Using the US Census Bureau Geocoder

Jack Walker, GIS & Data Associate, Brown University Library

May 3, 2023

`https://libguides.brown.edu/gis_data_tutorials/census`

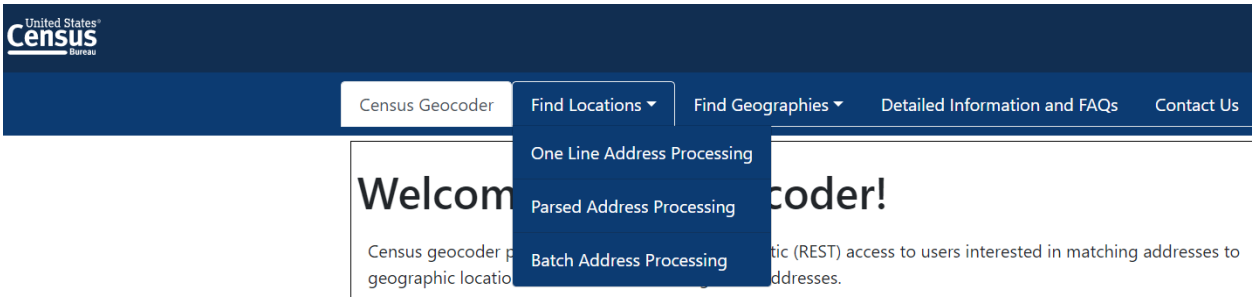
Introduction

Publicly accessible through the US Census Bureau, the Census Geocoder can provide longitude and latitude coordinates for up to 10,000 U.S. addresses at a time. The Geocoder matches addresses to street segments in the TIGER database that are assigned addresses ranges. It estimates where an address falls along that street, and returns coordinates just to the left or right side of the street in the NAD 83 coordinate system. This tutorial demonstrates how to use several Geocoder features.

1 The Geocoder Website

To begin, let's examine some of the Geocoder's key tools. Access the Geocoder at `https://geocoding.geo.census.gov/`.

Find Locations

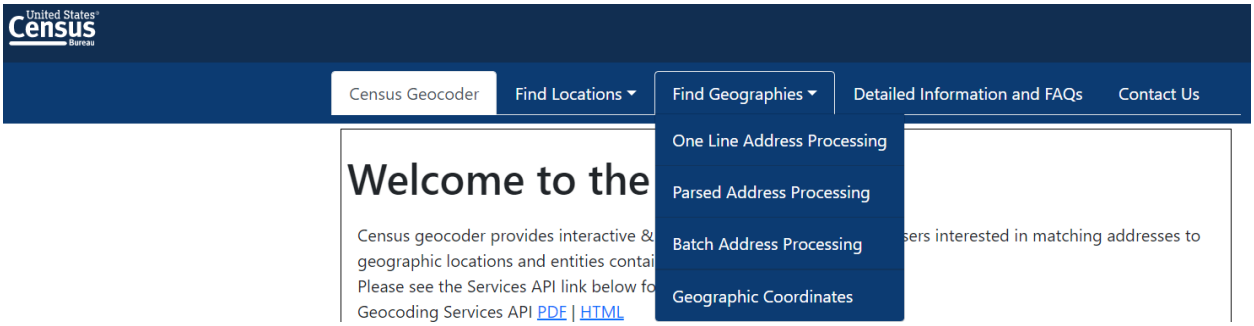


When users submit an address into *Find Locations*, the tool returns its corresponding longitude and latitude coordinates. You can access *Find Locations* in the navigation banner.

1. **Explore Find Locations:** In the navigation banner, hover your cursor over *Find Locations*. A drop-down will appear with the following options:
    - *One Line Address Processing*, which reads a single address in a single line of text.
    - *Parsed Address Processing*, which reads a single address through separate fields for house number and street name; city; state; and ZIP code.
    - *Batch Address Processing*, which reads up to 10,000 addresses in a single CSV file.

We'll explore these tools in greater depth throughout the tutorial.

Find Geographies



1. **Explore Find Geographies:** In the navigation banner, hover your cursor over *Find Geographies*. Another drop-down appears, but with a new option: *Geographic Coordinates*. This tool reads latitude and longitude coordinates and returns corresponding spatial information.

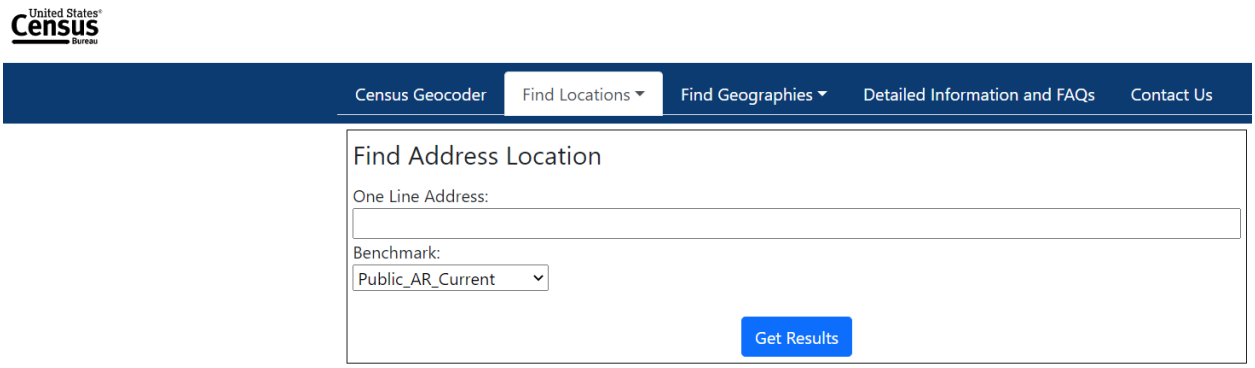
Find Locations and *Find Geographies* require the same input information. They only differ in that *Find Geographies* provides additional spatial information for each address: its state, county, census tract, and legislative district. For the remainder of this tutorial, we'll focus on *Find Locations*.

