Geographic Information for Vector Surveillance

Week 1 Day 2

Creating spatial data from entomological surveys

Learning objectives

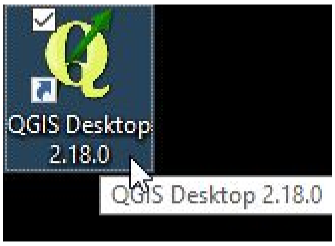
- able to add existing vector data (points, lines, polygons) to QGIS
- able to create spatial data from entomological surveys by :
    - transferring household-level data from paper forms to an Excel spreadsheet
    - creating block-level summaries in Excel using pivot tables
    - calculating block-level larvae/pupae indices in Excel
    - adding point-level data from an Excel spreadsheet into QGIS
    - joining data from an Excel spreadsheet to a shapefile in QGIS

Length : 1 day
Updated : September 2018

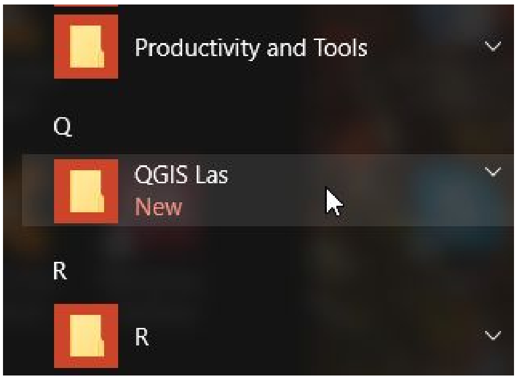
2.1. Adding vector data (points, lines, polygons) into QGIS

In Day 1, we explored the ready-made QGIS project file, and learnt how to visualise point, line, polygon data which has been saved in the format of a shapefile, plus raster data which has been saved in the format of a geoTiff. Today we will learn how to create your own QGIS project. We will start by learning how to add vector data which has been saved in the format of a shapefile into QGIS.

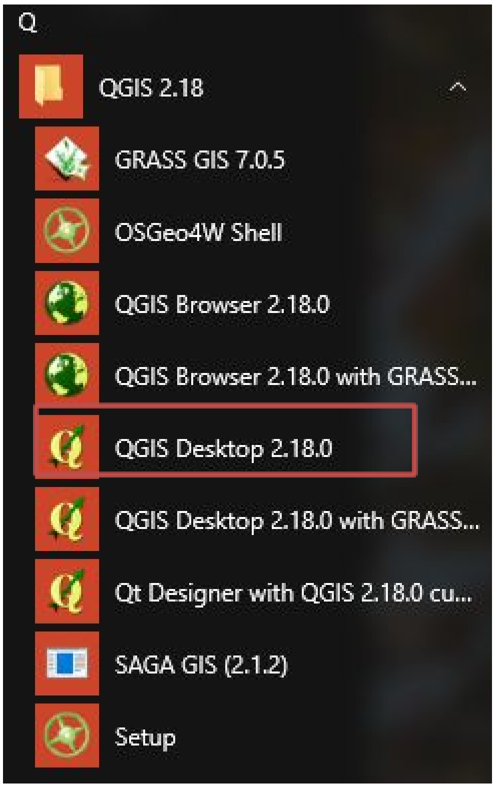
➔ Start QGIS by double clicking on the shortcut.



➔ OR go to the start menu from the windows icon at the lower left and select the QGIS folder, it may say Las or 2.18.

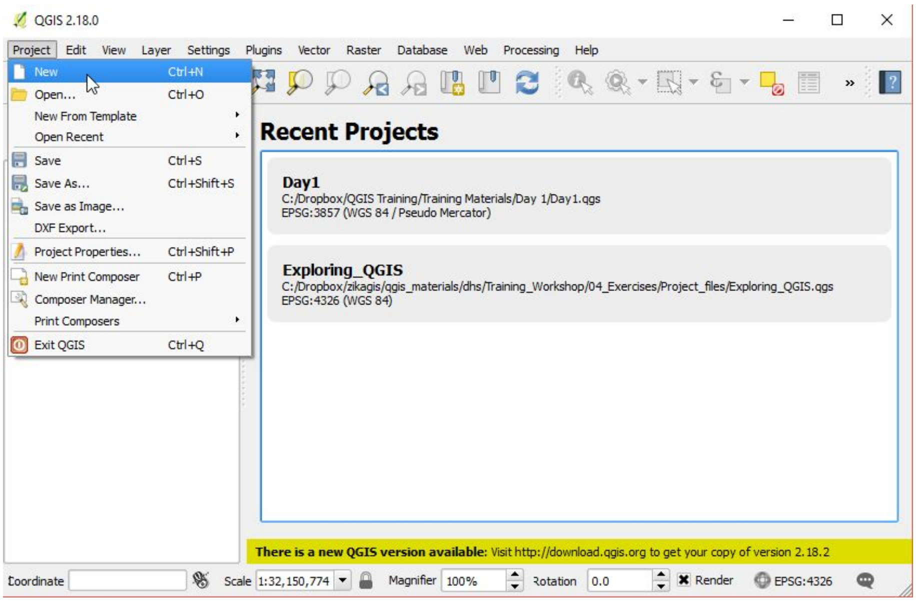


- ➔ Then select QGIS Desktop 2.18.0.


Q
QGIS 2.18
GRASS GIS 7.0.5
OSGeo4W Shell
QGIS Browser 2.18.0
QGIS Browser 2.18.0 with GRASS...
QGIS Desktop 2.18.0
QGIS Desktop 2.18.0 with GRASS...
Qt Designer with QGIS 2.18.0 cu...
SAGA GIS (2.1.2)
Setup

When QGIS starts it will usually give the option to open recent projects. This time, we will create a New, empty project.

- ➔ Left click Project, New



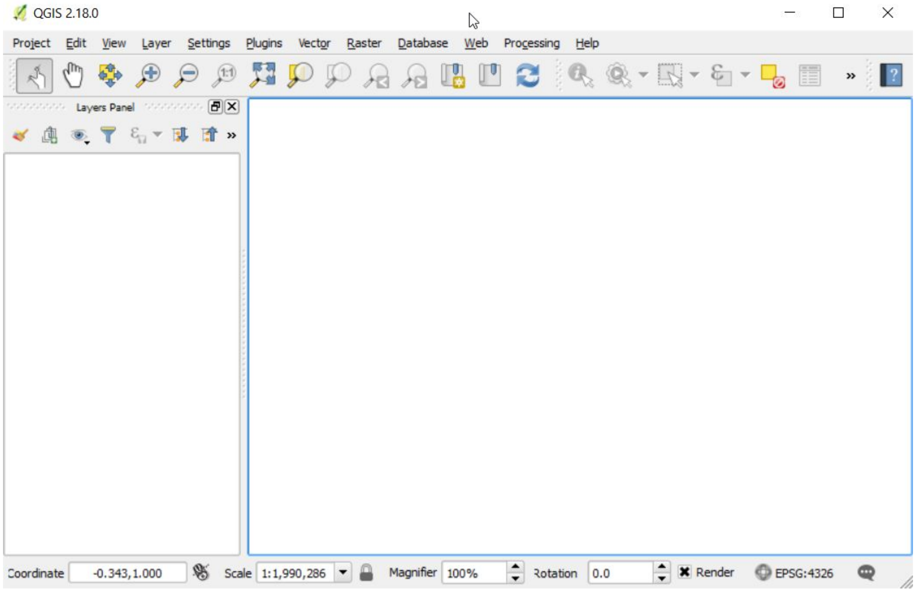
Day1
C:/Dropbox/QGIS Training/Training Materials/Day 1/Day1.qgs
EPSG:3857 (WGS 84 / Pseudo Mercator)

Exploring_QGIS
C:/Dropbox/zikagis/qgis_materials/dhs/Training_Workshop/04_Exercises/Project_files/Exploring_QGIS.qgs
EPSG:4326 (WGS 84)

There is a new QGIS version available: Visit http://download.qgis.org to get your copy of version 2.18.2

Scale 1:32,150,774 Magnifier 100% Rotation 0.0 EPSG:4326

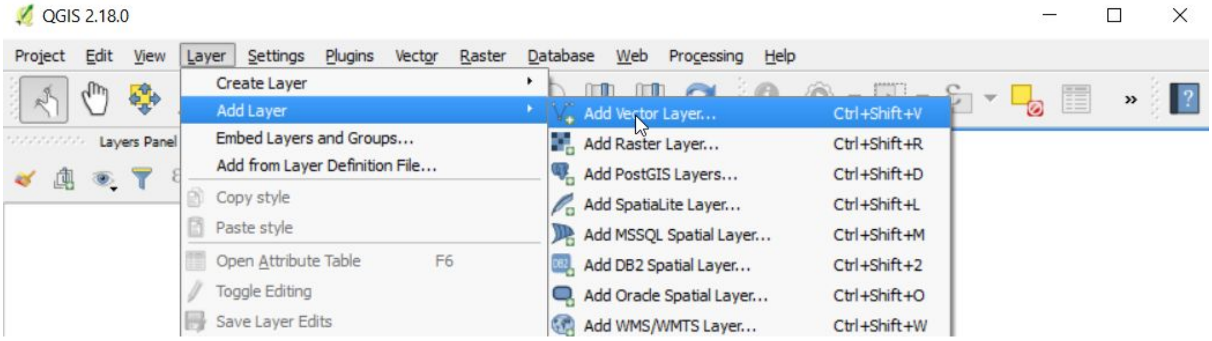
This should open up a blank project as shown below :



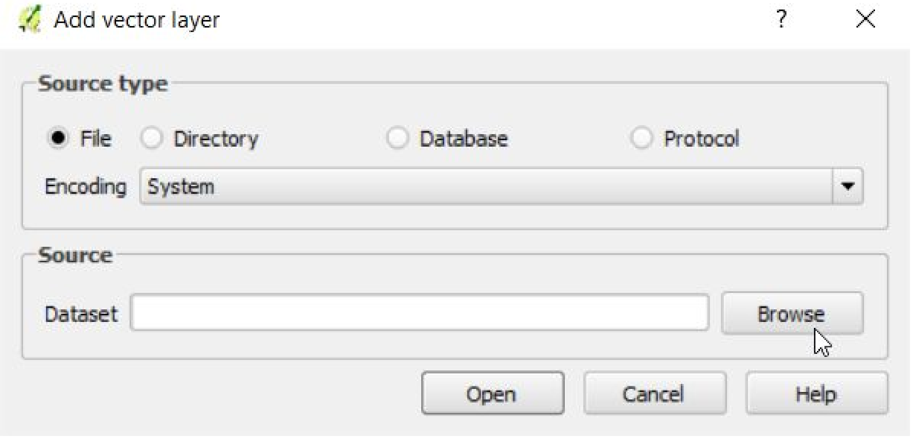
Coordinate -0.343,1.000 Scale 1:1,990,286 Magnifier 100% Rotation 0.0 EPSG:4326

Now we will add some vector layers to the project.

- ➔ Left click : Layer, Add Layer, Add Vector Layer


QGIS 2.18.0
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Layers Panel
Create Layer
Add Layer
Embed Layers and Groups…
Add from Layer Definition File…
Copy style
Paste style
Open Attribute Table F6
Toggle Editing
Save Layer Edits
Add Vector Layer… Ctrl+Shift+V
Add Raster Layer… Ctrl+Shift+R
Add PostGIS Layers… Ctrl+Shift+D
Add SpatiaLite Layer… Ctrl+Shift+L
Add MSSQL Spatial Layer… Ctrl+Shift+M
Add DB2 Spatial Layer… Ctrl+Shift+2
Add Oracle Spatial Layer… Ctrl+Shift+O
Add WMS/WMTS Layer… Ctrl+Shift+W

- ➔ click 'Browse' to find a file in your folders.



Add vector layer

Source type

● File ○ Directory ○ Database ○ Protocol

Encoding: System

Source

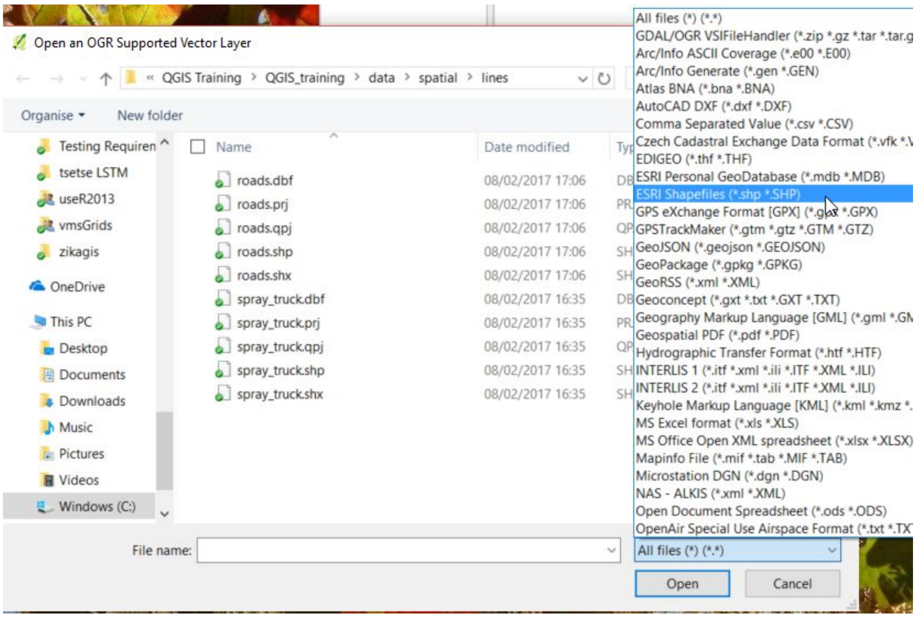
Dataset:

This should bring up a browse window. We will firstly add the road data that we looked at in yesterday's session into QGIS.

- ➔ navigate to the data folder for this training.
- ➔ select Spatial data, Lines (QGIS_training\data\spatial\lines).

Recall from yesterday's practical that shapefiles actually consist of a collection of files with the same name and different extensions. We want to add the file with the '.shp' extension to our QGIS project. You can use the dropdown menu at the bottom right of the browse window to display only those files with a .shp extension if you wish :

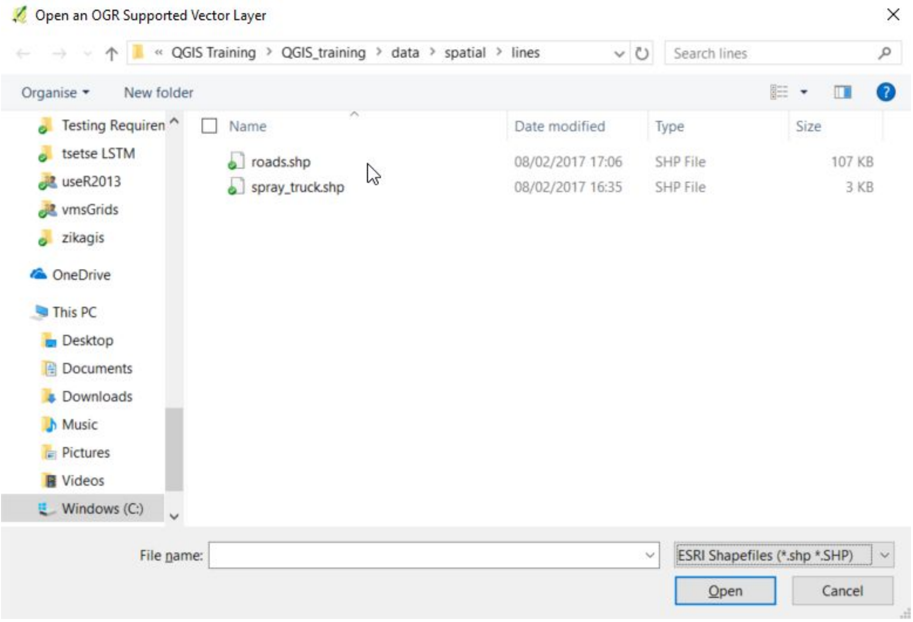
- ➔ Click All Files at lower right
- ➔ Select 'Esri Shapefiles' so that only these will be displayed


Open an OGR Supported Vector Layer
« QGIS Training › QGIS_training › data › spatial › lines
Organise ▾
New folder
Testing Requiren
tsetse LSTM
useR2013
vmsGrids
zikagis
OneDrive
This PC
Desktop
Documents
Downloads
Music
Pictures
Videos
Windows (C:)
File name:
All files (*) (*.*)
GDAL/OGR VSIFileHandler (*.zip *.gz *.tar *.tar.g
Arc/Info ASCII Coverage (*.e00 *.E00)
Arc/Info Generate (*.gen *.GEN)
Atlas BNA (*.bna *.BNA)
AutoCAD DXF (*.dxf *.DXF)
Comma Separated Value (*.csv *.CSV)
Czech Cadastral Exchange Data Format (*.vfk *.V
EDIGEO (*.thf *.THF)
ESRI Personal GeoDatabase (*.mdb *.MDB)
ESRI Shapefiles (*.shp *.SHP)
GPS eXchange Format [GPX] (*.gpx *.GPX)
GPSTrackMaker (*.gtm *.gtz *.GTM *.GTZ)
GeoJSON (*.geojson *.GEOJSON)
GeoPackage (*.gpkg *.GPKG)
GeoRSS (*.xml *.XML)
Geoconcept (*.gxt *.txt *.GXT *.TXT)
Geography Markup Language [GML] (*.gml *.GM
Geospatial PDF (*.pdf *.PDF)
Hydrographic Transfer Format (*.htf *.HTF)
INTERLIS 1 (*.itf *.xml *.ili *.ITF *.XML *.ILI)
INTERLIS 2 (*.itf *.xml *.ili *.ITF *.XML *.ILI)
Keyhole Markup Language [KML] (*.kml *.kmz *.
MS Excel format (*.xls *.XLS)
MS Office Open XML spreadsheet (*.xlsx *.XLSX)
Mapinfo File (*.mif *.tab *.MIF *.TAB)
Microstation DGN (*.dgn *.DGN)
NAS - ALKIS (*.xml *.XML)
Open Document Spreadsheet (*.ods *.ODS)
OpenAir Special Use Airspace Format (*.txt *.TX
All files (*) (*.*)
Open
Cancel

Name	Date modified	Typ
roads.dbf	08/02/2017 17:06	DB
roads.prj	08/02/2017 17:06	PR
roads.qpj	08/02/2017 17:06	QP
roads.shp	08/02/2017 17:06	SH
roads.shx	08/02/2017 17:06	SH
spray_truck.dbf	08/02/2017 16:35	DB
spray_truck.prj	08/02/2017 16:35	PR
spray_truck.qpj	08/02/2017 16:35	QP
spray_truck.shp	08/02/2017 16:35	SH
spray_truck.shx	08/02/2017 16:35	SH

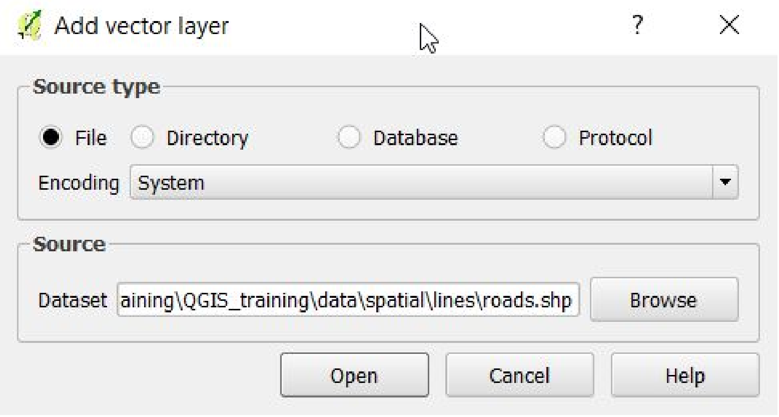
Once we have selected to just see ESRI Shapefiles then just a single entry for each shapefile will appear.

➔ double-click on the file 'roads.shp'



Name	Date modified	Type	Size
roads.shp	08/02/2017 17:06	SHP File	107 KB
spray_truck.shp	08/02/2017 16:35	SHP File	3 KB

➔ click ‘Open’



Add vector layer

Source type

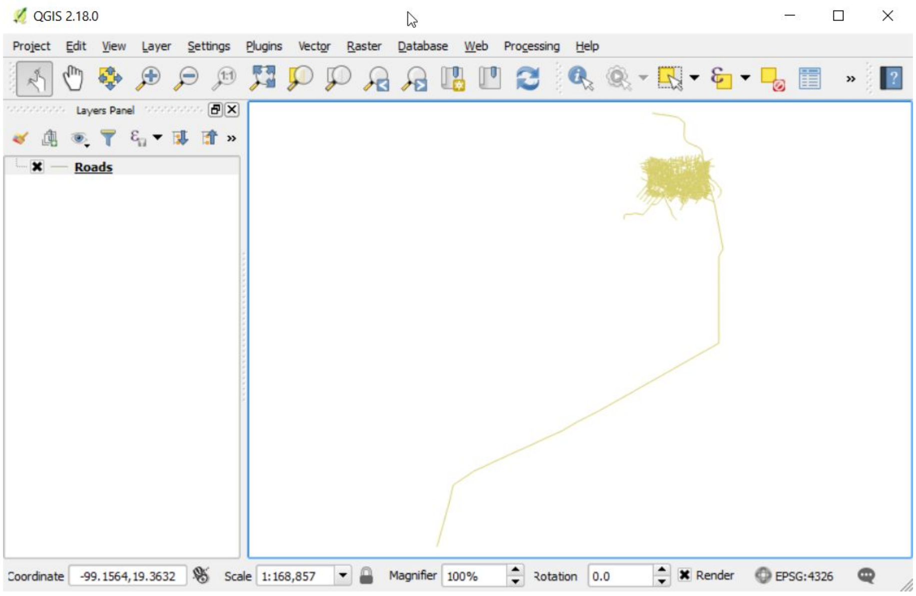
● File ○ Directory ○ Database ○ Protocol

Encoding System

Source

Dataset aining\QGIS_training\data\spatial\lines\roads.shp

You should now see the roads layer displayed in the map view on the right and the layers panel on the left. QGIS chooses different default colours depending on what you have done before so don't worry if your roads are a different colour.



