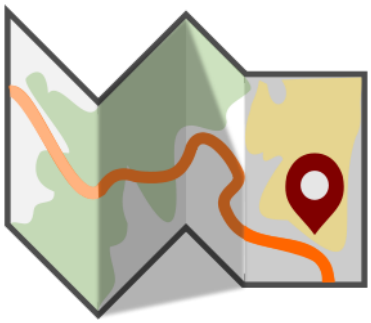
Nearest Neighbor Analysis

QGIS Tutorials and Tips



Author

Ujaval Gandhi

http://www.spatialthoughts.com

Nearest Neighbor Analysis

GIS is very useful in analyzing spatial relationship between features. One such analysis is finding out which features are closest to a given feature. QGIS has a tool called **Distance Matrix** which helps with such analysis. In this tutorial, we will use 2 datasets and find out which points from one layer are closest to which point from the second layer.

Overview of the task

Given the locations of all known significant earthquakes, find out the nearest populated place for each location where the earthquake happened.

Other skills you will learn

- How to do table joins in QGIS. (See *Performing Table Joins* for detailed instructions.)
- Using Query Builder to show a subset of features from a layer.
- Using MMQGIS plugin to create hub lines to visualize the nearest neighbors.

Get the data

We will use NOAA's National Geophysical Data Center's Significant Earthquake Database as our layer representing all major earthquakes. Download the tab-delimited earthquake data.

Natural Earth has a nice Populated Places dataset. Download the simple (less columns) dataset

For convenience, you may directly download a copy of both the datasets from the links below:

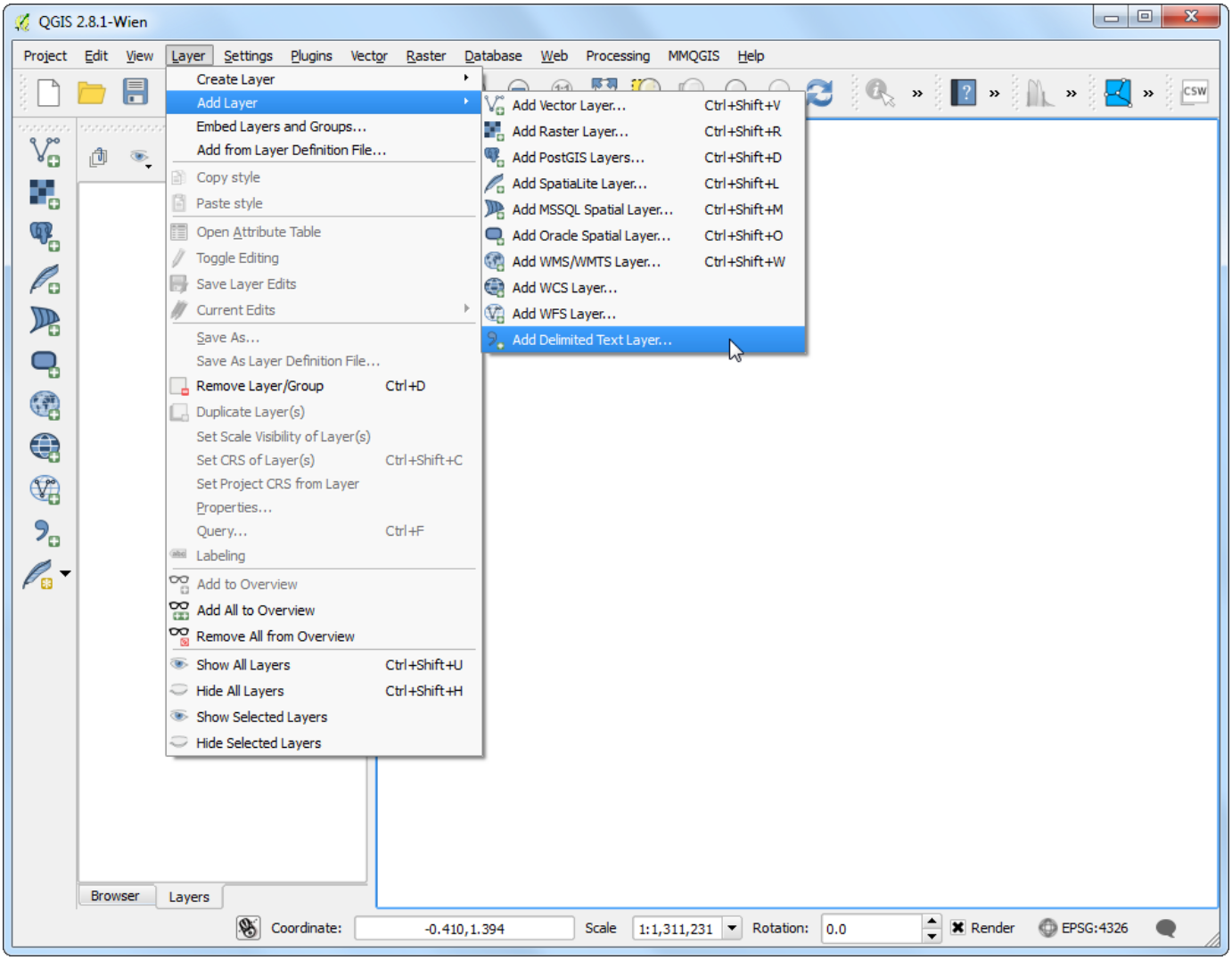
signif.txt

ne_10m_populated_places_simple.zip

Data Sources: [NGDC] [NATURALEARTH]

Procedure

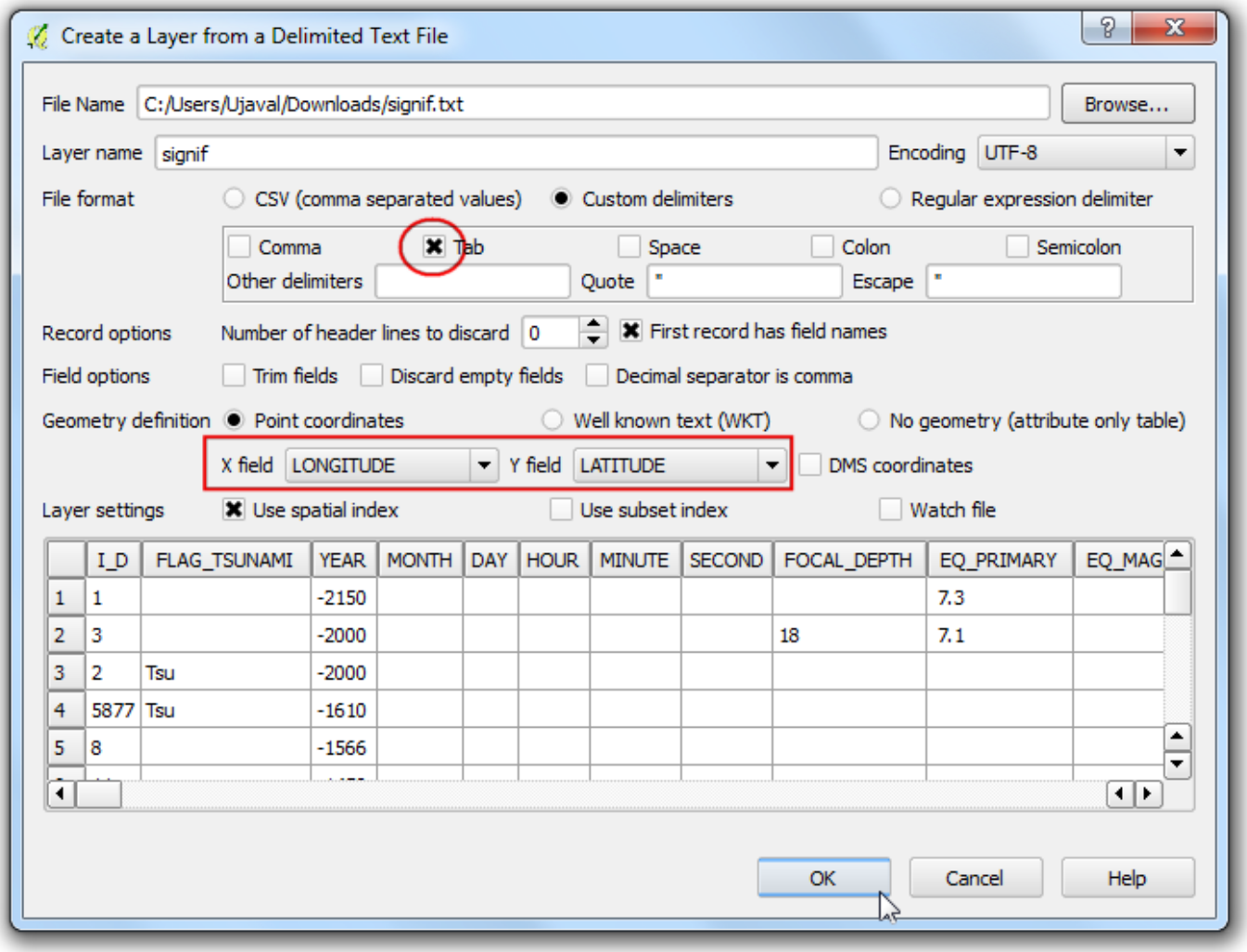
1. Open Layer ▸ Add Layer ▸ Add Delimited Text Layer and browse to the downloaded `signif.txt` file.



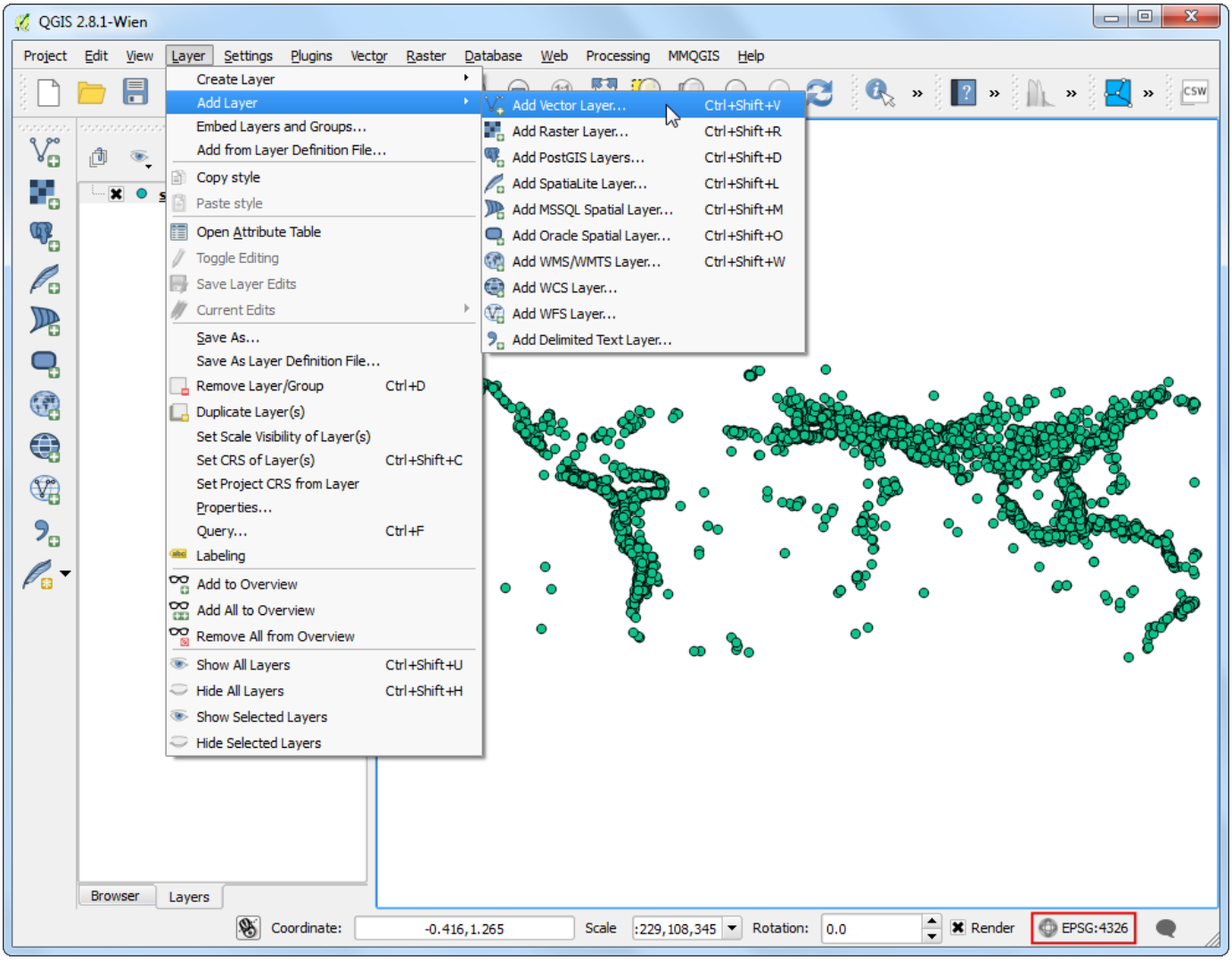
2. Since this is a *tab-delimited file*, choose Tab as the File format. The X field and Y field would be auto-populated. Click OK.

