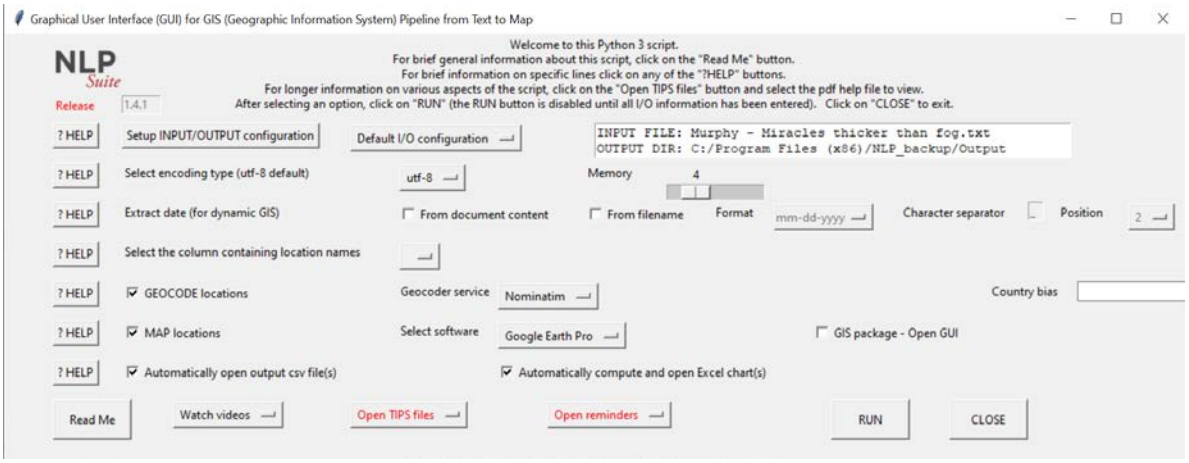
Geocoding

Geocoding: What is it?

Geocoding is the process of converting addresses (like “1600 Amphitheatre Parkway, Mountain View, CA”) into geographic coordinates (like latitude 37.423021 and longitude -122.083739), which you can use to place markers on a map, or position the map.

Geocoding in the NLP Suite

The NLP Suite provides two free approaches to geocoding locations, via Nominatim and via Google. These options are available from GIS_main.



Other NLP Suite scripts (e.g., GIS_Google_Earth_main, SVO_main), also have geocoding capabilities. GIS_Google_Earth_main and SVO_main rely on Nominatim as the default

geocoding option. If you wish to use Google for geocoding, please, use the GIS_main script.

Whichever option is chosen for geocoding (Nominatim, Google) the NLP Suite will carry out the geocoding automatically. But… other options are available. See below.

Nominatim

Nominatim is an open-source, freeware geocoding tool. Nominatim uses OpenStreetMap data to find locations on Earth by name and address (geocoding) (see below for additional information on OpenStreetMap). Nominatim can also do the reverse, find an address for any location on the planet.

Nominatim is the default geocoding option when running the GIS_main script.

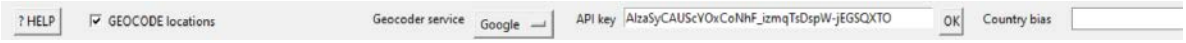
Nominatim limitations

If the Nominatim geocoder service exits with the error "**too many requests**," you can break up the csv location file and process each subfile for geocoding as normal csv files.

Google

When you select the Google geocoding tool the API key widget will be displayed. You will need to enter there the free Google API key (see the TIPS on Google API).

You will need to enter the API key only once. From then on, the Suite will use the saved key when you selected Google as the preferred geocoding tool.



Google limitations

With the basic Google Earth or Google Maps, you can find only one location at a time by typing its physical address or geographic coordinates (latitude and longitude) into the search box. **Google Earth Pro lets you import up to 2,500 locations at a time**. For larger datasets, you can break up the location data into batches of 2,500 items or less, then import each batch separately. To perform batch geocoding, you need to have the addresses or coordinates in one or more .csv (comma separated values) or .txt (plain text) files. You can export such files from Microsoft Excel and other spreadsheets. Check the "Importing Addresses" section of the Google Earth help documentation for information about formatting these files.

Faulty geocoding results and how to deal with it

Whether you use Google or Nominatim as your geocoding tool, after the geocoding and

mapping is done, please, check carefully the results. If you are geocoding locations such as Athens or Rome in Georgia, most likely they will be geocoded in Greece and Italy. If you specify the United States as the country bias, the geocoder may select Rome, New York, or Indiana, or Illinois, rather than Georgia. To make sure the geocoded Rome is in Georgia, you may need to edit the geocoded csv file, adding Georgia as the state, e.g., Rome, Georgia.

Databases/Gazetteers

There are also freeware available databases that one can download and then search for waypoints or build an application that can read an Excel file and batch geocode. The problem with these databases is that waypoints are given in decimal degrees and these will need to be converted since most GIS programs do not work with decimal waypoints. You are better off using the applications listed above and that, in any case, rely on the databases listed here.

OpenStreetMap

http://www.openstreetmap.org/about

OpenStreetMap is the database used by **OpenRefine**, and many other applications.

GeoNames

The **GeoNames** geographical database covers all countries and contains over eight million placenames that are available for download free of charge. See the GeoNames website at http://www.geonames.org/

WhosisOnFirst

WhoIsOnFirst is an online system for developers that is attempting to eliminate some of the errors in the previous two databases (https://whosonfirst.org/).