QGIS 2.18.0

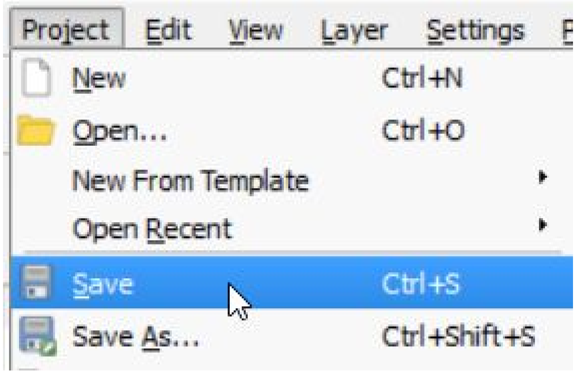
Layers Panel

☒ Roads

Coordinate -99.1564,19.3632 Scale 1:168,857 Magnifier 100% Rotation 0.0 ☒ Render EPSG:4326

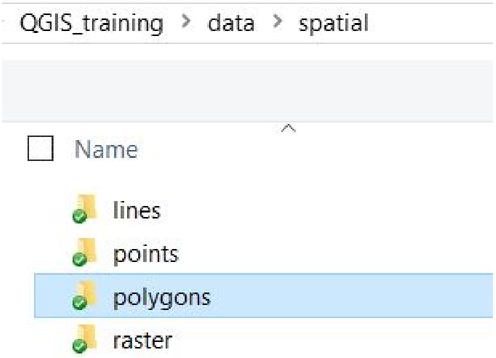
Once you have added layers to a project it is a good idea to save the project so that you can come back to it later.

- ➔ click Project, Save from the menu bar.
- ➔ browse to the folder : my_work/QGIS_projects/
- ➔ name the project : day2.qgs



Now we can add other layers to our project. We are going to firstly add the Mexico administrative boundary polygons, plus the block polygons to our map. To do this, repeat the previous steps :

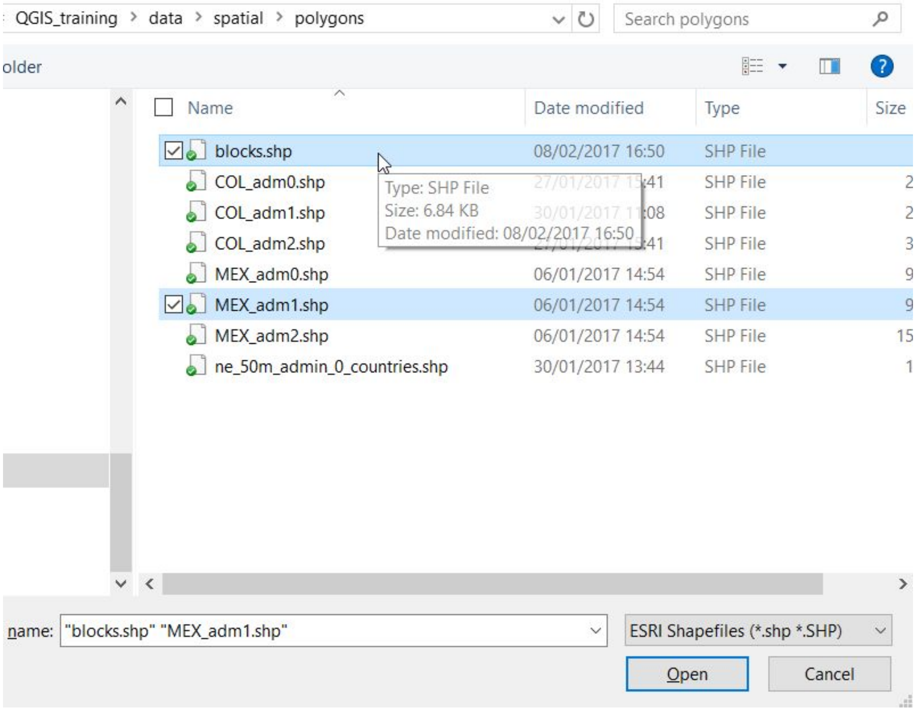
- ➔ Left click : Layer, Add Layer, Add Vector Layer
- ➔ click 'Browse' to find a file in your folders.
- ➔ go to the /data/spatial/polygons folder.



QGIS should remember your setting from earlier to just display one file per shapefile.

We wish to add two more layers to the QGIS project. We can use the ‘Ctrl’ button to select both so that they can be added to the project simultaneously.

- ➔ Hold down the Ctrl key
- ➔ Click on ‘blocks.shp’ and ‘MEX_adm1.shp’
- ➔ Click Open



QGIS_training > data > spatial > polygons

Search polygons

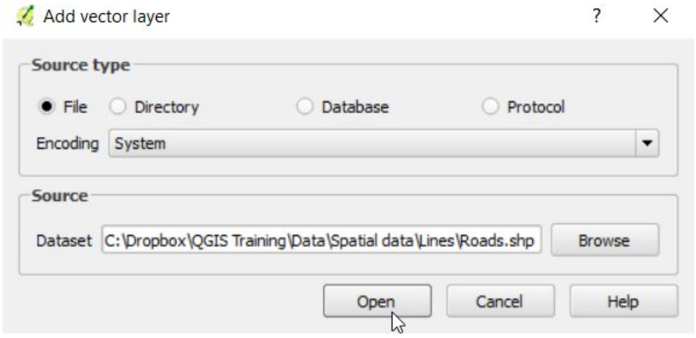
	Name	Date modified	Type	Size
☑	blocks.shp	08/02/2017 16:50	SHP File	
	COL_adm0.shp	27/01/2017 13:41	SHP File	2
	COL_adm1.shp	30/01/2017 11:08	SHP File	2
	COL_adm2.shp	27/01/2017 13:41	SHP File	3
	MEX_adm0.shp	06/01/2017 14:54	SHP File	9
☑	MEX_adm1.shp	06/01/2017 14:54	SHP File	9
	MEX_adm2.shp	06/01/2017 14:54	SHP File	15
	ne_50m_admin_0_countries.shp	30/01/2017 13:44	SHP File	1

Type: SHP File
Size: 6.84 KB
Date modified: 08/02/2017 16:50

name: "blocks.shp" "MEX_adm1.shp"

ESRI Shapefiles (*.shp *.SHP)

➔ Click Open again in the 'Add vector layer' window.



Add vector layer

Source type

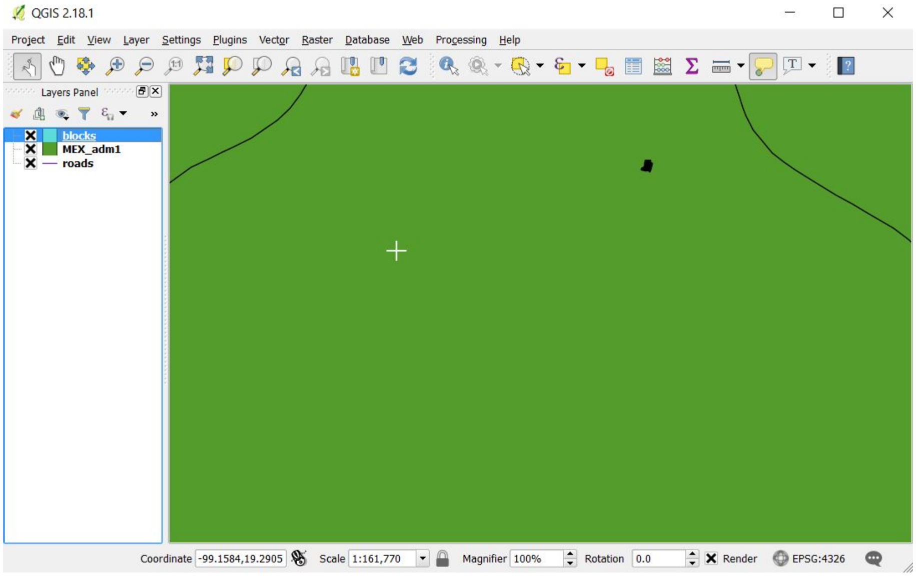
● File ○ Directory ○ Database ○ Protocol

Encoding: System

Source

Dataset: C:\Dropbox\QGIS Training\Data\Spatial data\Lines\Roads.shp

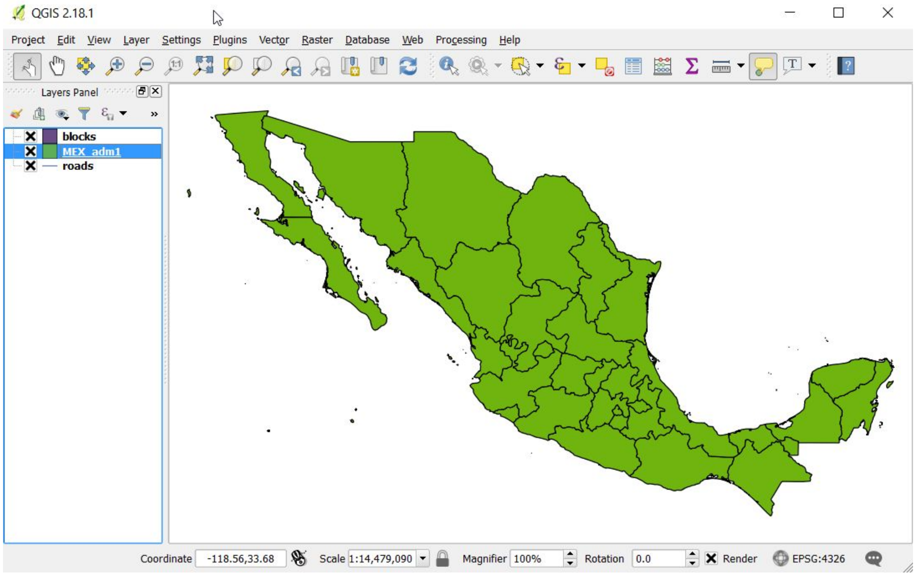
This should give you something like this (which doesn't look very informative!). Remember that your colors may be different and not to worry about that.


QGIS 2.18.1
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Layers Panel
☒ blocks
☒ MEX_adm1
☒ roads

Coordinate -99.1584,19.2905 Scale 1:161,770 Magnifier 100% Rotation 0.0 ☒ Render EPSG:4326

The colours that QGIS has allocated to the polygons can be seen in the Layers Panel on the left. You should see that the small box by MEX_adm1 is the same color as the background in the map view.

➔ To see the whole of the layer, right click on 'MEX_adm1' in the layers panel and select 'Zoom to layer'.

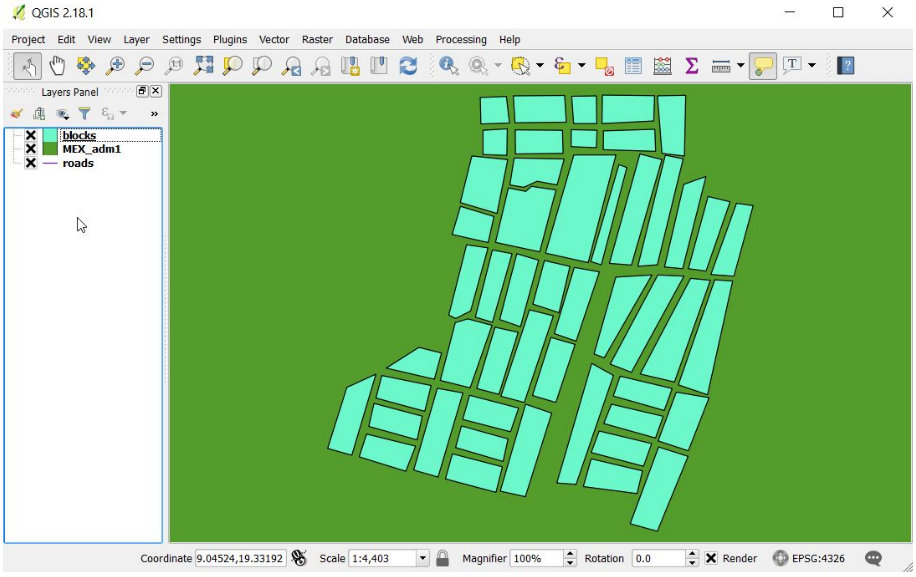

QGIS 2.18.1
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Layers Panel
☒ blocks
☒ MEX_adm1
☒ roads

Coordinate -118.56,33.68 Scale 1:14,479,090 Magnifier 100% Rotation 0.0 ☒ Render EPSG:4326

Now we wish to look more closely at the other layers.

➔ In the Layers Panel, right click on 'blocks'. Select 'Zoom to layer' as before.

Now you should see the familiar blocks we were looking at yesterday.



Coordinate 9.04524,19.33192 Scale 1:4,403 Magnifier 100% Rotation 0.0 ☒ Render EPSG:4326

Explore the map a little more, by reordering or turning off some of the layers.

We will return to this later in the session.

➔ click Project, Save from the Menu bar to save your progress.

2.2. Creating spatial data from paper-based entomology surveys

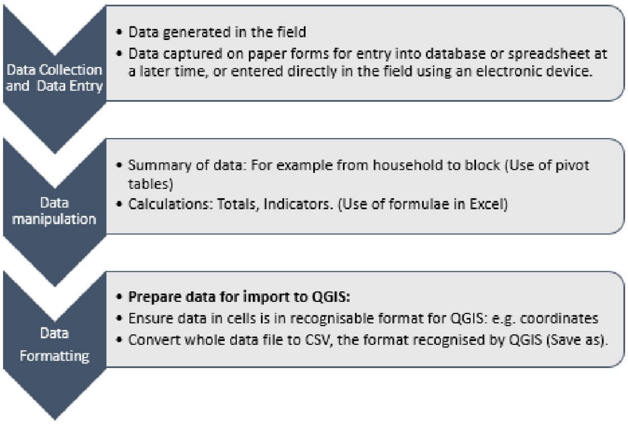
Any map you create, will only be as informative as the data behind it. It is therefore important to think about what data you collect, how you want to present it in the end, and the steps in between that will be required.

In our example, an entomology survey is being carried out at households, with data being recorded on paper forms by the survey team. This data is then going to be entered into an Excel spreadsheet. In this example, the GPS coordinates (longitude and latitude) were written on the paper form. Note that this might not always be the most practical way to record location information. We will be exploring GPS coordinates in more detail on Day 3, including ways in which coordinates can be recorded electronically.

Then, Excel will be used to summarise the results and calculate indicators. Finally, the data to be presented in a map will be saved in a format compatible with QGIS, ready to be joined to a shapefile.

The following figure summarises the steps we will be touching on in this module. The type and extent of data manipulation required will vary depending on the dataset and the desired outputs. We will just introduce some tools and examples that may be useful.

Figure 2: Data processing flow from field to mapping format.



Data Collection and Data Entry

- Data generated in the field
- Data captured on paper forms for entry into database or spreadsheet at a later time, or entered directly in the field using an electronic device.

Data manipulation

- Summary of data: For example from household to block (Use of pivot tables)
- Calculations: Totals, Indicators. (Use of formulae in Excel)

Data Formatting

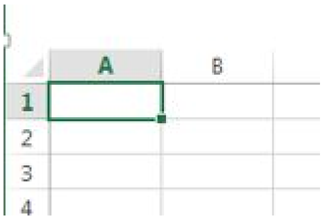
- **Prepare data for import to QGIS:**
- Ensure data in cells is in recognisable format for QGIS: e.g. coordinates
- Convert whole data file to CSV, the format recognised by QGIS (Save as).

2.2.1. Data entry in Excel

If you have used a computer for work, you are likely to have come across Excel. Microsoft Excel is commonly used throughout the world to record and manipulate small datasets. Excel could be a whole other training in itself, for this course we will just touch on some basic functionality that could be useful in handling data in preparation for presenting it as a map.

The appearance of Excel on your computer will vary depending on which version is installed on the computer you are using, but the basic functionality will be the same. Excel is not free, it comes as part of the Microsoft Office package often purchased and installed as standard with a new computer. There are free open source alternatives available online if required (Apache OpenOffice Calc has a pivot table equivalent called data pilots for example), but we will only consider Excel in our example.

When you open an Excel worksheet, the active cell will be highlighted showing your position in the worksheet. In the example here, this is denoted by the green border around the cell. Note that the corresponding column and row numbers will also be shaded for your reference, in this case A1.



A mouse can be used to navigate your worksheet, but it is often quicker and preferable when entering data to use the keyboard as much as possible. If you use the 'enter' key in Excel, you will move down the column one cell at a time (e.g. A1 to A2). If you use the 'tab' key you will move from left to right (e.g. A1 to B1). The arrow keys can also be used to move through the worksheet. Shortcuts are also available. One useful example is Ctrl + arrow key, this will take you to the either end of the data entered in your column or row, depending on which arrow you use.

It is possible to pre-configure your worksheet for data entry, for example to add validation, formulas, or formatting. This is beyond the scope of this training, but is something worth bearing in mind if there are a lot of data to be entered. It could reduce the amount of time spent cleaning your data afterwards. It is also worth bearing in mind when entering data into a worksheet created by somebody else, for example are you copy-pasting data or a formula? If the data collection form and data entry sheet can be set up to resemble each other (order of variables, direction of data in rows/columns) this will also facilitate data entry.

Before using your data, they should checked and cleaned if necessary. For example, are the numbers within the range you would expect? Have any dates been entered correctly and in the same format? Is there consistent spelling of any text fields?

2.2.2. Data manipulation and saving in CSV format

Sometimes raw data collected from the field needs to be summarised or used to calculate indicators before being ready to present to others. You will want your map to be clear and meaningful, so you may need to work a little on your dataset to prepare the results ready for display.

Formulas can be introduced in Excel to perform calculations which can save time if there are many replicates and reduce possible human error involved in manual calculation. They can also be used to introduce a specific character or combine cell contents, which we will come onto later.

Pivot tables can be a useful tool for examining a dataset. We will demonstrate how they can be used to summarise data.

QGIS will not recognise a regular Excel spreadsheet or a pivot table. Therefore once the data are ready for mapping, we will save them as a CSV file. CSV stands for ‘Comma Separated Values’ in English, the data are literally saved as lists of values divided by commas, or if your system is in Spanish by semi-colons. This makes them small files, good for transferring data between programmes. Any visual formatting such as colours and lines will not be retained.

Formulas in Excel

It is possible to do many things in Excel. We will look at how to perform simple calculations such as additions, divisions and multiplications, and then go on to introduce a logical function ‘if’. We will return to our entomological survey data as an example. Find and open the Excel file “household_survey_jul_sep.xlsx” in your USB (training_QGIS/ data/ entomology/). The file has three tabs of data. Click on the July tab. We are going to complete this data set by filling in values for the columns shaded yellow.

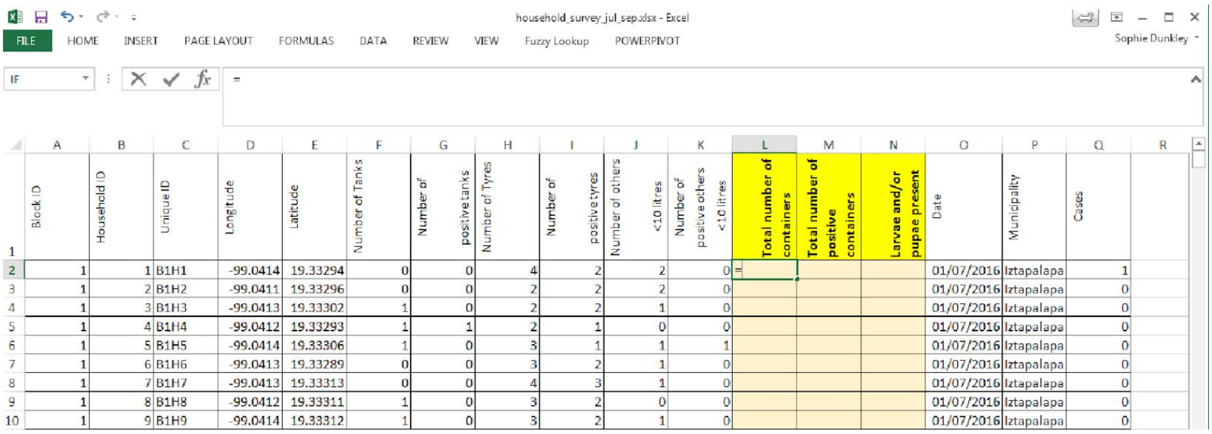
Addition

In our survey, the number of water tanks, tyres, plants, and other containers were registered, as were the number of each that were positive (had larvae and/or pupae present). Imagine we want to know the total number of containers for each household. Rather than do this calculation for each of the 500 households manually, we can create a formula and calculate all 500 rows of data much more quickly, and reduce the opportunities for introducing human error by doing the calculation once.

One way to start a formula in Excel is to type = in the cell where you want the result to appear. You can then select cells of data you are interested in by clicking on them, and type in the appropriate symbol.

When you have finished, hit the Enter key and the result of the calculation will be returned in that cell.

For example, in our spreadsheet, find cell L2, the first row of data for the column “Total number of containers”, and type the symbol “=”.

