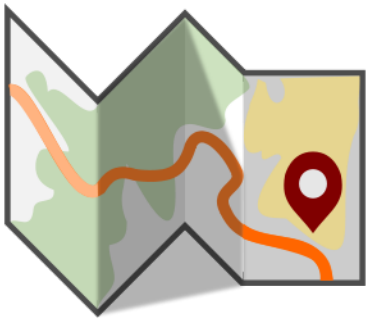
Nearest Neighbor Analysis

QGIS Tutorials and Tips



Ujaval Gandhi

http://google.com/+UjavalGandhi

This work is licensed under a Creative Commons Attribution 4.0 International License.

Nearest Neighbor Analysis

GIS is very useful is 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.)

Get the data

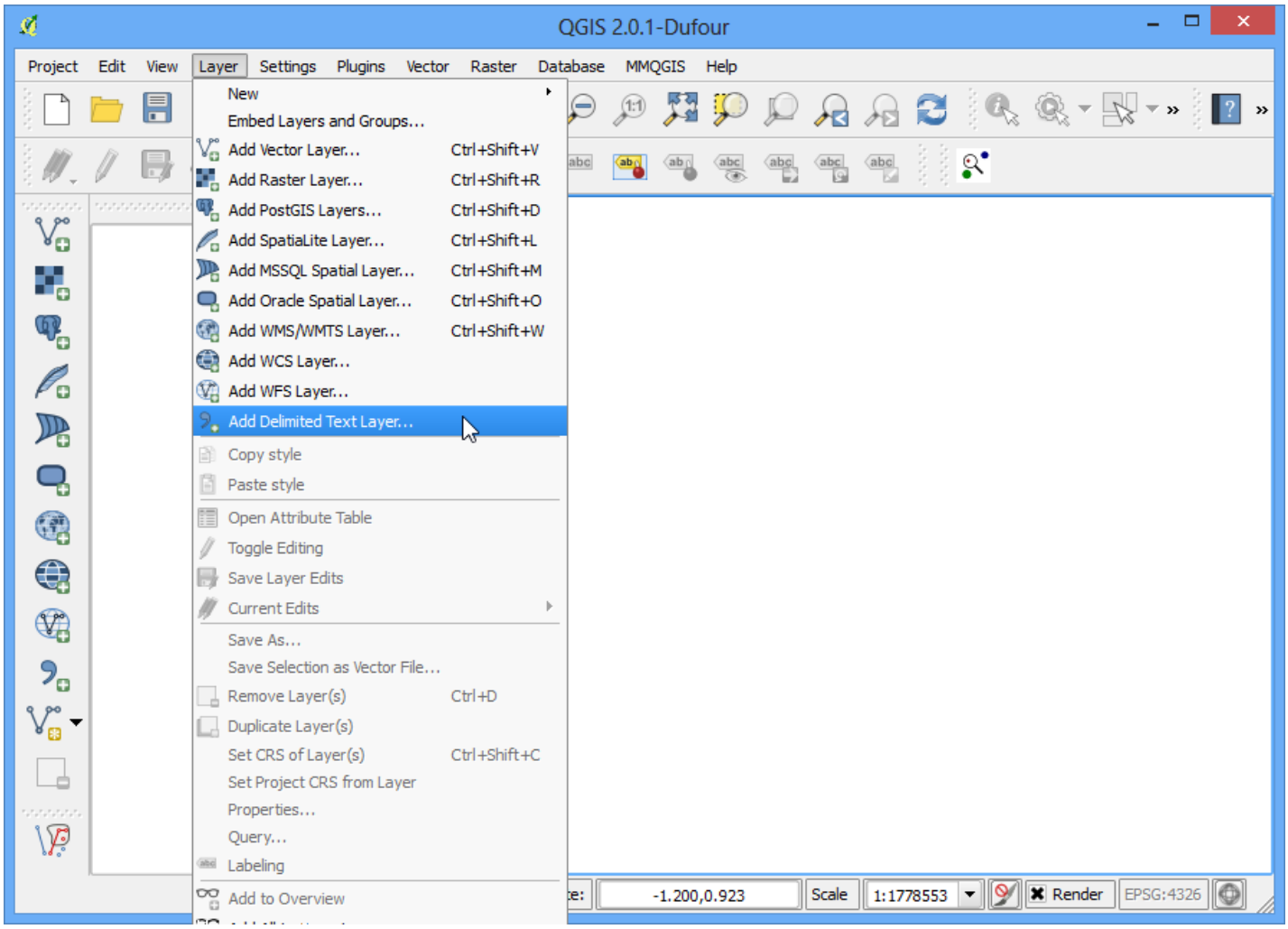
We will use NOAA's National Geophysical Data Center's Significant Earthquake Database as our layer represenging all major earthquakes. Download the tab-delimited earthquake data.

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

Data Sources: [NGDC] [NATURALEARTH]

Procedure

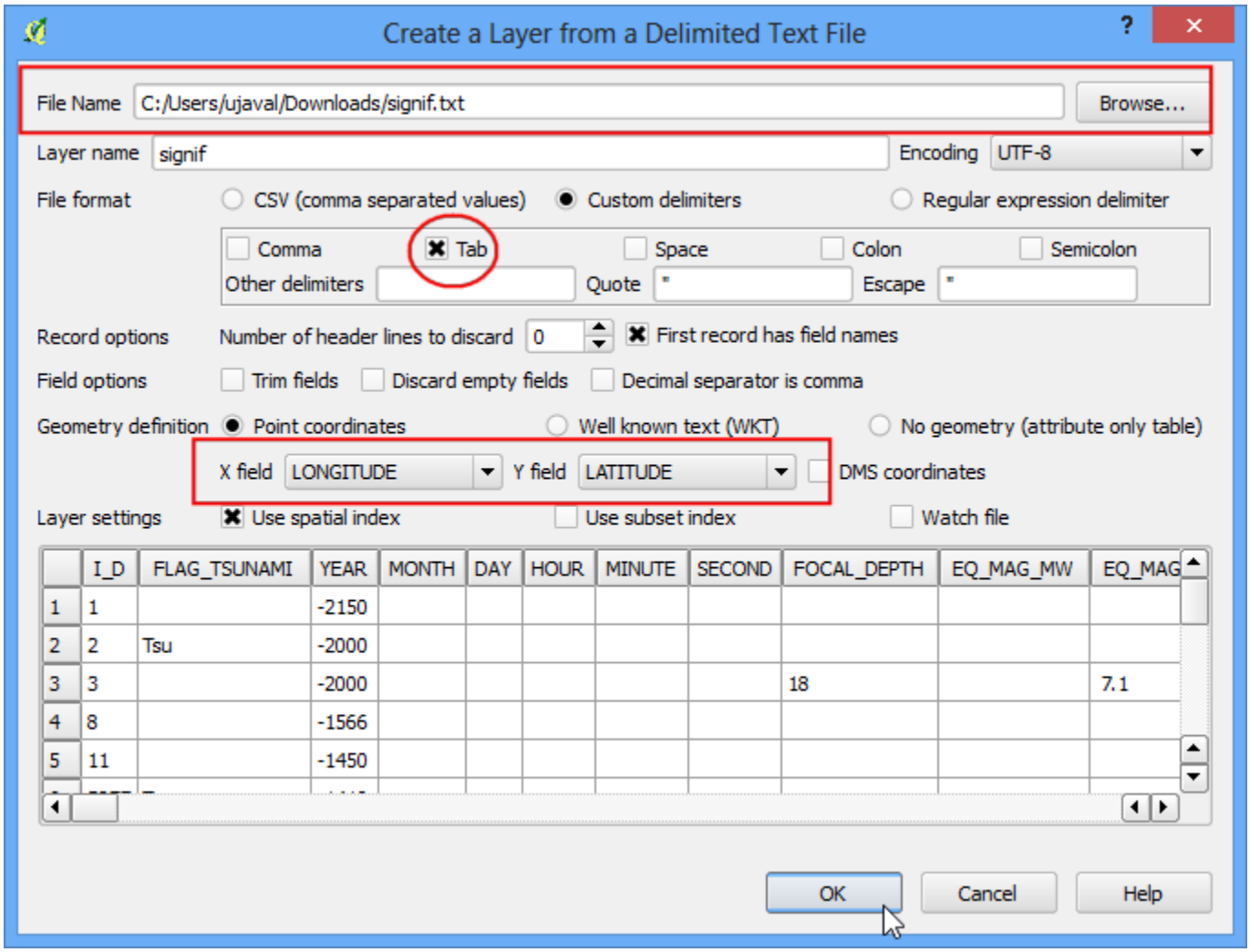
1. Open Layer ‣ Add Delimited Text Layer and browse to the downloaded ***signif.txt*** file.



