Mapping 2: Manipulating geographical data with QGIS

Hands-on at NICAR 2014, Baltimore, Mar 1

Peter Aldhous

peter@peteraldhous.com

apaldhous

In addition to displaying geographic data, QGIS is a powerful tool for processing data for use by other mapping applications. If you want to make online maps, using tools including <u>TileMill</u> and <u>Leaflet</u>, for instance, QGIS can help get your data into the right shape.

In this class we'll explore some useful QGIS data processing functions, including joining tables of data to existing shapefiles, simplifying geographic features so that online maps will render quickly, and how to save data in formats commonly used for web mapping. We'll also learn how to edit geographic data, and how to process data consisting of hundreds or thousands of closely spaced or overlapping points to give a more meaningful summary display.

Joining data to a shapefile

Launch QGIS Desktop, and import this shapefile:



(During this class, we're not going to set a specific projection, and will work throughout with the WGS84 lat-long default. So whenever you're prompted, accept this option.)

The imported shapefile contains the boundary data for the world's freshwater ecoregions, to which were going to join the data on threatened amphibians that we mapped in the previous class. If you open the attribute table, you'll see that it lacks the **THREAT_AMP** column that we used to create the thematic map.

The data is in <u>this file</u>, which I've saved in **DBF** format. You can create DBFs from any spreadsheet – I'd recommend using **Calc**, the free spreadsheet program avilable in <u>LibreOffice</u>. Simply open your spreadsheet in Calc, and select **File>Save As** File Type **dBASE (.dbf)**.

Here's the data, which has two columns, **THREAT_AMP** and **ECO_ID** (you can ignore the letter and numbers after the first comma, which are a description of the data):

	A	В	С	D
1	ECO ID.N.6.2	HREAT AMP,N,5,2		
2	312.00	95.00		
3	301.00	94.00		
4	302.00	68.00		
5	765.00	43.00		
6	715.00	37.00		
7	518.00	37.00		
8	201.00	35.00		
9	203.00	33.00		
10	305.00	33.00		
11	519.00	33.00		
12	202.00	32.00		
13	171.00	29.00		
14	304.00	29.00		
15	807.00	29.00		
16	505.00	28.00		
17	173.00	26.00		
18	206.00	26.00		
19	205.00	25.00		
20	581.00	25.00		
21	204.00	23.00		
22	174.00	23.00		
23	763.00	23.00		
24	165.00	22.00		

The shapefile contains the column **ECO_ID**, too, so we can use this field to join the threatened amphibian data to each ecoregion. If you've used relational databases, this is conceptually the same as an "inner join" query, which pulls together data from different tables where they match for a selected field.

First we need to add the DBF to the project, using Layer>Add vector layer. Notice how it appears in the Layers panel with an attribute table icon:



Right click on the shapefile in the **Layers** panel, select **Properties>Joins**, click **(*)** and fill in the dialog box as follows:

🔏 QGIS	2.0.1-	Dufour					(Lay	ver Properties - FEOW	/v1_1	NC													? X		23
Project	Edit	View	Layer	Se	tings P	lug	-		_																
					2			General		oins															
. 11	1				Par		2	🖌 Style		Join la	yer	Join	field	Target	field										
<u>ः १८ -</u> े भाष्य	00.000	00m	No.	and and a	Na	2	at	Labels																	
			LO		Lia	8		Fields																	
Vo		thre	atene	d am	phibians		9	Display																	
			EOWv	1 TN	<u>[</u>		8	Actions				6	Add vecto	or join		8	23	η							
Q.							•	Joins					Join layer		threaten	ned amphib	ians 🔻								NY N
Po								Diagrams					Join field		ECO_ID		-							Ĩ,	• ~
			Sho	rtest n	eh 1000		1	Metadata					Target field		ECO_ID		•								
	Star	t	210										Create a	in layer in ttribute in	virtual memo	ory field									
														ſ	OK	Ca	ancel								
4	Stop											L						IJ							4
10	Crite	rion			Length																				
	Leng	ath 🗌			congut																			N	4
	Time																								3
C C	C	alculate		Exp	ort																			1	1
3 20					ielp																				
V V																								1	
•							L			w															
-()-								Restore Defaul	t Style	2			Save As D	Default			Lo	ad Style .			Save S	Style	•		
¥ Toggles t	he edit	ing stat	of the	curre	nt layer													(ОК	Cancel	Appl	y	Help	43	26

Click Apply and OK to complete the join.

