QGIS: Improving Table Joins with CSVT Files

QGIS 3.22

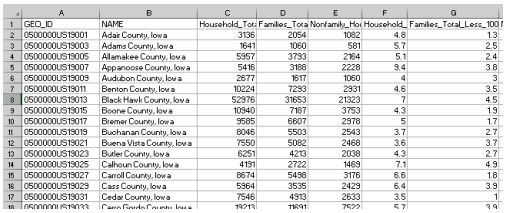
Welcome to the Essential GIS Task Sheet Series. This series supplements the Iowa State University Extension and Outreach Geospatial Technology Training Program's workshops and short courses by providing quick and easy instructions for performing a variety of mapping, data science, analysis and data visualization tasks.

A CSVT file contains a single line of text listing the field types used in a corresponding CSV file. When adding CSV files to QGIS you may notice all fields are treated as strings (text data); creating a CSVT file will help QGIS recognize the correct field types. Common field types you will come across are: string, for text data; integer, for whole numbers; real (sometimes called double), for numbers with decimals; and boolean, for yes/no or true/false data with only two options. This task sheet will teach you how to manage your comma separated value (CSV) files and improve joins by creating a CSVT file.

1. Getting Started & Downloading the Data

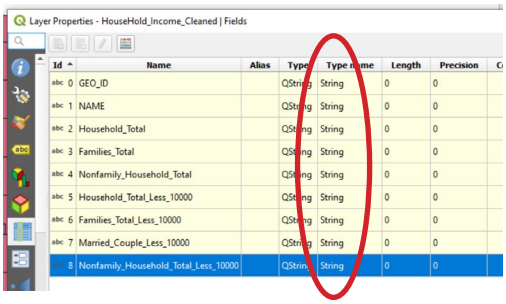
a. To download the data used in this task sheet, navigate to: https://isueogtp.github.io/GISTaskSheets/TaskSheetData/GISTP0030.zip.

b. When the download is complete, you will need to unzip the **GISTP0030** folder in order to access the file. The folder contains a CSV file and a shapefile of Iowa counties.



c. Open the **HouseHold_Income_Cleaned.csv** file in Microsoft Excel or your preferred spreadsheet application. Observe the data and notice it is a mix of text, whole numbers, and decimal numbers. Leave the spreadsheet open and continue to the next steps.

d. Open **QGIS**. Create a new project, and add the **Counties.shp** file.

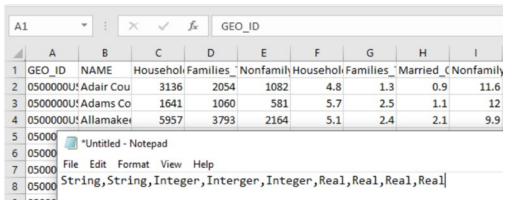


e. Add **HouseHold_Income_Cleaned** to QGIS, and open the **Layer Properties**. In the **Fields** tab, notice the **Type name** column says each field is a **String**. When a data type is incorrect, it can be difficult to join or symbolize.

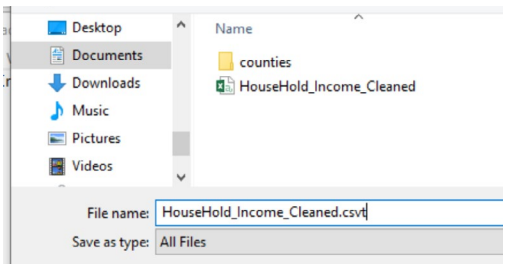
f. Remove the **HouseHold_Income_Cleaned** layer from QGIS.

2. Creating a CSVT File

a. Open a text editing application (Notepad on Windows or TextEdit on Mac) and arrange the windows such that both the text editor and the CSV file are visible.

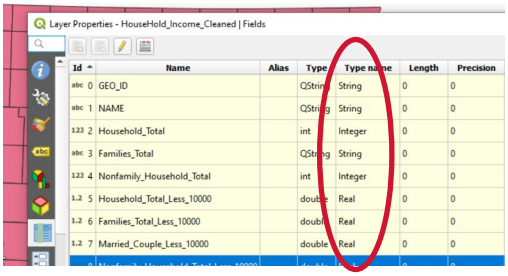


b. For each column in the spreadsheet, type the intended data type in your text editor, separating each type with a comma to match the example picture. *Notice: GEO_ID has both numbers and letters making that column string data rather than an integer or real number.*



c. When you finish entering the data types, save the file in the same location as **HouseHold_Income_Cleaned.csv**. Name the file **HouseHold_Income_Cleaned.csvt**.