Note

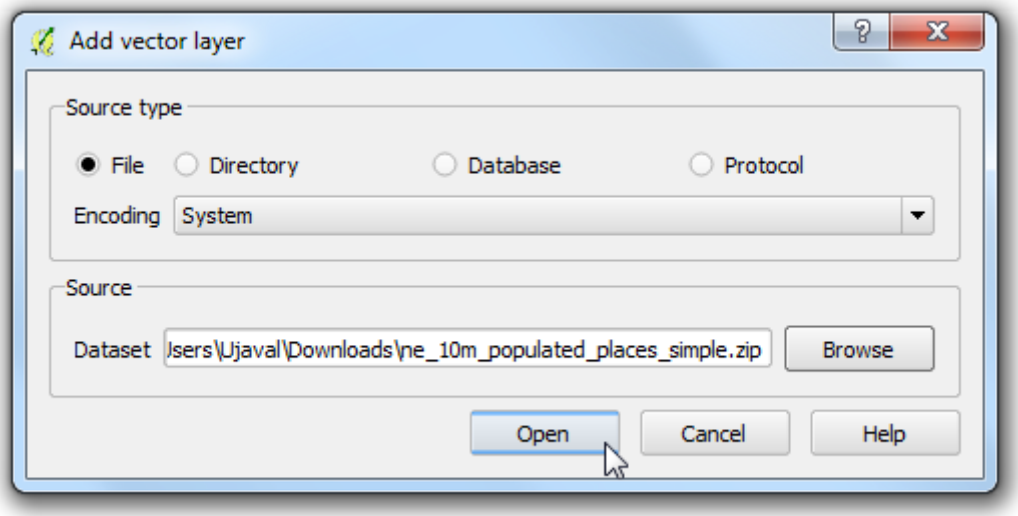
You may see some error messages as QGIS tries to import the file. These are valid errors and some rows from the file will not be imported. You can ignore the errors for the purpose of this tutorial.



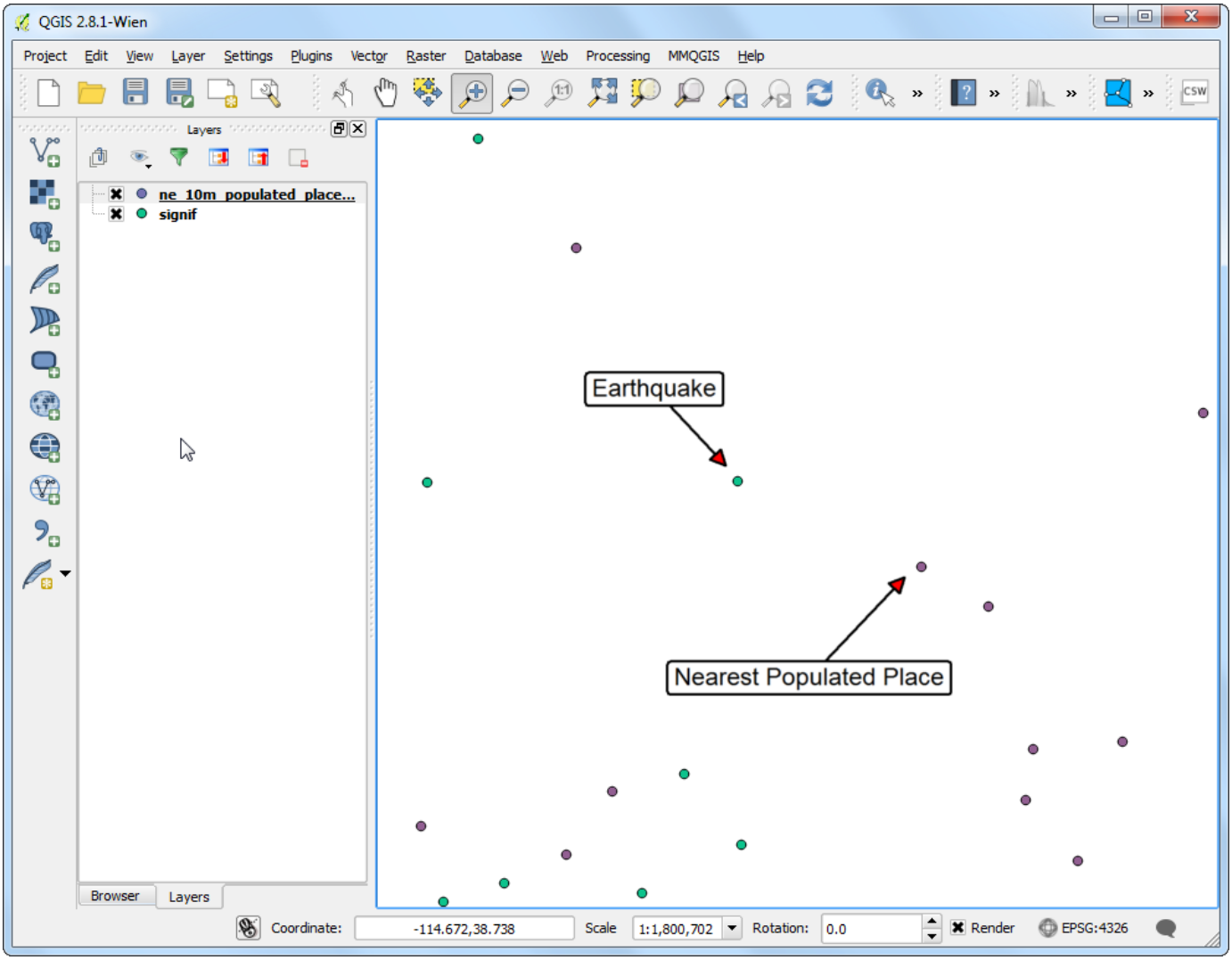
3. As the earthquake dataset has Latitude/Longitude coordinates, it will be imported with the default CRS of `EPSG: 4326`. Verify that is the case in the bottom-right corner. Let's also open the Populated Places layer. Go to Layer ▸ Add Layer ▸ Add Vector Layer.



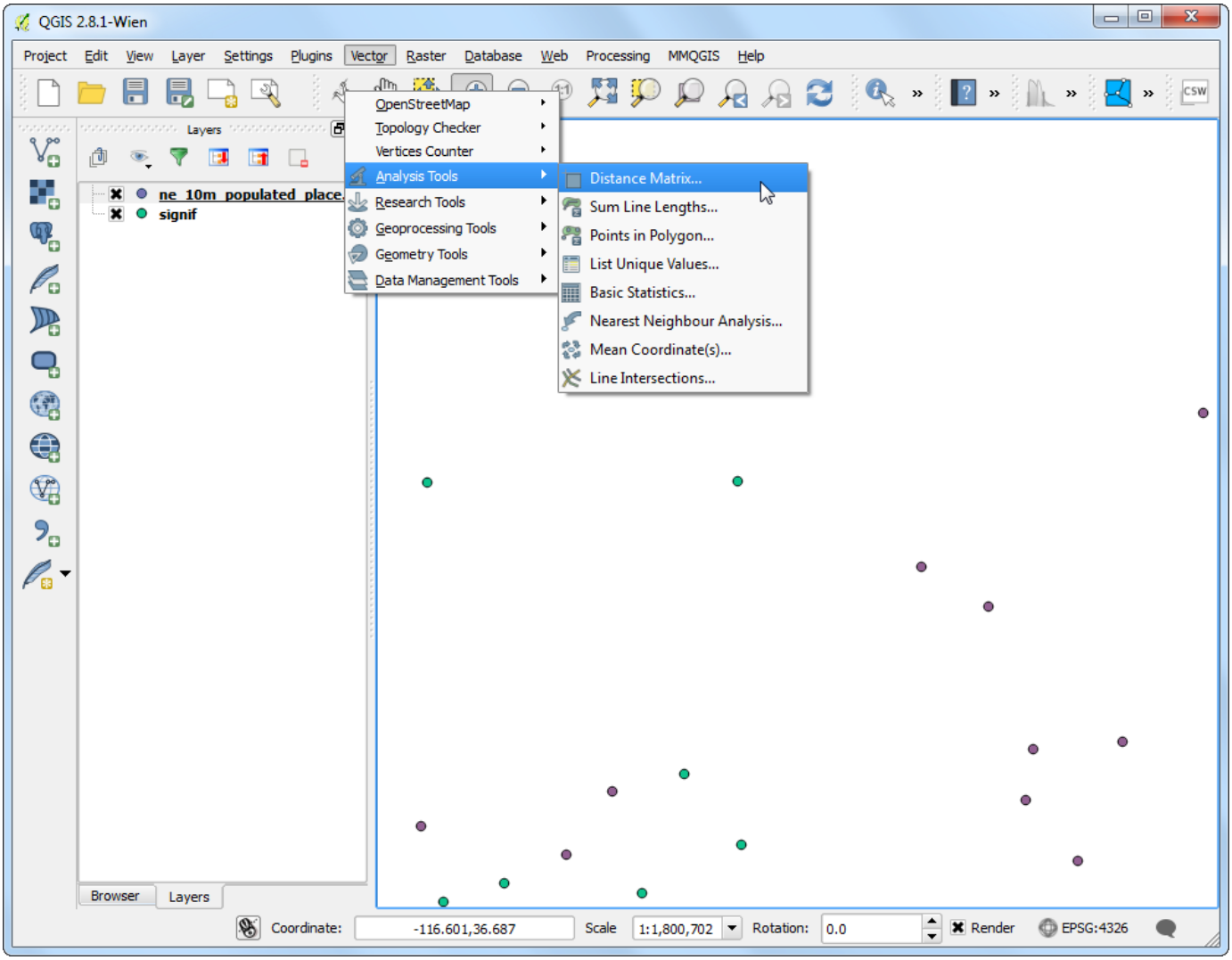
4. Browse to the downloaded `ne_10m_populated_places_simple.zip` file and click Open.



5. Zoom around and explore both the datasets. Each purple point represents the location of a significant earthquake and each blue point represents the location of a populated place. We need a way to find out the nearest point from the populated places layer for each of the points in the earthquake layer.



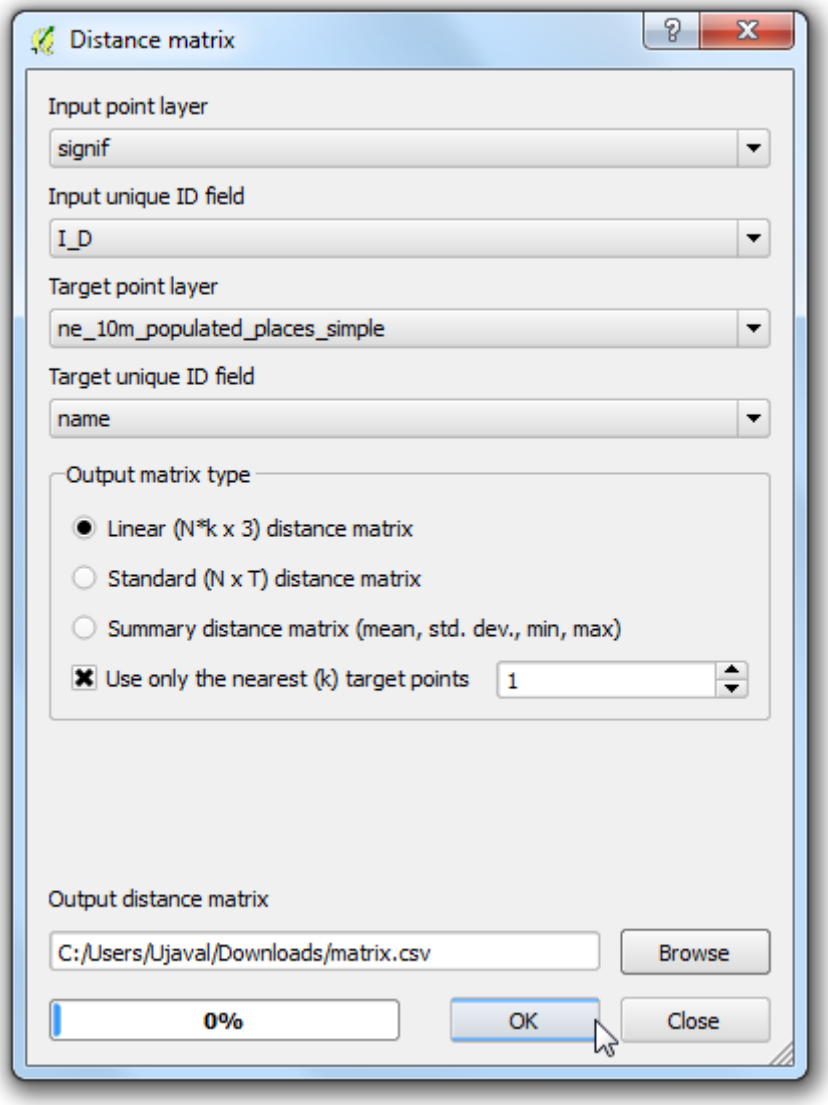
6. Go to Vector ▸ Analysis Tools ▸ Distance Matrix.



