

Practical 9

AIM: Validating Map data

Solution:

Software Used: QGIS Desktop 3.4.2 and Microsoft Excel 2013.

Datasets Used: The following datasets are used

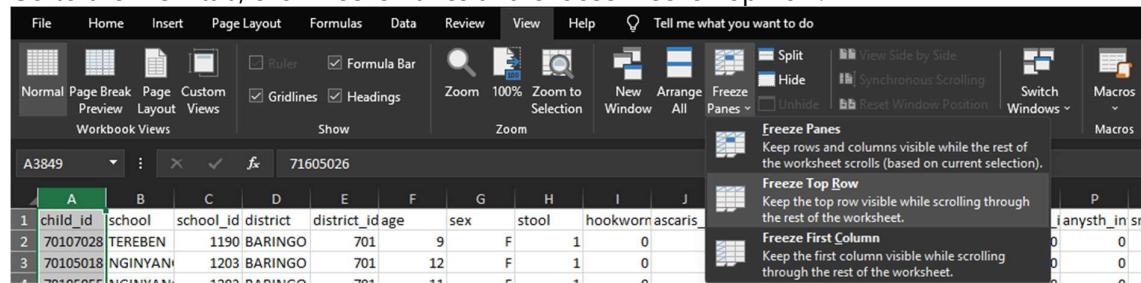
1. Kenya_epidemiological_data.xls
2. Kenya_epidemiological_dict.xls

Shape Files used: Kenya admin.

Format of the Database

Open Kenya_epidemiological_data.xls in excel.

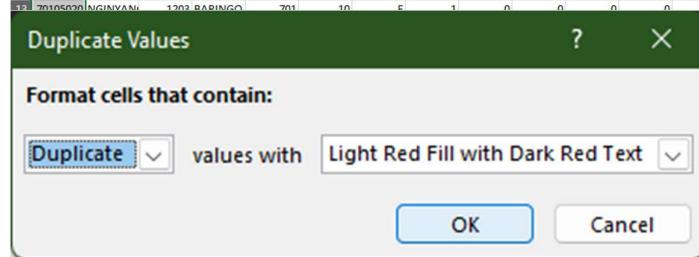
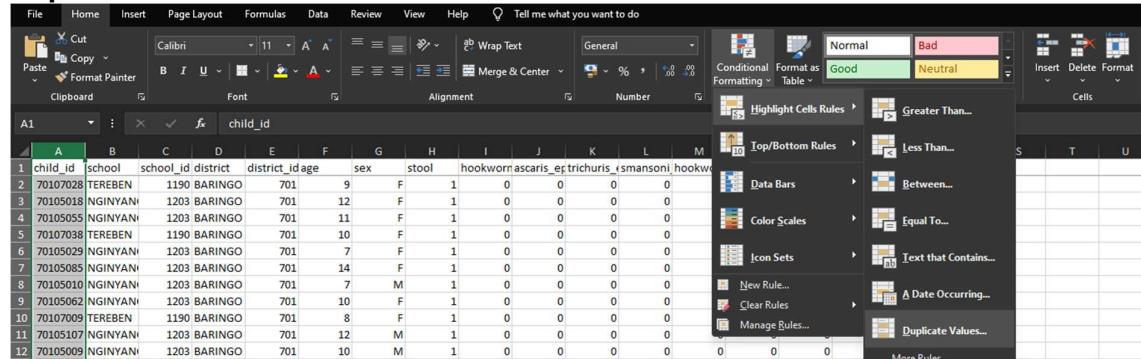
Go to the View tab, click Freeze Panes and choose Freeze Top Row.



The screenshot shows the Microsoft Excel ribbon with the 'View' tab selected. In the 'View' tab, the 'Freeze Panes' button is highlighted. A dropdown menu is open from this button, displaying three options: 'Freeze Panes', 'Freeze Top Row', and 'Freeze First Column'. The 'Freeze Top Row' option is currently selected. The main Excel interface shows a table with columns labeled A through J and rows numbered 1 through 12. The first column, 'child_id', contains unique identifiers for each row.

select the entire "child_id" column (first column),

Under Home Tab, click on **Conditional formatting >Highlight Cell Rules> Select Duplicate values**



The screenshot shows the Microsoft Excel ribbon with the 'Home' tab selected. The 'Conditional Formatting' button is highlighted, and a dropdown menu is open. The 'Highlight Cell Rules' option is selected, and the 'Duplicate Values...' option is highlighted. Below this, the 'Duplicate Values' dialog box is open, showing the condition 'Format cells that contain: Duplicate values with Light Red Fill with Dark Red Text'. There are 'OK' and 'Cancel' buttons at the bottom of the dialog box.

Removing Duplicates

Select all the columns of existing worksheet

Now go to Data Tab and select Remove Duplicates

The screenshot shows a Microsoft Excel spreadsheet titled "Kenya_epidemiological_data". The "Data" tab is selected. A "Remove Duplicates" dialog box is open, centered over the spreadsheet. The dialog box has a checkbox "My data has headers" which is checked. Below it, a list of columns is shown: child_id, school, school_id, district, district_id, age, sex, stool, hookworm, ascaris_eg, trichuris_smansoni, hookworm, ascaris_in_trichuris, anysth_in_smansoni, dewormed. At the bottom of the dialog box are "OK" and "Cancel" buttons.

Coding of variables

In the current worksheet, select the sex column.

Now type Ctrl+F and use Replace Function and Replace as follows

"M" to 1

"F" to 2

Please keep track of how many values are getting replace.

The screenshot shows a Microsoft Excel spreadsheet with the ribbon visible. The "Edit" tab is selected, and the "Replace..." option is highlighted under the "Cells" section of the ribbon. The main spreadsheet area shows the same data as the previous screenshot, with the "sex" column selected. The data includes columns for child_id, school, school_id, district, district_id, age, sex, stool, hookworm, ascaris_eg, trichuris_smansoni, hookworm, ascaris_in_trichuris, anysth_in_smansoni, and dewormed.

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Cut Copy Format Painter Clipboard Font Alignment Number Styles

G1 sex

child_id	school	school_id	district	district_id	age	sex	stool	hookworm	ascaris_eg	trichuris_smansoni	hookworm	ascaris_in	trichuris_lanysth_in	smansoni	dewormed
70107028	TEREBEN	1190	BARINGO	701	9	F	1	0	0	0	0	0	0	0	0
70105018	NGINYAN	1203	BARINGO	701	12	F	1	0	0	0	0	0	0	0	0
70105055	NGINYAN	1203	BARINGO	701	11	F	1	0	0	0	0	0	0	0	0
70107038	TEREBEN	1190	BARINGO	701	10	F	1	0	0	0	0	0	0	0	1
70105029	NGINYAN	1203	BARINGO	701	7	F	1	0	0	0	0	0	0	0	0
70105085	NGINYAN	1203	BARINGO	701	14	F	1	0	0	0	0	0	0	0	0
70105010	NGINYAN	1203	BARINGO	701	7	M	1	0	0	0	0	0	0	0	0
70105062	NGINYAN	1203	BARINGO	701	10	F	1	0	0	0	0	0	0	0	0
70107009	TEREBEN	1190	BARINGO	701	8	F	1	0	0	0	0	0	0	0	0
70105107	NGINYAN	1203	BARINGO	701	12	M	1	0	0	0	0	0	0	0	0
70105020	NGINYAN	1203	BARINGO	701	10	M	1	0	0	0	0	0	0	0	0
70107051	TEREBEN	1190	BARINGO	701	10	M	1	0	0	0	0	0	0	0	0
70107019	TEREBEN	1190	BARINGO	701	9	M	1	0	0	0	0	0	0	0	0
70107032	TEREBEN	1190	BARINGO	701	9	M	1	0	0	0	0	0	0	0	0
70107003	TEREBEN	1190	BARINGO	701	6	F	1	0	0	0	0	0	0	0	0
70107036	TEREBEN	1190	BARINGO	701	9	M	1	0	0	0	0	0	0	0	0
70107050	TEREBEN	1190	BARINGO	701	10	M	1	0	0	0	0	0	0	0	0

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Cut Copy Format Painter Clipboard Font Alignment Number Styles

G1 sex

child_id	school	school_id	district	district_id	age	sex	stool	hookworm	ascaris_eg	trichuris_smansoni	hookworm	ascaris_in	trichuris_lanysth_in	smansoni	dewormed
70107028	TEREBEN	1190	BARINGO	701	9	F	1	0	0	0	0	0	0	0	0
70105018	NGINYAN	1203	BARINGO	701	12	F	1	0	0	0	0	0	0	0	0
70105055	NGINYAN	1203	BARINGO	701	11	F	1	0	0	0	0	0	0	0	0
70107038	TEREBEN	1190	BARINGO	701	10	F	1	0	0	0	0	0	0	0	1
70105029	NGINYAN	1203	BARINGO	701	7	F	1	0	0	0	0	0	0	0	0
70105085	NGINYAN	1203	BARINGO	701	14	F	1	0	0	0	0	0	0	0	0
70105010	NGINYAN	1203	BARINGO	701	7	M	1	0	0	0	0	0	0	0	0
70105062	NGINYAN	1203	BARINGO	701	10	F	1	0	0	0	0	0	0	0	0
70107009	TEREBEN	1190	BARINGO	701	8	F	1	0	0	0	0	0	0	0	0
70105107	NGINYAN	1203	BARINGO	701	12	M	1	0	0	0	0	0	0	0	0
70105020	NGINYAN	1203	BARINGO	701	10	M	1	0	0	0	0	0	0	0	0
70107051	TEREBEN	1190	BARINGO	701	10	M	1	0	0	0	0	0	0	0	0
70107019	TEREBEN	1190	BARINGO	701	9	M	1	0	0	0	0	0	0	0	0
70107032	TEREBEN	1190	BARINGO	701	9	M	1	0	0	0	0	0	0	0	0
70107003	TEREBEN	1190	BARINGO	701	6	F	1	0	0	0	0	0	0	0	0

