Geocoding and buffering in QGIS

Kyle Walker

Wednesday, June 18, 2014

Step 1: Getting your data ready

To get started, please download the CSV available at this link: https://dl.dropbox.com/s/nce0umurze01s4q/target.csv. This is a CSV of Target stores in Texas, which you are going to geocode.

Geocoding refers to the conversion of a description of a place (e.g. an address) to actual geospatial coordinates that you can locate on a map. This what you do every time you search for something in Google Maps, and indeed you are going to use the Google Maps geocoding service through QGIS to locate your points. Be warned, however, that geocoding is an inexact science and can sometimes be wrong; you'll want to do some quality checking on your data after you've completed the geocoding process.

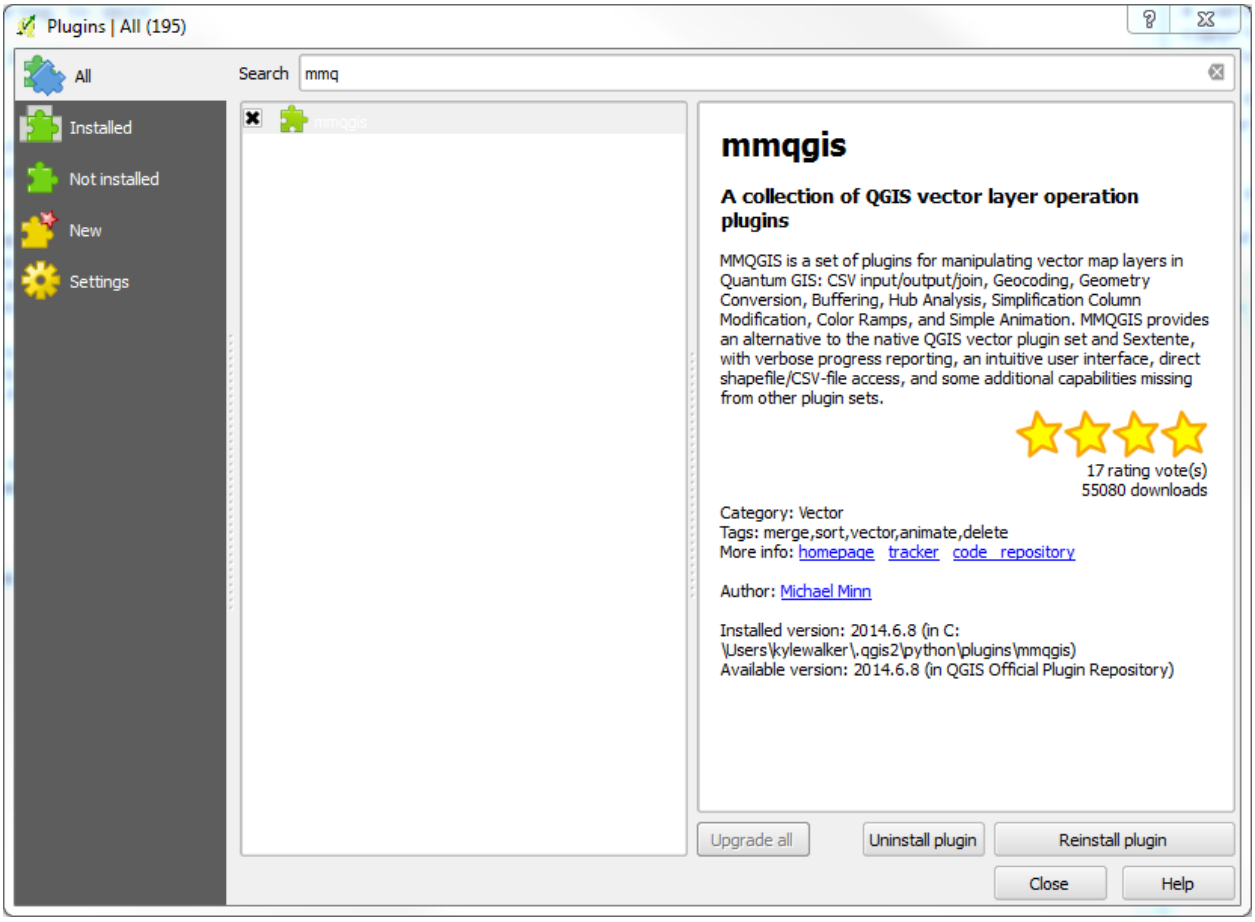
The text in this CSV file is in **UTF-8** format, which is the format required by the geocoding tool you are going to use. I've converted the file to UTF-8 for you, but this may be an issue you'll run into. Save the file to your computer.

