THE GEO PROJECT
GEOSPATIAL EDUCATION & OUTREACH

www.geoproject.msstate.edu

# GEO TUTORIAL

#ArcGIS PRO #QGIS

## JOINING TABLES WITH SPATIAL DATA

Krzysztof Raczynski
Flavia Xavier
John Cartwright

Geosystems Research Institute
Mississippi State University

MAY 2025

*This work was supported through funding by the National Oceanic and Atmospheric Administration Regional Geospatial Modeling Grant, Award # NA19NOS4730207.*

*GEOSYSTEMS RESEARCH INSTITUTE, MISSISSIPPI STATE UNIVERSITY, BOX 9627, MISSISSIPPI STATE, MS 39762-9652*

*The Geospatial Education and Outreach Project (GEO Project) is a collaborative effort among the Geosystems Research Institute (GRI), the Northern Gulf Institute (a NOAA Cooperative Institute), and the Mississippi State University Extension Service. The purpose of the project is to serve as the primary source for geospatial education and technical information for Mississippi.*

*The GEO Project provides training and technical assistance in the use, application, and implementation of geographic information systems (GIS), remote sensing, and global positioning systems for the geospatial community of Mississippi. The purpose of the GEO Tutorial series is to support educational project activities and enhance geospatial workshops offered by the GEO Project. Each tutorial provides practical solutions and instructions to solve a particular GIS challenge.*

# JOINING TABLES WITH SPATIAL DATA

Krzysztof Raczynski [1, 2, 4, 5, 6, 8] chrisr@gri.msstate.edu
Flavia Xavier [3, 7] flavia.xavier@msstate.edu
John Cartwright [7, 9, 10, 11] johnc@gri.msstate.edu

Geosystems Research Institute
Mississippi State University

CRediT: 1: Conceptualization; 2: Methodology; 3: Verification; 4: Resources; 5: Data Curation; 6: Writing - Original Draft; 7: Writing - Review; 8: Visualization; 9: Supervision; 10: Project administration; 11: Funding acquisition

## REQUIRED RESOURCES


- ArcGIS Pro 3+; QGIS 3+ (both programs are covered in this tutorial)

## FEATURED DATA SOURCES

- Click here to access dataset used in this tutorial (79.04 MB).

## OVERVIEW

Not all data is geospatial. In fact, the abundance of spatially referenced data observed today is a result of the tripling of the market in the last 10 years. This means that some data is still provided in the form of regular tables. This can be attributed to various factors, with the most common ones being an early start date of data collection (long-term datasets) or the size of the data itself. However, this does not preclude spatial analysis or the need for a lot of work to make data spatially related. In this tutorial, you will learn how to combine non-spatial data provided in the form of a table with a spatial data layer for analysis and mapping purposes.

Imagine that you are working for the Census Bureau, and you are tasked with creating a choropleth map of population change between 2010 and 2023 by county over the United States. You know that the Bureau provides all the necessary data in the online portal. However, the data is in tabular format. You decided to use this data with Census core TIGER data. In this tutorial, you will learn how to link the two datasets together for purposes of spatial analysis and map creation.

## DATA

To begin with, download the data available in the **Featured Data Sources** above and add it to a new project in the software of your choice (ArcGIS Pro and QGIS are covered in this tutorial). You can also try and use the links provided in the *Overview* section to find the necessary data yourself. To do so, you will need a table with the population information for 2010 and 2023, as well as a shapefile of *US Census counties* (Fig. 1). The download consists of five main features: two population data tables, two metadata files detailing column usage, and a

*shapefile* set for the US counties spatial data. The spatial dataset contains 3,235 features. Note that all files are packaged in a .zip file and must be extracted before use.