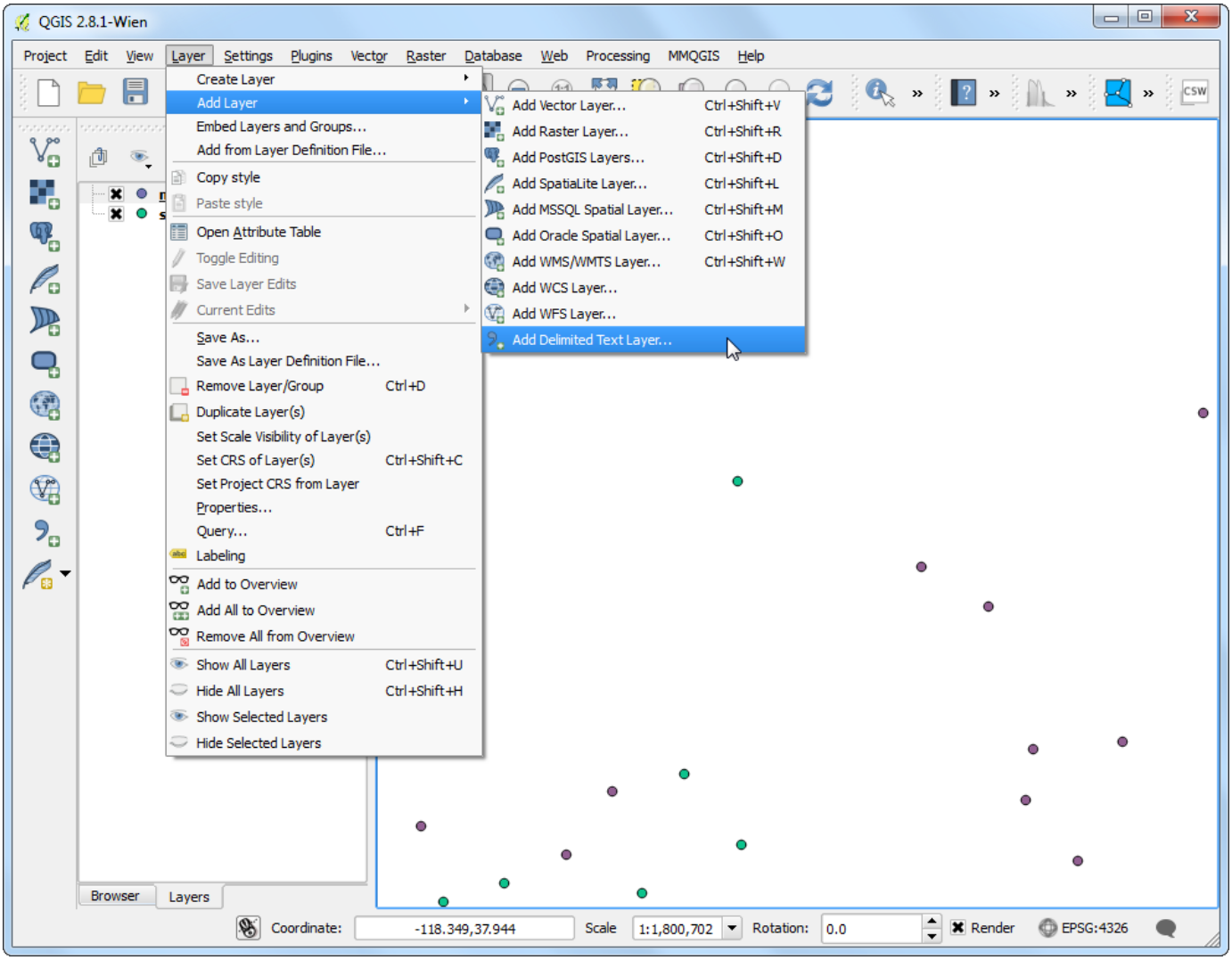
7. Here select the earthquake layer `signif` as the Input point layer and the populated places `ne_10m_populated_places_simple` as the target layer. You also need to select a unique field from each of these layers which is how your results will be displayed. In this analysis, we are looking to get only `1` nearest point, so check the Use only the nearest(k) target points, and enter 1. Name your output file `matrix.csv`, and click OK. Once the processing finishes, click Close.

Note

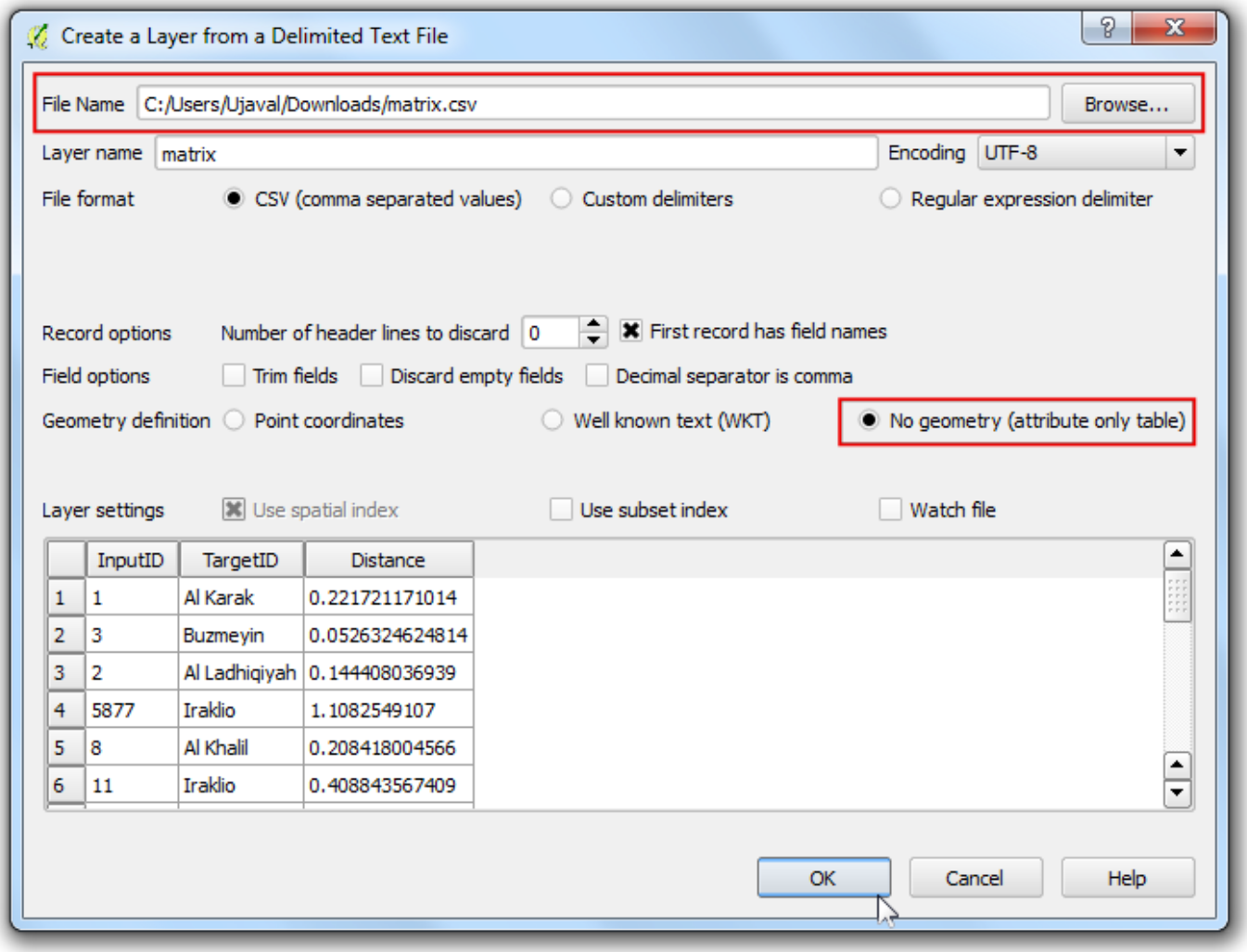
A useful thing to note is that you can even perform the analysis with only 1 layer. Select the same layer as both Input and Target. The result would be a nearest neighbor from the same layer instead of a different layer as we have used here.



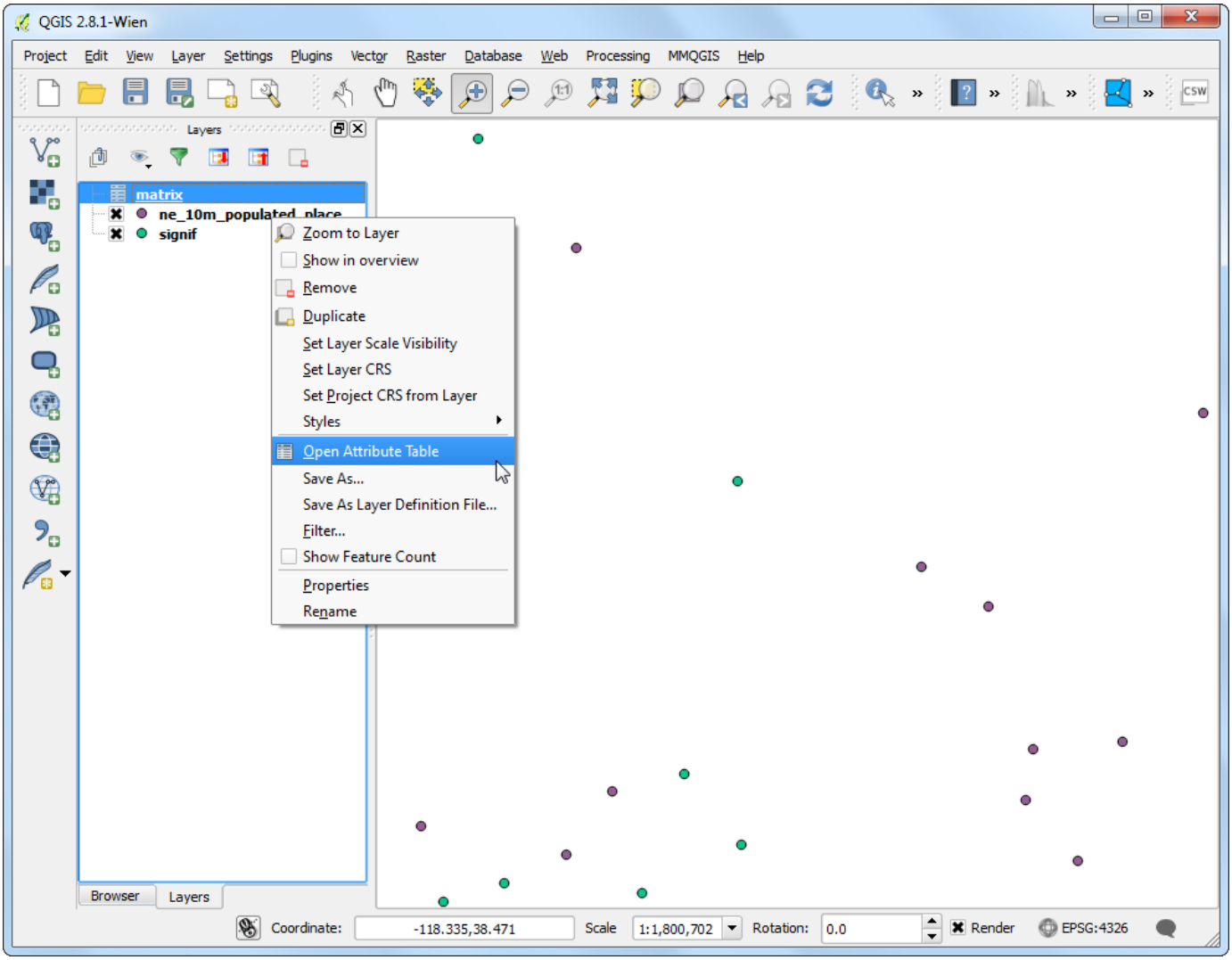
8. Once the processing finishes, click the Close button in the Distance Matrix dialog. You can now view the `matrix.csv` file in Notepad or any text editor. QGIS can import CSV files as well, so we will add it to QGIS and view it there. Go to Layer ▸ Add Layer ▸ Add Delimited Text Layer....



9. Browse to the newly created `matrix.csv` file. Since this file is just text columns, select No geometry (attribute only table) as the Geometry definition. Click OK.