Further geocoding options

Google Maps

Google Maps provides an API that will do geocoding for you via an HTTP request, where an API is an Application Programming Interface, i.e., a set of routines, protocols, and tools for building software applications.

https://developers.google.com/maps/documentation/geocoding/intro?utm_source=google&utm_medium=cpc&utm_campaign=2015-Geo-NA-GEO-MAB-GoogleSearchDev&utm_content=NBINNPSGeocoding&utm_term=dev&gclid=CIWdgpOUv8gCFdgRgQodnMgAlw

Google Earth Pro

Google Earth Pro (http://www.google.com/earth/)

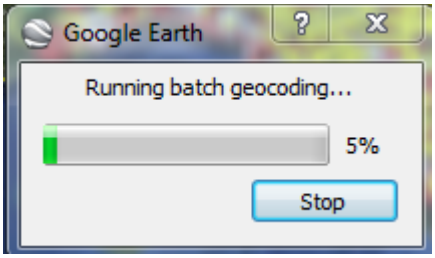
In January 2015, Google made available for free its Google Earth Pro. You can download it at http://www.google.com/earth/download/gep/agree.html

Note: Google Earth Pro requires a license key. If you do not have a key, use your email address and the key **GEPFREE** to sign in.

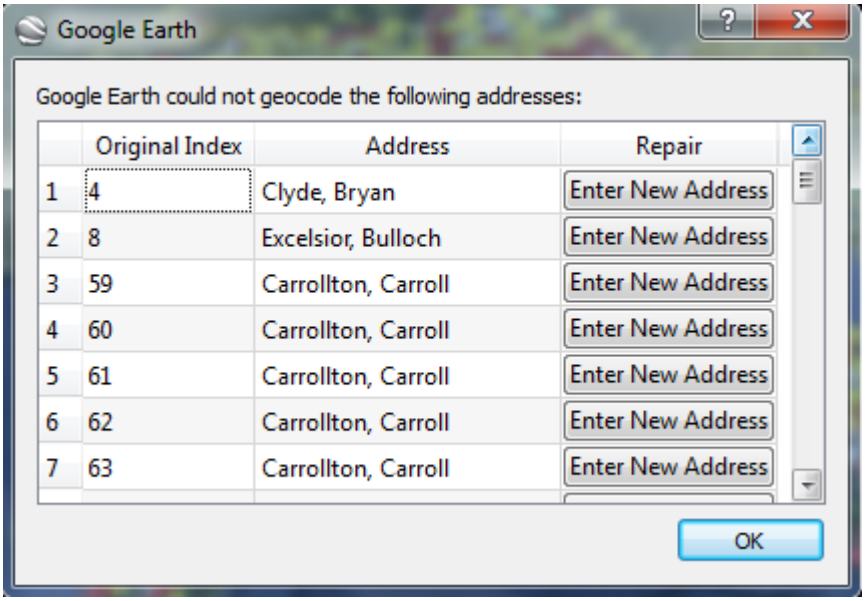
The software runs on computers with Windows (XP and later) and Apple (OS X v. 10.6.0 or later) operating systems.

From the Google Earth Pro menu, click on File and then Import to open a dialog box that let us select the file with the addresses. The Data Import Wizard lets you verify that the .csv or .txt file structure was correct. If the data contains latitude/longitude information, users would click Next to go to a second page, where they'd verify that the proper latitude/longitude fields have been selected.

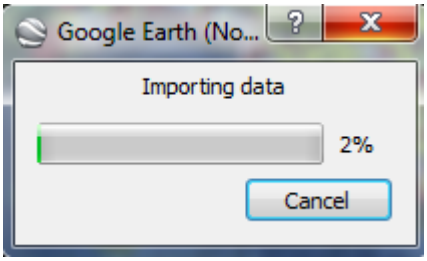
Otherwise, click the checkbox "This dataset does not contain latitude/longitude information, but street addresses", then click Next.



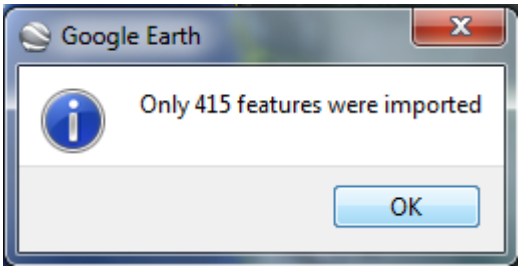
At the end of the geocoding process, you may get a form that displays all records that could not be geocoded



You will then get a progress form for data import



When errors are encountered, Google Earth will display the following message:



Features of Google Earth and Google Earth Pro

Features	Google Earth	Google Earth Pro
Print images	Screen resolution only	High resolution (up to 4800 x 3200 pixels)
Regionate large datasets	No	Yes
Batch geocode addresses	No	Automatically Geo-locate up to 2500 at a time
Import GIS data	No	ESRI shapefiles (.shp) and MapInfo (.tab) files
Import GIS images	Manually geo-locate	Automatically geo-locate
Import large image files	Up to max texture size	Super Image Overlays when larger than max texture size
Access demographic, parcel, and traffic data layers	No	Yes
Create premium movies	No	HD 1920 x 1080 pixels
Measure area of a polygon or circle	No	Yes
Map multiple points at once	No	Yes
Viewshed tool	No	Yes
Map-making tool	No	Yes

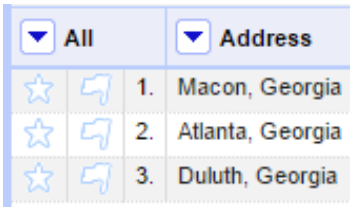
Table 1. Comparison of some key features of Google Earth and Google Earth Pro.

OpenRefine

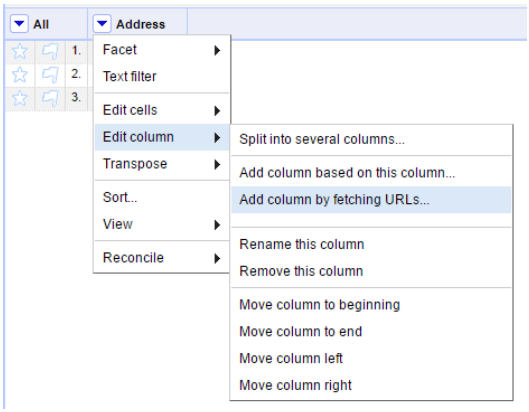
OpenRefine (http://openrefine.org/) allows you to geocode using the Google Geocoding API.