Verifying the plausibility of data

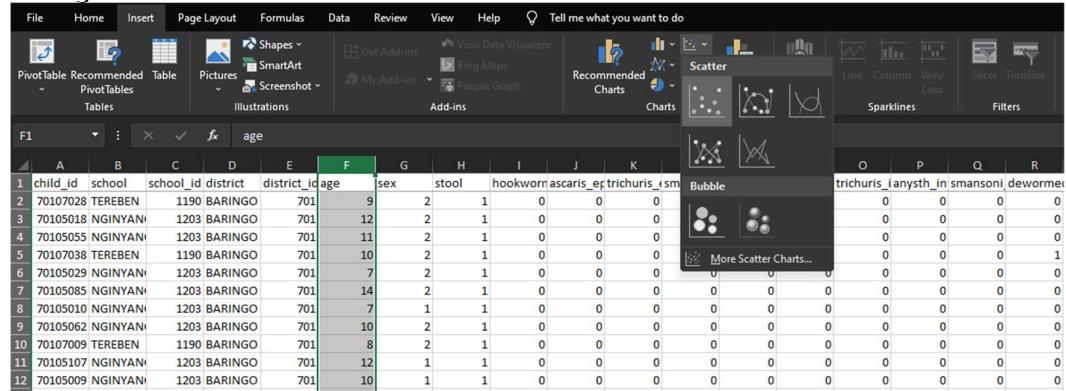
In this step, we perform two basic operations

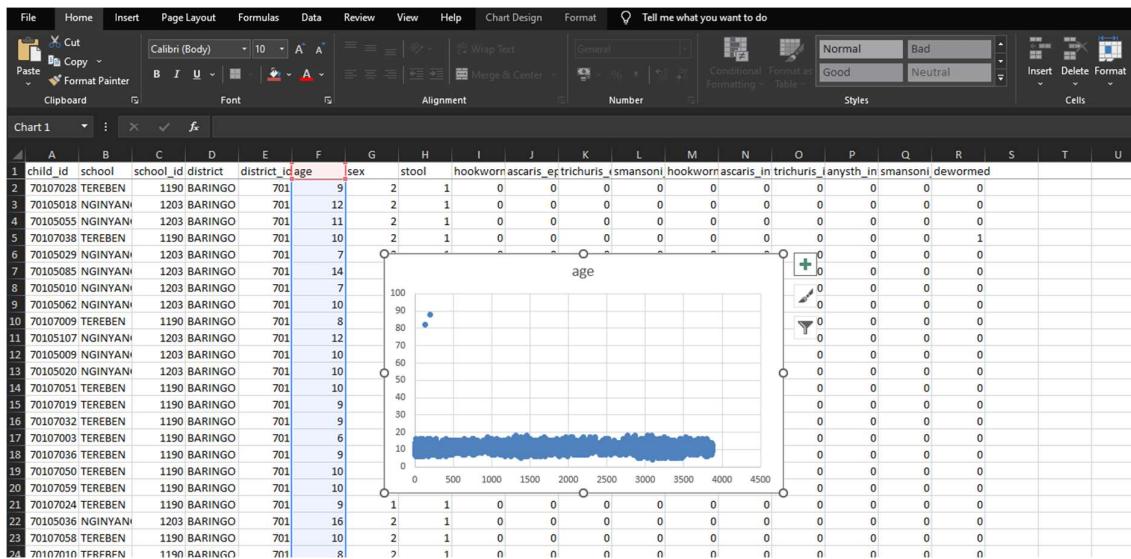
- A. Coding of variables
- B. Using a filter to detect outliers

A. Coding of variables

Select the age column in the existing worksheet.

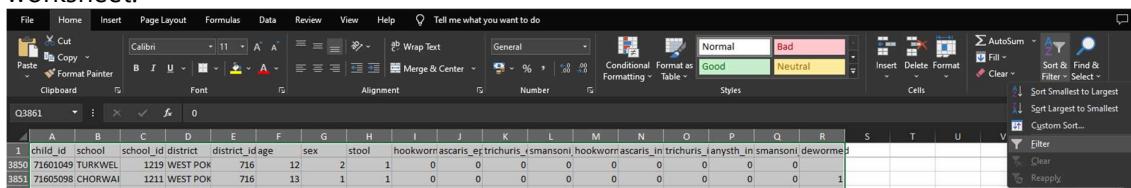
Now go to Insert tab and select Scatter. You will set chart as shown below



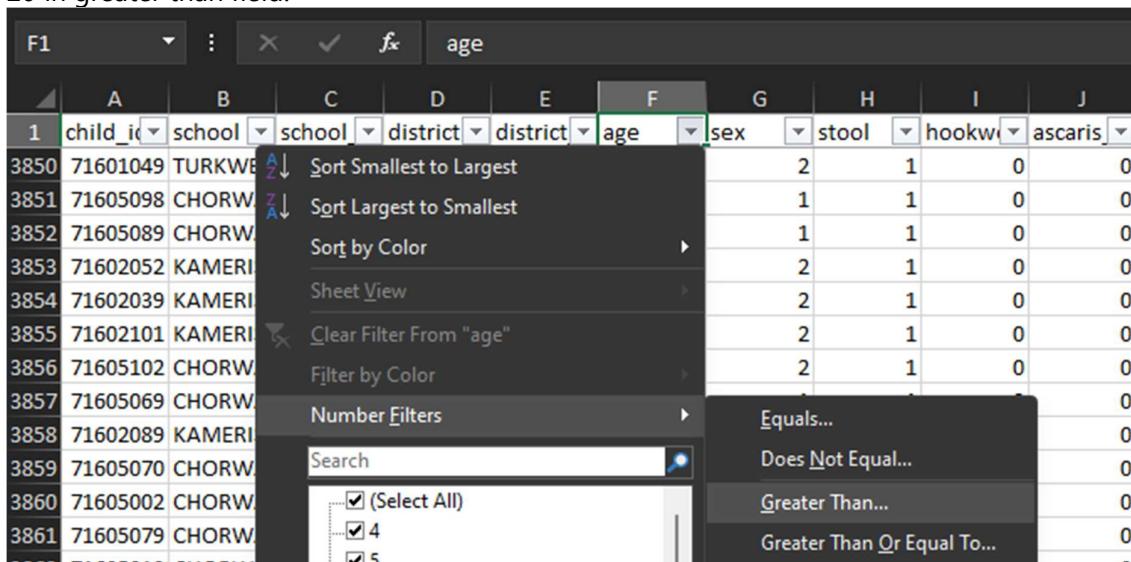


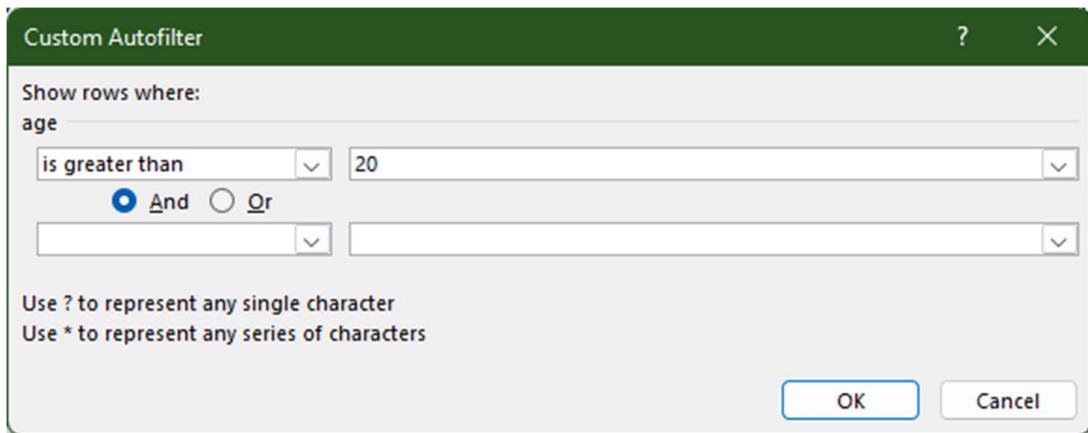
B. Using a filter to detect outliers.

Go to the Home Tab>Sort and Filter>Filter. Click and apply the filter to all the columns of the worksheet.



Now click on age filter and click on Number Filter> Greater Than option and type the value 20 in greater than field.