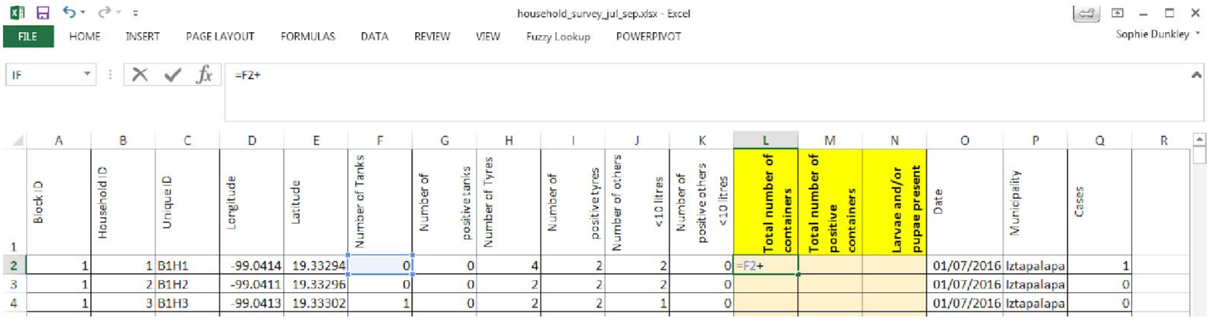


=

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres	Number of positive tyres	Number of others <10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**	**Larvae and/or pupae present**	Date	Municipality	Cases	
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	=			01/07/2016	Iztapalapa	1	
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0				01/07/2016	Iztapalapa	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0				01/07/2016	Iztapalapa	0	
5	1	4	B1H4	-99.0412	19.33293	1	1	2	1	0	0				01/07/2016	Iztapalapa	0	
6	1	5	B1H5	-99.0414	19.33306	1	0	3	1	1	1				01/07/2016	Iztapalapa	0	
7	1	6	B1H6	-99.0413	19.33289	0	0	3	2	1	0				01/07/2016	Iztapalapa	0	
8	1	7	B1H7	-99.0413	19.33313	0	0	4	3	1	0				01/07/2016	Iztapalapa	0	
9	1	8	B1H8	-99.0412	19.33311	1	0	3	2	0	0				01/07/2016	Iztapalapa	0	
10	1	9	B1H9	-99.0414	19.33312	1	0	3	2	1	0				01/07/2016	Iztapalapa	0	

Next, use your mouse to click in cell F2 which has the number of tanks observed at household 01.

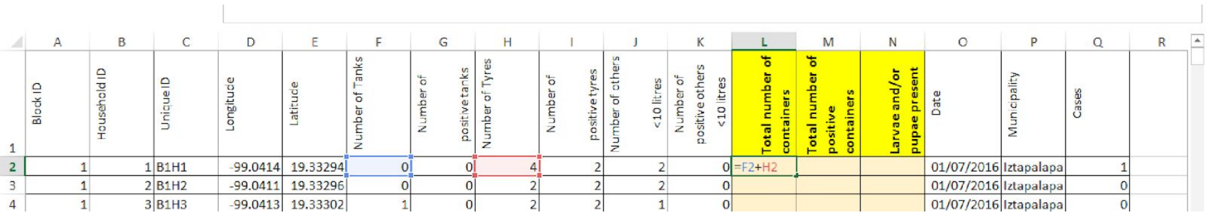
Type + (for an addition, substitute for the required symbol for other calculations)



=F2+

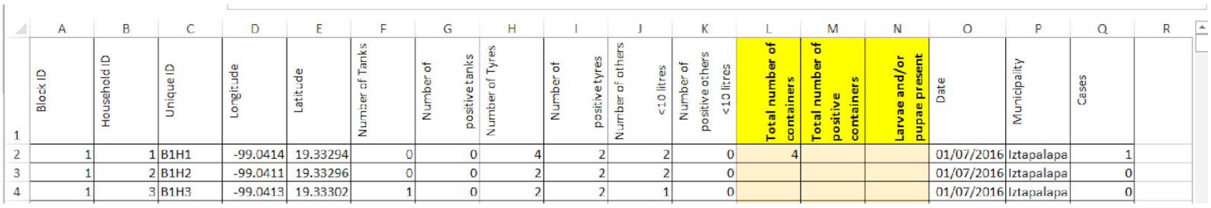
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres	Number of positive tyres	Number of others <10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**	**Larvae and/or pupae present**	Date	Municipality	Cases	
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	=F2+			01/07/2016	Iztapalapa	1	
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0				01/07/2016	Iztapalapa	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0				01/07/2016	Iztapalapa	0	

Then, click in cell H2 to select the number of tyres seen and add this.



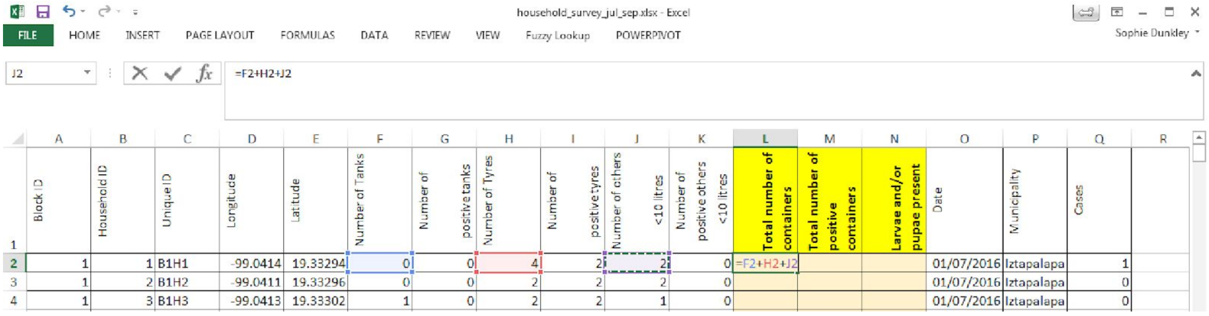
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres	Number of positive tyres	Number of others <10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**	**Larvae and/or pupae present**	Date	Municipality	Cases	
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	=F2+H2			01/07/2016	Iztapalapa	1	
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0				01/07/2016	Iztapalapa	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0				01/07/2016	Iztapalapa	0	

If you hit Enter at this point, you should see the result of the addition of these two cells in cell L2 where we have been writing the formula. In this case, 4, as 0+4=4.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres	Number of positive tyres	Number of others <10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**	**Larvae and/or pupae present**	Date	Municipality	Cases	
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	4			01/07/2016	Iztapalapa	1	
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0				01/07/2016	Iztapalapa	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0				01/07/2016	Iztapalapa	0	

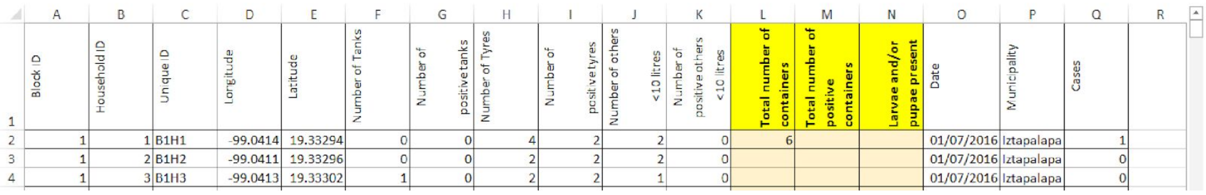
To edit the formula further, double-click in the cell, check that the cursor is flashing at the point where you wish to make the edit, and make your adjustments. For example, add the number of other containers by adding another "+" sign, and clicking in cell J2 to select the required data.



=F2+H2+J2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres	Number of positive tyres	Number of others <10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**	**Larvae and/or pupae present**	Date	Municipality	Cases	
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	=F2+H2+J2			01/07/2016	Iztapalapa	1	
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0				01/07/2016	Iztapalapa	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0				01/07/2016	Iztapalapa	0	

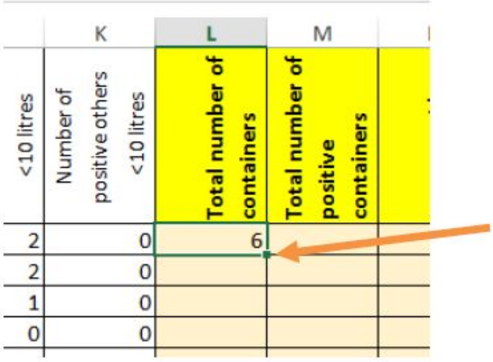
If you hit the Enter key, the total number of containers should now read 6.



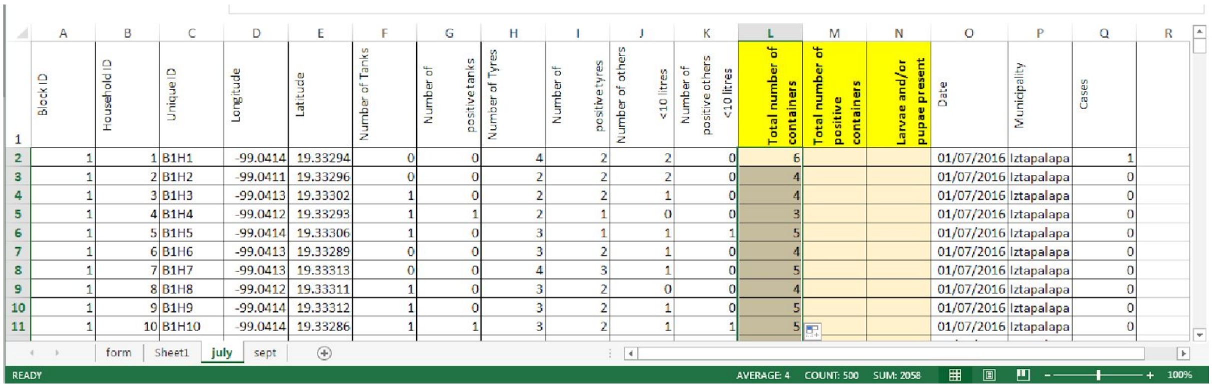
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres	Number of positive tyres	Number of others <10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**	**Larvae and/or pupae present**	Date	Municipality	Cases	
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	6			01/07/2016	Iztapalapa	1	
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0				01/07/2016	Iztapalapa	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0				01/07/2016	Iztapalapa	0	

It is good practice to verify if the result returned is at least approximately as expected. You can also double check that you have added all the required cells by double clicking in your result cell again. Note that as well as showing the formula, the cells included will be highlighted by this action, facilitating a check. The calculation for the other 499 rows of data in our survey will be based on the construction of the formula in the first row, so it is good to take the opportunity to check at this point.

Next, instead of repeating this 500 times, we can reproduce this formula for the remaining rows. One way of doing this is to click on the small square in the bottom right-hand side of the active cell, and drag down the column as far as you want results. Another is to double-click on this same square, which will automatically fill the column with this formula, for as far as there is continuous data in the rest of the sheet.



	K	L	M
<10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**
2	0	6	
2	0		
1	0		
0	0		



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres	Number of positive tyres	Number of others <10 litres	Number of positive others <10 litres	**Total number of containers**	**Total number of positive containers**	**Larvae and/or pupae present**	Date	Municipality	Cases	
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	6			01/07/2016	Iztapalapa	1	
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0	4			01/07/2016	Iztapalapa	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0	4			01/07/2016	Iztapalapa	0	
5	1	4	B1H4	-99.0412	19.33293	1	1	2	1	0	0	3			01/07/2016	Iztapalapa	0	
6	1	5	B1H5	-99.0414	19.33306	1	0	3	1	1	1	5			01/07/2016	Iztapalapa	0	
7	1	6	B1H6	-99.0413	19.33289	0	0	3	2	1	0	4			01/07/2016	Iztapalapa	0	
8	1	7	B1H7	-99.0413	19.33313	0	0	4	3	1	0	5			01/07/2016	Iztapalapa	0	
9	1	8	B1H8	-99.0412	19.33311	1	0	3	2	0	0	4			01/07/2016	Iztapalapa	0	
10	1	9	B1H9	-99.0414	19.33312	1	0	3	2	1	0	5			01/07/2016	Iztapalapa	0	
11	1	10	B1H10	-99.0414	19.33286	1	1	3	2	1	1	5			01/07/2016	Iztapalapa	0	

Exercise 2.1

Use the same addition process to calculate the Total number of positive containers, for all 500 rows in column M. We want to know the number of positive tanks, plus the number of positive tyres, plus the number of positive others.

When you have finished, save a copy of the Excel file in your “my work” folder (browse on your USB to training_QGIS/my_work/data/entomology), call it “my_household_containers”.

Logical Function “If”

We have seen that formulas can be built up as above, by directly typing the required symbols and selecting the cells to be included. A function is a predefined formula. There are numerous functions available in Excel that have been developed to facilitate common tasks. We are going to use the “If” function.

First of all, what are we trying to achieve? The third column in yellow in our example is “Larvae and/or pupae present”. Why might we want to know this? Remember that our example is about monitoring *Aedes aegypti.* There are three indices commonly used for this, the Stegomyia indices. Let us consider one of these, the house index, the formula for which is as follows:

House index (HI): percentage of houses infested with larvae and/or pupae

$$HI = 100 \times Number\ of\ inspected\ \hbar ouses\ wit\hbar\ at\ least\ one\ positive\ container\ /$$

$$Total\ number\ of\ inspected\ \hbar ouses$$

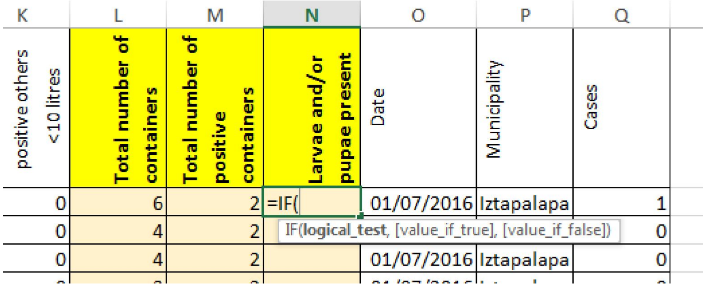
To be able to calculate this index, we are going to need to know the “number of inspected houses with at least one positive container”, or to put it another way, if there were larvae or pupae present for each and every household visited.

In exercise 2.1 above, we calculated the number of positive containers per household (per row). We can use this information to categorize the houses into those with at least one positive container and those with none. This is what we are going to do with the If function, ask Excel that **if** the value fir the total number of positive containers in column M is greater than zero, put a “1” in the column for larvae and/or pupae present, if not, put “0”.

There are different ways to do this in practice. Once you are familiar with a function and how the argument is constructed, you can type it directly into the cell as with other formulas, by typing =, then IF, then the rest of the conditions in parentheses. In our case it will look as follows:

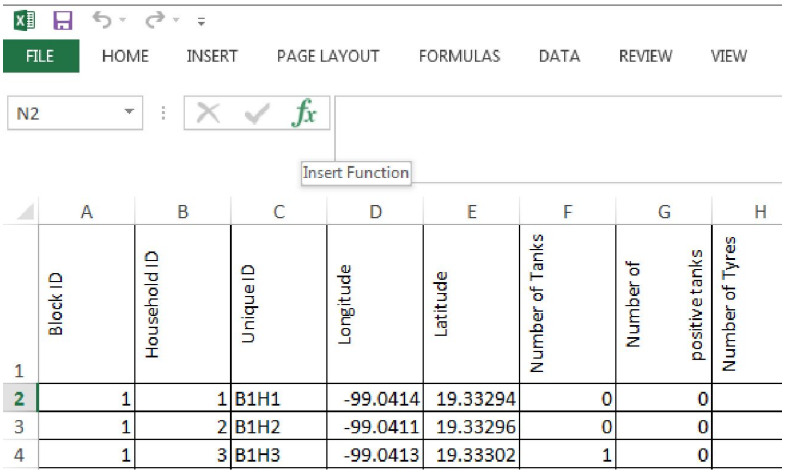
```
=IF(M2>0,1,0)
```

If you type in the function, when you open the parentheses you may notice the system display a message about the construction of the function, as below.



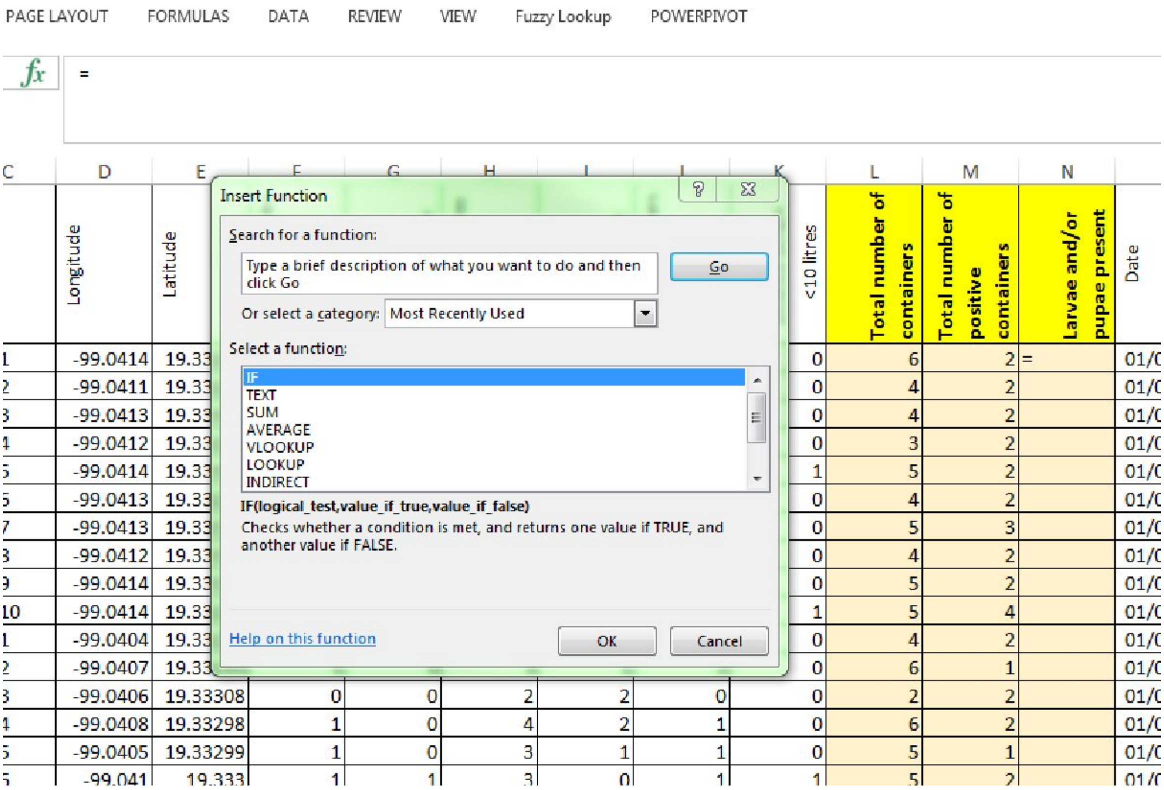
	K	L	M	N	O	P	Q
	positive others <10 litres	Total number of containers	Total number of positive containers	Larvae and/or pupae present	Date	Municipality	Cases
	0	6	2	=IF(	01/07/2016	Iztapalapa	1
	0	4	2	IF(**logical_test**, [value_if_true], [value_if_false])			0
	0	4	2		01/07/2016	Iztapalapa	0

Another way to do this is to use “Insert Function”, the shortcut for which can be found as seen here:



	A	B	C	D	E	F	G	H
1	Block ID	Household ID	Unique ID	Longitude	Latitude	Number of Tanks	Number of positive tanks	Number of Tyres
2	1	1	B1H1	-99.0414	19.33294	0	0	
3	1	2	B1H2	-99.0411	19.33296	0	0	
4	1	3	B1H3	-99.0413	19.33302	1	0	

Click in the cell where you want your formula to be, then click on the insert function shortcut. This will open the insert function window.



Search for a function:

Type a brief description of what you want to do and then click Go

Or select a category: Most Recently Used

Select a function:

IF
TEXT
SUM
AVERAGE
VLOOKUP
LOOKUP
INDIRECT

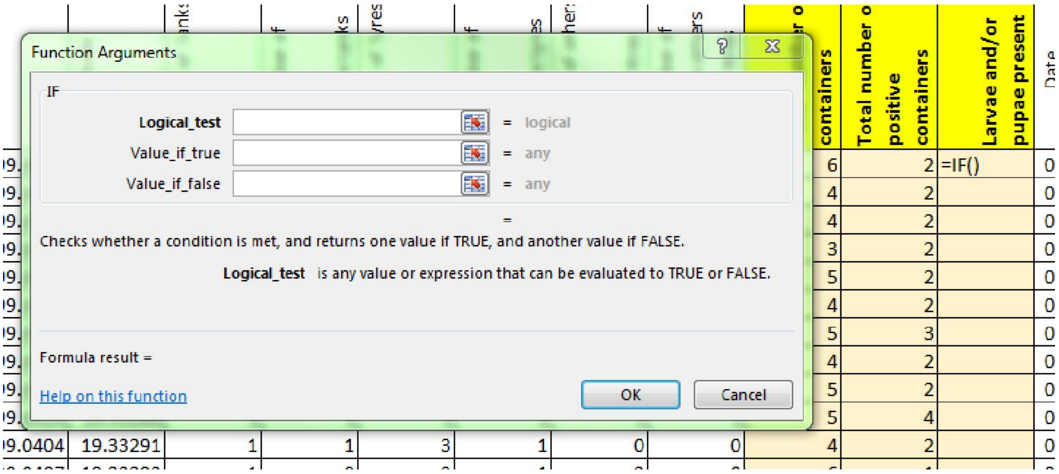
IF(logical_test,value_if_true,value_if_false)

Checks whether a condition is met, and returns one value if TRUE, and another value if FALSE.

