TUTORIAL 7.2 | HISTORICAL CENSUS DATA

Goals

- Download and use historical census data.
- Learn how to use the IPUMS platform.

Introduction

Using IPUMS, a database of census and other types of social data maintained by the University of Minnesota. IPUMS has census records dating back to 1790, in the US and in other countries. The records have been “harmonized” to correspond with contemporary shapefiles – typically the 2000 or 2010 shapefile for older records. The farther back you search, the less likely that you’ll find fine-grained data for an area, but IPUMS has census tract level data for some variables dating back decades, and even more extensive county level data

For this tutorial, we’ll be looking at historical commuter data – specially, method of transportation to work. This data is available at the census tract level dating back to 1960. The data is available for each decade, but we’ll look at the change in 20 year increments: 1960, 80, 2000, and 2020. You’ll need a separate shapefile for each of these years, and since IPUMS data are only available for the entire country you’ll need to downsize the spreadsheets and shapefiles for each year as well. In the end you’ll have an extensive dataset of commuter change across the country, which you’ll reduce to only your city.

As with other census data, you’ll need to make sure that the GEOIDs for the spreadsheet and shapefile match and can be joined.

Step 1: Create a Free IPUMS NHGIS account.

The account and data are free for educational and research purposes.

https://www.nhgis.org/

Once you've created the account, login to the NHGIS website.

Step 2: Filter NHGIS tables by variable and geography.

You might not know at first which data table specifically you're looking for, or even what kinds of data are available for which years and geographical levels. I like to start by filtering by geographic level, and then by topic, and lastly by year. You'll see that each filter helpfully reduces the options for the next – for instance, once you select a geography and topic, only the years with that information available show up in the filter list.

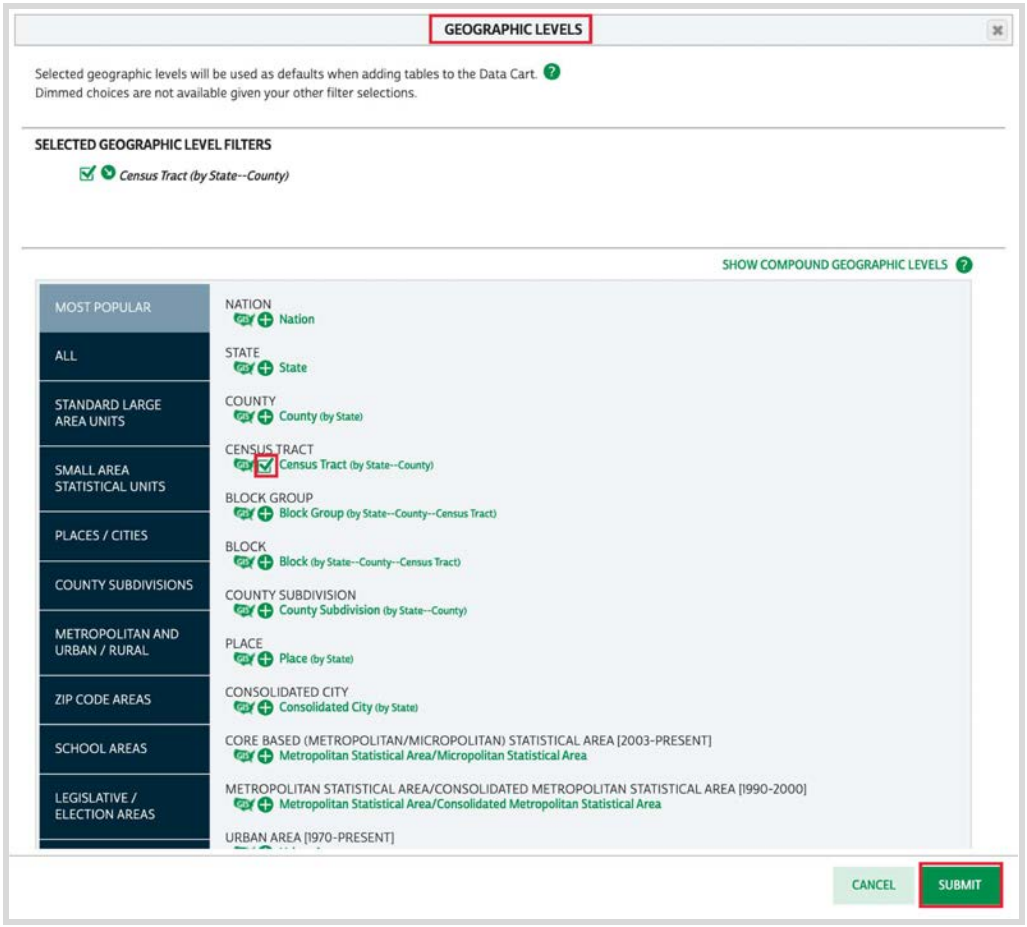
2a Click "**Get Data**", the green button on the NHGIS home page. Clicking "**Select Data**" at the top will also work.