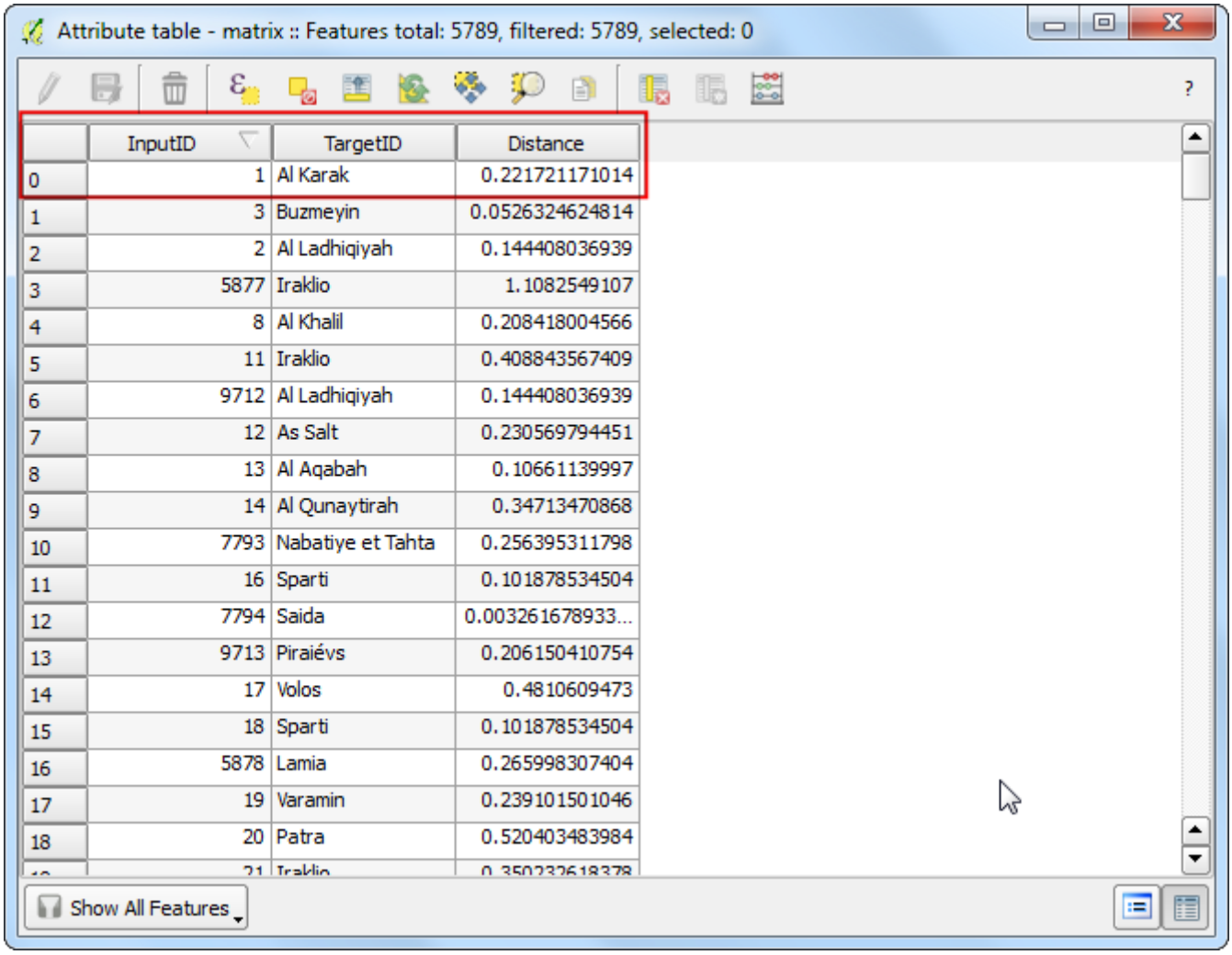
10. You will see the CSV file loaded as a table. Right-click on the table layer and select Open Attribute Table.



11. Now you will be able to see the content of our results. The InputID field contains the field name from the Earthquake layer. The TargetID field contains the name of the feature from the Populated Places layer that was the closest to the earthquake point. The Distance field is the distance between the 2 points.

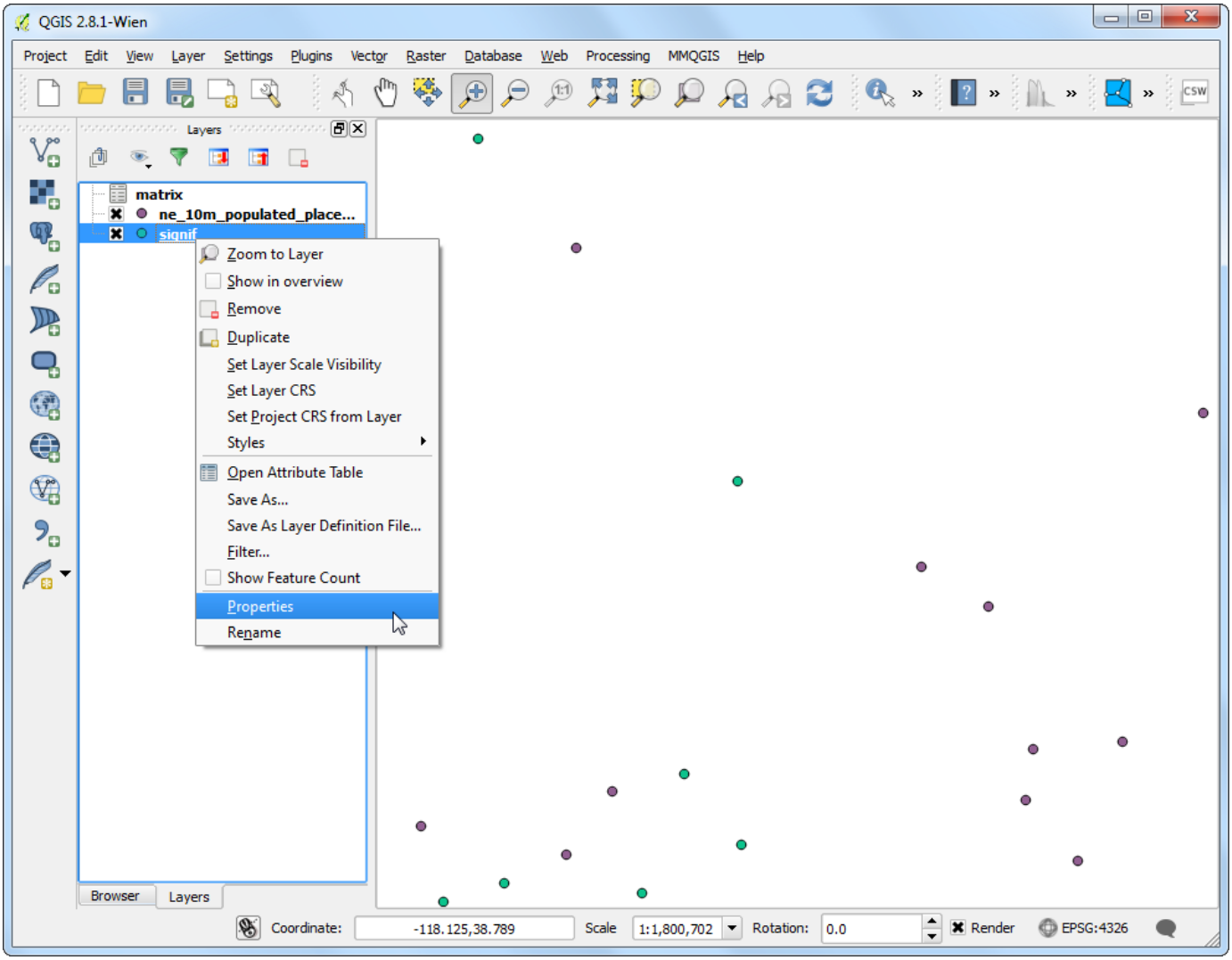
Note

Remember that the *distance* calculation will be done using the layers' Coordinate Reference System. Here the distance will be in *decimal degrees* units because our source layer coordinates are in degrees. If you want distance in meters, reproject the layers before running the tool.

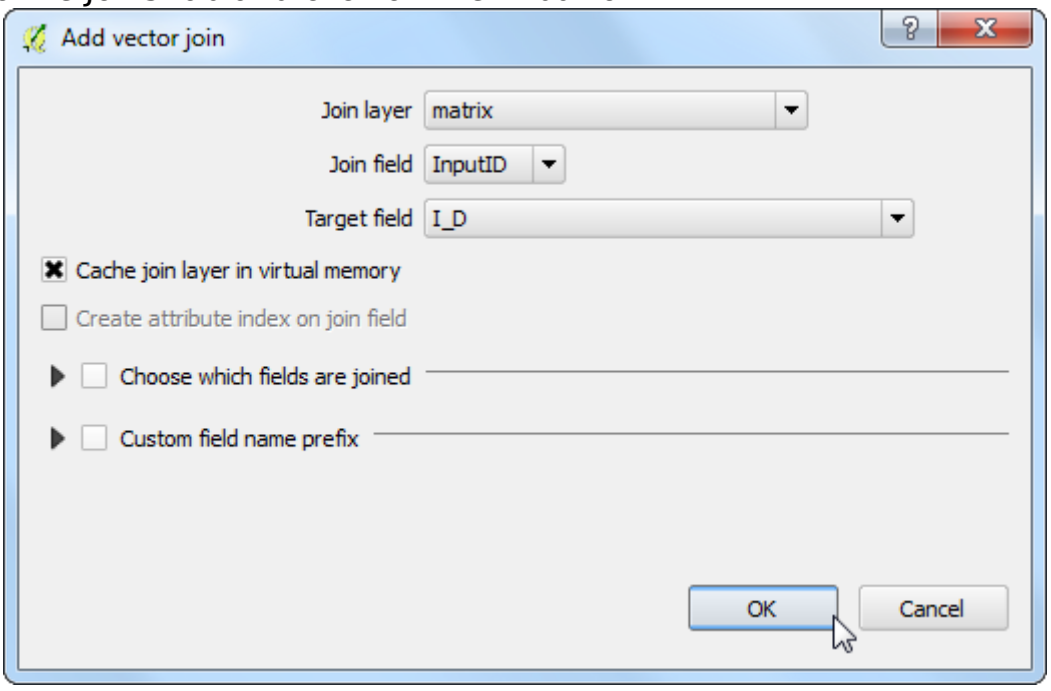


	InputID	TargetID	Distance
0	1	Al Karak	0.221721171014
1	3	Buzmeyin	0.0526324624814
2	2	Al Ladhiqiyah	0.144408036939
3	5877	Iraklio	1.1082549107
4	8	Al Khalil	0.208418004566
5	11	Iraklio	0.408843567409
6	9712	Al Ladhiqiyah	0.144408036939
7	12	As Salt	0.230569794451
8	13	Al Aqabah	0.10661139997
9	14	Al Qunaytirah	0.34713470868
10	7793	Nabatiye et Tahta	0.256395311798
11	16	Sparti	0.101878534504
12	7794	Saida	0.003261678933...
13	9713	Piraiévs	0.206150410754
14	17	Volos	0.4810609473
15	18	Sparti	0.101878534504
16	5878	Lamia	0.265998307404
17	19	Varamin	0.239101501046
18	20	Patra	0.520403483984
	21	Iraklio	0.350232618378

12. This is very close to the result we were looking for. For some users, this table would be sufficient. However, we can also integrate this results in our original Earthquake layer using a **Table Join**. Right-click on the Earthquake layer, and select Properties.