Open the shapefile's attribute table, and see that it contains the joined data on threatened amphibians in a new column at the right:

🐔 At	ttribute table - FEO	Wv1_TNC :: Featur	es total: 449, filtere	d: 449, selected: 0				3
		E	🚳 🐝 🞾				(?
	ECO_ID 🗸	ECOREGION	MHT_TXT	MHT_NO	OLD_ID	ECO_ID_U	threatened amphibians_THREAT_AMP	
0	103	Alaska & Cana	temperate coas	5	1.000000	30103	2.00	
1	120	Columbia Glaci	temperate upla	6	2.000000	30120	0.00	
2	121	Columbia Ungl	temperate floo	7	3.000000	30121	1.00	
3	122	Upper Snake	temperate upla	6	4.000000	30122	0.00	
4	123	Oregon & Nort	temperate coas	5	5.000000	30123	3.00	
5	125	Sacramento - S	temperate coas	5	6.000000	30125	7.00	
6	159	Southern Califo	xeric freshwater	4	7.000000	30159	3.00	
7	127	Bonneville	xeric freshwater	4	8.00000	30127	1.00	
8	126	Lahontan	xeric freshwater	4	9.000000	30126	2.00	
9	124	Oregon Lakes	xeric freshwater	4	10.000000	30124	1.00	
10	128	Death Valley	xeric freshwater	4	11.000000	30128	7.00	
11	130	Colorado	xeric freshwater	4	12.000000	30130	4.00	
12	129	Vegas - Virgin	xeric freshwater	4	13.000000	30129	1.00	
13	131	Gila	xeric freshwater	4	14.000000	30131	3.00	
14	132	Upper Rio Gran	temperate upla	6	15.000000	30132	1.00	
15	161	Guzman - Sam	xeric freshwater	4	16.000000	30161	2.00	
16	134	Rio Conchos	xeric freshwater	4	17.000000	30134	2.00	
17	133	Pecos	xeric freshwater	4	18.000000	30133	0.00	
18	163	Mayran - Viesca	xeric freshwater	4	19.000000	30163	1.00	
19	135	Lower Rio Gran	temperate floo	7	20.000000	30135	1.00	
20	137	Rio Salado	xeric freshwater	4	21.000000	30137	0.00	
21	136	Cuatro Cienegas	xeric freshwater	4	22.000000	30136	0.00	
22	138	Rio San Juan (xeric freshwater	4	23.000000	30138	0.00	
22	148	Unner Mississinni	temperate floo	7	24.000000	30148	0.00	
	Show All Features							

Here are *further instructions* on joining data to shapefiles, including from CSV files.

Now is a good time to save the joined shapefile. Right click on it in the Layers panel, and Save as an ESRI Shapefile with a suitable name:

1	🕺 Save vector lay	er as	8 23
	Format	ESRI Shapefile	-
	Save as	hwater ecoregions/FW_amphibs.shp	Browse
	Encoding	UTF-8	•
	CD.C	Layer CRS	•
	CRS	WGS 84	Browse
	Symbology export	No symbology	•
	Scale	1:50000	
	OGR creation opt	ions	
	Data source		
	Layer		
	Skip attribute	e creation e to map	
		OK Cancel	Help

We can also use the same dialog box to export the data in formats commonly used for web mapping, such as <u>geoJSON</u> or <u>KML</u>. But there's one problem: The boundary data is highly detailed, which will make the exported data file huge and sluggish to load.

Simplifying the geometry of your data

To solve that problem, select **Vector>Geometry Tools>Simplify geometries**, fill in the dialog box as follows, and click **OK**:

Simplify geometries	2 S 2
Input line or polygon layer	
FEOWv1_TNC	•
Use only selected features	5
Simplify tolerance	0.1000
Save to new file	Browse
Add result to canvas	
	0%
	OK Close

In practice, you will probably want to check the **Save to new file** option, and experiment with different numbers for **Simplify tolerance**, to get a visually acceptable result that will export as a file of reasonable size. (Your web developers may be able to advise on that.)