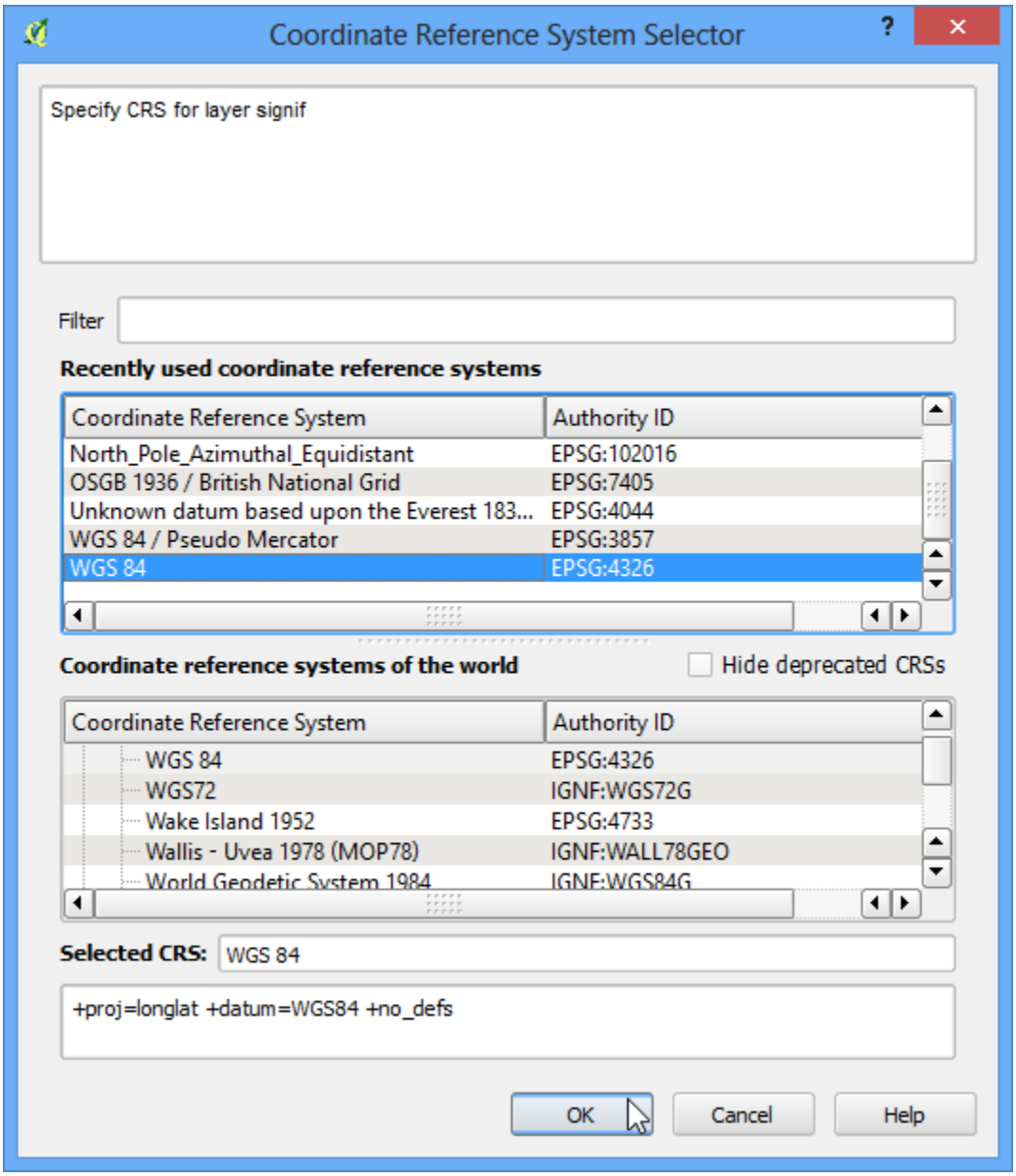
2. Since this is a tab-delimited file, choose Tab as the File format. The X field and Y field would be auto-populated. Click OK.

Note

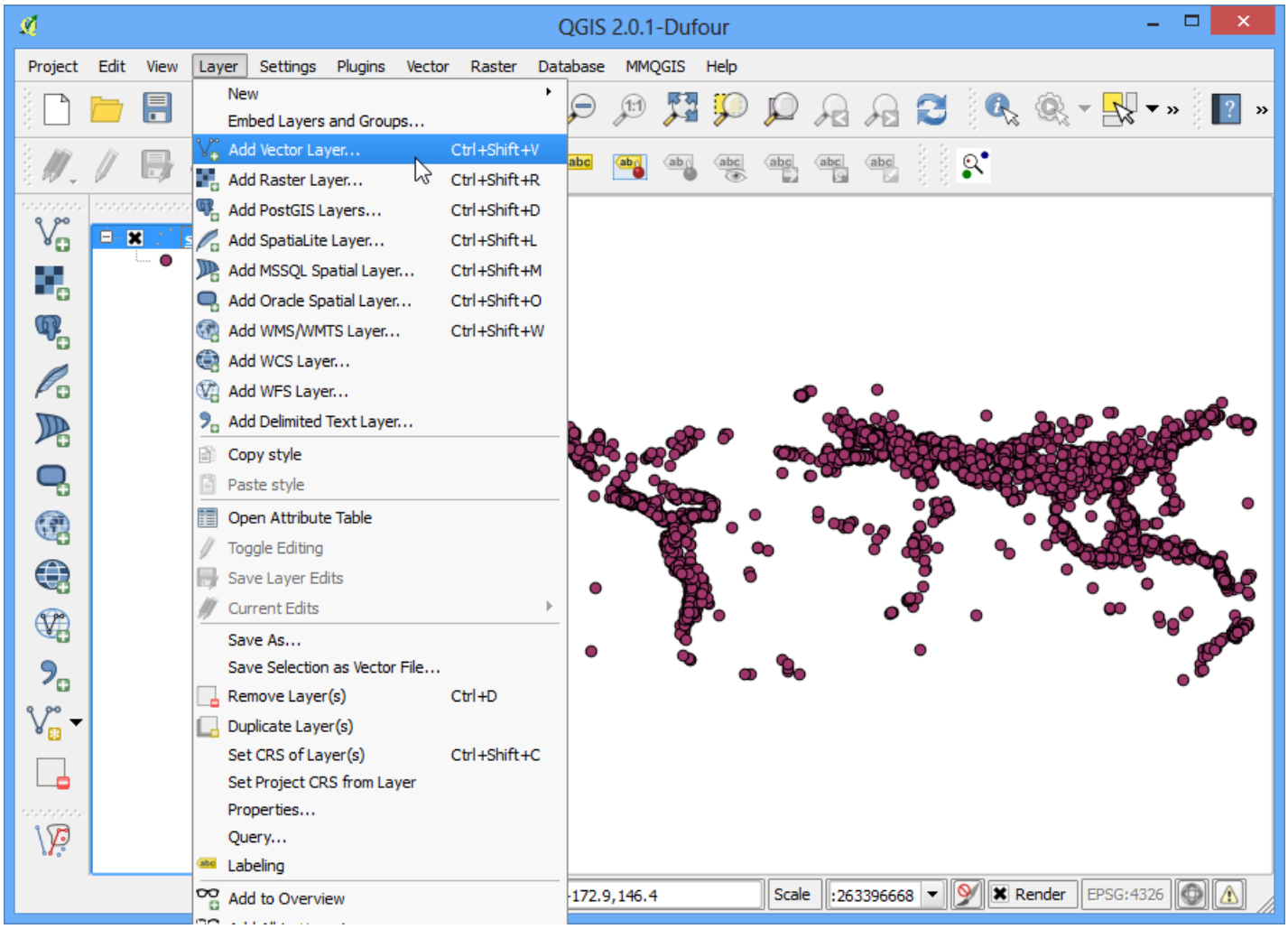
You may see some error messages as QGIS tries to import the file. These are valid errors and some rows from the file will not be imported. You can ignore the errors for the purpose of this tutorial.



