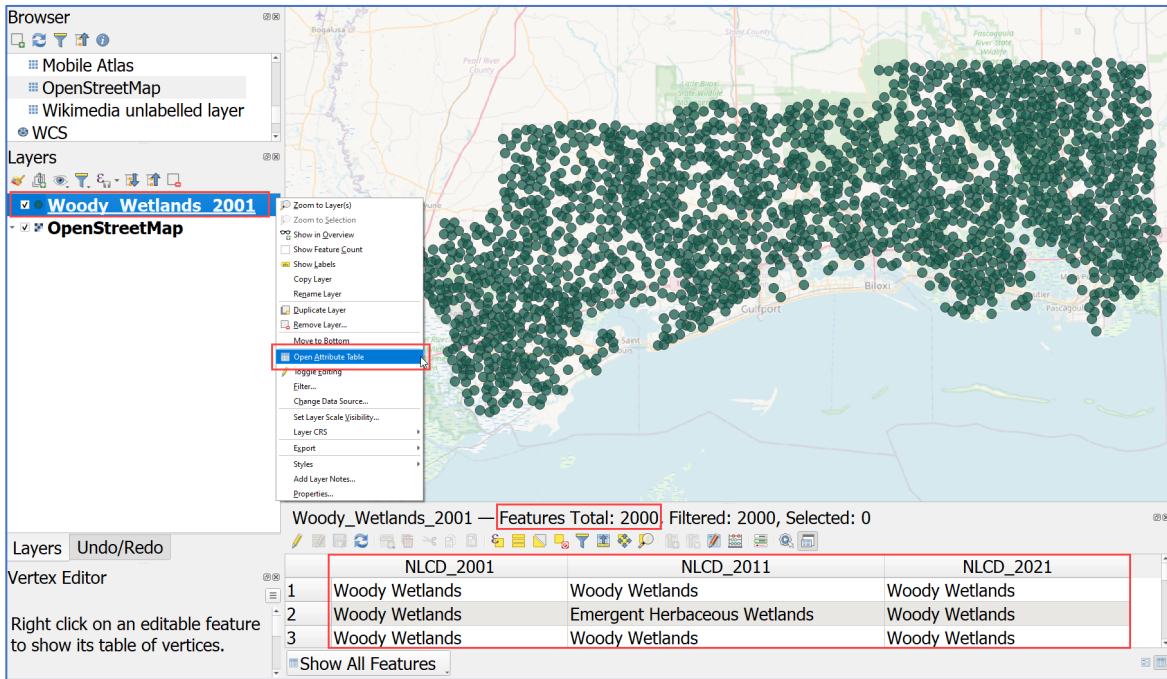
G. Take a minute and think how you would evaluate row values in the available **NLCD-related** columns to easily detect features that don't represent **Woody Wetlands** at all times during the analyzed period.

H. Don't close the attribute table since you will continue using it in the next step. **Save Project** 🖫 to ensure you don't lose your progress.

STEP 2. CONSTRUCT A COMPARISON EXPRESSION IN THE FIELD CALCULATOR

Hoping to uncover additional ideas, you posted the task challenges on the QGIS online forum. You received numerous helpful suggestions on how to compare the available NCLD-related fields and determine how many points classified as **Woody Wetlands** in 2001 represent the same land cover according to 2011 and 2021 classifications. You opted to automate this process by building a comparison expression in the **Field Calculator** tool. Below, you will discover a detailed description of how to achieve this, so keep reading!

A. To begin the process, start the **editing mode** 🖉 for the **Woody_Wetlands_2001** layer and open the **Field Calculator** 🧮. Fill out the required parameters, as described below. Ensure that no selection is present by verifying that the **Only update 0 selected features** option is grayed out.

B. Since you do not want to change any of the existing fields, verify that the box next to the **Create a new field** option is checked.

C. Name the output field **"Change"** and make sure to use the **Integer(32 bit)** as an **Output field type**. Note that the type of field you need to create relates directly to the type of attributes the calculation will return, and in this example, the attributes are an integer type. You can keep the default value for the **Output field length** option.

D. Place the cursor inside the **Expression** window and type in the following line of code:

```
("NLCD_2001" = "NLCD_2011") AND ("NLCD_2011" = "NLCD_2021")
```

E. Before you click the **OK** button, you need to be familiar with a couple more principles. The key idea behind the expression is to return a value of **1**, in case all **NCLD-related** fields are identical, and otherwise a value of **0**. This is commonly done with the conditional statement, but here, we will use a comparison statement instead. The comparison statements return Boolean values True (1) or False (0). More specifically, by incorporating the logical operator **AND** you will get a return value of **1** for true conditions (the NLCD-related fields are the same) and **0** for false conditions (the NLCD-related fields are different).