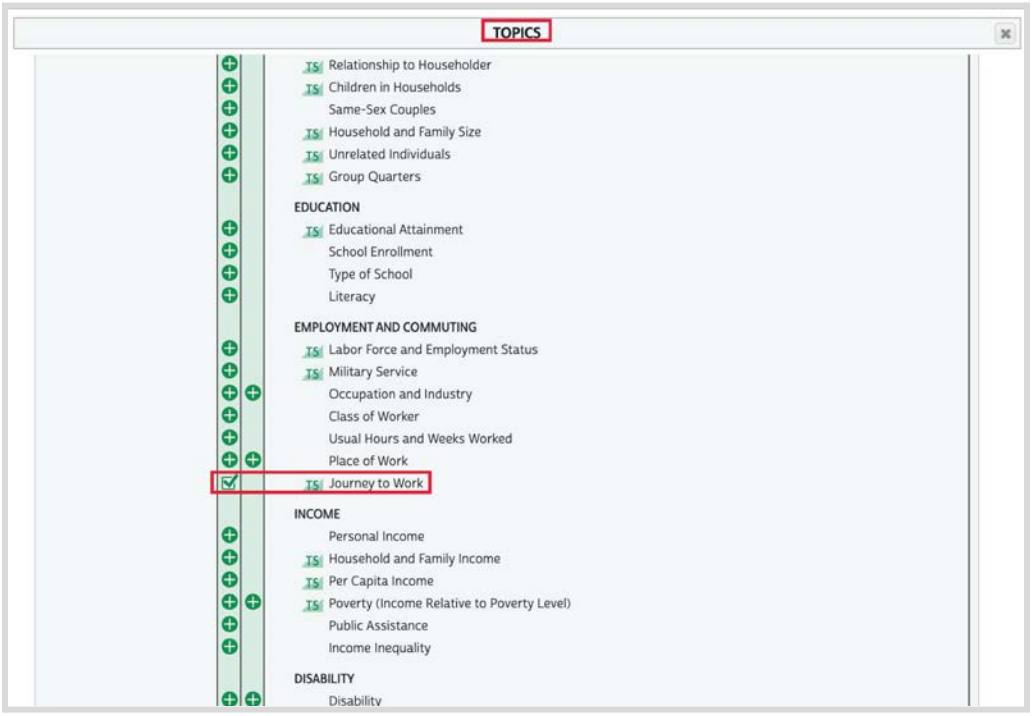
2b You will see the NHGIS search page. To show the tables we'll use, apply the following filters by clicking green plus symbol beside each filter:

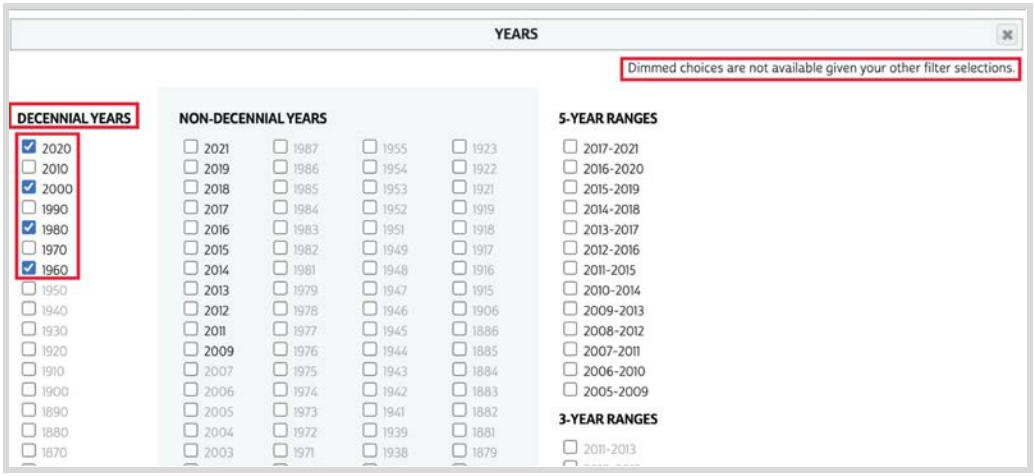
Geography Level: census tract

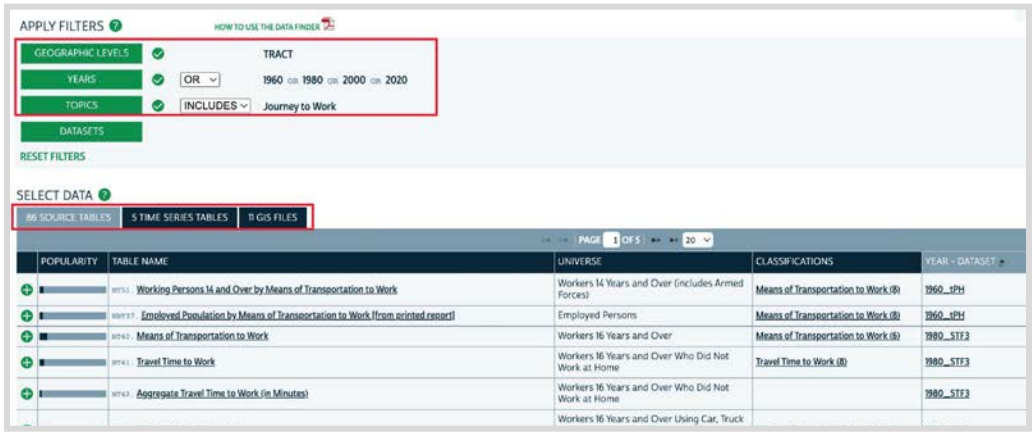
Topic: Under "Population", scroll down and select "Journey to Work"

Year: select 1960, 1980, 2000, and 2020


GEOGRAPHIC LEVELS
Selected geographic levels will be used as defaults when adding tables to the Data Cart.
Dimmed choices are not available given your other filter selections.
SELECTED GEOGRAPHIC LEVEL FILTERS
Census Tract (by State--County)
SHOW COMPOUND GEOGRAPHIC LEVELS
MOST POPULAR
ALL
STANDARD LARGE AREA UNITS
SMALL AREA STATISTICAL UNITS
PLACES / CITIES
COUNTY SUBDIVISIONS
METROPOLITAN AND URBAN / RURAL
ZIP CODE AREAS
SCHOOL AREAS
LEGISLATIVE / ELECTION AREAS
NATION
Nation
STATE
State
COUNTY
County (by State)
CENSUS TRACT
Census Tract (by State--County)
BLOCK GROUP
Block Group (by State--County--Census Tract)
BLOCK
Block (by State--County--Census Tract)
COUNTY SUBDIVISION
County Subdivision (by State--County)
PLACE
Place (by State)
CONSOLIDATED CITY
Consolidated City (by State)
CORE BASED (METROPOLITAN/MICROPOLITAN) STATISTICAL AREA [2003-PRESENT]
Metropolitan Statistical Area/Micropolitan Statistical Area
METROPOLITAN STATISTICAL AREA/CONSOLIDATED METROPOLITAN STATISTICAL AREA [1990-2000]
Metropolitan Statistical Area/Consolidated Metropolitan Statistical Area
URBAN AREA [1970-PRESENT]
CANCEL
SUBMIT


