Why would you join tables?

Let's say that you want to make a map showing the number of people who voted for each political candidate in an election. You have a table showing each county with the number of voters for each candidates, but there is no spatial coordinates attached to the data. E.g. no way to visually represent this in GIS software. Never fear! We know that it is easy to download boundary datasets for counties, states, countries, towns, etc. from various online sources. We can easily join the non-spatial table to these spatial datasets for visualization.

1. **Download GIS dataset of boundaries (towns, states, etc.)**

There are many different sources for boundary datasets:

- Natural Earth (https://www.naturalearthdata.com/downloads/): pick either large, medium, or small scale depending on the scale of your map, then click on the "Cultural" button. **Good for countries and administrative boundaries within countries.** Note: these are not always "neutral" representations of the boundaries, particularly for disputed regions (e.g. Crimea is displayed as part of Russia, although it is considered part of Ukraine).
- US Census Tiger files (https://www.census.gov/cgi-bin/geo/shapefiles/index.php): good for **states, counties, towns, school districts, zip codes, Census tracts and block groups**. Only for the United States. These are usually downloaded for the entire state, even if you only have a data table for one particular city, etc.

Or generally, you can search "*study area* boundary GIS data" and see what results come up.

You will likely want to download the data as a Shapefile. The data will be downloaded to your computer as a <u>zip file</u>- a compressed folder made up of 3-8 different files. All of these individual files come together to make one Shapefile. Make sure to right-click on the zip file name and select 'Extract All' to unzip it before use (or on a Mac, double-click into the folder).

2. **Bring boundary dataset into ArcGIS Pro**

In ArcGIS Pro, click on the Add Data button in the upper ribbon.
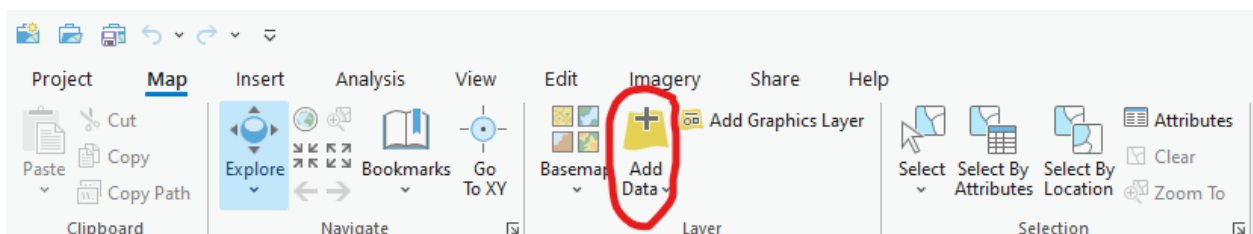
The layer should be displayed both on the map and in the Contents pane (at left). Right-click the layer's name in the Content pane and choose Attribute Table.

Take a look at the different field names (columns) inside the table. There will likely be one containing the name of each boundary. Depending on where your data came from, you might also have an ID number (usually GEOID for Census data). See the example table screenshot on the next page.

	FID	Shape	STATEFP	COUNTYFP	COUSUBFP	COUSUBNS	GEOID	NAME	NAMELSAD	LSAD
1	0	Polygon	33	013	19460	00873583	3301319460	Dunbarton	Dunbarton town	43
2	1	Polygon	33	013	27380	00873600	3301327380	Franklin	Franklin city	25
3	2	Polygon	33	013	60020	00873696	3301360020	Pembroke	Pembroke town	43
4	3	Polygon	33	013	00660	00873528	3301300660	Allenstown	Allenstown town	43
5	4	Polygon	33	013	06260	00873547	3301306260	Boscawen	Boscawen town	43
6	5	Polygon	33	013	06500	00873548	3301306500	Bow	Bow town	43
7	6	Polygon	33	011	01700	00873533	3301101700	Antrim	Antrim town	43
8	7	Polygon	33	011	31940	00873613	3301131940	Greenville	Greenville town	43
9	8	Polygon	33	011	44580	00873655	3301144580	Lyndeborough	Lyndeborough town	43
10	9	Polygon	33	011	51940	00873681	3301151940	New Ipswich	New Ipswich town	43
11	10	Polygon	33	009	07700	00873552	3300907700	Bristol	Bristol town	43
12	11	Polygon	33	009	74740	00873730	3300974740	Sugar Hill	Sugar Hill town	43
13	12	Polygon	33	009	61060	00873698	3300961060	Piermont	Piermont town	43

3. Find the linking key field in between your two tables

In order for tables to join, they have to have a common field ("linking key") between them. The field could have different names between the two tables (e.g. "NAME" and "County_Name"), but it needs to contain IDENTICAL information. We will usually use the ID or Name field for this purpose.

If using the Name field, pay special attention to formatting between the two tables. If there are any slight differences in names (e.g. "Dunbarton" and "Dunbarton town"), the computer will not recognize these values as identical and will not be able to join those records. If there is added punctuation in one, or extra spacing, this will also cause issues. You may need to carefully go through your tables and edit names to be identical for this to work.

