Dealing with tables

GEOG370 2020

Javier A. Arce-Nazario

GIS Vector layers contain:

- **Spatial information**
  - The interface (GUI) doesn't always show you all of the information in a vector file!
  - You can **query** the data to find it
  - For example:
    - X and Y location for each vertex
    - The direction of the lines (initial vertex and final vertex)
    - The boundaries of the polygon

- **Attribute table**
  - This can contain all kinds of data (not necessarily spatial)

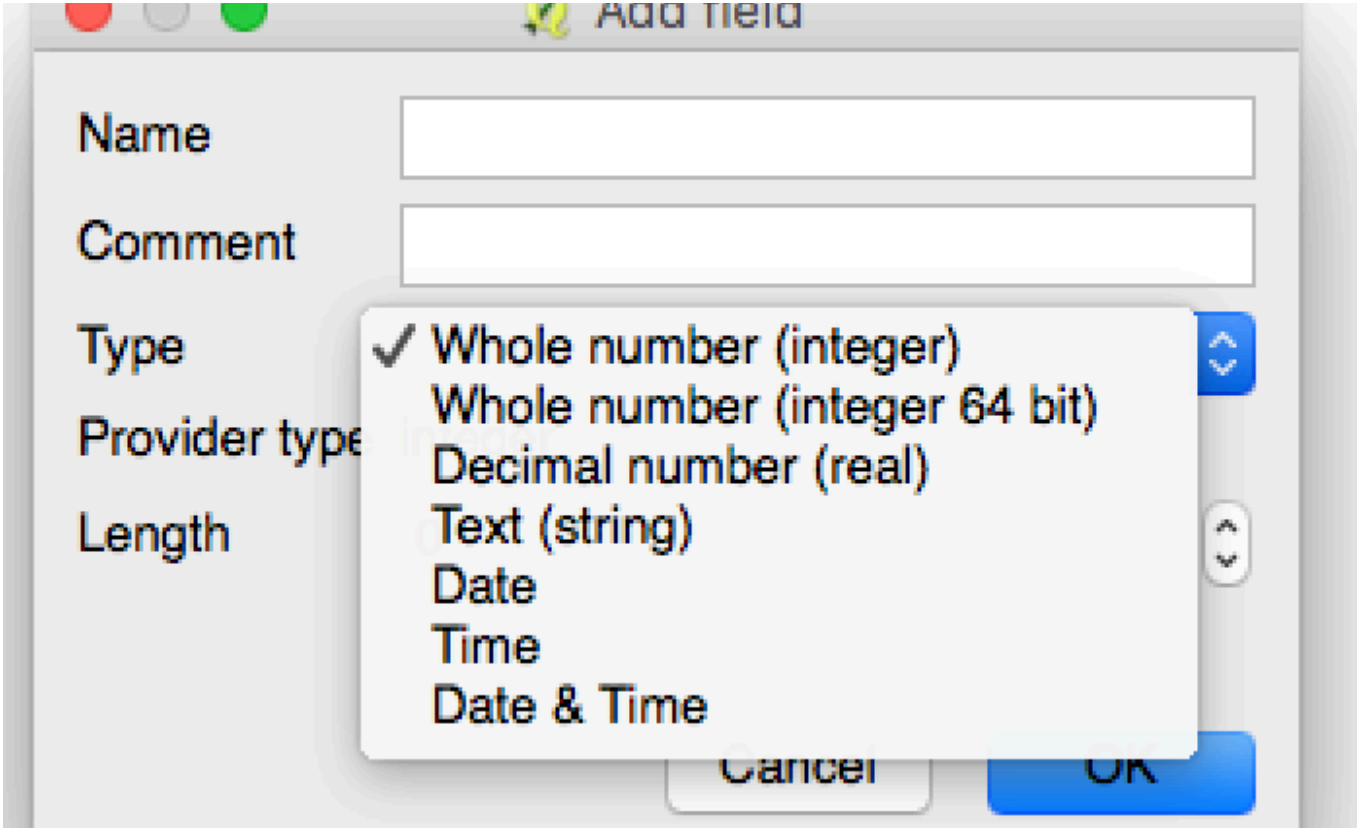
Attribute tables

- Attribute tables are composed of **fields** (columns) and **features** (rows)
  - For example, the **features** in a point shapefile are the points. X and Y are two **fields**.
- Every attribute table should have a field that is a *Unique Identifier (UID)* for each feature.