2 Single Address Processing

Let's return to *Find Locations* and examine the two tools that process a single address at a time: *One Line Address Processing* and *Parsed Address Processing*.

One Line Address Processing

1. **Find One Line Address Processing:** In the navigation banner, hover your cursor over *Find Locations* and select *One Line Address Processing*. This opens a webpage with two fields (labeled *One Line Address* and *Benchmark*) and a *Get Results* button.



2. **Examine fields:** Users can submit a single address into this tool in the following order:

    - House number & street name

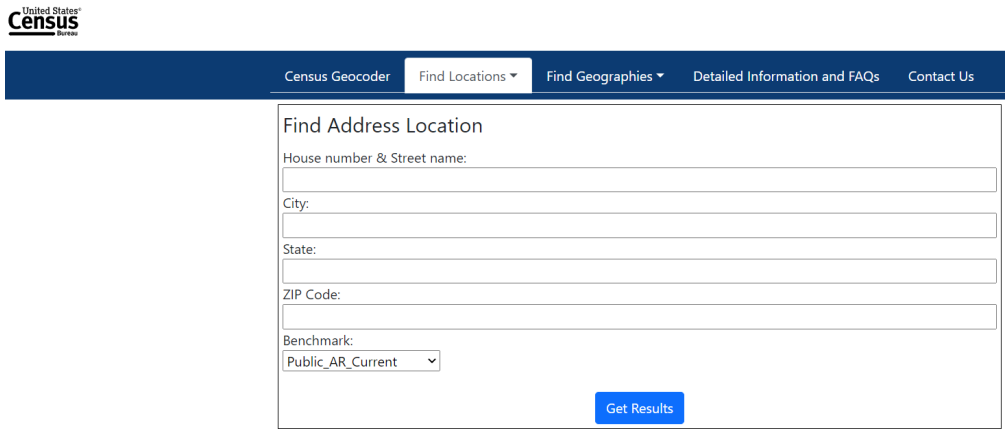
- City
- State abbreviation
- ZIP code

This information cannot be separated by commas or other punctuation. We must follow these formatting conventions each time we use the tool to ensure we receive the correct coordinates.

One Line Address Processing reads addresses formatted in just one line of text, but otherwise operates identically to another tool: *Parsed Address Processing*, which uses multiple input fields.

Parsed Address Processing

1. **Find Parsed Address Processing:** Hover your cursor over *Find Locations* in the navigation banner and select *Parsed Address Processing*. This opens a webpage with five fields (*House number & Street name*, *City*, *State*, *ZIP code*, and *Benchmark*) and a *Get Results* button.

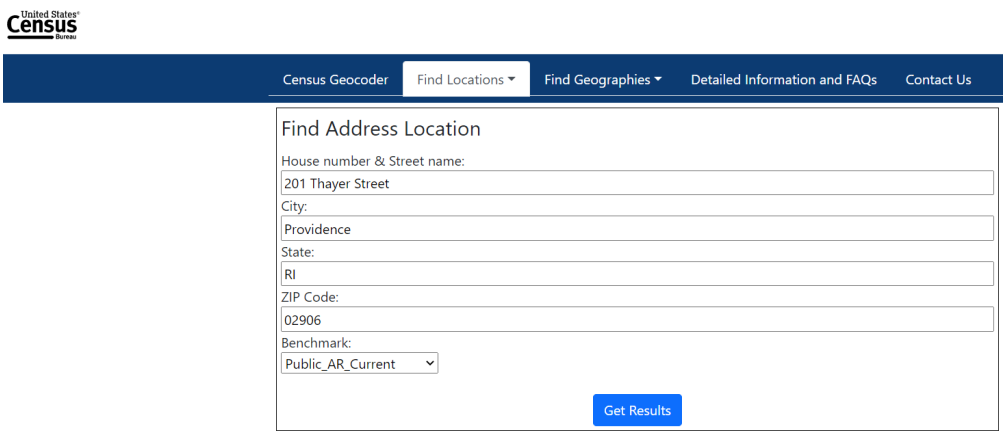


2. **Examine fields:** With *Parsed Address Processing*, the address components are separated into distinct fields.

3. **Submit an address:** Let's geocode the address for the Brown University Sciences Library. Formatted correctly, the library's address looks like this:

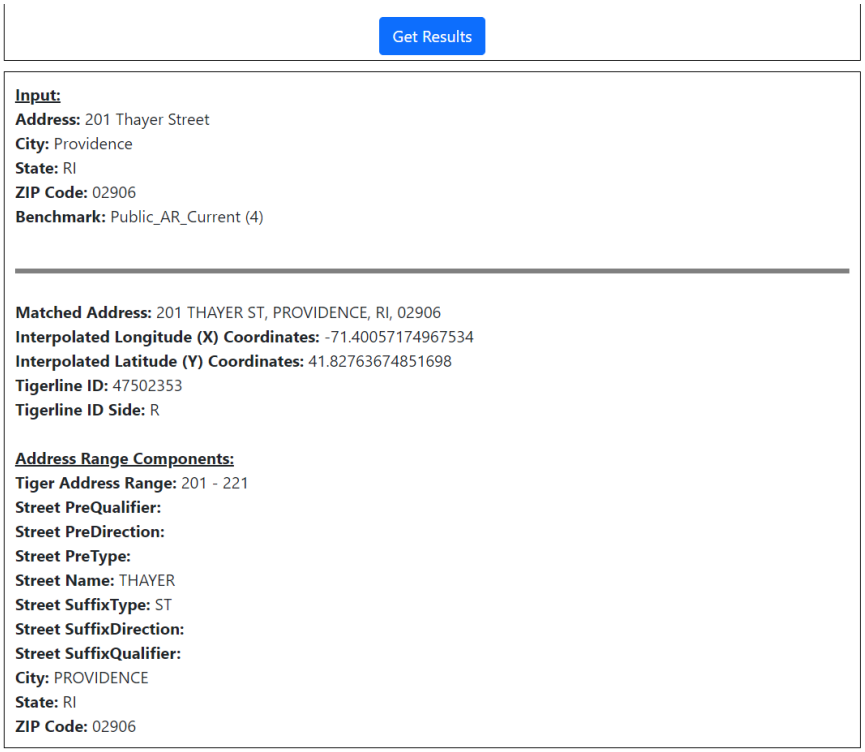
   201 Thayer St Providence RI 02912

   Enter each portion of this address into its corresponding field.



4. **Examine the Benchmark field:** *Benchmark* refers to the dataset we want the Geocoder to use. We'll keep our default, Public_AR_Current; the Census Bureau's most up-to-date street network database. Only change this field when working with data from previous years.

5. **Get results:** Click *Get Results*. The tool returns this:



6. **Analyze your results:** Examine what the Geocoder returned. *Matched Address* tells us the address from the Geocoder database that corresponds with the address we provided. If *Matched Address* differs from the actual address desired, run the tool again and ensure you follow the proper formatting conventions.

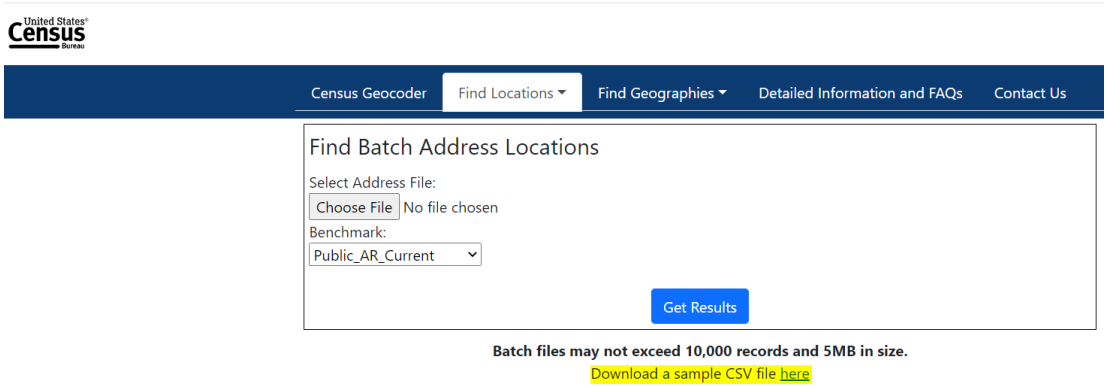
   The tool's output provides us with an *Interpolated Longitude (X) Coordinate* and an *Interpolated Latitude (Y) Coordinate*. This information tells us the longitude and latitude of the address we provided: -71.4006 and 41.8276, respectively.

You've now successfully used the Geocoder to find the longitude and latitude coordinates for an address! Let's try out a new tool that can read multiple addresses at once: *Batch Address Processing*.

3 Batch Address Processing

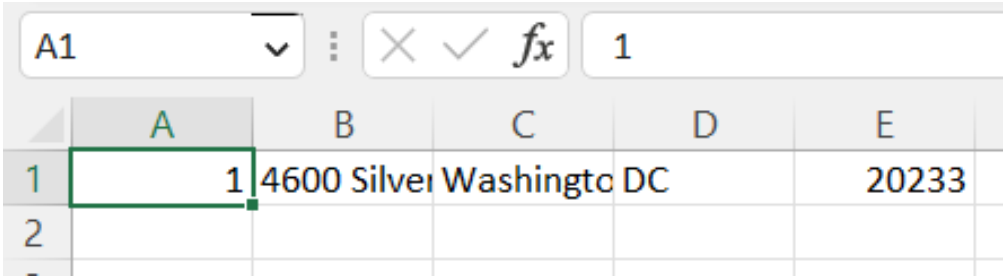
Basic Example

1. **Find Batch Address Processing:** In the navigation banner, hover your cursor over *Find Locations* and select *Batch Address Processing*. This opens a webpage with two fields (*Select Address File* and *Benchmark*) and a *Get Results* button.



2. **Download sample data:** Beneath *Get Results*, find the line of text that says "Download a sample CSV file here." It's highlighted in the image above. Click the hyperlink and download.

3. **Examine data:** Open the sample data on your desktop. It looks like this:



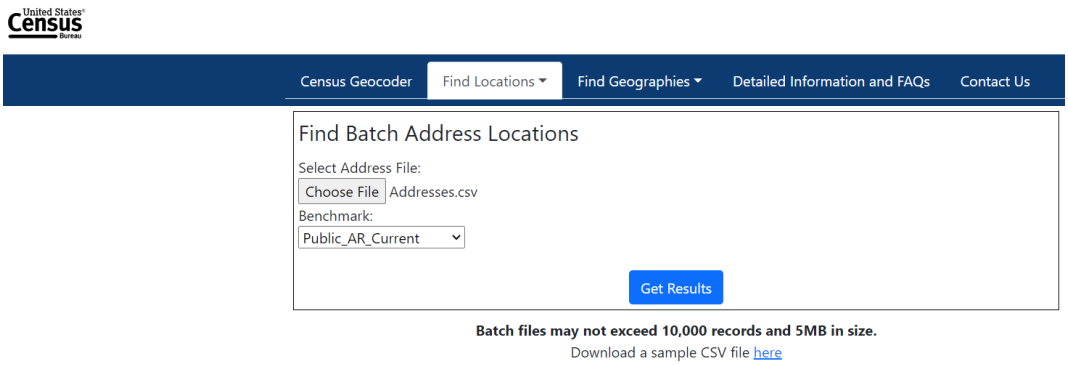
Like the other tools, *Batch Address Processing* is strict about address formatting. This sample CSV file demonstrates how all CSV files submitted to the tool must be formatted:

- *Unique ID number*
- *House number & street name*
- *City*
- *State abbreviation*
- *ZIP code*

Each unique address is contained in a unique row, and each part of an address gets a separate column. Every address has five columns in this exact order. Include no additional columns (or column names / headers), or the Geocoder will reject your file.

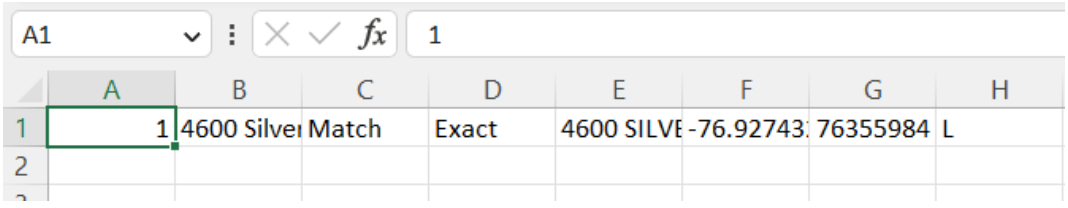
4. **Examine address IDs:** Most of the address information used in this tool is identical to the information we used for previous tools in this exercise. The exception, however, is that this tool requires each address entry to have a unique ID number. Our current dataset comes with a preexisting unique ID number in `Column A`. However, users might have to create a unique ID of their own if their dataset lacks such an identifier. The specific number of this ID does not matter, and simply numbers each entry for the tool to read. Nonetheless, we must ensure our ID numbers follow a coherent logic so that we don't get confused once we work with larger datasets; the simplest approach to this field is using a sequential integer: 1, 2, 3, etc.

5. **Run the tool:** Now that we've taken a look at what our sample CSV file contains, let's return to the *Batch Address Processing* tool. Under the field labeled *Select Address File*, click *Choose File*. This will open your desktop files. Find the sample CSV you downloaded (`Addresses.csv`) select it, and press *Open*. We now see the `Address.csv` beside the *Open File* button, indicating our sample CSV file has been uploaded.



Leave the benchmark set to its default, Public_AR_Current, and select *Get Results*.

6. **Analyze results:** The Geocoder returns a CSV file (`GeocodeResults.csv`) downloaded automatically through your browser. Find the file in the downloads folder of your computer and open it. The CSV file looks like this:



This table looks similar to our input table, but with some differences. `Column A` contains the unique ID number we set in our input table, and `Column B` contains the input address associated with that identifier. `Column C` indicates whether our input address matched an address in its own database, and `Column D` indicates whether these two addresses are an exact

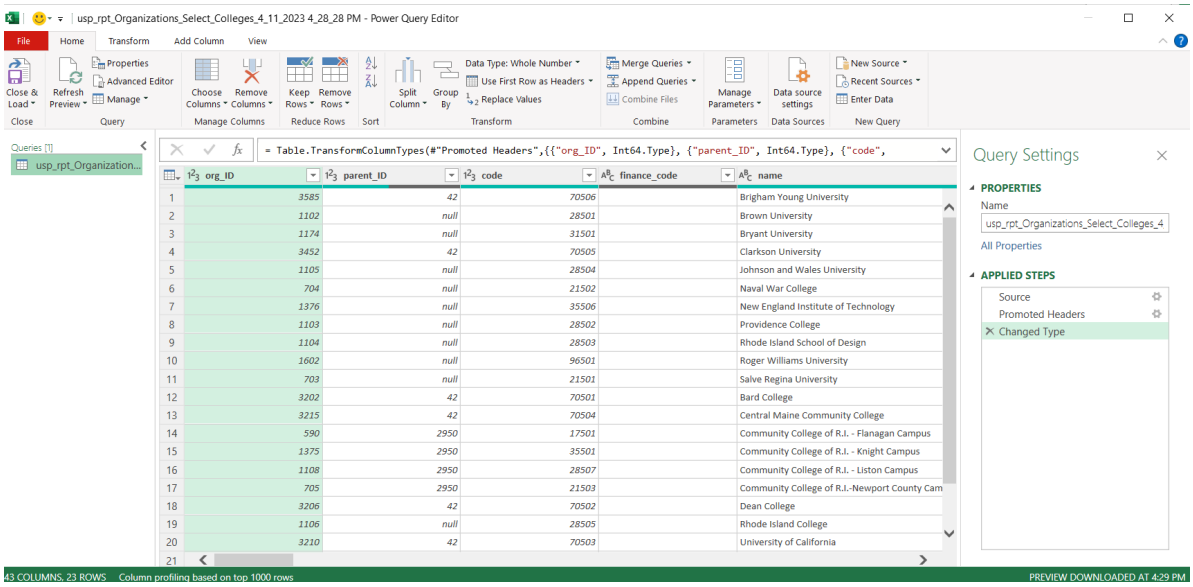
match, a non-exact match, a tie, or no match. Click on the value in `Row 1`, `Column E` and see that it is an exact match to your address in `Column B`.