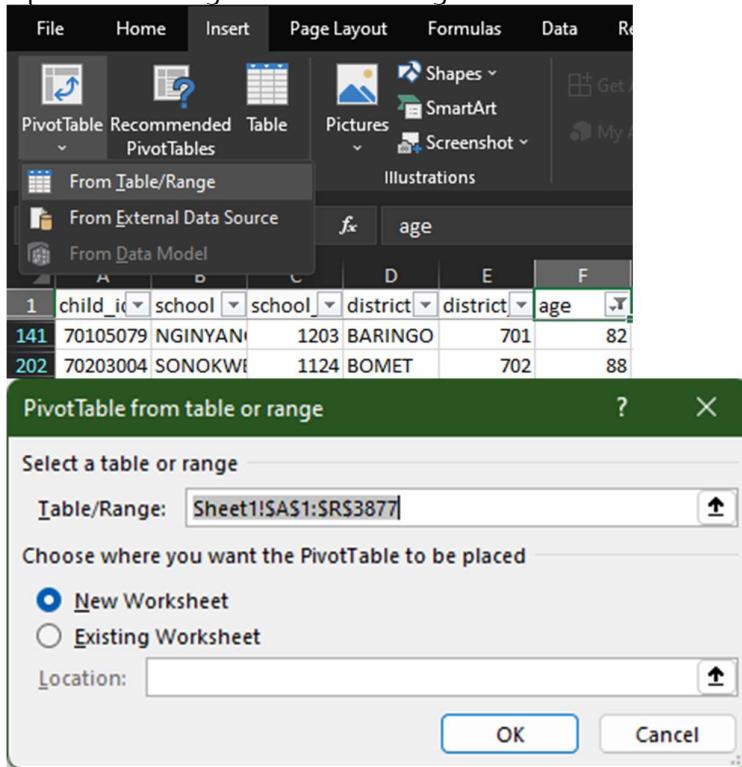
Logical Data Checks

In this step, we perform two basic operations

- A. Cross Tabulations
- B. Formulas

A. Cross tabulations

Open the existing worksheet. Now go to Insert Tab and select Pivot table function.



An empty table is inserted in a new sheet and a window will open on the right hand side named PIVOT TABLE FIELD LIST.

From the PivotTable Field List, drag the "stool" item and drop it into the "Row Label" field as shown above.

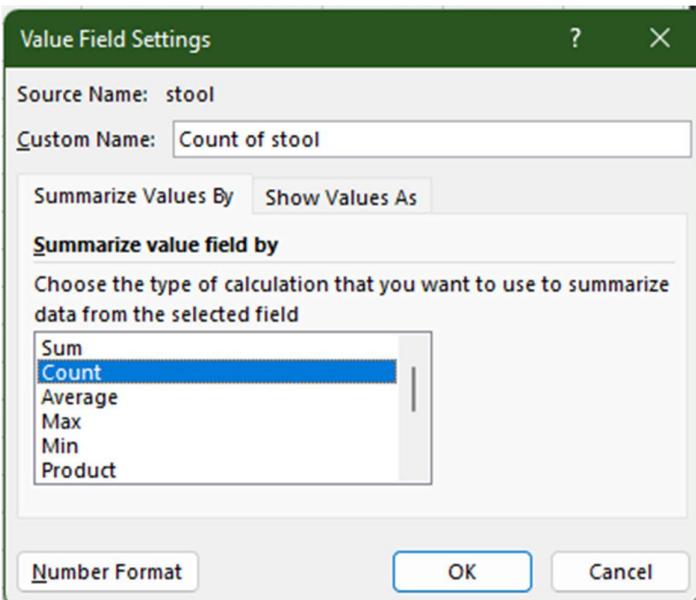
The screenshot shows the 'PivotTable Fields' dialog box. At the top, there's a search bar and a settings gear icon. Below the search bar is a list of fields: school_id, district, district_id, age, sex, stool, hookworm_epg, ascaris_epg, and tricouris_epg. The 'stool' field has a checked checkbox. A red arrow points from the 'stool' entry in the list down to the 'Rows' section below. The 'Rows' section contains a dropdown menu with 'stool' selected. To the right of the 'Rows' section are two other sections: 'Filters' and 'Columns'. At the bottom of the dialog box are 'Defer Layout Update' and 'Update' buttons.

Similarly, Click on anysth_inf and draw it into the "Column labels".

To include the count of Stool drag it in values box and in the table you might need to change the value field settings to count.

Click on the combo box Sum of stools and Click on Value Field Settings.

The screenshot shows the 'Value Field Settings...' dialog box. It has a 'Remove Field' button at the top left. The main area contains a dropdown menu with 'Sum of stool' selected. The background of the dialog box is dark grey.



Change the value in Summarize value filed by to Count and click OK. Table is updated with count values as shown below

	A	B	C	D
1	Drop Report Filter Fields Here			
2				
3	Count of stool	anysth_inf		
4	stool	0	1	Grand Total
5		0	2	
6		1	3410	464
7	Grand Total	3412	464	3876
8				

Formulas

Open the existing worksheet

Create a new column with the variable called check

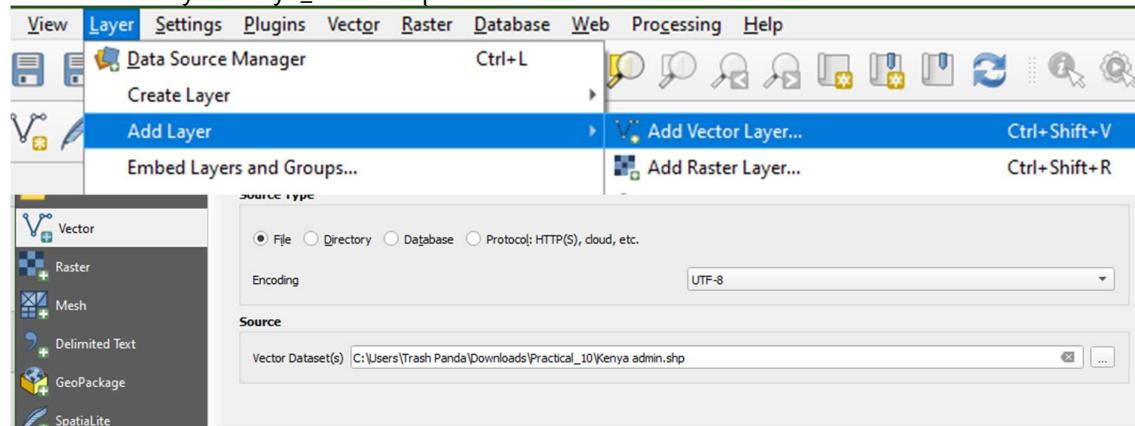
Type the following formula in S2 column of worksheet =IF(AND(H2=0, NOT(P2="")),1,0)

Now copy the formula to all other cells.

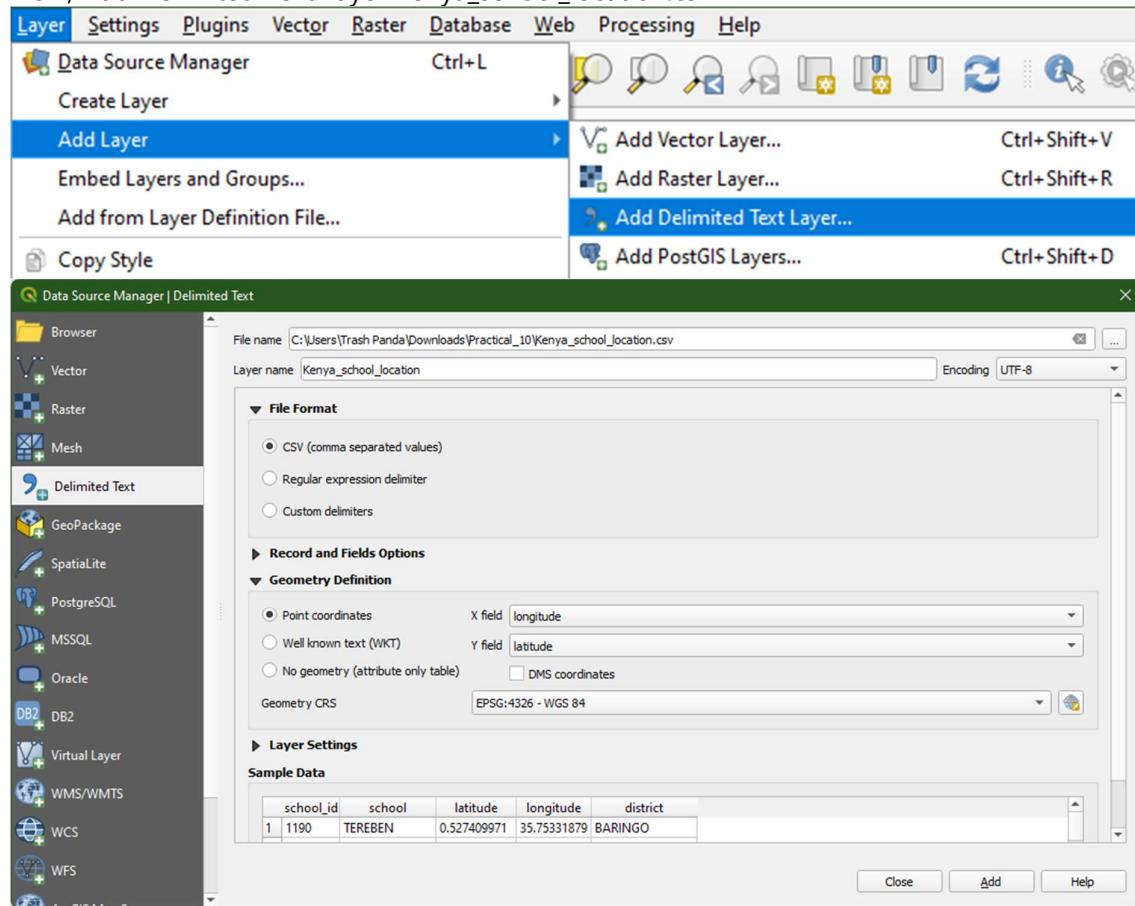
Verifying the coordinates of mapping data

