

Geocoding Lab Exercise

By Jeff Blossom: jblossom@cga.harvard.edu

June 3, 2020

Introduction

The ability to assign specific geographic locations to textual information (the process known as geocoding) is available to anyone with a computer and internet access. The relative ease of geocoding and resulting accuracy can vary widely depending on a number of factors. What is the nature of the data? How 'clean' is it and what format is it in? What geocoding technique will be used? Determining a geocoding strategy that best suits a particular need is not always clear.

The process of geocoding begins with comparing data in text or tabular form to a reference data table in geographic format. The reference table is a dataset that has already been mapped, with established map coordinates. When matches between the input data and the reference data are found, the corresponding map coordinates are assigned from the reference data to the input features, thus geocoding them. A geocoding service (also called an address locator) is a program that allows for a user to input a batch of data contained in a table, search for matches as compared to a reference table, and output the result in a map or GIS layer format. The key to confidently geocoding data lies in understanding the reference table which the data is being matched to, how a match is found, and the resulting spatial accuracy.

Objective: The goal of this exercise is to give everyone practice performing geocoding using different methods, and to gain insight regarding how to successfully geocode.

Geocoding lab

In this exercise you will geocode single addresses and cities using Bing Maps and Google Maps. You will geocode lists of multiple addresses and cities using ArcGIS Pro, Google Maps, and ArcGIS Online. You will create GIS datasets, online maps, and desktop maps of the results. You will learn how to evaluate the accuracy of the data using base maps.

Lab Setup: From the GIS Institute Summer 2020 page, download and unzip the **(lab data)** file under the **Day 2 Geocoding & Introduction to the Harvard GGA** topic. This ZIP file contains a folder named *Geocoding_lab*. Unzip/Extract this folder onto your Desktop.

1) Geocode one address with Bing and Google Maps

In an internet browser, go to <http://www.bing.com/maps/>

Type an address, such as the CGA's address "1737 Cambridge St. Cambridge, MA " into the

window at the top, and click the Search button: 

The map will zoom to this location, and the latitude, longitude values will be displayed at the bottom of the information window (see below). You can click Copy, and paste into a word or other document.

United States · MA · Middlesex Co.
Square
42.375583, -71.113257

To find the latitude/longitude of any point on the map, just right click anywhere on the map. The two numbers listed at the bottom of the pop-up window are the latitude, longitude of that location. *This could be helpful if you want to find the location of specific points, (for example a park bench) instead of just the location the geocoder returns.* Right click somewhere on the Bing Map and copy the coordinates. Then paste these coordinates into the search window of Google Maps at <http://maps.google.com>. The Google map will zoom to this same latitude/longitude coordinate. To view the latitude/longitude location of an address on Google Maps, you must right click on the location, and choose “What’s here?”. Try searching for other places, like city names, water bodies, or famous locations (such as the Eiffel Tower). Google Maps, Bing Maps and most online geocoders will be able to locate “non-address” data as well.

1.1 Geocode multiple cities with Google Maps

In Windows File Explorer, browse to the Geocoding_lab folder and double click the World_Livable_Cities.csv file. This will open it in Excel. Notice that this contains a list of city, country locations, and a “livability” ranking and index. This list was generated by Business Week magazine, and downloaded in 2012. You’ll use Google Maps to geocode these cities.

1.2 In a web browser, go to <https://drive.google.com/drive>. Login with your Google Account. Click **New > Google Sheet**. Click **File > Import**. In the Import File dialog, click **Upload**. Browse to your Geocoding_lab folder and select the **World_Livable_Cities.csv** file and choose **Open** and click **Import Data**. Choose to **Create new spreadsheet** and click **Import Data**. Now you’ve uploaded your .csv file into a Google Sheet. Click the **Open data** button to open the sheet.

1.3 Go to <http://maps.google.com>, and login if you aren’t already. Click the **Menu** button, it looks like this:



Then click **Your Places**. Click the **MAPS** option at the top, and **CREATE MAP** at the bottom. Now it is time to geocode your spreadsheet locations. Click the blue **Import** button underneath where it says Untitled layer. Then choose **Google Drive**, find your **World_Livable_Cities** sheet, click it, and click **Select**.