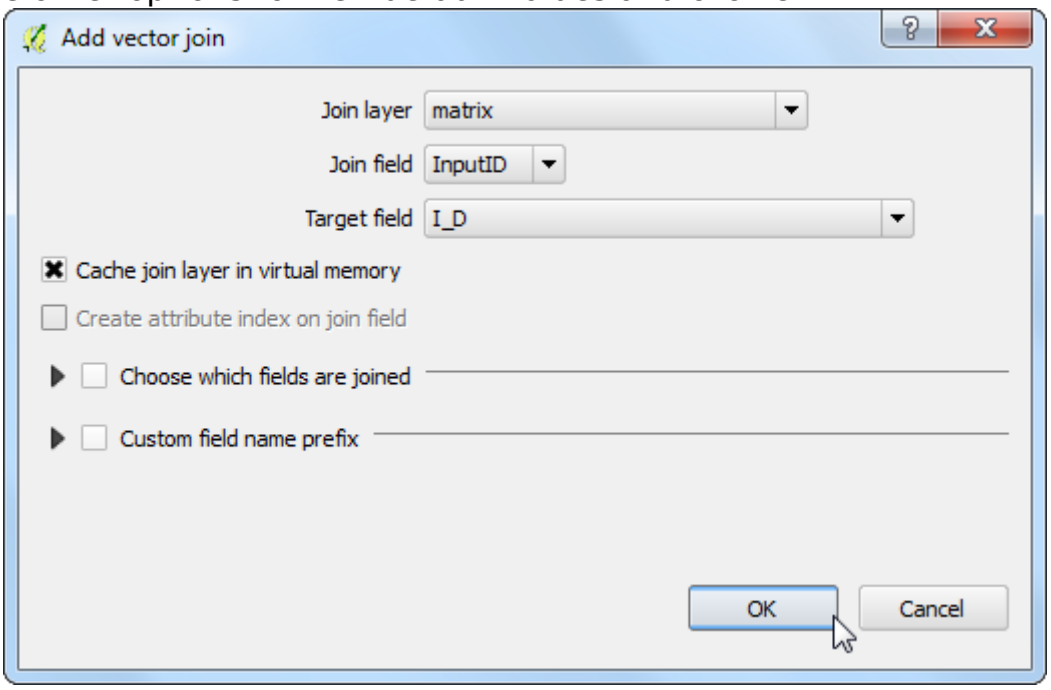


13. Go to the Joins tab and click on the + button.

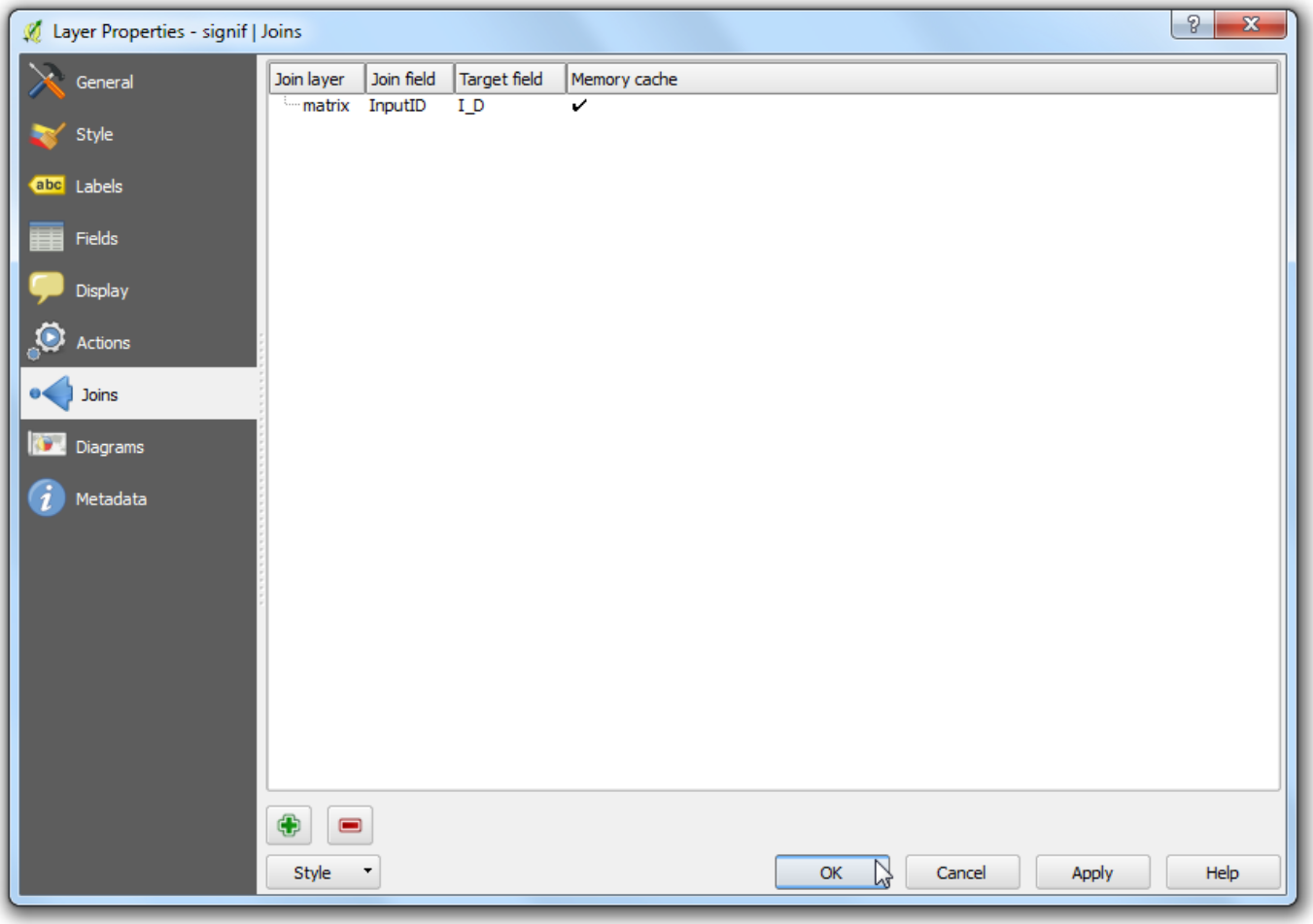


14. We want to join the data from our analysis result to this layer. We need to select a field from each of the layers that has the same values. Select `matrix`

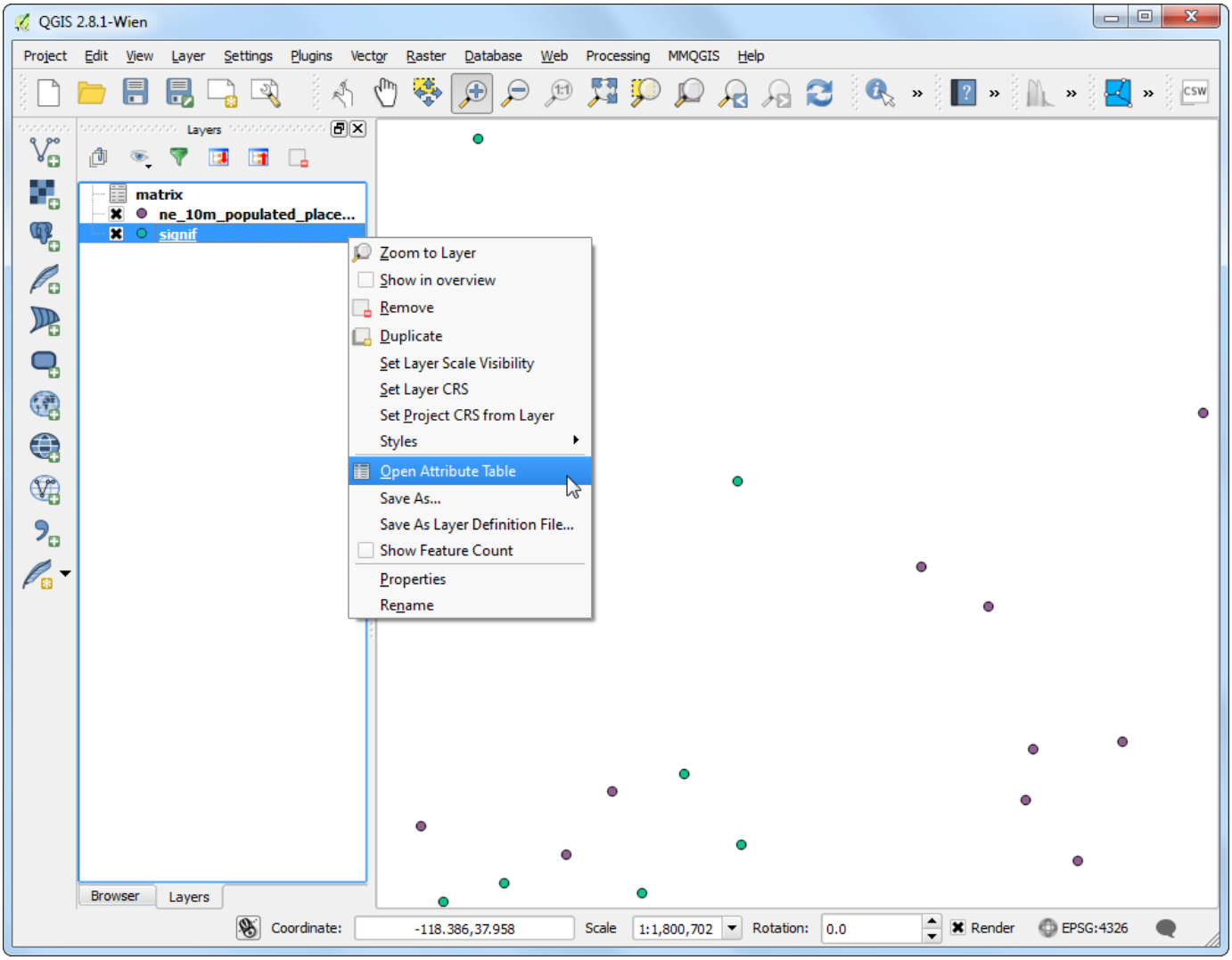
as the Join layer` and `InputID` as the Join field. The Target field would be `I_D`. Leave other options to their default values and click OK.



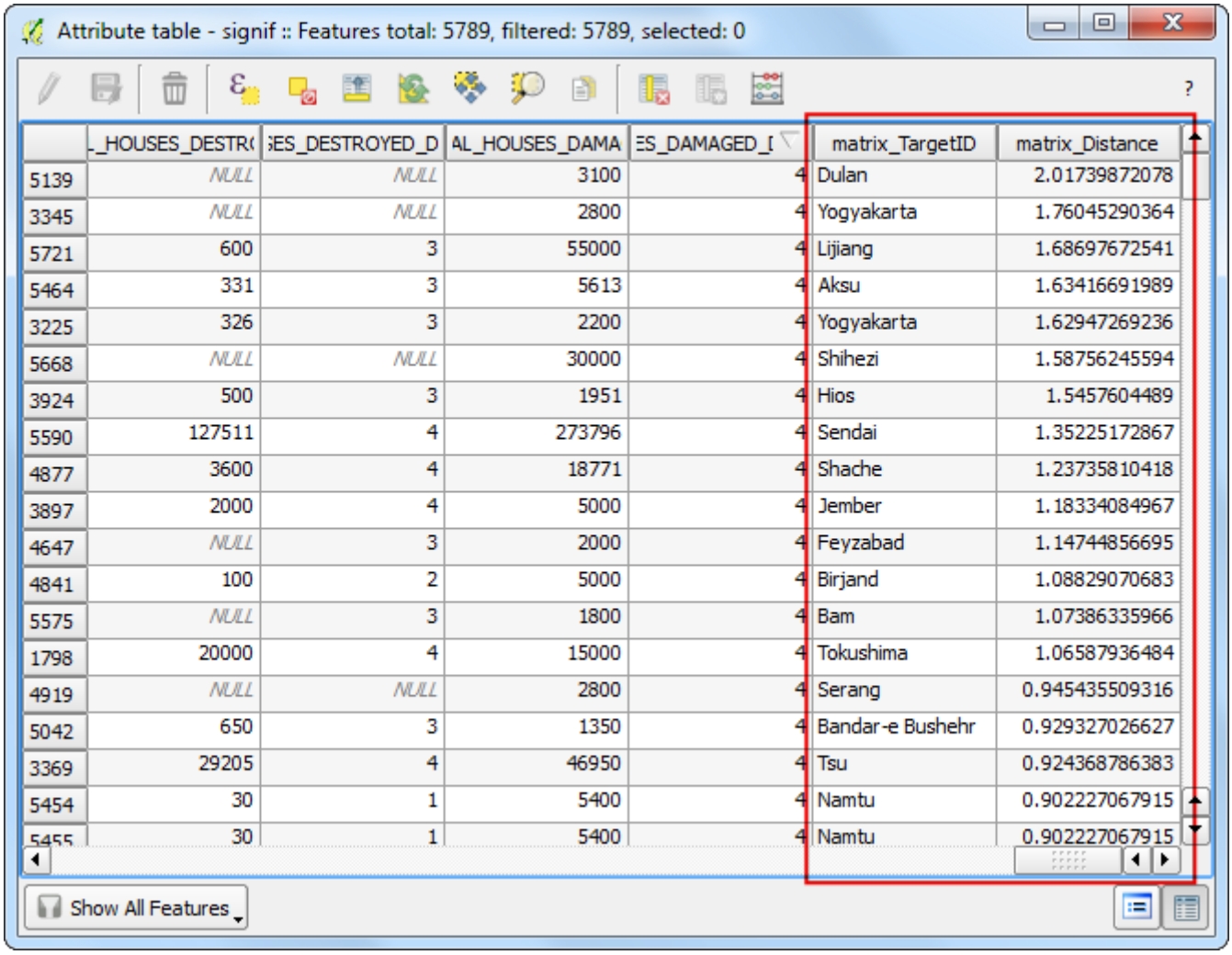
15. You will see the join appear in the Joins tab. Click OK.



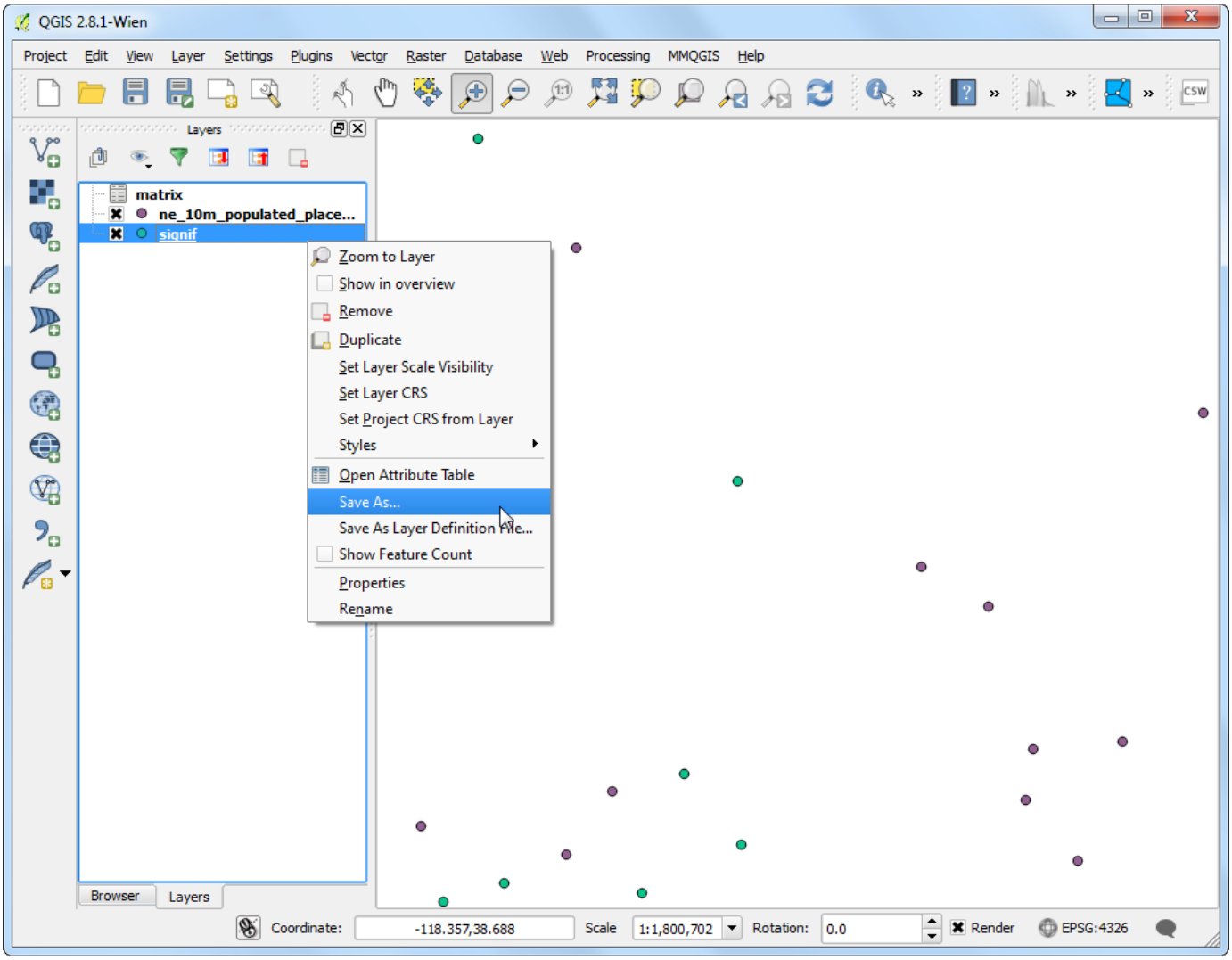
16. Now open the attribute table of the `signif` layer by right-clicking and selecting Open Attribute Table.