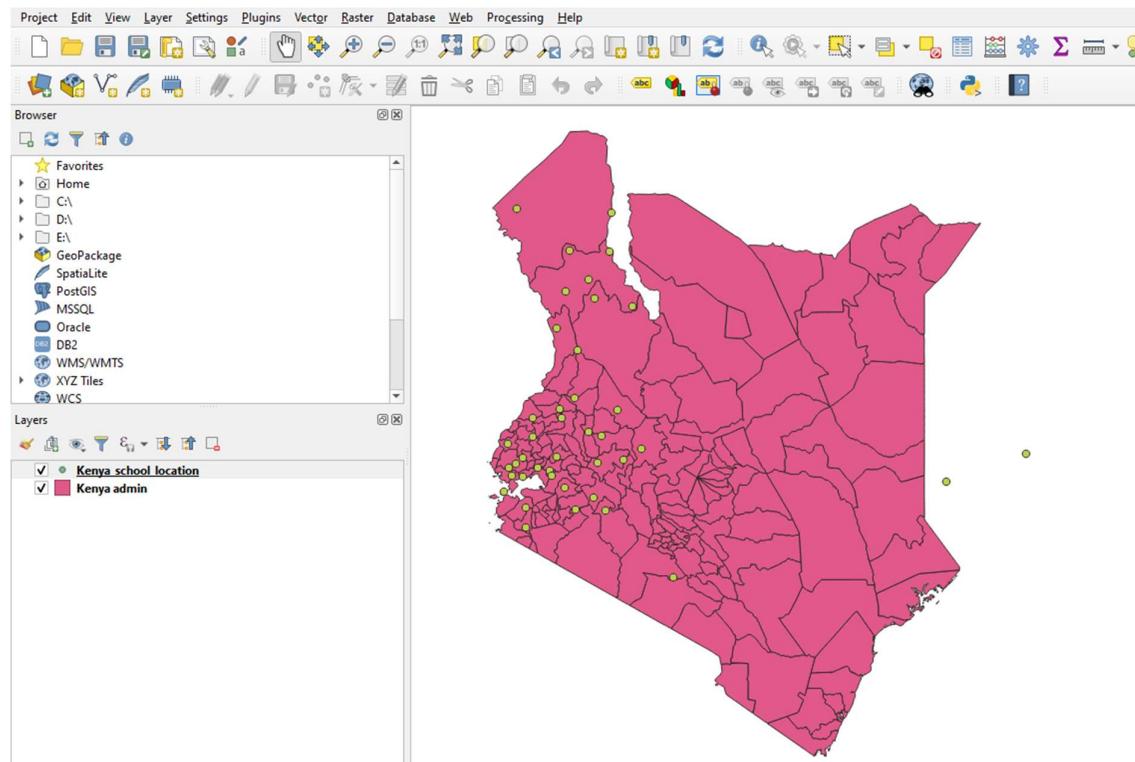
Create a New Project in QGIS.

Add Vector Layer Kenya_admin.shp



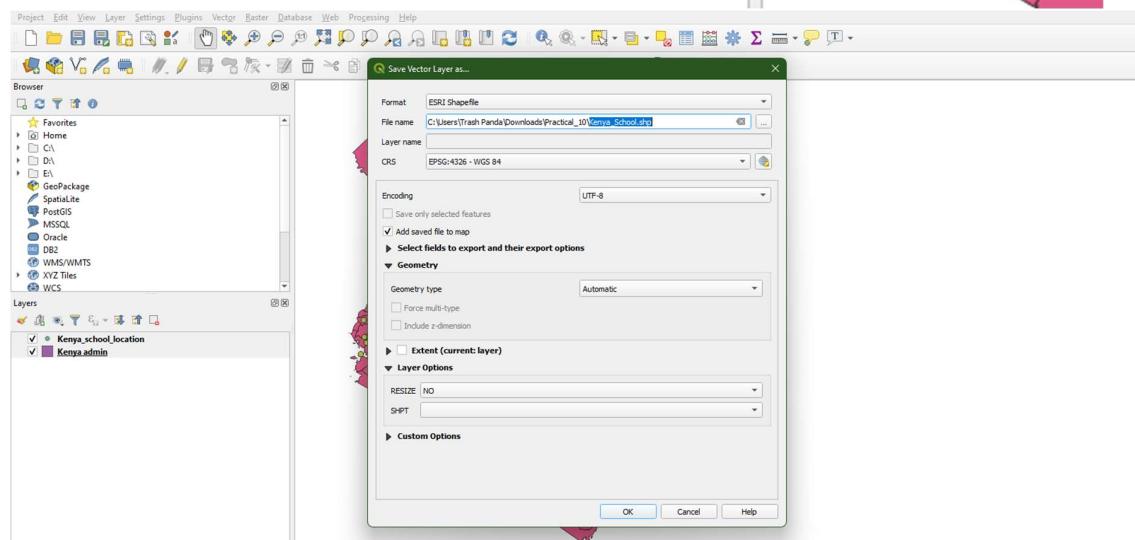
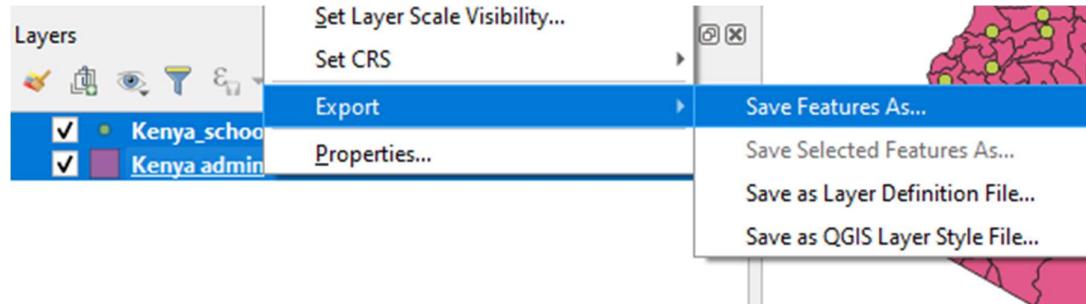
Now, Add Delimited Text Layer Kenya_school_location.csv





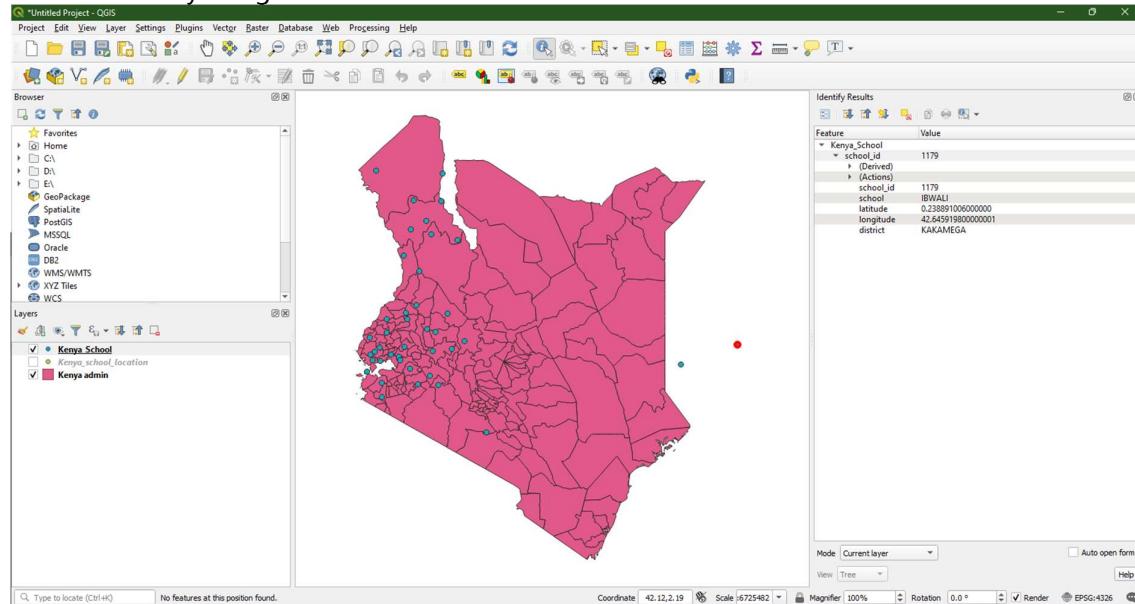
As you can see 2 points are not on the map.

To examine this, we need to save these layers as a Shapefile, to do that, select both the layers Kenya_school_location and Kenya admin, then right click on them and choose Save Features As...



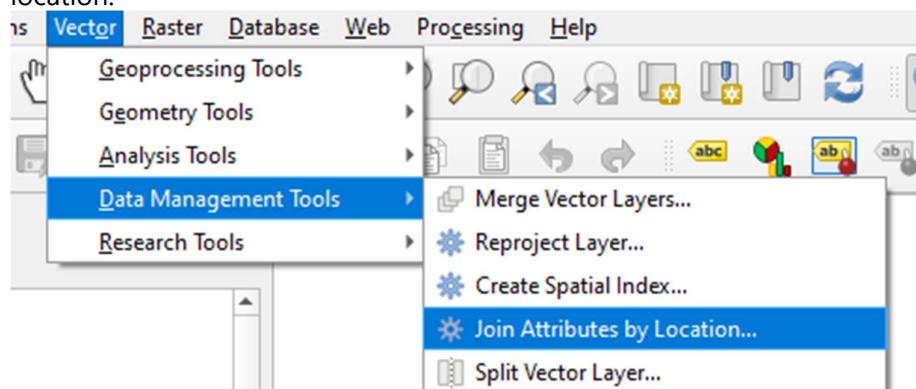
After this is done you can uncheck the Kenya_school_location in the layers section.