When prompted to “Choose columns to position your placemarks” click the box next to **City-Country** and click **Continue**. When prompted to “Choose a column to title your markers”, choose **City-Country** as well. Click **Finish**. These locations will be geocoded, and automatically added to your Google Map. You can save or share this map or export the geocoded locations to a .kml file for use in ArcGIS or other GIS software. To export these geocoded locations, click the

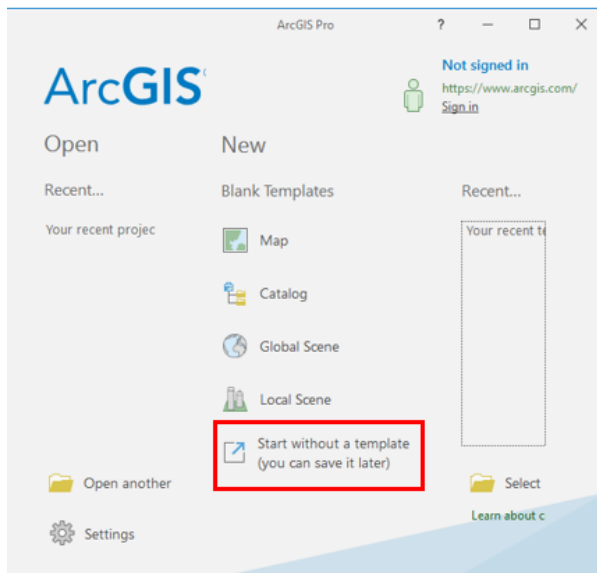
3 vertical dots button:  and choose **Export to KML/KMZ**.

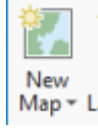
Batch geocoding with Google Maps is also possible for up to 2,000 addresses. For more than this, the Google Maps Platform API is required. More information and a Python script to access the API can be found on the CGA’s geocoding page: <https://gis.harvard.edu/geocoding>

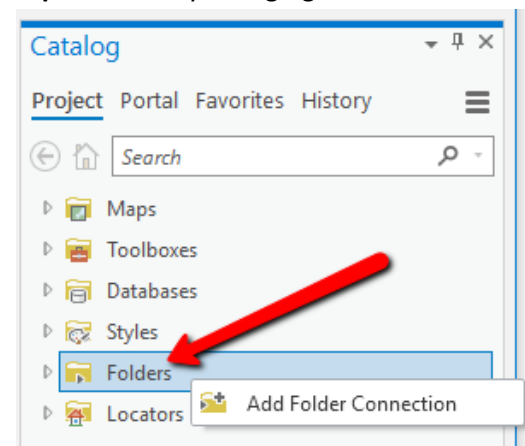
2) Geocode cities and addresses with the ArcGIS World Geocoder in ArcGIS Pro

In this exercise you will geocode 50 worldwide cities with ArcGIS Pro software

2.1 Open ArcGIS Pro by clicking **Start >> ArcGIS >> ArcGIS Pro** Choose “**Start without a template**” to begin (see screenshot below).



2.2 Create a new map and connect to a folder. Click the New Map button  to add a new map to your project. The World Topographic Base Map displayed is fed into ArcGIS Pro from a server online. You can change this base map by clicking the **Basemap** button. Try changing some of the base maps, and zooming in and out using your mouse scroll wheel to explore the different map styles. The **Catalog** pane (see screenshot at right) should be open, but if it's not, click View at the top and Catalog Pane to open it. In the Catalog pane, right click **Folders** and choose **Add Folder Connection**. On the left side of the Add Folder Connection screen click **Desktop**. On the right side of the screen click on **Geocoding_lab** and click OK. Now, in Catalog, double click Folders > Geocoding_lab and you'll see the World_Livable_Cities.csv file. Click on this file and drag it onto the map.



2.3 Log into ArcGIS Online. In order to geocode the world livable cities, access to a reference dataset is necessary. We'll use the ArcGIS World Geocoding Online Service for this, which requires logging into the ArcGIS Online server through ArcGIS Pro. To do this, click in the upper right of ArcGIS Pro where it says Not signed in, and choose Sign In. If you have a Harvard Key, follow step A below. If you don't have a Harvard Key, follow step B:

A) Click the **Enterprise login** button, and enter “harvard-cga” and click the blue **Continue** button. Click the blue button that says “**Harvard University...**”, and enter your Harvard Key information to login.

B) Enter **GIS_Institute** for the login and **ARCGIS2019** for the password and click SIGN IN. This signs you into a temporary account that will be active during the GIS Institute.

2.4 Geocode the world livable cities. In the Contents pane, right click the **World_Livable_Cities.csv** file and choose **Geocode Table**. In the Geocode Table interface, read through the steps, and then click **Go to Tool**. For the Input Locator click the dropdown arrow and choose **ArcGIS World Geocoding Service**. Change Input Address Fields to **Single Field**, and choose the **City_Country** field as the **Single Line Input**. Your screen should look like the one on the right →. Click **Run** at the bottom to geocode the cities.