d. Add the **HouseHold_Income_Cleaned.csv** file to **QGIS** again. This time, QGIS will utilize the CSVT file and display the data types accordingly. Open the **Layer Properties** for **HouseHold_Income_Cleaned Fields** and select the **Fields** tab. Now the **Type name** column will have the correct data types for all of the fields. *Note: In order for this process to work, the CSVT file must have the same name and be located in the same folder as its corresponding CSV file.*

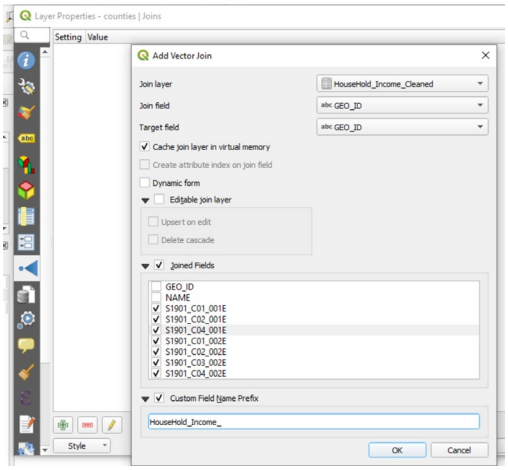


3. Joining Data

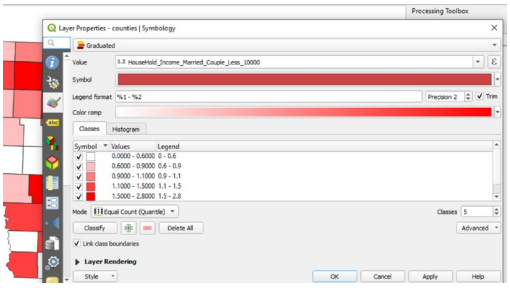
a. In QGIS, right-click on the **Counties** layer, choose **Properties**, and select the **Join** tab. Click the **Green Plus** icon (located at the bottom of the window) to open the **Add Vector Join** window. Set **Join Layer** to **HouseHold_Income_Cleaned**, **Join Field** as **GEO_ID,** and **Target field** as **GEO_ID**.

b. Make a few more adjustments by checking the box for **Joined Fields**, this will control what new fields will appear after the join. Check all the boxes except **GEO_ID** and **NAME**. Also check the **Custom Field Name Prefix** box, and set the name to **HouseHold_Income_**. Click **OK** to close the **Add Vector Join** window.

c. Click **OK** again to complete the join. The **Layer Properties** window will close.



d. Right-click the **Counties** layer and open the **Attribute table** to observe the results of the join. Columns from the **HouseHold_Income_Cleaned** file have been added to the **Counties** attribute table, and the text **HouseHold_Income_** has been added to the beginning of each column name.

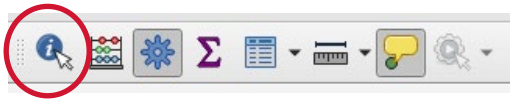


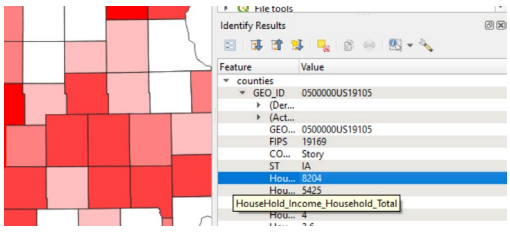
4. Symbolizing Data

a. Right-click on the **Counties** layer and open the **Layer Properties** window. Select the **Symbology** tab. Set the top button to **Graduated**, then set **Value** to **HouseHold_Income_Married_Couple_Less_10000**. Select a **Color Ramp** such as **Reds**. Set the **Mode** to **Equal Count (Quantile)** and click the **Classify** button. Finally, click **OK** and close the **Layer Properties**.

b. Select the **Identify Features** tool from the **Attributes Toolbar**.

c. Click on a county on the map. It will highlight in red and the **Identify Results** panel will open to display attributes for the selected county.





Contact: Jay Maxwell, Data Analyst, and Professor Christopher J. Seeger, PLA, GISP can be reached at geospatial@iastate.edu. Additional task sheets and information about the Geospatial Technology and Spatial Data Science Programs are available at www.extension.iastate.edu/communities/gis.