

# Using the US Census Geocoder

Developed by:

- Heather Squire [heather.squire@temple.edu](mailto:heather.squire@temple.edu)
- Josefina Peralta [tug88995@temple.edu](mailto:tug88995@temple.edu)

## Overview

In this lab we will be looking at 2016 foreclosure data for the City of Philadelphia and comparing it to demographic data at the Census tract level in order to see if there are any possible relationships to explore further. To do this you will first have to geocode the addresses of foreclosed homes using the US Census Geocoder.

## Data

[Download the data](#)

The data we will be using for this lab are: - 2016 Philadelphia Foreclosures (addresses gathered from Redfin, a real estate tracking website <http://www.redfin.com>) - 2010 Philadelphia Census Tracts with housing info from Lab 3 - 2015 American Community Housing Survey - ACS 2014-2016 Data from previous labs.

## Formatting Data for Geocoding

The US Census provides a valuable tool for easily geocoding addresses.

We are able to geocode individual addresses or *batch geocode* up to 1000 addresses. There are over 800 foreclosed homes in Philadelphia, so we will batch geocode using a spreadsheet. Addresses *must* be added in the following format to prevent errors:

Object ID					
A	B	C	D	E	
Object ID	Address	City	State	ZIP	D
1	2135 Walnut St.	Philadelphia	PA	19103	
2	212 S 24th St.	Philadelphia	PA	19103	
3	1901 Walnut St.	Philadelphia	PA	19103	
4	4158 W Girard Ave.	Philadelphia	PA	19104	
5	4224 Wallace St.	Philadelphia	PA	19104	

Once all addresses have been added to your spreadsheet, the spreadsheet must be saved in **CSV** format (Comma Separated Values; .csv). This has already been done in the data above. The format is important if you plan to geocode through the Census Geocoder in the future.

## Geocoding with the Census Geocoder

- Go to <https://geocoding.geo.census.gov/>
- Under “Find Geographies Using...” go to **Address Batch**
- Upload the 2016 Philadelphia Foreclosures CSV file
- Hit **Get Results** [Note: The Benchmark/Vintage are automatically set for current data.] Results may take a few minutes.
- Download the data file called **GeocodeResults.csv**

The result of the Geocode should be in the following format:

	A	B	C	D	E	F	G	
1	338	2506 E Birch St., Philadel	Match	Exact	2506 E BIRCH S	-75.11121	39.98354	R
2	339	1835 E Cambria St., Philad	Match	Exact	1835 E CAMBR	-75.11967	39.991966	L
3	332	3139 Custer St., Philadel	Match	Exact	3139 CUSTER S	-75.11661	39.995937	R
4	333	3141 Custer St., Philadel	Match	Exact	3141 CUSTER S	-75.1166	39.995964	R
5	330	2083 Clarence St., Philade	Tie					
6	331	3127 Custer St., Philadel	Match	Exact	3127 CUSTER S	-75.116646	39.995773	R
7	336	2031 E Auburn St., Philad	Match	Exact	2031 E AUBUR	-75.11867	39.988937	L
8	337	2076 E Birch St., Philadel	Match	Exact	2076 E BIRCH S	-75.11606	39.989403	R
9	334	3309 D St., Philadelphia, F	Match	Exact	3309 D ST, PHI	-75.120865	39.999367	R
10	335	1846 E Atlantic St., Phila	Match	Exact	1846 E ATLANT	-75.10468	40.000015	R
11	531	33 Farson St., Philadelphia	Match	Non_Exact	33 N FARSON S	-75.22179	39.960182	R
12	748	2501 S Colorado St., Phila	Match	Exact	2501 S COLOR	-75.176384	39.920006	L
13	747	2224 Mountain St., Philad	Match	Exact	2224 MOUNTAI	-75.18277	39.92164	L

Note: Addresses are exact or non\_exact, but both are able to be mapped. The Census geocoder conveniently transfers the zip code information into census tracts. Below are the resulting fields from the geocoder:

#### Finding Geographies Using Option

Column	Column Name	Column Description
1	ID	ID from original address list
2	Address1	Address from original address list
3	Matching Result 1	Results indicating whether or not there was a match for the address (Match, tie, no match)
4	Matching Result 2	Results indicating if the match is exact or not (Exact, non-exact)
5	Address2	Address the original address matches to
6	Latitude, Longitude	Interpolated latitude and longitude for the address
7	TIGER/Line ID	Unique ID for the edge the address falls on in the MAF/TIGER database
8	Side	Side of the street address in on (L for left and R for right)
9	State	State FIPS Code
10	County	County FIPS Code
11	Census Tract	Census Tract Code
12	Census Block	Census Block Code

In order to transfer this into QGIS, we have to separate Latitude and Longitude (Column F) into two separate columns. In an Excel spreadsheet, Select (Column F) and then click on **Text to Columns** under the **Data** tab.

