

Working with NHGIS and Joins in QGIS

Description:

Many of you will be using census data in your final projects, or you will be downloading data in tabular formats (.csv) that you need to "join" to shapefiles. This tutorial will run you through the basics of this process. Since so many of you will be using the National Historical Geographic Information System (NHGIS) for your projects, I will be using that as an example. This process, however, can be adapted for any dataset that arrives in separate tabular and shapefile formats.

Sources

For census data after 2000, you can use the census bureau's "American Fact Finder." <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>

Use the "guided search" to download tabular data specific to your project. You can also find "boundary files" (i.e. shapefiles) for enumeration units after 2000 by using the Tiger/Line Shapefile portal. <https://www.census.gov/geo/maps-data/data/tiger-line.html>

For census data before 2000 (with a few exceptions), you will need to use the NHGIS portal.

Other sites to explore:

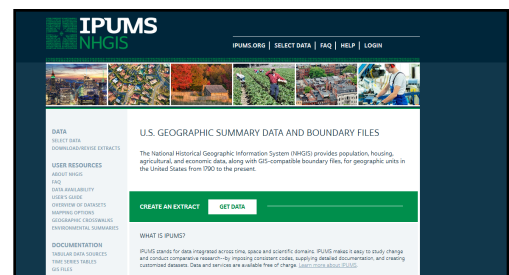
- Diversity and Disparities Project (Brown University) <https://s4.ad.brown.edu/Projects/Diversity/Researcher/MapLTDB.htm>

- Placing Segregation (Omaha 1870 data) <https://dsps.lib.uiowa.edu/placingsegregation/>

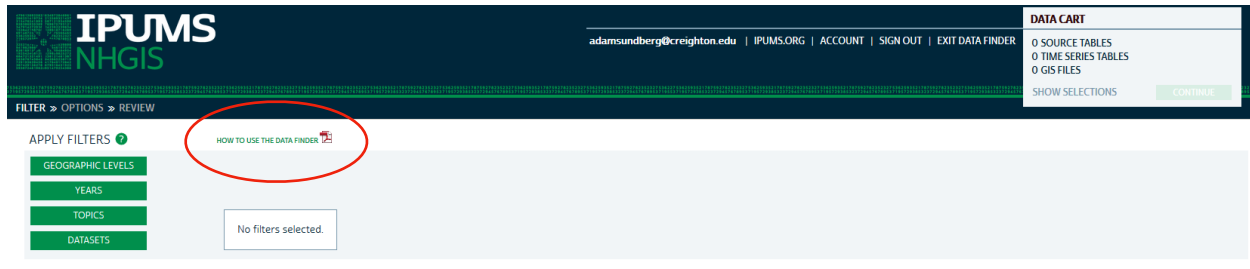
- Urban Historical Transition Project (Omaha 1880 data) <https://s4.ad.brown.edu/Projects/UTP/index.htm>

Using the NHGIS Browser

1. Go to www.nhgis.org
2. Sign up for an account
3. Once you have signed in, click "get data."



4. For this example, I will be looking for census block data (finer resolution than county data) or the number of African Americans living in Omaha in 1950. For a more in depth look at other options, use NHGIS's "using the nhgis data finder" tutorial available on this first page. <https://www.nhgis.org/sites/www.nhgis.org/files/>



using the [nhgis data finder.pdf](#)