C	D Longitude	E Latitude	F	G	H	I	J	K <10 litres	L Total number of containers	M Total number of positive containers	N Larvae and/or pupae present	Date
1	-99.0414	19.33						0	6	2	=	01/(
2	-99.0411	19.33						0	4	2		01/(
3	-99.0413	19.33						0	4	2		01/(
4	-99.0412	19.33						0	3	2		01/(
5	-99.0414	19.33						1	5	2		01/(
5	-99.0413	19.33						0	4	2		01/(
7	-99.0413	19.33						0	5	3		01/(
3	-99.0412	19.33						0	4	2		01/(
9	-99.0414	19.33						0	5	2		01/(
10	-99.0414	19.33						1	5	4		01/(
1	-99.0404	19.33						0	4	2		01/(
2	-99.0407	19.33						0	6	1		01/(
3	-99.0406	19.33308	0	0	2	2	0	0	2	2		01/(
4	-99.0408	19.33298	1	0	4	2	1	0	6	2		01/(
5	-99.0405	19.33299	1	0	3	1	1	0	5	1		01/(
5	-99.041	19.333	1	1	3	0	1	1	5	2		01/(

This will offer you a search facility and a select list by category (in this above example, Most Recently Used). It also displays an explanation for the currently selected function (highlighted in blue) and structure of that function.

Select IF and click OK. The Function Arguments window will be displayed.



IF

Logical_test		= logical
Value_if_true		= any
Value_if_false		= any

=

Checks whether a condition is met, and returns one value if TRUE, and another value if FALSE.

Logical_test is any value or expression that can be evaluated to TRUE or FALSE.

Formula result =

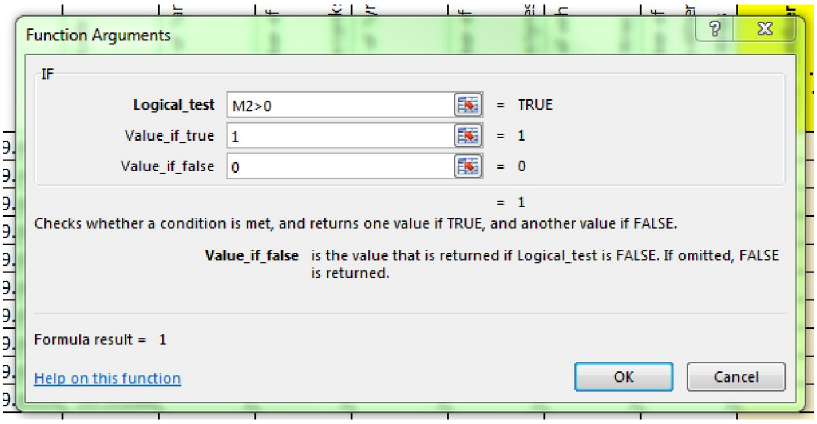
								containers	Total number of positive containers	Larvae and/or pupae present	Date
19.								6	2	=IF()	0
19.								4	2		0
19.								4	2		0
19.								3	2		0
19.								5	2		0
19.								4	2		0
19.								5	3		0
19.								4	2		0
19.								5	2		0
19.								5	4		0
9.0404	19.33291	1	1	3	1	0	0	4	2		0

This can be used to construct our formula. Excel will automatically add in the required syntax. If you look in cell N2, you will see that it has already started.

Using this window, we just need to fill in the three boxes as follows.

The **logical test** is the check we are asking the system to perform. Remember that what we want to know is if the household has at least one positive container, so what we will ask is if cell M2 (total number of positive containers) is greater than 0. You can type this all in “M2>0” or select the cell M2 and type in the >0 part.

Then we have **value if true** and **value if false**. This is where we instruct the system what to do with the results of the logical test. If it is true, that M2>0, what should it do? In our example we are going to ask that it return a value of “1” if it is true, and “0” if it is false. Simply type these numbers into the given boxes. You could use text for these values if useful in your situation to display this, make it return “Yes” or “No” or “at least one positive” for instance. In our example, 1 and 0 are relatively easy to understand, and may also facilitate summary later, knowing that the number 1 is equivalent to one positive household.



IF

Logical_test M2>0 = TRUE

Value_if_true 1 = 1

Value_if_false 0 = 0

= 1

Checks whether a condition is met, and returns one value if TRUE, and another value if FALSE.

Value_if_false is the value that is returned if Logical_test is FALSE. If omitted, FALSE is returned.

Formula result = 1

When you have entered this information, click OK.
The result of the formula should be displayed in cell N2, in this case “1”.
The function can be copied for the rest of the column as before, by double clicking on the small square in the bottom right-hand corner. Do this to fill in the results for the whole of column N.
You should now have data filled in for the three yellow columns.

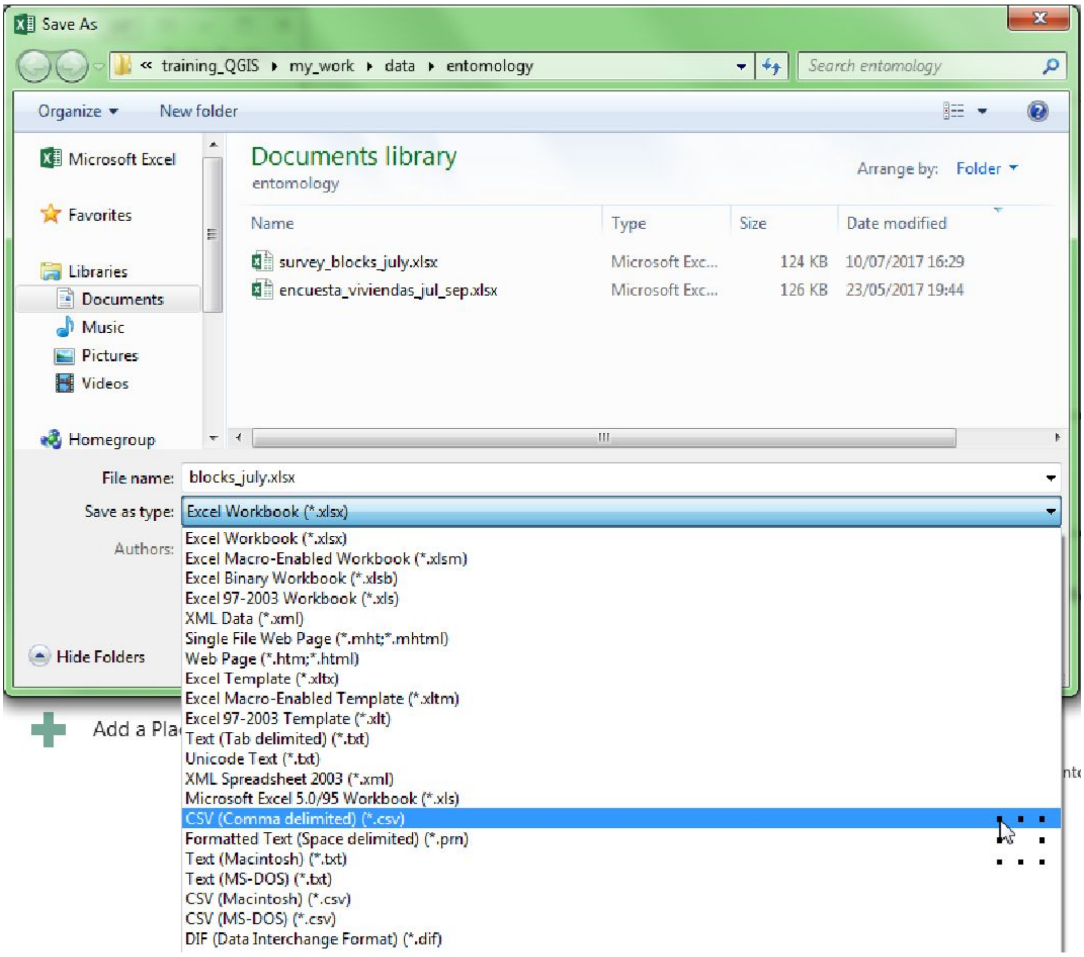
We will return to formulas later, to calculate the Stegomyia indices mentioned earlier. We will also introduce pivot tables as a way of summarising data, but first we will introduce CSV files.

Data Format – Saving as a CSV file

We have some data in an Excel spreadsheet. We want to visualise this as a map. We are going to use QGIS to do this, so we need to get our data from Excel into QGIS. QGIS does not recognise the Excel Workbooks in .xlsx format. Like many other softwares, it uses a format called CSV. Therefore we are going to change our file from .XLSX to .CSV.

CSV stands for “comma separated values”, also known as a comma delimited file. CSV files save the information as lists of the values (text or numbers), separated one from another by commas (or in languages where commas are used to denominate a decimal place, the values can instead be separated by a semi-colon to avoid confusion in interpretation of numbers). CSV files are commonly used for data transfer as they are simple and relatively small.

To convert our data from .xlsx to .csv we will use the “Save As” option in Excel (File/Save As) and simply change the file type in the drop down “save as type” option from “Excel Workbook (*.xlsx)”, to “CSV (Comma delimited)(*.csv)”.


Save As
« training_QGIS ▸ my_work ▸ data ▸ entomology
Search entomology
Organize ▾
New folder
Microsoft Excel
Favorites
Libraries
Documents
Music
Pictures
Videos
Homegroup
Documents library
entomology
Arrange by: Folder ▾
Authors:
Hide Folders
Add a Pla
Excel Workbook (*.xlsx)
Excel Macro-Enabled Workbook (*.xlsm)
Excel Binary Workbook (*.xlsb)
Excel 97-2003 Workbook (*.xls)
XML Data (*.xml)
Single File Web Page (*.mht;*.mhtml)
Web Page (*.htm;*.html)
Excel Template (*.xltx)
Excel Macro-Enabled Template (*.xltm)
Excel 97-2003 Template (*.xlt)
Text (Tab delimited) (*.txt)
Unicode Text (*.txt)
XML Spreadsheet 2003 (*.xml)
Microsoft Excel 5.0/95 Workbook (*.xls)
CSV (Comma delimited) (*.csv)
Formatted Text (Space delimited) (*.prn)
Text (Macintosh) (*.txt)
Text (MS-DOS) (*.txt)
CSV (Macintosh) (*.csv)
CSV (MS-DOS) (*.csv)
DIF (Data Interchange Format) (*.dif)

Name	Type	Size	Date modified
survey_blocks_july.xlsx	Microsoft Exc...	124 KB	10/07/2017 16:29
encuesta_viviendas_jul_sep.xlsx	Microsoft Exc...	126 KB	23/05/2017 19:44

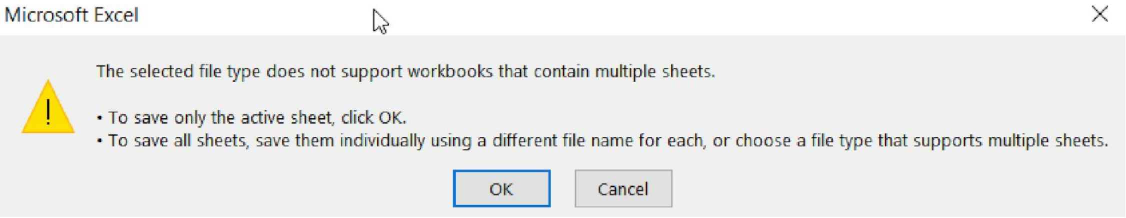
File name: blocks_july.xlsx

Save as type: Excel Workbook (*.xlsx)

- ➔ Have your July data in your my_household_containers file open
- ➔ Go to File, Save As
- ➔ Navigate to your USB, my_work/data/entomology
- ➔ Give your file a new name “houses_july”
- ➔ Go to save as type and scroll through the dropdown menu to find and select CSV.
- ➔ Click Save.

You will get a series of warning messages, this is normal as explained below.

CSVs do not support workbooks that contain multiple sheets:



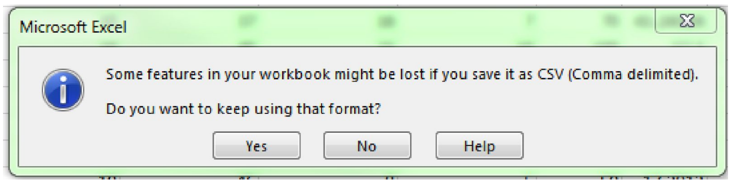
The selected file type does not support workbooks that contain multiple sheets.

• To save only the active sheet, click OK.
• To save all sheets, save them individually using a different file name for each, or choose a file type that supports multiple sheets.

- ➔ click ‘OK’.

Remember we are asking Excel to save this data as a list of values separated by commas. It essentially does not know it should represent the separation between worksheets.

Some features in your workbook might be lost if you save it as a CSV:



Microsoft Excel

Some features in your workbook might be lost if you save it as CSV (Comma delimited).

Do you want to keep using that format?

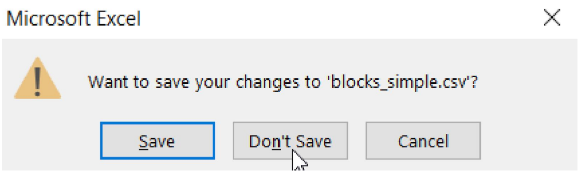
→ click Yes.

CSV is a relatively simple data format, efficient for data transfer as we have said. Excel workbooks allow you to display your data using colours, lines, text formatting and other features. These will not be retained in your CSV file, so this is just a warning to check that you are aware of this.

Your file should now have been converted and saved. Slightly confusingly, it is still readable in Excel.

Close your spreadsheet using the X button at the top right of the Excel window. You will get this slightly worrying message box when you close:

Do you want to save changes to your file?



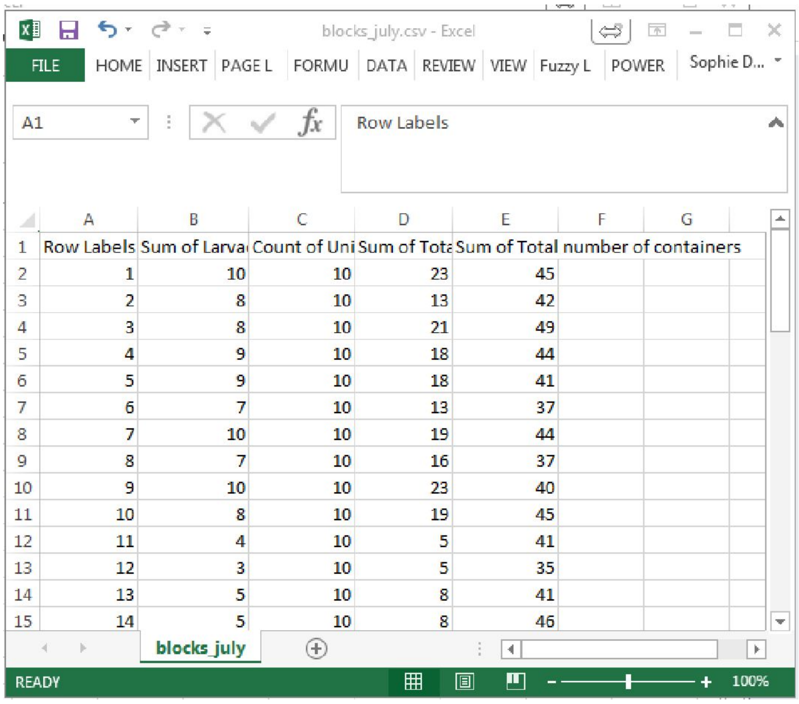
Microsoft Excel

Want to save your changes to 'blocks_simple.csv'?

→ Press 'Don't Save'

We have saved the file already. If you click Save, you will enter into a cycle of saving your file and may never escape.

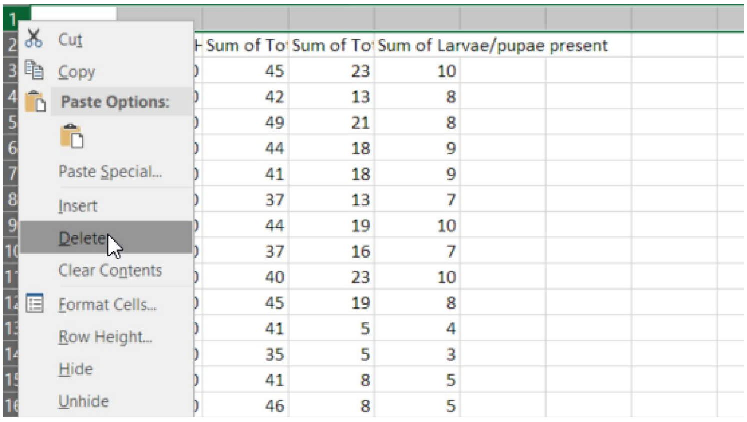
Check your new CSV file is there. Re-open it by browsing for it in your my_work folder (my_work/data/entomology/my_household_data.csv). Your file should open in Excel, and look similar to the screenshot below:



	A	B	C	D	E	F	G
1	Row Labels	Sum of Larva	Count of Uni	Sum of Tota	Sum of Total number of containers		
2	1	10	10	23	45		
3	2	8	10	13	42		
4	3	8	10	21	49		
5	4	9	10	18	44		
6	5	9	10	18	41		
7	6	7	10	13	37		
8	7	10	10	19	44		
9	8	7	10	16	37		
10	9	10	10	23	40		
11	10	8	10	19	45		
12	11	4	10	5	41		
13	12	3	10	5	35		
14	13	5	10	8	41		
15	14	5	10	8	46		

This data is now in a format recognised by QGIS and could be added to a project. There are just a couple more peculiarities to bear in mind.

The first row of data will be your variable names in QGIS. If you have a blank row at the top of the spreadsheet, you will therefore need to remove it. To do this, right-click on the number 1 next to the row, and select Delete. If your first row already contains the variable names, continue.



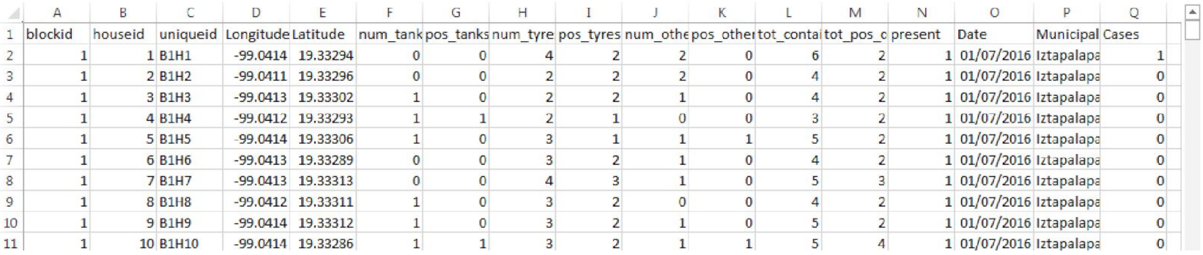
Sum of To	Sum of To	Sum of Larvae/pupae present
45	23	10
42	13	8
49	21	8
44	18	9
41	18	9
37	13	7
44	19	10
37	16	7
40	23	10
45	19	8
41	5	4
35	5	3
41	8	5
46	8	5

QGIS variable names are limited to 10 characters in length. The next step for us is to shorten the variable names longer than this. The data will not be rejected with longer names, the variable names will simply be cut after the first ten characters. This might make them difficult to understand, so it is easier to give them names of your choosing so that you can find and understand your data once it is introduced into QGIS.

