Nearest Neighbor Analysis

QGIS Tutorials and Tips

Ujaval Gandhi ujaval@qgistutorials.com

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

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.

ø	Create a Layer from a Delimited Text File ? ×												
File I	Name	C:/Use	rs/ujaval/Do	wnloads	s/signif.txt							Browse	
Laye	er name	e signi	f							Enco	oding UTF-8	-	
File	File format OCSV (comma separated values) OCustom delimiters ORegular expression delimiter										n delimiter		
			Comma						ice	Colon Semicolon			
			Other del	imiters		_	(Quote 📳		Escape	•		
Rec	ord opt	ions	Number of	fheader	lines to d	scard	0	🗧 🗙 Firs	t record ha	as field names			
Field	d optior	IS	Trim fie	elds	Discard	empty	fields	Decima	separator	is comma			
Geo	metry o	definitio	n 🖲 Point (coordina	tes		0 N	ell known	text (WKT)		geometry (attribu	te only table)	
			X field LC	ONGITU	DE	-	field L	ATITUDE	•	- 🗌 DMS coordi	nates		
Laye	er setti	ngs	X Use sp	atial ind	ex		l	Jse subset	index	W	atch file		
	I_D	FLAG	_TSUNAMI	YEAR	MONTH	DAY	HOUR	MINUTE	SECOND	FOCAL_DEPTH	EQ_MAG_MW	EQ_MAG	
1	1			-2150									
2	2	Tsu		-2000									
3	3			-2000						18		7.1	
4	8			-1566									
5	11			-1450								▲	
1		I			l	I	I	I	I	I	I		
										ОК	Cancel	Help	

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

	Coordinate Reference	System Selector	?	×
Specify CRS for layer	r signif			
Filter Recently used coo	ordinate reference systems	5		
Coordinate Refere	ence System	Authority ID	[•	•
North_Pole_Azimu OSGB 1936 / Britis Unknown datum	ıthal_Equidistant h National Grid based upon the Everest 183	EPSG:102016 EPSG:7405 EPSG:4044		
WGS 84 / Pseudo	Mercator	EPSG:3857		
WGS 84		EPSG:4326		-
•				_
Coordinate refere	nce systems of the world	Hide	deprecated CRS	s
0 0 0 0	ance System	Authority ID	ĺ.	
Coordinate Refere	ance system	Authonity iD		
WGS 84	ince system	EPSG:4326		
WGS 84	are system	EPSG:4326 IGNF:WGS72G	[
WGS 84 WGS 72 Wake Islan	d 1952	EPSG:4326 IGNF:WGS72G EPSG:4733		
Coordinate Refere WGS 84 WGS72 Wake Islan Wallis - Uv	d 1952 rea 1978 (MOP78)	EPSG:4326 IGNF:WGS72G EPSG:4733 IGNF:WALL78GEO		
Coordinate Refere WGS 84 WGS72 WGS72 Wake Islan Wallis - Uv World Geo	d 1952 rea 1978 (MOP78) ndetic System 1984	EPSG:4326 IGNF:WGS72G EPSG:4733 IGNF:WALL78GEO IGNF:WGS84G		
Coordinate Refere WGS 84 WGS72 Wake Islan Wallis - Uv World Geo Selected CRS: W	id 1952 rea 1978 (MOP78) idetic System 1984	EPSG:4326 IGNF:WGS72G EPSG:4733 IGNF:WALL78GEO IGNE:WGS84G		
Coordinate Refere WGS 84 WGS72 Wake Islan Wallis - Uv World Geo Selected CRS: We +proj=longlat +dat	id 1952 rea 1978 (MOP78) idetic System 1984 GS 84 um=WGS84 +no_defs	EPSG:4326 IGNF:WGS72G EPSG:4733 IGNF:WALL78GEO IGNE:WGS84G		

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.

Ø								C	GIS	2.0.1	-Dufc	our								- 5	×	
Project	Edit	View	Layer S	Settings	Plugins	Vector	Raster	Data	abase	MMQ	GIS	Help										
			New Embe	ed Layers	and Group)S		•	\bigcirc	1:1		Ç	\square	\mathcal{A}	R	8		Q.	-	▼ »	2	**
÷	/	₿	Add F	Vector Lay Raster La	yer	6	Ctrl+Shift+ Ctrl+Shift+	R	abc	aby	abr	(abc ()	abc	abc	abc		Q					
V	- ,	K / S	Add 9	SpatiaLite MSSQL Sp	Layer batial Layer		Ctrl+Shift+ Ctrl+Shift+	L M														
G			Add (Oracle Sp WMS/WM WCS Lave	atial Layer. TS Layer er		Ctrl+Shift+ Ctrl+Shift+	o w														
			Add V	WFS Laye Delimited	r Text Layer				20		æ	•	đ	1			•	22		2020	S S	
Q			Copy	style style	Tabla																•	
			/ Toggl	le Editing Layer Ed	its				•					~ Q P	8		•	~			1	
₩.			// Curre Save	As	an Vector	File		•	•		Do o		\$		•	•	•			990	and the second	
"₀ V° -			Remo	ove Layer cate Laye	(s) r(s)	1	Ctrl+D							•						٠		
			Set C Set P	CRS of Lay Project CR	/er(s) S from Lay	/er	Ctrl+Shift+	с														
\ X			Quer Mabel	y ling												51/2						
			Add t	to Overvie	ew				172.9	,146.4			Scale	263	339666	8 ▼	9 🗶 F	lender	EPSG:	4326		

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.