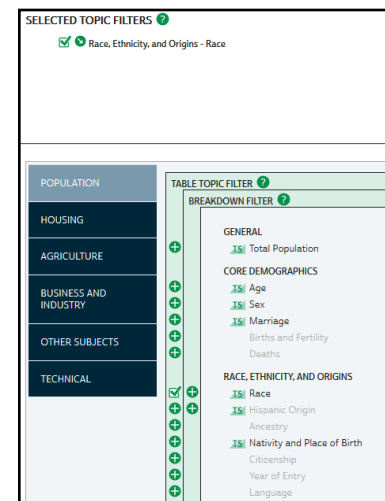
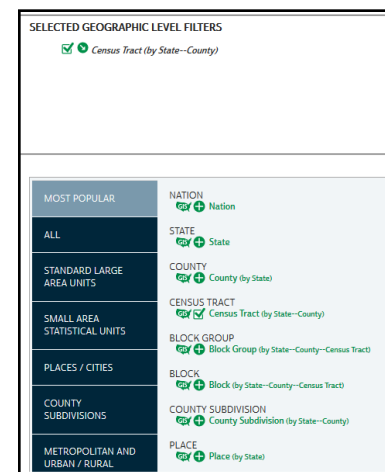
5. Click "geographic levels." We need to filter available data since we're only interested in census block data. Then click "submit." You'll note that there are quite a few "source tables" "time series tables," and "GIS files." Too many in fact. Especially since we are only interested in race information in 1950.

6. Click "years" in the filter panel. Click "1950" in the "decennial years" column. The census that is taken every 10 years will usually provide the most complete information. Click submit.

7. You'll now see that there are quite a bit less to choose from. Only 2 "GIS" files (i.e. boundary shapefiles of census blocks for 1950) and 54 source tables. Since we're interested in race, let's apply one more filter. Under "topics" select race. Click "submit."

8. Now there are only six tables. Population by race sounds appropriate so click the green plus sign to the left of it. You'll note that your "data cart" in the upper left now contains a source table. This is your database file (which will produce your attribute table) when you import it into QGIS.

9. We still need a shapefile of what the census blocks looked like in 1950. Click the "GIS Files" tab under select data. Click the first option for 1950 and you will see 1 GIS



file enter your data cart. Click "continue" on the data cart.

10. On the following screen, you will be prompted to "select geographic levels." Confirm that you want to select census blocks by clicking submit. Then press "continue" on the data cart again.

11. This brings you to the "review and submit" portion of the process. You want to download your source table in .csv format "(best for GIS)" and make sure to add a description so you can keep track of what you're doing for later.

12. Click submit. This will prompt NHGIS to prepare your data. When it is complete, they will send an email to the account you used to set up your login. It should take a couple minutes.

13. When the email arrives, follow the link provided back to the NHGIS portal. You will see a spreadsheet with all of your data requests, descriptions, and link to the zipped files you requested.

DATA CART
CLEAR X

1 SOURCE TABLE

0 TIME SERIES TABLES

1 GIS FILE

SHOW SELECTIONS
CONTINUE

REVIEW AND SUBMIT
?

TABLE FILE STRUCTURE

☒ Comma delimited (best for GIS)

☐ Include additional descriptive header row (best for spreadsheets)

☐ Fixed width (best for statistical packages)

SOURCE TABLES

1 source table

0.31 MB

TIME SERIES TABLES

None selected

GIS FILES

1 GIS file

DESCRIPTION

1950 census block, population by race

SUBMIT

Your NHGIS extract is ready. To retrieve your extract, go to the link below.

Extract: nhgis0026

Link: <http://data2.nhgis.org/downloads>

Description: 1950 census block, population by race

EXTRACTS HISTORY							
The download progress of your latest extract may be seen by refreshing your internet browser.							
EXTRACT NUMBER	DATE CREATED	DOWNLOAD TABLE DATA	DOWNLOAD GIS DATA	STATUS	REVISE EXTRACT	RESUBMIT	DESCRIPTION (CLICK TO EDIT)
26	2018-Mar-10	tables (198 KB)	gis (58 MB)	complete	revise	resubmit	1950 census block, population by race