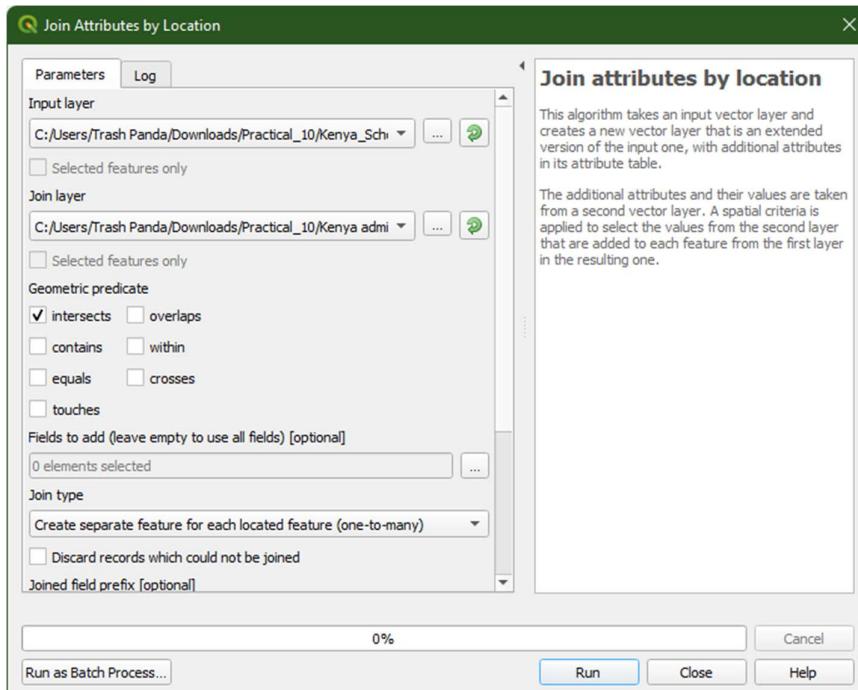
Let's try and get the details of these 2 points that are not on the map, select the Kenya_schools layer, click on the Identify Features Tool button and then click on the points outside of Kenya to get their details.



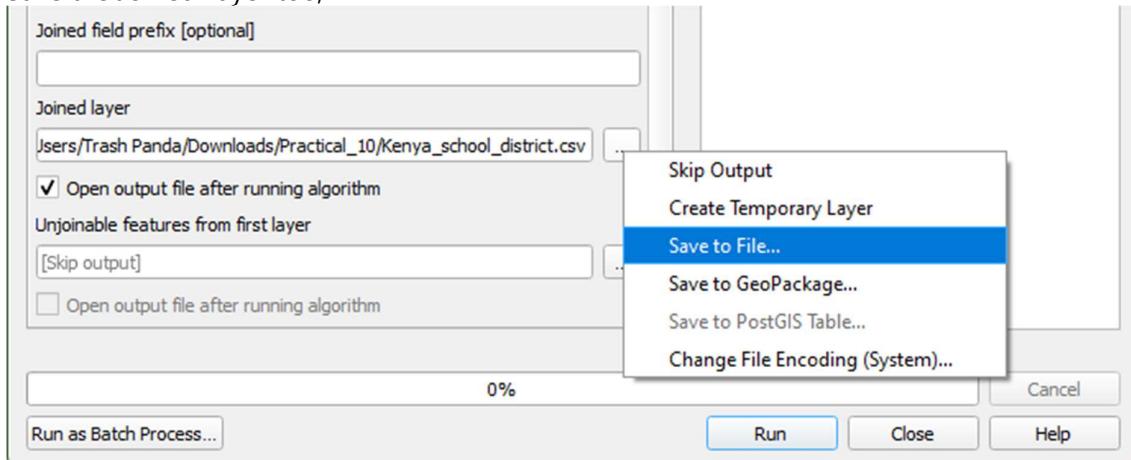
Now we want to add the district information to the map. Therefore, we will join information based on the geographical localization.

In the menu click on Vector, then Data Management Tools, then select Join attributes by location.





Save the Joined Layer too,



Now Click run.

After it is done, Open the saved Kenya_school_district.csv.

school_id	school	latitude	longitude	district	LVID	ISO_CTRY	STATECO	NAME	LVA	UPPERLVN/UPPERLVN/A
1	1163 USULA	0.0898	34.3679	SIAYA	601 KE	KE	SIAYA		2	6 NYANZA
2	1156 MUR MAL	0.00975	34.26051	SIAYA	601 KE	KE	SIAYA		2	6 NYANZA
3	1148 SIAGE	-0.1139	34.3082	BONDO	602 KE	KE	BONDO		2	6 NYANZA
4	1150 KARUNGA	-0.04272	34.91285	KISUMU	604 KE	KE	KISUMU E		2	6 NYANZA
5	1145 ALWALA	-0.1388	34.48393	KISUMU	605 KE	KE	KISUMU V		2	6 NYANZA
6	1230 NATOLE	3.51651	35.87964	TURKANA	701 KE	KE	TURKANA		2	7 RIFT VALLEY
7	1224 KANGIRIS	2.61852	36.2592	TURKANA	701 KE	KE	TURKANA		2	7 RIFT VALLEY
8	1226 LOCHOR E	2.86972	35.17767	TURKANA	701 KE	KE	TURKANA		2	7 RIFT VALLEY
9	1227 NAOTIN	3.0574	35.54712	TURKANA	701 KE	KE	TURKANA		2	7 RIFT VALLEY
10	1231 MAKUTAN	3.53587	35.24499	TURKANA	701 KE	KE	TURKANA		2	7 RIFT VALLEY
11	1233 NARIOKO	4.44145	35.91398	TURKANA	702 KE	KE	TURKANA		2	7 RIFT VALLEY
12	1234 AIC COPID	4.2063	34.39046	TURKANA	702 KE	KE	TURKANA		2	7 RIFT VALLEY
13	1225 LOCHOR E	2.7578	35.64711	TURKANA	703 KE	KE	TURKANA		2	7 RIFT VALLEY
14	1219 TURKVEL	1.9191	35.37047	WEST POK	704 KE	KE	WEST POK		2	7 RIFT VALLEY
15	1222 KAMERIS	2.2752	35.02739	WEST POK	705 KE	KE	POKOT N		2	7 RIFT VALLEY
16	1211 CHORWAI	1.3398	35.31615	WEST POK	706 KE	KE	POKOT CE		2	7 RIFT VALLEY
17	1204 TOROR	0.956572	35.07867	TRANS NZ	710 KE	KE	TRANS NZ		2	7 RIFT VALLEY
18	1190 TEREBEN	0.52741	35.75332	BARINGO	713 KE	KE	BARINGO		2	7 RIFT VALLEY
19	1192 NGINYANI	0.9392	36.02051	NABINGO	715 KE	KE	EAST POK		2	7 RIFT VALLEY
20	1169 NGENDAL	0.141355	36.11529	KOIBATEK	716 KE	KE	KOIBATEK		2	7 RIFT VALLEY
21	1164 KAPCHOLI	0.09255	35.68682	KOIBATEK	716 KE	KE	KOIBATEK		2	7 RIFT VALLEY
22	1193 SONGETO	0.5923	35.55437	KEIYO	721 KE	KE	KEIYO		2	7 RIFT VALLEY
23	1178 KIMONDII	0.19257	35.03781	NANDI	723 KE	KE	NANDI CE		2	7 RIFT VALLEY
24	1180 GATAMI	0.313997	36.41159	LAIKIPIA	729 KE	KE	LAIKIPIA I		2	7 RIFT VALLEY
25	1133 TETA	-0.46973	35.62071	NAKURU	733 KE	KE	MOLO		2	7 RIFT VALLEY
26	1123 ILPOLTON	-0.68108	35.81882	NAROK	734 KE	KE	NAROK N		2	7 RIFT VALLEY
27	1141 KABOKYE	-0.31438	35.16176	KERICHO	739 KE	KE	KERICHO		2	7 RIFT VALLEY
28	1124 SONOKWI	-0.66587	35.33916	BOMET	743 KE	KE	BOMET		2	7 RIFT VALLEY

Now, open Kenya_school_location.csv and make the following changes, set:

IBWALI: Longitude 34.6459198

SIWOT: Longitude 35.35437012

school_id	school	latitude	longitude	district
1	1190 TEREBEN	0.52741	35.75331879	BARINGO
2	1203 NGINYANI	0.9392	36.02051163	BARINGO
3	1124 SONOKWI	-0.66587	35.33916092	BOMET
4	1148 SIAGE	-0.1139	34.30820084	BONDO
5	1189 MABUSSI	0.51736	34.63671112	BUNGOMA
6	1182 INDOLI	0.390609	34.24145889	BUSIA
7	1175 SHIATSAL	0.17746	34.47684097	BUTERE/MUMIAS
8	1125 KOGWANI	-0.63851	34.52355957	HOMA BAY
9	1091 ENKIRGIRI	-1.76414	36.92058945	KAJADAO
10	1179 IBWALI	0.238891	34.6459198	KAKAMEGA
11	1193 SONGETO	0.5923	35.55437088	KEIYO
12	1141 KABOKYE	-0.31438	35.16175842	KERICHO
13	1143 SIWOT	-0.21165	35.35437012	KERICHO
14	1145 ALWALA	-0.13884	34.48392868	KISUMU
15	1150 KARUNGA	-0.04272	34.91284943	KISUMU
16	1169 NGENDAL	0.141355	36.1152916	KOIBATEK
17	1164 KAPCHOLI	0.09255	35.68682098	KOIBATEK

Save above modified file as Kenya_school_location_2.csv

Preparing data for mapping

Open Kenya_epidemiological_data_2.xls, select the entire sheet, go to Insert tab to create new Pivot Table.

Tick New Worksheet to tell Excel that you want to place the table in a new sheet.