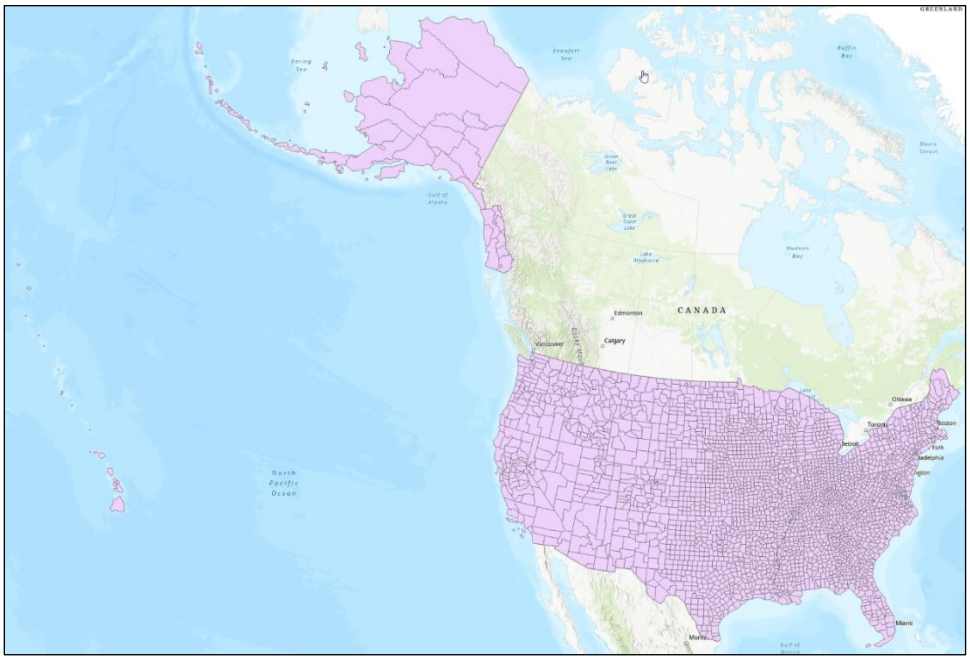


***Fig. 1.*** *Input data presenting counties across United States*

## ANALYZING THE PROBLEM

To calculate the change in population, we need to subtract values of the population in 2023 and 2010. Then, to create a Choropleth map, we need to assign this value to each county. Although the first part can be done outside of GIS software, the entire process can be done using it without any other processing solutions.

Practically any tabular data can be joined with spatial information, as long as one condition is met. Each data set must have a common attribute (column) that describes the same objects. This can be any information: a number, text, or character combination, as long as it is unique; that is, each information fits only one object. For example, if only the name of the city is provided, it can't be used, as there are cases where there is more than one; for example, Franklin in the US is the name of 31 different cities. However, instead of a city name, a zip code can be used, as this is a unique identifier. In our case, the data provided contains the name of the geographic area, but even more conveniently, tabular data contains a unique ID assigned to each case in the *GEO_ID* column. The same ID values are assigned in the spatial data layer in the *GEOIDFQ* attribute (the names of the attributes do not have to match, but the values do). With data prepared like that, we can join all the needed data.

## ARCGIS PRO JOIN

With the *tl_2024_us_county* layer and the two main data tables (*ACSST5Y2023.S0701-Data* and *ACSST5Y2010.S0701-Data*) already imported into your project, you may notice that tables, as non-spatial data, are not displayed on the map, and the list says *Standalone Tables* (Fig. 2). Now that all the data is in place, follow the steps:

A. Right-click on the *counties* layer in the *Contents* pane and select ***Joins and Relates*** from the submenu. Then choose ***Add Join***.
B. A new window will pop up with the join setup (Fig. 3). Here we need to set the following:
   a. for *Input Table,* set the counties layer as ***tl_2024_us_county***;
   b. for *Input Field,* select ***GEOIDFQ,*** which is attribute that contains the unique identifiers for each county;
   c. for *Join Table* choose the census population table from 2010: ***ACSST5Y2010.S0701-Data***

d. set the *Join Field* as ***GEO_ID***;
e. mark the ***Keep all input records*** setting.

C. You can test the join before performing the operation by pressing ***Validate Join*** button. This will display a small window with a message indicating the number of matching rows in each table. If you selected the correct attributes, you should see: "*The input table has 3235 and the join table has 3222 records*". Otherwise, the message will say "*WARNING 003237: Join had no matches*" indicating there are no matching pairs of values.
D. Once the validation is finished, click *OK* button.
E. Open the attribute table for the *counties* layer by right-clicking on it in the *Contents* pane and selecting *Attribute Table*. Notice that, in addition to the original 20 attributes, the reminder of the table contains data from *ACSST5Y2010.S0701-Data* table (see Note on ArcGIS Pro Table Join below on limiting the number of columns joined).
F. Now repeat the procedures A-E for the *ACSST5Y2023.S0701-Data* table.