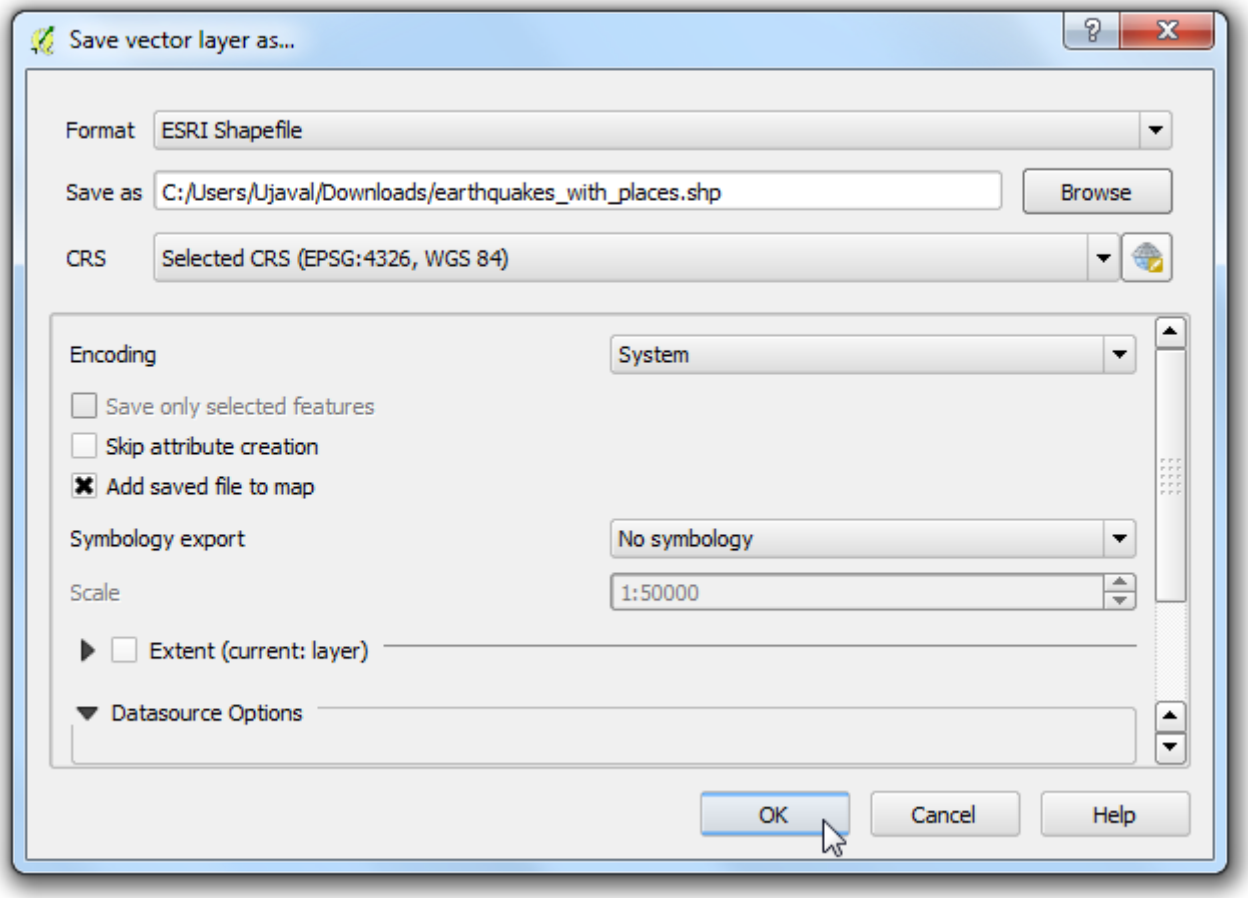
17. You will see that for every Earthquake feature, we now have an attribute which is the nearest neighbor (closest populated place) and the distance to the nearest neighbor.



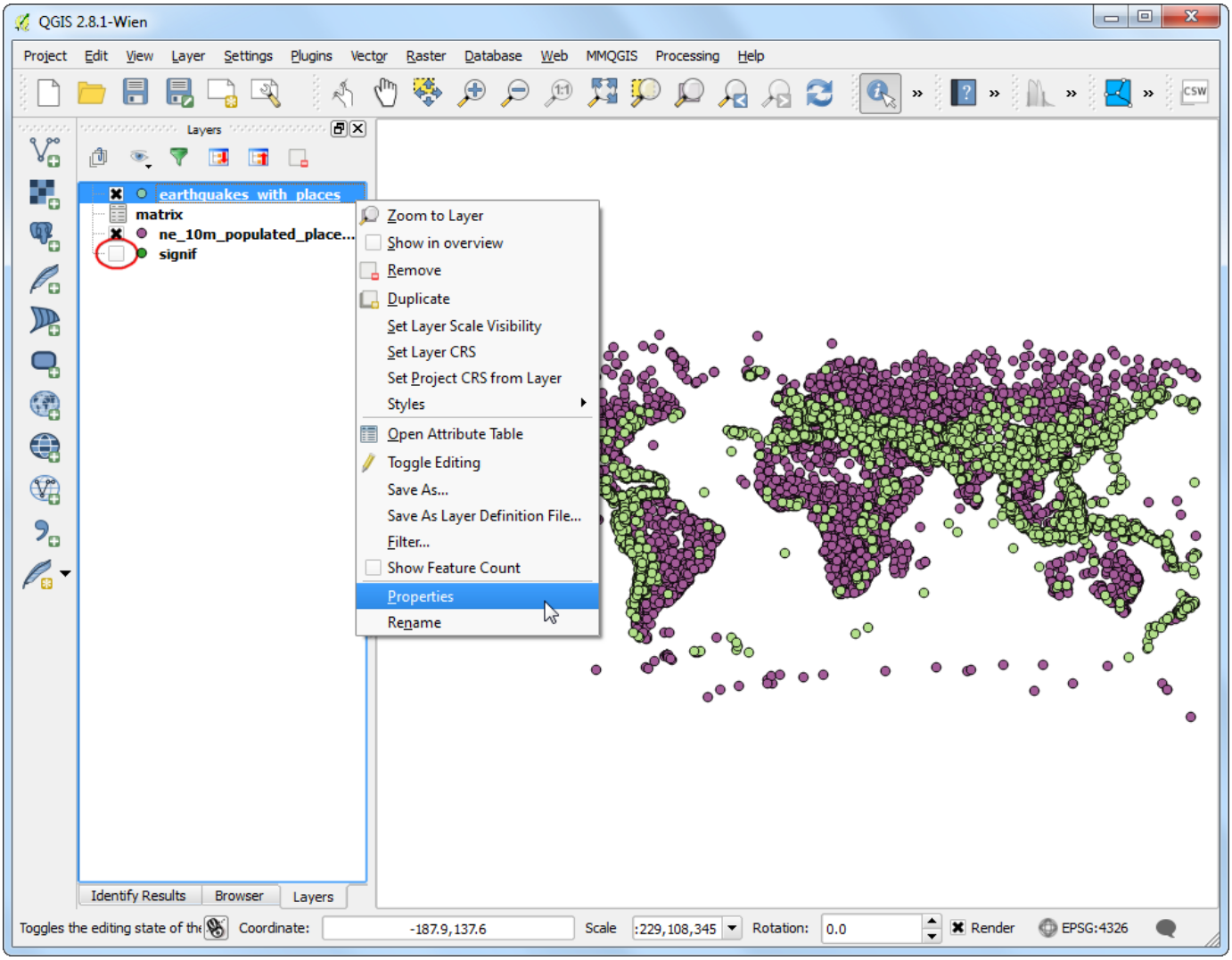
18. We will now explore a way to visualize these results. First, we need to make the table join permanent by saving it to a new layer. Right-click the `signif` layer and select Save As....



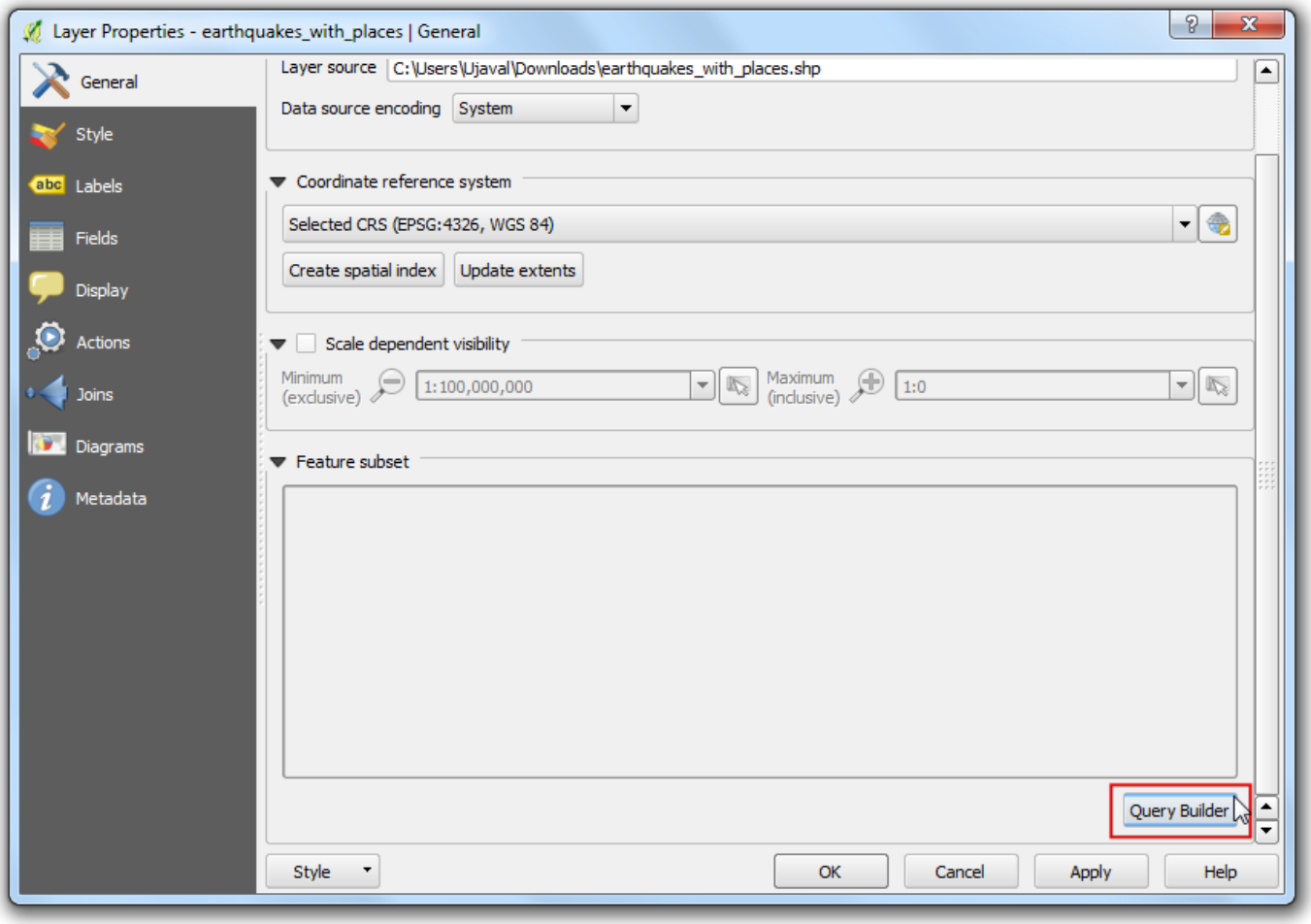
19. Click the Browse button next to Save as label and name the output layer as `earthquake_with_places.shp`. Make sure the Add saved file to map box is checked and click OK.



20. Once the new layer is loaded, you can turn off the visibility of the `signif` layer. As our dataset is quite large, we can run our visualization analysis on a subset of the data. QGIS has a neat feature where you can load a subset of features from a layer without having to export it to a new layer. Right-click the `earthquake_with_places` layer and select Properties.