TOPICS
Relationship to Householder
Children in Households
Same-Sex Couples
Household and Family Size
Unrelated Individuals
Group Quarters
EDUCATION
Educational Attainment
School Enrollment
Type of School
Literacy
EMPLOYMENT AND COMMUTING
Labor Force and Employment Status
Military Service
Occupation and Industry
Class of Worker
Usual Hours and Weeks Worked
Place of Work
Journey to Work
INCOME
Personal Income
Household and Family Income
Per Capita Income
Poverty (Income Relative to Poverty Level)
Public Assistance
Income Inequality
DISABILITY
Disability

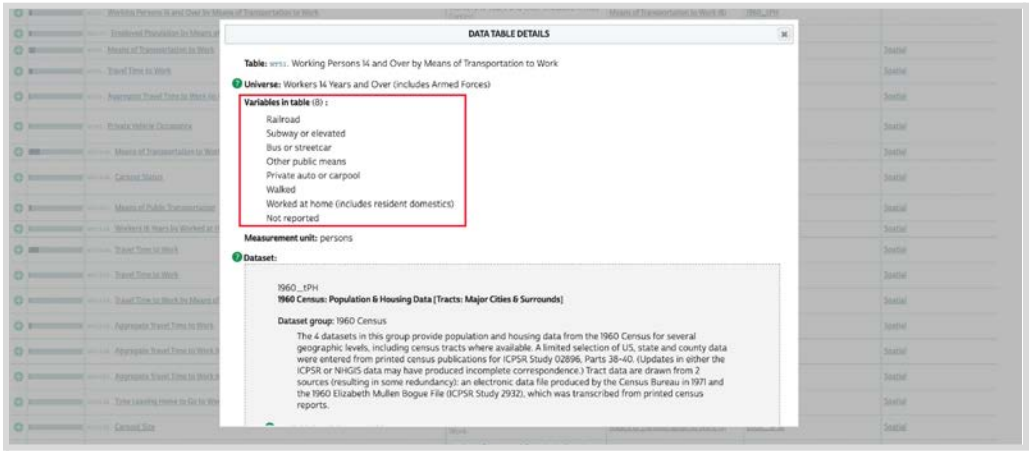




You should now see a list of around 86 source tables, 5 time series tables, and 11 GIS files returned.

Step 3: Select the tables and GIS files. Download the data.

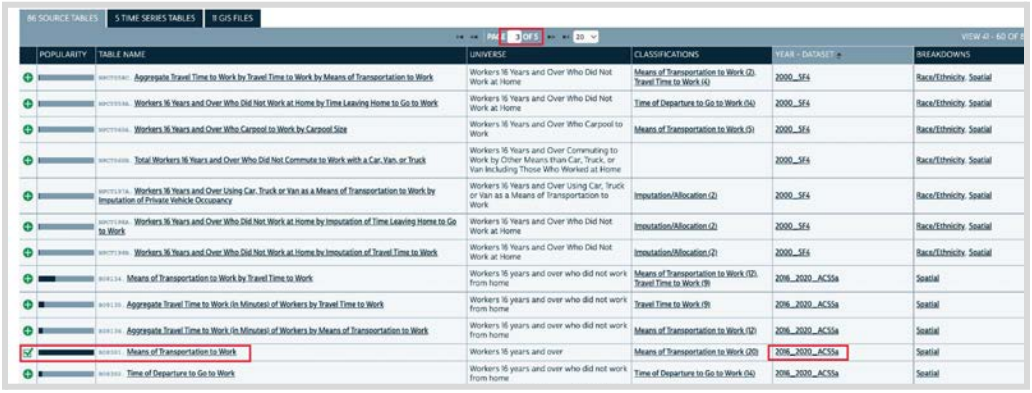
You'll see that many source tables have similar names. We want only the data about worker commute – not travel time to work, not travel time by type of transportation. To check what kind of data a table has you can click on the Table Name to pull up a window with variables, units, available geographic levels, and some other useful information. Note: when in doubt about which tables to use, check the "popularity" column. The more popular table is usually the safer bet.



3a Click the check mark beside the tables for each decade which relate to work transportation method. You'll see that the variable name changes:

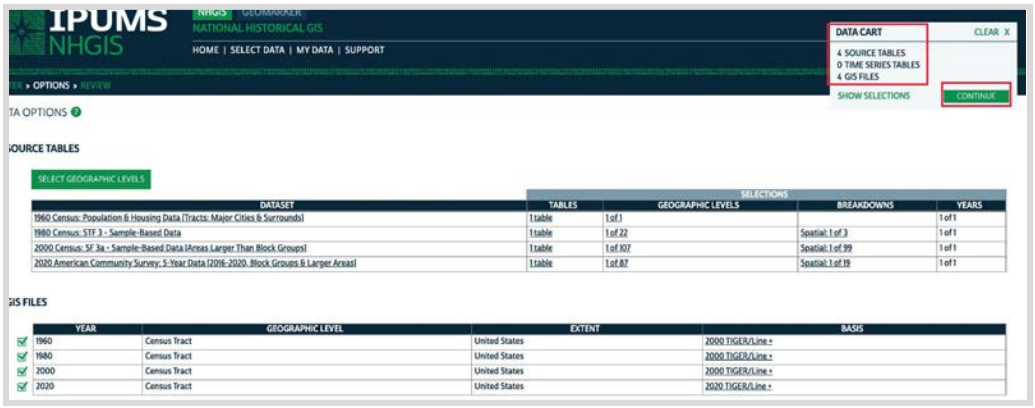
Working Persons 14 and Over by Means of Transportation to Work (1960)
Means of Transportation to Work (1980 and 2000 and 2020)

Note: the tables are automatically sorted in date order, so the 2020 tables will follow several pages of 2000 results.