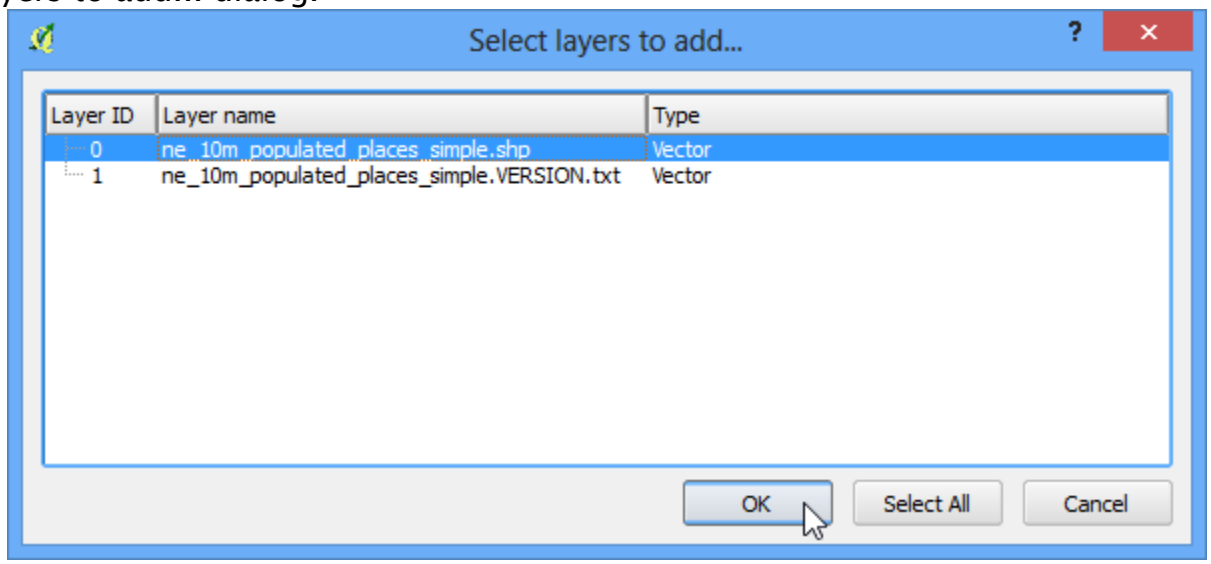
3. As the earthquake dataset has Latitude/Longitude coordinates, choose WGS 84 EPSG:436 as the CRS in the Coordinate Reference System Selector dialog.



4. The earthquake point layer would now be loaded and displayed in QGIS. Let's also open the Populated Places layer. Go to Layer ‣ Add Vector Layer.

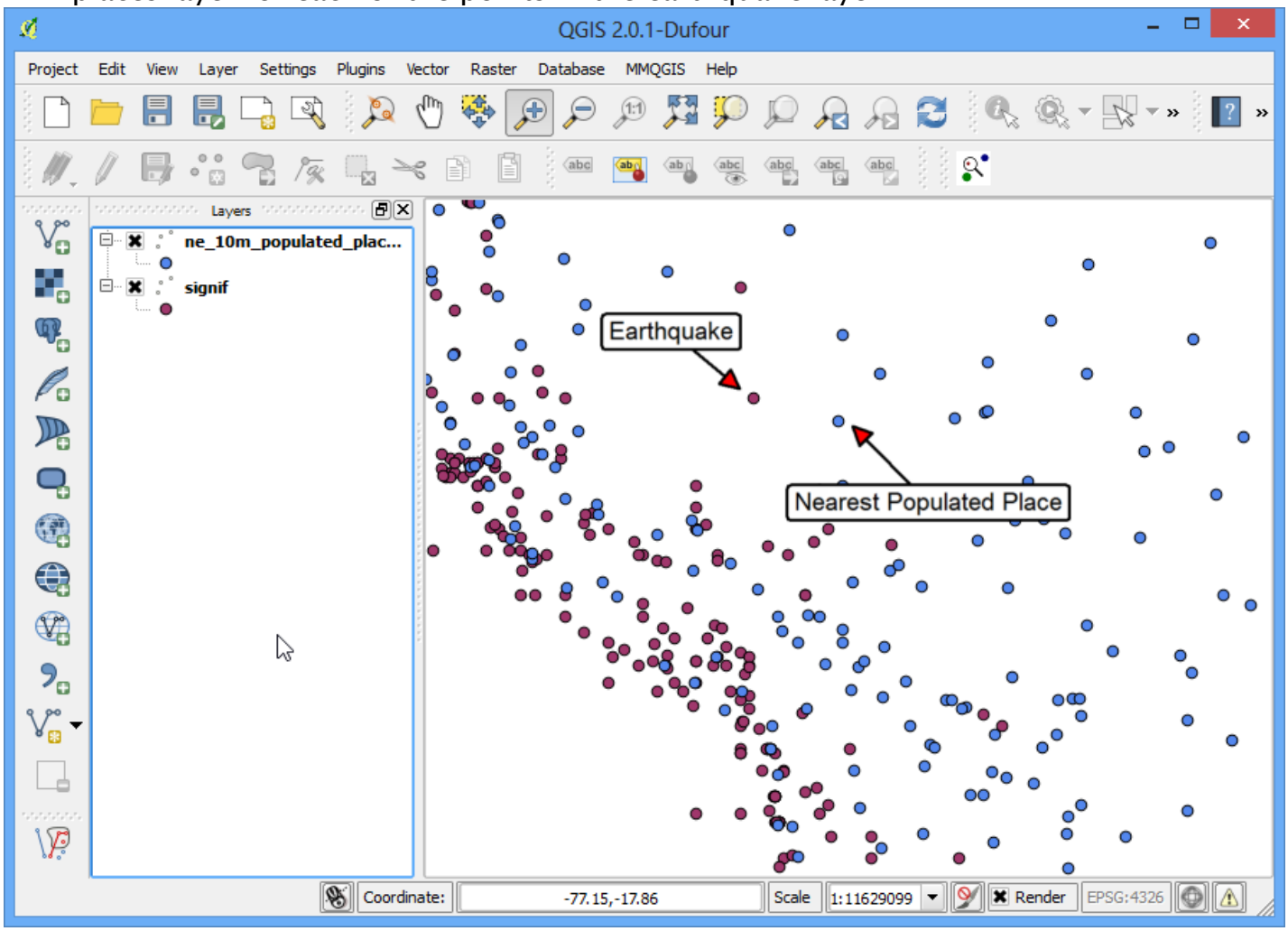


5. Browse to the downloaded ***ne_10m_populated_places_simple.zip*** file and click Open. Select the ***ne_10m_populated_places_simple.shp*** as the layer in the Select layers to add... dialog.

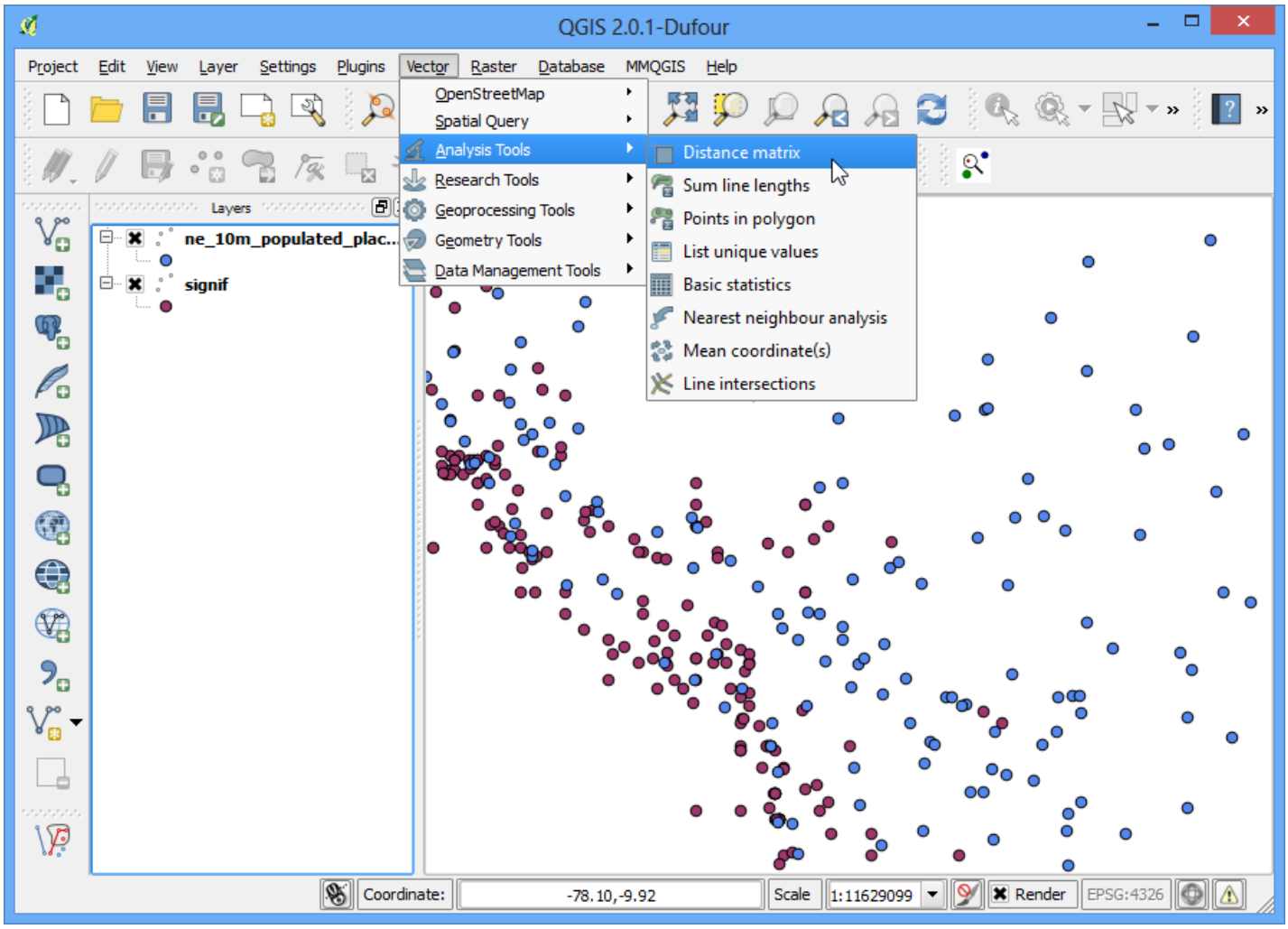


6. Zoom around and explore both the datasets. Each purple point represents the location of a significant earthquake and each blue point represents the location of a

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.