Common data types for fields

- Text (String)
- Numbers
  - Integer
  - Real
- Date-Time (different formats)

In QGIS when you create a new field you can select the following types….



Working with attribute tables

- QGIS and most GIS software use SQL
  - SQL stands for Structured Query Language, which is a language designed to manage data in relational databases.

Using SQL in GIS we can do a lot of things, but in this class I will only show you how to

1. Make simple data query of the attribute table
2. Join an attribute table of a vector layer to a data table from another non-spatial source

Simple Queries

For text:

- `"COUNTYFP" = '017'`
- `"COUNTYFP" IS '017'`
- `"COUNTYFP" IS NOT '017'`

For numeric data

- `"ALAND" = 2264613777`
- `"ALAND" < 2264613777`
- `"ALAND" > 2264613777`
- `"ALAND" >= 2264613777`
- `"ALAND" <= 2264613777`
- `"ALAND" <> 2264613777`

Simple Queries

Other Boolean operators (AND, OR)

```
"STATEFP" =  ‘45’  OR "STATEFP" =  ‘54’
```

```
"COUNTYFP"   =   ‘017’ AND  "STATEFP" =  ‘37’
```

There are many more operators (but I will not cover them in this course)

Join table

- Joining data is typically used to append the fields of one table to those of another through an attribute or field common to both tables. (ESRI definition)

Simple Join Example

Team member #1 is taking notes

E.g.: The soil characteristics of a location

Team member #2 is taking GPS points

E.g.: The GPS id and location

Simple Join Example

Team member #1's table

Location ID	Soil Type
1	Sandy
2	Clayish
3	Loamy
4	Clayish

Team member #2's table

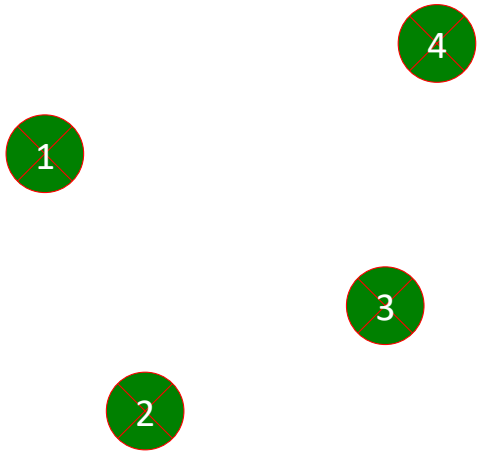
Location ID	Lat	Lon
1	15.20	-31.27
2	13.13	-31.10
3	14.10	-30.30
4	16.20	-30.98

Simple Join Example

Joined Tables

Location ID	Lat	Lon	Soil Type
1	15.20	-31.27	Sandy
2	13.13	-31.10	Clayish
3	14.10	-30.30	Loamy
4	16.20	-30.98	Clayish

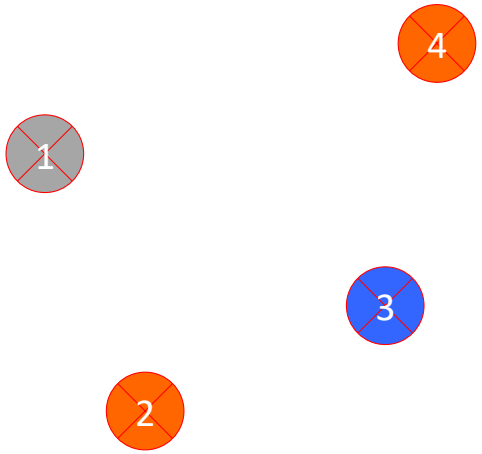
Simple Join Example



Location ID	Lat	Lon
1	15.20	-31.27
2	13.13	-31.10
3	14.10	-30.30
4	16.20	-30.98

Location ID	Soil Type
1	Sandy
2	Clayish
3	Loamy
4	Clayish

Simple Join Example



Location ID	Lat	Lon	Soil Type
1	15.20	-31.27	Sandy
2	13.13	-31.10	Clayish
3	14.10	-30.30	Loamy
4	16.20	-30.98	Clayish