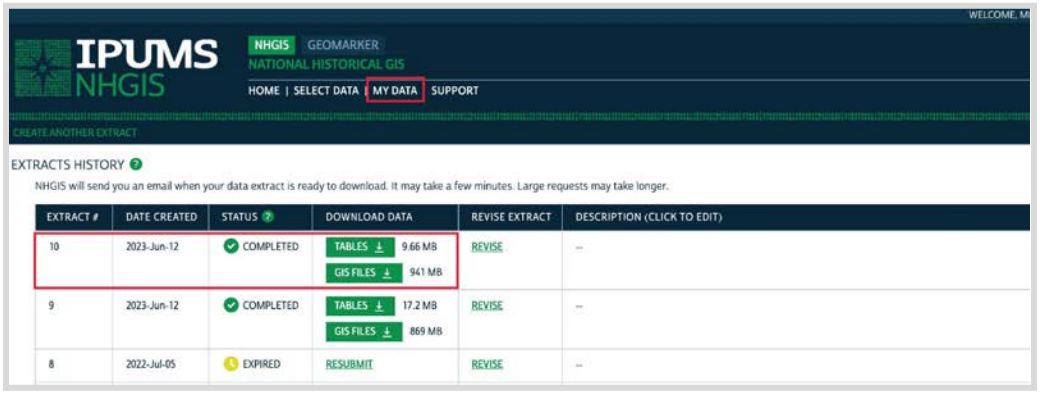


3b Next, go to the “GIS Files” tab. Select the files with the “2000 TIGER File” Basis. Select one for each of the 4 decades: 1960, 80, 2000, and 2020.

3c Your “Data Cart” in the upper righthand corner should show 4 Source Tables and 4 GIS Files. Click “Continue”, review the files, and click “Continue” again. On the next page, click “Submit”.



3d You’ll be taken to the “Extracts History” page. Once the “Status” column shows “Completed”, download the tables and GIS files. Processing your request should take a few minutes.

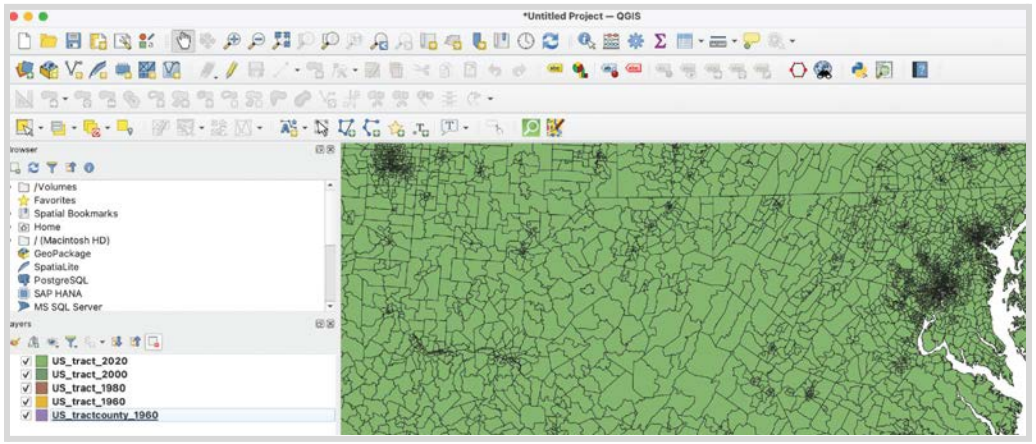


Step 4: Import and join in QGIS.

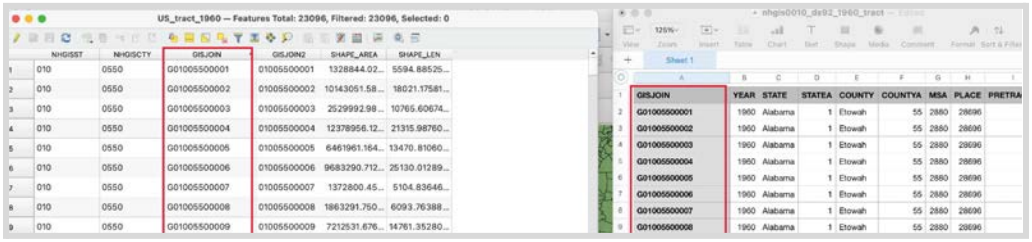
4a First, open a new QGIS file. Set the CRS to “UTM 17N”.

4b Unzip and Import your TIGER shapefiles.

NOTE: the census tracts are divided into different geographical categories depending on the decade. You can import and check each file. You’ll see that in 1960, Roanoke and Lynchburg were not divided into census tracts, but by 1980 they were. For each decade, import the GIS file which contains your city (tract or county).



4c Open the Attribute Table for the 1960 shapefile and scroll to the “GIS Join” columns. Open the 1960 csv file you downloaded and check that the GIS Join column and the data table’s GIS Join columns match. Sort both columns ascending to check.