To do so, first import a file with at least the list of addresses you wish to geocode. OpenRefine supports the following file formats: TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents.

We have imported a list of three cities in Georgia that we'd like to geocode. It is important that both city and state are listed together in the same column formatted as: City, State

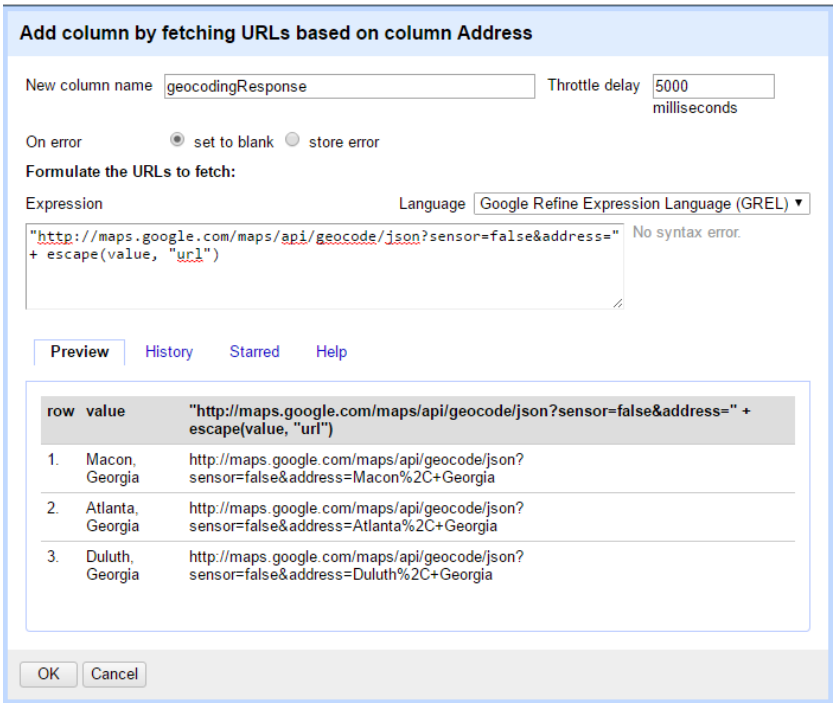


Now, click on the arrow in the column containing your addresses with the City, State format and select **Edit column** > **Add column by fetching URLs**…

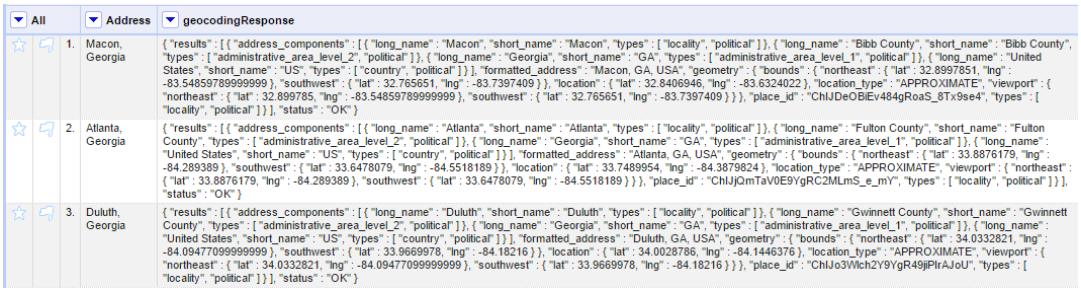


In the form that pops up, give your new column a name such as “geocodingResponse”. Paste the following into the “Expression” box:

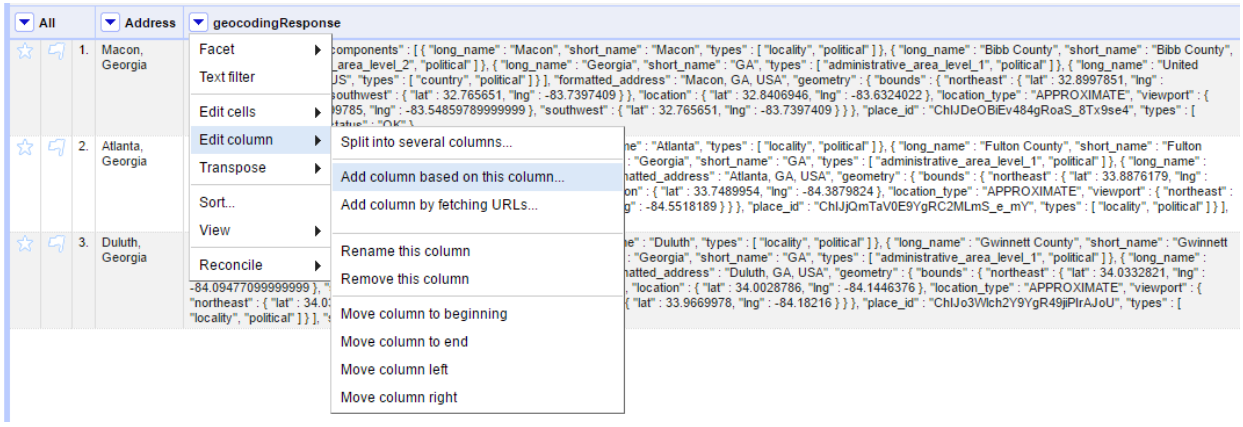
```
"http://maps.google.com/maps/api/geocode/json?sensor=false&address=" +
escape(value, "url")
```