We have provided suggestions for the shortened variable names in the table below :

Variable name – Used in Excel spreadsheet	**Short name – to be used in csv files**	**Description**
Block ID	blockid	Block ID, ranging from 1-50
Household ID	houseid	House ID within each block, ranging from 1-10
Unique ID	uniqueid	Unique household ID in the format BxxHxx
Number of tanks	num_tanks	Number of tanks found at the house
Number of positive tanks	pos_tanks	Number of tanks at the house where larvae/pupae were found
Number of tyres	num_tyres	Number of tyres found at the house
Number of positive tyres	pos_tyres	Number of tyres at the house where larvae/pupae were found
Number of others <10 litres	num_other	Number of other containers of less than ten litre capacity
Number of positive others <10 litres	pos_other	Number of other containers of less than ten litre capacity where larvae/pupae were found
Total number of containers	tot_contai	Total number of containers recorded at the house (variable to be calculated)
Total number of positive containers	tot_pos_co	Total number of larvae/pupae positive containers found at the house (variable to be calculated)
Larvae and/or pupae present	present	Indicator of larvae/pupae present (1= present, 0=absent) (variable to be calculated)
Date	date	Date of collection in DD/MM/YYYY format
Municipality	municipali	Municipality in which the household is situated
Cases	cases	Number of laboratory-confirmed Zika virus disease cases reported amongst those people residing in the surveyed house

➔ Change the column names to the short version



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	blockid	houseid	uniqueid	Longitude	Latitude	num_tank	pos_tanks	num_tyre	pos_tyres	num_othe	pos_other	tot_contai	tot_pos_c	present	Date	Municipal	Cases
2	1	1	B1H1	-99.0414	19.33294	0	0	4	2	2	0	6	2	1	01/07/2016	Iztapalapa	1
3	1	2	B1H2	-99.0411	19.33296	0	0	2	2	2	0	4	2	1	01/07/2016	Iztapalapa	0
4	1	3	B1H3	-99.0413	19.33302	1	0	2	2	1	0	4	2	1	01/07/2016	Iztapalapa	0
5	1	4	B1H4	-99.0412	19.33293	1	1	2	1	0	0	3	2	1	01/07/2016	Iztapalapa	0
6	1	5	B1H5	-99.0414	19.33306	1	0	3	1	1	1	5	2	1	01/07/2016	Iztapalapa	0
7	1	6	B1H6	-99.0413	19.33289	0	0	3	2	1	0	4	2	1	01/07/2016	Iztapalapa	0
8	1	7	B1H7	-99.0413	19.33313	0	0	4	3	1	0	5	3	1	01/07/2016	Iztapalapa	0
9	1	8	B1H8	-99.0412	19.33311	1	0	3	2	0	0	4	2	1	01/07/2016	Iztapalapa	0
10	1	9	B1H9	-99.0414	19.33312	1	0	3	2	1	0	5	2	1	01/07/2016	Iztapalapa	0
11	1	10	B1H10	-99.0414	19.33286	1	1	3	2	1	1	5	4	1	01/07/2016	Iztapalapa	0

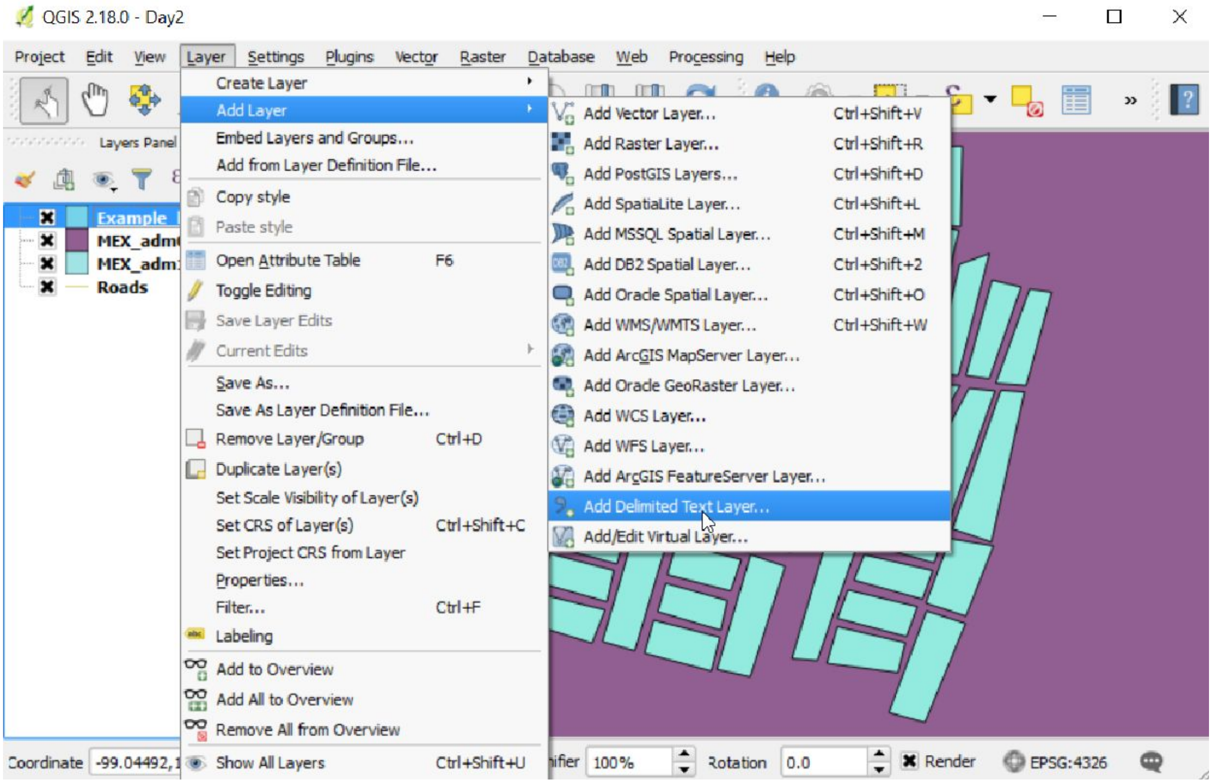
If you know that you will be transferring your data into QGIS you could avoid this step by giving your variables short names from the start. Note however that this will be important when importing pivot table data. We will come back to this.

Save and close your file. We will now see how to add this data to a map.

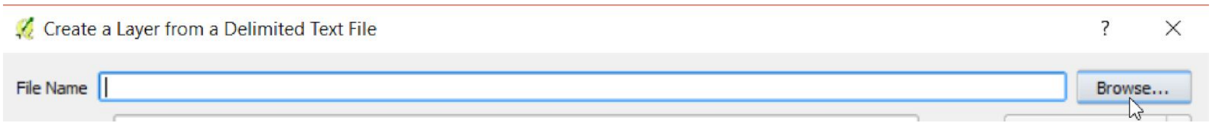
2.2.3. Adding data to a map by creating point data from a CSV

We shall now add houses_july.csv to QGIS.

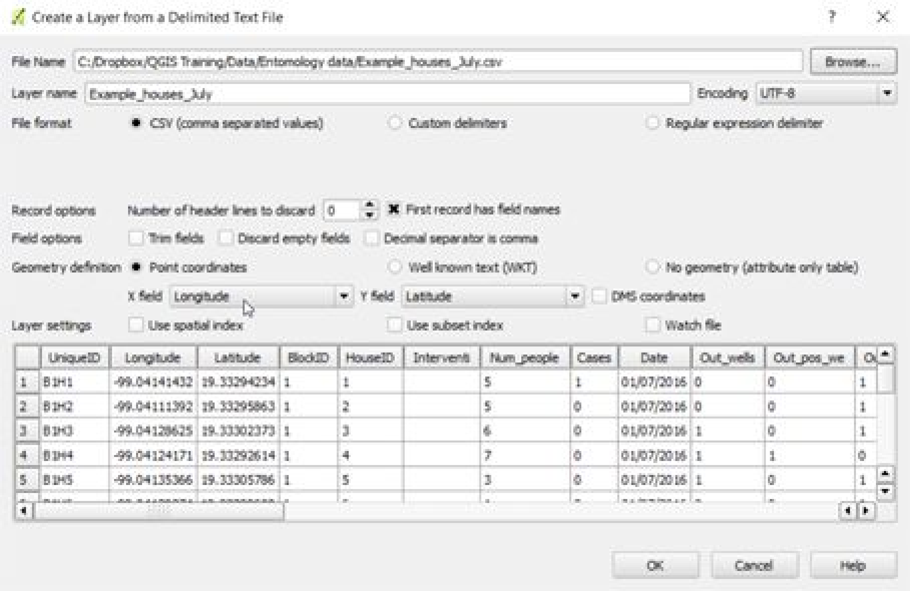
- ➔ Open up the QGIS project day2.qgs from earlier
- ➔ from the Menu bar and select Layer, Add Layer, Add Delimited Text Layer


QGIS 2.18.0 - Day2
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Create Layer
Add Layer
Embed Layers and Groups...
Add from Layer Definition File...
Copy style
Paste style
Open Attribute Table F6
Toggle Editing
Save Layer Edits
Current Edits
Save As...
Save As Layer Definition File...
Remove Layer/Group Ctrl+D
Duplicate Layer(s)
Set Scale Visibility of Layer(s)
Set CRS of Layer(s) Ctrl+Shift+C
Set Project CRS from Layer
Properties...
Filter... Ctrl+F
Labeling
Add to Overview
Add All to Overview
Remove All from Overview
Show All Layers Ctrl+Shift+U
Add Vector Layer... Ctrl+Shift+V
Add Raster Layer... Ctrl+Shift+R
Add PostGIS Layers... Ctrl+Shift+D
Add SpatiaLite Layer... Ctrl+Shift+L
Add MSSQL Spatial Layer... Ctrl+Shift+M
Add DB2 Spatial Layer... Ctrl+Shift+2
Add Oracle Spatial Layer... Ctrl+Shift+O
Add WMS/WMTS Layer... Ctrl+Shift+W
Add ArcGIS MapServer Layer...
Add Oracle GeoRaster Layer...
Add WCS Layer...
Add WFS Layer...
Add ArcGIS FeatureServer Layer...
Add Delimited Text Layer...
Add/Edit Virtual Layer...
Layers Panel
Example
MEX_adm
MEX_adm
Roads
Coordinate -99.04492,1
ifier 100%
Rotation 0.0
Render
EPSG:4326

- ➔ Press 'Browse' in the resulting window and find houses_july.csv.



This should cause the window to look like this :



Create a Layer from a Delimited Text File

File Name: C:/Dropbox/QGIS Training/Data/Entomology data/Example_houses_July.csv

Layer name: Example_houses_July

Encoding: UTF-8

File format: ● CSV (comma separated values) ○ Custom delimiters ○ Regular expression delimiter

Record options: Number of header lines to discard 0 ☒ First record has field names

Field options: □ Trim fields □ Discard empty fields □ Decimal separator is comma

Geometry definition: ● Point coordinates ○ Well known text (WKT) ○ No geometry (attribute only table)

X field: Longitude Y field: Latitude □ DMS coordinates

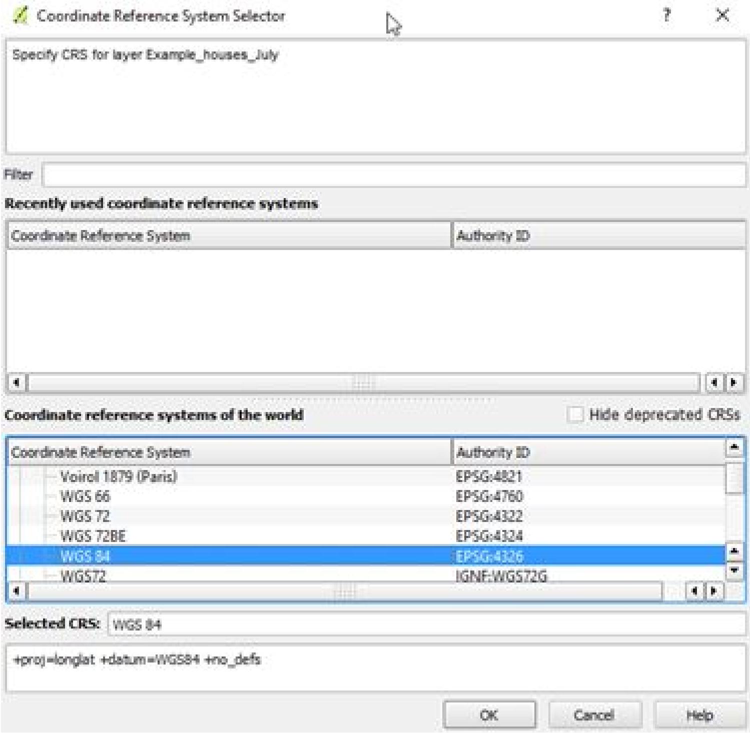
Layer settings: □ Use spatial index □ Use subset index □ Watch file

	UniqueID	Longitude	Latitude	BlockID	HouseID	Intervent	Num_people	Cases	Date	Out_wells	Out_pos_we	O
1	B1H1	-99.04141432	19.33294234	1	1		5	1	01/07/2016	0	0	1
2	B1H2	-99.04111392	19.33295863	1	2		5	0	01/07/2016	0	0	1
3	B1H3	-99.04128625	19.33302373	1	3		6	0	01/07/2016	1	0	1
4	B1H4	-99.04124171	19.33292614	1	4		7	0	01/07/2016	1	1	0
5	B1H5	-99.04135366	19.33305786	1	5		3	0	01/07/2016	1	0	1

The choice boxes half way down labelled ‘X field’ and ‘Y field’ determine which columns are used for the X and Y coordinates of the points. If, as in our case, there are columns called Longitude and Latitude then QGIS will assume that these contain the coordinates. If this doesn’t happen then you can change the choice boxes to select the correct column names.

➔ click OK

This will bring up a 'Coordinate Reference System Selector' (CRS) window. Select the default option of 'WGS 84' by clicking OK. This is the same CRS used in other layers in our project. We will be learning more about Coordinate Reference Systems during Day 3.



Coordinate Reference System Selector

Specify CRS for layer Example_houses_July

Filter

Recently used coordinate reference systems

Coordinate Reference System	Authority ID

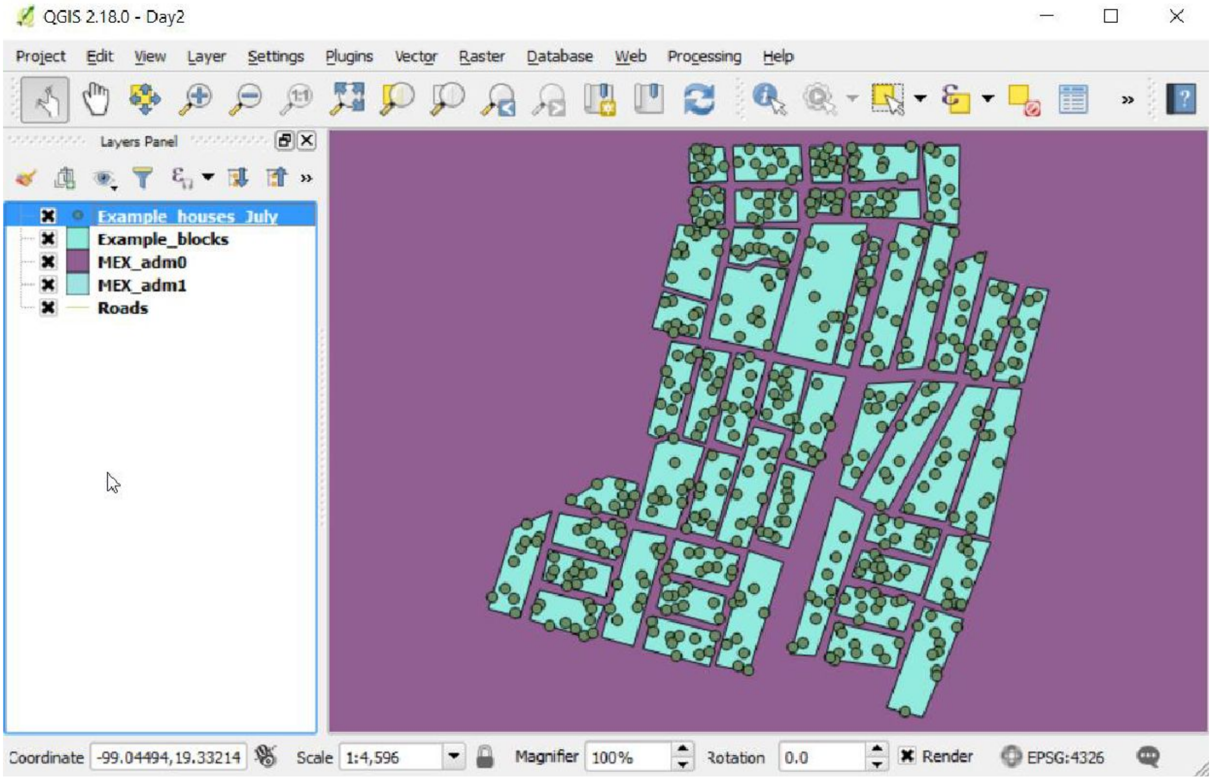
Coordinate reference systems of the world □ Hide deprecated CRSs

Coordinate Reference System	Authority ID
Voirol 1879 (Paris)	EPSG:4821
WGS 66	EPSG:4760
WGS 72	EPSG:4322
WGS 72BE	EPSG:4324
WGS 84	EPSG:4326
WGS72	IGNF:WGS72G

Selected CRS: WGS 84

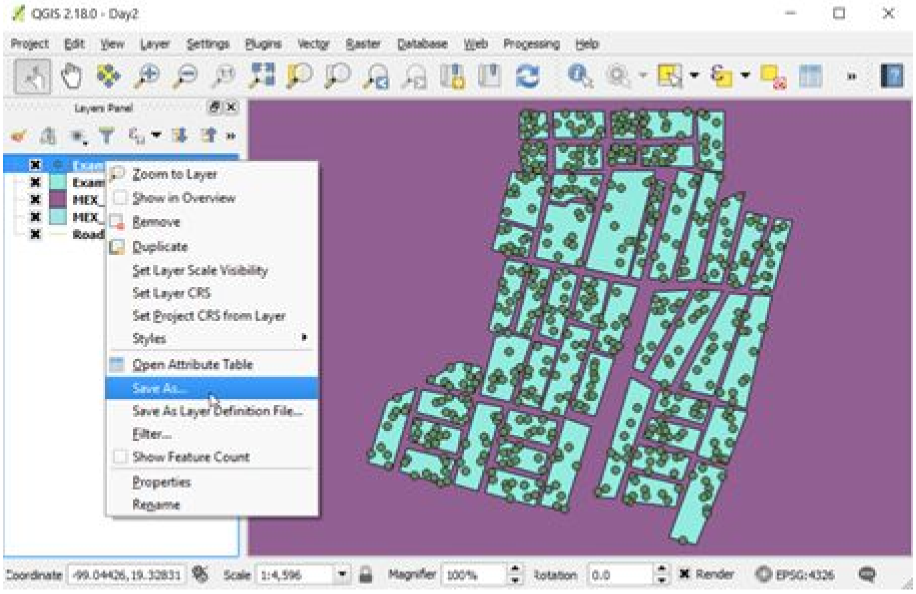
+proj=longlat +datum=WGS84 +no_defs

You should see that the points from the csv file have been added to the map:



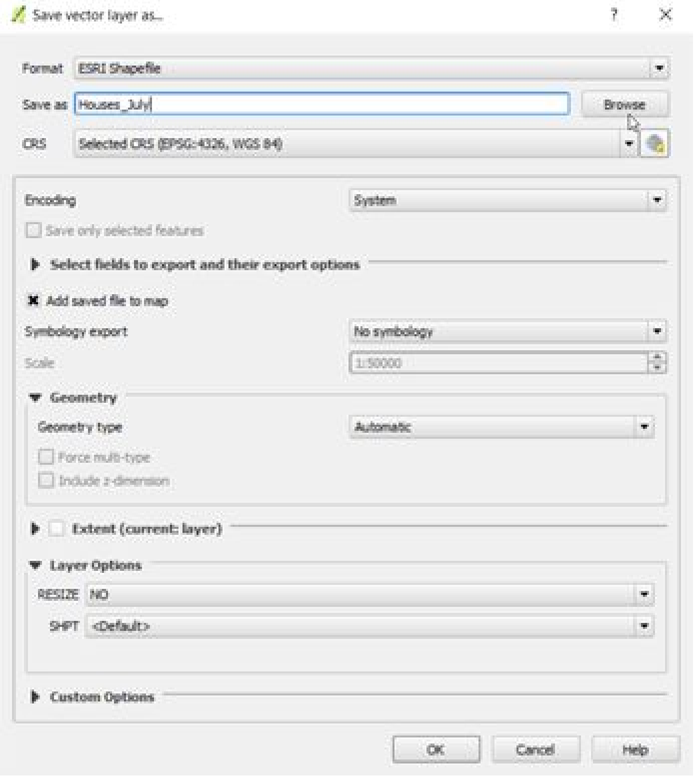
So far we have only created a link between the QGIS project and the csv file. We now need to save the new point layer as a shapefile.

- ➔ Right click on the new layer name in the Layers Panel
- ➔ select 'Save As'



This brings up a window entitled ‘Save vector layer as..’.

- ➔ Make sure selected Format is *ESRI Shapefile* (using dropdown menu at the right if necessary),
- ➔ Type in a file name (houses_july)
- ➔ Browse to my_work/data/points
- ➔ Click OK.



Save vector layer as...

Format: ESRI Shapefile

Save as: Houses_July

CRS: Selected CRS (EPSG:4326, WGS 84)

Encoding: System

□ Save only selected features

Select fields to export and their export options

☒ Add saved file to map

Symbology export: No symbology

Scale: 1:50000

Geometry

Geometry type: Automatic

□ Force multi-type

□ Include z-dimension

□ Extent (current: layer)

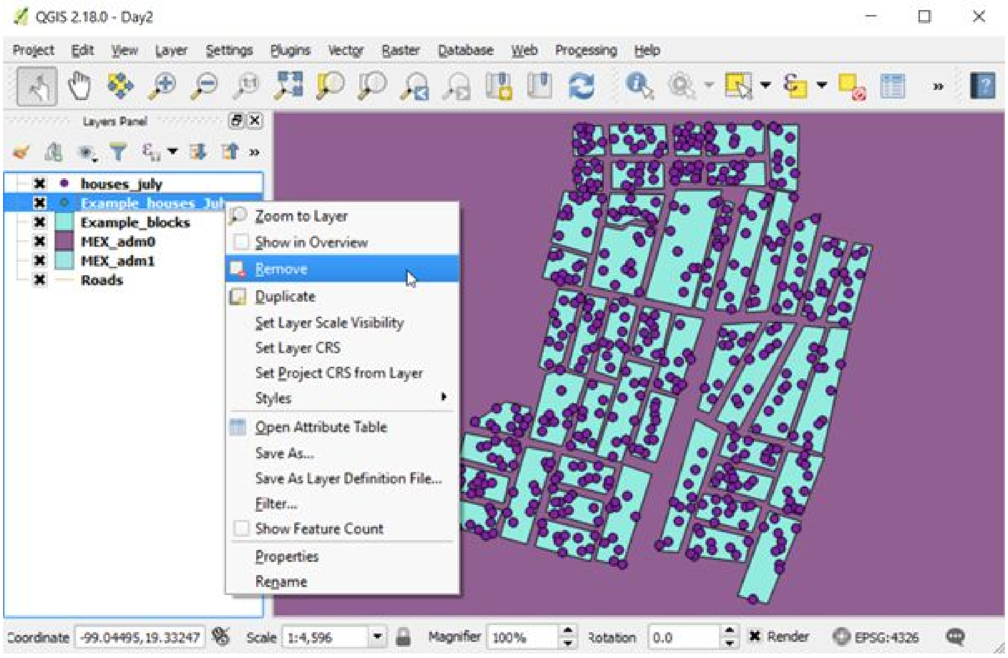
Layer Options

RESIZE: NO

SHPT: <Default>

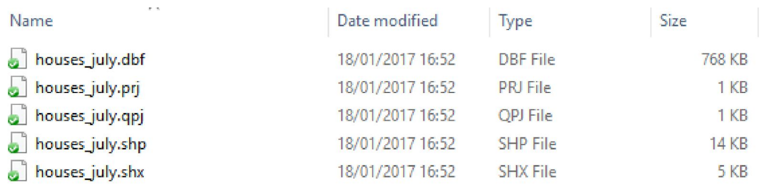
Custom Options

The newly created shapefile is added to the top of the Layers Panel. We can now remove the link to the csv file by Right clicking on it and selecting remove:



Coordinate -99.04495,19.33247 Scale 1:4,596 Magnifier 100% Rotation 0.0

If you look at the folder where you saved the shapefile in Windows File Explorer, you will see that the collection of files that make up a shapefile have been created:



Name	Date modified	Type	Size
houses_july.dbf	18/01/2017 16:52	DBF File	768 KB
houses_july.prj	18/01/2017 16:52	PRJ File	1 KB
houses_july.qpj	18/01/2017 16:52	QPJ File	1 KB
houses_july.shp	18/01/2017 16:52	SHP File	14 KB
houses_july.shx	18/01/2017 16:52	SHX File	5 KB

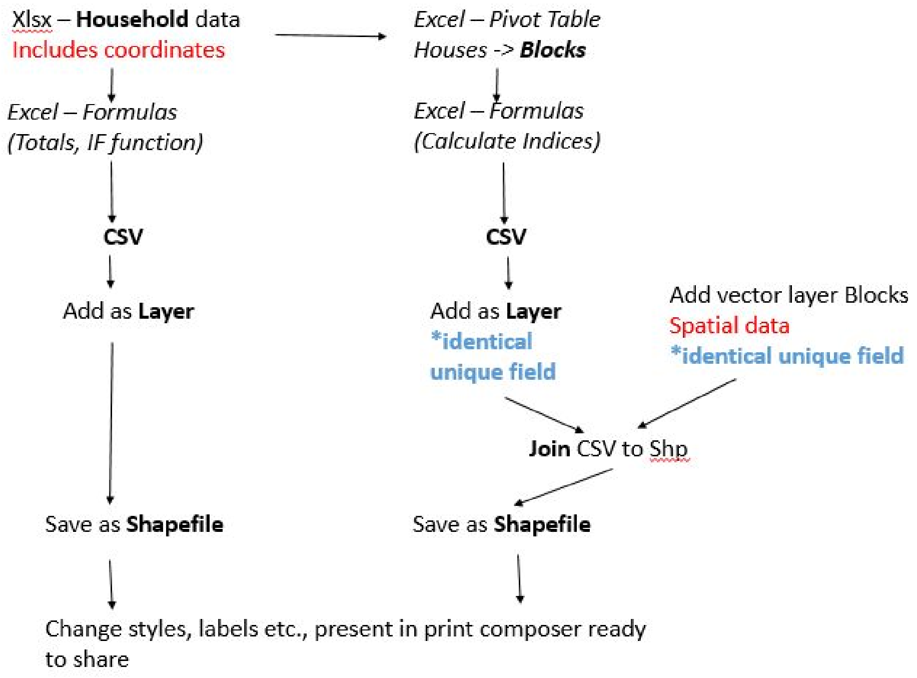
- ➔ Open the attribute table to see your data in QGIS.
- ➔ Try changing the style of your points to show one of your new variables.