For this reason, people often use ID numbers as the linking value. If using census data, the IDs can often be different between table and GIS dataset (e.g. "15000000US3301306500 vs. 3301306500). See the handout "Census Data Download Guide" for more information about how to do this.

Take a note of the names of the fields for both the GIS boundary layer and the non-spatial table that contains this linking key.

4. Join your tables

In the Contents pane at left, right-click on your GIS boundary layer and select Joins and Relates -> Add Join. In the tool window, leave the input table as your boundary layer and set the join table to be your CSV. In the input field, specify the name of the column containing the linking field for your GIS boundary layer. In the join field, specify the name of the linking key column for your non-spatial table.

Click on 'Validate Join' and scroll to the bottom. If it works, you should see a message like 'Join succeeded' with the number of matching features. Don't worry too much if not all features

matched, especially if your table and boundary data are at different scales (e.g. all counties in state vs. all counties in one city).

If the join doesn't work, this is likely due to formatting issues in the linking key, or perhaps incorrect data types (e.g. one field of ID numbers is type 'integer' while the other is type 'text'). Make sure that your linking key values definitely match between tables. Hover your cursor over the names of your linking key fields and wait until a pop-up about that column appears. Check the value next to Type. If it is different between the two fields, this could be a problem- especially if one is Text and the other numeric (Double/Integer).

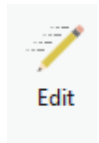| GEOID	NAME
330	**GEOID**
330	
330	Type: Text (10)
330	Default: <Null>
330	Read-Only: No
330	Nullable: No
330	Indexed: No
	Required: No
330	

If doing this for your Excel table, this is an easy fix- open the file in Excel, select all values in that field, right-click and choose Format Cells, then set the Category to the correct data type.

Edit

If doing this in ArcGIS Pro, you could add a new field into the table (in your attribute table, click Add (if this is grayed out, go to the Edit tab in the upper ribbon and click the big pencil button first

⊞ tl_2020_33_cousub ✕

Field: ⊞ Add | Calculate   Selection: ▦ Select By Attributes ⊕ Zoom To ▦ Switch ▤ Clear ▦ Delete ▤ Copy

	FID	Shape	STATEFP	COUNTYFP	COUSUBFP	COUSUBNS	GEOID	NAME	NAMELSAD
1	0	Polygon	33	013	19460	00873583	3301319460	Dunbarton	Dunbarton town
2	1	Polygon	33	013	27380	00873600	3301327380	Franklin	Franklin city
3	2	Polygon	33	013	60020	00873696	3301360020	Pembroke	Pembroke town
4	3	Polygon	33	013	00660	00873528	3301300660	Allenstown	Allenstown town
5	4	Polygon	33	013	06260	00873547	3301306260	Boscawen	Boscawen town
6	5	Polygon	33	013	06500	00873548	3301306500	Bow	Bow town
7	6	Polygon	33	011	01700	00873533	3301101700	Antrim	Antrim town
8	7	Polygon	33	011	31940	00873613	3301131940	Greenville	Greenville town

Add a new field to your table with some reasonably short name, and make sure to set the Data Type to match (for IDs, we would usually use integer (Short/Long type)).

	☑	☐	AWATER	AWATER	Double	☐	☐	Numeric		14	0
	☑	☐	INTPTLAT	INTPTLAT	Text	☐	☐			0	0
	☑	☐	INTPTLON	INTPTLON	Text	☐	☐			0	0
	☑	☐	ID		Short ▾	☐	☐				

Save your table edits (right-click the green square to the right of the new field name and choose Save'), then go back to your attribute table. Scroll all the way to the right to see the new field, then right-click its name and choose Calculate Field. Double-click on the name of the linking key field that you want to copy in the Fields list scroll, and you should see the box for ID = fill in below. Hit OK, then try the join again.

5. Export the joined tables into one combined dataset

At this point, the computer is still reading your tables as two separate files on the computer, rather than one unified dataset. If you were to close your project and open it again, or try to go on and do analysis, the changes may not be saved.

Right-click on the name of your boundary dataset in the Contents pane, then choose Data -> Export Features. Save the output feature class somewhere you will find it again with a reasonable name. Then remove the old dataset (right-click -> Remove) and go forward with your joined data.

Calculate Field   ?   ×

 ⓘ   This tool modifies the Input Table   ×

 ⓘ   No pending edits.
    ↶ ↷ 🖫 🗒

Input Table
tl_2020_33_cousub   ⌄   📁

Field Name (Existing or New)
ID   ⌄ ⚙

Expression Type
Python   ⌄

* Expression

Fields   ▼ Helpers   ▼

STATEFP   ⌃   .as_integer_ratio()   ⌃
COUNTYFP   .capitalize()
COUSUBFP   .center()
COUSUBNS   .conjugate()
GEOID   .count()
NAME   .decode()
NAMELSAD   .denominator()
LSAD   ⌄   ⌄

Insert Values   ⌄   * / + - =

ID =

!GEOID!

Code Block   ⌄

Enable Undo 🔘   Apply   OK