Set the throttle delay to 500 milliseconds. Hit OK. The next step may take some time, but you will eventually have a new column looking something like this:

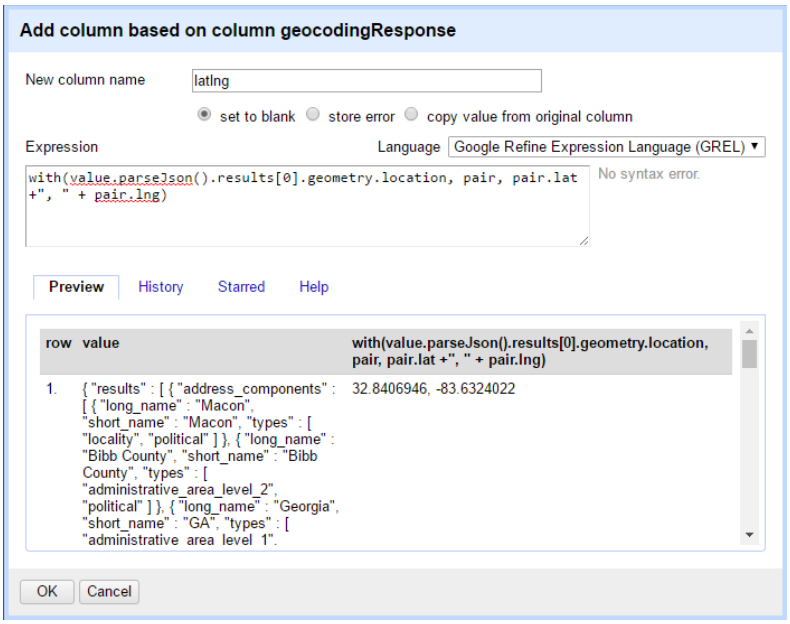


In order to clean up this data, click on the arrow in your newly-created column and select **Edit column** > **Add column based on this column**…



In the form that pops up, give your new column a name such as “latlng”. Then, paste the following into the “Expression” box:

```
with(value.parseJson().results[0].geometry.location, pair, pair.lat +", " + pair.lng)
```



Hit OK. You will now have a third column containing the coordinates of your addresses. To clean up further, you can remove the geocodingResponse column (**Edit column** > **Remove this column**) to get the final result:

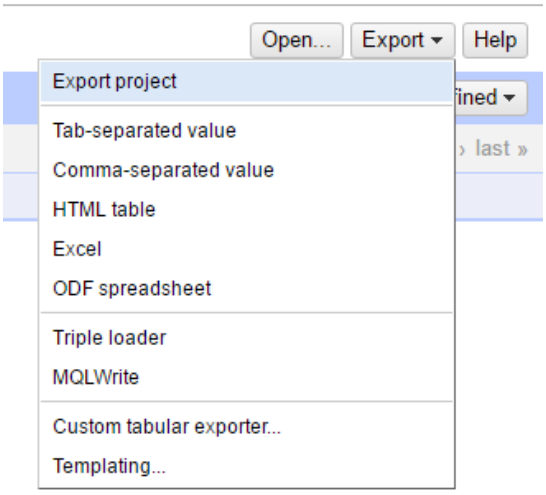
All		Address	latlng
1.		Macon, Georgia	32.8406946, -83.6324022
2.		Atlanta, Georgia	33.7489954, -84.3879824
3.		Duluth, Georgia	34.0028786, -84.1446376

If you wish to have latitude and longitude in separate columns, you can do so by going to **Edit column** > **Split into several columns**…

The comma serves as the delimiter, so you can leave the settings as they are and hit OK. OpenRefine will then split the "latlng" column into two columns, labeled "latlng 1" and "latlng 2". You can rename them for clarity (**Edit column** > **Rename this column**), such as in the following screenshot:

All		Address	lat	lng
1.		Macon, Georgia	32.8406946	-83.6324022
2.		Atlanta, Georgia	33.7489954	-84.3879824
3.		Duluth, Georgia	34.0028786	-84.1446376

You can then export this data to various formats:

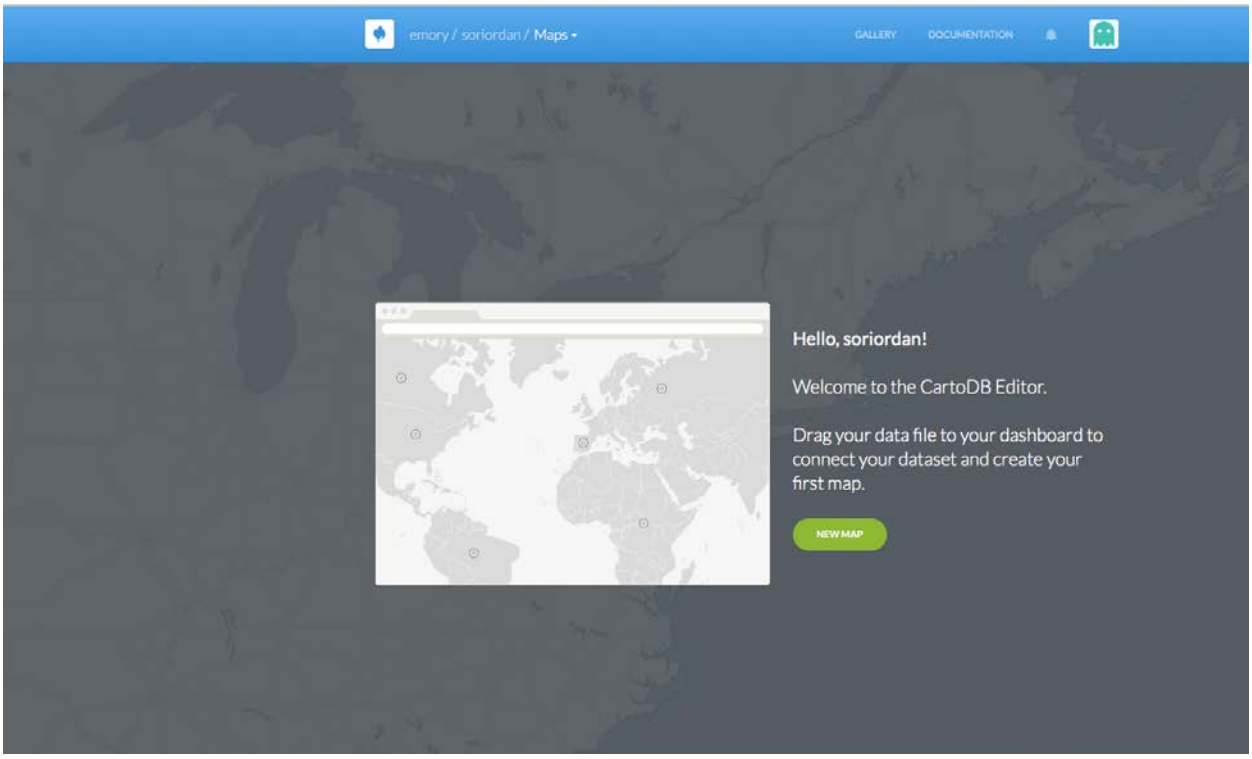


CartoDB & QGIS

You can also use a combination of CartoDB and QGIS (both free software) to geocode placenames and export coordinates, respectively. CartoDB is a web-based map-making and GIS analysis software that is free to use for the task at hand.

Geocoding in CartoDB

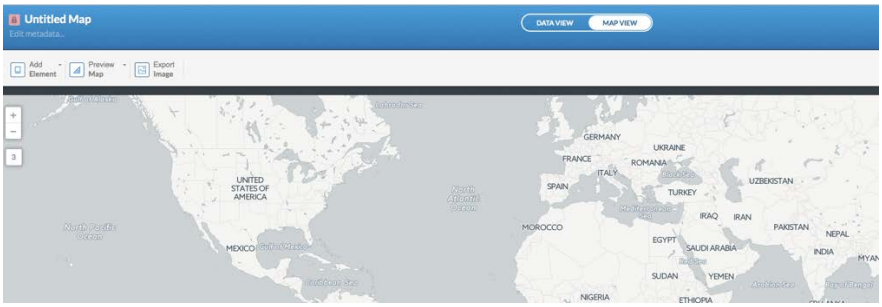
Open the CartoDB website (link) in your default browser. Create a free account and you will enter a page similar to this:



You can now import a csv file with the locations to be geocoded into CartoDB. Either drag and drop the spreadsheet onto this page or select **New Map** and **Connect Your Dataset**.

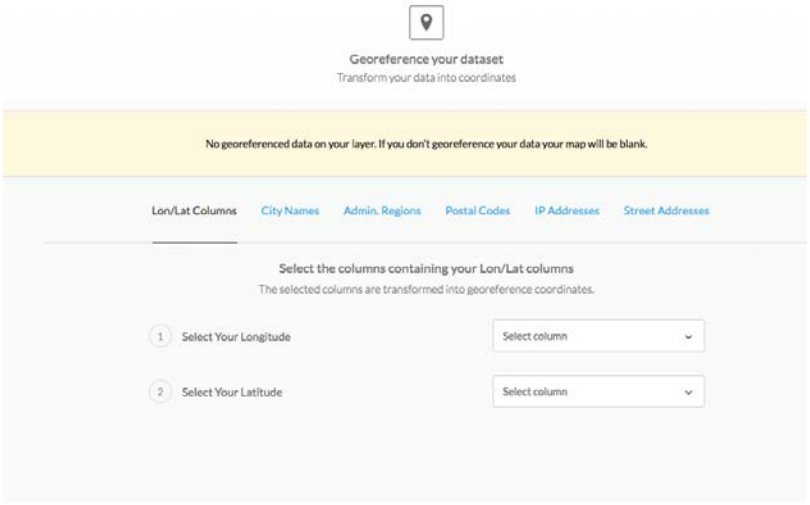
****Note**** If you already have a CartoDB account with existing datasets/maps, then select **New Map** at the top right of the screen and follow the instructions above.

Either way, once your spreadsheet is uploaded into CartoDB you will be shown blank world map in the **Map View** and your spreadsheet in the **Data View**.

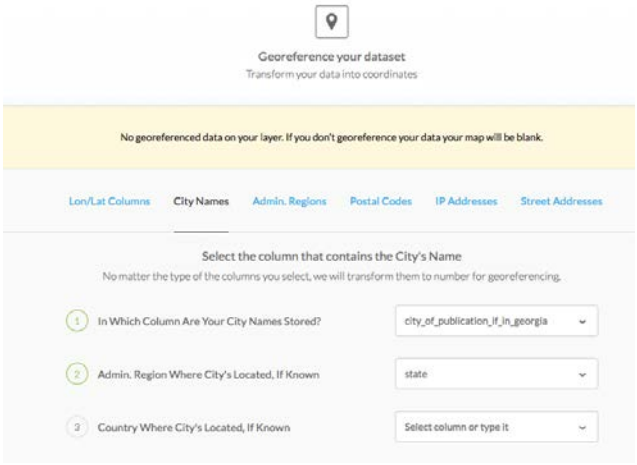


The map is blank because your spreadsheet/data is not georeferenced (creating latitude/longitude points based on locations, in this case, city names).

To Georeference your data, go to the **Edit** button at the top right of the page and select **Georeference Layer**. Once there you will enter a new screen. This is where you will specify which columns to link to geographic coordinates.