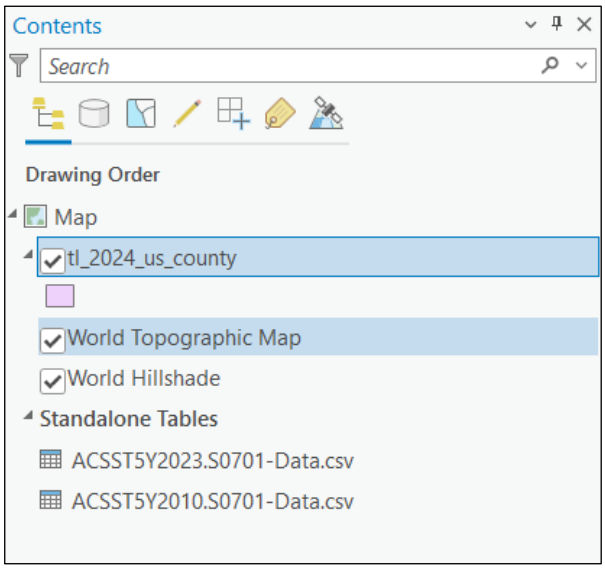


***Fig. 2.*** *Nonspatial data tables in ArcGIS Pro are displayed in the Contents pane as Standalone Tables.*

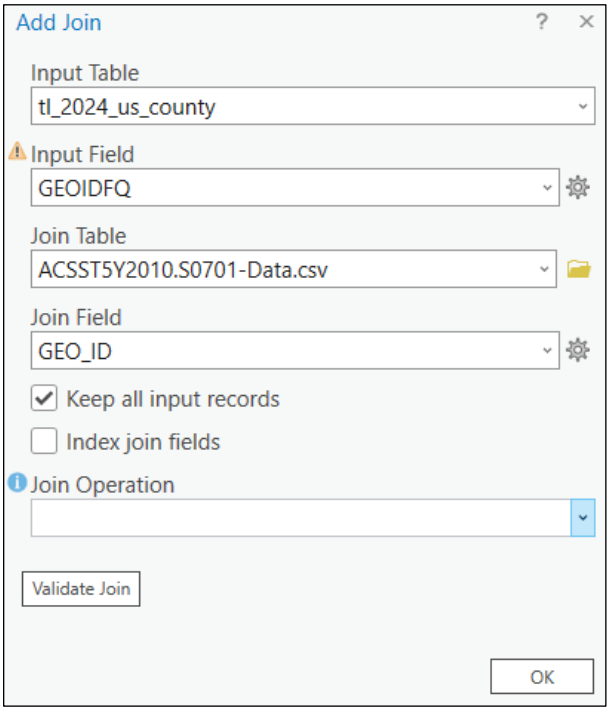


***Fig. 3.*** *ArcGIS Pro settings in batch process allow to choose the parameter that will change in each batch.*

Now that we have successfully joined both tables to our county data, we can compute the population change and create the choropleth map.

Open the *attribute table* of the *counties* layer, to which both tables are joined. Click *Add* to create a new field and name it *pop_change*, with the Data Type *Double*. Click *Save* in the *Fields* tab to apply the changes. Now, you are ready to calculate the population change in the newly created field.

In the *attribute table,* click the *Calculate* tool. A new window will be opened, where you can set the details for the calculation in the field we have just created. Before we can start constructing an expression, we need to verify which fields we will be using. Thanks to the metadata files, a quick peek tells us that the argument ***S0701_C01_001E*** contains information on the total population. This information is exactly what we need.

Let's get back to the *Calculate Field* tool to prepare the expression. You can use the list to select corresponding field names or just type them in. To indicate to ArcGIS PRO the use of a field value, we use the exclamation (**!**) character. Also, notice how, if you chose an attribute from a joined table, it does not just indicate the name of the attribute, but it is preceded by the name of the joined table separated by the dot. To present the percentage change in the population, we need to subtract the population in 2010 from the population in 2023 and then divide the result by the population in 2010 multiplied by 100 (i.e., pop_change = (pop2023 – pop2010)/pop2010 * 100). This will give us the percentage change in the population in relation to 2010. The expression should look as follows:

```
(!ACSST5Y2023.S0701-Data.csv.S0701_C01_001E! -
!ACSST5Y2010.S0701-Data.csv.S0701_C01_001E!) /
!ACSST5Y2010.S0701-Data.csv.S0701_C01_001E! * 100
```