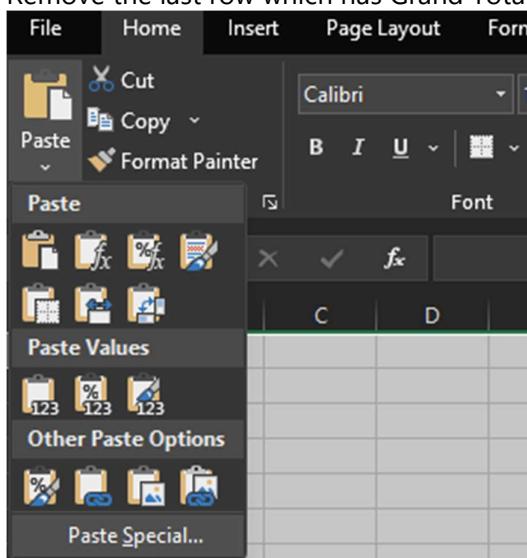
	child_id	school	district	age	sex	stool	hookwi	ascaris	trichuri	smansc	hookwi	ascaris	trichuri	anysth	smansc	dewori	check
2	70107045	TEREBEN	1190	BARINGO	701	11	2	0	0	0	0	0	0	0	0	0	1
3	61110005	KARUNGA	1150	KISUMU	611	7	1	0	0	0	0	0	0	0	0	0	0
4	70107028	TEREBEN	1190	BARINGO	701	9	2	1	0	0	0	0	0	0	0	0	0
5	70105018	NGINYANI	1203	BARINGO	701	12	2	1	0	0	0	0	0	0	0	0	0
6	70105055	NGINYANI	1203	BARINGO	701	11	2	1	0	0	0	0	0	0	0	0	0
7	70107038	TEREBEN	1190	BARINGO	701	10	2	1	0	0	0	0	0	0	0	0	1
8	70105029	NGINYANI	1203	BARINGO	701	7	2	1	0	0	0	0	0	0	0	0	0
9	70105085	NGINYANI	1203	BARINGO	701	14	2	1	0	0	0	0	0	0	0	0	0
10	70105010	NGINYANI	1203	BARINGO	701	7	1	1	0	0	0	0	0	0	0	0	0
11	70105062	NGINYANI	1203	BARINGO	701	10	2	1	0	0	0	0	0	0	0	0	0
12	70107009	TEREBEN	1190	BARINGO	701	8	2	1	0	0	0	0	0	0	0	0	0
13	70105107	NGINYANI	1203	BARINGO	701	12	1	1	0	0	0	0	0	0	0	0	0
14	70105009	NGINYANI	1203	BARINGO	701	10	1	1	0	0	0	0	0	0	0	0	0
15	70105020	NGINYANI	1203	BARINGO	701	10	2	1	0	0	0	0	0	0	0	0	0
16	70107051	TEREBEN	1190	BARINGO	701	10	1	1	0	0	0	0	0	0	0	0	0
17	70107019	TEREBEN	1190	BARINGO	701	9	1	1	0	0	0	0	0	0	0	0	0
18	70107032	TEREBEN	1190	BARINGO	701	9	1	1	0	0	0	0	0	0	0	0	0
19	70107003	TEREBEN	1190	BARINGO	701	6	2	1	0	0	0	0	0	0	0	0	0
20	70107036	TEREBEN	1190	BARINGO	701	9	1	1	0	0	0	0	0	0	0	0	0
21	70107050	TEREBEN	1190	BARINGO	701	10	1	1	0	0	0	0	0	0	0	0	0
22	70107059	TEREBEN	1190	BARINGO	701	10	2	1	0	0	0	0	0	0	0	0	0
23	70107024	TEREBEN	1190	BARINGO	701	9	1	1	0	0	0	0	0	0	0	0	0
24	70105036	NGINYANI	1203	BARINGO	701	16	2	1	0	0	0	0	0	0	0	0	0

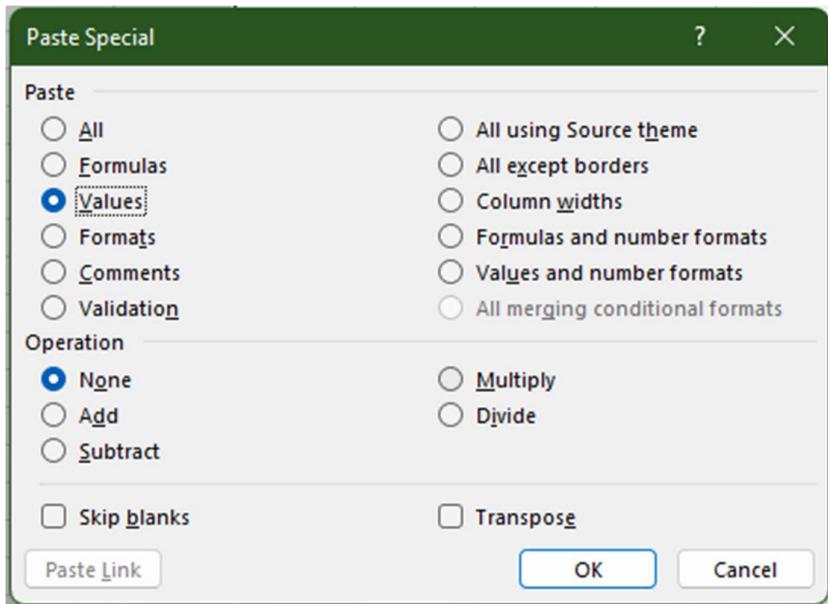
Now click on school_id to drag and drop it in the "Row labels" field at the bottom.

Add district_id to "Σ Values" and click on it, a drop down list will open, click on Value Field Settings. Choose Average as type of calculation, because all children in the same school will have same district_id.

drag and drop child_id into the "Σ Values" field, click on Value Field Settings and choose Count to summarize the results. So, we come to know how many children per school are infected. Similarly, Drag and drop anysth_inf into the field, click on Value Field Settings and choose Sum. As, infected is 1 and not affected is 0, the sum will give us the total number of infected children.

Now copy the table, open a new Excel file and paste the values into the new spreadsheet. Therefore, click Paste in the Home tab, then choose Paste special and paste only the values. Remove the last row which has Grand Total and other values.





Remove the last row which has Grand Total and other values.

Save this new data table as Kenya_school_STH_surveys.xlsx

school_id	Average	Count	o	Sum	of	Sum	of	Sum	of	Sum	of	smanson_inf
1091	704	105	0	0	0	0	0	0	0	0	0	0
1119	613	109	22	18	45	21	8					
1123	712	105	3	0	3	2	0					
1124	702	110	17	9	34	27	0					
1125	610	107	34	8	39	11	1					
1133	710	109	0	1	1	0	0					
1139	616	106	1	3	6	2	68					
1141	706	108	1	2	13	13	0					
1143	706	109	0	0	0	0	0					
1145	611	109	3	3	6	1	8					
1146	614	66	10	2	14	3	2					
1148	609	106	15	4	23	7	10					
1150	611	63	5	3	9	2	7					
1155	808	110	1	3	9	8	0					
1156	615	69	15	8	30	17	2					
1163	615	110	5	10	25	19	3					
1164	707	108	0	0	1	1	0					
1169	707	108	3	1	4	0	0					
1175	803	108	15	28	37	19	0					
1178	711	108	1	1	20	19	0					
1179	804	106	6	3	22	18	0					
1180	708	23	0	0	0	0	0					
1182	802	106	30	1	32	2	0					
1189	801	70	22	0	22	0	0					
1190	701	71	0	2	2	0	0					
1193	705	109	0	0	0	0	0					
1198	805	106	2	1	6	3	0					
1199	806	106	26	2	29	9	0					

Now we need to know the total count of infected children, that is, the prevalence of STH.

So, we create a new column: sth_prev which we are going to be dividing the number of infected children by the number of children and multiplying by hundred to obtain a value in percent (%).

Formula: =F2/C2*100

SUM	:	X	✓	f(x)	=F2/C2*100							
	A	B	C	D	E	F	G	H	I			
1	school_id	Average of distr	Count of child_id	Sum of hookworm	Sum of trichuri	Sum of anysth_inf	Sum of ascaris_inf	Sum of smanson	sth_prev			
2	1091	704	105	0	0	0	0	0	0=F2/C2*100			
3	1119	613	109	22	18	45	21	8				

Additionally, you could calculate the 95% confidence interval (CI) of your prevalence.

You will have to create a new column called "Cllow" (lower confidence limit) and

"Clup"(upper confidence limit)

You can calculate the limits by typing the formulas as follows and then copying to the other cells:

For Cllow: =I2 - (1.96*(SQRT((I2*(100-I2)/C2))))

For Clup: =I2 + (1.96*(SQRT((I2*(100-I2)/C2))))

I2 = sth_prev and C2 = Count of child_id.