To review slightly, so far today you have

- Used formulas in Excel to calculate totals and manipulate your data
- Saved this data in a CSV format ready for import into QGIS
- Added your data as a CSV layer in QGIS
- Saved this layer as a Shapefile, ready for mapping.

The data we used in this example could be mapped directly because included in the original spreadsheet were coordinates – spatial information that QGIS could recognise and use to plot the data as points.

Viewed as a schematic below, we have followed the route on the left-hand-side.



We will now move on to consider the right-hand side of this flow diagram. In the shapefile created so far, each point represents the location of a house, and each of these houses has data associated with it. This data can now be displayed at these locations in a map. Now imagine we want to display this as aggregated data. We want to visualise the situation at block level. If we had a spreadsheet of block level data, we could save this as a csv and add this as a layer in QGIS. In our scenario, we do not! However, we are going to return to Excel to introduce another tool that can be useful in summarising data, Pivot Tables.

Pivot Tables

Pivot tables are a tool in Excel which can be useful for exploring and summarizing your data. What you choose to summarize and how, for example, can be changed very quickly and easily.

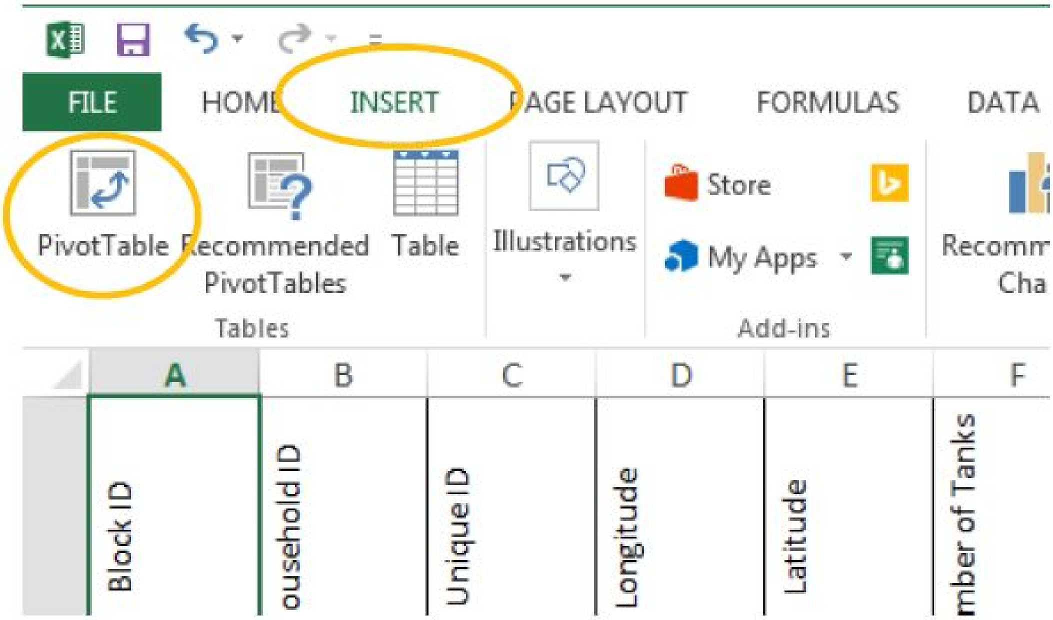
In our example, we want to summarize our data from household level to block level. We are going to look at how a pivot table can be used to do this.

- ➔ Open the Excel sheet “my_household_containers.xlsx” from your my_work folder, if it is not still open from earlier.
- ➔ Open the tab of July data we completed the yellow columns for.

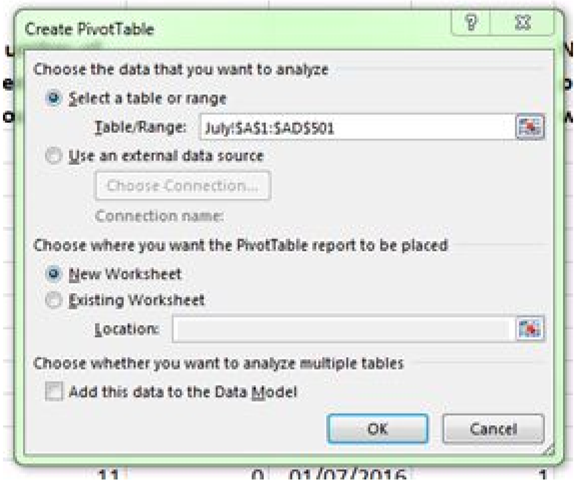
If we look at the first two columns, A and B, we can remind ourselves of the survey design and dataset we are working with. If you recall, a total of 500 households were visited, 50 blocks of 10 houses. The results are currently presented as one row of data per household. The IDs in the first two columns show block numbers 1 to 50, and houses 1 to 10 in each block. If you wanted this data at block level, what would you do? Summarise all 17 columns 50 times, each block separately? This would be quite time consuming. Let’s look at how this can be done in a pivot table.

To create a pivot table of the July data :

- ➔ click on any cell within your worksheet
- ➔ click on “Insert” “PivotTable” from the Menu bar



This will present you with a “Create PivotTable” window. This will give you the option to select which data is included and where the pivot table is placed.



Create PivotTable

Choose the data that you want to analyze

● Select a table or range

Table/Range: July!$A$1:$AD$501

○ Use an external data source

Choose Connection...

Connection name:

Choose where you want the PivotTable report to be placed

● New Worksheet

○ Existing Worksheet

Location:

Choose whether you want to analyze multiple tables

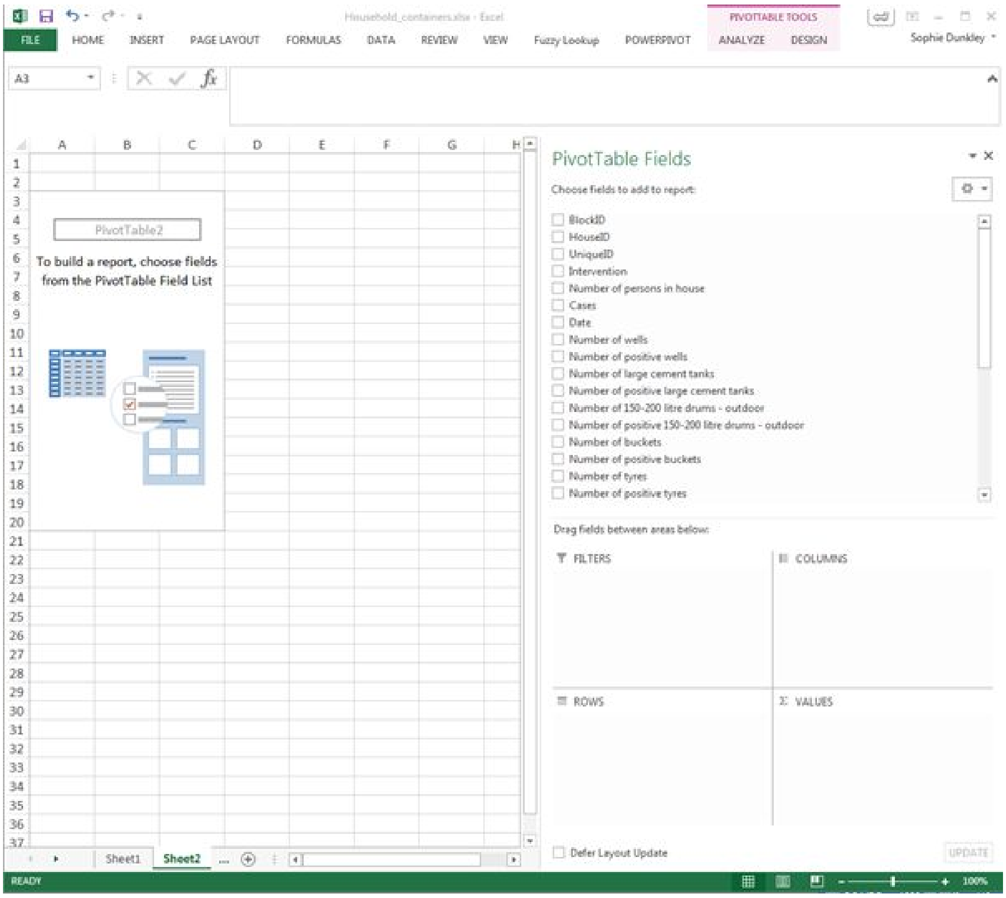
□ Add this data to the Data Model

OK Cancel

Usually your whole table of data will be selected by default. Which data is selected is shown by a dotted line. There is an option to select which data is used, by clicking on the button to the right, and clicking and dragging your mouse over the desired table.

- ➔ Check that the whole table is selected for the data to analyze
- ➔ Make sure that ‘new worksheet’ is selected for where the pivot table will be placed
- ➔ Click OK.

You will be presented with a new worksheet that looks something like this:



PivotTable Fields

Choose fields to add to report:

□ BlockID
□ HouseID
□ UniqueID
□ Intervention
□ Number of persons in house
□ Cases
□ Date
□ Number of wells
□ Number of positive wells
□ Number of large cement tanks
□ Number of positive large cement tanks
□ Number of 150-200 litre drums - outdoor
□ Number of positive 150-200 litre drums - outdoor
□ Number of buckets
□ Number of positive buckets
□ Number of tyres
□ Number of positive tyres

Drag fields between areas below:

FILTERS
COLUMNS
ROWS
VALUES

□ Defer Layout Update
UPDATE

You will find all your variables (all your column headings) listed as selectable options in the "PivotTable Fields" window on the right hand side. These can be added to your table by ticking the check-boxes and/or dragging them (by holding down your left mouse button) into one of the four areas in the bottom left hand side of the window.

➔ Try ticking and un-ticking boxes under 'PivotTable fields'

Note how they appear in the areas below. Also note that the fields you select will appear and disappear from the spreadsheet on the left-hand side.

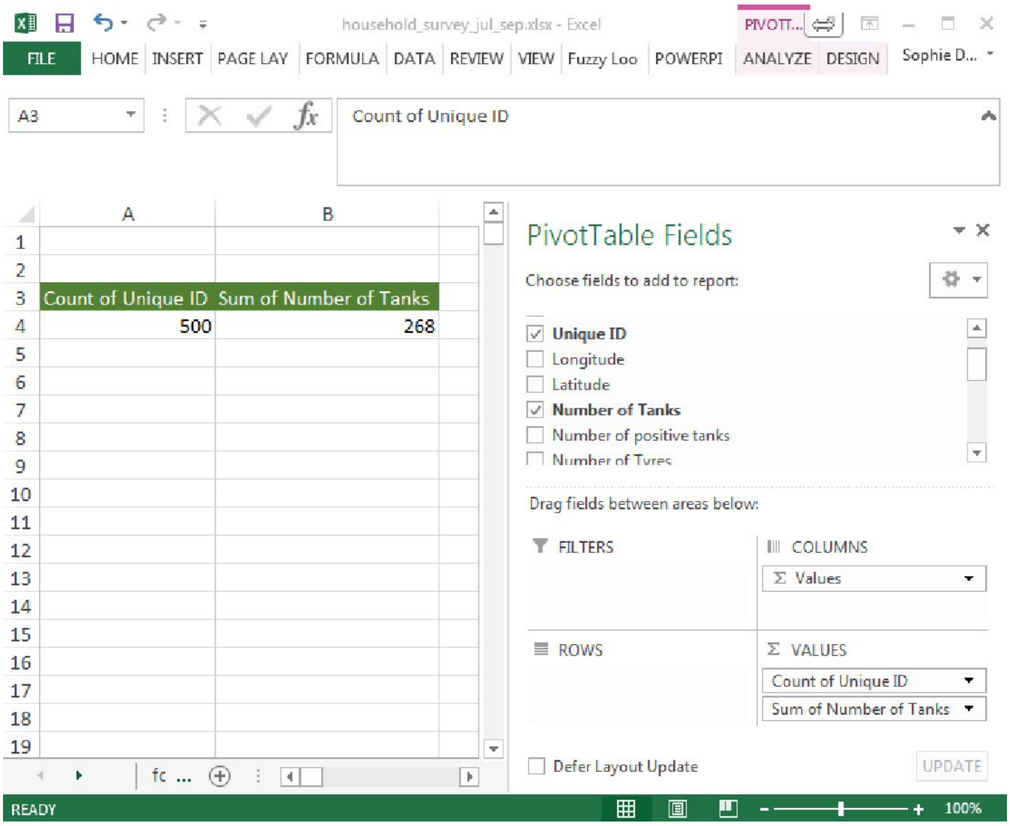
➔ For example, click on "Unique ID", if it appears in the "Rows", you will see a list of each unique household identifier on the left hand side.

➔ Drag the Unique ID field from the Rows area, to Values.

It will change to "Count of Unique ID", and on the left hand side the number 500 should be displayed. It has summarised the data, by counting each ID in the list for our 500 households.

➔ check the box for "Number of tanks".

It appears in the “Values” area as “Sum of Number of tanks” and also on the left, with a number, 268. This number is the total number of tanks for all the data selected, all 500 households.



Count of Unique ID

	A	B
3	Count of Unique ID	Sum of Number of Tanks
4	500	268

PivotTable Fields

Choose fields to add to report:

☑ Unique ID
□ Longitude
□ Latitude
☑ Number of Tanks
□ Number of positive tanks
□ Number of Tyres

Drag fields between areas below:

FILTERS

COLUMNS: Σ Values

ROWS

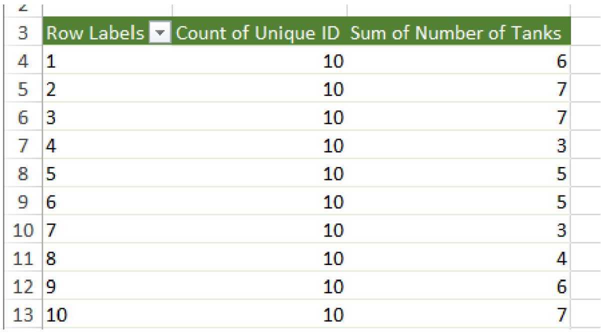
VALUES: Count of Unique ID; Sum of Number of Tanks

□ Defer Layout Update

UPDATE

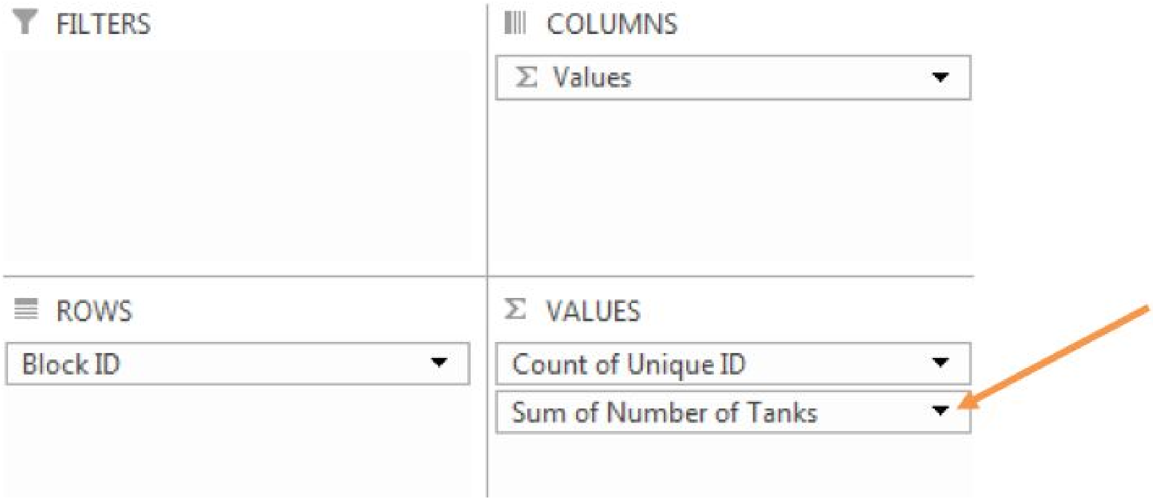
➔ Now, drag “Block ID” into the Rows area.

You should now have 50 rows of data, labelled 1-50, the IDs of the blocks. The Count of Unique ID column is showing that there are 10 households in each block. The Sum of Number of Tanks has now been totalled for each block. If you scroll down you will also see the grand totals of 500 and 268 displayed.



3	Row Labels ▾	Count of Unique ID	Sum of Number of Tanks
4	1	10	6
5	2	10	7
6	3	10	7
7	4	10	3
8	5	10	5
9	6	10	5
10	7	10	3
11	8	10	4
12	9	10	6
13	10	10	7

Note that we can specify how the data in the Values area should be summarised. So far we have used the options “Count” and “Sum”. The small triangles at the right-hand side of the field names in the areas indicate the presence of a menu. Clicking on this triangle or the field in general in one of the four areas will display the menu.



FILTERS

COLUMNS

Σ Values

ROWS

Block ID

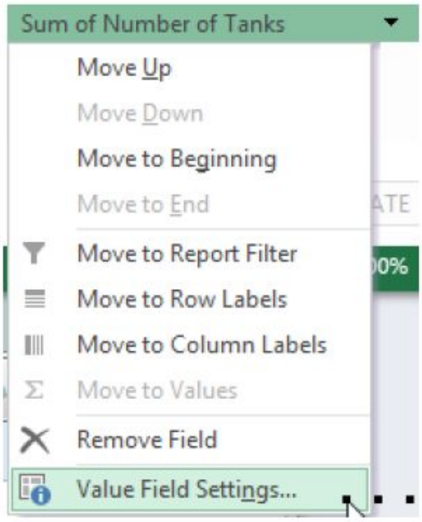
Σ VALUES

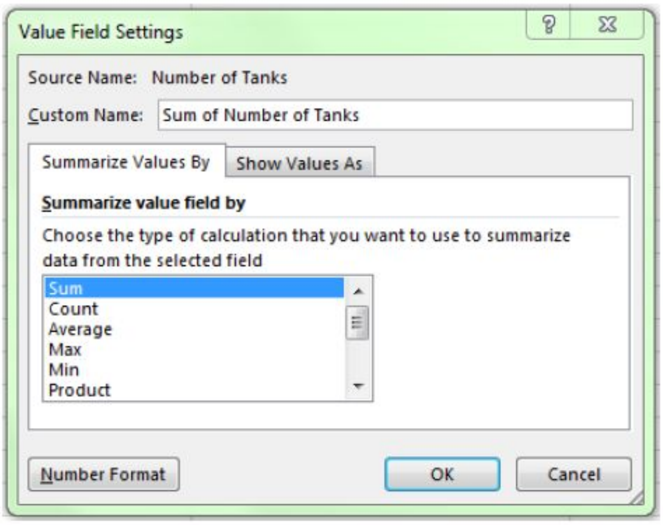
Count of Unique ID

Sum of Number of Tanks

One of the options in this menu is Value Field Settings. This will open up a window that gives us various calculation options to choose from.

- ➔ try changing Sum of Number of Tanks to Average Number of Tanks.