Note: If the above formula fails, modify it to manually convert the joined fields from text to numeric values:

```
(int(!ACSST5Y2023.S0701-Data.csv.S0701_C01_001E!) -
int(!ACSST5Y2010.S0701-Data.csv.S0701_C01_001E!)) /
int(!ACSST5Y2010.S0701-Data.csv.S0701_C01_001E!) * 100
```

If you see a warning message stating that certain rows were set to NULL due to a *TypeError*, this means that some of the joined fields contained NULL or non-numeric values. This affects only those specific rows, and you can proceed with creating a map showing the changes of population per county once the calculations are finished (Fig. 5).

## NOTE ON ARCGIS PRO TABLE JOIN

In the previous section, we learned how to join a table with a spatial layer. This type of join exists only in the memory of the program, meaning the values from the table are not permanently assigned to the data. Additionally, this option only allows you to join all fields at once. Sometimes, if many attributes are in the table, like in our case, this might be bothersome. If you would like to join a specific column or make a join permanent, then you should use *Join Field* (from *Data Management Tools*; you can look for it in the *Geoprocessing* pane). The setup will be the same, with the difference that in the *transfer fields*, you can choose which attributes will be added to the *input table*. Note that this is a permanent join that alters the original table.

Alternatively, once the tables are added to the project, you can open them and choose *view fields* from the menu. There you can deselect all redundant columns and leave only a selected few that will be added to the layer. This can significantly lower the number of columns joined.

## QGIS JOIN

After importing the *tl_2024_us_county* layer to a new project, import the two main data tables: *ACSST5Y2023.S0701-Data* and *ACSST5Y2010.S0701-Data* (you can just drag and drop the files from the folder to the *Layers* panel). Notice that the tables, as non-spatial data are not displayed on the map, and the icon on the list shows a table, without the option to turn off/on the layer. Now that all the data is in place, follow the steps:

A. Right-click on the *counties* layer and select *Properties* from the context menu.
B. Change tab to *Joins*.
C. On the bottom of the window, select the plus icon to open *Add Vector Join* window.
D. In the newly opened window, we need to setup the join parameters (Fig. 4):
   a. set the *join layer* to ***ACSST5Y2010.S0701-Data***;

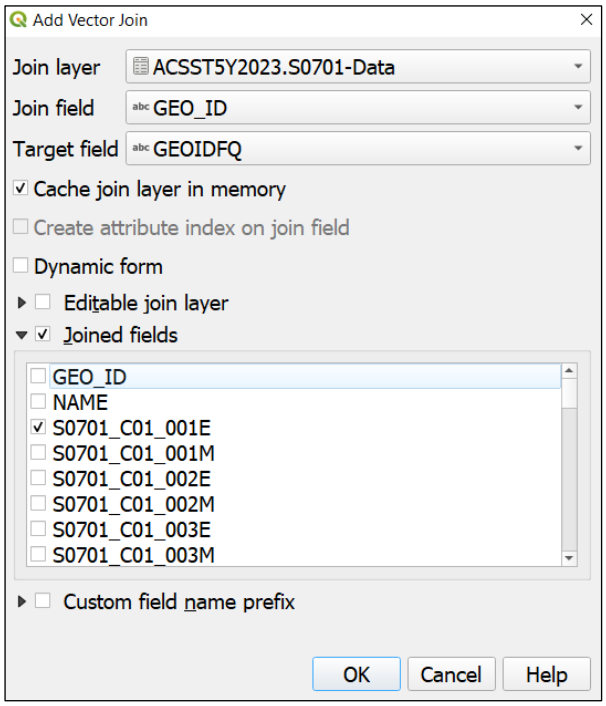


***Fig. 4.*** *Join settings in QGIS allows to define table columns to be joined as attributes to the spatial layer.*

b. set *join field* to ***GEO_ID*** – this is the column in the table that contains the identifiers;
c. set the *target field* to ***GEOIDFQ*** – this is the attribute in the spatial data, that contains matching identifiers;
d. to limit the number of joined fields to one, representing the total population, check the *joined fields* setting and select ***S0701_C01_001E*** attribute from the list. This will ensure that only a single column is added to the table;
e. once all the settings are ready, click *OK*. This will bring you back to the layer properties, where information on the joined layer is displayed. Now, you can repeat the procedures a-e for the 2023 table.