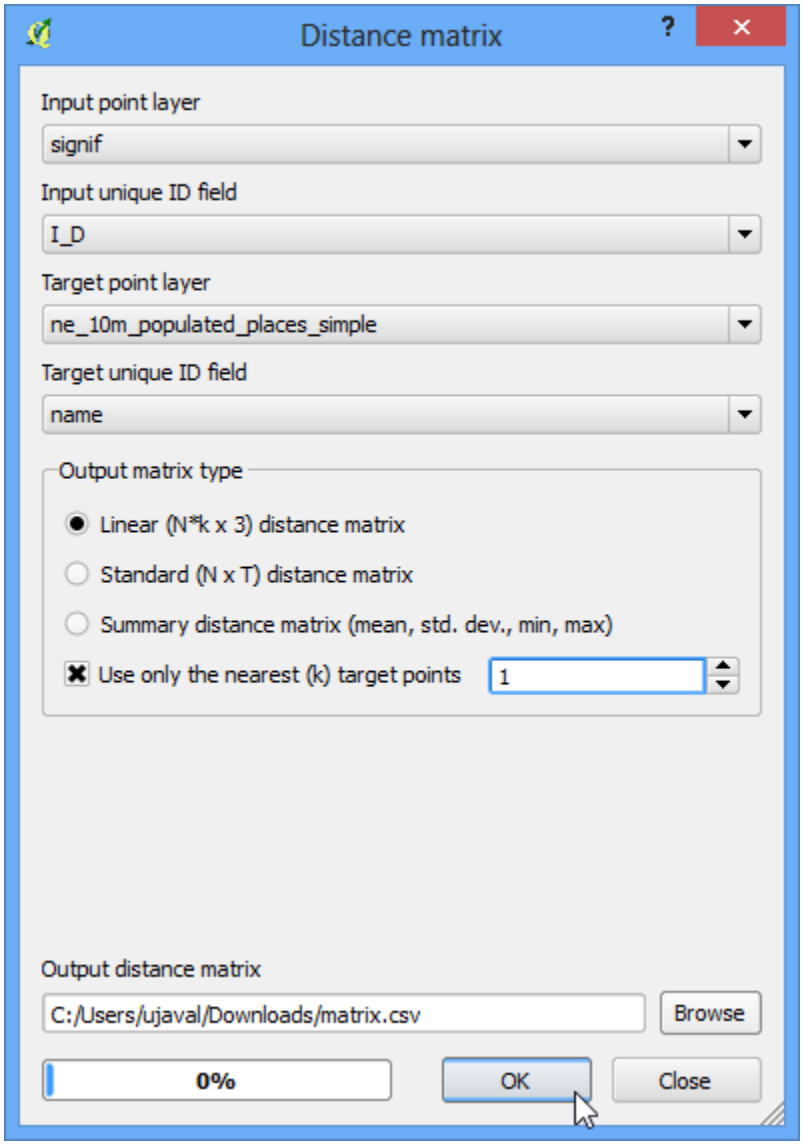
7. Go to Vector › Analysis Tools › Distance Matrix.



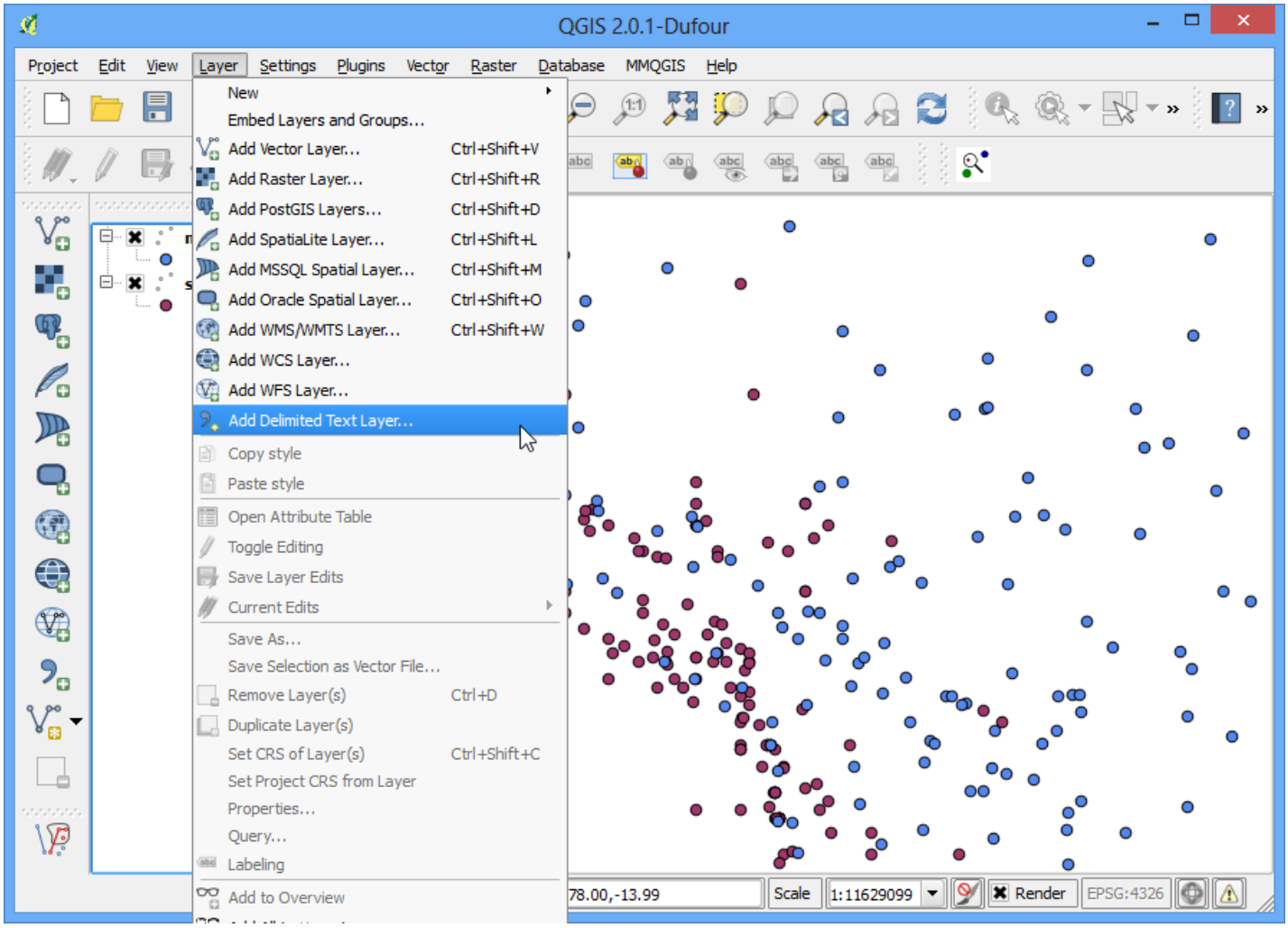
8. Here select the earthquake layer ***signif*** as the Input point layer and the populated places ***ne_10m_populated_places_simple*** as the target layer. You also need to select a unique field from each of these layers which is how your results will be displayed. In this analysis, we are looking to 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.