You should see a result like this, which will greatly reduce the size of an exported geoJSON or KML file:



We're finished with this data, so select **Project>New** and **Discard** the current project to start again with a fresh screen.

Editing geographical data

Now we're going to explore some of QGIS's data editing functions. First import <u>this</u> <u>shapefile</u> of the countries of the world by selecting **Vector>Add Vector Layer**:

🥖 Open an OGR Suppor	ted Vec	tor Layer		
QGIS ▼ S CGIS	► Dat	ta 🕨 world countries 🔍 🍫	Search world countries	Q
Organize 🔻 New	folder			
🖳 Recent Places	*	Name	Date modified	Туре
E 1.1		ne_50m_admin_0_countries.dbf	11/5/2012 12:41 AM	DBF File
Cibraries		ne_50m_admin_0_countries	11/5/2012 12:41 AM	PRJ File
		ne_50m_admin_0_countries.README	11/7/2012 11:15 PM	Firefox H1
Pictures		ne_50m_admin_0_countries	11/5/2012 12:41 AM	SHP File
Videos		ne_50m_admin_0_countries.shx	11/5/2012 12:41 AM	SHX File
La viacos	=	ne_50m_admin_0_countries.VERSION	11/7/2012 11:15 PM	Text Docu
🤣 Homegroup				
💻 Computer				
🚢 Local Disk (C:)				
🚽 Home on 'psf' (\	(
🖵 Host on 'psf' (Z:)	+ -			Þ
F	ile nam	e: ne_50m_admin_0_countries 🔹 🖌	Il files (*)	•
			Open Can	cel

Next add <u>this CSV file</u> of points showing Ramsar Convention sites deemed important for amphibian conservation, from the previous class, using Layer>Add Delimited Text Layer:

File	Name Y:/Dropbox/NICAR/QGIS/Data/r	amsar amphibian site:	S.CSV			Browse
Lay	er name ramsar amphibian sites				Encoding UTF-8	•
File	format	ated values) 🛛 Cu	istom delimiters		Regular expression	delimiter
Rec Field Geo	ord options Number of header lines d options Trim fields Dis metry definition Point coordinates	; to discard 0 🔶 card empty fields 🗌 We	First recor Decimal separa	d has field name ator is comma /KT)	s) No geometry (attribut	e only table)
Lay	er settings	Y field Lat	e subset index	▼ DMS	Watch file	
	Site name	Designation date	Latitude	Longitude	Total site area (ha)	
1	Réserve Intégrale du Lac Tonga	11/4/1983	36.88333333	8.516666667	2700	
	Marais de la Macta	2/2/2001	35.68333333	-0.166666667	44500	
2	Oasis de Tamantit et Sid Ahmed Timmi	2/2/2001	27.75	0.25	95700	
2 3	Petit Loango	30-12-1986	-2.3	9.616666667	480000	
2 3 4		1	2 21000007	10.08333333	220000	
2 3 4 5	Setté Cama	30-12-1986	-2./1000000/	101000000000		
2 3 4 5 6	Setté Cama Ain Elshakika	30-12-1986 5/4/2000	32.76666667	21.35	33	

We're going to edit the points to make a shapefile containing only sites in the United States. The CSV file contains no information on country, but we can use the countries shapefile to select the points within the US.