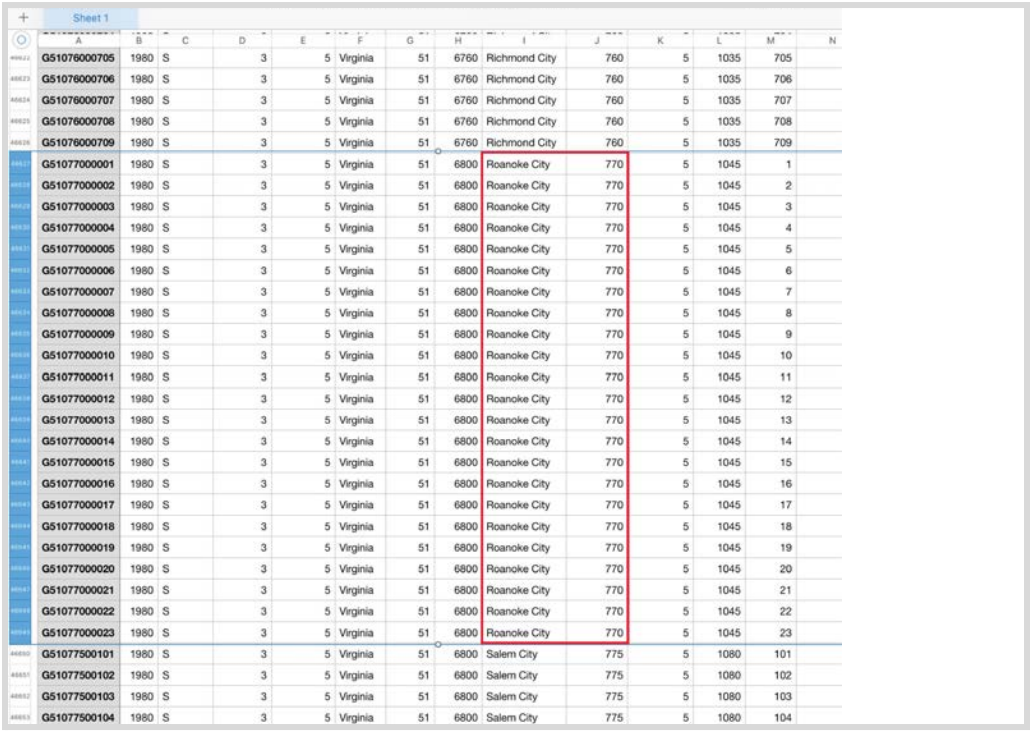


NOTE: if your city is included in the “tractcounty” instead of the “tract” file, you’ll need to download the “county” csv for that year instead. For instance, Roanoke does not appear in the 1960 tract shapefile but does appear in the tractcounty file, so for that year I need to go back to the NHGIS website and download the “county” level geography for the 1960 “Working Persons 14 and Over by Means of Transportation to Work” table.

4d Reduce each csv table before importing it to QGIS.

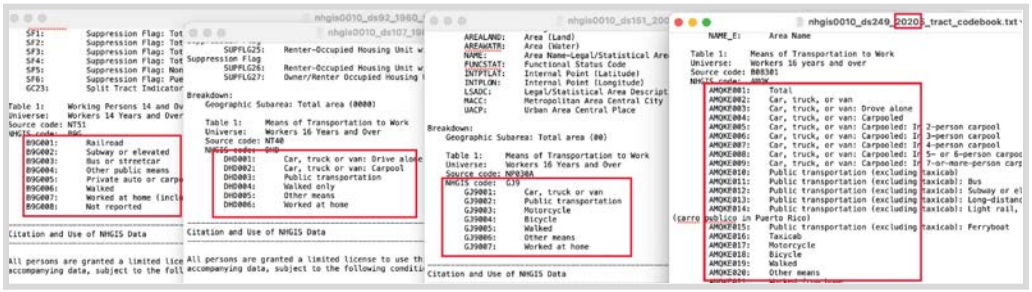
First, sort the table by “State” and copy / paste the Virginia rows into a new spreadsheet. Copy / paste the header column from the first spreadsheet as well. Note: from 1980 onward you can further reduce the spreadsheet to only your county, using the following codes in the relevant county column, “CountyA”, or “CountyFP”, or “NHGISCity”:

Roanoke: 770
Richmond: 760
Lynchburg: 680



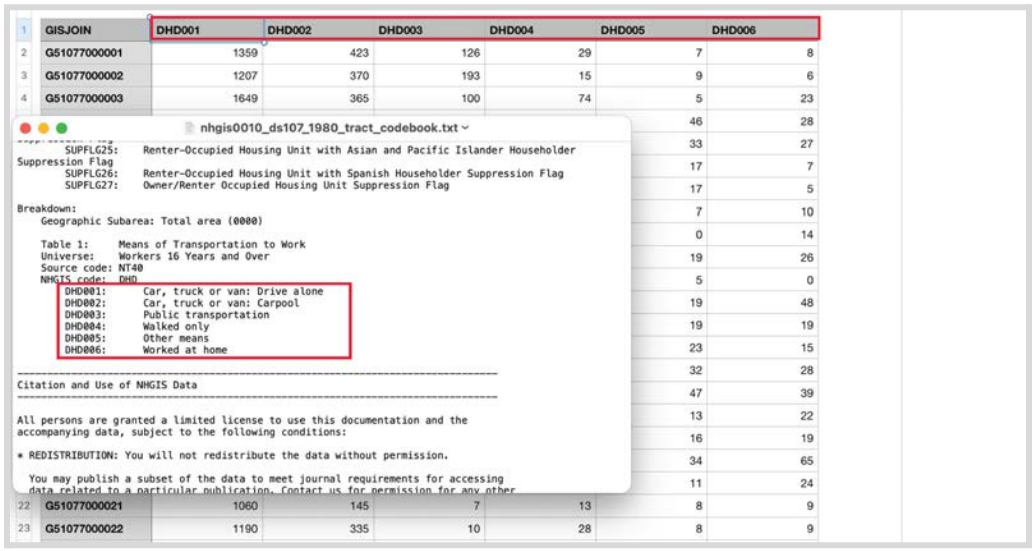
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	G51076000705	1980	S	3	5	Virginia	51	6760	Richmond City	760	5	1035	705	
	G51076000706	1980	S	3	5	Virginia	51	6760	Richmond City	760	5	1035	706	
	G51076000707	1980	S	3	5	Virginia	51	6760	Richmond City	760	5	1035	707	
	G51076000708	1980	S	3	5	Virginia	51	6760	Richmond City	760	5	1035	708	
	G51076000709	1980	S	3	5	Virginia	51	6760	Richmond City	760	5	1035	709	
	G51077000001	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	1	
	G51077000002	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	2	
	G51077000003	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	3	
	G51077000004	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	4	
	G51077000005	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	5	
	G51077000006	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	6	
	G51077000007	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	7	
	G51077000008	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	8	
	G51077000009	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	9	
	G51077000010	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	10	
	G51077000011	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	11	
	G51077000012	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	12	
	G51077000013	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	13	
	G51077000014	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	14	
	G51077000015	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	15	
	G51077000016	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	16	
	G51077000017	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	17	
	G51077000018	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	18	
	G51077000019	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	19	
	G51077000020	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	20	
	G51077000021	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	21	
	G51077000022	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	22	
	G51077000023	1980	S	3	5	Virginia	51	6800	Roanoke City	770	5	1045	23	
	G51077500101	1980	S	3	5	Virginia	51	6800	Salem City	775	5	1080	101	
	G51077500102	1980	S	3	5	Virginia	51	6800	Salem City	775	5	1080	102	
	G51077500103	1980	S	3	5	Virginia	51	6800	Salem City	775	5	1080	103	
	G51077500104	1980	S	3	5	Virginia	51	6800	Salem City	775	5	1080	104	