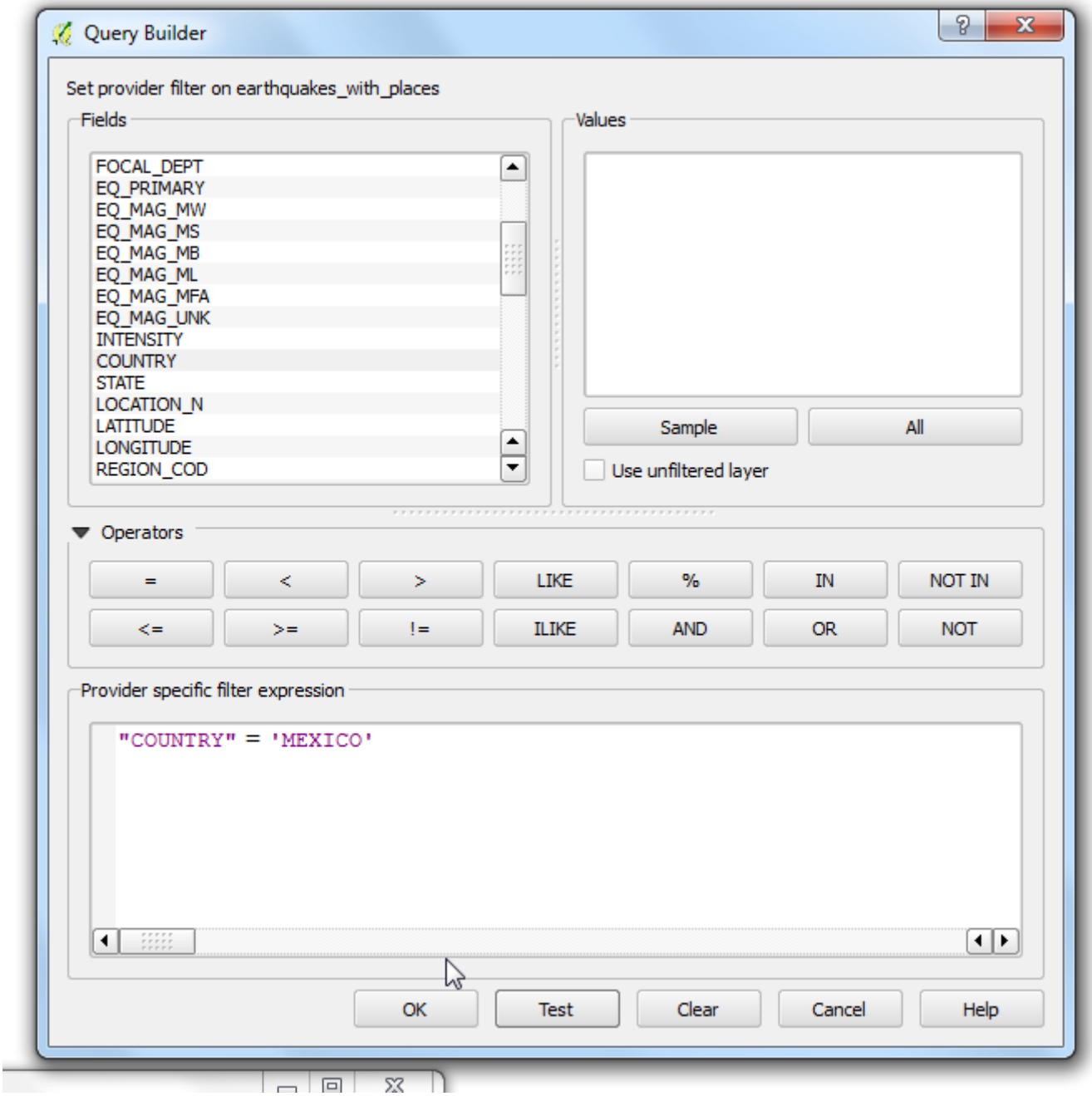


21. In the General tab, scroll down to the Feature subset section. Click Query Builder.

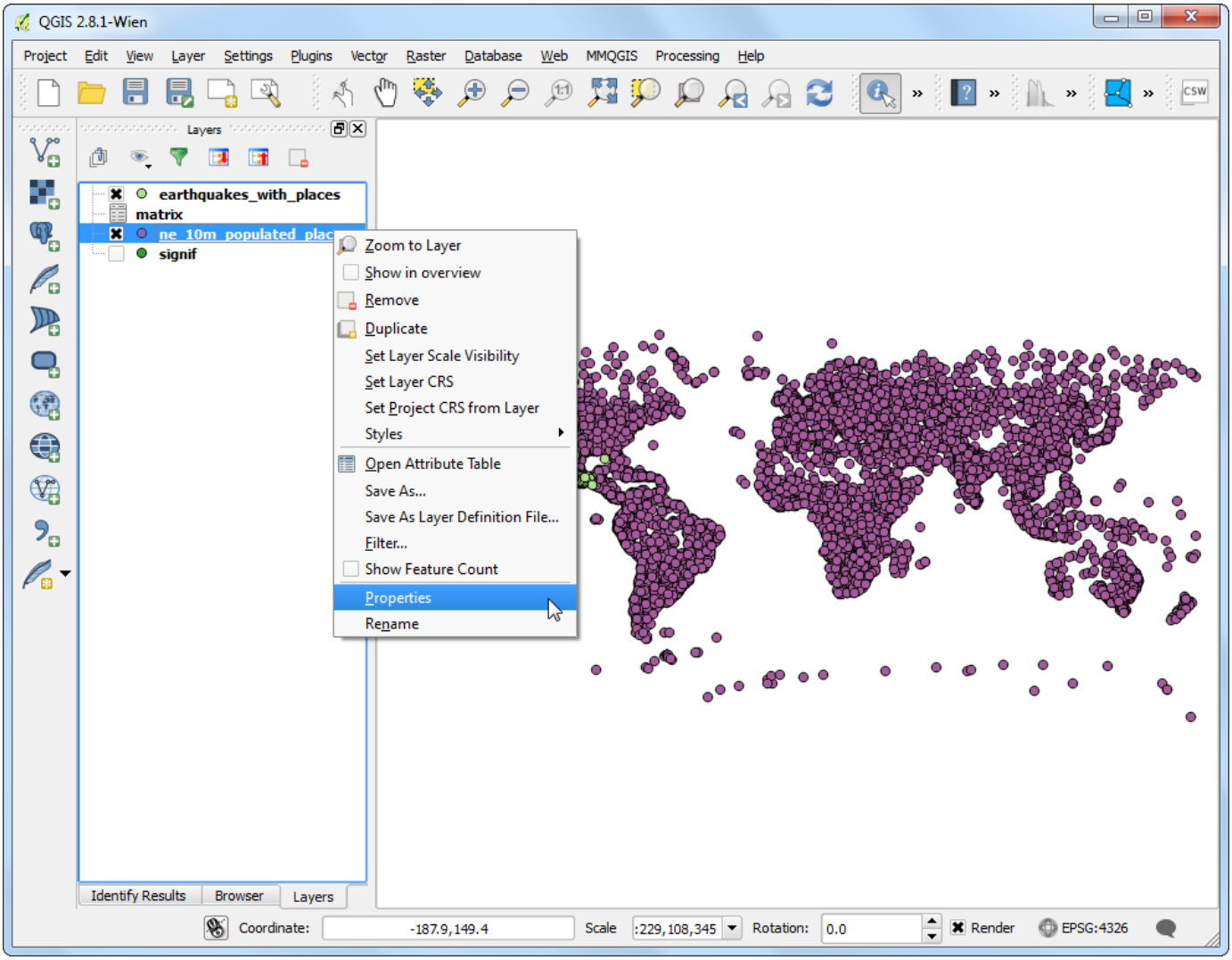


22. For this tutorial, we will visualize the earthquakes and their nearest populated places for Mexico. Enter the following expression in the Query Builder dialog.

```
"COUNTRY" = 'MEXICO'
```

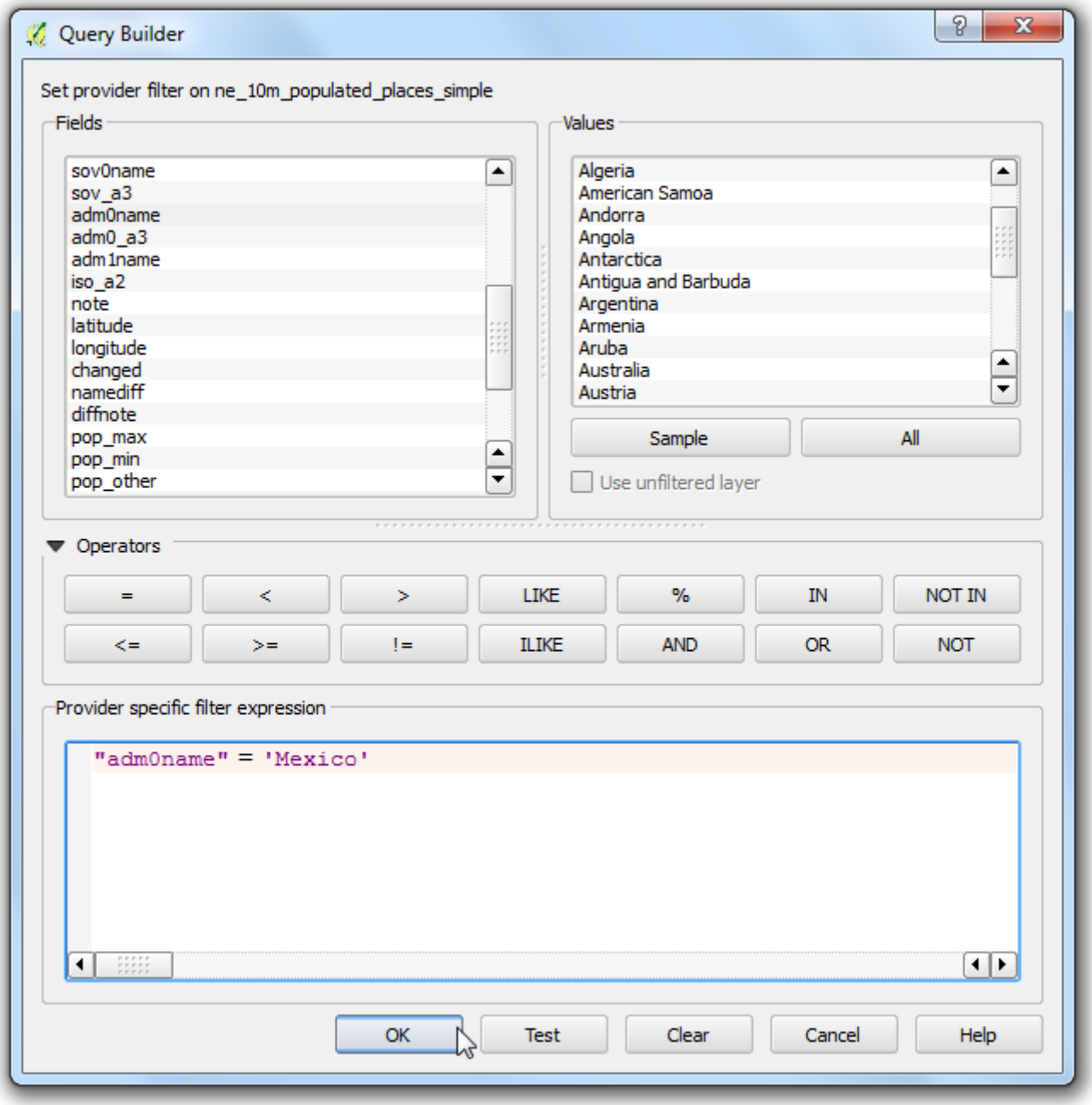


23. You will see that only the points falling within Mexico will be visible in the canvas. Let's do the same for the populated places layer. Right-click on the `ne_10m_populated_places_simple` layer and select Properties.