**Text to Columns**

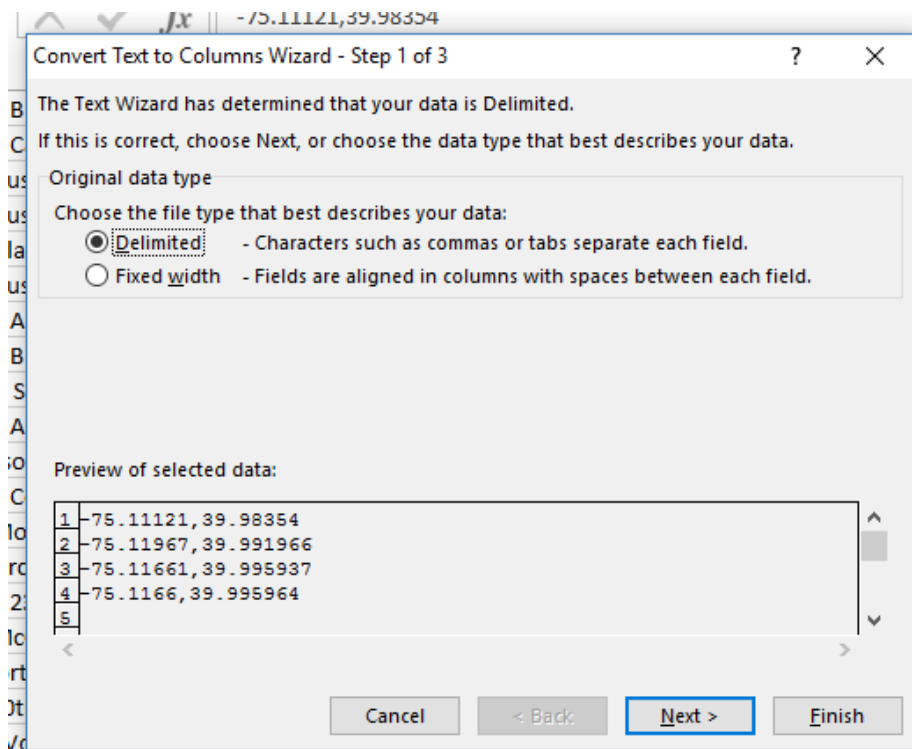
Split a single column of text into multiple columns.

For example, you can separate a column of full names into separate first and last name columns.

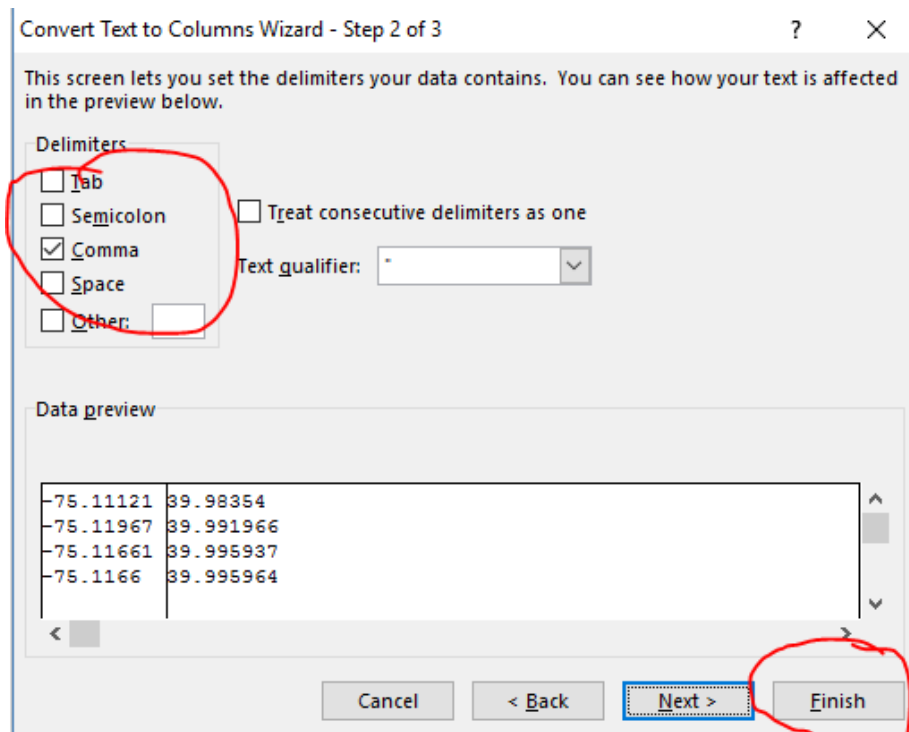
You can choose how to split it up: fixed width or split at each comma, period, or other character.