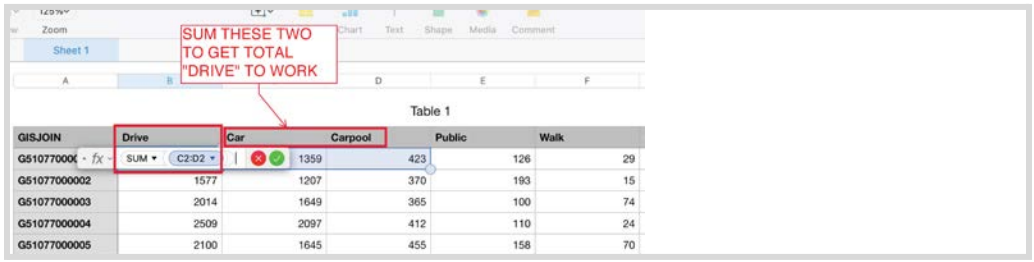
Second, in the new reduced spreadsheet, delete every column except the GISJoin column and the columns with census codes (B9G001, B9G002, DHD001, etc.). For each table, refer to the accompanying “codebook” text file which contains the full name of each data column in the table.



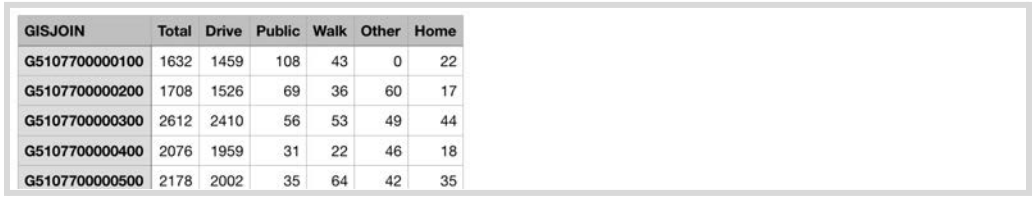
Third, rename the census columns according to the codebook file. Note that, though the exact name of transportation methods changes between the decades, you should use the same name across all years: for instance, “Walk”, “Drive”, etc. If these names match between all 4 tables you’ll be able to more easily share styles between them in QGIS. In some cases, you may want to combine multiple columns using the “SUM” function (make sure not to remove columns, since you will need the full count for a “total” column).. I recommend simplifying the columns to “Walk”, “Drive”, “Public” (public transit), “Other” (including bike), “Home” (work from home), and “None” (eg no response). Note that in the codebook text file, the column descriptions indicate categories, and after colon a

subcolumn. So, for the 2020 file, “Car, truck or van: Drove alone” is a subset of “Car, truck, or van”. When totalling these columns, don’t include the subcolumns or you’ll double count.





Lastly, add a column which totals the other census columns. Fill out this column with the function “=SUM(*cell range*)”, without the quotation marks. For instance, =SUM(C2:F2). Pull the function down the entire column by clicking and dragging from the bottom.