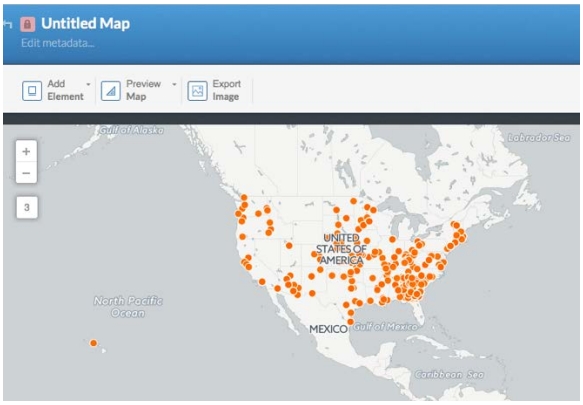


Go to the **City Names** tab and then select the column where the city names are located, enter the state name in the **Admin Region** field and Country name in the **Country** field. The more fields you can fill out, the more accurate your georeferencing will be.



Hit the **Continue** button at the bottom of the screen and select the **Georeference your data with points** option. CartoDB will then begin quickly georeferencing your spreadsheet.

Shortly after, data points (the place names) will now be layered over a world map.



Go back to the **Data View** at the top of the screen to view the actual dataset underlying the map. Your spreadsheet will now have values (lon/lat data) in the **the_geom** column (2nd column)

****Note**** When you first ingested your spreadsheet, the geom column existed, but had null values.

Unfortunately, you cannot simply export the longitude/latitude data via a spreadsheet. What you need to do is export the data in a shapefile. You can do this by hitting **Edit** in the top right of the screen and selecting **Export Layer**. Select SHP (shapefile) and download the zipped file. Unzip the file and you will have a folder with 5 files within. Make sure you keep the five files together in that folder. With that, we turn to the other software, QGIS.

Alternatively, if you would prefer not to use QGIS, you can export the data as a KML file and convert the KML to an Excel spreadsheet. This process is described in the "KML to Excel" TIPS file.

Geocoding in QGIS

QGIS is a free, open-source GIS software that is an alternative to ArcGIS. It is useable on Windows, Macs and Linux. Here is the download link/instructions on how to download QGIS (link).

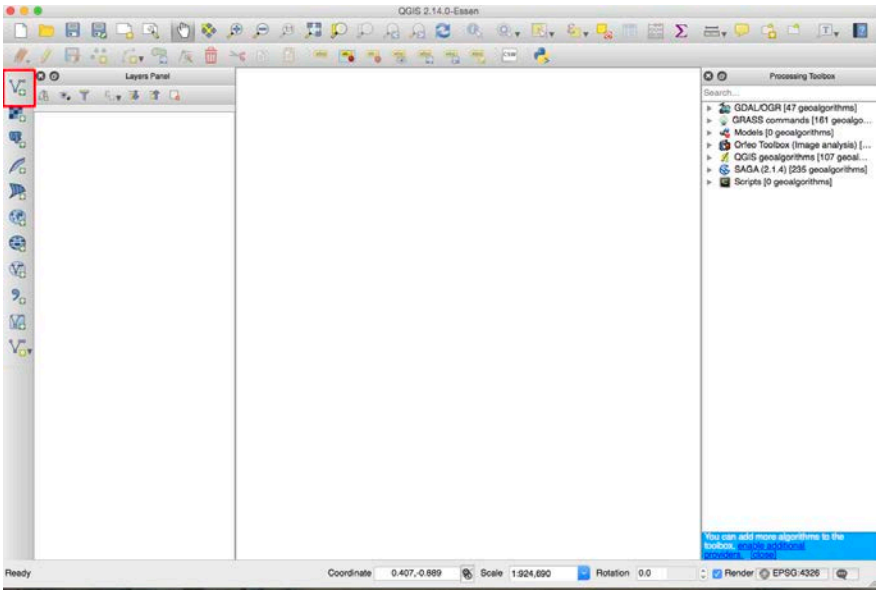
You can geocode data in QGIS but you will need to download and install the **plugin "mmqgis"**, which has batch geocoding capabilities.

You have the option of using either Google Maps or OpenStreetMap as the web service to locate an addresses, or cities, or states.