`Column F` contains interpolated longitude and latitude coordinates for the matched address. `Column G` contains this address's tigerline ID, and `Column H` contains its tigerline ID side, left or right; these are unique identifiers for street segments in the Census TIGER Line database.

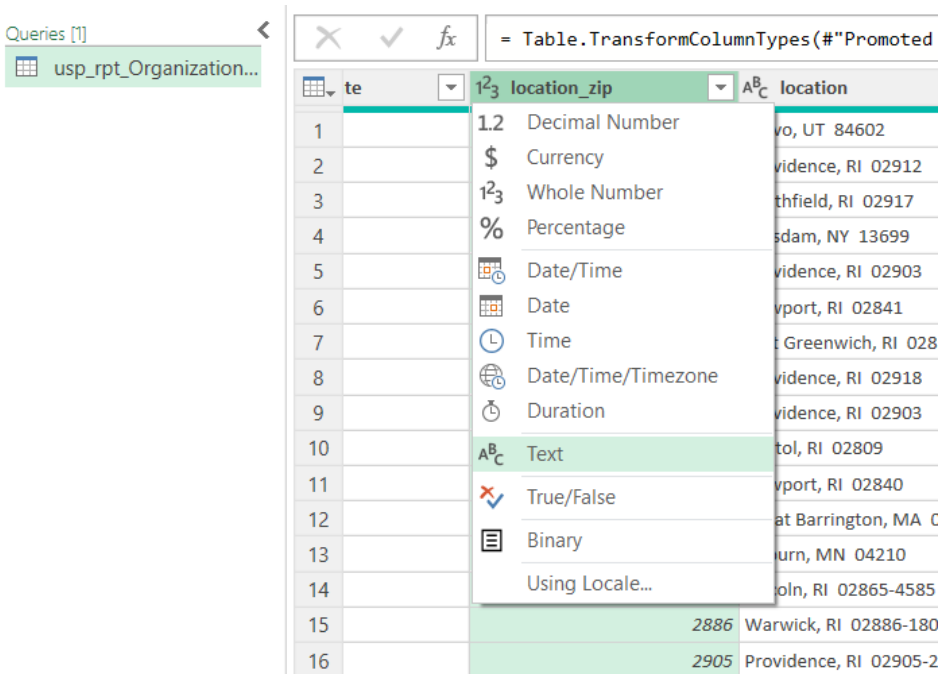
A Workflow with Sample Data

Now, let's use *Batch Address Processing* with our own dataset: all colleges and universities in the state of Rhode Island, from a directory produced by the RI Department of Education. The sample data is available for download from the link where you accessed this tutorial. We will demonstrate some basic data processing techniques to prepare data for geocoding.

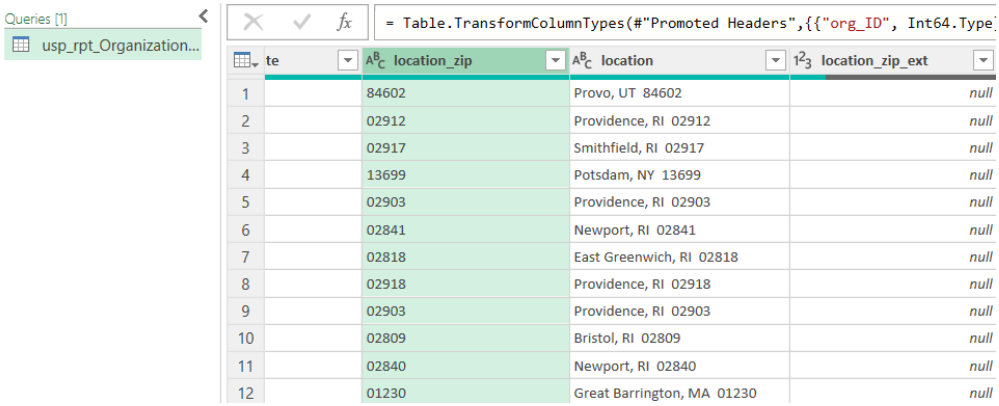
1. **Import the file into Excel:** The file you downloaded is in CSV format. Do not doubleclick the CSV to open it in Excel; we need to import the data to preserve our ZIP Codes formatting. To begin, let's import it into Excel. Open Excel from your desktop and open a blank workbook. Navigate to *Get Data - From File - From Text/CSV*, select the CSV file from your computer's directory, and select *Open*. This will show us a preview of the data in this file. Recall from previous steps that the Geocoder requires only a select few address components and a specific row order. We'll have to modify this dataset to get it to the Geocoder's convention, so select *Transform Data*. Your screen now looks like this:



2. **Correct the ZIP code data:** Scroll to the column entitled `location_zip` in your spreadsheet. The ZIP codes in this column begin with 0, but the data type for the column is designated as a *Whole Number* 123. This means the first digit of any ZIP code that begins with 0 is cut off. To fix this, we need to transform this column's data type into *Text* ABC. Find the *Whole Number* 123 symbol in the `location_zip` column, click on it, then select *Text* ABC - *Replace current*.