Ø	Select layers	to add	?	×
Layer ID 0 1	Layer name ne_10m_populated_places_simple.vERSION.txt	Type Vector Vector		
		OK Select All	Can	cel

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.

2 Distance matrix	?	×
Input point layer		
signif		-
Input unique ID field		
I_D		-
Target point layer		
ne_10m_populated_places_simple		-
Target unique ID field		
name		-
Output matrix type		
Linear (N*k x 3) distance matrix		
O Standard (N x T) distance matrix		
 Summary distance matrix (mean, std. dev., min, max) 		
Use only the nearest (k) target points		Ð
		00050
		owse
0% OK	Clos	se

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.

ø			Create a	Layer from a Delimited Text File	? ×						
File I	Name C:/	Jsers/ujaval/Do	ownloads/matrix.csv		Browse						
Laye	er name 🛛 m	atrix			Encoding UTF-8						
File f	format	CSV (comma separated va	alues) 🔿 Custom delimiters	O Regular expression delimiter						
Perr	Decend actions										
Field	Record options Number of header lines to discard 0 🚽 🕱 First record has field names										
Field	options		elos Discard en	npty fields Decimal separator is comma							
Geor	metry defini	ition 🔘 Point	coordinates	 Well known text (WKT) 	 No geometry (attribute only table) 						
Laye	er settings	Use s	oatial index	Use subset index	Watch file						
	InputID	TargetID	Distance								
1	1	Al Karak	0.221721171014								
2	2	Al Ladhiqiyah	0.144408036939								
3	3	Buzmeyin	0.0526324624814								
4	8	Al Khalil	0.208418004566								
5	11	Iraklio	0.408843567409								
6	5877	Iraklio	1.1082549107		▲						
	OK Cancel Help										
					~~						

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.

ø		Attribute table	e - matrix :: Feat	ures total: 5727, filtered: 5727, selected: 0		×
		E 😼 🔳	🗟 😻 🞾 [?
	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	2		
19	21	Iraklio	0.350232618378	~5		
20	22	Kavala	1.1152439462			
21	9652	Rajkot	0.717056768568			÷.
		T 1 ¹ 1	0.51007152534		_	
SI SI	now All Features					

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.

M.		Layer	Properties - sig	gnif			?	×
General	Joins							_
😻 Style	Join layer	Join field	Target field					
abc Labels								
Fields								
🥟 Display								
Actions								
0 Joins								
Diagrams								
👔 Metadata								
Restore Default Style	e	Save As Default		Load Style		Save Style		•
				ОК	Cancel	Apply	Help	

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.

×.	Add vector join ? ×								
Join layer	matrix								
Join field	InputID								
Target field	Target field I_D								
🗶 Cache jo	in layer in virtual memory								
Create a	ttribute index on join field								
	OK Cancel								

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

Я.		Layer	Properties - signif	? ×
General	Joins			
Style	Join layer	Join field	Target field	
	· matrix	InputID	I_D	
Labels				
Fields				
Display				
Actions				
• Joins				
Diagrams				
🕜 Metadata				
	(]		
Restore Default Shul		Save As Default	Load Style	Sava Stala
Restore Default Style		Save As Default		Save Style
			OK Cancel	Apply Help

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.

ø		Attribute table	e - signif :: Featu	ures total: 5727,	filtered: 5727,	selected: 0		×
	8	è 🖪 🖪 🛙	🗟 😽 🞾 🛛					?
	DAMAGE_DESCR	L_HOUSES_DESTR	ES_DESTROYED_E	L_HOUSES_DAM	SES_DAMAGED_D	matrix_TargetID	matrix_Distance	Þ
0	NULL	NULL	NULL	NULL	NUL	Al Karak	0.221721171014	Ŧ
1	NULL	NULL	NULL	NULL	NULL	Al Ladhiqiyah	0.14440803693	9
2	1	NULL	1	NULL	NULL	Buzmeyin	0.0526324624814	4
3	NULL	NULL	NULL	NULL	NULL	Al Khalil	0.20841800456	5
4	NULL	NULL	NULL	NULL	NULL	Iraklio	0.40884356740	9
5	3	NULL	NULL	NULL	NULL	Iraklio	1.108254910	7
6	3	NULL	NULL	NULL	NULL	Al Ladhiqiyah	0.14440803693	3
7	NULL	NULL	NULL	NULL	NULL	As Salt	0.23056979445	1
8	NULL	NULL	NULL	NULL	NULL	Al Aqabah	0.1066113999	7
9	NULL	NULL	NULL	NULL	NULL	Al Qunaytirah	0.3471347086	3
10	NULL	NULL	NULL	NULL	NULL	Nabatiye et Tahta	0.25639531179	3
11	NULL	NULL	NULL	NULL	NULL	Sparti	0.101878534504	4
12	NULL	NULL	NULL	NULL	NULL	Saida	0.0032616789332	1
13	NULL	NULL	NULL	NULL	NULL	Piraiévs	0.206150410754	4
14	NULL	NULL	NULL	NULL	NULL	Volos	0.481060947	3
15	1	NULL	1	NULL	NULL	Sparti	0.101878534504	4
16	3	NULL	3	NULL	NULL	Lamia	0.265998307404	4
17	3	NULL	NULL	NULL	NULL	Varamin	0.23910150104	5
18	3	NULL	3	NULL	NULL	Patra	0.520403483984	4
19	1	NULL	NULL	NULL	NULL	Iraklio	0.350232618378	3
20	NULL	NULL	NULL	NULL	NULL	Kavala	1.1152439462	2
21 ◀	NULL	NHH	NULL	NULL	NHH	Raikot	0 71705676856	
Sł	now All Features				6			