With the countries shapefile highlighted in the Layers panel, click the Select Single Feature icon:



Then click anywhere within the borders of the US, and you should see a screen like this:



Now select **Vector>Geoprocessing Tools>Intersect** and fill in the dialog box as follows, selecting a suitable name for the output shapefile:

🚀 Intersect	8 22
Input vector layer	
ramsar amphibian sites	-
Use only selected features	
Intersect layer	
ne_50m_admin_0_countries	-
X Use only selected features	
Output shapefile	
Y:/Dropbox/NICAR/QGIS/Data/US ramsar.shp	Browse
0% OK	Close
Output shapefile Y:/Dropbox/NICAR/QGIS/Data/US ramsar.shp OK	Browse

Click **OK**, then accept the option to add it to map. Hide the full Ramsar sites layer by unchecking it, and you should see a screen like this, showing our new shapefile:



The **Vector** menu in QGIS offers a rich set of tools for manipulating geographical data. Our **Intersect** option created a new shapefile with fields from both shapefiles in its attribute table. **Clip** is similar, but would not have added fields from the countries shapefile to the new shapefile's attribute table. **Buffer** creates a new shapefile bounding a zone a defined distance from a set features of interest – you might use this, for instance, to see which parts of your city lie within a mile of a park, or a library. Find out more about the options from the <u>QGIS manual</u>, and explore their functions for yourself.

We can also edit data in QGIS directly. With the countries shapefile selected in the layers panel, click the **Toggle Editing** icon:

Then open the attribute table for the countries shapefile and click the **Invert Selection** icon:

This will select all countries apart from the US, which we can then delete from the shapefile by clicking the **Delete Selected Features** icon:

Click **OK**, close the attribute table, then click the **Toggle Editing** icon once more to exit editing mode. Save your changes, and you should now have a map showing just the US.

This is just a brief intoduction to QGIS's data editing functions – see <u>the manual</u> for more possibilities.

We're now finished with this data, so select **Project>New** and **Discard** the current project to start again with a fresh screen.

Hexagonal binning: a better option than throwing many points at a map

In the final part of this class, we'll explore how to make a sensible map display from data that consists of hundreds or even thousands of closely spaced or overlaid points. I used this option to make <u>these maps</u> of the Syrian civil war.

Open <u>this shapefile</u>, showing the location of distinct violent events in Syria from the first quarter of 2013:



If you open the shapefile's atribute table, you'll see that it contains more than 10,000 points, many of which overlay one another. We're going to create a hexagonal grid covering the same area, and then count the points in each cell in the grid.

One of the strengths of QGIS is that it has an active community of open-source software developers producing useful plugins for specific geodata processing tasks. To make our grid, we need to install one of these plugins, called **MMQGIS**.

Select **Plugins>Manage and Install Plugins,** then select **Get More** and type **mmqgis** into the **Search** box. Select the result that appears, then click **Install Plugin**. An **MMQGIS** menu should now have appeared.

Next, use the **Zoom In** and **Zoom Out** tools to give some blank space around the points:





This will ensure that the grid layer will completely cover all the points.

To create the grid layer select **MMQGIS>Create>Create Grid Layer** and fill in the dialog box as follows, selecting an appropriate name for the output shapefile:

Y:/Dropbox/NICAR/QGIS/Da	ta/grid.shp Browse
Center X	▼ 39
Width	H Spacing
12.5676121385	0.15
Middle Y	▼ 35
Height	V Spacing
6.88257711881	0.173205161514
Havagan	(nelveen)

In practice, you'll need to experiment with the **H Spacing** value to produce hexagons of a suitable size. The **V spacing** value will adjust automatically.

Once the shapefile has loaded, drag the points shapefile above the grid shapefile in the **Layers** panel, and you should see a screen like this, confirming that all our points are in the grid:



Now we'll make a new shapefile with a count of the points in each cell of the grid. Select **Vector>Analysis Tools>Points in polygon** and fill in the dialog box as follows, selecting an appropriate name for the new shapefile:

Input polygon vector layer	
grid.shp	
Input point vector layer	
2013_Q1	
Output Shapefile	
Output Shapefile Y:/Dropbox/NICAR/QGIS/Data/syria_grid.sh	np Brov

Click **OK**, and once the data has processed, accept the option to add the new shapefile to the map. Open its attribute table, and you'll see that there is a column **PNTCNT**, giving the number of violent events in each grid cell. You can now save this new shapefile in whatever format you need for your mapping application of choice.

Next steps

For more on hexagonal binning, and pointers on styling the resulting shapefile in TileMill to make an online map, see <u>this tutorial</u>. For instructions on how to make web maps with Leaflet using data exported as geoJSON, see <u>here</u> and <u>here</u>.