Geocode Table

When using the ArcGIS World Geocoding Service

Input Table: World_Livable_Cities.csv

Input Locator: ArcGIS World Geocoding Service

Input Address Fields: Single Field

Locator Field: Data Field

Single Line Input: City_Country

Output: World_Livable_Cities_Geocoded

☒ Add output to map after completion

Preferred Location Type: Routing Location

Country: All

Category:

Run

2.5 Inspect the geocoded results. In the Geocoding Completed pop up window click **No** when prompted to "Start rematch process?". A new layer named **World_Livable_Cities_Geocoded** will be added to your map. Inspect the results of your geocoding by clicking on a dot on the map to identify it, and then comparing it's location to the base map. **Right click** the **World_Livable_Cities_Geocoded** layer and choose **Attribute table** to open it's table and inspect the additional fields the geocoding process added.

Open the attribute table and note the **Match_addr** field. This field contains the exact address the input was matched to. Also note the **Addr_type** field. This specifies what locator was used to match the address.

Addr_type = AddressPoint means the address was mapped using a location at the exact address location.

Addr_type = StreetAddress means the address was mapped using street range interpolation.

Addr_type = NameStreet indicates a match to the street only, placed halfway along the length of the street.

Addr_type = StreetInt indicates a match to the intersection of two streets.

Addr_type = Zipcode and Locality indicate a match to the geographic center of the ZIP code and City, respectively.

Geocoding results can be refined and adjusted by right clicking the layer and choosing **Data > Rematch Addresses**. This will open the Rematch Addresses dialog box. Here you can view the Unmatched and Tied addresses, edit the input address information, and change the matched locations, if desired. Your instructor will demonstrate some of this functionality.

2.6 Geocode 1,000 LA County crime addresses

Now you'll geocode 1,000 addresses of crime locations in Los Angeles County. This is part of a dataset downloaded from the LA County GIS Data Portal:

<https://egis-lacounty.hub.arcgis.com/>

2.7 Add the LA_crime_2005.csv file to your map by dragging it onto the map from the Catalog pane. In the Contents pane, right click the **LA_crime_2005.csv** file and choose **Geocode Table**. In the Geocode Table interface, click **Go to Tool**. For the Input Locator click the dropdown arrow

and choose **ArcGIS World Geocoding Service**. Change Input Address Fields to **Single Field**, and choose the **ADDRESS** field as the **Single Line Input**. For Country, choose **United States**. Click **Run** at the bottom to geocode the addresses. Inspect the result on the map and open the resulting attribute table. Experiment with **Data > Rematch Addresses** if you want to practice refining the locations of your addresses.

3) Geocode 2,000 addresses with ArcGIS Online

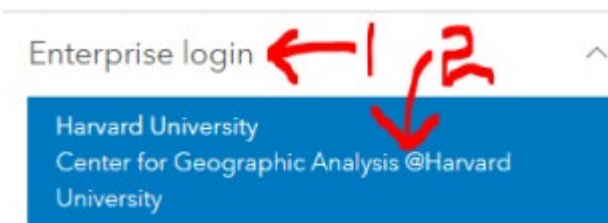
3.1 Go to the Harvard ArcGIS Online Enterprise account page:

<https://harvard-cga.maps.arcgis.com/home/index.html>

3.2 Click Sign In at the upper right:



3.3 Click Enterprise login, and then the blue button named “Harvard University Center for Geographic Analysis @Harvard University”. See below.



3.4 Login with your Harvard Key, including the two step verification.

3.5 Once logged in, click **Map** at the top – this will create a new blank map for you.

3.6 Click **Add > Add Layer from file**, and click the **Choose File** button. Browse to your Geocoding_lab folder, and choose a .csv file that contains the fields Location and Count and choose **IMPORT LAYER**. This will open the Add CSV Layer dialog box.

3.7 At the top of the Add CSV Layer box, specify **Addresses or Places** for the Locate features by choice. Next to the Location field name, click under Location Fields where it says “Not used”, and change this to “Address or Place”. Make sure the other fields are specified as “Not used”. Click **ADD LAYER**, and the geocoder will run.

The geocoded results are now a feature layer in ArcGIS Online available for mapping and analysis. This feature layer can also be accessed through ArcGIS Pro and ArcMap.