F. Note that you can type the entire expression, including the field names into the **Expression window** manually, but to eliminate potential syntax errors, it is easier to build it by double-clicking on the names shown under the **Fields and Values** menu. This way names of the attribute fields will be automatically placed between double straight quotes ("). If you use a single quote (') instead, this will indicate text, and the expression will not be able to reference attribute values and will ultimately fail. This is commonly overlooked syntax rule, so don't forget that in QGIS field calculations, single and double quotes can't be used interchangeably.

G. Note a series of buttons located directly under the **Expression** window. Test these buttons to find out how they can help you generate expressions.

H. If needed, correct the syntax errors. This should be evident as you will see a red text warning indicating that your **Expression is Invalid**. Use options described above to help you make the required corrections. When all the expression problems are resolved, a **Preview** of expected calculation result will appear. This means that you are ready to run your expression.

I. One last time, verify that your inputs correspond to the figure shown on the next page and click the **OK** button to run the calculation.

J. When completed, scroll down through the attribute table rows, and visually evaluate the results. If done correctly, a new field named **"Change"** is added to the table and attributed with values of **0** and **1**.

K. Assuming the operation was successful, it is important to save the edits ▣. At this point it might be a good idea to save the project ▣ as well. You may have noticed that the save icons are very similar and can be easily mixed up. The one small difference is a tiny red pencil at the corner of the save edits' icon.

L. **Select features using an expression** where the **"Change"** column values are set to the attribute indicating false conditions (**0**). Their spatial spread should be identical to the distribution of yellow points shown in the figure on the right. It is clear that majority of the inspected points (**1,849**) remained unchanged. Only **151** points have changed the land cover class during the

analyzed period. Now, you are ready to design the final map and share the results with your clients from the Mississippi Wetland Association.
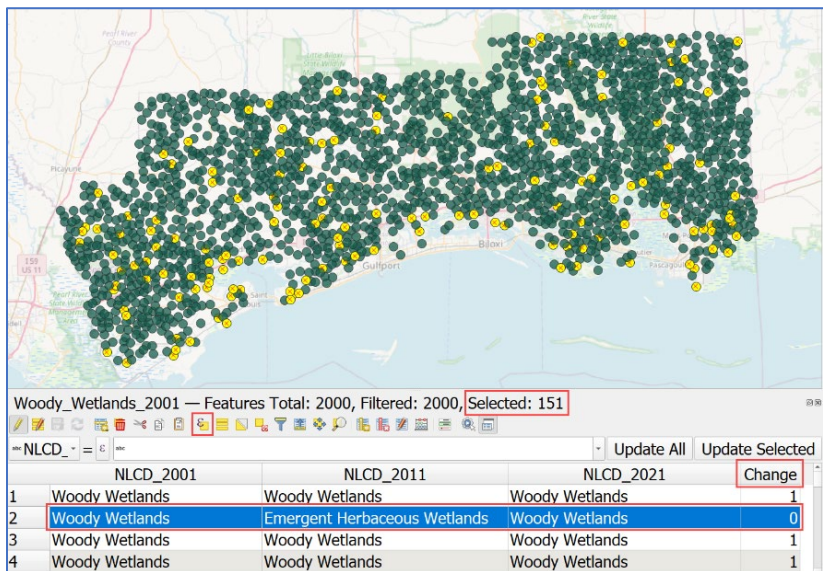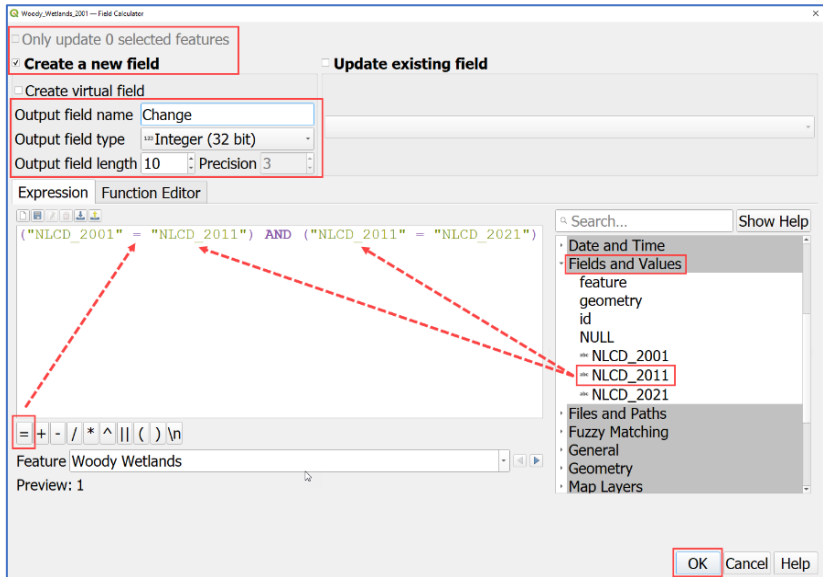
STEP 3. ALTERNATIVE SOLUTIONS