14. Click both "tables" and "gis" to download the data you ordered. They will arrive in "zipped format" so make sure to **unzip them**. Then cut and paste them into your project folder on your computer (**do not work with them out of your temporary downloads folder**). There should be two files associated with the csv file. Open the .txt file. This is your codebook. The database file will include columns that do not make sense without this codebook, so do not lose it.

```

Tables:
1. Population by Race
  Universe: Persons
  Source code: NT2
  NHGIS code: B03
-----
Data Dictionary
-----
Context Fields
GISJOIN: GIS Join Match Code
YEAR: Data File Year
STATE: State Name
STATEA: State Code
COUNTY: County Name
COUNTYA: County Code
PRETRACTA: Census Tract Prefix Code
TRACTA: Census Tract Code
POSTTRACTA: Census Tract Suffix Code
AREANAME: Area Name

Table 1: Population by Race
Universe: Persons
Source code: NT2
NHGIS code: B03
B03001: White
B03002: Negro
B03003: Other non-white
  
```

15. Your shapefile will have a number of different files in them. Make sure to cut and paste **ALL OF THEM** into your project folder. We now have all of the data we need to visualize our census data. Time to bring it into Q.

16. Open QGIS. Lets add our .csv file first. This is not a vector file or a raster file. It's a "delimited text layer." To add it, click the apostrophe symbol. To add your .csv, select "browse" and add the .csv file you just downloaded. File format is "CSV," and it has "no geometry" (i.e. its not a shapefile). Click ok.



Create a Layer from a Delimited Text File

File Name:

Layer name: Encoding:

File format: ☒ CSV (comma separated values) ☐ Custom delimiters ☐ Regular expression delimiter

Record options: Number of header lines to discard: ☒ First record has field names

Field options: ☒ Trim fields ☐ Discard empty fields ☐ Decimal separator is comma

Geometry definition: ☐ Point coordinates ☐ Well known text (WKT) ☒ No geometry (attribute only table)

Layer settings: ☐ Use spatial index ☐ Use subset index ☐ Watch file

	GISJOIN	YEAR	STATE	STATEA	COUNTY	COUNTYA	PRETRACTA	TRACTA	POSTTRACTA	AREANAME
1	G01007300001	1950	Alabama	01	Jefferson	073		0001		STCTY-01073 TRACT- 0001 IN BIR
2	G01007300002	1950	Alabama	01	Jefferson	073		0002		STCTY-01073 TRACT- 0002 IN BIR
3	G01007300003	1950	Alabama	01	Jefferson	073		0003		STCTY-01073 TRACT- 0003 IN BIR
4	G01007300004	1950	Alabama	01	Jefferson	073		0004		STCTY-01073 TRACT- 0004 IN BIR
5	G01007300005	1950	Alabama	01	Jefferson	073		0005		STCTY-01073 TRACT- 0005 IN BIR