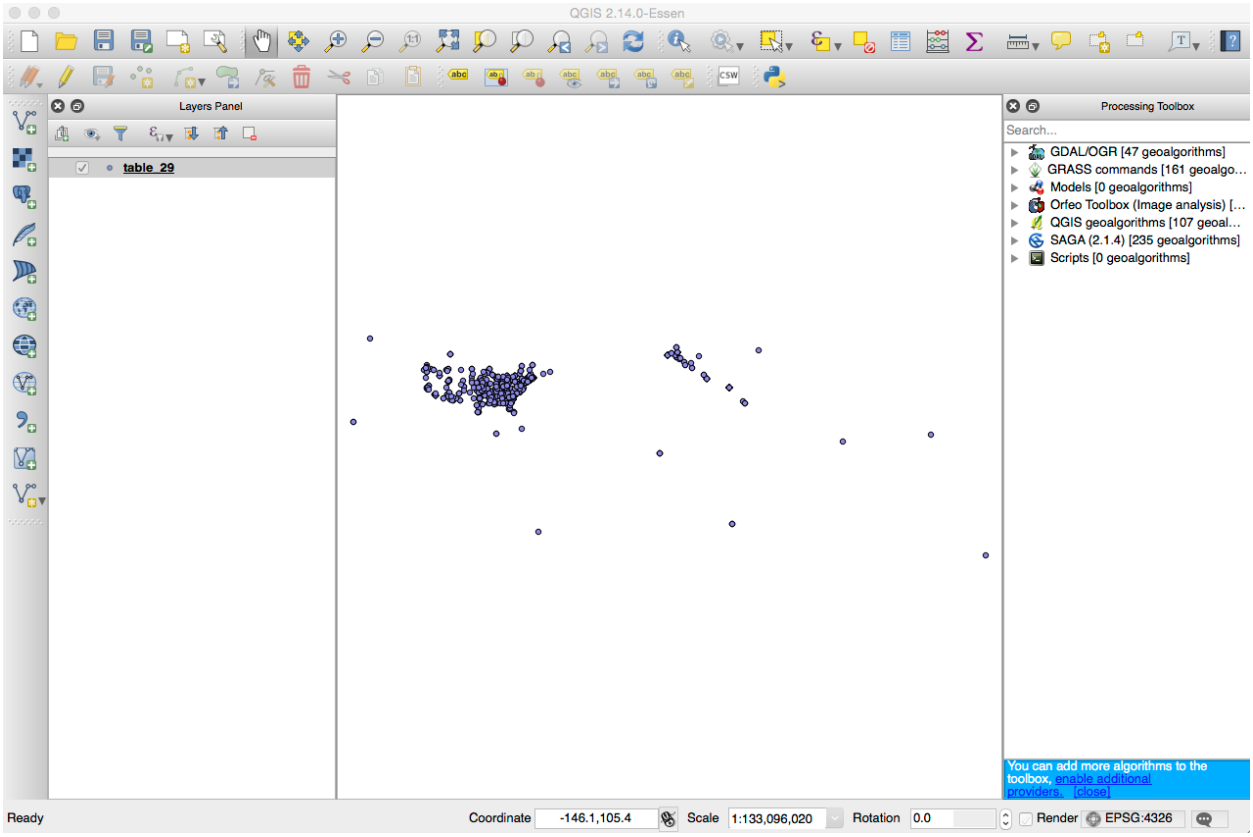
Warning!!! Geocoding is VERY slow. Be patient!!! Most importantly, don't give up thinking that QGIS has frozen up.

QGIS – Exporting the Longitute/Latitude file from the Shapefile

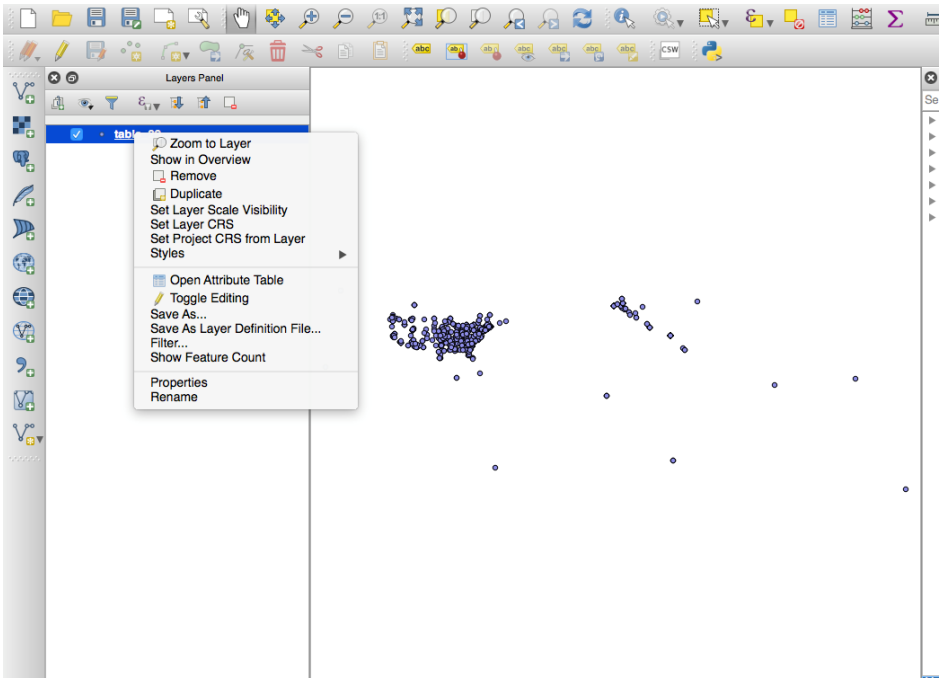
Once you have QGIS installed, open it. You should enter a blank home screen. Select **Add Vector Layer** from the left panel (highlighted in red)