Cleaning and importing data

- Headers
  - Make sure that the first row has the names of the columns and that the second row (onward) has data.
- Data type consistency within a column
  - Make sure that the data type is numbers in a column that represents numbers
  - Make sure that the dates are all in the same format
  - Make sure that real numbers are all real numbers
- Unique categories
  - Make sure that similar objects are described in exactly the same way. (example: ‘Mercedes Benz’ is not the same as ‘Mercedes’)

Headers

Wrong

	A	B	C	D	E
1	GEO.id	GEO.id2	GEO.display-	HC01_EST_V	HC01_MOE_
2	Id	Id2	Geography	Households;	Households;
3	0500000US3	37001	Alamance Co	61545	683
4	0500000US3	37003	Alexander Cc	13581	315

This table will result in all columns being considered as text

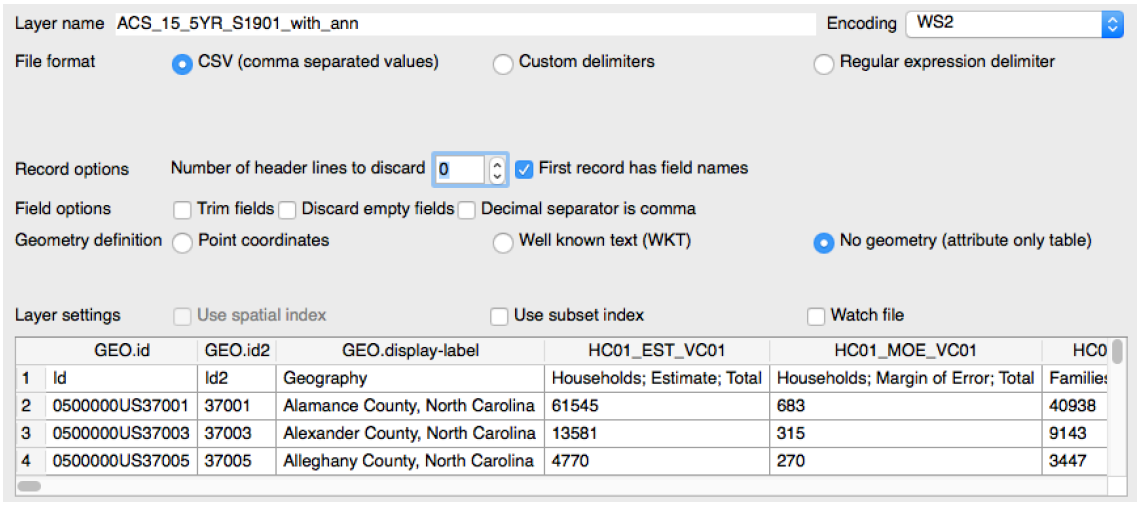
Correct

	A	B	C	D	E
1	GEO.id	GEO.id2	GEO.display-	HC01_EST_V	HC01_MOE_
2	0500000US3	37001	Alamance Co	61545	683
3	0500000US3	37003	Alexander Cc	13581	315

The issue was solved by erasing the second row.