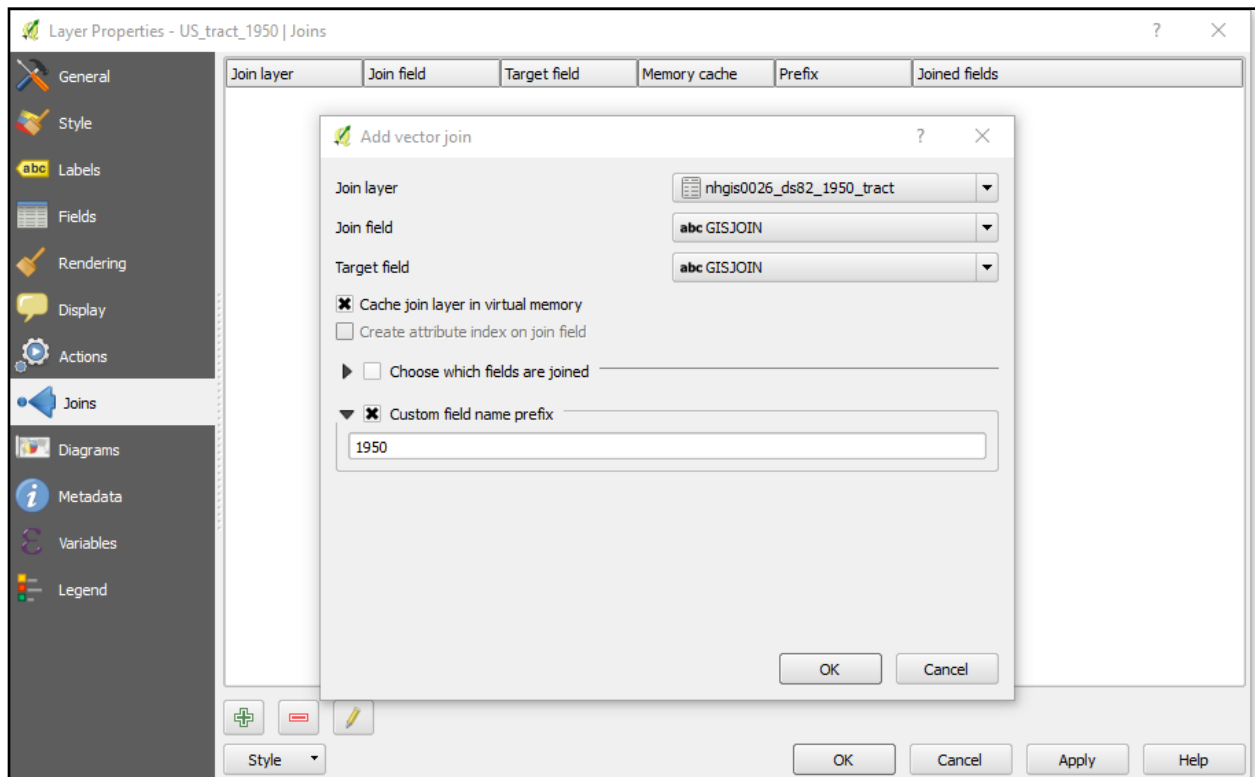
17. Your database is now added. Nothing should be visualized, but the file should now appear in the layers panel.



18. Now let's add the shapefile. Click the "add vector layer" icon. Add US_tract_1950.shp. Zoom to layer. You'll note that only major metropolitan areas are visualized, including Omaha.

19. Now we need to *join* the table to the shapefile layer. Double click your shapefile layer to open the "layer properties" dialogue box. On the left tab, select "joins." Click the green plus sign at the bottom. Join layer should be your csv file. Join field should be "GISJOIN" and target field should be "GISJOIN." This is telling QGIS that you want

to use the column that is shared between the two datasets to join your tabular data to the shapefile. Click "custom field name prefix, and rename it 1950. Click "ok" and "ok."

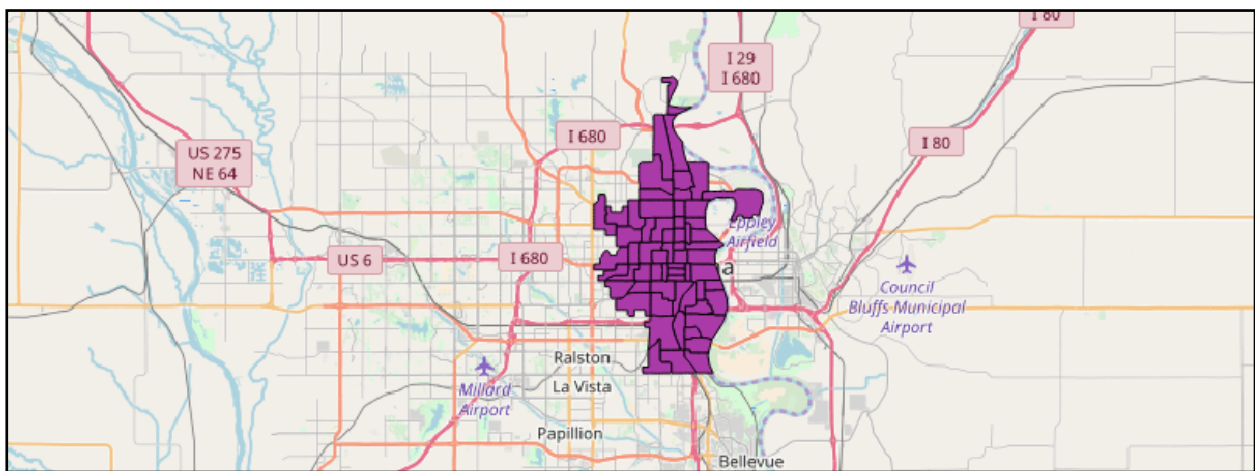
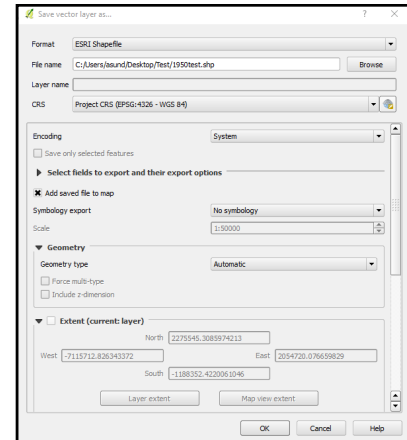