Although this is an optional step, it might be well worth your time. As you may expect, there are multiple correct ways to write an expression that would achieve identical results to the ones described above. For example, you could employ the **LIKE** operator to search for a specific substring or use the conditional statement. Listed below are a few examples of these alternative solutions you can test. To make things easier, there is no need to create a new calculation field. You can run these expressions on the already-existing **"Change"** column. Simply check the option next to **Update existing field** and select the appropriate field name from the drop-down menu. While testing, fill out the remaining parameters as described in step 2. Each time you run a new expression, make sure to verify your results and clear any potential selections you make.

A. The first solution relies on the **LIKE** operator and the **%** wildcard symbol representing any or none characters. The combination of the two is commonly used to look for a subset of text in a longer string. Note that the main structure of the expression is the same as the comparison statement described in step 2. Also, as mentioned earlier, double quotes (**"**) indicate names of the attributes, while single quotes (**'**) indicate text. In the example below, we are searching for the text **'Woody'** in a longer string listed in the attribute **"NLCD_2011"**. You need to know that the **LIKE** operator is case-sensitive. This means that searching for **'woody'** vs. **'Woody'** will not produce the same results. Additionally, our expression is designed to look for the text **'Woody'** with any characters before and after. Hypothetically, this means that evaluated strings such as **'This Woody Park'**, **'Woody Park'** or **'Woody'** are seen the same. Keep these concepts in mind while testing this expression.

```
("NLCD_2011" LIKE '%Woody%') AND ("NLCD_2021" LIKE '%Woody%')
```

B. The second solution starts with the **if()** operator, analogous to the **if()** function you may have seen in Microsoft Excel. The basic form is as follows: if(condition, value if true, value if false). Initially, we define the logical condition, then we provide a value returned when the condition is true (1), followed by a value when the condition is false (0). All the structural expression elements must be separated by a comma and enclosed in parentheses.

```
if("NLCD_2011" = "NLCD_2001" AND "NLCD_2021" = "NLCD_2001", 1, 0)
```

C. You can easily convert the expression shown above into a more elaborate solution. Multiple conditions can be combined into a longer series. For example, we can add a second condition accounting for a change detected in 2011 but no change in 2021 (restored to the original Woody Wetlands class). In that case, we would like to return a value of 2. This means, some rows that would originally turn 0, would now be marked as 2.

```
if("NLCD_2011" = "NLCD_2001" AND "NLCD_2021" = "NLCD_2001", 1,
if("NLCD_2001" = "NLCD_2021", 2, 0))
```

Once you view the results, you may wonder why some cases of 1 were not changed to 2 while running the second condition. That is because our expression is considered from left to right, till the first true outcome occurs. More specifically, this means that when the comparing algorithm already produced a value of 1 for some row, it will skip the second condition entirely. This approach allows you to share more specific information with your client, especially if they would be interested to learn whether the change in land cover was permanent.

We have reached the end of our GEO Tutorial on incorporating a comparison statement into the QGIS field calculator. You can use concepts presented here to explore similarities and differences among values in any multiple columns. Minor adjustments allow for the use of similar expressions when analyzing strings, dates, integers, and other numeric values. In addition to detecting land cover change, these types of calculations could be helpful when evaluating the results of various classification methods, assessing datasets from various sources, or identifying typos in revised text attributes.

You have successfully finished this tutorial, but hopefully, you are up for a challenge. For example, it might be a good exercise to create a similar expression using just two of the available NLCD-related fields or add an extra fictitious field and expand the expression to more than three fields. Use this tutorial as your guide and have fun unlocking the hidden potential of field calculator expressions on your own!