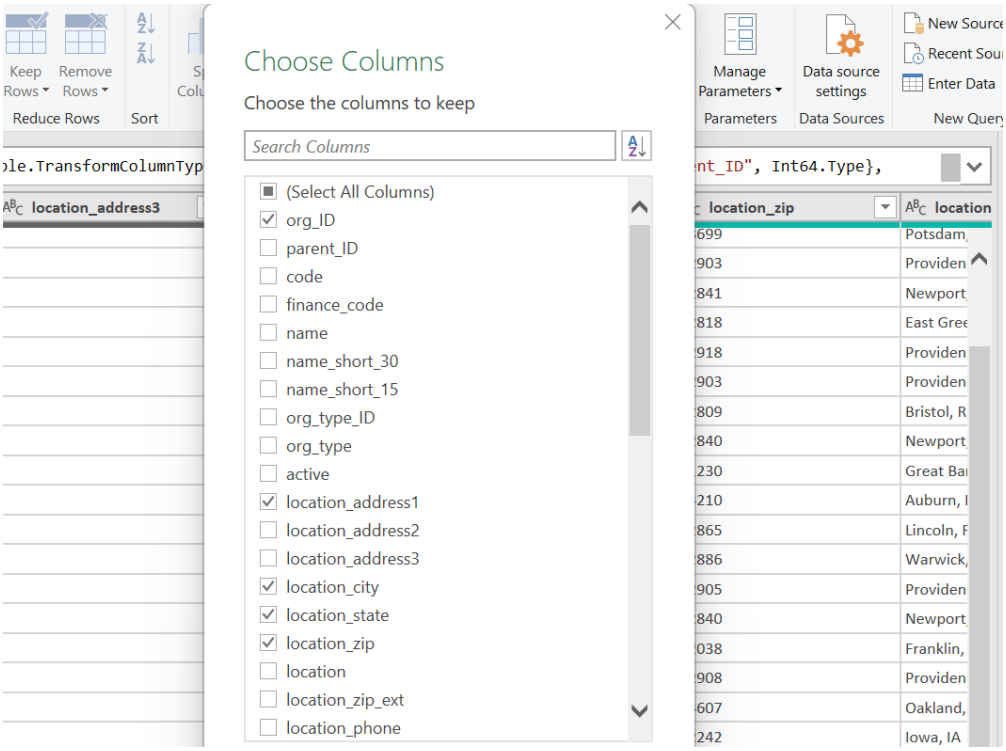


The data in this column will now be read as text, so each ZIP code now begins with a 0.



3. **Choose desired columns:** Our CSV file has 48 columns; let's get rid of those we don't need. Still in the same editor, navigate to *Home* - *Choose Columns* (which lists every file column's data field) and uncheck it. Scrolling through the list, find and check the following fields:

    - org_ID
    - location_address1
    - location_city
    - location_state
    - location_zip

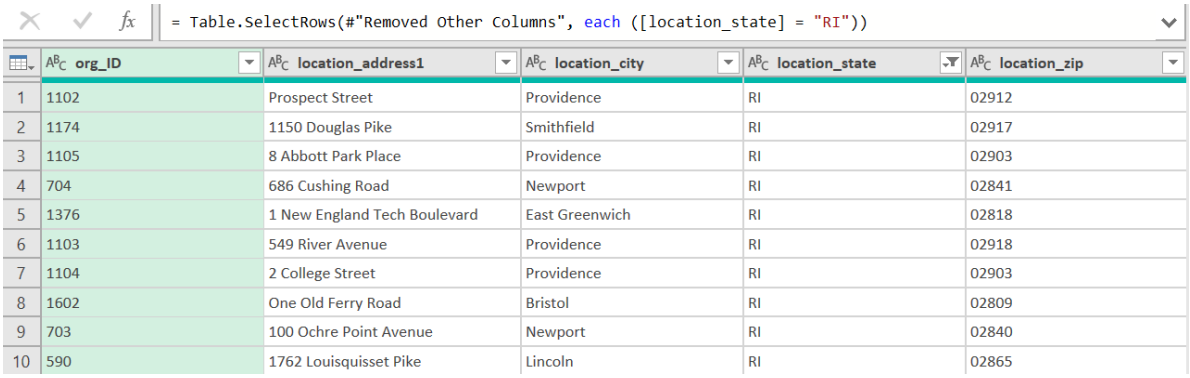


Press *OK*. Your table now looks like this:

```
= Table.SelectColumns(#"Changed Type",{"org_ID", "location_address1", "location_city", "location_state",
```

	org_ID	location_address1	location_city	location_state	location_zip
1	3585	Brigham Young University	Provo	UT	84602
2	1102	Prospect Street	Providence	RI	02912
3	1174	1150 Douglas Pike	Smithfield	RI	02917
4	3452	8 Clarkson Ave	Potsdam	NY	13699
5	1105	8 Abbott Park Place	Providence	RI	02903
6	704	686 Cushing Road	Newport	RI	02841
7	1376	1 New England Tech Boulevard	East Greenwich	RI	02818
8	1103	549 River Avenue	Providence	RI	02918
9	1104	2 College Street	Providence	RI	02903
10	1602	One Old Ferry Road	Bristol	RI	02809
11	703	100 Ochre Point Avenue	Newport	RI	02840
12	3202	84 Alford Road	Great Barrington	MA	01230

4. **Remove out-of-state entries:** Our file includes all Rhode Island colleges and universities, plus out-of-state schools with educational opportunities in Rhode Island. Let's focus only on colleges and universities based in state. Still in the editor, find the `location_state` column and select the arrow in the column header. We'll now find and remove schools that aren't located in state. Uncheck *Select All*, select *RI* and then press *OK*. Your file looks like this:



	org_ID	location_address1	location_city	location_state	location_zip
1	1102	Prospect Street	Providence	RI	02912
2	1174	1150 Douglas Pike	Smithfield	RI	02917
3	1105	8 Abbott Park Place	Providence	RI	02903
4	704	686 Cushing Road	Newport	RI	02841
5	1376	1 New England Tech Boulevard	East Greenwich	RI	02818
6	1103	549 River Avenue	Providence	RI	02918
7	1104	2 College Street	Providence	RI	02903
8	1602	One Old Ferry Road	Bristol	RI	02809
9	703	100 Ochre Point Avenue	Newport	RI	02840
10	590	1762 Louisquisset Pike	Lincoln	RI	02865