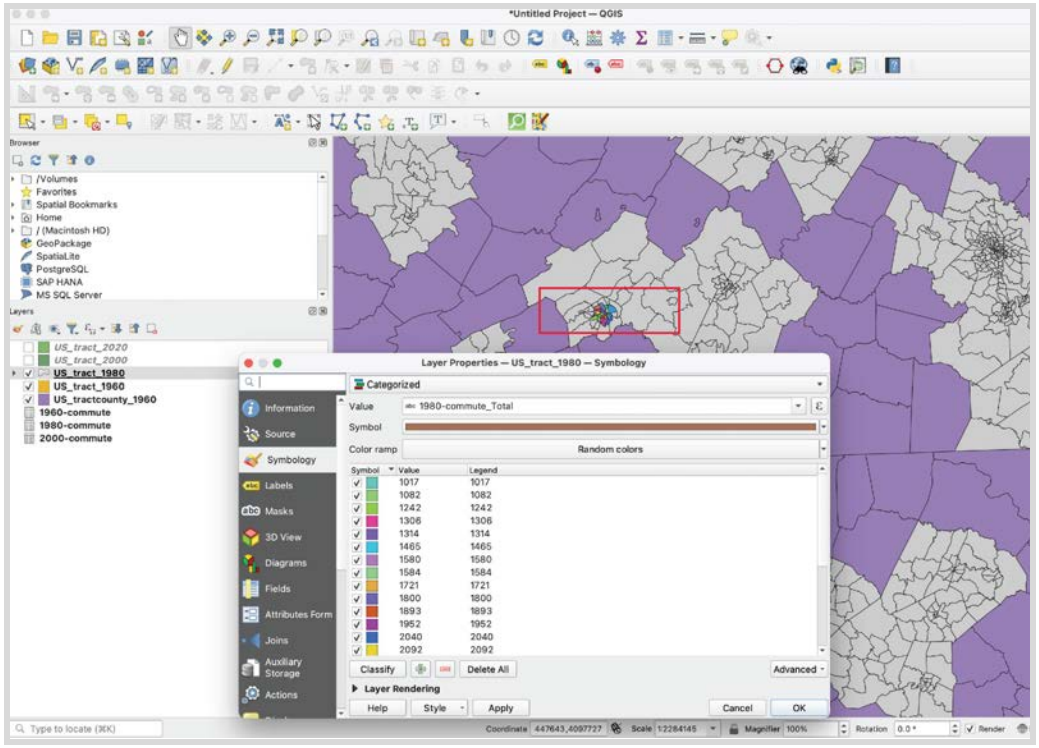


GISJOIN	Total	Drive	Public	Walk	Other	Home
G5107700000100	1632	1459	108	43	0	22
G5107700000200	1708	1526	69	36	60	17
G5107700000300	2612	2410	56	53	49	44
G5107700000400	2076	1959	31	22	46	18
G5107700000500	2178	2002	35	64	42	35

Save the new file as a csv. Repeat for the remaining 3 tables.

4d Once all the data table and shapefile GEOID columns match for their respective years (1960 to 1960, 1980 to 1980, and so on), import the 4 data tables to QGIS.

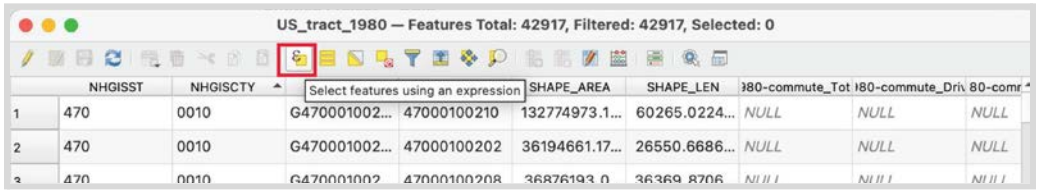
4e Join each data table to its respective shapefile. Check the Attribute Tables of the shapefiles to make sure the new columns joined correctly and are not Null – but note that **only the tracts around your city will have data**, since you deleted everything else. You can also check by styling the data.

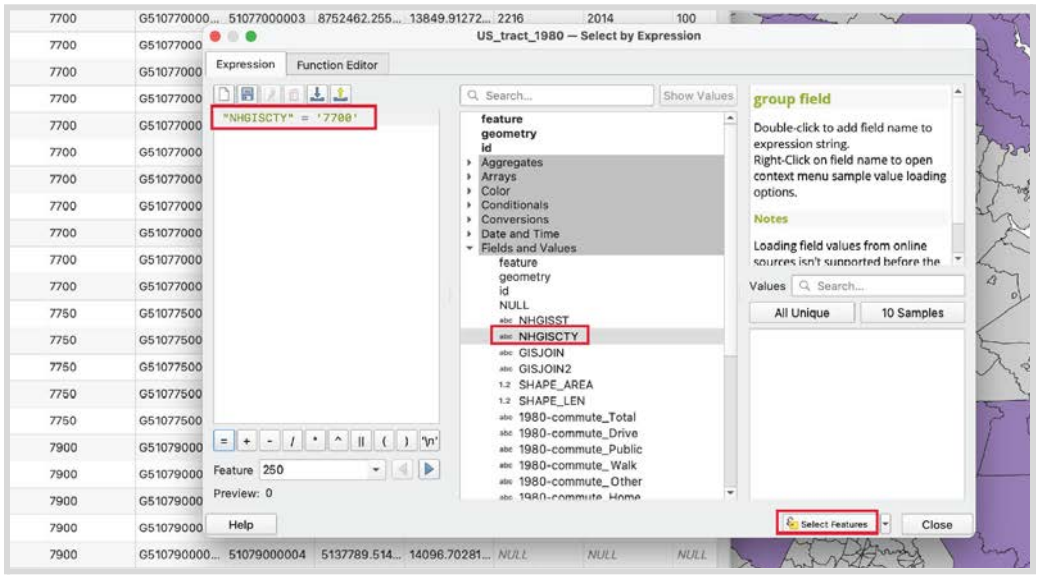

*Untitled Project — QGIS
Browser
/Volumes
Favorites
Spatial Bookmarks
Home
/ (Macintosh HD)
GeoPackage
SpatiaLite
PostgreSQL
SAP HANA
MS SQL Server
Layers
US_tract_2020
US_tract_2000
US_tract_1980
US_tract_1960
US_tractcounty_1960
1960-commute
1980-commute
2000-commute
Layer Properties — US_tract_1980 — Symbology
Information
Source
Symbology
Labels
Masks
3D View
Diagrams
Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Categorized
Value
1980-commute_Total
Symbol
Color ramp
Random colors
Symbol
Value
Legend
1017
1082
1242
1306
1314
1465
1580
1584
1721
1800
1893
1952
2040
2092
Classify
Delete All
Advanced
Layer Rendering
Help
Style
Apply
Cancel
OK
Type to locate (⌘K)
Coordinate
447643,4097727
Scale
1:2284145
Magnifier
100%
Rotation
0.0
Render

