

Last saved on 12/31/2024 - 09:26

Is latest revision Yes

Current state Published

Join by Attributes (Part 1): One-to-One Joins

- View
- Edit
- Delete
- Revisions
- Clone
- Translate

Signed in as scheeva

Account settings

Sign out

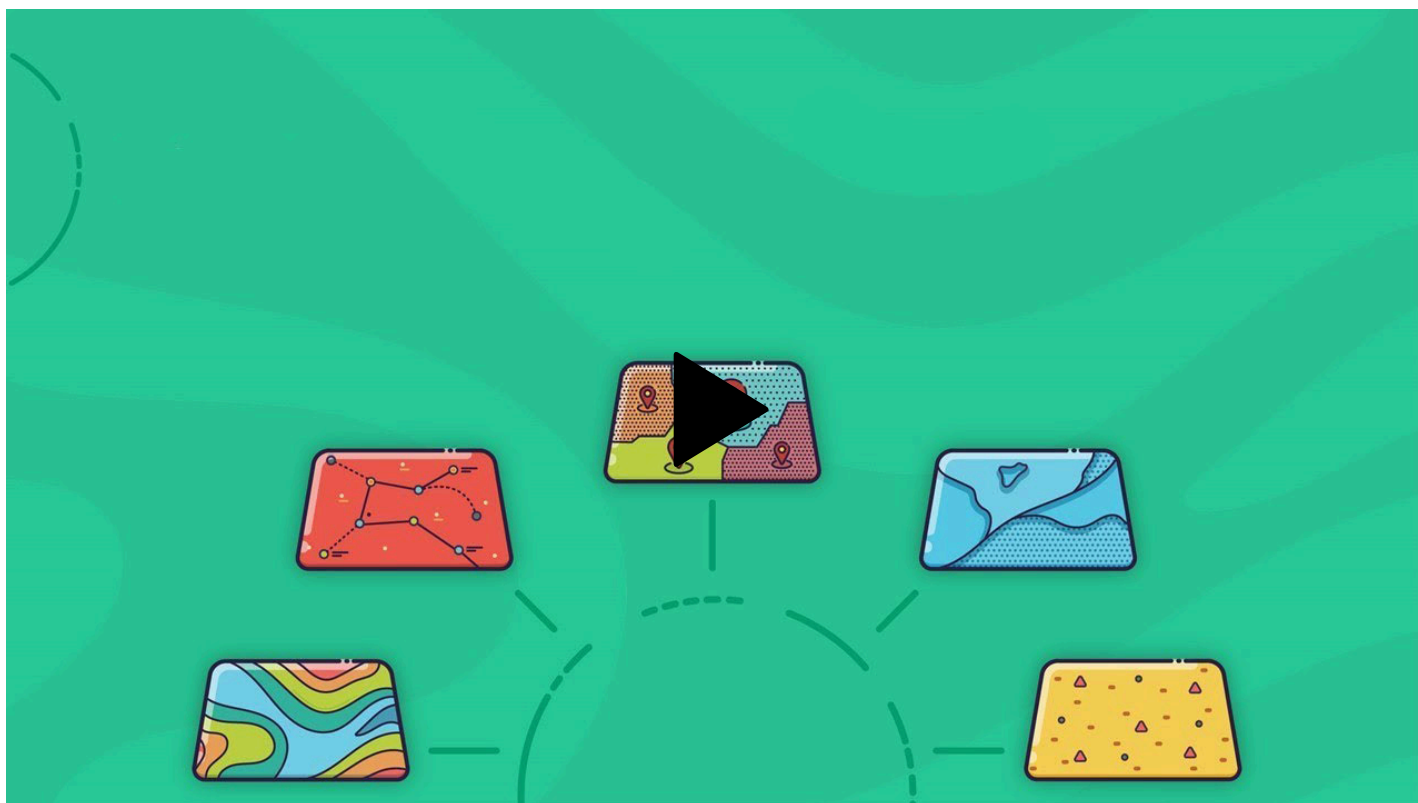
UUID Link: [uuid-link:node:94280eb4-3a85-4daa-923e-d71c4f59ef8f]

Catalogue number: **89200005**

Issue number: 2020012

Release date: November 20, 2020

QGIS Demo 12



▼ Join by Attributes (Part 1) - One-to-One Joins - Video transcript

(The Statistics Canada symbol and Canada wordmark appear on screen with the title: "Join by Attributes (Part 1) - One-to-One Joins")

So in this tutorial we'll introduce joining datasets by attributes, specifically linking tables to vector datasets for analysis and visualization. This is a powerful way to examine tabulated variables, linking them to vector geometries via common entries - in this case a column in the table and a matching field within the vector. There are two types of attribute joins, with slightly different procedures. Today we'll cover the first, the one-to-one join, where there is one row for each corresponding feature or geometry. For a successful join the entries must match perfectly. Thus, numeric identifiers are best due to complications with text such as special characters, spacing and case-sensitivities. Copying and matching entries between datasets is another method to improve the likelihood of a successful join.