Headers

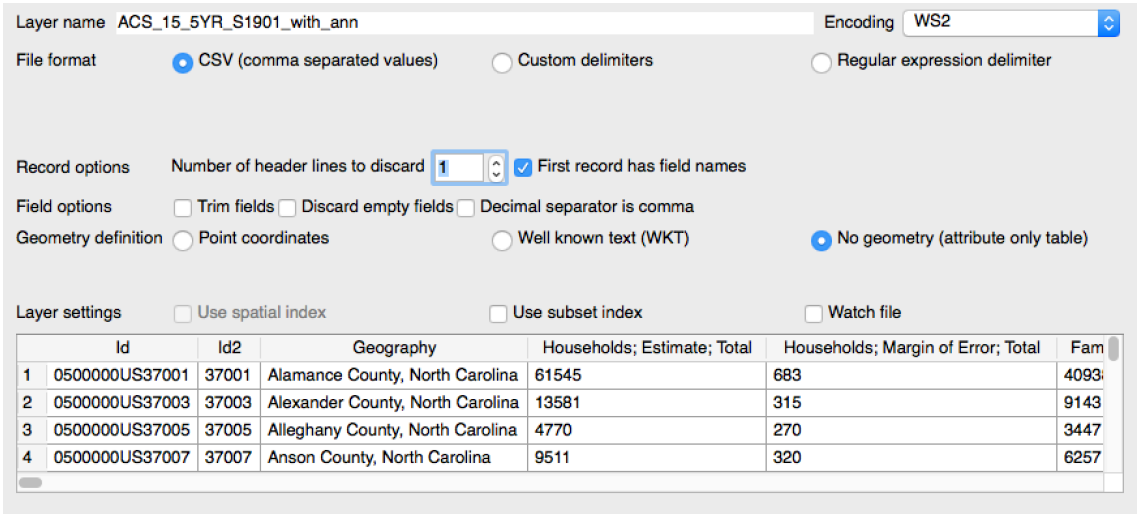
Wrong



	GEO.id	GEO.id2	GEO.display-label	HC01_EST_VC01	HC01_MOE_VC01	HC0
1	Id	Id2	Geography	Households; Estimate; Total	Households; Margin of Error; Total	Familie
2	0500000US37001	37001	Alamance County, North Carolina	61545	683	40938
3	0500000US37003	37003	Alexander County, North Carolina	13581	315	9143
4	0500000US37005	37005	Alleghany County, North Carolina	4770	270	3447

This will result in all columns being considered as text

Correct



	Id	Id2	Geography	Households; Estimate; Total	Households; Margin of Error; Total	Fam
1	0500000US37001	37001	Alamance County, North Carolina	61545	683	4093
2	0500000US37003	37003	Alexander County, North Carolina	13581	315	9143
3	0500000US37005	37005	Alleghany County, North Carolina	4770	270	3447
4	0500000US37007	37007	Anson County, North Carolina	9511	320	6257

The issue was solved by indicating which row should be considered the first row

Data type consistency

ID	Temp
01	24
02	23.3
03	e24.1
4	22

Wrong

ID	Temp	est
01	24.0	0
02	23.3	0
03	24.1	1
04	22.0	0

Correct

ID	Temp
1	24.0
2	23.3
3	-9999
4	22.0

Correct

Unique categories

Wrong

Date.Of.Stop	Time.Of.Stop	Year	Make
4/9/15	22:02:00	2007	BMW
8/13/13	11:14:00	1999	BWM
2/15/13	8:23:00	2006	MERCEDES
5/1/16	2:43:00	2015	VOLKSWAGON
10/21/16	23:57:00	2016	BMW
3/23/14	14:11:00	2012	BMW
2/7/14	21:25:00	2007	BMW
8/26/14	7:10:00	2000	VOLKSWAGEN

This example would have 5 car make categories

BMW, BWM, MERCEDES, VOLKSWAGON, and VOLKSWAGEN

Correct

Date.Of.Stop	Time.Of.Stop	Year	Make
4/9/15	22:02:00	2007	BMW
8/13/13	11:14:00	1999	BMW
2/15/13	8:23:00	2006	MERCEDES
5/1/16	2:43:00	2015	VOLKSWAGEN
10/21/16	23:57:00	2016	BMW
3/23/14	14:11:00	2012	BMW
2/7/14	21:25:00	2007	BMW
8/26/14	7:10:00	2000	VOLKSWAGEN

This example would have 3 car make categories

BMW, MERCEDES, and VOLKSWAGEN