[Tell me more](#)

Under “Original data type” select **Delimited** and then click **Next**



Choose the **Comma Delimiter**, then Press **Finish**. You now have two separate columns with Latitude and Longitude.



Rename the columns according to the Census Output results in Excel to make transferring data to QGIS easier. *Make sure the Latitude and Longitude are correctly labeled.*

	A	B	C	D	E	F	G	H
	ID	Address 1	Matching Result 1	Matching Result 2	Address2	Latitude	Longitude	Side
	338	2506 E Birch	Match	Exact	2506 E BIRCI	-75.11121	39.98354	R
	339	1835 E Camb	Match	Exact	1835 E CAM	-75.11967	39.991966	L
	332	3139 Custer	Match	Exact	3139 CUSTEI	-75.11661	39.995937	R
	333	3141 Custer	Match	Exact	3141 CUSTEI	-75.1166	39.995964	R
	330	2083 Clarent	Tie					

## Mapping Results

Add the 2010 Census Tracts to your map.

Go to Layer > Add Layer > Add Delimited Text Layer , or use the button



Add your Geocode Result CSV file. Your dialog box should look like the following:

**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)

X field:  Y field:  ☐ DMS coordinates

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

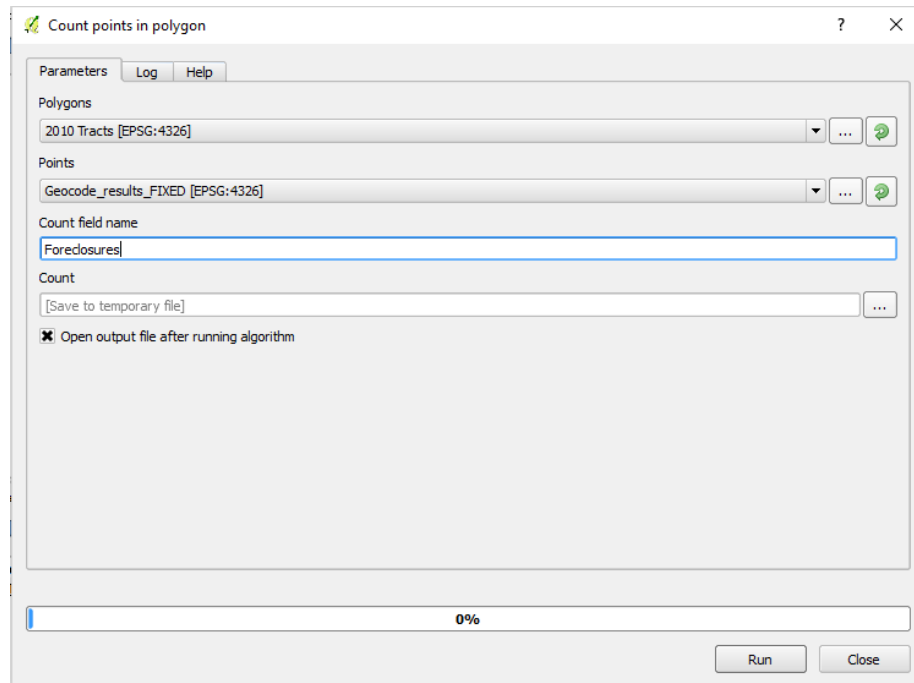
	Object ID	Address 1	Matching Result 1	Matching Result 2	Add
1	338	2506 E Birch St., Philadelphia, PA, 19134	Match	Exact	2506 E BIRCH ST, PHIL
2	339	1835 E Cambria St., Philadelphia, PA, 19134	Match	Exact	1835 E CAMBRIA ST, PI
3	332	3139 Custer St., Philadelphia, PA, 19134	Match	Exact	3139 CUSTER ST, PHIL
4	333	3141 Custer St., Philadelphia, PA, 19134	Match	Exact	3141 CUSTER ST, PHIL
5	330	2083 Clarence St., Philadelphia, PA, 19134	Tie		