Note

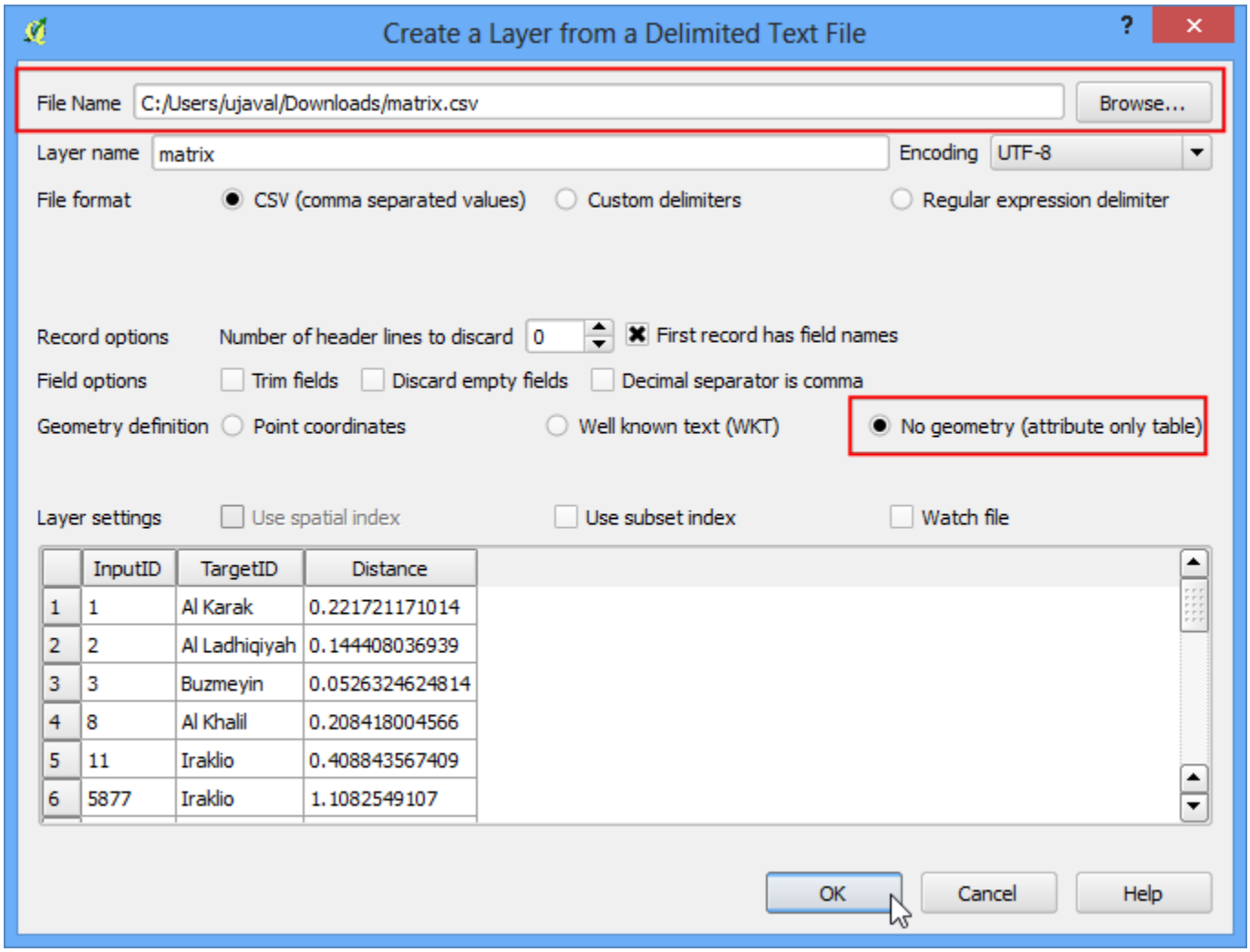
A useful thing to note is that you can even perform the analysis with only 1 layer. Select the same layer as both Input and Target. The result would be a nearest neighbor from the same layer instead of a different layer as we have used here.



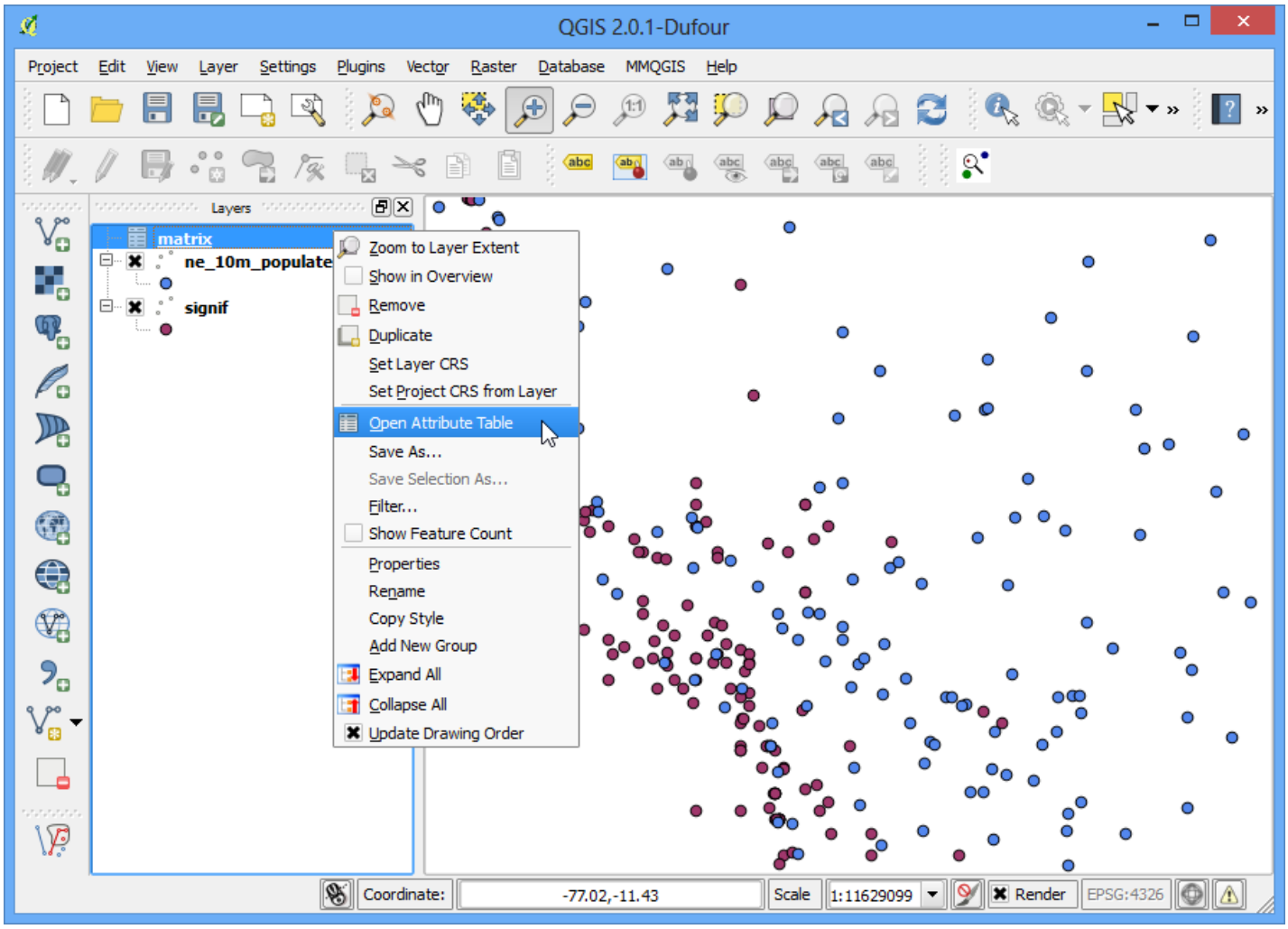
9. Once your file is generated, you can view it in Notepad or any text editor. QGIS can import CSV files as well, so we will add it to QGIS and view it there. Go to Layer ‣ Add Delimited Text Layer....



10. 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.