Important Geocoding Note: If addresses cannot be loaded onto a server due to HIPAA compliance or other restrictions, then these online Esri and Google geocoders cannot be used. Instead, a local copy of the 2019 Esri geocoder (for the USA only) can be downloaded

for use within ArcGIS Desktop or ArcGIS Pro software. Your instructor will demonstrate this. See more on downloading this and other local geocoders at the CGA's geocoding website: <https://gis.harvard.edu/geocoding>

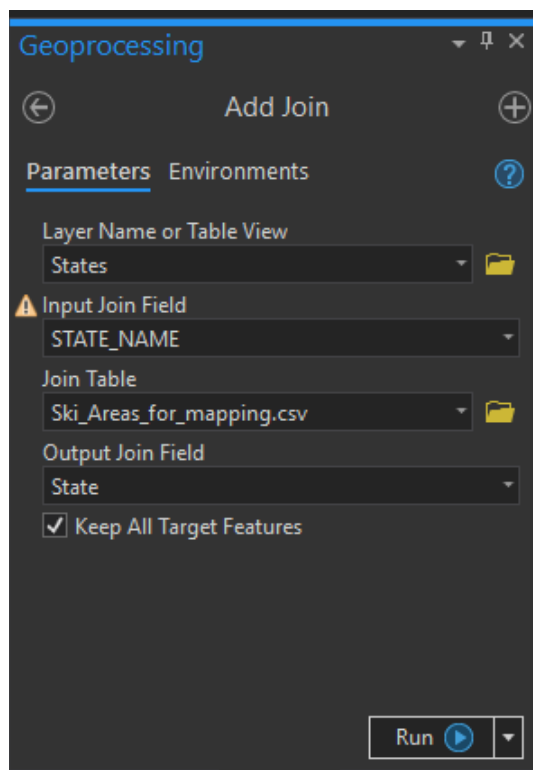
4) Geocode to an existing dataset.

So far you've geocoded addresses and cities into point locations on the map. Often you may want to geocode data to non-point features, such as administrative areas. In this step, you'll geocode a list of US ski resort counts per state to an existing States polygon shapefile. This list of ski resorts is found in the `Ski_Areas_for_mapping.csv` file. Double click it to open in Excel to examine the contents.

4.1 In ArcGIS Pro in the **Catalog** pane, find the **Ski_Areas_for_mapping.csv** file in the `Geocoding_lab` folder, and drag it onto the map. Next, find the **States** shapefile in the `Geocoding_lab\GIS_Data` folder. Click it and drag it onto the map. Right click **States** and choose **Zoom to Layer**.

4.2 Right click the **States** layer and choose **Attribute Table**. Note that the `STATE_NAME` field contains the name of the states. This field is what you'll use to geocode the Ski Areas to. Close the States attribute table.

Right click the **States** Layer and choose **Joins and Relates > Add Join**. In the Add Join dialog, specify to join from the `STATE_NAME` input field to the `State` field in the **Ski_Areas_for_mapping.csv** table. Make sure to check the box next to **Keep All Target Features**. Your Add Join dialog window should look like the one below:



Click **Run** to execute the Join.

4.3 Right click the **States** layer and choose **Attribute Table**.

Scroll over to the far right of the table, and you should see the **Ski_areas** field at the very end.

Now you can symbolize your map by this Ski Areas statistic by **right clicking the States layer**, choosing **Symbology**, and in the Symbology pane changing **Single Symbol** to **Graduated Colors**. Then choose the **Ski_areas** as the Field to symbolize by.

Conclusion - General tips to keep in mind when geocoding:

- Clearly determine your geocoding purpose, and the accuracy level required to meet this need. If you just need to geocode your data to the city or country level, don't complicate things with address information. However, if address level accuracy is necessary, take the proper steps to ensure confidence in your result.
- Have an 'iterative' mindset. Geocoding may reveal errors or typos in your data, or expose the shortcomings of the geocoding method used. Be prepared to re-geocode, and refine your data and geocoding process accordingly – several geocoding iterations may be necessary to achieve the desired result.
- Don't treat one basemap as the gospel. Compare your results to more than one basemap. For example, if geocoded in ArcMap, import the results to Google Earth to see if they match GE's basemap.

Additional notes:

- Many cities and some states have built custom geocoders from their internal street map that run in ArcMap, and make them available for use. These often will produce better results than the ArcGIS or Google geocoders.
- If you have private or confidential data that is now allowed online, you must use a local geocoder that runs on your C drive (The ArcGIS geocoder, or one from a particular city). Another option for local geocoding is the DeGAUSS geocoder, a free, open source software that runs on multiple operating systems:

<https://github.com/cole-brokamp/DeGAUSS/wiki/Geocoding-with-DeGAUSS>

For the latest on geocoding at the CGA visit our geocoding site:

<https://gis.harvard.edu/geocoding>

For GIS and geocoding questions and help contact the CGA at: <http://gis.harvard.edu/contactus>