Value Field Settings

Source Name: Number of Tanks

Custom Name: Sum of Number of Tanks

Summarize Values By | Show Values As

Summarize value field by

Choose the type of calculation that you want to use to summarize data from the selected field

Sum
Count
Average
Max
Min
Product

If you look at the data displayed in the table, you can see that for example the total of 6 tanks present in Block 1 have been divided by the 10 households, to give an average (mean) of 0.6 tanks per household in Block 1. Whether this information is useful to you is another question, but you have various options at your disposal.

Dragging options between the Rows and Columns areas will change the organisation of how your data is displayed.

The Filters area allows further exploration of your data by one or more variables. For example we could look at only the data for households that had cases, or only those where larvae were present. We are not going to use this option this time, so will not go into detail on Filters now.

Exercise 2.2

Returning to our survey example, we want to summarise our data by block, and we want to have the information necessary to feed into the calculation of the Stegomyia indices. Use a pivot table to do this for the July example data.

We will work through exercise 2.2 step by step below.

With the July tab open in the my_household_containers Excel file, insert a pivot table in a new worksheet.

We know that we want to summarise the data by block, so we will drag the Block ID field into the Rows area, so that each row in our table will represent a block.

Next we need to think about what data we require to calculate the indices. Here is a reminder of their formulae:

House index (HI): percentage of houses infested with larvae and/or pupae

$$HI = 100 \times \frac{\textit{Number of inspected houses with at least one positive container}}{\textit{Total number of inspected houses}}$$

Container Index (CI): percentage of containers infested with larvae or pupae

$$CI = 100 \times \frac{\textit{Number of positive containers}}{\textit{Total number of inspected containers}}$$

Breteau Index (BI): number of positive containers per 100 houses inspected

$$BI = 100 \times \frac{\textit{Number of positive containers}}{\textit{Total number of inspected houses}}$$

Go through them one by one:

- *Number of inspected houses with at least one positive container*

If you recall, this is why we created the larvae and/or pupae present data. If the total number of positive containers for a household is greater than zero, there will be a 1 in this field, (if not, a 0). Therefore, each “1” is a house with at least one positive container. Check the Larvae and/or pupae present box, and check that the “Sum of” is selected in the Values area to give us the total number of positive houses per block.

- *Total number of inspected houses*

For this, we can use a Count of the Unique IDs which we know that each household inspected has. Same number for HI and BI.

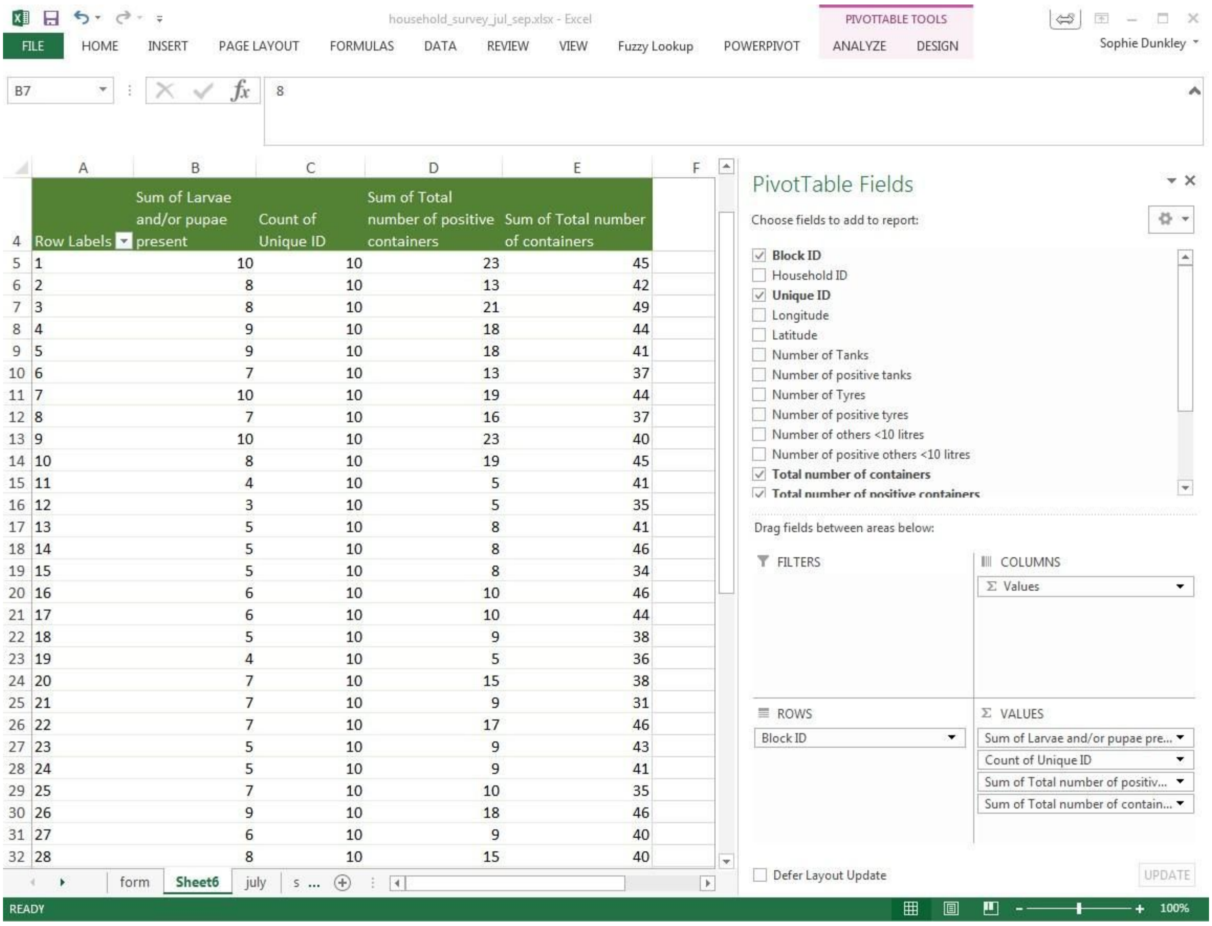
- *Number of positive containers*

Sum of the Total number of positive containers. Same variable for CI and BI.

- *Total number of inspected containers*

Sum of the Total number of containers.

Your selection should look something like this:



Row Labels	Sum of Larvae and/or pupae present	Count of Unique ID	Sum of Total number of positive containers	Sum of Total number of containers
1	10	10	23	45
2	8	10	13	42
3	8	10	21	49
4	9	10	18	44
5	9	10	18	41
6	7	10	13	37
7	10	10	19	44
8	7	10	16	37
9	10	10	23	40
10	8	10	19	45
11	4	10	5	41
12	3	10	5	35
13	5	10	8	41
14	5	10	8	46
15	5	10	8	34
16	6	10	10	46
17	6	10	10	44
18	5	10	9	38
19	4	10	5	36
20	7	10	15	38
21	7	10	9	31
22	7	10	17	46
23	5	10	9	43
24	5	10	9	41
25	7	10	10	35
26	9	10	18	46
27	6	10	9	40
28	8	10	15	40

PivotTable Fields

Choose fields to add to report:

☑ Block ID
☐ Household ID
☑ Unique ID
☐ Longitude
☐ Latitude
☐ Number of Tanks
☐ Number of positive tanks
☐ Number of Tyres
☐ Number of positive tyres
☐ Number of others <10 litres
☐ Number of positive others <10 litres
☑ Total number of containers
☑ Total number of positive containers

Drag fields between areas below:

FILTERS

COLUMNS: Σ Values

ROWS: Block ID

VALUES: Sum of Larvae and/or pupae pre... / Count of Unique ID / Sum of Total number of positiv... / Sum of Total number of contain...

☐ Defer Layout Update

These are the values required to calculate the indices, summarised by block.

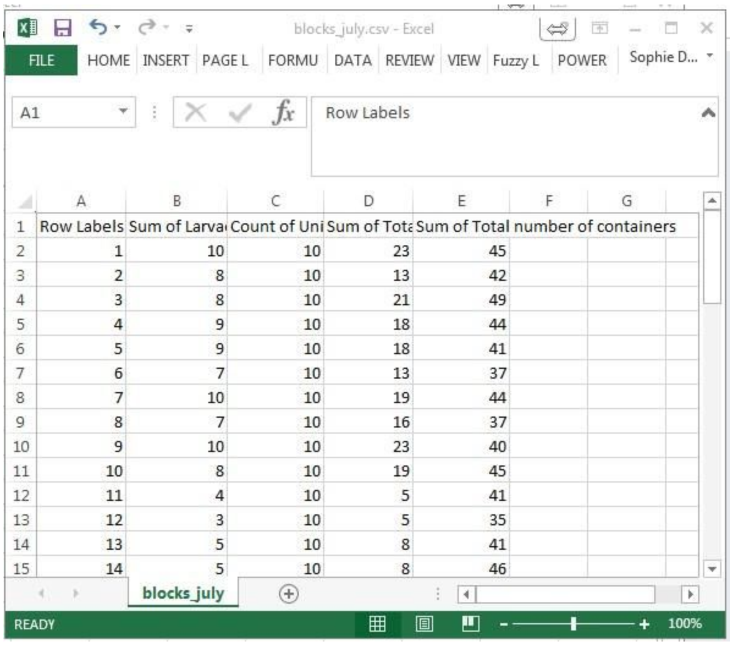
➔ Click Save to ensure the changes to your workbook are recorded.

Data Format – Saving as a CSV file

For our example, we only wanted to use the pivot table to summarise the data to block level. Now this is done, we can save the data as a CSV file (as before).

➔ With your pivot table datasheet open, select ‘File/Save As’

- ➔ Navigate to the folder ‘my_work/data/entomology’
- ➔ Name the new file “blocks_july”
- ➔ Use the drop down menu of ‘Save as type’ to select “CSV (Comma delimited) (*.csv)”
- ➔ Click Save. You will get the same warning messages as before. Click through them as previously.
- ➔ Close your spreadsheet using the X button at the top right of the Excel window. You will get the warning box again, press 'Don't save' because we have saved the file in the CSV format already.
- ➔ Reopen the csv file my_work/data/entomology/blocks_july.csv. Your file should open in Excel, and look similar to the screenshot below:



	A	B	C	D	E	F	G
1	Row Labels	Sum of Larva	Count of Uni	Sum of Tota	Sum of Total number of containers		
2	1	10	10	23	45		
3	2	8	10	13	42		
4	3	8	10	21	49		
5	4	9	10	18	44		
6	5	9	10	18	41		
7	6	7	10	13	37		
8	7	10	10	19	44		
9	8	7	10	16	37		
10	9	10	10	23	40		
11	10	8	10	19	45		
12	11	4	10	5	41		
13	12	3	10	5	35		
14	13	5	10	8	41		
15	14	5	10	8	46		

- ➔ If you have a blank row at the top of the spreadsheet, you will need to delete it.

The next step for us is to shorten the variable names in preparation for import to QGIS, as before.

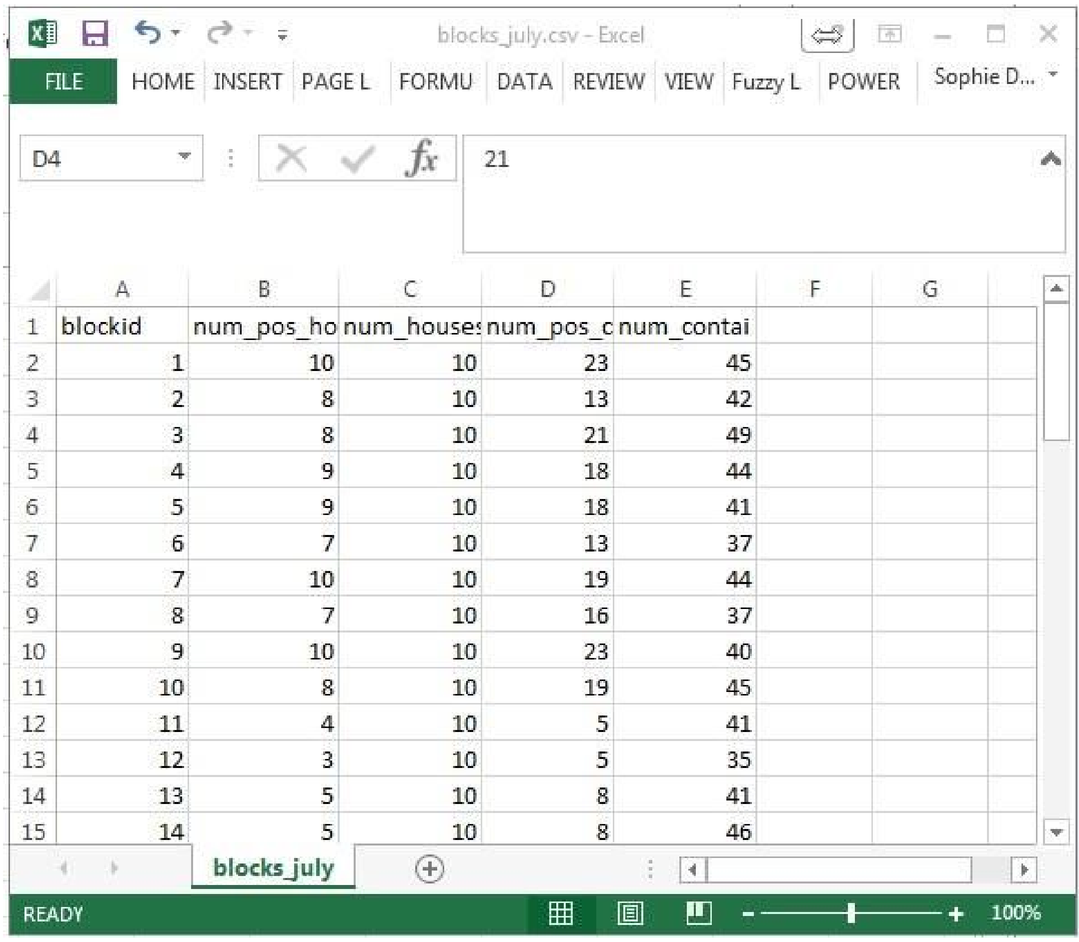
Note additionally that the pivot table tool automatically changes or adds to column headings making the names even longer or lose their meaning along the way (for example “Row labels” which we know in this case are the block IDs, or “Sum of Total number of containers”). The first ten characters of the variables: “Sum of Total number of positive containers” and “Sum of Total number of containers”, would be exactly the same if cut automatically to 10 characters: “Sum_of_Tot”. Therefore, if we change these column names to short, meaningful names now we will avoid this problem in the future.

- ➔ Rename your columns based on the suggestions in the table below.

Note that the order of your columns may be different, depending on the order of your variables in your pivot table. It is important to name your variables correctly and make sure you verify that you are using the correct one and not just the cell reference.

Variable name – Used in Excel spreadsheet Pivot Table	**Short name – used in csv files**	**Description**
Block ID	blockid	Block ID, ranging from 1-50
Count of Household ID	num_houses	The total number of houses surveyed in each block
Sum of Total number of containers	num_contai	The total number of containers in each block
Sum of Total number of positive containers	num_pos_co	The total number of larvae/pupae positive containers in each block
Sum of Number of houses with larvae/pupae present	num_pos_ho	The total number of houses in the block with larvae/pupae present
House index (HI): percentage of houses infested with larvae and/or pupae	hi	Household index = number of positive houses[num_pos_ho]/total number of houses[num_houses]
Container index (CI): percentage of water-holding containers infested with larvae or pupae	ci	Container index = number of positive containers[num_pos_co]/number of containers[num_contai]
Breteau index (BI): number of positive containers per 100 houses inspected.	bi	Breteau index = 100*Number of positive containers[num_pos_co]/Number of houses[num_houses]

Your CSV file should now look similar to the one below.



D4 | 21

	A	B	C	D	E	F	G
1	blockid	num_pos_ho	num_houses	num_pos_c	num_contai		
2	1	10	10	23	45		
3	2	8	10	13	42		
4	3	8	10	21	49		
5	4	9	10	18	44		
6	5	9	10	18	41		
7	6	7	10	13	37		
8	7	10	10	19	44		
9	8	7	10	16	37		
10	9	10	10	23	40		
11	10	8	10	19	45		
12	11	4	10	5	41		
13	12	3	10	5	35		
14	13	5	10	8	41		
15	14	5	10	8	46		

We are now ready to calculate the Stegomyia indices. To do this, we return to using formulae.

Calculating the Stegomyia indices

We shall start by adding the column headings for the three indices to your table i.e. 'hi', 'ci' and 'bi'. We will then use formulae to calculate each of the Stegomyia indices. These can be typed directly into the cells and copied for the rest of the column as we have done previously.

A few new things to note:

- The symbol for multiply in an Excel formula is an asterisk *
- To divide you use /
- Brackets ( ) can be used to clarify the order of actions in the calculation. The part of the sum in brackets will be carried out first.
- To use a number rather than a cell reference (e.g. to multiply by 100), simply type the number into the equation.
- Pay attention to your variable names, rather than the cell references, to make sure you use the correct variable.

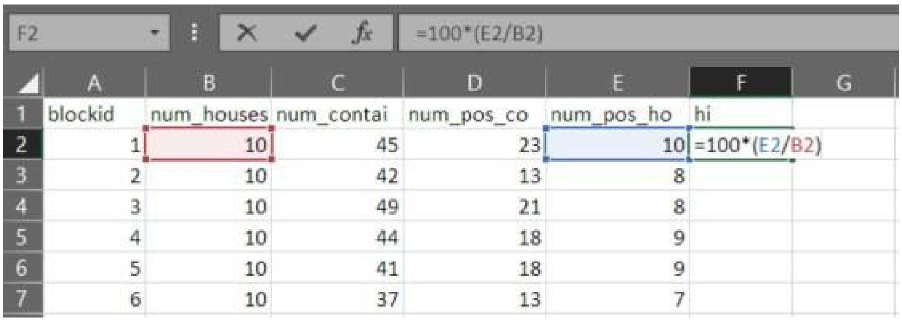
So, for the House Index:

$$\textit{Number of inspected houses with at least one positive container}$$

$$HI = 100 \times \frac{}{Total\ number\ of\ inspected\ houses}$$

- ➔ Type **=100*E2/B2’**

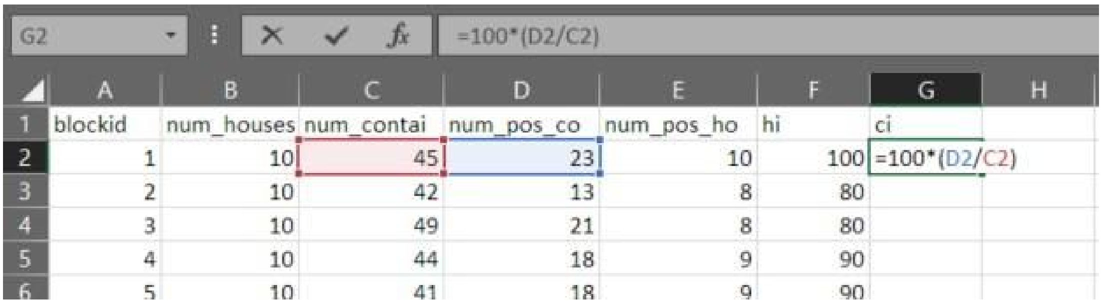
E2 in this example, is the number of positive houses, and B2 is the number of houses inspected.


F2 =100*(E2/B2)

	A	B	C	D	E	F	G
1	blockid	num_houses	num_contai	num_pos_co	num_pos_ho	hi	
2	1	10	45	23	10	=100*(E2/B2)	
3	2	10	42	13	8		
4	3	10	49	21	8		
5	4	10	44	18	9		
6	5	10	41	18	9		
7	6	10	37	13	7		

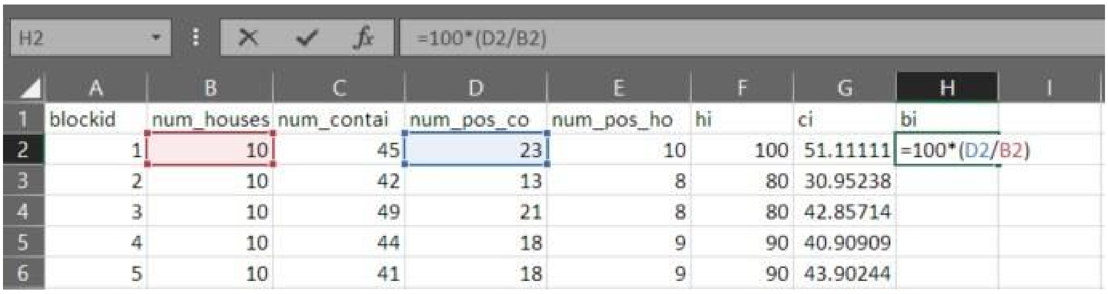
- ➔ Press Enter, then double-click the small box at the bottom of the newly filled cell the fill in the rest of the column with the formula.
- ➔ Repeat for the container index and Breteau index.