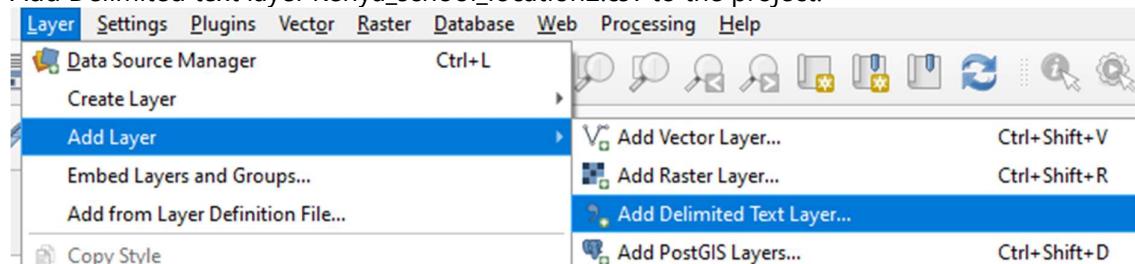
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	school_id	Average of dist	Count of child_id	Sum of hookworm	Sum of trichuris	Sum of anysth_inf	Sum of ascaris_inf	Sum of smanson_sth_prev	Cllow	Clup					
2	1091	704	105	0	0	0	0	0	32.04141186	50.52739548					
3	1119	613	109	22	18	45	21	8	41.28440367	6.04378163					
4	1123	712	105	3	0	3	2	0	2.857142857	-0.329495915					
5	1124	702	110	17	9	34	27	0	30.90909091	22.27307769	39.54510413				
6	1125	610	107	34	8	39	11	1	36.44859813	27.32917925	45.56801701				
7	1133	710	109	0	1	1	0	0	0.917431193	-0.872466475	2.707228861				
8	1139	616	106	1	3	6	2	68	5.660377358	1.261185543	10.05956917				
9	1141	706	108	1	2	13	13	0	12.03703704	5.900068215	18.17400586				
10	1143	706	109	0	0	0	0	0	0	0	0	0	0		
11	1145	611	109	3	3	6	1	8	5.504587156	1.222943085	9.786231227				
12	1146	614	66	10	2	14	3	2	21.21212121	11.34919333	31.0750491				
13	1148	609	106	15	4	23	7	10	21.69811321	13.85117491	29.54505151				
14	1150	611	63	5	3	9	2	7	14.28571429	5.644726688	22.92670188				
15	1155	808	110	1	3	9	8	0	8.181818182	3.059707188	13.30392018				
16	1156	615	69	15	8	30	17	2	43.47826087	31.78123727	55.17528447				
17	1163	615	110	5	10	25	19	3	22.772727273	14.89574462	30.55880083				
18	1164	707	108	0	0	1	1	0	0.925925926	-0.880467429	2.732319281				
19	1169	707	108	3	1	4	0	0	3.70370370	0.141923601	7.265483806				
20	1175	803	108	15	28	37	19	0	34.25925926	25.30869377	43.20982475				
21	1178	711	108	1	1	20	19	0	18.51851852	11.19233223	25.84468481				
22	1179	804	106	6	3	22	18	0	20.75471698	13.03416668	28.47526728				
23	1180	708	23	0	0	0	0	0	0	0	0	0	0		
24	1182	802	106	30	1	32	2	0	30.18867925	21.44914834	38.92821015				
25	1189	801	70	22	0	22	0	0	31.42857143	20.55327833	42.30386453				
26	1190	701	71	0	2	2	0	0	2.816901408	-1.0317457	6.665548517				
27	1193	705	109	0	0	0	0	0	0	0	0	0	0		
28	1198	805	106	2	1	6	3	0	5.660377358	1.261185543	10.05956917				
29	1199	806	106	26	2	29	9	0	27.35849057	18.87173615	35.84524498				

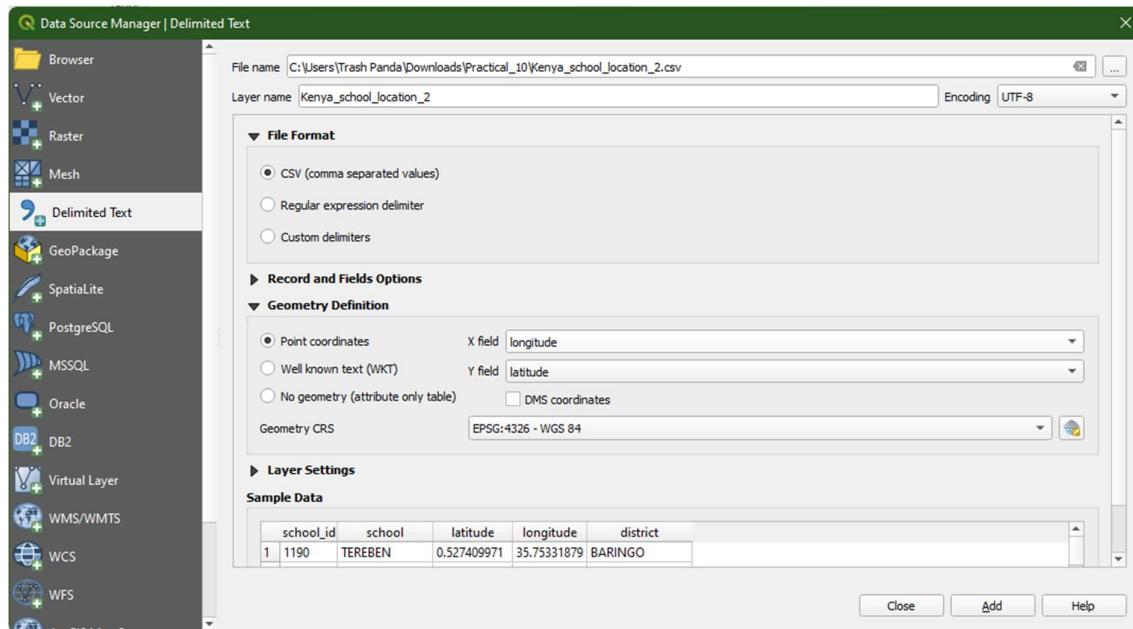
Save this file as Kenya_epidemiological_school.csv

Now that all the datasets are ready, let's add them.

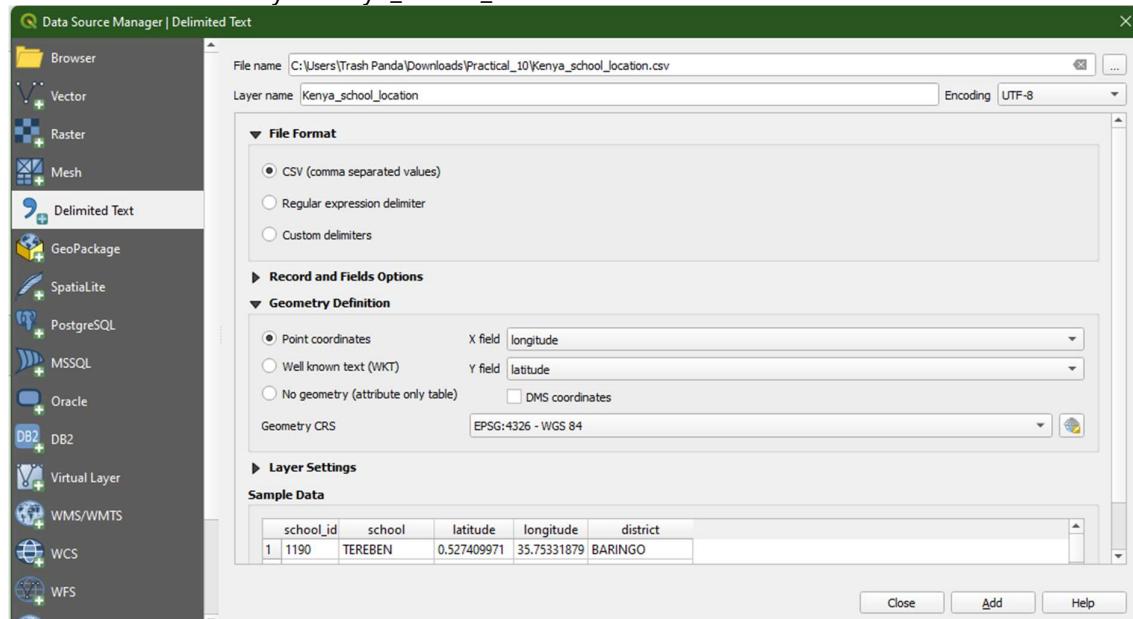
Open a new QGIS project.

Add Delimited text layer Kenya_school_location2.csv to the project.