For the tutorial we'll use the Population and Dwelling Highlight tables - downloaded previously. They are ideally formatted, as the join information is readily available and boundary changes between census collections have been accounted for – requiring no external formatting.

So we can load table datasets in to QGIS using the established procedures, double left clicking or dragging and dropping into the Layers Panel. So with the tables and corresponding boundaries loaded we'll open the Layer Property box of the Census Division layer which we'll use to demonstrate the procedures. We can use the Joins tab to perform one-to-one joins and to create our join click the Plus icon.

So the Join layer is the dataset that we'd like to join. And the join field is the specific column or field used to link the datasets. While the target field is the corresponding field within the vector containing matching entries, in this case the Unique Census Division Identifier.

We'll also check the following boxes, and specify the fields to join – as otherwise all fields are joined by default. Specifically we'll add the population and total private dwelling counts and percent change fields, as well as the land area, population density and two population rank fields. We'll remove the custom prefix to retain the original column names.

Now in the Source Fields tab we can see that the columns are joined temporarily and have been misattributed as text field types – a default in QGIS. Additionally, the joined field names exceed the limits of the file format we're using. So to permanently join the datasets we would have to export the dataset to a new layer. But to accomplish all three tasks at once we'll use the Refactor Fields tool.

In the Processing toolbox we can search the tool and double-left click to open it. So we'll close the tool description. And using the drop-downs we can change the field types for the joined columns.

Specifically we'll use integer for the count variables, which will be whole numbers, while for the percent change field, land area and population density fields we'll use double - the equivalent for the decimal number field type. Then we need to specify the parameters - using a length of 12 and, for our double columns, a precision of 2. Finally we'll rename our fields to abbreviated headers less than 10 characters in length - the limit of the shapefile format we're using. So feel free to use abbreviations that are most interpretable to you.

Now we'll save the file - providing it an output directory and name. I am storing it in the Joins demo folder and calling it JPopCD for Joined Population - Census Divisions. Once complete, refresh the Browser panel and load the joined dataset into the Layers Panel.

So with our fields correctly attributed - we can now visualize the joined variables, applying a graduated symbology to the numeric fields - specifically in this case we'll use the percent population changes field. We'll select an appropriate colour ramp for visualization and for the Census Division layer we can use Pretty Breaks to establish the value ranges for visualization - clicking OK.

Now we can examine the spatial variations in the joined variables across Divisions within Canada.

We recommend practicing these procedures on your own by repeating with the Census Subdivisions layer to familiarize yourself with the workflow. Once again specify the Join layer and the join information in this case using the Census Subdivision unique identifier to link the datasets.

For the demo, I've loaded a Joined Subdivision layer I created earlier. Due to the data distribution at this level Quantile (Equal Count) was the method used to establish the break values for visualization. But toggling back and forth between the layers demonstrates how these procedures enable variables of interest to be assessed at multiple scales relatively quickly and easily.

And on that note - if we want to determine trends at a broader level we can use the Aggregate tool. So we specify the layer to be aggregated, and then the Group by Expression drop-down enables us to select a field to use in aggregating both the geometries and attributes of the layer. Since we are gonna use the Provin...Unique Provincial Identifier, the first three Census Division fields are redundant - so we'll remove them. We can specify the operator applied in aggregating the fields from the drop-down. So we'll use First Value for the first two text fields and for the percent changes we'll use Mean. Finally we'll remove the Population Rank fields. And we'll save this layer in the same directory calling it JAgPrPop for Joined Aggregated Provincial Populations. Once it's complete we'll refresh the Browser Panel once again loading the layer.

And now we'll recalculate the percent population change field with the Field Calculator. So we'll add a bracket, and then subtract the Population in 2016 from that in 2011, close the bracket and divide it by the baseline in this case being the population in 2011 multiplied by 100. Once again we can use the symbology tab to visualize the variables at three separate levels.

So the final item I'd like to discuss is the Graphics drop-down in the Processing Toolbox, which can be used to quickly assess data distributions, such as the feature counts in different categories using

histograms and boxplots, or variable relations between joined variables – using the Scatterplot tool. So we could specify the 2016 Population as our independent variable and the Total Private Dwellings in the same year as the dependent. Once run, we can click on the Hyperlink in the Results Viewer to examine the relation. Unsurprisingly there is a very strong positive relationship between the total population and number of private dwellings.

So that concludes the one-to-one join - when there is one entry for each corresponding feature - using the Joins tab. The procedures can be applied to join tables or link two or more vector datasets together - enabling the visualization of variables of interest. We also learned how to apply the Refactor Fields tool to alter field types of joined data and use the Aggregate tool to examine trends at broader levels. We could iterate these procedures to combine multiple variables for examining changing variable relations between locations or over time. In the next demo we'll examine how to perform a one-to-many join.

(The words: "For comments or questions about this video, GIS tools or other Statistics Canada products or services, please contact us: statcan.sisagrequestssrsrequetesag.statcan@canada.ca" appear on screen.)

(Canada wordmark appears.)

Date modified:

2024-12-31