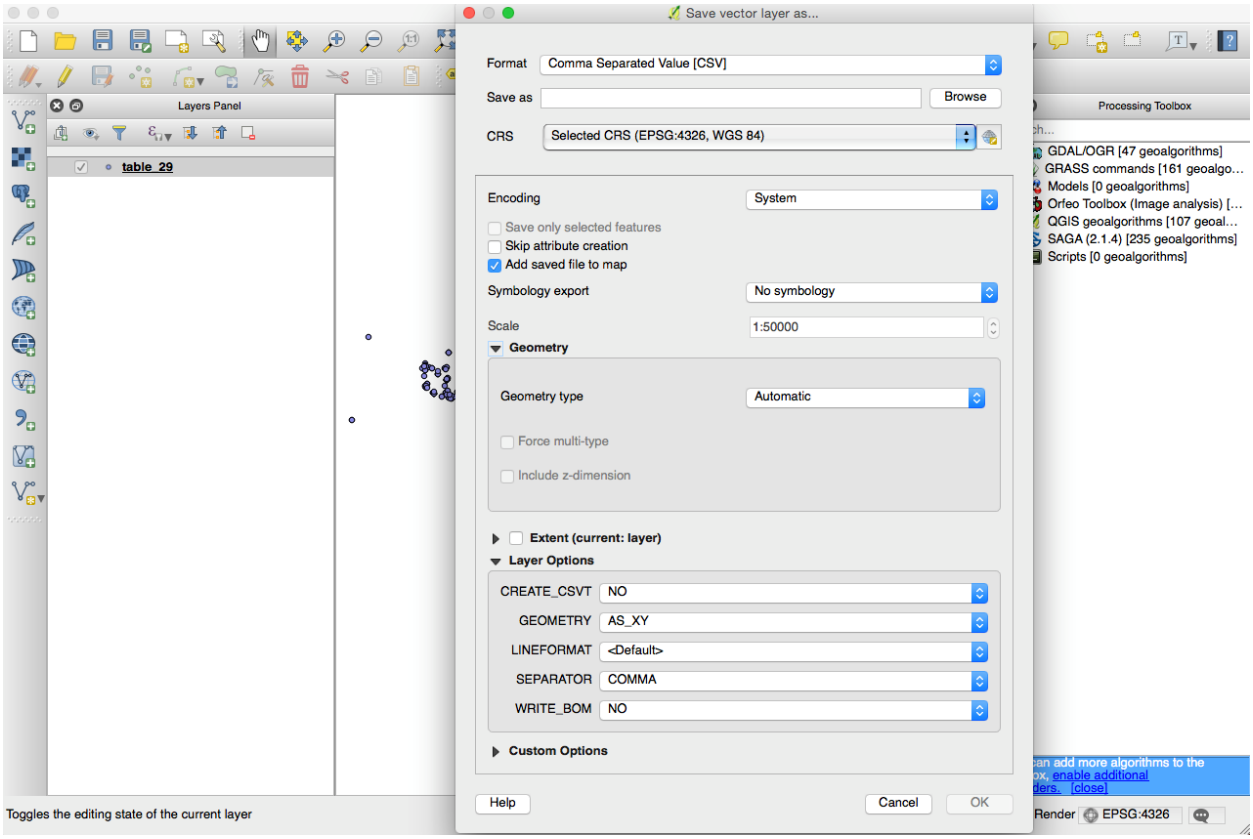
Select the shape file (file ending in .shp) you should get a screen like this.



What you want to do is right click the shape file on the left-side panel (in the screen above the filename is Table_29) and select **Save As**



You can then select where you want to save the file on your computer. Make sure the settings are the same as in the screen below. It is important to save it as a CSV file.



When you open the newly created CSV file, you should see your latitude in the X Column (first column) and the longitude in the Y Column (2nd column).

	A	B	C	D	E	F
1	X	Y	cartodb_id	newspaper_	city_of_pu	county
2	-81.97484	33.47097	1068	The Augusta Chronicle	Augusta	Augusta-Rich
3	-81.97484	33.47097	1067	Augusta Chronicle	Augusta	Richmond
4	-116.20345	43.6135	1096	Idaho Daily Statesman	Boise	Ada
5	-80.84313	35.22709	1133	The Charlotte News and Evening Chronicle	Charlotte	Mecklenburg
6	-83.72017	33.99261	61	Barrow Times	Winder	Barrow
7	-80.84313	35.22709	1134	The Charlotte Observer	Charlotte	Mecklenburg
8	-104.82025	41.13998	1135	Wyoming State Tribune	Cheyenne	Laramie
9	-87.65005	41.85003	1136	Chicago Daily Tribune	Chicago	Cook
10	-116.20345	43.6135	1097	Idaho Statesman	Boise	Ada
11	-87.65005	41.85003	1140	Daily Inter Ocean	Chicago	Cook
12	-87.65005	41.85003	1139	Daily Herald	Chicago	Cook
13	-76.61219	39.29038	1070	The Afro-American	Baltimore	Baltimore
14	-76.61219	39.29038	1069	Afro American	Baltimore	Baltimore
15	-87.65005	41.85003	1141	The Day book	Chicago	Cook
16	-77.03637	38.89511	1534	The Evening Critic	Washington DC	
17	-81.69541	41.4995	1153	Cleveland Call and Post	Cleveland	Cuyahoga
18	-84.45689	39.162	1152	The Cincinnati Commercial	Cincinnati	Hamilton
19	-83.35461	32.38683	1161	Atorning Oregonian	Cochran	Bleckley
20	-104.82136	38.83388	1163	Colorado Springs Gazette	Colorado Springs	El Paso
21	-82.99879	39.96118	1167	Columbus Daily Enquirer-Sun	Columbus	Muscogee
22	-82.99879	39.96118	1169	Columbus Enquirer Daily	Columbus	Muscogee
23	-88.88531	30.39603	1085	The Biloxi Herald	Biloxi	Harrison
24	-77.03637	38.89511	1536	The National Tribune	Washington DC	
25	-82.99879	39.96118	1173	The Colombus Enquirer	Columbus	Muscogee
26	-109.92841	31.44815	1091	The Bisbee Daily Review	Bisbee	Cochise
27	-82.99879	39.96118	1176	The Columbus Enquirer	Columbus	Franklin
28	-84.4102	33.66011	460	The Statesman	Hapeville	Fulton
29	-77.38277	44.16682	1076	Belleville News Democrat	Belleville	St. Clair
30	-78.88947	38.06847	127	The True Citizen	Waynesboro	Burke
31	-76.61219	39.29038	1072	The Baltimore AmericanA316	Baltimore	Baltimore
32	0	0	1079	The bemidji daily pioneer	Bemdiji	Beltrami
33	-77.03637	38.89511	1533	The Daily Critic	Washington DC	
34	0	0	1074	Mountain advocate	Barbourvilee	Knox
35	-84.38798	33.749	1064	Atlanta Daily World	Atlanta	Fulton
36	0	0	32	Tallapoosa Journal Beacon	Tallapoosa	Haralson
37	0	0	33	Tallapoosa Journal	Tallapoosa	Haralson
38	-84.38798	33.749	1066	The Fulton County News	Atlanta	Fulton
39	-84.38798	33.749	1065	The Atlanta Constitution	Atlanta	Fulton