E. Once both tables are added, click *OK*.
F. To calculate the percentage change in population open *Attribute Table* (right-click the *counties* layer and select *Open Attribute Table*). Click on the *Open Field Calculator* icon on the ribbon of the new window.
G. Define the following parameters:
   a. *output field name*, e.g. = ***pop_change***;
   b. *output field type* = ***Decimal number (real)***;
   c. in the *expression* box, type field names or use the middle panel and locate them under Fields and Values tab, as follows[1]:

```
("ACSST5Y2023.S0701-Data_S0701_C01_001E" –
"ACSST5Y2010.S0701-Data_S0701_C01_001E") /
"ACSST5Y2010.S0701-Data_S0701_C01_001E" * 100
```

Click OK:

Use *Style* properties to create a final map based on the attribute we have just computed (Fig. 5).

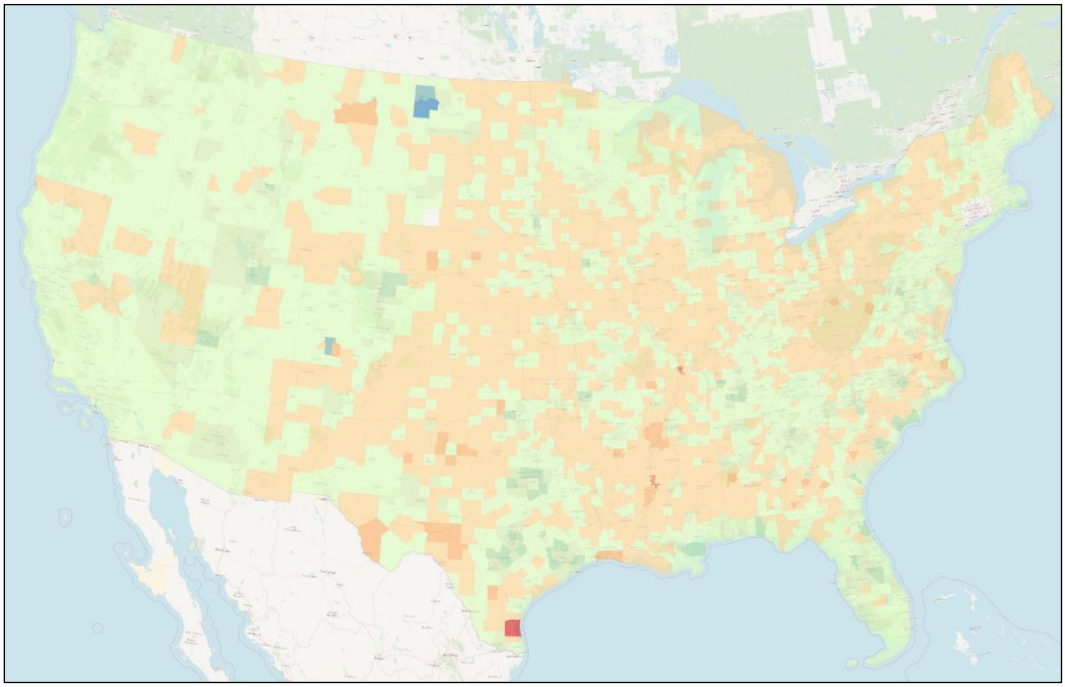

***Fig. 5.*** *Part of the results, over the contiguous United States.*

[1] If you see red warning in result preview saying the syntax is invalid after copying the formula, manually replace characters like ' – ' (minus sign) or ' / ' (division sign) using your keyboard. Sometimes styling can affect how the software sees the character and incorrectly interprete it.

## CONCLUSION

You have completed this tutorial! This tutorial demonstrated how to join non-spatial tabular data with spatial data using unique identifiers, thus enabling meaningful spatial analysis for visualization. By linking Census population data to geographic county boundaries, we calculated population change and created a choropleth map. This process highlights the importance of data integration in GIS, allowing for deeper insights from otherwise disconnected datasets. Understanding how to perform table joins ensures that analysts can work with a variety of data sources, thus improving decision-making and mapping capabilities. Whether using ArcGIS Pro or QGIS, these techniques are essential for leveraging spatial relationships in real-world applications.