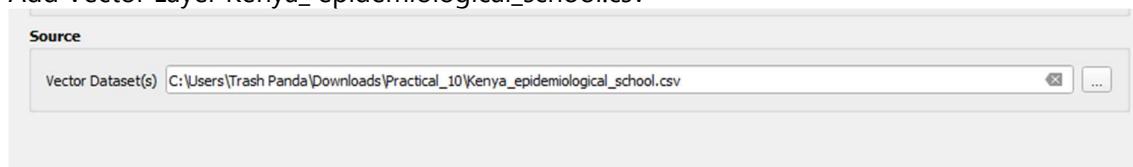




Add Delimited text Layer Kenya_school_location.csv

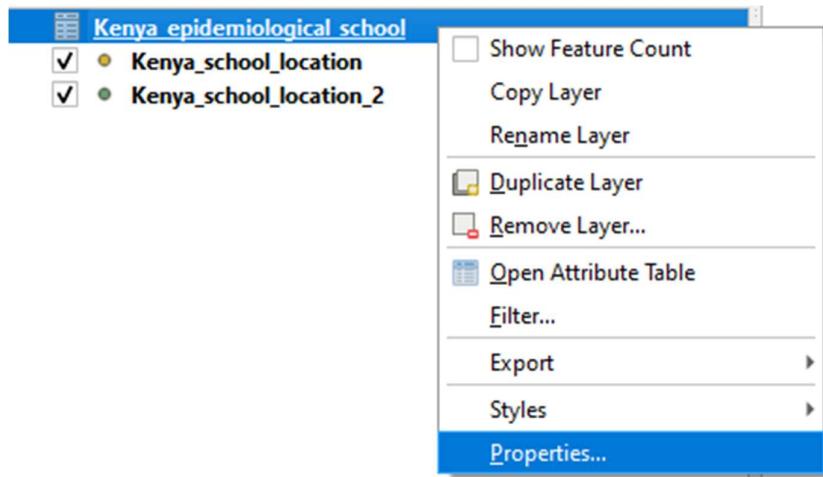


Add Vector Layer Kenya_epidemiological_school.csv

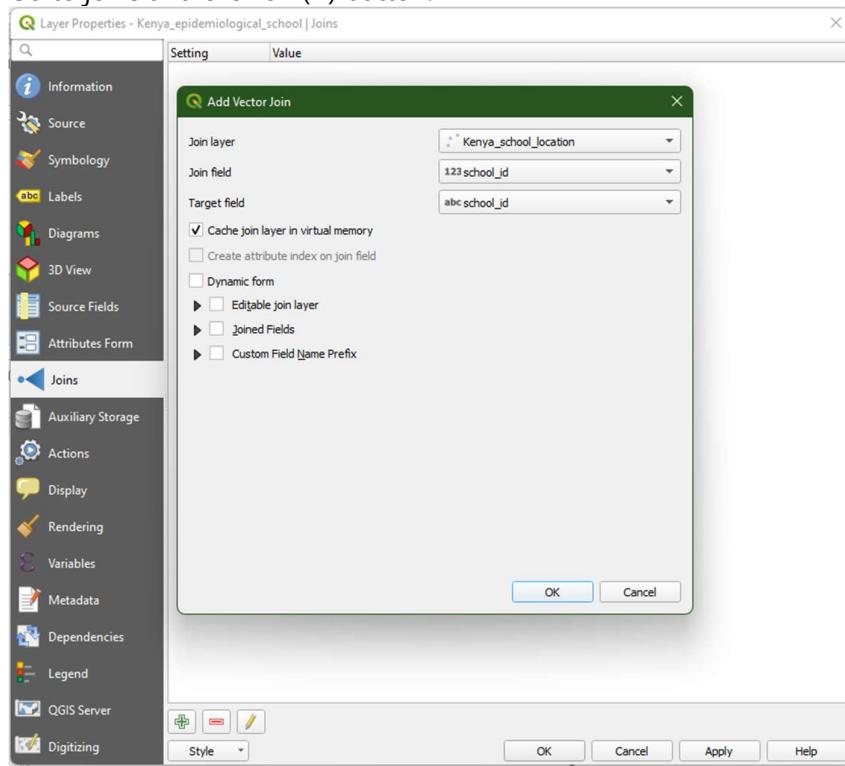


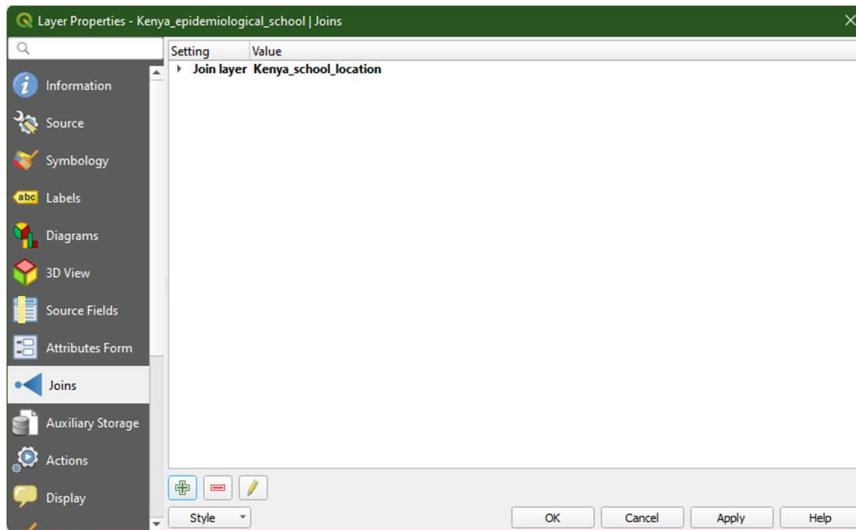
Now Join the data.

Right click on Kenya_epidemiological_school layer and click on Properties.

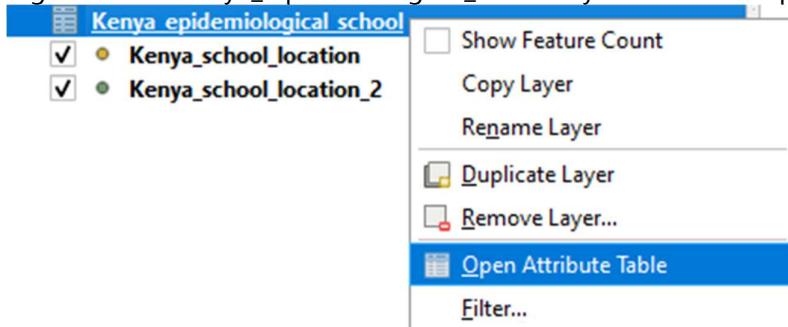


Go to joins and click on (+) button.



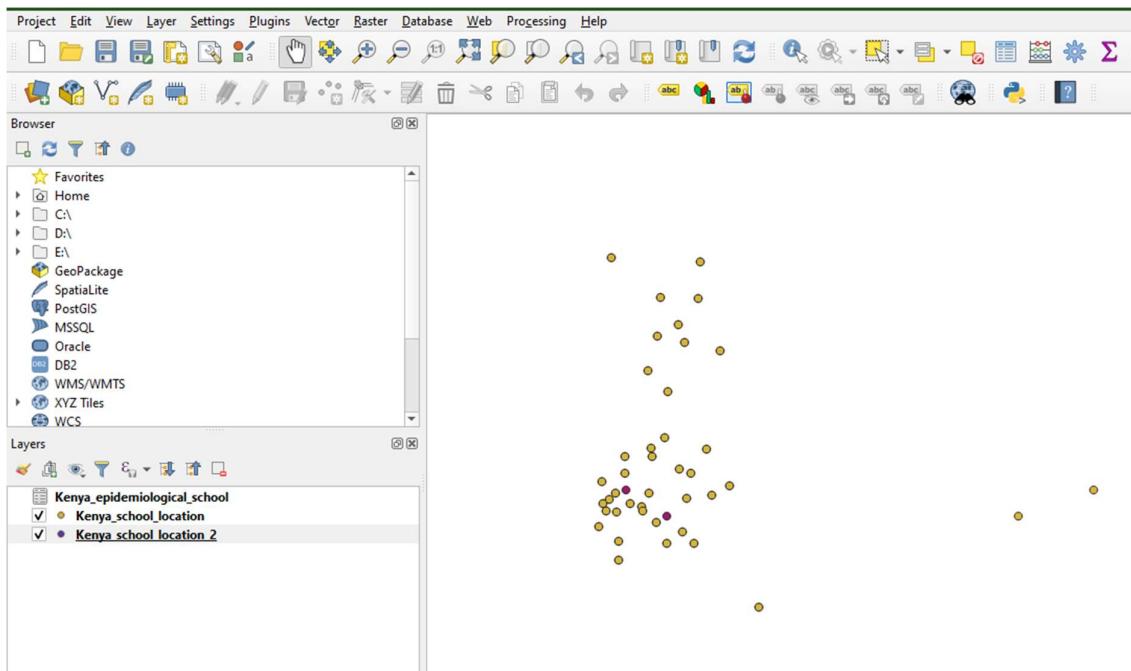


Right click on Kenya_epidemiological_school layer and select Open Attribute Table.



You will see that the epidemiological variables are included.

	school_id	verage of district_id	Count of child_id	im of hookworm	sum of trichuris_in	Sum of anysth_inf	Sum of ascaris_inf	um of smansoni_ir	sth_prev	Cllow	Clup	_school_location_sch	sch
1	1141	706	108	1	2	13	13	0	12.03703704	5.900058215	18.17400586	KABOKYEK	
2	1150	611	63	5	3	9	2	7	14.28571429	5.644726688	22.92670188	KARUNGA	
3	1119	613	109	22	18	45	21	8	41.28440367	32.04141186	50.52739548	ORUBA	
4	1156	615	69	15	8	30	17	2	43.47826087	31.78123727	55.17528447	MUR MALANGA	
5	1155	808	110	1	3	9	8	0	8.181818182	3.059707188	13.30392918	TIGOI	
6	1125	610	107	34	8	39	11	1	36.44859813	27.32917925	45.56801701	KOGWANG	
7	1175	803	108	15	28	37	19	0	34.25925926	25.30869377	43.20982475	SHIATSALA	
8	1124	702	110	17	9	34	27	0	30.90909091	22.27307769	39.54510413	SONOKWEK	
9	1182	802	106	30	1	32	2	0	30.18867925	21.44914834	38.92821015	INDOLI	
10	1189	801	70	22	0	22	0	0	31.42857143	20.55327833	42.30386453	MABUUSI RCEA	
11	1199	806	106	26	2	29	9	0	27.35849057	18.87173615	35.84524498	SACHO RC	
12	1163	615	110	5	10	25	19	3	22.72727273	14.89574462	30.55880083	USULA	
13	1204	714	110	19	0	25	17	0	22.72727273	14.89574462	30.55880083	TOROR	
14	1148	609	106	15	4	23	7	10	21.69811321	13.85117491	29.54505151	SIAGE	
15	1179	804	106	6	3	22	18	0	20.75471699	13.03416668	28.47526728	IBWALI	
16	1146	614	66	10	2	14	3	2	21.21212121	11.34919333	31.0750491	MAREGA	



Now, Finally Select all the layers, right-click on them and in Export select Save Features As.