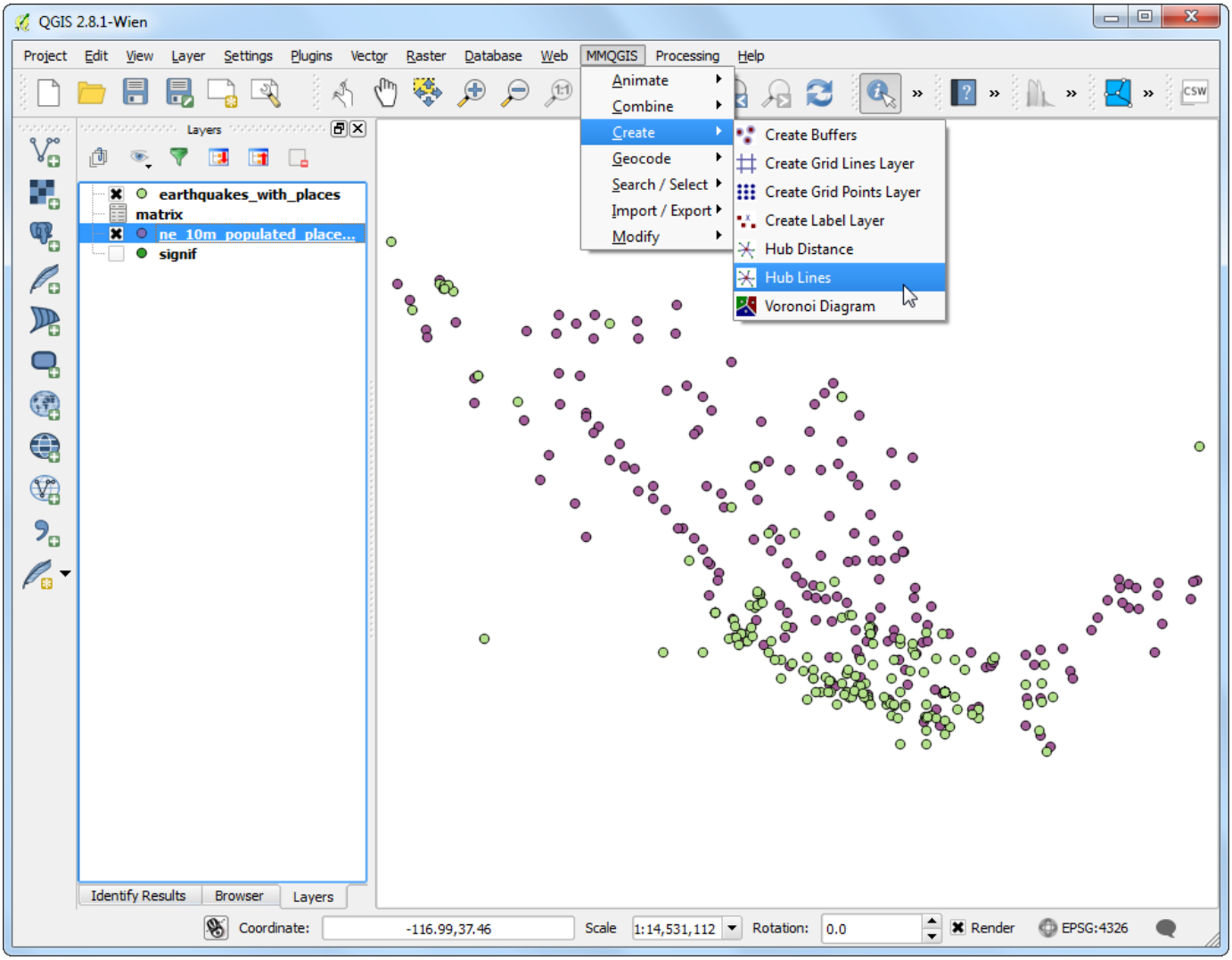


24. Open the Query Builder dialog from the General tab. Enter the following expression.

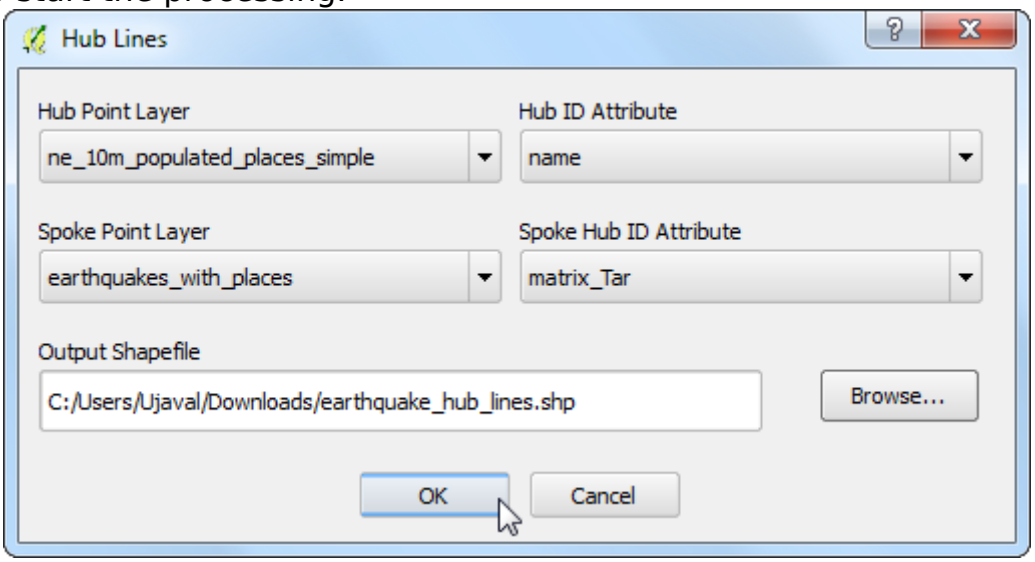
```
"adm0name" = 'Mexico'
```



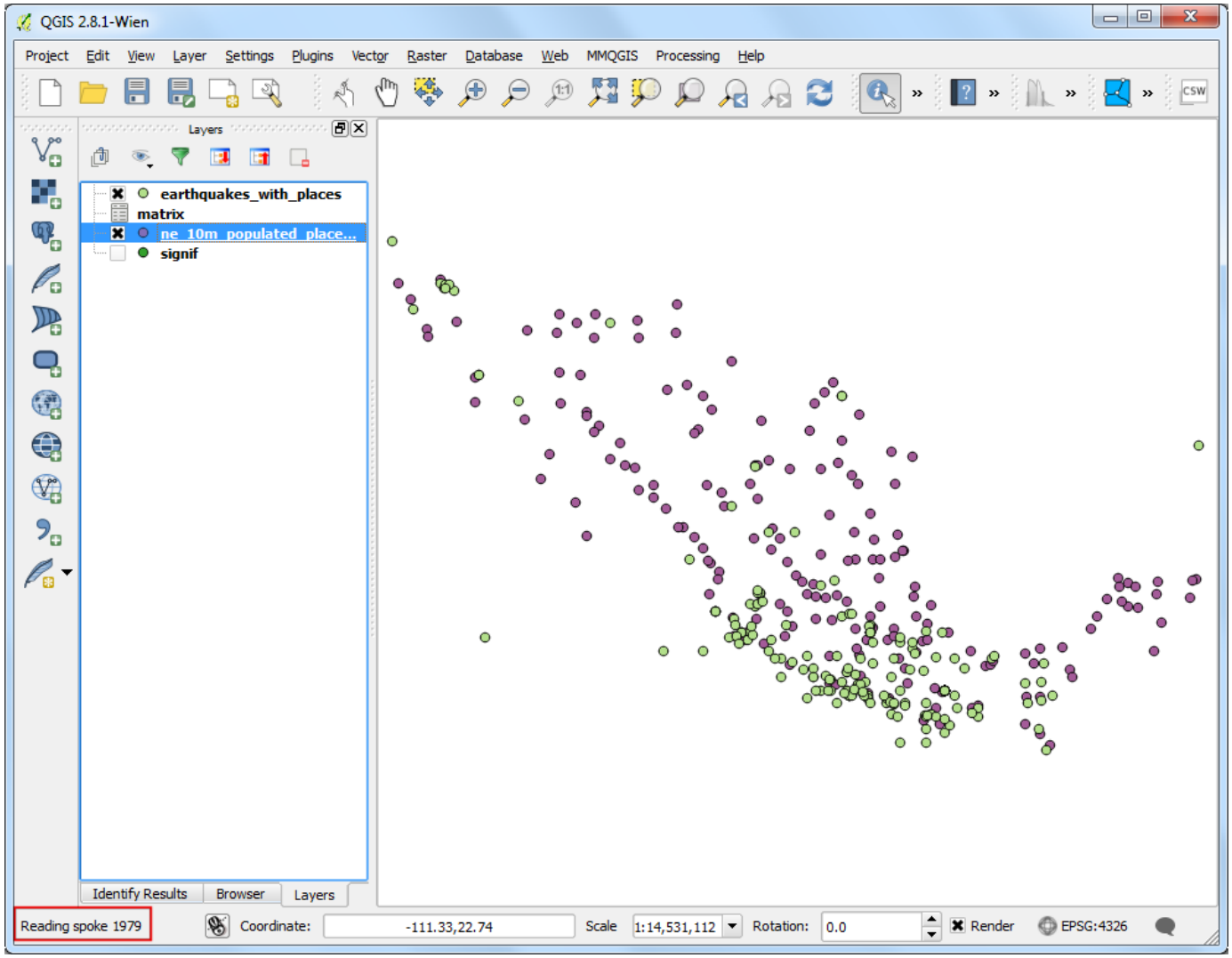
25. Now we are ready to create our visualization. We will use a plugin named `MMQGIS`. Find and install the plugin. See *Using Plugins* for more details on how to work with plugins. Once you have the plugin installed, go to MMQGIS ▸ Create ▸ Hub Lines.



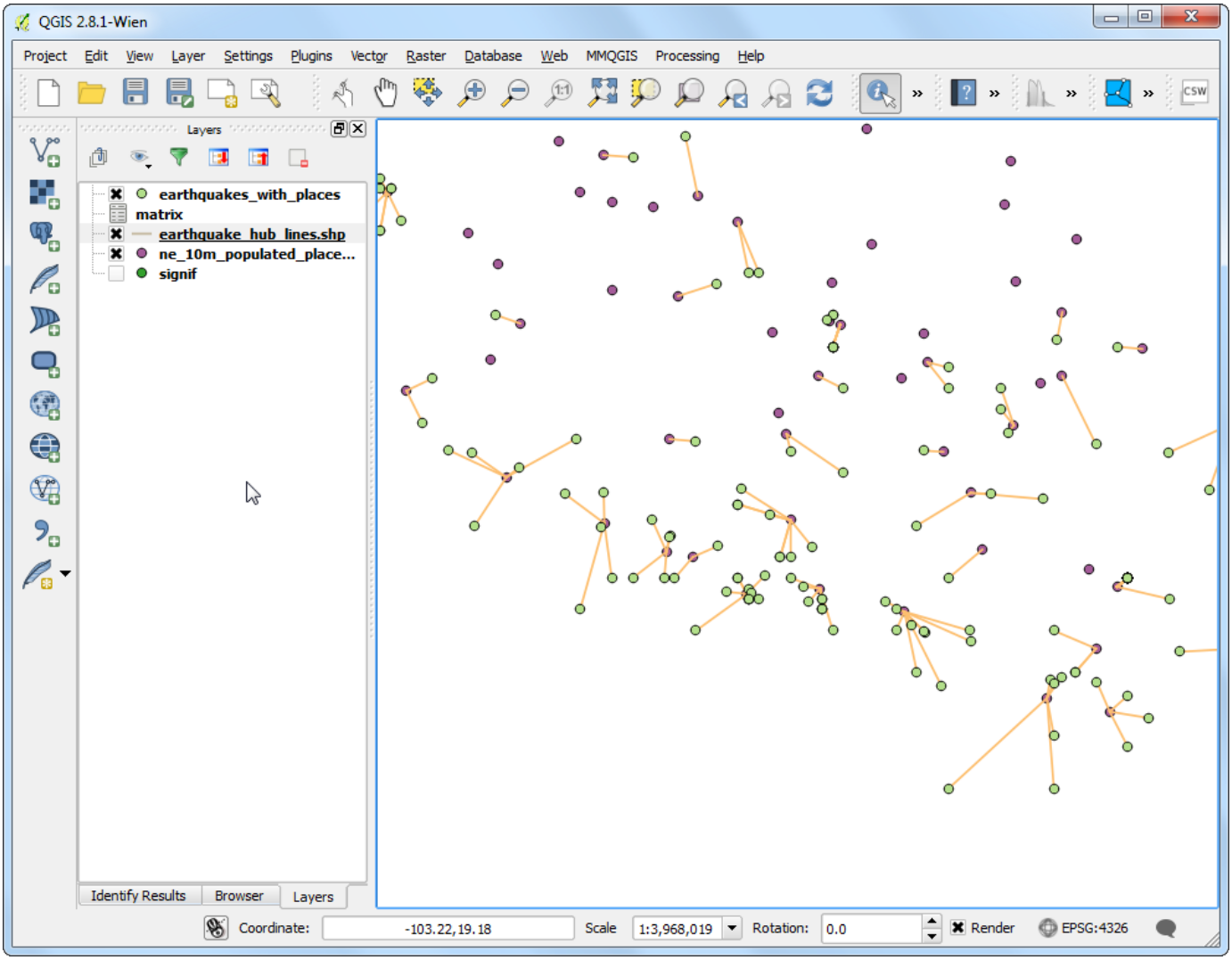
26. Select `ne_10m_populated_places_simple` as the Hub Point Layer and `name` as the Hub ID Attribute. Similarly, select `earthquake_with_places` as the Spoke Point Layer and `matrix_Tar` as the Spoke Hub ID Attribute. The hub lines algorithm will go through each of earthquake points and create a line that will join it to the populated place which matches the attribute we specified. Click Browse and name the Output Shapefile as `earthquake_hub_lines.shp`. Click OK to start the processing.



27. The processing may take a few minutes. You can see the progress on the bottom-left corner of the QGIS window.



28. Once the processing is done, you will see the `earthquake_hub_lines` layer loaded in QGIS. You can see that each earthquake point now has a line that connects it to the nearest populated place.


QGIS 2.8.1-Wien
Project Edit View Layer Settings Plugins Vector Raster Database Web MMQGIS Processing Help
Layers
earthquakes_with_places
matrix
earthquake_hub_lines.shp
ne_10m_populated_place...
signif
Identify Results
Browser
Layers
Coordinate:
-103.22,19.18
Scale
1:3,968,019
Rotation:
0.0
Render
EPSG:4326