20. Now, open your attribute table for the shapefile. You'll see that our data is now joined.

GISJOIN	GISJOIN2	SHAPE_AREA	SHAPE_LEN	1950YEAR	1950STATE	1950STATEA	1950COUNTY	1950COUNTYA	1950PRETRACTA	1950TRACTA	1950POSTTRACTA	1950AREANAME	1950B0J001	1950B0J002	1950B0J003
55007900129	55007900129	609036.2793130...	3724.864664080...	1950	Wisconsin	55	Milwaukee	79		129		STCTY-55079 TR...	546	0	0
55007909M0215	55007909M0215	1573475.737810...	6987.54029159000	1950	Wisconsin	55	Milwaukee	79	SM	215		STCTY-55079 TR...	2359	2	0
5500790MC0201	5500790MC0201	1729083.892010...	7016.383336090...	1950	Wisconsin	55	Milwaukee	79	MC	201		STCTY-55079 TR...	3000	0	0
5500790MC0199	5500790MC0199	1208094.573499...	6828.575045020...	1950	Wisconsin	55	Milwaukee	79	MC	199		STCTY-55079 TR...	2710	4	0
5500790MC0200	5500790MC0200	2574556.842319...	9154.490416200...	1950	Wisconsin	55	Milwaukee	79	MC	200		STCTY-55079 TR...	2302	0	0
55007900018	55007900018	372006.4062999...	2566.680446140...	1950	Wisconsin	55	Milwaukee	79		18		STCTY-55079 TR...	4346	12	43
55007900017	55007900017	581355.0514020...	3199.947742180...	1950	Wisconsin	55	Milwaukee	79		17		STCTY-55079 TR...	2115	3	41
55007900019	55007900019	530196.7905389...	3251.161306610...	1950	Wisconsin	55	Milwaukee	79		19		STCTY-55079 TR...	2859	573	56
55007900219	55007900219	3952320.151399...	12435.48138120...	1950	Wisconsin	55	Milwaukee	79		219		STCTY-55079 TR...	4335	87	5
5500790MC0168	5500790MC0168	4902635.157829...	10218.76569270...	1950	Wisconsin	55	Milwaukee	79	MC	168		STCTY-55079 TR...	2674	0	0
5500790MC0165	5500790MC0165	2089262.267519...	6277.58612615000	1950	Wisconsin	55	Milwaukee	79	MC	165		STCTY-55079 TR...	1605	1	0
5500790MC0166	5500790MC0166	1010246.691039...	5380.799579520...	1950	Wisconsin	55	Milwaukee	79	MC	166		STCTY-55079 TR...	2140	22	0
5500790WB0156	5500790WB0156	856093.1421369...	4067.77820289000	1950	Wisconsin	55	Milwaukee	79	WB	156		STCTY-55079 TR...	4864	7	3
5500790WB0155	5500790WB0155	2201914.243460...	7796.910662110...	1950	Wisconsin	55	Milwaukee	79	WB	155		STCTY-55079 TR...	5076	12	1
550079050158	550079050158	14065474.54389...	19620.25550730...	1950	Wisconsin	55	Milwaukee	79	S	158		STCTY-55079 TR...	4907	18	2
5500790G0204	5500790G0204	1707245.966759...	7800.668886670...	1950	Wisconsin	55	Milwaukee	79	G	204		STCTY-55079 TR...	2752	0	0
55007900107	55007900107	1043628.868410...	4215.552798210...	1950	Wisconsin	55	Milwaukee	79		107		STCTY-55079 TR...	3590	2	0
55007900105	55007900105	800326.5986200...	5679.770689340...	1950	Wisconsin	55	Milwaukee	79		105		STCTY-55079 TR...	3149	0	4