5. **Accept changes and remove column headers:** Our data is now ready to be inserted into the table! In the editor, select *Home - Close & Load*. Your spreadsheet looks like this:

	org_ID	location_address1	location_city	location_state	location_zip
2	1102	Prospect Street	Providence	RI	02912
3	1174	1150 Douglas Pike	Smithfield	RI	02917
4	1105	8 Abbott Park Place	Providence	RI	02903
5	704	686 Cushing Road	Newport	RI	02841
6	1376	1 New England Tech Boulevard	East Greenwich	RI	02818
7	1103	549 River Avenue	Providence	RI	02918
8	1104	2 College Street	Providence	RI	02903
9	1602	One Old Ferry Road	Bristol	RI	02809
10	703	100 Ochre Point Avenue	Newport	RI	02840

Let's clean up this spreadsheet so it's ready for the Geocoder. As you know, we cannot include column headers in data for the Geocoder. We want to remove our column headers and table formatting entirely. To do this, open a new sheet in the same Excel book. In our original sheet, copy all text except for the column headers; or, in other words, highlight `Columns A` through `E` and `Rows 2` through `17`. In the `A1` cell of our new sheet, right-click and select *Paste Values (V)*. Delete our original sheet. Your new sheet looks like this:

	A	B	C	D	E
1	1102	Prospect S	Providenc	RI	02912
2	1174	1150 Doug	Smithfield	RI	02917
3	1105	8 Abbott F	Providenc	RI	02903
4	704	686 Cushi	Newport	RI	02841
5	1376	1 New Eng	East Greer	RI	02818
6	1103	549 River	Providenc	RI	02918
7	1104	2 College S	Providenc	RI	02903
8	1602	One Old F	Bristol	RI	02809
9	703	100 Ochre	Newport	RI	02840
10	590	1762 Louis	Lincoln	RI	02865

6. **Clean addresses:** Finally, we need to change some addresses to ensure each contains a numerical house number for the Geocoder to read. Some of our addresses have a spelled-out house number, and our first entry (Brown University) and fifteenth entry (University of Rhode Island) lack a house number altogether. Change the text in these cells as follows:

- `B1`: Prospect Street -> 1 Prospect Street
- `B8`: One Old Ferry Road -> 1 Old Ferry Road
- `B12`: One Hilton Street -> 1 Hilton Street
- `B13`: One John H. Chafee Blvd. -> 1 John H. Chafee Blvd
- `B15`: Green Hall -> 35 Campus Avenue

7. **Submit your file to the Geocoder:** We can now submit our file to the Geocoder. Save your changes to the CSV file and close it. Reopen the Geocoder in your browser. Hover your cursor over *Find Locations* in the navigation banner and select *Batch Address Processing*. Under *Select Address File*, select *Choose File* and select the CSV file you modified. Leave *Benchmark* set to Public_AR_Current, and select *Get Results*.

8. **Review return file:** Once again, the Geocoder returns a CSV file that your browser downloads automatically. Once downloaded, open this new CSV file. It now looks like this:

	A	B	C	D	E
1	RECORD ID NUMBER	INPUT ADDRESS	TIGER ADDRESS RANGE MATCH INDICATOR	TIGER MATCH TYPE	TIGER OUTPUT ADDRESS
2					
3	1220	35 Campus Avenue, Kingston, RI, 02881	No_Match		
4	2385	80 Washington Street, Providence, RI, 02903	Match	Exact	80 WASHINGTON ST, PROVIDENCE, RI, 02903
5	1174	1150 Douglas Pike, Smithfield, RI, 02917	Match	Exact	1150 DOUGLAS PIKE, SMITHFIELD, RI, 02917
6	590	1762 Louisquisset Pike, Lincoln, RI, 02865	Match	Non_Exact	1762 OLD LOUISQUISSET PIKE, LINCOLN, RI, 02865
7	1602	1 Old Ferry Rd, Bristol, RI, 02809	Tie		
8	1106	600 Mt. Pleasant Avenue , Providence, RI, 02908	Match	Exact	600 MOUNT PLEASANT AVE, PROVIDENCE, RI, 02908
9	1105	8 Abbott Park Place, Providence, RI, 02903	No_Match		
10	1104	2 College Street, Providence, RI, 02903	Match	Exact	2 COLLEGE ST, PROVIDENCE, RI, 02903
11	1103	549 River Avenue, Providence, RI, 02918	Match	Non_Exact	549 RIVER AVE, PROVIDENCE, RI, 02908
12	1102	1 Prospect Street, Providence, RI, 02912	Match	Exact	1 PROSPECT ST, PROVIDENCE, RI, 02912
13	1376	1 New England Tech Boulevard, East Greenwich, RI, 02818	No_Match		
14	1375	400 East Avenue, Warwick, RI, 02886	Match	Exact	400 EAST AVE, WARWICK, RI, 02886
15	703	100 Ochre Point Avenue, Newport, RI, 02840	Match	Exact	100 OCHRE POINT AVE, NEWPORT, RI, 02840
16	704	686 Cushing Road, Newport, RI, 02841	No_Match		
17	705	1 John H. Chafee Blvd, Newport, RI, 02840	Match	Exact	1 JOHN H CHAFEE BLVD, NEWPORT, RI, 02840
18	1108	1 Hilton Street, Providence, RI, 02905	Match	Exact	1 HILTON ST, PROVIDENCE, RI, 02905

9. **Analyze new data:** To review, this is what each column of the new CSV file contains:

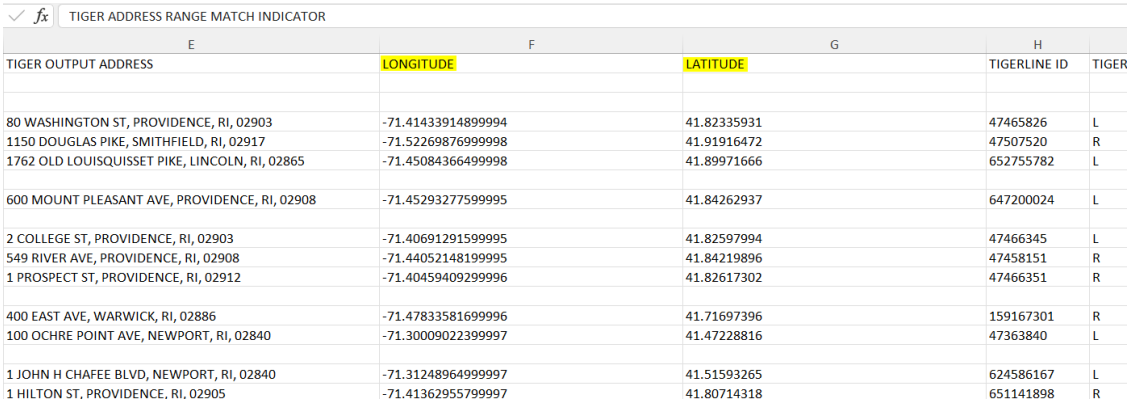
   - `Column A`: The unique ID number we provided in our input CSV file
   - `Column B`: The input address
   - `Column C`: Tme match result
   - `Column D`: Whether the two addresses are an exact or non-exact match
   - `Column E`: The matched address
   - `Column F`: The interpolated longitude and latitude coordinates of the matched address
   - `Column G`: The address's tigerline ID
   - `Column H`: The address's tigerline ID side

   Some of our data did not return coordinates; unfortunately, this is common when working with address data, and we often must tweak our datasets repeatedly in order to process them properly. We'll discuss this further in the tutorial's Conclusion.

10. **Separate longitude and latitude columns:** We have one last step to complete before we can use this data in a GIS application: separating the longitude and latitude coordinates. Currently, these coordinates are both in `Column F`, separated by a comma. First, let's add a

new column to the right of `Column F`. Select `Column G`, then *Home - Insert - Insert Sheet Columns* in the navigation banner. A new, blank column appears as `Column G`; this is where our latitude coordinates will be inserted.

Now, to separate the latitude and longitude coordinates, select *Data - Text to Columns* in the navigation banner. Select *Delimited* to indicate commas will separate our coordinates into separate fields, and press *Next*. Now, deselect *Tab* and select *Comma* as the delimiter. Select *Next*. Select *Text* as our column data format to ensure no digits get cut off, and select *Finish*. Save your spreadsheet. Your file should now look like this:



fx TIGER ADDRESS RANGE MATCH INDICATOR

E	F	G	H	
TIGER OUTPUT ADDRESS	LONGITUDE	LATITUDE	TIGERLINE ID	TIGER
80 WASHINGTON ST, PROVIDENCE, RI, 02903	-71.41433914899994	41.82335931	47465826	L
1150 DOUGLAS PIKE, SMITHFIELD, RI, 02917	-71.52269876999998	41.91916472	47507520	R
1762 OLD LOUISQUISSET PIKE, LINCOLN, RI, 02865	-71.45084366499998	41.89971666	652755782	L
600 MOUNT PLEASANT AVE, PROVIDENCE, RI, 02908	-71.45293277599995	41.84262937	647200024	L
2 COLLEGE ST, PROVIDENCE, RI, 02903	-71.40691291599995	41.82597994	47466345	L
549 RIVER AVE, PROVIDENCE, RI, 02908	-71.44052148199995	41.84219896	47458151	R
1 PROSPECT ST, PROVIDENCE, RI, 02912	-71.40459409299996	41.82617302	47466351	R
400 EAST AVE, WARWICK, RI, 02886	-71.47833581699996	41.71697396	159167301	R
100 OCHRE POINT AVE, NEWPORT, RI, 02840	-71.30009022399997	41.47228816	47363840	L
1 JOHN H CHAFEE BLVD, NEWPORT, RI, 02840	-71.31248964999997	41.51593265	624586167	L
1 HILTON ST, PROVIDENCE, RI, 02905	-71.41362955799997	41.80714318	651141898	R

You have now successfully completed *Batch Address Processing* through the Census Geocoder!

4 Conclusion

In this tutorial, we examined the different ways we can use the US Census Geocoder to retrieve coordinates for a specified address. As you can see from our *Batch Address Processing* result, datasets often require adjustment to produce matches. One of our input addresses was tied, meaning it had more than one matched addresses; in future iterations, we should try to find a standardized address for this location to rematch it and get one result. Others produced no match at all. These require a deeper reworking, including cross-referencing our data with online information from Google Maps and school websites to find correct addresses. Continue modifying your datasets with the Geocoder to attain better match results.

In the last example, we used data that consisted of just address information, as the Geocoder only accepts a unique ID and address data. If you had additional attributes that you want to associate with your results (like the name of the colleges in this example): make sure you have a unique ID first, then separate the data by copying the ID and address fields into a separate sheet, and save it as a CSV for geocoding. Once you have results, you can join them back to your original dataset using the unique ID. You can use Excel's VLOOKUP function, or other software that allows you to join tables (GIS, stats packages, or relational databases). If your dataset has more than 10,000 records, you will need to split it up into subsets, geocode each individually, and aggregate the results at the end.