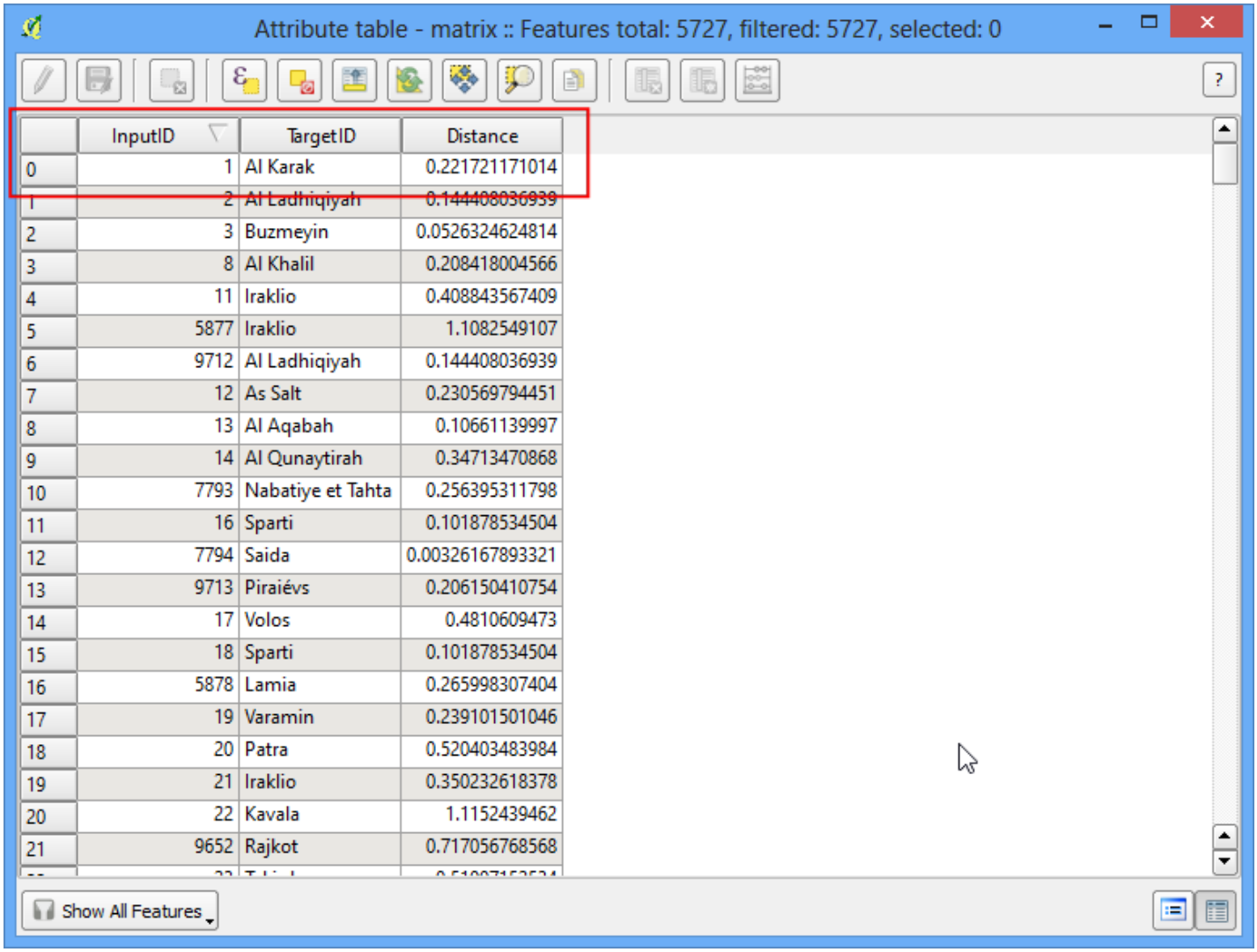
11. You will the the CSV file loaded as a table. Right-click on the table layer and select Open Attribute Table.



12. 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.

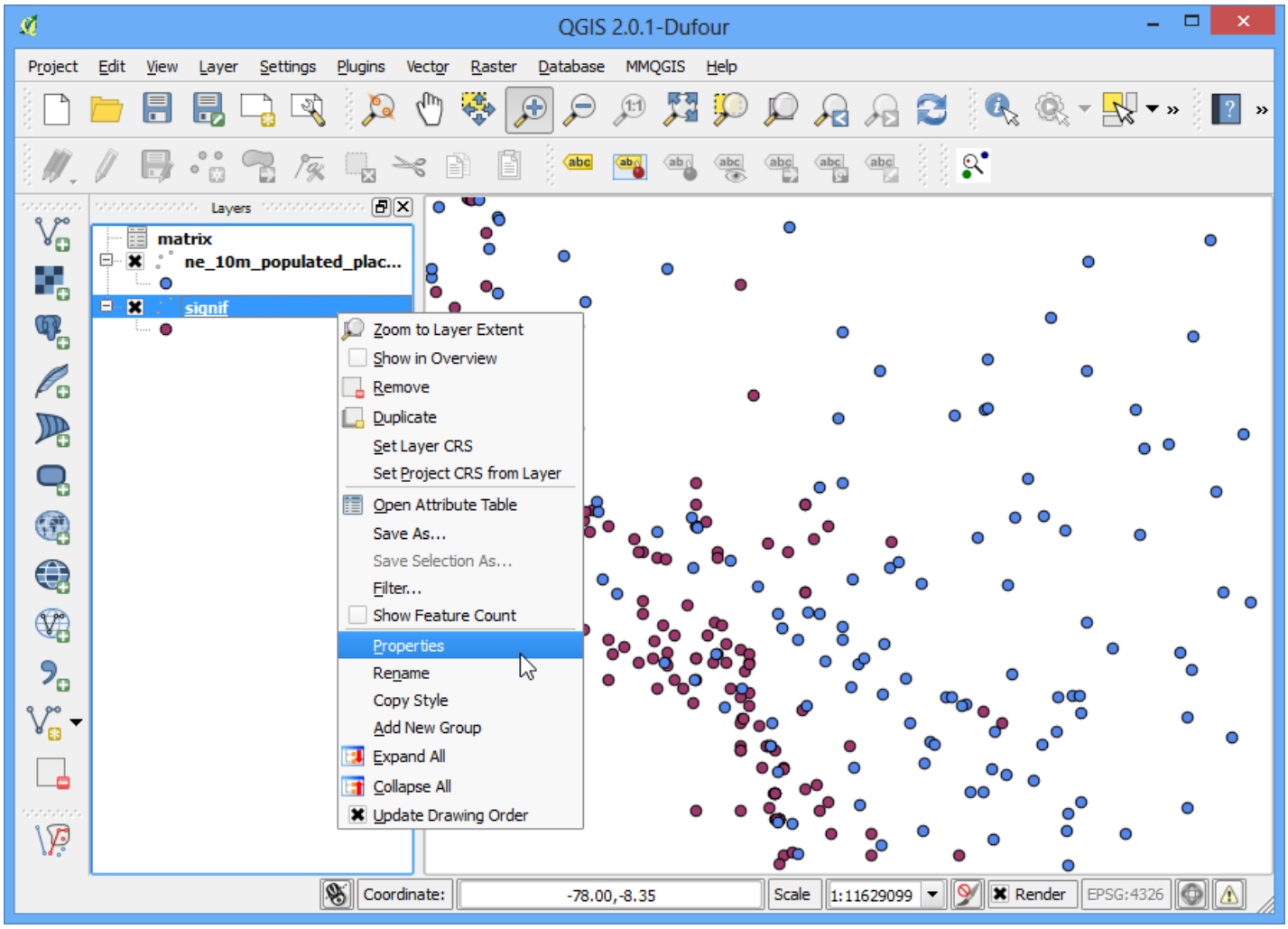
Note

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.