55007900141	55007900141	1862358.446680...	6459.596168030...	1950	Wisconsin	55	Milwaukee	79		141		STCTY-55079 TR...	2246	0	0
55007900123	55007900123	480362.6856360...	2843.4377112910...	1950	Wisconsin	55	Milwaukee	79		123		STCTY-55079 TR...	4643	1	3
55007900118	55007900118	564704.3279770...	3004.153483100...	1950	Wisconsin	55	Milwaukee	79		118		STCTY-55079 TR...	5929	1	8
55007900122	55007900122	702874.6015909...	3419.488788000...	1950	Wisconsin	55	Milwaukee	79		122		STCTY-55079 TR...	6219	2	2
55007900119	55007900119	462592.7305749...	2776.765286760...	1950	Wisconsin	55	Milwaukee	79		119		STCTY-55079 TR...	4374	0	6
55007900117	55007900117	456295.6111520...	2789.217280720...	1950	Wisconsin	55	Milwaukee	79		117		STCTY-55079 TR...	5258	5	0
55007900116	55007900116	493564.2920949...	2925.763792950...	1950	Wisconsin	55	Milwaukee	79		116		STCTY-55079 TR...	5570	3	20
55007900103	55007900103	1309847.319679...	5434.820152140...	1950	Wisconsin	55	Milwaukee	79		103		STCTY-55079 TR...	4366	30	5
55007900101	55007900101	602347.1270840...	3137.35214783000	1950	Wisconsin	55	Milwaukee	79		101		STCTY-55079 TR...	3850	0	1
55007900102	55007900102	479391.7305189...	2869.283582270...	1950	Wisconsin	55	Milwaukee	79		102		STCTY-55079 TR...	3121	0	0
55007900076	55007900076	1086292.429400...	4319.638801860...	1950	Wisconsin	55	Milwaukee	79		76		STCTY-55079 TR...	4126	0	3
55007900046	55007900046	516706.2833720...	1882.72934061000	1950	Wisconsin	55	Milwaukee	79		46		STCTY-55079 TR...	5873	0	0

21. Right now, this join is only temporary, so let's save it as a new file. Right click your shapefile, click "save as", make sure to save it to your project folder and select the correct CRS. Click ok. You can remove your original shapefile and csv file from the layers panel now.

22. Let's zoom in on Omaha. If you don't know exactly where it is, add a basemap. You'll see that only eastern Omaha is visualized. More and more of Omaha will be included as you look to data from the 60s, 70s, etc.



23. Let's style the map to visualize the number of african americans by census block. Double click your shapefile and click "style," "graduated," column "1950B0J002" (I knew this would be the number of Af Americans by looking at the txt file codebook associated with my csv file). Click quantile as your mode and click classify. We now have a racial choropleth map. These are raw numbers, however, so make sure that you create a new column with total population and another new column that divides number of Af American inhabitants by that total population for a normalized column of data that you can use in a choropleth.