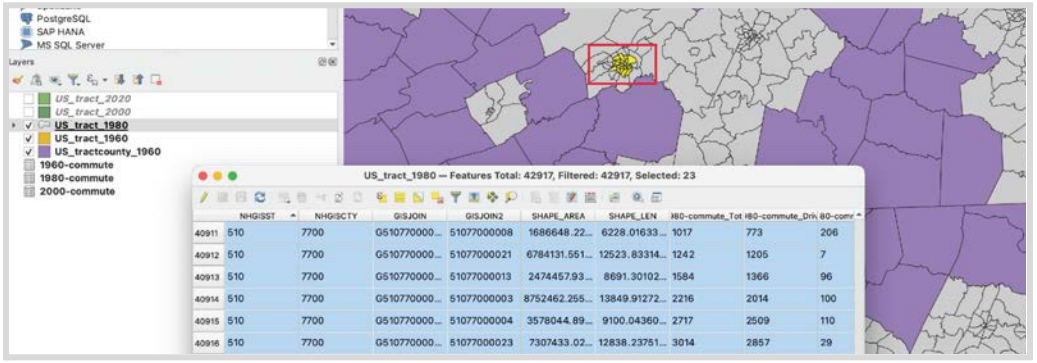
Step 5: Reduce the shapefiles to your city of interest.

5a Open the Attribute table of the 1960 shapefile. Click “Select by Expression” from the upper toolbar. Select “NHGISCITY” and set it equal to your city’s code with an additional 0 at the end:

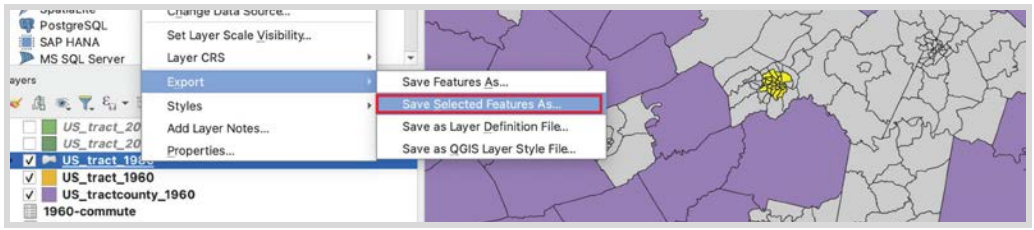
Roanoke: 7700
Richmond: 7600
Lynchburg: 6800







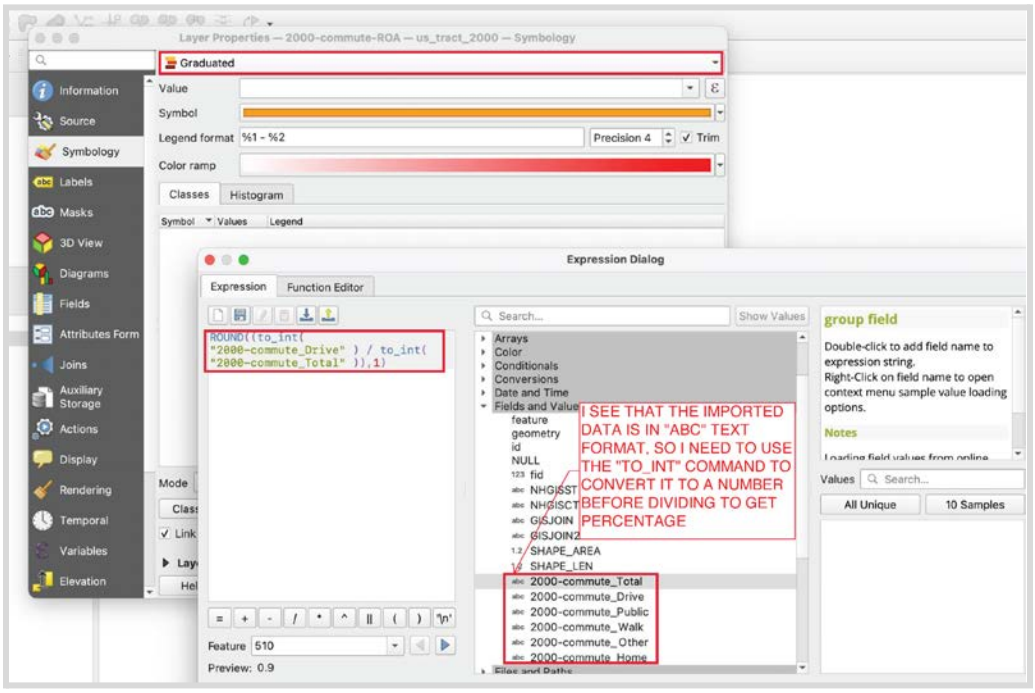
5b If you see the correct selection in your map, right click on the Shapefile layer, go to “Export”, and then “Save Selected Features As”. The new reduced shapefile will be added to your map.



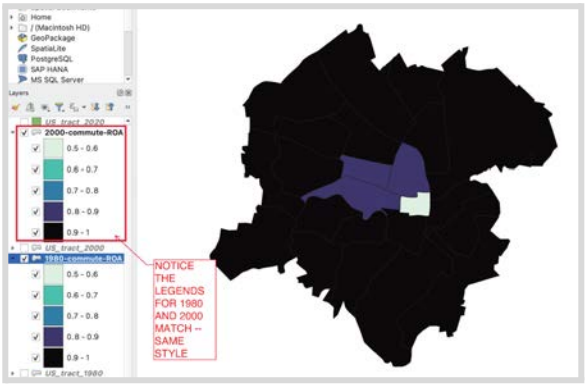
5c Repeat these steps for the other three shapefiles.

Step 6: Create a Print Layout to compare the four decades.

6a Using the “Total” and “Drive” columns, show the percentage of Car commuters for each decade. You may need to use the “to_int” function if your data is text instead of numbers.



6b Be sure to use the same color scheme for each decade. Use the “Save Style As” and “Load Style” functions.



6c In the print layout, include one map for each decade. Include some observations about trends. As always, include a title, legend, scale, north arrow, map labels, your name, and sources.