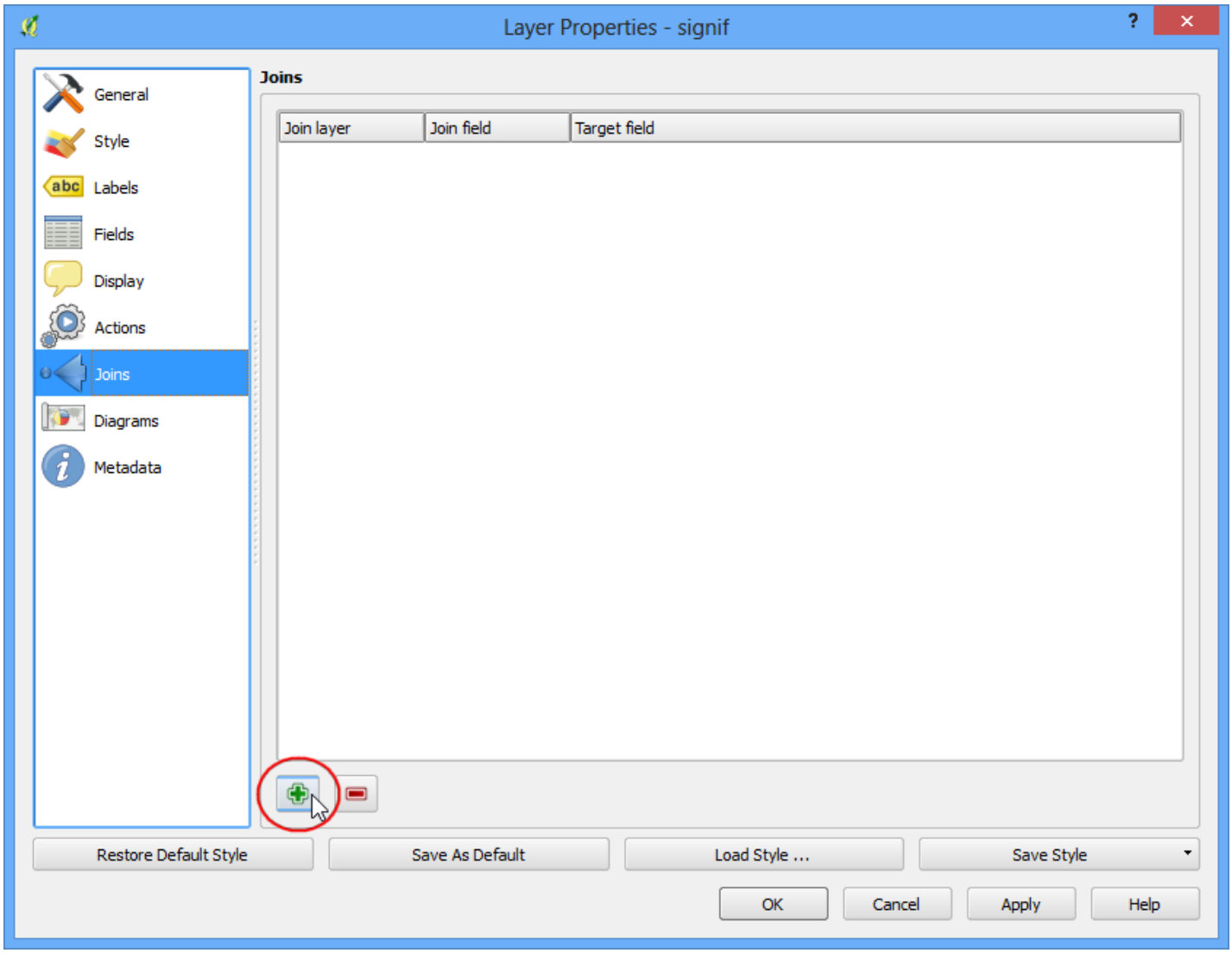


	InputID	TargetID	Distance
0	1	Al Karak	0.221721171014
1	2	Al Ladhiqiyah	0.144408036939
2	3	Buzmeyin	0.0526324624814
3	8	Al Khalil	0.208418004566
4	11	Iraklio	0.408843567409
5	5877	Iraklio	1.1082549107
6	9712	Al Ladhiqiyah	0.144408036939
7	12	As Salt	0.230569794451
8	13	Al Aqabah	0.10661139997
9	14	Al Qunaytirah	0.34713470868
10	7793	Nabatiye et Tahta	0.256395311798
11	16	Sparti	0.101878534504
12	7794	Saida	0.00326167893321
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
19	21	Iraklio	0.350232618378
20	22	Kavala	1.1152439462
21	9652	Rajkot	0.717056768568

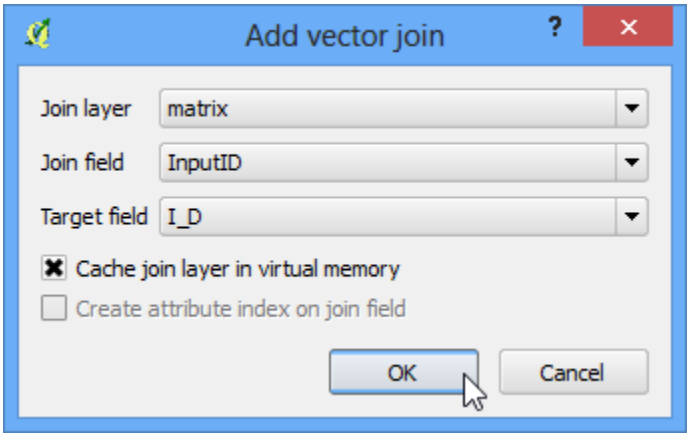
13. This is very close to the result we were looking for. For some users, this table would be sufficient. However, we can also integrate this results in our original Earthquake layer using a Table Join. Right-click on the Earthquake layer, and select Properties.



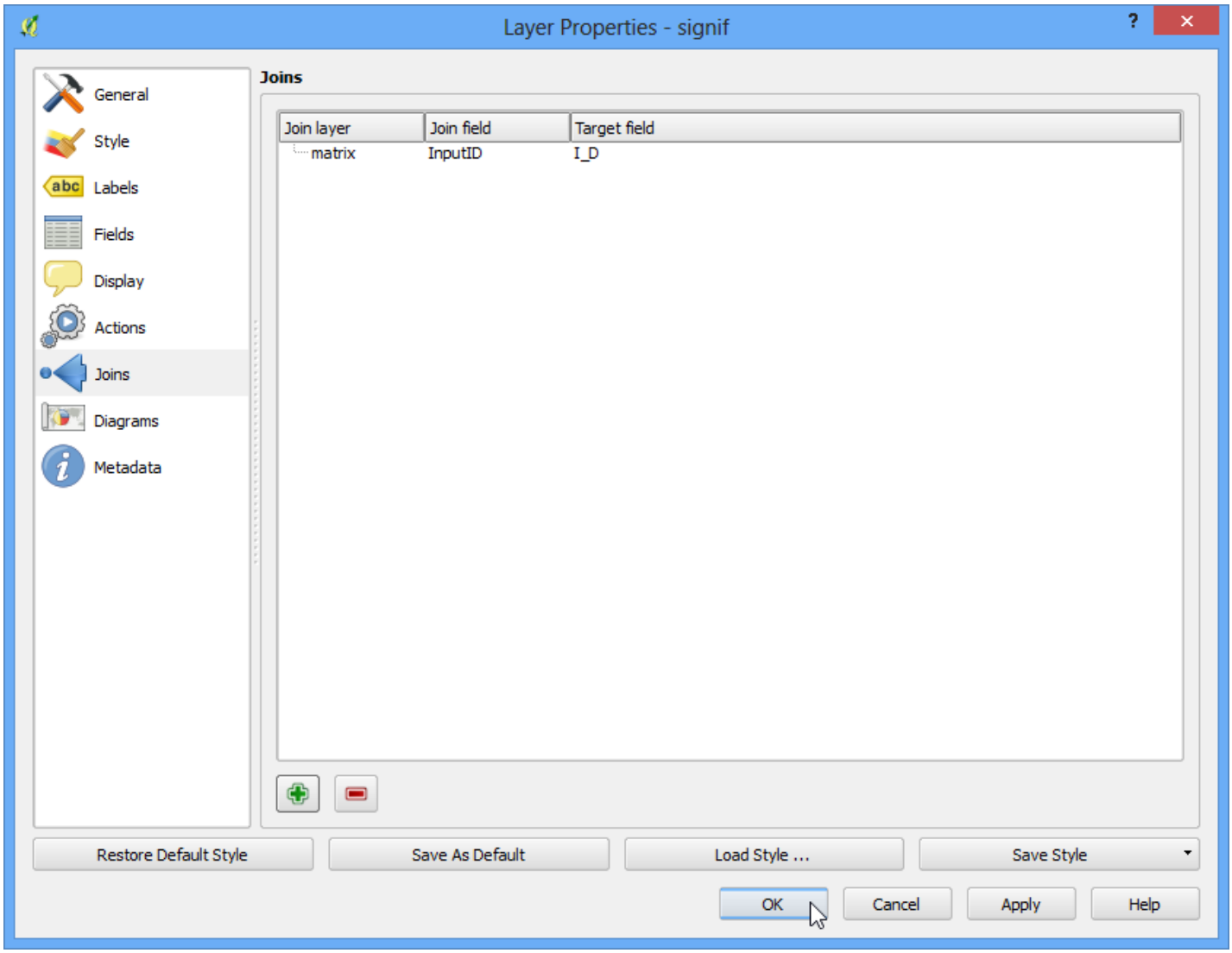
14. Go to the Joins tab and click on the + button.