Hit OK.

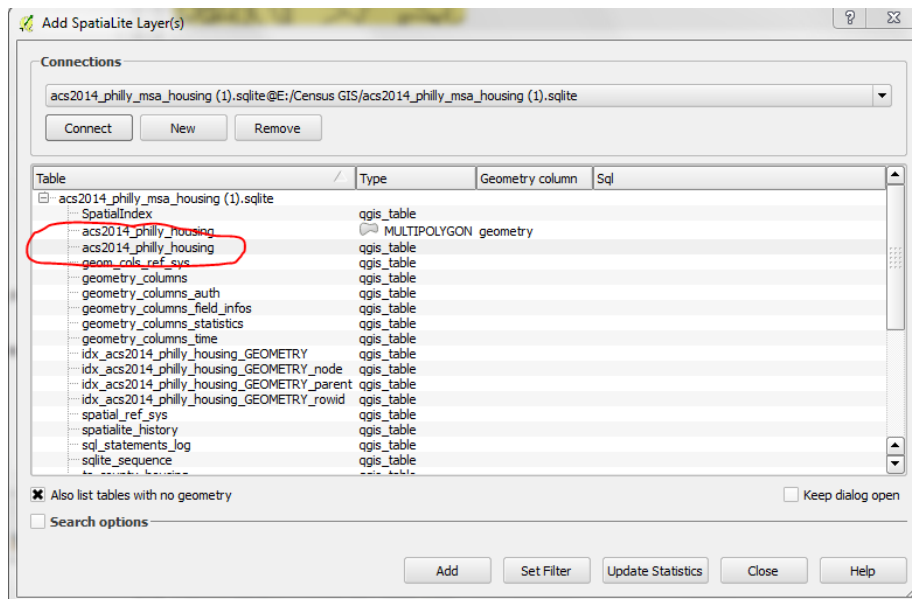
[Note: If the projection is missing, the 2.16 version of QGIS assumes a projection with Latitude and Longitude which is the appropriate projection for this lab exercise. If you are using an older version of QGIS, it will prompt you for a projection, which in this case is WSG 84.]

We have created a point shapefile from this data. We want to transfer that information to create a count of foreclosures per census tract. As in a previous lab, we will be using the Count Points to Polygon tool.

Go to **Toolbox > Geoprocessing > Vector > Statistics > Count Points in Polygon**. It should look like the following:



Rather than the the actual number of foreclosures, we are interested in mapping the foreclosure rate, which compares the number of foreclosures in the census tract to the total number of residential housing units in that tract. We are interested in the 2010 adjusted housing rate per census tract which we created in Lab 3. In order to compare the two, we have to **join** our lab data from Lab 3 by connecting to the **acs2014\_philly\_msa\_housing** Spatialite database in QGIS. Select **Also list tables with no geometry** and then select **acs2014\_philly\_housing** and click Add.



Now we are able to join the housing data to our tracts shapefile. Perform a spatial join based on Census tract.

## Assignment

Create two side-by-side choropleth maps comparing foreclosure rate in Philadelphia with some other demographic attribute along census tracts, which you think may be related to foreclosures (ie: race, income, poverty rate, etc.). Use the ACS 2014 data from previous labs. Make sure to set an appropriate projection. Include a title and legend that adequately communicate the relationship your two maps describe.

*Hint: In order to create the choropleth map for foreclosure rate, make sure to use the Expression Dialog under the Graduated Symbolology section.*