Excel does not encode CSVs as UTF-8 by default, and the tool will not work otherwise. To convert a CSV to UTF-8 in Windows, right-click the CSV and choose ***Open with > Notepad****. Then, save the file with the* ***Save As*** *command, making sure that UTF-8 is selected from the "Encoding" drop-down menu.*

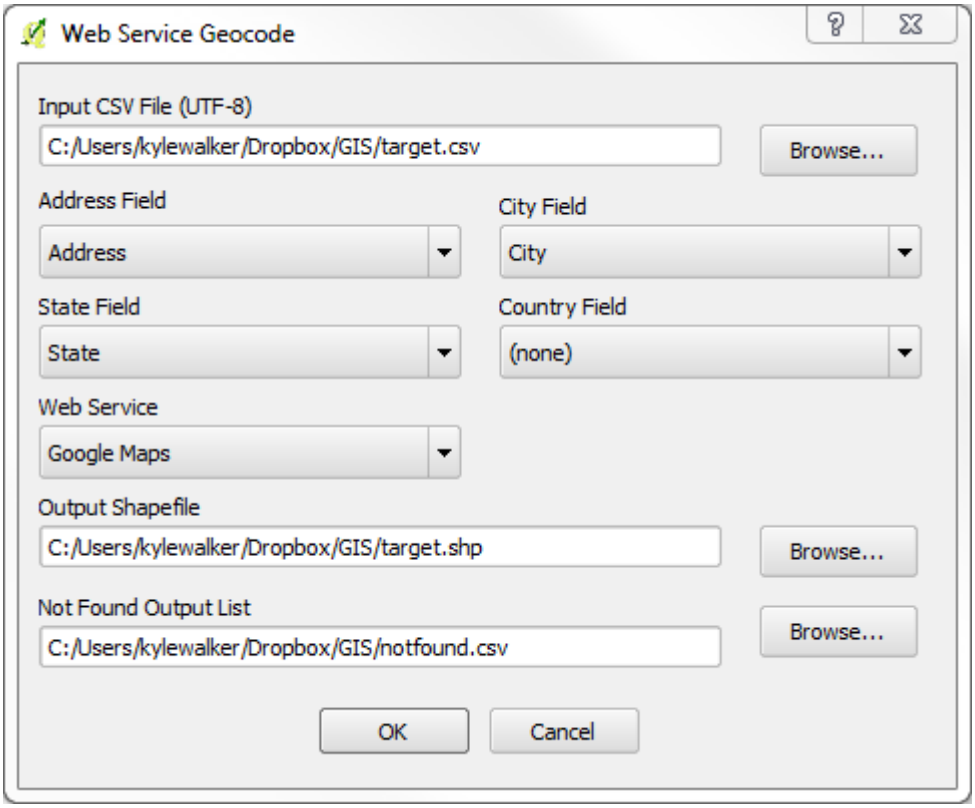
A brief note when you are doing this yourself. The MMQGIS plugin requires that your CSV be formatted a specific way to work. The street address should be in one column, the city in another, and the state in another. If they are all together in the same column, the geocoder will not run correctly.

Step 2: Geocoding the addresses

Now, start up QGIS, and look for the **Plugins** drop-down menu. Click **Manage and Install Plugins**. Click **All**, and search for the MMQGIS plugin.



Install the plugin; once it is installed, you should see a drop-down menu for MMQGIS available to you. Click **MMQGIS>Geocode>Geocode CSV with Google / OpenStreetMap**. Make sure all the settings look similar to those in the image below. You'll be using the Google geocoding service, and saving your results to a shapefile. Click OK to run the tool; it should take a couple minutes to run through all the addresses (Google places limits on how fast you can run geocoding queries against their servers).



If the tool has run correctly, you should see points representing the location of Target stores show up on your map, and vaguely be able to make out the shape of Texas. To assess the accuracy of your geocoding, right-click your new shapefile in the "Layers" window and choose **Open Attribute Table**. This will show you the "attributes" of your data, which appear like a database table. The fields "addrtype" and "addrlocat" will tell you how the addresses were geocoded. "ROOFTOP" means that the geocoder was able to locate the specific building; this is the most accurate. Some stores may just have approximate locations, meaning that the geocoder was not able to locate the precise address.

Step 3: Projecting your data

Spatial data are somewhat different from the other types of data you're going to be working with in your job. They are unique in that they have *geometry*, which consist of X & Y coordinate pairs that define how the data represent features on the Earth's surface – in other terms, where they are located. The trick is, there are many different ways to define coordinates – these are called ***coordinate systems*** – which provide a reference for making measurements on part, or all, of the Earth's surface. Related to coordinate systems are ***map projections***, which refer to the way in which your geographic data are then displayed on your screen (or on a map).

Coordinate systems are complicated – many advanced GIS students struggle with the topic – and they are beyond the scope of this tutorial. The thing is, however, if you get your coordinate system wrong, you are likely to make incorrect inferences with your spatial data. If you continue with your GIS work, I strongly recommend picking up a book on the topic and working through it.

Here's the general gist of it, however. Your Target stores have been geocoded to a **geographic coordinate system**, which uses latitude and longitude coordinates on a three-dimensional surface. If you were to try to calculate distances around these stores now, it would honor this coordinate system and calculate it in terms of decimal degrees (you don't want this). As such, you'll want to **project** your data to a two-dimensional surface that uses distance units we are more familiar with (meters).

To do this, right-click "target" in the Layers pane and choose **Save As**. Keep the **Format** as ESRI Shapefile. Define an output shapefile in your folder. Change **CRS** to **Selected CRS**. Now, click **Browse**. In the **Filter** box, type `3082`. This is a coordinate system that works for all of Texas, Texas Centric Lambert Conformal. Select it, and click **OK**. Check **Add saved file to map**, and click **OK**.