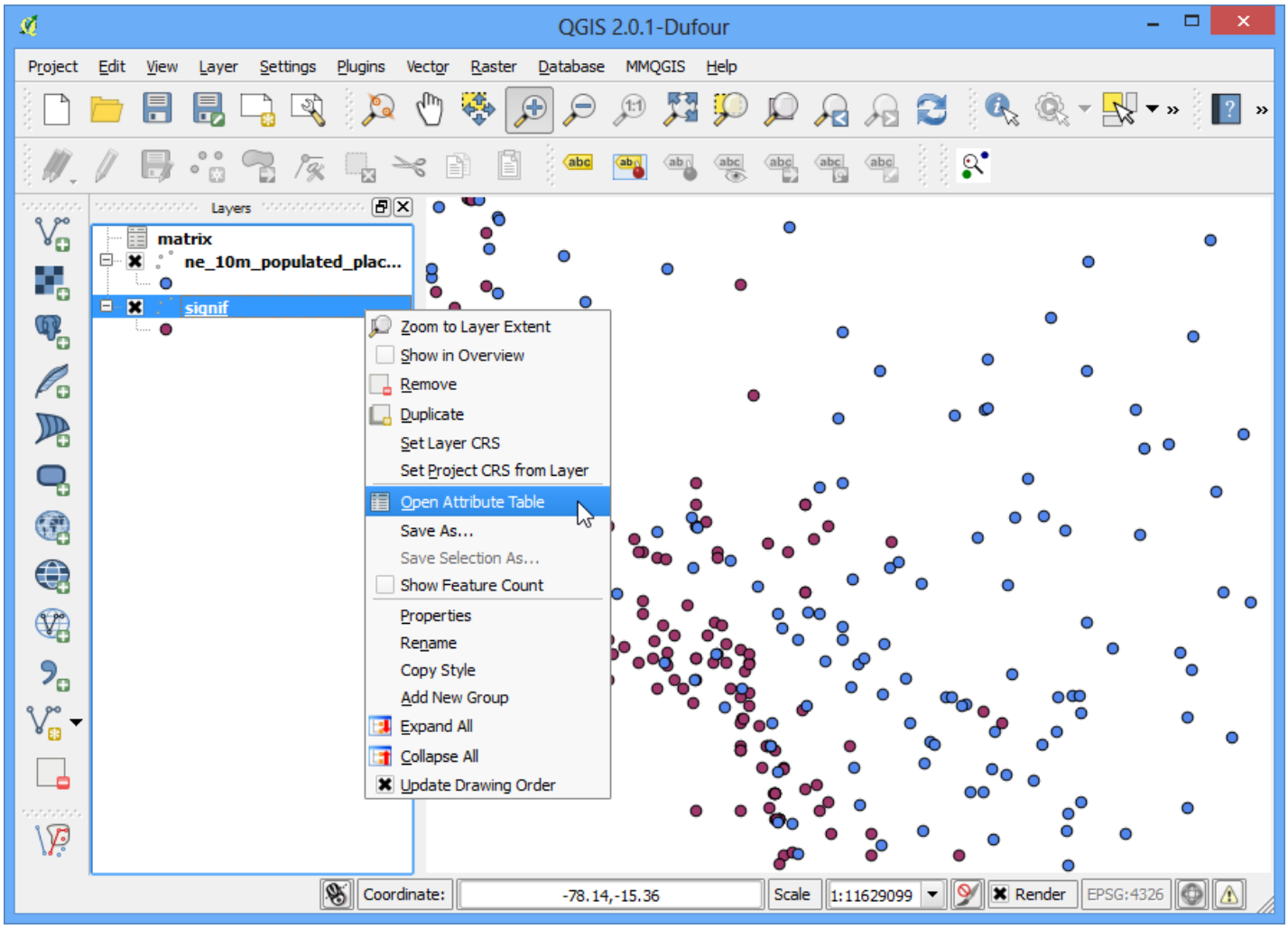
15. We want to join the data from our analysis result (***matrix.csv***) to this layer. We need to select a field from each of the layers that has the same values. Select the fields as shown below.



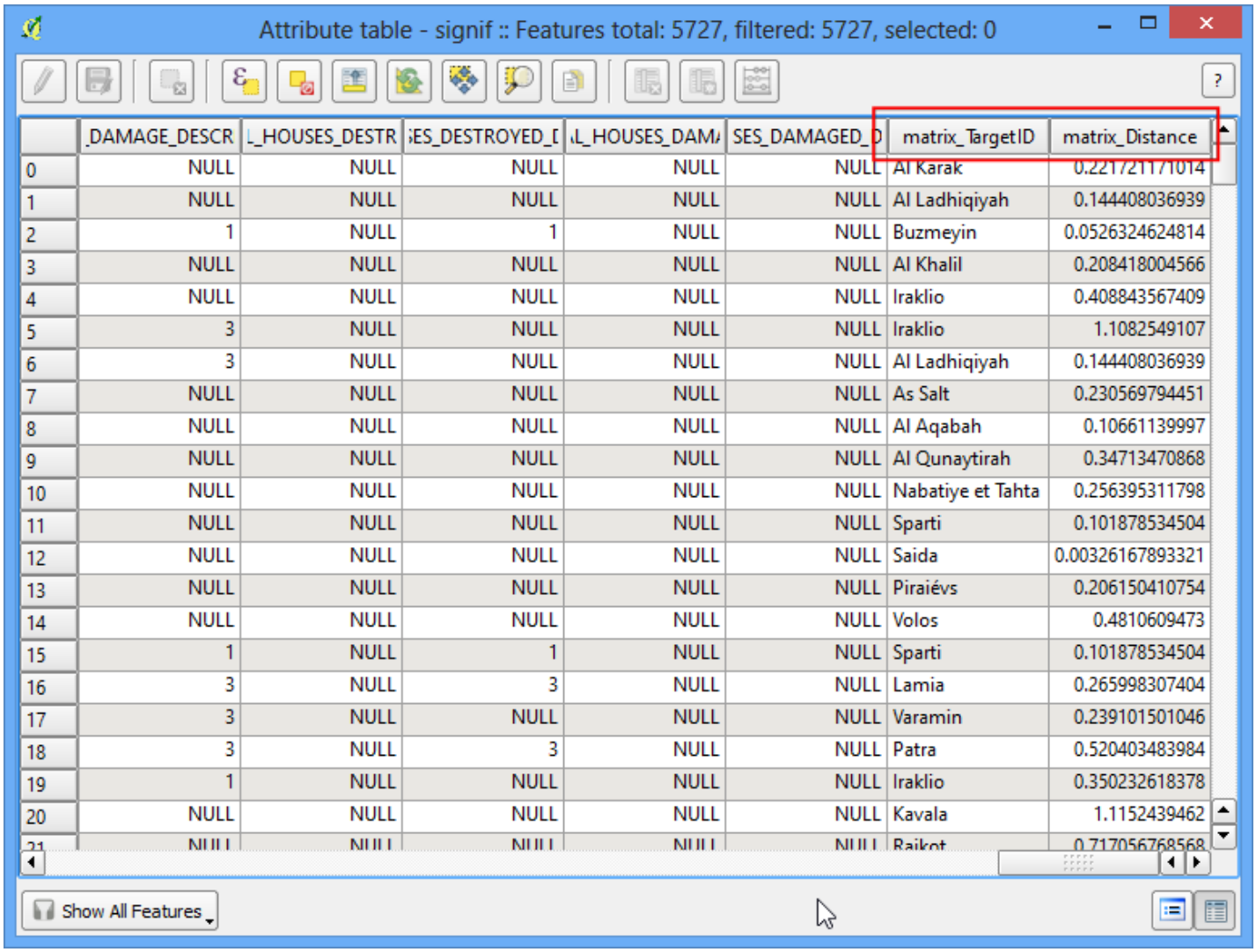
16. You will see the join appear in the Joins tab. Click OK.



17. Now open the attribute table of the Earthquakes layer by right-clicking and selecting Open Attribute Table.



18. 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.



	_DAMAGE_DESCR	L_HOUSES_DESTR	ES_DESTROYED_[	L_HOUSES_DAM/	SES_DAMAGED_D	matrix_TargetID	matrix_Distance
0	NULL	NULL	NULL	NULL	NULL	Al Karak	0.221721171014
1	NULL	NULL	NULL	NULL	NULL	Al Ladhiqiyah	0.144408036939
2	1	NULL	1	NULL	NULL	Buzmeyin	0.0526324624814
3	NULL	NULL	NULL	NULL	NULL	Al Khalil	0.208418004566
4	NULL	NULL	NULL	NULL	NULL	Iraklio	0.408843567409
5	3	NULL	NULL	NULL	NULL	Iraklio	1.1082549107
6	3	NULL	NULL	NULL	NULL	Al Ladhiqiyah	0.144408036939
7	NULL	NULL	NULL	NULL	NULL	As Salt	0.230569794451
8	NULL	NULL	NULL	NULL	NULL	Al Aqabah	0.10661139997
9	NULL	NULL	NULL	NULL	NULL	Al Qunaytirah	0.34713470868
10	NULL	NULL	NULL	NULL	NULL	Nabatiye et Tahta	0.256395311798
11	NULL	NULL	NULL	NULL	NULL	Sparti	0.101878534504
12	NULL	NULL	NULL	NULL	NULL	Saida	0.00326167893321
13	NULL	NULL	NULL	NULL	NULL	Piraiévs	0.206150410754
14	NULL	NULL	NULL	NULL	NULL	Volos	0.4810609473
15	1	NULL	1	NULL	NULL	Sparti	0.101878534504
16	3	NULL	3	NULL	NULL	Lamia	0.265998307404
17	3	NULL	NULL	NULL	NULL	Varamin	0.239101501046
18	3	NULL	3	NULL	NULL	Patra	0.520403483984
19	1	NULL	NULL	NULL	NULL	Iraklio	0.350232618378
20	NULL	NULL	NULL	NULL	NULL	Kavala	1.1152439462
21	NULL	NULL	NULL	NULL	NULL	Rajkot	0.717056768568