For the container index, the formula will be something like: =100*(D2/C2)


G2 =100*(D2/C2)

	A	B	C	D	E	F	G	H
1	blockid	num_houses	num_contai	num_pos_co	num_pos_ho	hi	ci	
2	1	10	45	23	10	100	=100*(D2/C2)	
3	2	10	42	13	8	80		
4	3	10	49	21	8	80		
5	4	10	44	18	9	90		
6	5	10	41	18	9	90		

For the Breteau index: =100*(D2/B2)


H2 =100*(D2/B2)

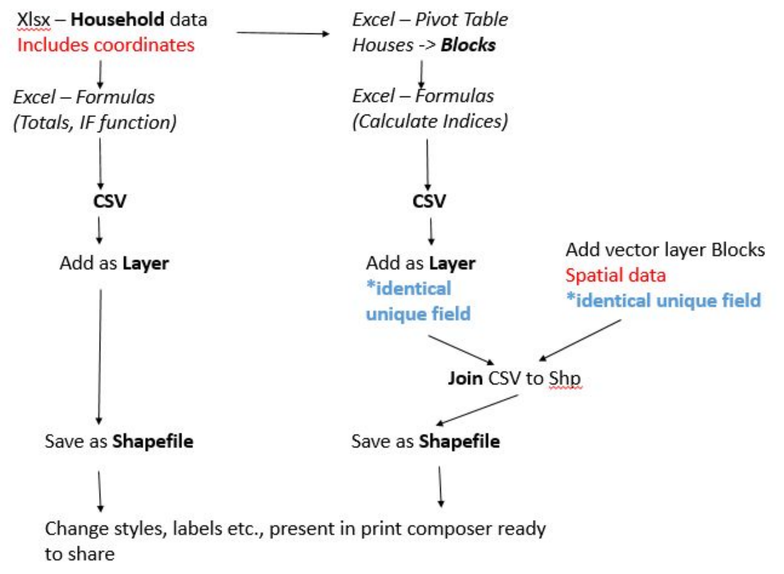
	A	B	C	D	E	F	G	H	I
1	blockid	num_houses	num_contai	num_pos_co	num_pos_ho	hi	ci	bi	
2	1	10	45	23	10	100	51.11111	=100*(D2/B2)	
3	2	10	42	13	8	80	30.95238		
4	3	10	49	21	8	80	42.85714		
5	4	10	44	18	9	90	40.90909		
6	5	10	41	18	9	90	43.90244		

- ➔ Once all three columns are complete, save and close the file in preparation for loading it into QGIS.

Exercise 2.3 (Optional)

If you wish to repeat the whole process for practice, the September data is available in the Excel file household_survey_ju_sep.xlsx in your USB drives under data/entomology. Create formulae to fill in the total and presence/absence of larvae in the yellow columns; use a pivot table to summarise the data from household level to blocks, save it as a CSV, create further formulae to calculate the indices.

Let’s review where we are. Remember our flow diagram.



You have used pivot tables to summarise data from household to block level.

You have used formulas again, this time to calculate indices.

You have saved your data as a CSV ready to add to QGIS.

Consider the block data you have created. Does it have any spatial data? Is there anything that tells QGIS how to plot this information? To map this block level data, we are going to look at joining data from a CSV to a shapefile.

Join data from a CSV to an existing shapefile

In the previous section we added a point layer from a csv file and saved that as a shapefile. Now we will add the block-level csv data. To do this we are going to use a slightly different approach, called ‘joining’. The joining process involved linking two or more datasets using a unique identification variable that needs to be present in all datasets being joined. For example, suppose I have two spreadsheets as below. One spreadsheet has two columns named country and climate, the other has two columns named nation and favourite sport (this is a very simple example and you may not agree with the data!).

country	climate
UK	cold
Colombia	hot
USA	mixed

nation	favourite sport
Colombia	football
USA	baseball
UK	football

If we join these two spreadsheets using the first column as the unique joining variable, then we obtain a single spreadsheet as seen below.

country	climate	favourite sport
UK	cold	football
Colombia	hot	football
USA	mixed	baseball

There are a number of points to note:

~ the joining columns do not have to have the same name

~ the rows do not have to be in the same order

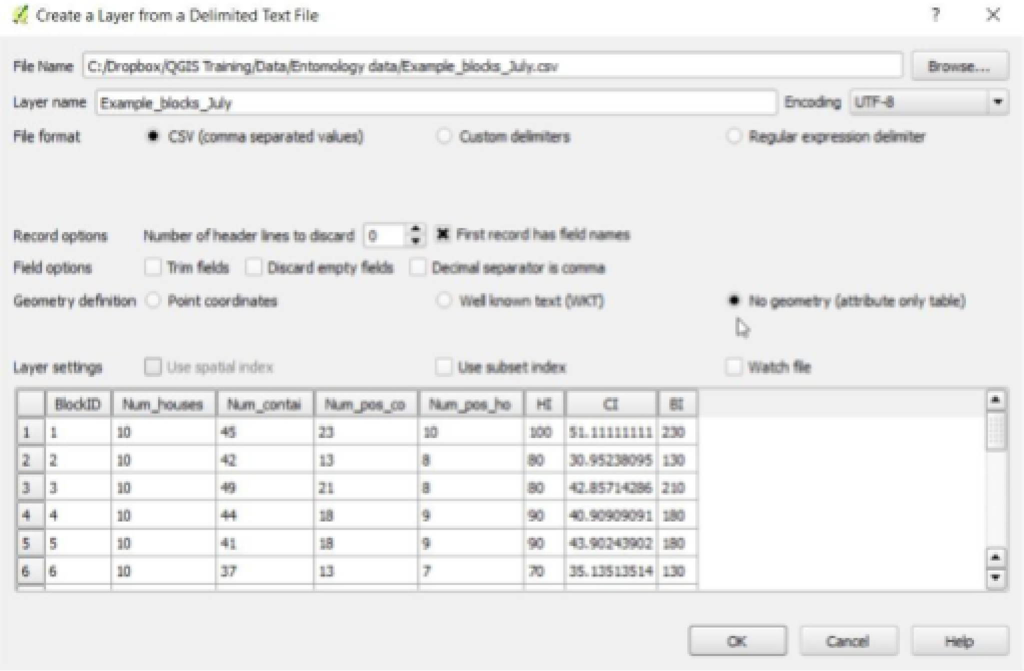
~ the data columns can contain repeated values (e.g. football in this example)

~ the join columns (country and nation in this case) must not contain repeated values ~ only values that are identical will be joined

We can do the same thing with spatial data. For example, we have a block polygons file 'blocks.shp' which contains spatial information but no vector surveillance information, and we have a CSV file 'blocks_july.csv' which contains vector surveillance information but no spatial information. Both data sources however contain a unique block ID variable which allows us to ‘join’ the spatial information and the vector surveillance information into one file. Note, that when joining two or more sources, the joining variable does not necessarily have to have the same column name, but must contain at least some of the same values.

To join the block polygons with the blocks_july.csv, we first need to add the CSV file to the project.

➔ As before, from the Menu bar select: Layer, Add Layer, Add Delimited Text Layer.
➔ Browse to the blocks csv file you created earlier.
➔ There are no spatial coordinates contained in this CSV file, so under the options for ‘Geometry definition’, select the 'No geometry (attribute only table)' button on the right.
➔ click OK.



Create a Layer from a Delimited Text File

File Name: C:/Dropbox/QGIS Training/Data/Entomology data/Example_blocks_July.csv

Layer name: Example_blocks_July

Encoding: UTF-8

File format: ● CSV (comma separated values) ○ Custom delimiters ○ Regular expression delimiter

Record options: Number of header lines to discard 0 ☒ First record has field names

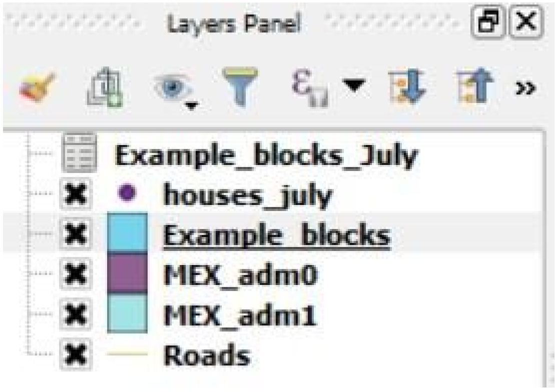
Field options: □ Trim fields □ Discard empty fields □ Decimal separator is comma

Geometry definition: ○ Point coordinates ○ Well known text (WKT) ● No geometry (attribute only table)

Layer settings: □ Use spatial index □ Use subset index □ Watch file

	BlockID	Num_houses	Num_contai	Num_pos_co	Num_pos_ho	HI	CI	BI
1	1	10	45	23	10	100	51.11111111	230
2	2	10	42	13	8	80	30.95238095	130
3	3	10	49	21	8	80	42.85714286	210
4	4	10	44	18	9	90	40.90909091	180
5	5	10	41	18	9	90	43.90243902	180
6	6	10	37	13	7	70	35.13513514	130

You should see that the new layer is added to the Layers Panel at the top with a different icon.

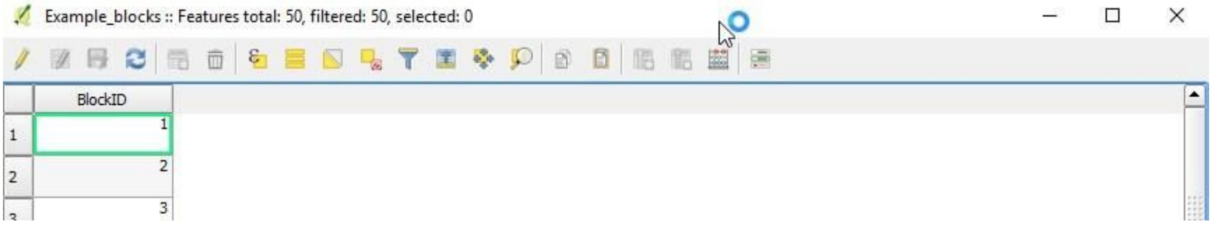


To join the csv block data onto the block layer, we must identify the column in each that contains the ID of the features we want to join the two data sources with.

QGIS calls the join field from the shapefile the ‘Target Field,’ and the join field from the csv layer the ‘Join field.’

➔ Open the attribute tables for the two blocks layers (right click, select Open Attribute Table)

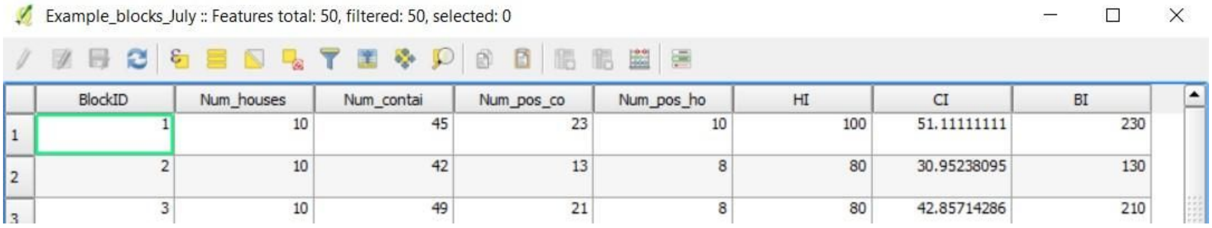
You should see that the shapefile layer has one column named BlockID,



Example_blocks :: Features total: 50, filtered: 50, selected: 0

	BlockID
1	1
2	2
3	3

and the layer from the csv also has a BlockID column and several others :

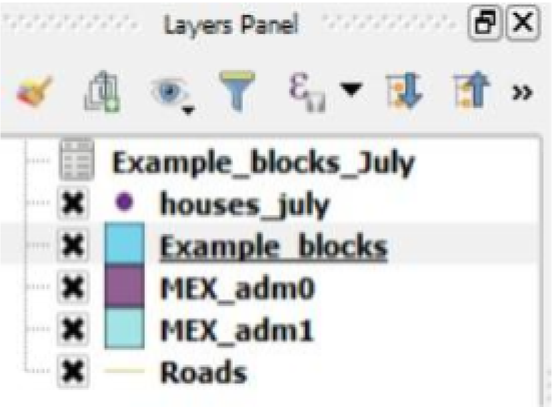


Example_blocks_July :: Features total: 50, filtered: 50, selected: 0

	BlockID	Num_houses	Num_contai	Num_pos_co	Num_pos_ho	HI	CI	BI
1	1	10	45	23	10	100	51.11111111	230
2	2	10	42	13	8	80	30.95238095	130
3	3	10	49	21	8	80	42.85714286	210

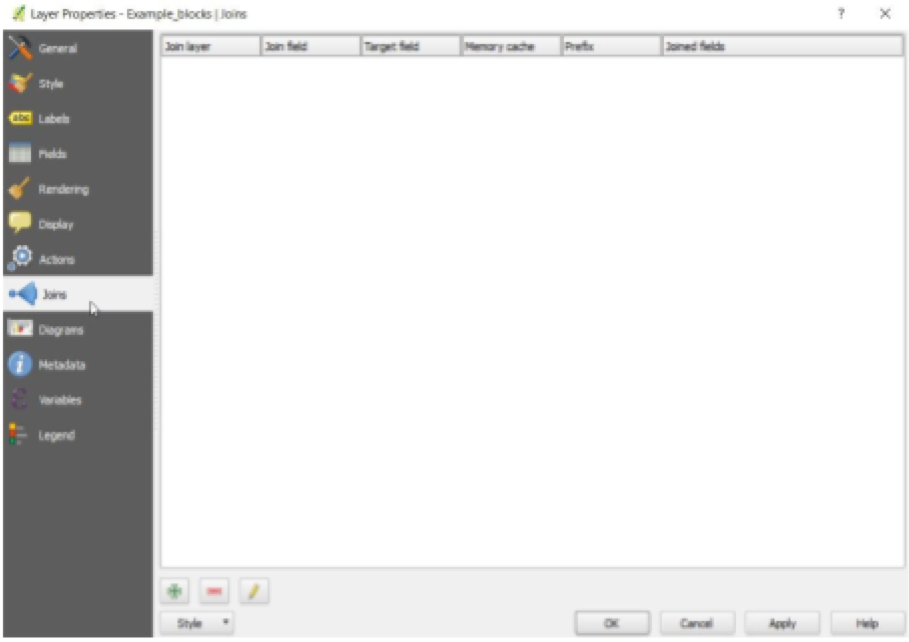
We can now proceed to joining the two datasets using the BlockID variable. Go to Properties for the blocks shapefile layer.

- ➔ Right click on ‘Example_blocks’ in the Layers Panel
- ➔ select Properties



Example_blocks_July
houses_july
Example_blocks
MEX_adm0
MEX_adm1
Roads

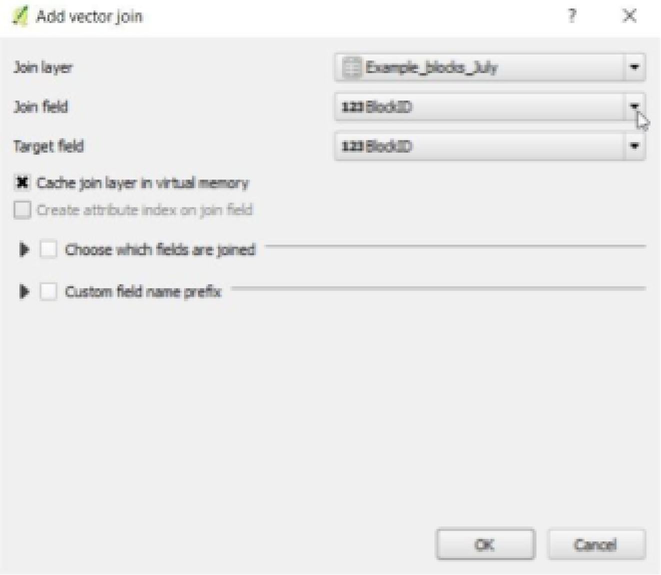
- ➔ From the Layer Properties window, select 'Joins' on the left hand side
- ➔ click the green '+' button lower down to add a join.



- ➔ In the Add vector join window that appears, select the CSV layer *blocks_july* as the join layer. Note that this may be automatically selected.

QGIS also automatically selects possible Join fields and Target fields based on the names of the columns in both the shapefile layer and the selected join layer.

- ➔ check that the selected Join and Target fields are 'BlockID'



Join layer: Example_blocks_July

Join field: 123 BlockID

Target field: 123 BlockID

☒ Cache join layer in virtual memory

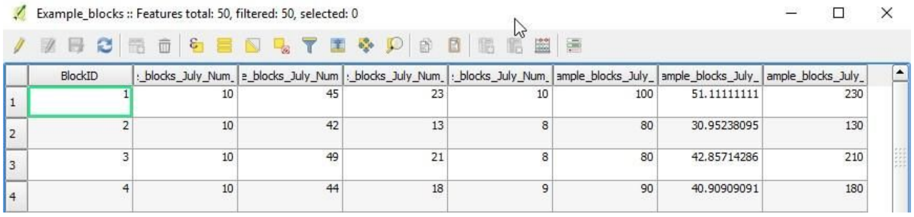
☐ Create attribute index on join field

☐ Choose which fields are joined

☐ Custom field name prefix

- ➔ Click OK until all windows are closed.
- ➔ Right click on the blocks layer within the Layers Panel and open the Attribute table.

You should see that the CSV data has been added into the Attribute table of the block polygon layer.



Example_blocks :: Features total: 50, filtered: 50, selected: 0

	BlockID	:_blocks_July_Num_	e_blocks_July_Num	:_blocks_July_Num_	:_blocks_July_Num_	ample_blocks_July_	ample_blocks_July_	ample_blocks_July_
1	1	10	45	23	10	100	51.11111111	230
2	2	10	42	13	8	80	30.95238095	130
3	3	10	49	21	8	80	42.85714286	210
4	4	10	44	18	9	90	40.90909091	180

As we noted earlier, the data displayed in the attribute table have not been added into the shapefile. Instead, there is just a link between the shapefile and the csv. Hence, if you delete the csv file the join would be broken.

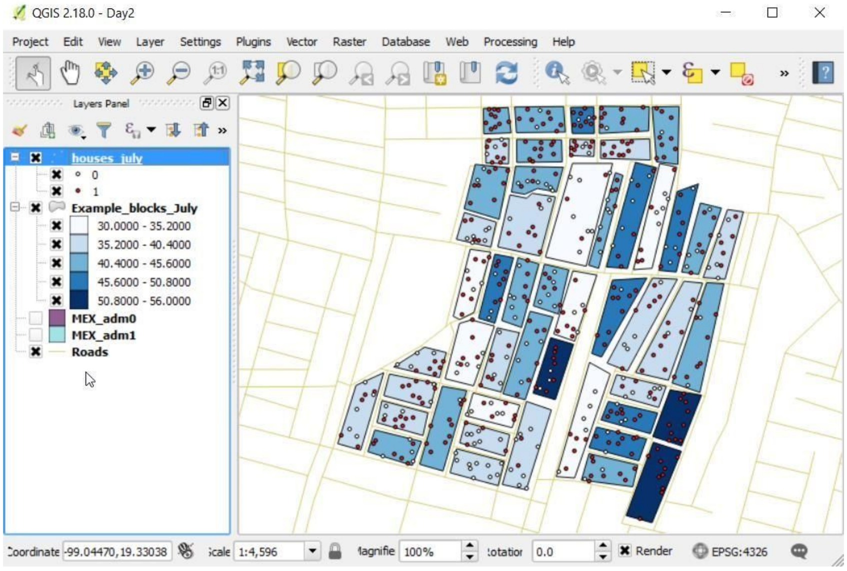
- ➔ To save the joined data into a new shapefile, right click on the layer name 'blocks' in the Layers panel and choose 'Save as'
- ➔ Save as an ESRI Shapefile named 'blocks_july' in ‘data/spatial/polygons/’

- ➔ Remove un-needed layers (blocks_july.csv and blocks) by right clicking on their names in Layers Panel and select 'Remove' as we did before.

Now we can make the map more useful using the techniques we learnt yesterday. Try going through the steps below as a reminder of yesterday’s material, look back at yesterday’s manual to help you.

- ➔ Right click on the name of the blocks layer and open its attribute table. Decide on a column that you would like to see displayed on the map (i.e. the values will be used to alter the appearance of each block). Hint: decide on one where there is variation in the values.
- ➔ Right click on the layer name again, open 'Properties' and select 'Style'.
- ➔ Choose 'Graduated' as the top option and set the column to the one you decided earlier.
- ➔ Choose a different 'Color ramp' if you like.
- ➔ Click Classify & OK
- ➔ Repeat this process for the houses layer, look at the attribute table, decide on a column and use that column to set the style of the point data map.
- ➔ untick the MEX_adm layers which don't contribute to the map at this zoom level.

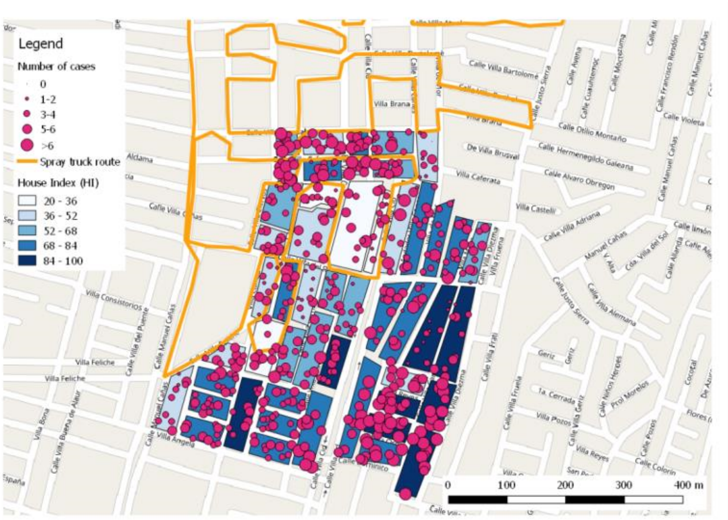
You should have a map that looks similar to the one below. Note that your map may look slightly different according to the symbology choices you have made.



➔ Save your project by Project, Save from the Menu bar (or you can use Ctrl S from the keyboard).

Exercise 2.4

In August, a spray truck sprayed part of the area covered by the example entomological survey. You can find the route the truck took in the *Lines* subfolder (spray_truck.shp). An entomological survey was then repeated in each of the 10 households per block, and the results are recorded in the Excel spreadsheet household_containers.xlsx in the *September* tab. Repeat the steps above to create a map of September's entomology survey data to visualise whether the spray truck has had any impact. Your map should look similar to the one below.



End of Day 2